DIFFUSION MAPS

MARIA CAMERON

Contents

1. Principal Component Analysis

Principal Component Analysis (PCA), originally formulated by Pearson (1901), is a widely used tool for data analysis in both natural and social sciences. A classical reference for the PCA is the book by I. T. Jolliffe [1] available online. Here I give just a brief overview.

Let $\eta \in \mathbb{R}^D$ be a vector random variable. Let

$$X = \begin{bmatrix} x_1^\top & \to \\ \vdots & \\ x_n^\top & \to \end{bmatrix}$$

be an $n \times D$ matrix of samples of η. The goal of the PCA is to map the samples of η from the high-dimensional space $\mathbb{R}^D$ into a low-dimensional space $\mathbb{R}^d$, while retaining as much variation present in the samples as possible.

1.1. Derivation.

We would like to replace η with a random variable $\xi \in \mathbb{R}^d$ whose components are linear combinations of the components of η,

$$\xi = (w_1^\top \eta, \dots, w_d^\top \eta)$$

where the coefficients w_i, $i = 1, \dots, d$ maximize the variance $\mathsf{Var}(w_i^\top \eta)$ while satisfy the constraints

$$w_i^\top w_i = 1, \quad i = 1, \dots, d, \quad \text{and} \quad \mathsf{Cov}(w_i\eta, w_j\eta) = 0.$$

The condition $\mathrm{Cov}(w_i\eta, w_j\eta) = 0$ means that the components of ξ should be uncorrelated.

First we find PCA 1, $\xi_1 = w_1\eta$ using Lagrange multipliers. Let

$$C := \mathsf{Cov}(\eta) = (E[(\eta_i - E[\eta_i])(\eta_j - E[\eta_j])])_{i,j=1}^{D}$$

be the covariance matrix of η. It is $D \times D$, symmetric positive definite provided that all components of η have nonzero variances. Then

$$\mathsf{Var}(\xi_1) = E[(w_1^\top \eta - w_1^\top E[\eta])^2] = w_1^\top C w_1.$$

The Lagrangian function is given by

$$L(w, \lambda) = w_1^\top C w_1 - \lambda(w_1^\top w_1 - 1),$$

where λ is the Lagrange multiplier. Differentiating L with respect to w_1 we get

$$Cw_1 - \lambda w_1 = (C - I\lambda)w_1 = 0.$$

Hence w_1 must be an eigenvector of C corresponding to the eigenvalue λ. To decide which eigenvalue we pick, we recall that we are maximizing

$$w_1^\top C w_1 = \lambda w_1^\top w = \lambda.$$

Hence, we pick λ_1, the largest eigenvalue of C, and the corresponding eigenvector w_1.

Next we will look for w_2. The zero covariance condition gives:

$$\mathrm{Cov}(w_1\eta, w_2\eta) = w_1^\top C w_2 = \lambda_1 w_1^\top w_2 = 0.$$

Hence w_2 must be orthogonal to w_1. The Lagrangian function is

$$L(w_2, \lambda, \phi) = w_2^\top C w_2 - \lambda(w_2^\top w_2 - 1) - \phi w_1^\top w_2.$$

Differentiating it with respect to w_2 we get:

$$Cw_2 - \lambda w_2 - \phi w_1 = (C - I\lambda)w_2 - \phi w_1 = 0.$$

Taking a dot product of this equation with w_1 we get: $0 - 0 - \phi = 0$ which forces ϕ to be zero. Hence, λ and w_2 must be an eigenpair corresponding to the second largest eigenvalue of C.

Proceeding in a similar manner, we find that w_k is the eigenvector of C corresponding to its kth largest eigenvalue.

The variables $\xi_i = w_i^\top \eta$ where w_i, $i = 1, \dots, d$ are the eigenvectors of the covariance matrix C corresponding to the d largest eigenvalues, are called *the principal components*. The vector w_i, $i = 1, \dots, d$, is called *the vector of coefficients or loadings for the ith principal component*.

1.2. **Calculation in practice.** In practice, when we are dealing with data, the data points $x_i \in \mathbb{R}^D$ are interpreted as samples of a vector random variable. The covariance matrix is not known. To approximate it, we compute the $D \times D$ data covariance matrix. First we need to center the data so that column means are all zero:

$$Y := X - \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \left[\frac{1}{n} \sum_{i=1}^{n} x_{i1}, \ldots, \frac{1}{n} \sum_{i=1}^{n} x_{iD} \right].$$

Then the covariance matrix is given by

$$C := \frac{1}{n} Y^\top Y.$$

Its eigendecomposition with eigenvalues ordered in the decreasing order is

$$C = W \Lambda W^\top, \quad \text{where} \quad \Lambda = \operatorname{diag}\{\lambda_1, \ldots, \lambda_D\}$$

and $\lambda_1 \geq \ldots \geq \lambda_D \geq 0$. The d eigenvectors corresponding to the d largest eigenvalues will be the desired loadings. The coordinates of the data points in the principal component space will be $z_i := Y w_i$, $i = 1, \ldots, d$.

Now let us connect the PCA with the SVD. Let

$$Y = U \Sigma W^\top$$

be an SVD of Y. Then

$$Y^\top Y = W \Sigma^2 W^\top \equiv nC = n W \Lambda W^\top.$$

Hence the first d columns of $W = [w_1, \ldots, w_D]$ are the vectors of coefficients for the first d principal components, and the principal components are $Y w_i$.

2. Diffusion maps

While the PCA is a power tool when the data points lie near a d-dimensional hyperplane in $\mathbb{R}^D$, it might fail to give a nice embedding if the data are instead located near some d-dimensional curved manifold. To handle this case, Coifman and Lafon (Yale University, 2006) introduced the so-called *diffusion maps* [2]. The key idea of this approach is to devise a discrete-time Markov chain on the data points and define the distances between remote points using the stochastic matrix of this Markov chain. This approach is robust to noisy data and is capable of adequately representing complex geometries of data structures. This dimensional reduction technique has been successfully applied to problems arising in protein dynamics (e.g. [3, 4]). This approach is that it requires providing a scaling parameter ϵ. A number of approaches for choosing it have been developed: see e.g. [5, 6].

2.1. **Background: properties of stochastic matrices.** An $n \times n$ matrix P is called *stochastic* if its entries are nonnegative and its row sums are equal to 1. The entries $P_{i,:}$ can be interpreted as the transition probabilities from state i: p_{ij} is the probability that the system currently at state i will go next to state j.

Definition 1. *We say that a sequence of random variables $(X_k)_{k\geq 0}$, $X_k : \Omega \to S \subset \mathbb{Z}$, is a Markov chain with initial distribution λ and stochastic matrix $P = (p_{ij})_{i,j\in S}$ if*

(1) *X_0 has distribution $\lambda = \{\lambda_i \mid i \in S\}$ and*
(2) *the Markov property holds:*

$$\mathbb{P}(X_{k+1} = i_{k+1} \mid X_k = i_k, \dots, X_0 = i_0) = \mathbb{P}(X_{k+1} = i_{k+1} \mid X_k = i_n) = p_{i_k i_{k+1}}.$$

We will write a probability distribution as a row vector. One can show that if λ is the initial probability distribution, then the probabiliby distribution after one step becomes λP, in two steps λP^2, in k steps λP^k, and so on. A probability distribution π is *invariant* if

$$\pi P = \pi \quad \text{and} \quad \sum_{i=1}^{n} \pi_i = 1.$$

If μ is a row vector with n entries such that $\mu P = \mu$, we say that μ is an *invariant probability measure*. Note that μ does not need to sum up to 1[1].

We will limit ourselves to a special kind of Markov chains arising in diffusion maps:

- The number of states is finite: $|S| = n$.
- The stochastic matrix P is irreducible and aperiodic. *Irreducibility* means that any state can be reached from any state in a finite number of jumps, i.e, for any pair i, j, $(P^t)_{ij} > 0$ for some $t \in \mathbb{N}$. *Aperiodcity* means that for any state i and for all sufficiently large t, there is a nonzero probability of returning to i in t steps, i.e, $(P^t)_{ii} > 0$ for all large enogh t and for all i. In this case, one can prove that there exists a *unique invariant probability distribution* π, and for any initial probability distribution λ we have
$$\lim_{k\to\infty} \lambda P^k = \pi.$$
- The Markov chain is time-reversible, or, equivalently, possesses the property of detailed balance: if π is the invariant probability distribution then
$$\pi_i P_{ij} = \pi_j P_{ji},$$
The detailed balance means that, on average, the number of transitions from i to j per time unit is the same as that from j to i.

