Lecture: Vector Diffusion Maps & Anisotropic Diffusion Maps

Foundations of Data Science: Algorithms and Mathematical Foundations

Mihai Cucuringu
mihai.cucuringu@stats.ox.ac.uk

CDT in Mathematics of Random System
University of Oxford

27 September, 2023

Vector Diffusion Maps

Anisotropic diffusion maps

Recall: Diffusion Maps

- consider a set of N points $V = \{x_1, x_2, \ldots, x_N\}$ in an p-dimensional space $\mathbb{R}^p$
- each point (typically) characterizes an image (or an audio stream, text string, etc.)
- if two images x_i and x_j are similar, then $||x_i - x_j||$ is small
- a popular measure of similarity between points in $\mathbb{R}^p$ is defined using the Gaussian kernel

$$w_{ij} = e^{-||x_i - x_j||^2/\epsilon}, \qquad (i,j) \in E$$

  so that the closer x_i is from x_j, the larger w_{ij}
- the matrix $W = (w_{ij})_{1 \le i,j \le N}$ is symmetric and has positive coefficients
- to normalize W, we define the diagonal matrix D, with $D_{ii} = \sum_{j=1}^{N} w_{ij}$ and define L by

$$L = D^{-1} W$$

  such that every row of L sums to 1.

Diffusion Maps

- interpre the eigenvectors as functions over our data set
- the *diffusion map* maps points from the original space to the first k eigenvectors of L ($k << p$), $\mathcal{L} : V \mapsto \mathbb{R}^k$

$$\mathcal{L}_t(x_j) = (\lambda_1^t \psi_1(j), \lambda_2^t \psi_2(j), \ldots, \lambda_k^t \psi_k(j)) \tag{1}$$

- the Euclidean distance in the diffusion map space

$$||\mathcal{L}(x_i) - \mathcal{L}(x_j)||^2 = \sum_{r=1}^{N-1} \left(\lambda_r^t \psi_r(i) - \lambda_r^t \psi_r(j)\right)^2 \tag{2}$$

- can be shown to equal the diffusion distance $D_t^2(i,j)$

$$L_{ij}^t = Pr\{x(t) = x_j | x(0) = x_i\}$$

- quantify the similarity between two points according to the evolution of their probability distributions

$$D_t^2(i,j) = \sum_{k=1}^{N} (L_{ik}^t - L_{jk}^t)^2 \frac{1}{d_k},$$

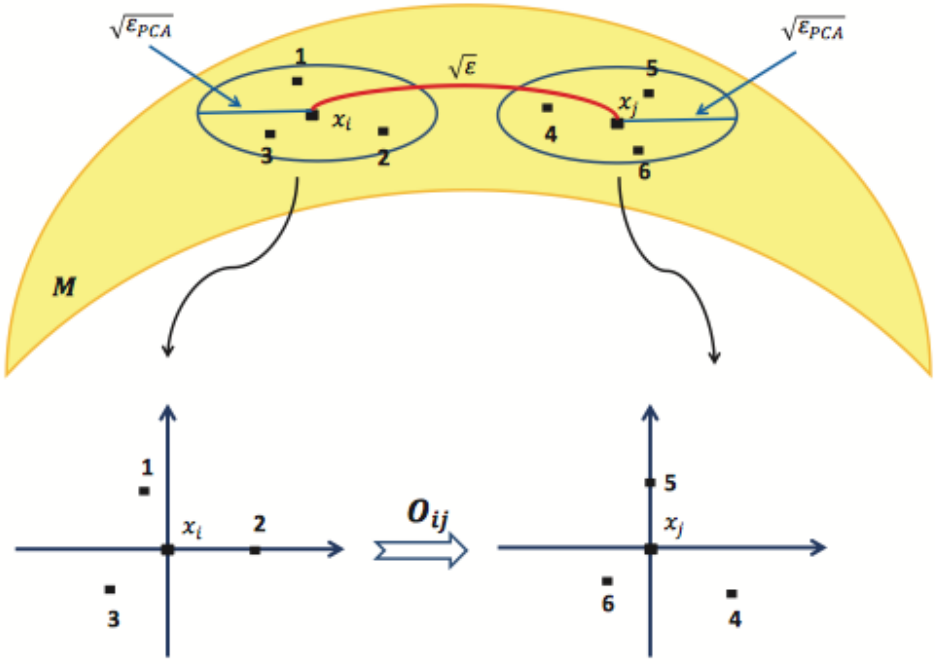
- $D_t(i,j)$ is the *diffusion distance* at time t

