Selecting the independent coordinates of manifolds with large aspect ratios

Yu-Chia Chen [1] Marina Meilă [1]

[1]University of Washington {yuchaz, mmp2}@uw.edu

UNIVERSITY of WASHINGTON

Introduction

Problem Manifold learning (ML) algorithms **fail** apparently or suffer from artifacts when data manifold is long and thin, i.e., when it has **aspect ratio** > 2. The problem lies with the selection of (Diffusion Map) eigenvectors, and it is called **Independent Eigen-coordinates Search (IES)** problem.

What we do

- Formulate the problem mathematically, show that a solution exists (for Diffusion Map).
- Introduce a data driven loss $\mathfrak{L}$ and **Independent eigen-coordinates search (IES)** algorithm.
- Results on real and synthetic data, showing the problem is pervasive.
- Limit of $\mathfrak{L}$ for $n \to \infty$.

Motivating example: eigenvalues/functions of $\Delta_{\mathcal{M}}$ on 2D long strip
Measurement of the strip (width, height) = (W, H). Here $\phi_{1,0}, \phi_{0,1}$ should be chosen.

$$\lambda_{k_1,k_2} = \left(\frac{k_1\pi}{W}\right)^2 + \left(\frac{k_2\pi}{H}\right)^2$$

$$\phi_{k_1,k_2}(w,h) = \cos\left(\frac{k_1\pi w}{W}\right)\cos\left(\frac{k_2\pi h}{H}\right)$$

Sorted in ascending order by λ, the first two eigenvalues are $\lambda_{1,0}$ and $\lambda_{2,0}$ if $W/H > 2$, while $\lambda_{0,1}$ is the $\lceil W/H \rceil^{\text{th}}$ eigenvalue (see Figure 1).

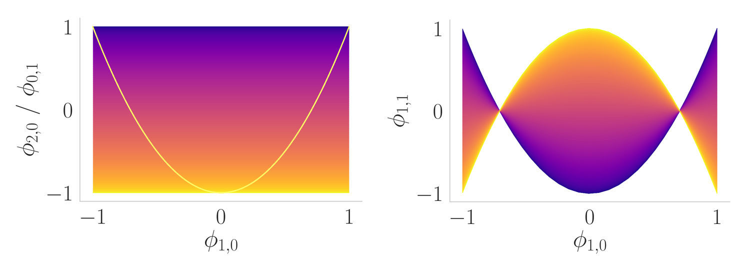


Figure 1. Example of repeated eigen-directions problem.

IES problem [4]

- Defect on a family of *local, spectral* embedding algorithms: LE, DM, LLE, LTSA, HLLE.
- Coordinates of the embedding might not be functionally independent to each other.

Situations when a mapping $\phi(\mathcal{M})$ can fail to be invertible

- *(Global) functional dependency*: $\operatorname{rank} D\phi < d$ on an open subset or all of $\mathcal{M}$ (yellow curve in 1a).
- The *knot*: $\operatorname{rank} D\phi < d$ at an isolated point (Figure 1b).
- The *crossing*: $\phi : \mathcal{M} \to \phi(\mathcal{M})$ is not invertible at $\mathbf{x}$, but $\mathcal{M}$ can be covered with open sets U such that the restriction $\phi : U \to \phi(U)$ has full rank d (Figure 2).
- Combinations of these three exemplary cases can occur.

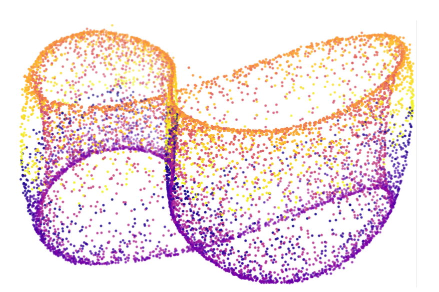


Figure 2. The *crossing*.

Existence of solution [1] However, s, the number of eigenfunctions needed, may exceed the *Whitney embedding dimension* ($\leq 2d$), and that s may depend on injectivity radius, aspect ratio, etc.