2.1.1. *Spectral decomposition.* The detailed balance property can be written in the matrix form:

$$\Pi P = P^\top \Pi, \quad \text{where} \quad \Pi := \mathsf{diag}\{\pi_1, \dots, \pi_n\}.$$

Note that the detailed balance condition $\Pi P = P^\top \Pi$ implies that ΠP is symmetric. Indeed, its transpose is $P^\top \Pi$. Hence, the stochastic matrix P is decomposable as

$$P = \Pi^{-1} \tilde{K}, \quad \text{where } \tilde{K} \text{ is symmetric.}$$

Furthermore, P has one eigenvalue equal to 1. The corresponding right eigenvector is $r_0 = [1, \dots, 1]^\top$ (as row sums are all 1), while the corresponding left eigenvector is is π, the

[1]Note that is the set of states is infinite, invariant measure may exist while invariant distribution does not exist. For example, consider a symmetric random walk on $\mathbb{Z}$.

invariant distribution (as $\pi P = \pi$). All other eigenvalues of P are less than 1 in absolute value. The fact that they do not exceed 1 in absolute value readily follows for Gershgorin circle theorem saying that the eigenvalues of a matrix A are located within the union of Gershgorin discs $D(a_{ii}, R_i) \subset \mathbb{C}$, where $R_i = \sum_{j \neq i} |a_{ij}|$. Each such disc has center on the real axis in the interval $[0, 1]$ and has radius at most 1. The fact that all other eigenvalues are less than 1 in absolute value follows from aperiodicity and irreducibility.

Exercise Prove this.

The detailed balance condition $\Pi P = P^\top \Pi$ implies that P is similar to a symmetric matrix

$$\Pi^{1/2} P \Pi^{-1/2} = \Pi^{-1/2} (\Pi P) \Pi^{-1/2}.$$

Hence all eigenvalues of P are real. Furthermore, let

$$V \Lambda V^\top$$

be the spectral decomposition of $\Pi^{1/2} P \Pi^{-1/2}$. Then

$$V^\top \Pi^{1/2} P \Pi^{-1/2} V = (\Pi^{-1/2} V)^{-1/2} P (\Pi^{-1/2} V) = \Lambda.$$

Hence

$$P = (\Pi^{-1/2} V) \Lambda (\Pi^{-1/2} V)^{-1}$$

is the eigendecomposition of P. Denoting the matrix $\Pi^{-1/2} V$ of right eigenvectors of P by R, we express $V = \Pi^{1/2} R$. Hence, the matrix $L = (\Pi^{-1/2} V)^{-1} = V^\top \Pi^{1/2}$ of left eigenvectors of P is expressed via R and Π as:

$$L = V^\top \Pi^{1/2} = R^\top \Pi.$$

Hence, the eigendecomposition of P is

$$P = R \Lambda R^\top \Pi. \tag{1}$$

Since $RL = LR = I$, we have

$$R^\top \Pi R = I. \tag{2}$$

2.2. **A basic construction of a diffusion map.** First, we present the most basic diffusion map algorithm corresponding to $\alpha = 0$ in [2]. This construction is very similar to the construction of Laplacian eigenmap by Belkin and Niyogi (2003) [7].

Let $X = (x_{ik})$ be an $n \times D$ matrix of data. The rows x_i, $i = 1, \ldots, n$, of X represent data points lying in $\mathbb{R}^D$.

- First we compute the squared-distance matrix between the data points:

$$\Delta(i, j) = \sum_{k=1}^{D} (x_{ik} - x_{jk})^2.$$

- Next, we pick a scaling parameter ϵ and define the diffusion kernel, an $n \times n$ matrix $K = (k_{ij})$ where

$$k_{ij} = \exp\left(-\frac{\Delta(i, j)}{\epsilon} \right).$$

A good choice of ϵ is very important. ϵ should be comparable to squared distances from the data points to their neighbors. In practice, pick a reasonable initial guess for ϵ and then tune it experimentally. One way to pick an initial ϵ is the following. We find row minima among off-diagonal entries for the matrix Δ . Then we find the mean of these minima and set ϵ to be double this mean:

```
for i = 1 : N
    drowmin(i) = min(d(i,setdiff(1:N,i)));
end
epsilon = 2*mean(drowmin);
```

Then, if the result is not satisfactory, keep increasing the factor by which the mean of row minima is multiplied in the last line until the embedding starts making sense.

- Convert the diffusion kernel K into a stochastic matrix $P = (p_{ij})$ by dividing each row of K by the corresponding row sum:

$$P = Q^{-1}K \quad \text{where} \quad Q := \mathsf{diag}\left\{\sum_{j=1}^{n} k_{1j}, \dots, \sum_{j=1}^{n} k_{nj}\right\} := \mathrm{diag}\{q_1, \dots, q_n\}.$$

Indeed, all entries of the resulting matrix P are nonnegative, and its row sums are one.

Note that the diagonal entries of Q constitute an invariant probability measure. Indeed:

$$[q_1, \dots, q_n]Q^{-1}K = [1, \dots, 1]K = [q_1, \dots, q_n]$$

as $K = K^\top$ and its row sums are 1. To obtain the invariant probability distribution, we normalize $[q_1, \dots, q_n]$ so that it sums up to one:

$$\pi = \frac{q}{\sum_{i=1}^{n} q_i} \quad \text{where} \quad q := [q_1, \dots, q_n].$$

- Let us take tth power of the matrix P and denote its entries by $p_{ij}^t \equiv (P^t)_{ij}$. The entry p_{ij}^t is the probability to transition from i to j in t steps, $t \in \mathbb{N}$. A family of *diffusion distances* indexed by $t \in \mathbb{N}$ is defined by

(3)
$$\boxed{D_t(x_i, x_j)^2 := \sum_{m=1}^{n} \frac{1}{\pi_m} \left| p_{im}^t - p_{jm}^t \right|^2 .}$$

Hence, the diffusion distance is a weighted l_2 distance between rows i and j of the matrix P^t. Note that $D_t(x_i, x_j)$ will be small if there is a large number of short paths connecting x_i and x_j, which makes transition for either of them to any state x_m approximately equally likely. The power t plays the role of a scale parameter. Let us list interesting features of the diffusion distance:

  - Since it reflects the connectivity of the data at a given scale, points are closer if they are highly connected in the graph. Therefore, this distance emphasizes the notion of a cluster.
  - The quantity $D_t(x_i, x_j)$ involves summing over all paths of length t connecting x_i and x_j. This number is very robust to noise perturbation, unlike the geodesic distance.

- The family of *diffusion maps* $\Psi_t : \mathbb{R}^D \to \mathbb{R}^{n-1}$ indexed by $t \in \mathbb{N}$ from the data space $\mathbb{R}^D$ to the diffusion space $\mathbb{R}^{n-1}$ is defined so that the Euclidean distances in the diffusion space are equal to the diffusion distances:

$$\|\Psi_t(x_i) - \Psi_t(x_j)\| = D_t(x_i, x_j).$$

Let

$$P = R\Lambda L \equiv R\Lambda R^\top \Pi$$

be the spectral decomposition of P with ordered eigenvalues:

$$1 = \lambda_0 > |\lambda_1| \geq |\lambda_2| \geq \ldots \geq |\lambda_{n-1}|.$$

The diffusion map Ψ_t is defined by:

$$\Psi_t(x_i) := \begin{bmatrix} \lambda_1^t r_1(i) \\ \vdots \\ \lambda_{n-1}^t r_{n-1}(i) \end{bmatrix}, \tag{4}$$

where $R := [r_0, r_1, \ldots, r_{N-1}]$ is the matrix of right eigenvectors of P normalized so that $R^\top \Pi R = I$. Respectively, λ_m^t, is the tth power of λ_m, $m = 1, \ldots, n-1$. Note that since P is irreducible and aperiodic (as $P_{ii} > 0$, $i = 1, \ldots, n$) by construction, $\lambda_0 = 1$ and $r_0 = [1, \ldots, 1]^\top$. In Eq. (4), $r_m(i)$ denotes the ith entry of the vector r_k. In other words, $\Psi_t(x_i)$ is the transposed ith row of the matrix

$$(R_1 \Lambda_1^t)^\top \equiv [\lambda_1^t r_1, \lambda_2^t r_2, \ldots, \lambda_{n-1}^t r_{n-1}]^\top.$$

Note that we remove the first column of R because it consists of all ones and hence is not informative.