Vector Diffusion Maps

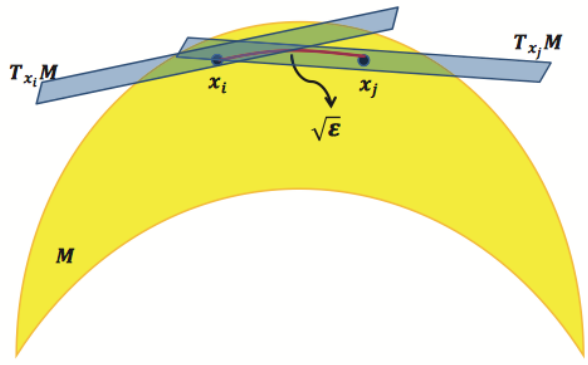
Anisotropic diffusion maps

Vector Diffusion Maps (VDM) (Singer and Wu, 2011)

- relationships between data points are represented as a weighted graph
- weights w_{ij} describe the affinities between data points
- together with linear orthogonal transformations O_{ij}
- the additional information captured by the orthogonal matrices O_{ij} can be exploited in such a way that the final low-dimensional representation of the graph is
  - significantly more robust to noise
  - extremely effective in discovering the structure of the underlying manifold
- the orthogonal matrices O_{ij} are obtained via a 2-step procedure



(a)



(b)

A. Singer, H.-T. Wu, "Vector Diffusion Maps and the Connection Laplacian", Communications on Pure and Applied Mathematics, 65 (8), pp. 1067–1144 (2012)

- symmetric matrix S ($n \times n$ blocks, each of size $d \times d$)

$$S(i,j) = \begin{cases} w_{ij} O_{ij} & \text{if } (i,j) \in E \\ 0_{d \times d} & \text{if } (i,j) \notin E, \end{cases} \tag{3}$$

- build the graph Connection Laplacian

$$L = D^{-1} S$$

Vector Diffusion Maps (VDM) (Singer & Wu)

- Step 1: perform local Principal Component Analysis
- gives an approximation for the orthonormal basis O_i for the tangent space of the manifold at point x_i
- the matrix O_i is a $p \times d$ matrix with orthonormal columns $O_i^T O_i = I_{d\times d}$, where d is the ambient space dimension
- Step 2: perform a local alignment by solving

$$O_{ij} = \text{argmin}_{O\in O(d)} ||O - O_i^T O_j||,$$

which can be easily computed via SVD of $O_i^T O_j$

- perform the above two steps for each pair of points situated close enough on the manifold
- build the symmetric matrix

$$S(i,j) = \begin{cases} w_{ij} O_{ij} & \text{if } (i,j) \in E \\ 0_{d\times d} & \text{if } (i,j) \notin E, \end{cases} \tag{4}$$

composed of $n \times n$ blocks, each of size $d \times d$

Vector Diffusion Maps (VDM) (Singer and Wu)

- build the graph Connection Laplacian

$$L = D^{-1}S$$

- the vector diffusion mapping is defined as

$$V_t : i \mapsto \left((\lambda_l \lambda_r)^t \langle v_l(i), v_r(i) \rangle\right)_{l,r=1}^{nd}, \tag{5}$$

- the vector diffusion distance is given by
  $d^2_{VDM}(i,j) = \langle V_t(i), V_t(i) \rangle + \langle V_t(j), V_t(j) \rangle - 2\langle V_t(i), V_t(j) \rangle$
- VD mapping and distances can be well approximated by using only the top eigenvalues and eigenvectors of the Connection Laplacian $D^{-1}S$
- use this embedding as a low-dimensional representation of our given data
- adjust this framework depending on the target application