Backgrounds

Laplacian eigenmap/diffusion map algorithm [2]

1. Build neighborhood graph $G(V, E)$ with $V = [n]$, $E = \{(i,j) \in V^2 : \|\mathbf{x}_i - \mathbf{x}_j\| \leq 3\varepsilon\}$.
2. Compute kernel matrix $[\mathbf{K}]_{ij} = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/\varepsilon^2)$ and the *renormalized graph Laplacian*
$$\mathbf{L} = \mathbf{I} - \mathbf{W}^{-1}\mathbf{D}^{-1}\mathbf{K}\mathbf{D}^{-1}, \text{ where } \mathbf{D} = \operatorname{diag}(\mathbf{K}\mathbf{1}_n) \text{ and } \mathbf{W} = \operatorname{diag}\left(\mathbf{D}^{-1}\mathbf{K}\mathbf{D}^{-1}\mathbf{1}_n\right)$$
3. An m dimensional embedding is obtained from the 2^{nd} to $m+1^{\text{th}}$ principal eigenvectors of $\mathbf{L}$.
   - We will show that the coordinates chosen by the criteria will **not** give us an optimal embedding.

The pushforward Riemannian metric [6] Associate with $\phi(\mathcal{M})$ a *pushforward Riemannian metric* $g_{*\phi}$ that preserves the geometry of $(\mathcal{M}, g)$. Here $g_{*\phi}$ is defined by

$$\langle \mathbf{u}, \mathbf{v} \rangle_{g_{*\phi}(\mathbf{x})} = \left\langle D\phi^{-1}(\mathbf{x})\mathbf{u}, D\phi^{-1}(\mathbf{x})\mathbf{v} \right\rangle_{g(\mathbf{x})}$$

for all $\mathbf{u}, \mathbf{v} \in \mathcal{T}_{\phi(\mathbf{x})}\phi(\mathcal{M})$

- $\mathcal{T}_{\mathbf{x}}\mathcal{M}, \mathcal{T}_{\phi(\mathbf{x})}\phi(\mathcal{M})$ are tangent subspaces.
- $D\phi^{-1}(\mathbf{x})$ maps vectors from $\mathcal{T}_{\phi(\mathbf{x})}\phi(\mathcal{M})$ to $\mathcal{T}_{\mathbf{x}}\mathcal{M}$.
- $g_{*\phi}(\mathbf{x}_i)$ in local coordinate is a PSD matrix $\mathbf{G}(i)$
$$\langle \mathbf{u}, \mathbf{v} \rangle_{g_{*\phi}(\mathbf{x}_i)} = \mathbf{u}^\top \mathbf{G}(i)\mathbf{v}$$
- Coordinate $\mathbf{U}(i)$ and distortion $\mathbf{\Sigma}(i)$ are from the SVD of co-metric $\mathbf{H}(i) = \text{pseudo\_inv}(\mathbf{G}(i))$.

Algorithm 1: Riemannian metric estimation

RMETRIC($\mathbf{Y}, \mathbf{L}, d$)
for *all* $\mathbf{y}_i \in \mathbf{Y}$, $k = 1 \to m$, $l = 1 \to m$ **do**
  $[\tilde{\mathbf{H}}(i)]_{kl} = \sum_{j \neq i} L_{ij}(y_{jl} - y_{il})(y_{jk} - y_{ik})$
end
for $i = 1 \to n$ **do**
  $\mathbf{U}(i), \mathbf{\Sigma}(i) \leftarrow$ REDUCEDRANKSVD($\tilde{\mathbf{H}}(i), d$)
  $\mathbf{H}(i) = \mathbf{U}(i)\mathbf{\Sigma}(i)\mathbf{U}(i)^\top$
  $\mathbf{G}(i) = \mathbf{U}(i)\mathbf{\Sigma}^{-1}(i)\mathbf{U}(i)^\top$
end
Return: $\mathbf{G}(i), \mathbf{H}(i) \in \mathbb{R}^{m\times m}$, $\mathbf{U}(i) \in \mathbb{R}^{m\times d}$, $\mathbf{\Sigma}(i) \in \mathbb{R}^{d\times d}$, for $i \in [n]$

Related works

1. Analysis on the sufficient conditions for failure (Goldberg et al., 2008 [4]).
2. Functionally independent coordinates (Blau & Michaeli, 2017; Dsilva et al., 2018 [3]).
3. Sequential spectral decomposition (Gerber et al., 2007; Blau & Michaeli, 2017).