Proposition 1.

$$D_t(x_i, x_j)^2 = \sum_{m=1}^{n-1} \lambda_m^{2t} |r_m(i) - r_m(j)|^2, \tag{5}$$

i.e., the diffusion distance in the data space equals the Euclidean distance squared in the diffusion space.

We will prove this proposition after we finish the description of the construction.

- The diffusion maps allow us to do dimensional reduction by keeping only the first few components of $\Psi_t(\cdot)$. Often it is desirable to keep only the first two or three entries of $\Psi_t(\cdot)$ as then the diffusion map is readily visualizable. To make the dimension of the embedding space justified, we introduce an accuracy parameter $\delta \in (0, 1)$ and define the number of terms to keep:

$$s(\delta, t) = \max\{m \in \mathbb{N} \text{ such that } |\lambda_m|^t > \delta |\lambda_1|^t\}. \tag{6}$$

Then, up to relative precision δ, we have:

$$D_t(x_i, x_j)^2 = \sum_{m=1}^{s(\delta,t)} \lambda_m^{2t} |r_m(i) - r_m(j)|^2, \tag{7}$$

and

$$\Psi_t(x_i) = \begin{bmatrix} \lambda_1^t r_1(i) \\ \vdots \\ \lambda_{s(\delta,t)}^t r_{s(\delta,t)}(i) \end{bmatrix}. \tag{8}$$

This allows us to determine the power t for embedding into $\mathbb{R}^d$ as follows. We pick $\delta \in (0,1)$, for example, $\delta = 0.2$, and then define t so that t is the smallest integer such that

$$\left(\frac{|\lambda_1|}{|\lambda_d|}\right)^t \leq \delta \;\Rightarrow\; t = \mathsf{ceil}\left[\frac{\log(\delta)}{(\log(|\lambda_d|) - \log|\lambda_1|)}\right]. \tag{9}$$

Once the appropriate power for the desired dimension of embedding space (2 or 3) is found, we can define diffusion maps (abusing the term) to 2D or 3D diffusion spaces by

$$\Psi_t(x_i) = \begin{bmatrix} \lambda_1^t r_1(i) \\ \lambda_2^t r_2(i) \end{bmatrix} \quad \text{and} \quad \Psi_t(x_i) = \begin{bmatrix} \lambda_1^t r_1(i) \\ \lambda_2^t r_2(i) \\ \lambda_3^t r_3(i) \end{bmatrix}. \tag{10}$$

Now let us prove Proposition 1.

Proof. Let us redefine the diffusion kernel K as

$$K \to \left(\sum_{i=1}^{N} q_i\right)^{-1} K.$$

Then the stochastic matrix P with the new K can be decomposed as

$$P = \Pi^{-1} K.$$

P is similar to the symmetric matrix

$$A := \Pi^{1/2} P \Pi^{-1/2} = \Pi^{1/2}\Pi^{-1} K \Pi^{-1/2} = \Pi^{-1/2} K \Pi^{-1/2}.$$

Hence, the eigenvalues of A coincide with those of P. Let

$$A = \Phi \Lambda \Phi^\top = \sum_{k=0}^{n-1} \lambda_k \phi_k \phi_k^\top.$$

be an eigendecomposition of A where Φ is orthogonal, and the diagonal entries of Λ, the eigenvalues, are ordered in the decreasing order. Then the desired eigendecomposition of P can be obtained as follows:

$$P = \Pi^{-1/2} A \Pi^{1/2} = \Pi^{-1/2} \Phi \Lambda \Phi^\top \Pi^{1/2} =: R \Lambda L = \sum_{k=0}^{n-1} \lambda_k r_k l_k, \tag{11}$$

where $r_k := \Pi^{-1/2}\phi_k$, the columns of R, are the right eigenvectors of P, and $l_k := \phi_k^\top \Pi^{1/2}$, the rows of L, are the left eigenvectors of P. It can be readily verified that the left and right eigenvectors satisfy the following conjugacy relationships:

$$\sum_{m=1}^{n} \pi_m r_i(m) r_j(m) = r_i^\top \Pi r_j = \phi_i^\top \Pi^{-1/2} \Pi \Pi^{-1/2} \phi_j = \phi_i \phi_j = \delta_{i,j}, \tag{12}$$

$$\sum_{m=1}^{n} \frac{l_i(m) l_j(m)}{\pi_m} = l_i \Pi^{-1} l_j^\top = \phi_i^\top \Pi^{1/2} \Pi^{-1} \Pi^{1/2} \phi = \phi_i \phi_j = \delta_{i,j}. \tag{13}$$

Eq. (11) allows us to write entries of P^t as

$$p_{im}^t = \sum_{k=0}^{n-1} \lambda_k^t r_k(i) l_k(m). \tag{14}$$

Plugging p^t_{im} and p^t_{jm} into (3) defining $D_t(i,j)$, we get:

$$\begin{aligned} D_t(i,j)^2 &= \sum_{m=1}^{n} \frac{1}{\pi_m} \left[\sum_{k=0}^{n-1} \lambda_k^t r_k(i) l_k(m) - \lambda_k^t r_k(j) l_k(m) \right]^2 \\ &= \sum_{m=1}^{n} \frac{1}{\pi_m} \sum_{k=0}^{n-1} \left[\lambda_k^t r_k(i) l_k(m) - \lambda_k^t r_k(j) l_k(m) \right]^2 \\ &+ \sum_{m=1}^{n} \sum_{k=0}^{n-1} \sum_{s \neq k} \frac{l_k(m) l_s(m)}{\pi_m} \lambda_k^t \lambda_s^t [r_k(i) - r_k(j)][r_s(i) - r_s(j)]. \end{aligned}$$

Let us show that the second term in this sum is zero. Rearranging the order of summation and using (13) we get

$$\sum_{k=0}^{n-1} \sum_{s \neq k} \lambda_k^t \lambda_s^t [r_k(i) - r_k(j)][r_s(i) - r_s(j)] \sum_{m=1}^{n} \frac{l_k(m) l_s(m)}{\pi_m} = 0.$$

Returning to the first term, we calculate:

$$\begin{aligned} D_t(i,j)^2 &= \sum_{m=1}^{n} \frac{1}{\pi_m} \sum_{k=0}^{n-1} \left[\lambda_k^t r_k(i) l_k(m) - \lambda_k^t r_k(j) l_k(m) \right]^2 \\ &= \sum_{m=1}^{n} \sum_{k=0}^{n-1} \frac{[l_k(m)]^2}{\pi_m} \lambda_k^{2t} \left[r_k(i) - r_k(j) \right]^2 \\ &= \sum_{k=0}^{n-1} \lambda_k^{2t} \left[r_k(i) - r_k(j) \right]^2 \sum_{m=1}^{n} \frac{[l_k(m)]^2}{\pi_m} \\ &= \sum_{n=0}^{N-1} \lambda_n^{2t} \left[r_n(i) - r_n(j) \right]^2. \end{aligned}$$

Finally, we take into account that since $r_0 = [1, \ldots, 1]^\top$, $r_0(i) - r_0(j) = 0$. Therefore,

$$D_t(i,j) = \sum_{k=1}^{n-1} \lambda_k^{2t} \left[r_k(i) - r_k(j) \right]^2$$

as desired. □

2.3. **Illustrative examples.**

2.3.1. *Swiss Roll.* First we make an approximately uniform mesh of points on the Swiss Roll as shown in Fig. 1(a). The number of points is $n = 1060$. We set $\delta = 0.2$ and find the values for ϵ and t as described above:

$$\epsilon = 0.7717928, \quad t = 147.$$

The points are sorted and colored according to the approximate geodesic distance to the data point closest to the origin. The matrix P^t is displayed in Fig. 1(b). The absolute eigenvalues of P^t starting from $|\lambda_1|$ are shown in Fig. 1(c). The embedding to 3D is in Fig. 1(d). The Swiss Roll has been mostly unrolled.

Next, we repeat this experiment by adding noise to the data:

```
noisestd = 0.4;
X = X + noisestd*randn(size(X)); % perturb by Gaussian noise
```

Setting $\delta = 0.2$ as before, we find:

$$\epsilon = 0.6108029, \quad t = 300.$$

The results are shown in Fig. 2.