Vector Diffusion Maps

Anisotropic diffusion maps

Anisotropic diffusion maps

- detect intrinsic slow variables in high-dimensional stochastic chemical reaction networks. It combines:
- anisotropic diffusion maps (ADM): A. Singer, R. Erban, I. G. Kevrekidis, and R. R. Coifman, PNAS (2009)
- with approximations based on the chemical Langevin equation (CLE)
- without any a-priori knowledge of the the slow variable

M. Cucuringu and R. Erban, *Detecting slow variables and their stationary distribution in continuous time Markov chains and dynamic data via anisotropic diffusion maps*, SIAM Journal on Scientific Computing, 39(1), B76-B101, arXiv: 1504.01786

Illustrative example - system

- consider the following system

$$X_2 \underset{k_2}{\overset{k_1}{\rightleftarrows}} X_1 + X_2, \qquad \emptyset \underset{k_4}{\overset{k_3}{\rightleftarrows}} X_1, \qquad X_1 + X_1 \underset{k_6}{\overset{k_5}{\rightleftarrows}} X_2, \tag{6}$$

- involving two molecular species X_1 and X_2
- reactions $R_1, R_2, \ldots, R_6$ have the propensity functions

$$\alpha_1(t) = k_1 X_2(t), \quad \alpha_2(t) = k_2 X_1(t) X_2(t)/V, \quad \alpha_3(t) = k_3 V,$$
$$\alpha_4(t) = k_4 X_1(t), \quad \alpha_5(t) = k_5 \frac{X_1(t)(X_1(t)-1)}{V}, \quad \alpha_6(t) = k_6 X_2(t)$$

- V denotes the system volume (we used $V = 8$) with dimensionless parameters

$$k_1 = 32,\ k_2 = 0.04V;\ k_3 V = 1475;\ k_4 = 19.75;\ k_5 = 10V;\ k_6 = 4000;$$

- R_5, R_6 occur on a much faster timescale than R_1, R_2, R_3, R_4
- $S = X_1 + 2X_2$ (invariant with respect to all fast reactions)