Loss function based on volume

Loss function Chosen independent coordinates $S_*(\zeta) = \operatorname{argmax}_{S\subseteq[m];|S|=s;1\in S} \mathfrak{L}(S;\zeta)$

$$\mathfrak{L}(S;\zeta) = \underbrace{\frac{1}{n}\sum_{i=1}^{n}\log\sqrt{\det\left(\mathbf{U}_S(i)^\top\mathbf{U}_S(i)\right)}}_{\mathfrak{R}_1(S)=\frac{1}{n}\sum_{i=1}^n \mathfrak{R}_1(S;i)} - \underbrace{\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{d}\log\|\mathbf{u}_k^S(i)\|_2}_{\mathfrak{R}_2(S)=\frac{1}{n}\sum_{i=1}^n \mathfrak{R}_2(S;i)} - \zeta\sum_{k\in S}\lambda_k \quad (1)$$

1. Start with larger set $[m] = \{1, \cdots, m\}$ of eigenvector of $\mathbf{L}$, find coordinates $S \subseteq [m]$ with $|S| = s$ and force the slowest varying coordinate to always be chosen, i.e., $1 \in S$.
2. Projected volume of a unit parallelogram in $\mathcal{T}_{\phi_S(\mathbf{x}_i)}\phi_S(\mathcal{M})$, $\operatorname{Vol}(i;S) = \frac{\sqrt{\det(\mathbf{U}_S(i)^\top\mathbf{U}_S(i))}}{\prod_{k=1}^d \|\mathbf{u}_k^S(i)\|_2}$
3. ϕ_S is *not* an isometry
   - Remove the local distortions $\mathbf{\Sigma}(i)$ introduced by ϕ from the estimated rank of ϕ at $\mathbf{x}$.
4. Regularization term, consisting of the sum of eigenvalues $\sum_{k\in S}\lambda_k$ of the graph Laplacian $\mathbf{L}$, is added to penalize the high frequency coordinates.

Computation

- Time complexity is $\mathcal{O}(nm^{s+3}) \to$ brute force search for small s.
- $\mathfrak{R}_1, \mathfrak{R}_2$ in (1) are submodular set functions $\to$ optimizing over difference of submodular functions for large s.
- [3] has quadratic dependency on sample size n (see also Figure 3).

Regularization path and choosing ζ
Define the *leave-one-out regret* of point i

$$\mathfrak{D}(S,i) = \mathfrak{R}(S_*^i;[n]\backslash\{i\}) - \mathfrak{R}(S;[n]\backslash\{i\})$$
$$\text{with } S_*^i = \operatorname{argmax}_{S\subseteq[m];|S|=s;1\in S}\mathfrak{R}(S;i)$$

$\mathfrak{D}$ is the gain in $\mathfrak{R}$ if all the other points $[n]\backslash\{i\}$ choose the un-regularized optimal coordinates in terms of point i.

$$\zeta' = \max_{\zeta\geq 0} \text{PERCENTILE}\left(\{\mathfrak{D}(S_*(\zeta),i)\}_{i=1}^n, \alpha\right) \leq 0$$