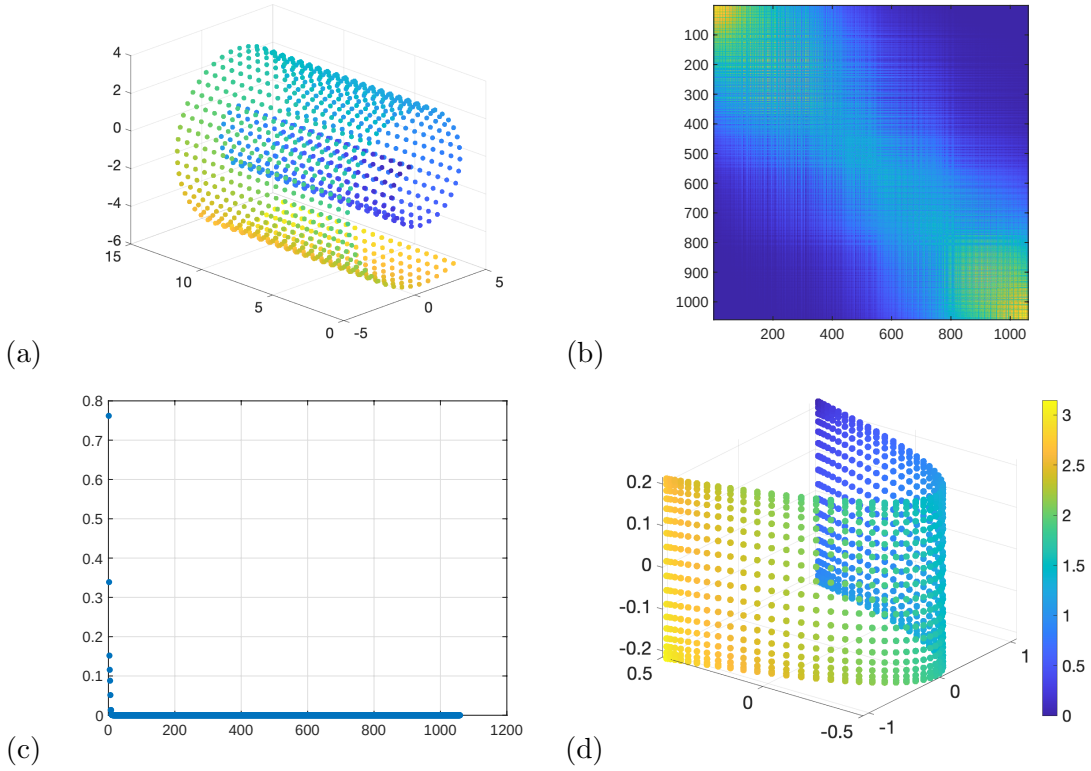

FIGURE 1. (a): The Swiss Roll dataset with points arranged into a quasi-uniform mesh. (b): The matrix P^t for $\epsilon = 0.7717928$, $t = 147$. (c): The absolute eigenvalues of the eigenvalues of P^t starting from $|\lambda_1|$. (d): Diffusion map to 3D.

Finally, we take the same Swiss Roll data that we used for the isomap experiments with Gaussian noise of standard deviation 0.8. With $\delta = 0.2$, we found

$$\epsilon = 2.104531, \quad t = 1705.$$

The results are shown in Fig. 3.

2.3.2. *Pacman.* Let us consider a data set consisting of 200 images depicting the Pacman. This example is similar to the one in this article[1]. Each image is 65×65 pixels either black (color = 0) or white (color = 255). The images differ from each other by the angle

[1]While this article offers a nice exposition, I do not recommend to rely on it as it contains a number of errors in important formulas. For example, Eqs. (6) and (7) contain errors, the comments following Eq. (9) are misleading, etc.

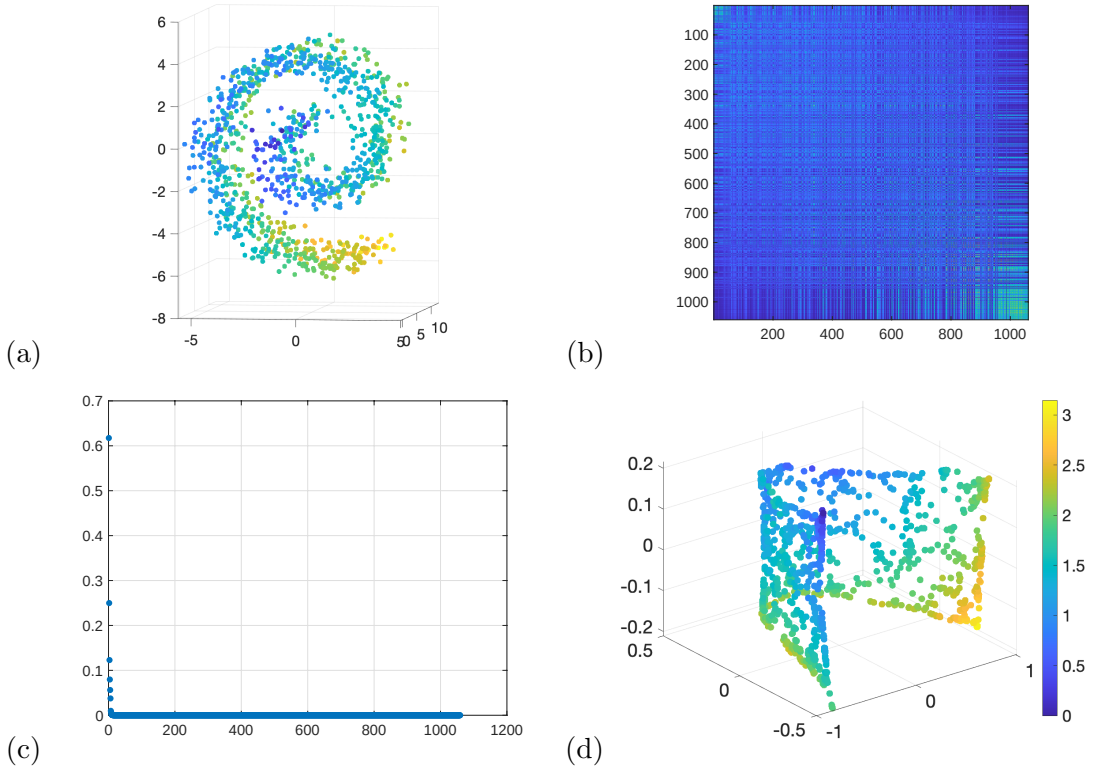

FIGURE 2. (a): The Swiss Roll dataset with points arranged into a quasi-uniform mesh and perturbed by Gaussian noise with standard deviation 0.4. (b): The matrix P^t for $\epsilon = 0.6108029$, $t = 300$. (c): The absolute eigenvalues of the eigenvalues of P^t starting from $|\lambda_1|$. (d): Diffusion map to 3D.

of rotation of the Pacman around the center of the image. The angles of rotation are

$$\alpha_i = \frac{2\pi i}{200}.$$

A sample of 20 such images is shown in Fig. 4(a). This dataset is naturally embedded into $\mathbb{R}^{65^2} = \mathbb{R}^{4225}$ space. Note that $D > N$ in this case. The PCA mapping into 3D applied to this dataset is shown in Fig. 4(b). The absolute eigenvalues and the embedding into 3D are shown in Figs. 4(c) and (d) respectively. Both the PCA and the diffusion map show that the set of images is well-approximated by a 1D manifold as we would expect.

2.3.3. *Cat-in-the-hat.* A similar example with a more complex image of the Cat-in-the-hat is shown in Fig. 5. Each image is 500×500. The double-loop formed by the mapped data