Illustrative example - transition diagram

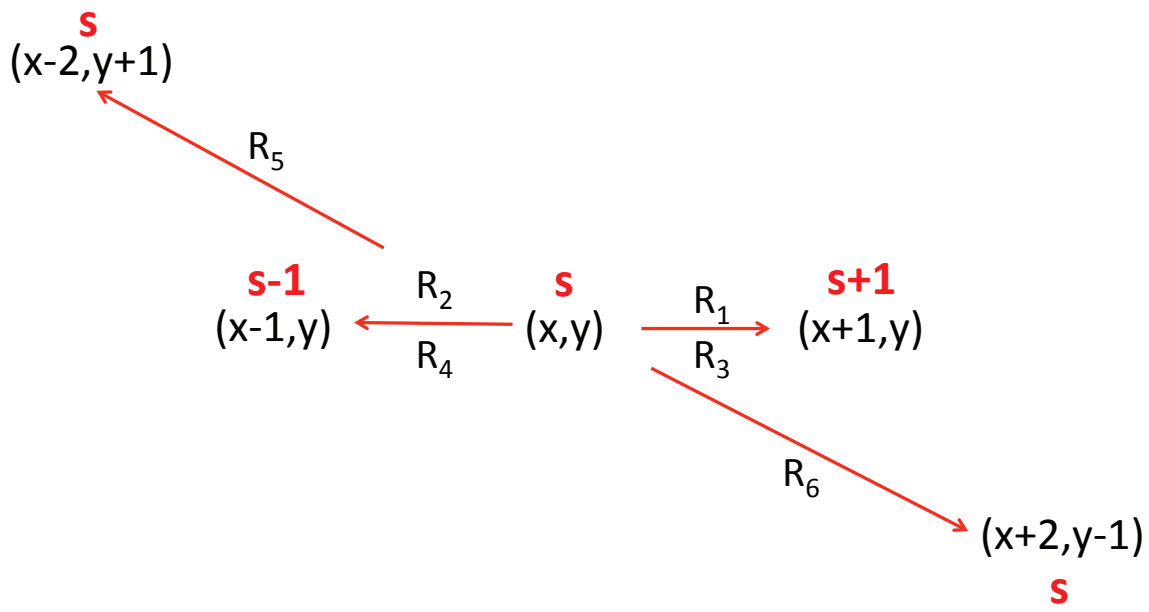
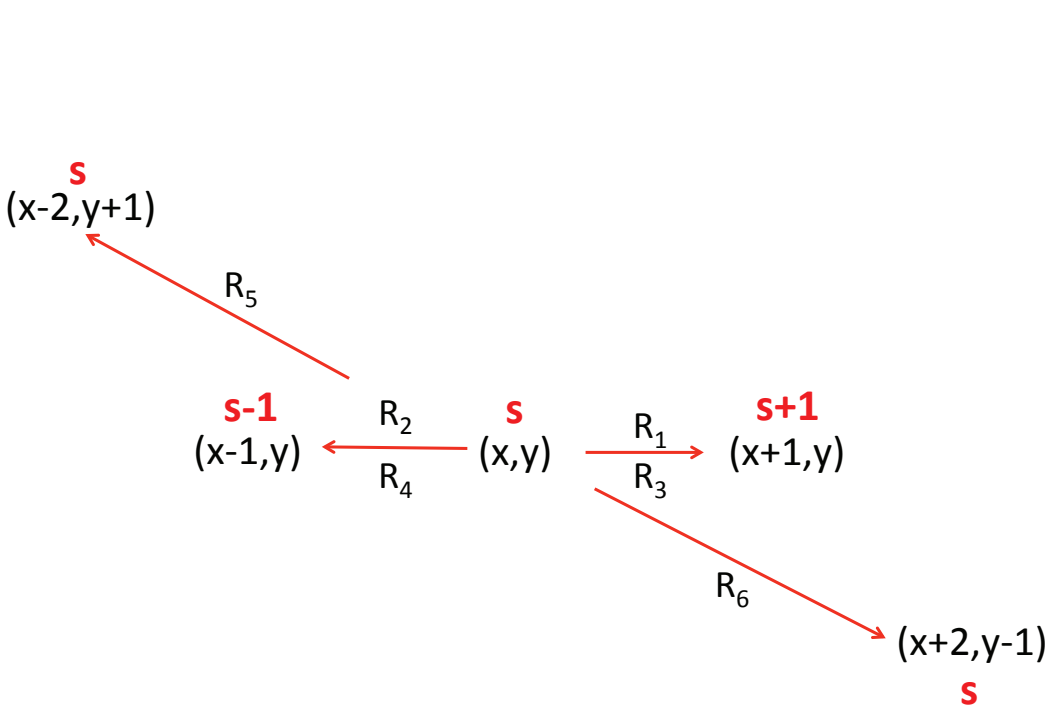