Algorithm 2: Indep. Eigencoordinates Search

INDEIGENSEARCH($\mathbf{X}, \varepsilon, d, s, \zeta$)
$\mathbf{Y} \in \mathbb{R}^{n\times m}, \mathbf{L}, \boldsymbol{\lambda} \in \mathbb{R}^m \leftarrow$ DIFFMAP($\mathbf{X}, \varepsilon$)
$\mathbf{U}(i), \cdots, \mathbf{U}(n) \leftarrow$ RMETRIC($\mathbf{Y}, \mathbf{L}, d$)
for $S \in \{S' \subseteq [m] : |S'| = s, 1 \in S'\}$ **do**
  $\mathfrak{R}_1(S) \leftarrow 0; \mathfrak{R}_2(S) \leftarrow 0$
  for $i = 1, \cdots, n$ **do**
    $\mathbf{U}_S(i) \leftarrow \mathbf{U}(i)[S,:]$
    $\mathfrak{R}_1(S) \mathrel{+}= \frac{1}{2n}\cdot\log\det\left(\mathbf{U}_S(i)^\top\mathbf{U}_S(i)\right)$
    $\mathfrak{R}_2(S) \mathrel{+}= \frac{1}{n}\cdot\sum_{k=1}^d \log\|u_k^S(i)\|_2$
  end
  $\mathfrak{L}(S;\zeta) = \mathfrak{R}_1(S) - \mathfrak{R}_2(S) - \zeta\sum_{k\in S}\lambda_k$
end
$S_* = \operatorname{argmax}_S \mathfrak{L}(S;\zeta)$
Return: Independent eigencoordinates set S_*

Limit of loss $\mathfrak{L}$

Theorem (Limit of $\mathfrak{R}$) Let $j_S(\mathbf{y}) = 1/\operatorname{Vol}(\mathbf{U}_S(\mathbf{y})\mathbf{\Sigma}_S^{1/2}(\mathbf{y}))$; $\tilde{j}_S(\mathbf{y}) = \prod_{k=1}^d \left(\|u_k^S(\mathbf{y})\|\sigma_k(\mathbf{y})\right)^{1/2}\Big)^{-1}$.
Under the following **assumptions**: (i) The manifold $\mathcal{M}$ is compact of class $\mathcal{C}^3$, and there exists a set S, with $|S| = s$ so that ϕ_S is a smooth embedding of $\mathcal{M}$ in $\mathbb{R}^s$, (ii) The data are sampled from a distribution on $\mathcal{M}$ continuous w.r.t. $\mu_{\mathcal{M}}$ with density p, and (iii) The estimate of $\mathbf{H}_S$ in Algorithm 1 computed w.r.t. the embedding ϕ_S is consistent, we have $\lim_{n\to\infty}\frac{1}{n}\sum_i \ln\mathfrak{R}(S,\mathbf{x}_i) = \mathfrak{R}(S,\mathcal{M})$, with

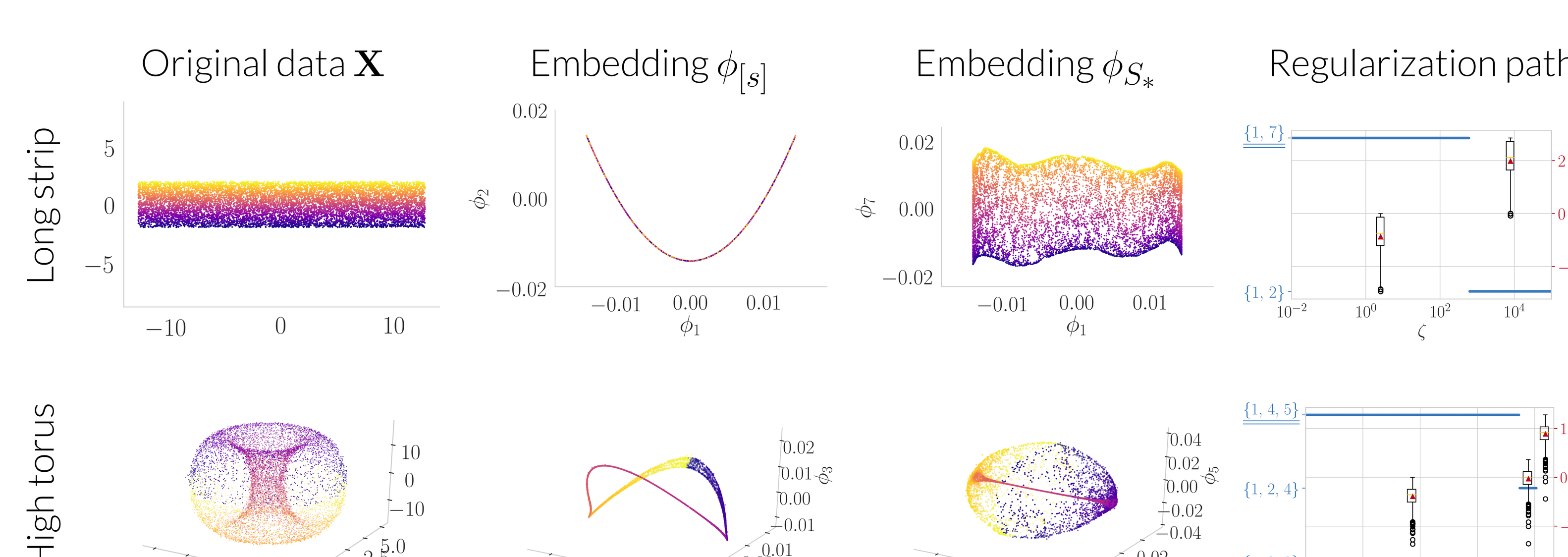
$$\mathfrak{R}(S,\mathcal{M}) = -\int_{\phi_S(\mathcal{M})} \ln\frac{j_S(\mathbf{y})}{\tilde{j}_S(\mathbf{y})} p(\phi_S^{-1}(\mathbf{y})) j_S(\mathbf{y}) d\mu_{\phi_S(\mathcal{M})}(\mathbf{y}) \stackrel{\text{def}}{=} -D(pj_S\|p\tilde{j}_S)$$