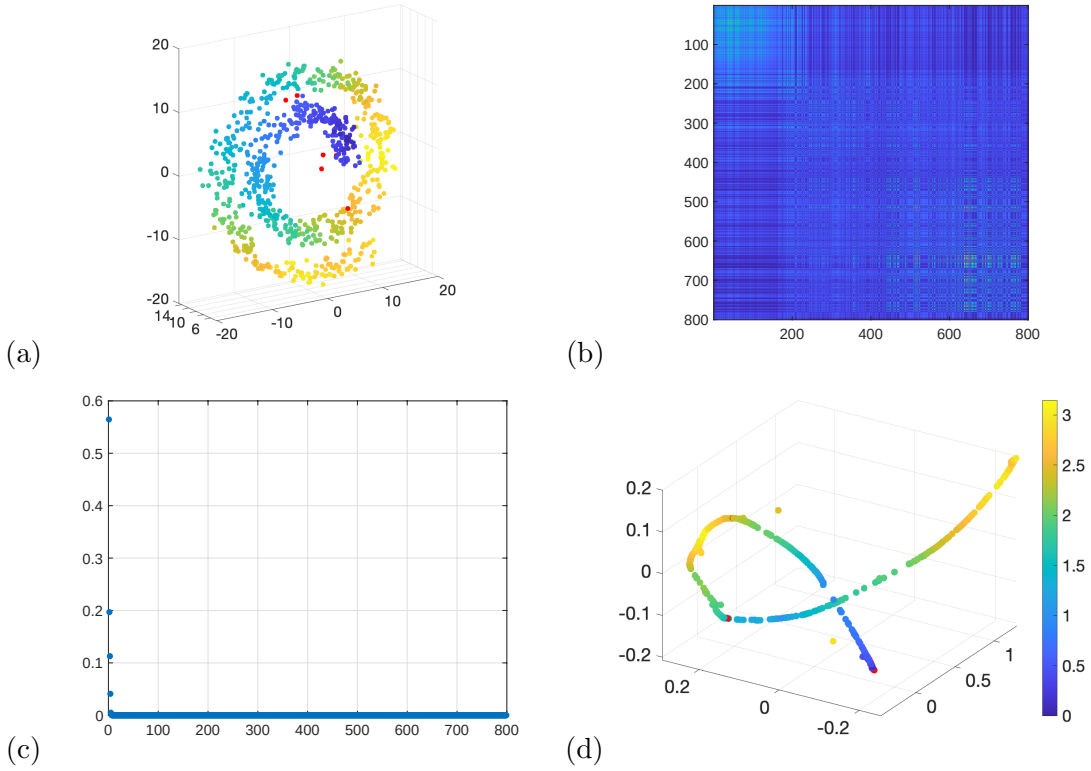

FIGURE 3. (a): The Swiss Roll dataset used for the experiments with isomap perturbed by Gaussian noise with standard deviation 0.8. (b): The matrix P^t for $\epsilon = 2.104531$, $t = 1705$. (c): The absolute eigenvalues of the eigenvalues of P^t starting from $|\lambda_1|$. (d): Diffusion map to 3D.

is caused by the fact that the image rotated by π is closer to the original image than those rotated by an angle between $\pi/6$ and $5\pi/6$.

2.4. **Removing the effect of nonuniform sampling.** Coifman and Lafon in [2] proposed a more general construction than the one described in Section 2.2: a whole family of diffusion maps parametrised by three parameters:

- ϵ, the local scale parameter,
- $t \in \mathbb{N}$, the number of steps done by the constructed Markov chain, and
- $\alpha \in \mathbb{R}$, a renormalization parameter.

The most interesting values of α are 0, $1/2$, and 1. The construction defined in Section 2.2 corresponds to $\alpha = 0$. The illustrative examples suggest that this construction handles noisy but uniformly distributed data pretty well, but nonuformity (even due to sampling from uniform distributions) leads to ugly embeddings like in Figs. 3(d) and 5(d).

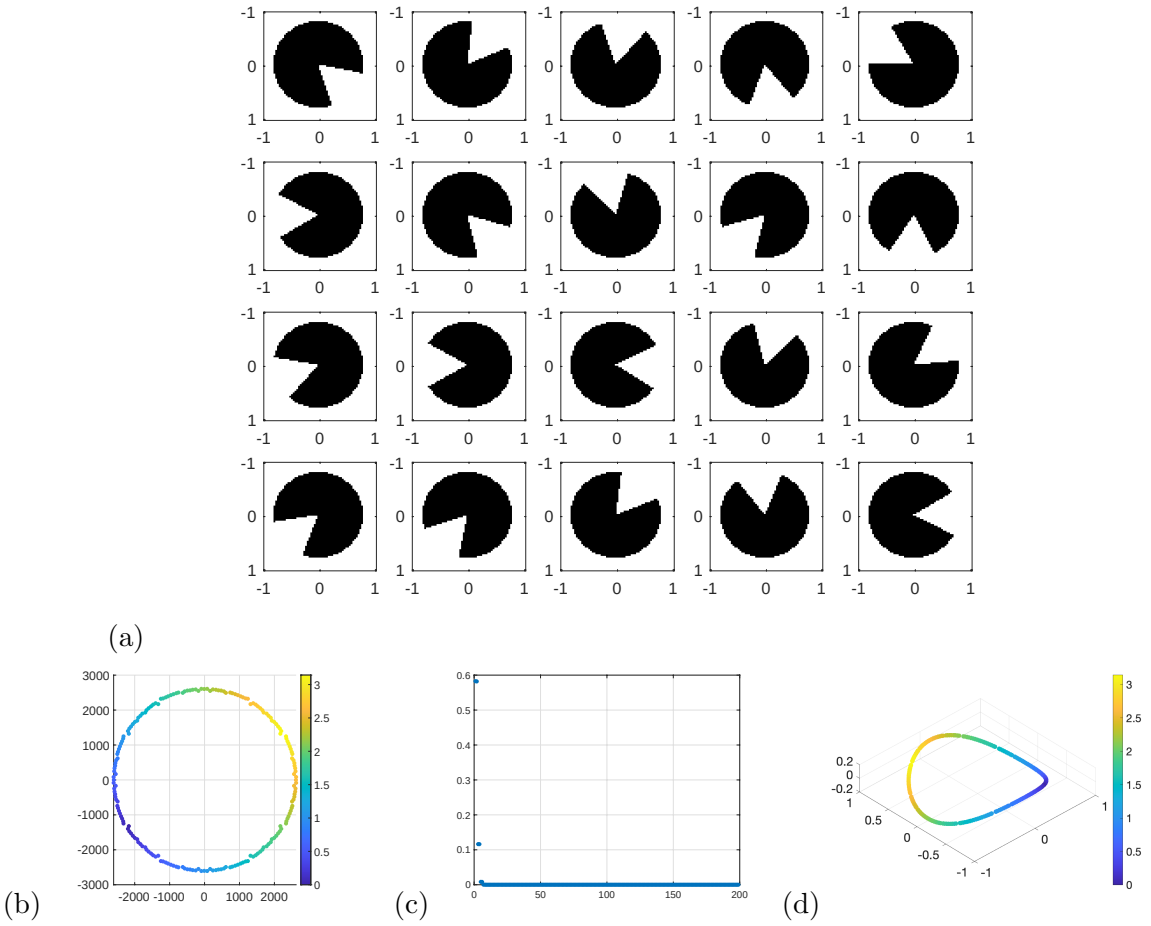

FIGURE 4. The dataset consists of 200 images of the Pacman rotated around the center of the image by angles $\alpha_i = 2\pi i/200$. (a): A sample of 20 data points. (b): The PCA mapping into 3D. (c): The absolute eigenvalues of the eigenvalues of P^t starting from $|\lambda_1|$. Here: $\delta = 0.5$, $\epsilon = 2335698$, $t = 187$. (d): Diffusion map to 3D.

The matrix P for a given $\alpha > 0$ is constructed as follows. Here we assume that the kernel is Gaussian while in [2] a more general kernel is allowed provided that it is rotation-invariant and has exponential decay.