Figure: *Transition diagrams for a chemical system.*

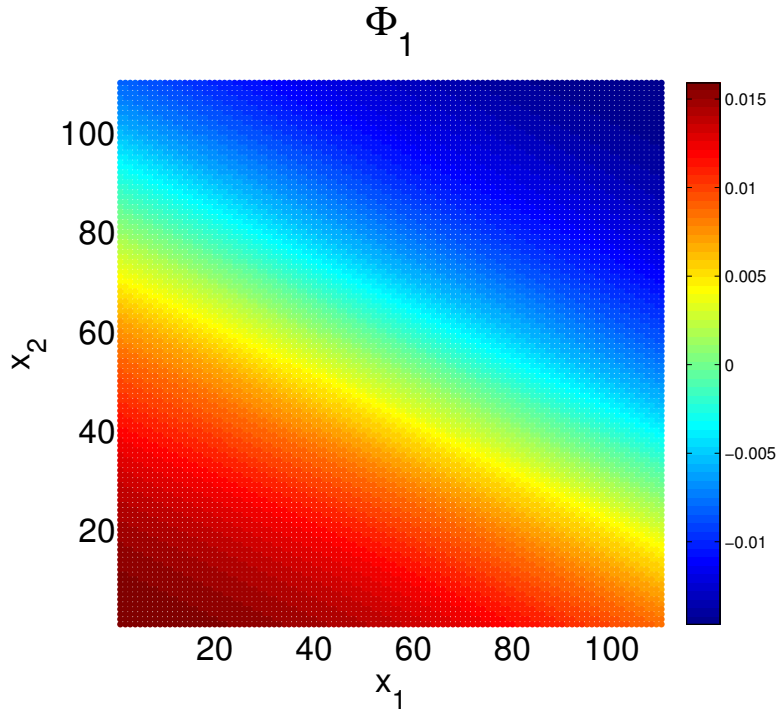
Here x denotes X_1, and y denotes X_2

$S = X_1 + 2X_2$ (invariant with respect to all fast reactions)

Illustrative example



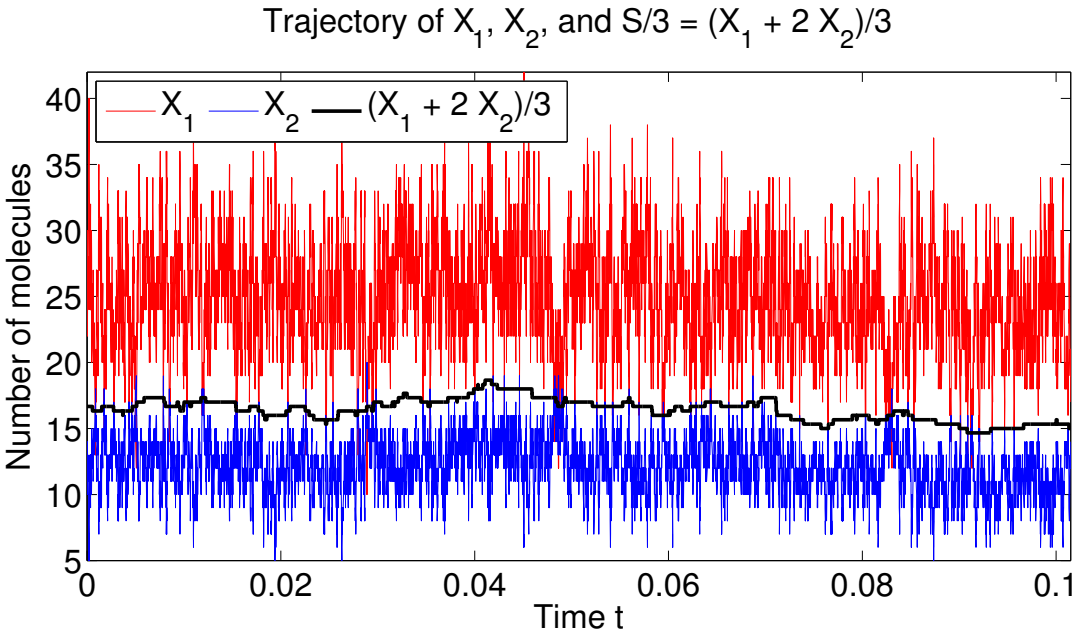
(a)