Because $j_S \geq \tilde{j}_S$ the divergence D is always positive.

The limit of regularization term $\phi_k^\top\mathbf{L}\phi_k \to \int_{\mathcal{M}}\|\operatorname{grad}\phi_k(\mathbf{x})\|_2^2 d\mu(\mathcal{M})$ when ϕ_k satisfies the Neumann boundary condition.

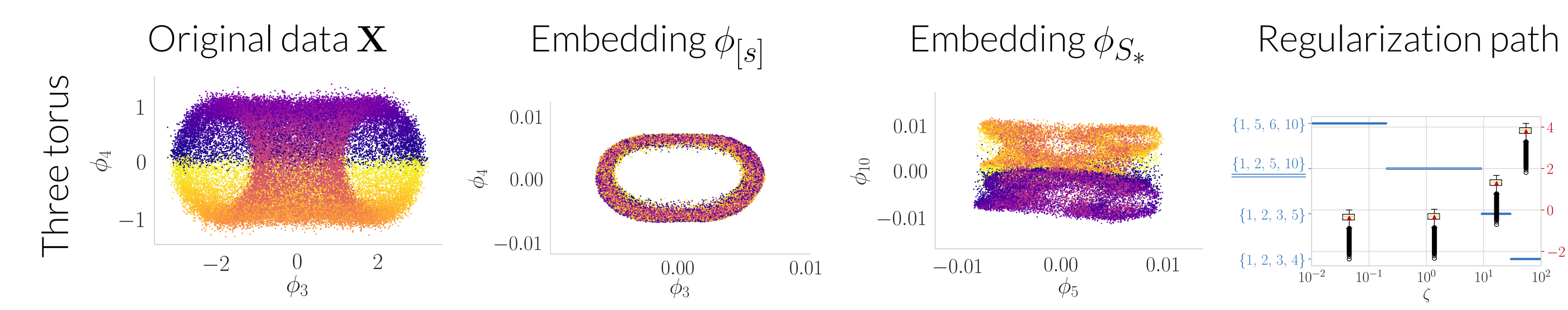
Experiments

Synthetic dataset — long strip and high torus



Experiments (cont.) & Discussion

Synthetic dataset — three torus



Real dataset

	n	D	deg_{avg}	(s,d)	t (sec)	S_*
SDSS (Abazajian et al. 2009)	299k	3750	144.91	(2, 2)	106.05	**(1, 3)**
Aspirin (Chmiela et al. 2017)	212k	244	101.03	(4, 3)	85.11	**(1, 2, 3, 7)**
Ethanol	555k	102	107.27	(3, 2)	233.16	**(1, 2, 4)**
Malondialdehyde	993k	96	106.51	(3, 2)	459.53	(1, 2, 3)
CH3Cl (Fleming et al. 2016)	23k	34	91.84	(3, 2)	8.37	**(1, 4, 6)**