- **Step 1.** Set a rotation-invariant kernel

$$k_\epsilon(x, y) = \exp\left[-\|x - y\|^2/\epsilon\right].$$

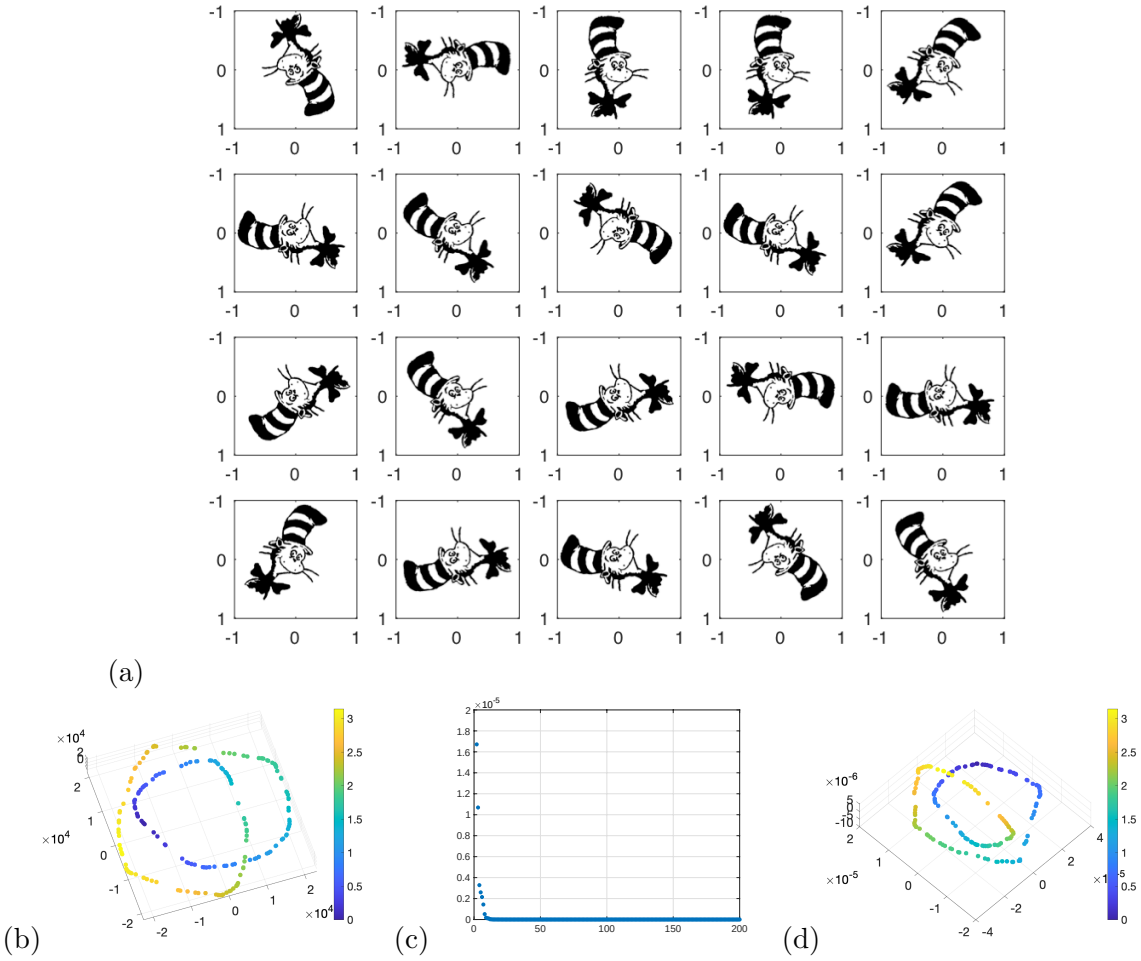

FIGURE 5. The dataset consists of 200 images of the Cat-in-the-hat rotated around the center of the image by angles uniformly distributed in $(0, 2\pi)$. (a): A sample of 20 data points. (b): The PCA mapping into 2D. (c): The absolute eigenvalues of the eigenvalues of P^t starting from $|\lambda_1|$. Here: $\delta = 0.2$, $\epsilon = 7.318159 \cdot 10^7$, $t = t$. (d): Diffusion map to 3D.

- **Step 2.** Calculate row sums $q_\epsilon(x) = \sum_y k_\epsilon(x, y)$ and right-normalize the kernel:

$$k_\epsilon^{(\alpha)}(x, y) = \frac{k_\epsilon(x, y)}{q_\epsilon^\alpha(y)}, \quad \text{or} \quad K = KQ^{-\alpha}. \tag{15}$$

- Calculate row sums
$$d_\epsilon^{(\alpha)}(x) = \sum_y k_\epsilon^{(\alpha)}(x,y)$$
and define the stochastic matrix by left normalization:

$$P_{\epsilon,\alpha} = \left[D_\epsilon^{(\alpha)}\right]^{-1} K_\epsilon^{(\alpha)}, \tag{16}$$

where
$$D_\epsilon^{(\alpha)} = \mathsf{diag}\left\{d_\epsilon^{(\alpha)}(x_1),\dots,d_\epsilon^{(\alpha)}(x_n)\right\}, \quad K_\epsilon^{(\alpha)} = (k_\epsilon^{(\alpha)}(x_i,x_j))_{i,j=1}^n.$$

The rest of the construction is the same as in Section 2.2 except for the formula for the embedding is modified to:

$$\Psi_t(x_i) = \begin{bmatrix} \lambda_1^{(1-\alpha)t} r_1(i) \\ \vdots \\ \lambda_{s(\delta,t)}^{(1-\alpha)t} r_{s(\delta,t)}(i) \end{bmatrix}. \tag{17}$$

This construction is given in [2] in a continuous setting. This allows one to consider the limit $\epsilon \to 0$. The main theoretical result in [2] is the theorem relating this limit to an α-dependent differential operator. It follows from this result that, *in order to compensate for a noninuform distribution of data, one needs to take* $\alpha = 1$.

3. Approximation of differential operators by means of diffusion maps

Suppose that the data points $x_i \in \mathbb{R}^d$, $i = 1,\dots,n$, are sampled from a pdf $q(x)$. Let $\mathcal{G}(x,y)$ be any function of two variables, and $f(y)$ be any function of one variable. Then we define the matrix $G = (\mathcal{G}(x_i,x_j))_{i,j=1}^n$ and the vector $\mathbf{f} = [f(x_1),\dots,f(x_n)]^\top$. Then by the strong law of large numbers we have

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^n G(x_i,x_j) f(x_j) = \mathbb{E}_y\left[\mathcal{G}(x_i,y)f(y)\right] = \int_{\mathbb{R}^d} \mathcal{G}(x_i,y)f(y)q(y)dy. \tag{18}$$

Now let us take $\mathcal{G}(x,y) = k_\epsilon(x,y) \equiv \exp(-\|x-y\|^2/\epsilon)$ and define

$$\mathcal{P}_\epsilon(x,y) = \frac{k_\epsilon(x,y)}{\int_{\mathbb{R}^d} k_\epsilon(x,y)q(y)dy} \quad \text{and} \quad P_\epsilon(x_i,x_j) = \frac{k_\epsilon(x_i,x_j)}{\sum_{j=1}^n k_\epsilon(x_i,x_j)}. \tag{19}$$

Then the matrix-vector product $P_\epsilon \mathbf{f}$ approximates the following integral:

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{q_\epsilon(x_i)} \sum_{j=1}^n k_\epsilon(x_i,x_j)f(x_j) &= \lim_{n\to\infty} \frac{\frac{1}{n}\sum_{j=1}^n k_\epsilon(x_i,x_j)f(x_j)}{\frac{1}{n}\sum_{j=1}^n k_\epsilon(x_i,x_j)} \\ &= \frac{\mathbb{E}_y\left[k(x_i,y)f(y)\right]}{\mathbb{E}_y\left[k(x_i,y)\right]} = \frac{\int_{\mathbb{R}^d} k_\epsilon(x_i,y)f(y)q(y)dy}{\int_{\mathbb{R}^d} k_\epsilon(x_i,y)q(y)dy} = \int_{\mathbb{R}^d} \mathcal{P}_\epsilon(x_i,y)f(y)q(y)dy \end{aligned} \tag{20}$$

This shows us how we can relate integral operators and matrix operators as the number of sample points tends to infinity. It is easier to manipulate integrals than sums as some

integrals we will encounter can be calculated analytically and give a nice and compact result. Therefore, we will first obtain the desired approximations for differential operators in terms of integral operators, and then switch to discrete sample points and matrix operators.

3.1. **Definitions and the main theorem.**

- Let $\alpha \in \mathbb{R}$ be a parameter. We will be mostly interested in $\alpha = 0$, $\alpha = 1/2$, and $\alpha = 1$ and these values lead familiar operators as we will see later. We define the kernel function

$$k_\epsilon(x, y) = e^{-\|x-y\|^2/\epsilon}, \quad x, y \in \mathbb{R}^d. \tag{21}$$

- Let $q(y)$ be a pdf. We define

$$q_\epsilon(x) = \int_{\mathbb{R}^d} k_\epsilon(x, y) q(y) dy. \tag{22}$$

Then we right-normalize the kernel depending on the parameter α:

$$k_{\epsilon,\alpha}(x, y) = \frac{k_\epsilon(x, y)}{q_\epsilon^\alpha(y)}. \tag{23}$$

Then we set

$$d_{\epsilon,\alpha}(x) = \int_{\mathbb{R}^d} k_{\epsilon,\alpha}(x, y) q(y) dy \tag{24}$$

and define the transition kernel

$$p_{\epsilon,\alpha}(x, y) = \frac{k_{\epsilon,\alpha}(x, y)}{d_{\epsilon,\alpha}(x)}. \tag{25}$$

- Next, we define the following integral operator:

$$P_{\epsilon,\alpha} f(x) = \int_{\mathbb{R}^d} p_{\epsilon,\alpha}(x, y) f(y) q(y) dy. \tag{26}$$

- Finally, we define the generator

$$L_{\epsilon,\alpha} f(x) := \frac{P_{\epsilon,\alpha} f(x) - f(x)}{\epsilon}. \tag{27}$$

The main theorem proven in [2] shows what this generator is in the limit $\epsilon \to 0$.