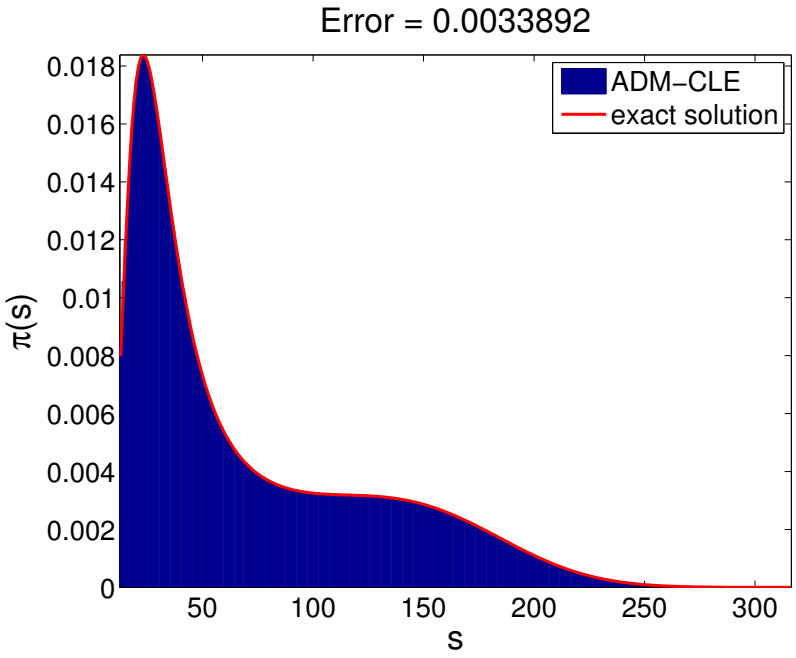
(b)

Figure: *Left: transition diagram for a 2-dimensional chemical system (with two molecular species) and reactions $R_1, \ldots, R_6$. Right: coloring of the nodes of G (states of the observable space) according to their corresponding entry in the top eigenvector Φ_1 of L.*

Illustrative example



(a)



(b)

Figure: *Left: trajectories of the* CS*, showing the slow behavior of the variable* $S = X_1 + 2X_2$ *in contrast to the fast behavior of variables* X_1 *and* X_2*. Right: the final estimated stationary distribution of the slow variable S, computed without knowledge of the slow variable (blue histograms). Red solid lines shows the exact solution.*

Fast and slow variables

- assume that $s = s(x_1, x_2) = x_1 + 2x_2$ and $f = f(x_1, x_2) = x_1$ are the slowly and rapidly changing variables, respectively
- together they define a mapping $g : (x_1, x_2) \mapsto (s, f)$
  - from the observable state variables x_1 and x_2 in the accessible space $\mathcal{O}$
  - to the "dynamically meaningful" (but in more complicated examples inaccessible) slow variable s and the fast accessible variable f, both in space $\mathcal{H}$
- in other words, g maps $(x_1, x_2) \mapsto (x_1 + 2x_2, x_1)$, and conversely its inverse $h := g^{-1} : (s, f) \mapsto (f, \frac{s-f}{2})$.

Contribution

1. avoid local bursts of simulations at each point to estimate the local covariances via (analytical) CLE approximation

2. build a sparse ellipsoid-like neighborhood graph at each point in the data set (leads to a sparse graph Laplacian) by exploiting the spectrum of each local covariance matrix
   - associate a state (i.e., node in the initial graph) to each possible combination of pairs of states (x_1, x_2) (not feasible whenever the range of the variables is large)
   - avoid computing the similarity between all $\binom{n}{2}$ pairs of points in the domain

3. introduced an unsupervised spectral-based method for inferring the slow variable

4. proposed a Markov-based approach for estimating the stationary distribution of the slow variable

Covariance-dependent distance

- The Σ-dependent distance between two $\mathcal{O}$-states is given by

$$d_{\Sigma}^{2}\left((x_1,x_2)^{(i)},(x_1,x_2)^{(j)}\right)=\frac{1}{2}\left((x_1,x_2)^{(i)}-(x_1,x_2)^{(j)}\right)\times \left(\Sigma^{-1}_{(x_1,x_2)^{(i)}}+\Sigma^{-1}_{(x_1,x_2)^{(j)}}\right)\left((x_1,x_2)^{(i)}-(x_1,x_2)^{(j)}\right)^{T} \quad (7)$$