Selected eigenvectors ↑

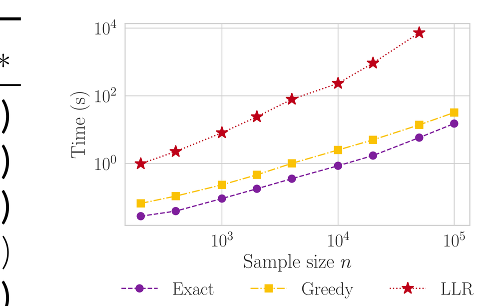


Figure 3. Runtime on 2D long strip, compared with [3].

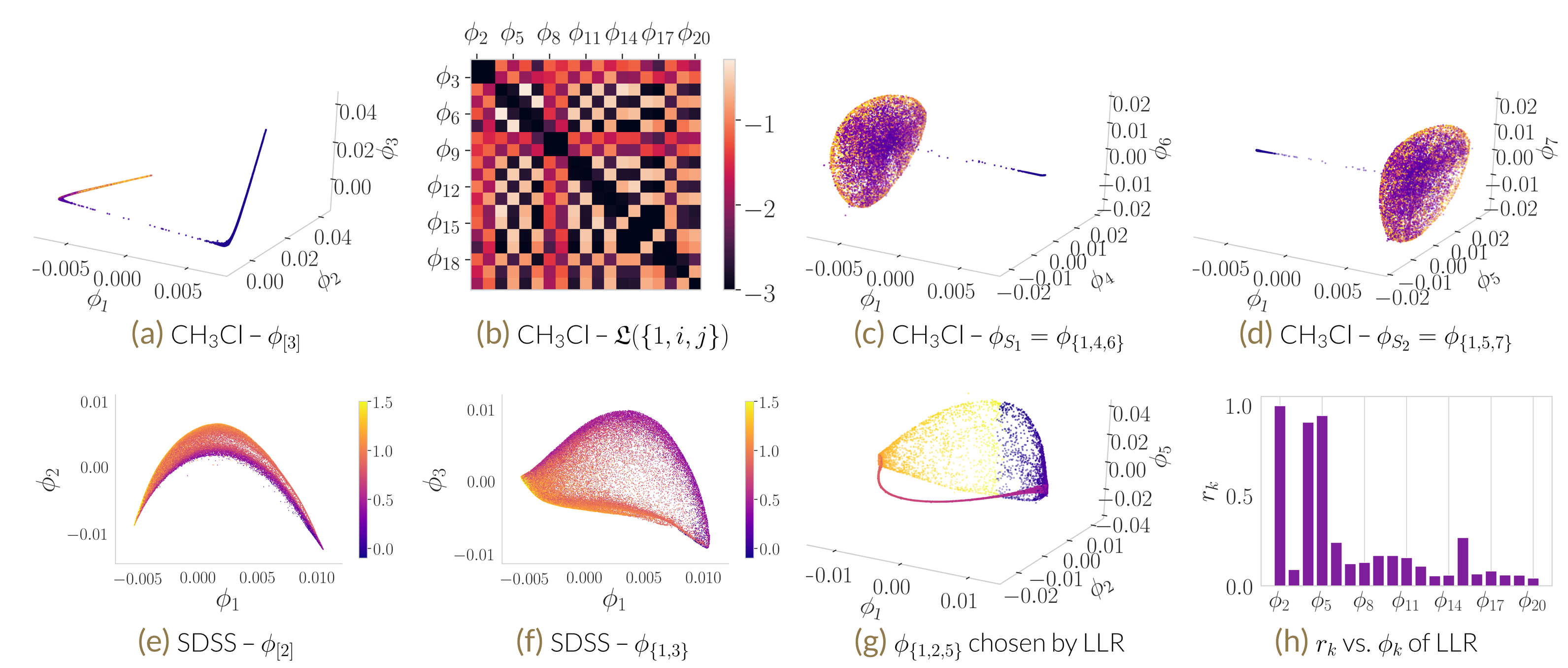
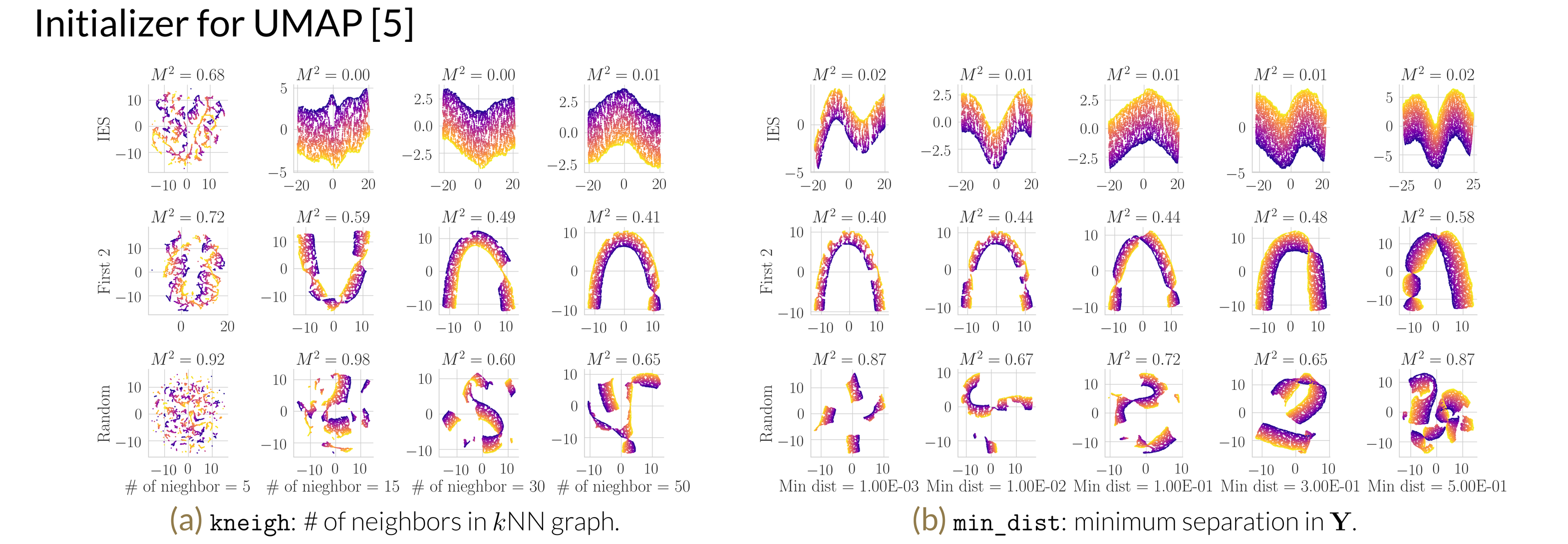