Theorem 1. *The generator defined in (27) approximates the following differential operator:*

$$\lim_{\epsilon \to 0} L_{\epsilon,\alpha} f(x) := \frac{\Delta\left(f q^{1-\alpha}\right) - f \Delta\left(q^{1-\alpha}\right)}{4 q^{1-\alpha}}. \tag{28}$$

Prior to proving this theorem, let us calculate the right-hand side of (28) for $\alpha = 0$, $1/2$, and 1.

$\alpha = 0$**:**

$$\begin{aligned}\frac{\Delta(fq) - f\Delta(q)}{4q} &= \frac{\nabla\cdot\nabla(fq) - f\Delta(q)}{4q} \\ &= \frac{q\Delta f + 2\nabla f\cdot\nabla q + f\Delta q - f\Delta(q)}{q} = \frac{1}{4}\left(\Delta f + 2\frac{\nabla f\cdot\nabla q}{q}\right).\end{aligned} \tag{29}$$

$\alpha = 1/2$**:** Suppose that q is the Gibbs density, i.e., $q(y) = Z^{-1}\exp(-\beta U(y))$. Then

$$\frac{\Delta\left(fe^{-\beta U/2}\right) - f\Delta\left(e^{-\beta U/2}\right)}{4e^{-\beta U/2}} = \frac{1}{4}\left(-\beta\nabla U\cdot\nabla f + \Delta f\right) = \frac{\beta}{4}Lf, \tag{30}$$

where L is the generator for the SDE

$$dx = -\nabla U(x)dt + \sqrt{2\beta^{-1}}dw.$$

$\alpha = 1$**:** In this case, one can readily see that the right-hand side of (28) is $1/4\Delta f$.

3.1.1. *Proof of Theorem 1.* The proof of Theorem 1 given below is a simplification of the proof in Appendix B in [2] for the case where the manifold $\mathcal{M}$ and the space X are merely $\mathbb{R}^d$. The proof makes use of the following

Lemma 1. *Let G_ϵ be an integral operator defined by*

$$G_\epsilon f(x) = \int_{\mathbb{R}^d} k_\epsilon(x,y)f(y)dy. \tag{31}$$

Then for every four times continuously differentiable function $f(x)$ that grows not faster than a polynomial as $\|x\|\to\infty$, the following expansion in ϵ takes place:

$$G_\epsilon f(x) = (\pi\epsilon)^{d/2}\left(f(x) + \frac{\epsilon}{4}\Delta f + O(\epsilon^2)\right). \tag{32}$$

Proof. To obtain (32), we collect the following building blocks:

- We will use the following Gaussian integrals:

$$\begin{aligned}&\int_{\mathbb{R}^d} e^{-\|z\|^2/\epsilon}dz = (\pi\epsilon)^{d/2};\\ &\int_{\mathbb{R}^d} z_i e^{-\|z\|^2/\epsilon}dz = 0, \quad i = 1,\dots,d;\\ &\int_{\mathbb{R}^d} z_i^2 e^{-\|z\|^2/\epsilon}dz = \frac{\epsilon}{2}(\pi\epsilon)^{d/2}, \quad i = 1,\dots,d;\\ &\int_{\mathbb{R}^d} z_i^3 e^{-\|z\|^2/\epsilon}dz = 0, \quad i = 1,\dots,d;\\ &\int_{\mathbb{R}^d} z_i^4 e^{-\|z\|^2/\epsilon}dz = \frac{3\epsilon^2}{4}(\pi\epsilon)^{d/2}, \quad i = 1,\dots,d.\end{aligned}$$

- We do the variable change: $z := y - x$. Writing $y = x + z$, we Taylor-expand $f(x+z)$ around $f(x)$:

$$f(x+z) = f(x) + \nabla f(x)^\top z + \frac{1}{2}z^\top\nabla\nabla f(x)z + r_3(z) + r_4(z),$$

where $r_3(z)$ is a homogeneous third degree polynomial in z, and $r_4(z)$ is $O(z^4)$ in a small ball around 0.

- We assume that ϵ is small. Consider integrals of the form

$$J := \int_{\mathbb{R}^d \setminus B_{2\epsilon^\mu}(0)} e^{-\|z\|^2/\epsilon} \phi(z) dz,$$

where $\mu \in (0, 1/2)$, $B_{2\epsilon^\mu}(0)$ is a ball of radius $2\epsilon^\mu$ centered at 0, and $|\phi(z)| \leq A + B\|z\|^k$, $k > 0$. Note that this estimate is valid for $r_4(z)$. Then switching to the d-dimensional generalization of polar coordinates (see e.g. this note), we obtain the following estimate:

$$|J| \leq C \int_{2\epsilon^\mu}^{\infty} r^k r^{d-1} e^{-r^2/\epsilon} dr \leq C\mathcal{R}(\epsilon) e^{-4\epsilon^{2\mu-1}},$$

where $\mathcal{R}$ is a finite linear combination of powers of ϵ, positive and/or negative. Since $2\mu - 1 < 0$, $|J|$ tends to zero faster than any power of ϵ as $\epsilon \to 0$.

Now we calculate:

$$\begin{aligned}
G_\epsilon f(x) &= \int_{\mathbb{R}^d} e^{-\|z\|^2/\epsilon} f(x+z) dz \\
&= \int_{\mathbb{R}^d} e^{-\|z\|^2/\epsilon} \left[f(x) + \nabla f(x)^\top z + \frac{1}{2} z^\top \nabla\nabla f(x) z + r_3(z) + r_4(z) \right] dz \\
&= (\pi\epsilon)^{d/2} \left[f(x) + \frac{\epsilon}{4} \Delta f(x) \right] + \int_{B_{2\epsilon^\mu}(0)} e^{-\|z\|^2/\epsilon} r_4(z) dz + \int_{\mathbb{R}^d \setminus B_{2\epsilon^\mu}(0)} e^{-\|z\|^2/\epsilon} r_4(z) dz \\
&= (\pi\epsilon)^{d/2} \left[f(x) + \frac{\epsilon}{4} \Delta f(x) + O(\epsilon^2) \right] + \int_{\mathbb{R}^d \setminus B_{2\epsilon^\mu}(0)} e^{-\|z\|^2/\epsilon} \phi(z) dz,
\end{aligned}$$

where $|\phi(z)| < A + B\|z\|^k$ is the difference between $r_4(z)$ and the fourth-order term in the Taylor expansion of $f(x+z)$ at $z = 0$. The last integral decays faster than any power of ϵ as $\epsilon \to 0$. Hence, (32) readily follows. □

Proof. (Theorem 1.) Lemma 1 implies that

$$q_\epsilon(x) = \int_{\mathbb{R}^d} k_\epsilon(x,y) q(y) dy = (\pi\epsilon)^{d/2} \left[q(x) + \frac{\epsilon}{4} \Delta q(x) + O(\epsilon^2) \right]. \tag{33}$$

Therefore,

$$q_\epsilon^{-\alpha}(x) = (\pi\epsilon)^{-\alpha d/2} q^{-\alpha}(x) \left[1 - \frac{\alpha\epsilon}{4} \frac{\Delta q(x)}{q(x)} + O(\epsilon^2) \right]. \tag{34}$$

Then

$$\begin{aligned} K_{\epsilon,\alpha}f(x) &= \int_{\mathbb{R}^d} \frac{k_\epsilon(x,y)}{q_\epsilon^\alpha(y)} q(y)f(y)dy \\ &= \int_{\mathbb{R}^d} k_\epsilon(x,y)\left[q_\epsilon^{-\alpha}(y)q(y)f(y)\right]dy \\ &= (\pi\epsilon)^{-\alpha d/2}\int_{\mathbb{R}^d} k_\epsilon(x,y)\left[q^{1-\alpha}(y)f(y)\right]\left(1-\frac{\alpha\epsilon}{4}\frac{\Delta q(y)}{q(y)}+O(\epsilon^2)\right)dy \\ &= (\pi\epsilon)^{d/2-\alpha d/2}\left[q^{1-\alpha}f+\frac{\epsilon}{4}\Delta\left(q^{1-\alpha}f\right)-q^{1-\alpha}f\frac{\alpha\epsilon}{4}\frac{\Delta q}{q}+O(\epsilon^2)\right] \end{aligned} \tag{35}$$