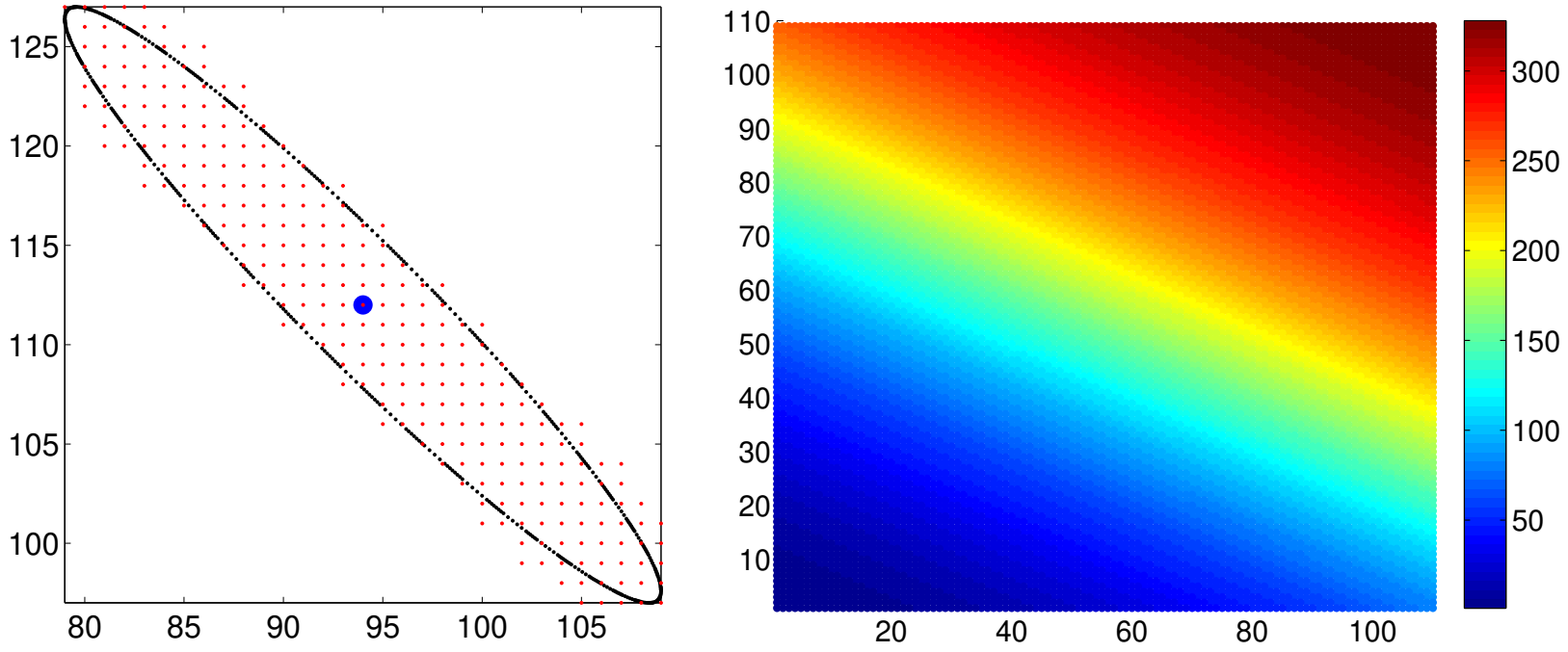
- represents a second order approximation of the Euclidean distance in the inaccessible $(s,\tau f)$-space

$$d_{\Sigma}^{2}[(x_1,x_2)^{(i)},(x_1,x_2)^{(j)}]\approx (s^{(i)}-s^{(j)})^2+\tau^2(f^{(i)}-f^{(j)})^2\approx (s^{(i)}-s^{(j)})^2 \quad (8)$$

Anisotropic diffusion maps

$$W_{ij} = \exp(-d_{\Sigma}^2(x^{(i)}, x^{(j)})/\epsilon^2) \tag{9}$$

Random-walk Laplacian: $L = D^{-1}W$



(a) Local neighborhood graph (b) Top eigenvector of the ADM

Figure: Left: the local neighborhood graph at a node; its shape is an ellipsoid whose axis ratio is given by the ratio of the eigenvalues of the covariance matrix. Right: Eigenvector coloring of the nodes in the state space according to their corresponding entry in the top eigenvector of the anisotropic diffusion map.