Figure 4. Experimental result – real datasets & comparison. LLR is the method by [3].

Initializer for UMAP [5]



Discussion & Future works

- Defect of sequential search (see Figure 4g & 4h).
- Extension to LTSA & HLLE with gradient estimation by coefficient Laplacian (Ting & Jordan, 2018).
- Manifold optimization on the Grassmannian.

References

[1] Jonathan Bates. The embedding dimension of laplacian eigenfunction maps. *Applied and Computational Harmonic Analysis*, 37(3):516–530, 2014.

[2] R. R. Coifman and S. Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 30(1):5–30, 2006.

[3] Carmeline J Dsilva, Ronen Talmon, Ronald R Coifman, and Ioannis G Kevrekidis. Parsimonious representation of nonlinear dynamical systems through manifold learning: A chemotaxis case study. *Applied and Computational Harmonic Analysis*, 44(3):759–773, 2018.

[4] Yair Goldberg, Alon Zakai, Dan Kushnir, and Ya'acov Ritov. Manifold learning: The price of normalization. *Journal of Machine Learning Research*, 9(Aug):1909–1939, 2008.

[5] Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.

[6] D. Perrault-Joncas and M. Meila. Non-linear dimensionality reduction: Riemannian metric estimation and the problem of geometric discovery. *ArXiv e-prints*, May 2013.