Observing that $d_{\epsilon,\alpha}(x) = K_{\epsilon,\alpha}1$, i.e., we need to use $f \equiv 1$ to get $d_{\epsilon,\alpha}(x)$, we calculate the operator $P_{\epsilon,\alpha}$:

$$\begin{aligned} P_{\epsilon,\alpha}f(x) &= \int_{\mathbb{R}^d} \frac{k_{\epsilon,\alpha}(x,y)}{d_{\epsilon,\alpha}(x)} f(y)q(y)dy \\ &= \frac{q^{1-\alpha}f+\frac{\epsilon}{4}\Delta\left(q^{1-\alpha}f\right)-q^{1-\alpha}f\frac{\alpha\epsilon}{4}\frac{\Delta q}{q}+O(\epsilon^2)}{q^{1-\alpha}+\frac{\epsilon}{4}\Delta\left(q^{1-\alpha}\right)-q^{1-\alpha}\frac{\alpha\epsilon}{4}\frac{\Delta q}{q}+O(\epsilon^2)} \\ &= \left[f+\frac{\epsilon}{4}\left(\frac{\Delta(q^{1-\alpha}f)}{q^{1-\alpha}}-\alpha f\frac{\Delta q}{q}\right)+O(\epsilon^2)\right]\left[1-\frac{\epsilon}{4}\left(\frac{\Delta(q^{1-\alpha})}{q^{1-\alpha}}-\alpha\frac{\Delta q}{q}\right)+O(\epsilon^2)\right] \\ &= f+\frac{\epsilon}{4}\left(\frac{\Delta(q^{1-\alpha}f)}{q^{1-\alpha}}-\alpha f\frac{\Delta q}{q}-f\frac{\Delta(q^{1-\alpha})}{q^{1-\alpha}}+\alpha f\frac{\Delta q}{q}\right)+O(\epsilon^2) \\ &= f+\frac{\epsilon}{4}\left(\frac{\Delta(q^{1-\alpha}f)}{q^{1-\alpha}}-f\frac{\Delta(q^{1-\alpha})}{q^{1-\alpha}}\right)+O(\epsilon^2) \end{aligned} \tag{36}$$

Finally, we compute the operator $L_{\epsilon,\alpha}$ and take the limit $\epsilon \to 0$:

$$\lim_{\epsilon\to 0}\frac{P_{\epsilon,\alpha}f(x)-f(x)}{\epsilon} = \frac{1}{4}\left(\frac{\Delta(q^{1-\alpha}f)}{q^{1-\alpha}}-f\frac{\Delta(q^{1-\alpha})}{q^{1-\alpha}}\right). \tag{37}$$

□

3.2. **Calculation of the committor function on a discrete dataset.** Suppose we have generated a set of data points $\{x_i\}_{i=1}^n$ by integrating the SDE

$$dx = -U(x)dt + \sqrt{2\beta^{-1}}dw$$

by means of the Euler-Maruyama method. We assume that the trajectory is long enough so that the sample points are distributed according to the invariant pdf which is the Gibbs measure $q(x) = Z^{-1}\exp(-\beta U(x))$. We choose an appropriate value of ϵ, define a kernel

$$k_\epsilon(x,y) = \exp(-\|x-y\|^2/\epsilon)$$

and the corresponding kernel matrix

$$K(i,j) = \exp(-\|x_i - x_j\|^2/\epsilon), \quad i,j = 1,\ldots,n.$$

Then we compute the vector

$$q(i) = \sum_{j=1}^{n} K(i,j), \quad i = 1, \dots, n,$$

and form the matrix

$$Q := \mathsf{diag}\{q(1), \dots, q(n)\}.$$

Next, we right-normalize the kernel corresponding to $\alpha = 1/2$:

$$K_{\epsilon,1/2} = KQ^{-1/2}.$$

Next, we normalize the rows of the new kernel $K_{\epsilon,1/2}$ and obtain a stochastic matrix

$$P_{\epsilon,1/2} := D^{-1}K_{\epsilon,1/2}, \quad \text{where} \quad D = \mathsf{diag}\left\{\sum_{j=1}^{n} K_{\epsilon,1/2}(1,j), \dots, \sum_{j=1}^{n} K_{\epsilon,1/2}(n,j)\right\}.$$

Finally, we pick sets $A, B \subset \mathcal{I}$, $A \cup B = \emptyset$, $\mathcal{I} := \{1, \dots, n\}$, and setup the committor equation:

$$\sum_{j=1}^{n} P_{\epsilon,1/2}(i,j)c(j) = c(i), \quad i \in \mathcal{I}\backslash(A \cup B), \tag{38}$$

$$c(i) = 0,\ i \in A, \quad c(i) = 1,\ i \in B. \tag{39}$$

Let us explain (38). Taking the limit $n \to \infty$ and applying the strong law of large numbers, we get:

$$\begin{aligned}
\lim_{n\to\infty} \sum_{j=1}^{n} P_{\epsilon,1/2}(i,j)c(j) - c(i) &= \lim_{n\to\infty} \frac{\frac{1}{n}\sum_{j=1}^{n} K_{\epsilon,1/2}(i,j)c(j)}{\frac{1}{n}\sum_{j=1}^{n} K_{\epsilon,1/2}(i,j)} - c(i) \\
&= \lim_{n\to\infty} \frac{\frac{1}{n}\sum_{j=1}^{n} \frac{K(i,j)c(j)}{\frac{1}{n}\left[\sum_{l=1}^{n} K(j,l)\right]^{1/2}}}{\frac{1}{n}\sum_{j=1}^{n} \frac{K(i,j)}{\frac{1}{n}\left[\sum_{l=1}^{n} k_\epsilon(j,l)\right]^{1/2}}} - c(i) \\
&= \frac{\mathbb{E}_y\left[\frac{k_\epsilon(x_i,y)c(y)}{(\mathbb{E}_z[k_\epsilon(y,z)])^{1/2}}\right]}{\mathbb{E}_y\left[\frac{k_\epsilon(x_i,y)}{(\mathbb{E}_z[k_\epsilon(y,z)])^{1/2}}\right]} - c(x_i).
\end{aligned}$$

Now we recall that

$$\mathbb{E}_z[k_\epsilon(x_i,z)] = \int_{\mathbb{R}^d} k_\epsilon(x_i,z)q(z)dz \equiv q_\epsilon(x_i),$$
$$\mathbb{E}_z[k_\epsilon(y,z)] = \int_{\mathbb{R}^d} k_\epsilon(y,z)q(z)dz \equiv q_\epsilon(y),$$
$$k_{\epsilon,1/2}(x_i,y) = \frac{k_\epsilon(x_i,y)}{q_\epsilon^{1/2}(y)},$$
$$\mathbb{E}_y[k_{\epsilon,1/2}(x_i,y)c(y)] = \int_{\mathbb{R}^d} k_{\epsilon,1/2}(x_i,y)c(y)q(y)dy,$$
$$\mathbb{E}_y[k_{\epsilon,1/2}(x_i,y)] = \int_{\mathbb{R}^d} k_{\epsilon,1/2}(x_i,y)q(y)dy.$$

Therefore, comparing these equations with the definitions in Section 3.1, we see that

$$\lim_{n\to\infty}\sum_{j=1}^{n} P_{\epsilon,1/2}(i,j)c(j) - c(i) = \int_{\mathbb{R}^d} p_{\epsilon,1/2}(x_i,y)c(y)q(y)dy - c(x_i) = \epsilon L_{\epsilon,1/2}c(x_i).$$

Dividing by ϵ and taking limit $\epsilon \to 0$, we get $1/4\beta Lc(x_i) = 0$, $x_i \in \mathcal{I}\backslash(A \cup B)$, which, together with the boundary conditions (39), is the boundary-value problem for the committor function $c(x)$.

References

[1] I. Jolliffe, *Principal component analysis*. Springer-Verlag New York, 1986.

[2] R. Coifman and S. Lafon, "Diffusion maps," *Appl. Comput. Harmon. Anal.*, vol. 21, pp. 5–30, 2006.

[3] A. L. Ferguson, A. Z. Panagiotopoulosa, P. G. Debenedettia, and K. I. G., "Systematic determination of order parameters for chain dynamics using diffusion maps," *PNAS*, vol. 107, no. 31, pp. 13597–13602, 2010.

[4] S. B. Kim, C. J. Dsilva, I. G. Kevrekidis, and P. G. Debenedetti, "Systematic characterization of protein folding pathways using diffusion maps: Application to trp-cage miniprotein," *J. Chem. Phys*, vol. 142, p. 085101, 2019.

[5] T. Berry, D. Giannakis, and J. Harlim, "Nonparametric forecasting of low-dimensional dynamical systems," *Physical Review E*, vol. 91, no. 3, p. 032915, 2015.

[6] T. Berry and J. Harlim, "Variable bandwidth diffusion kernels," *Applied and Computational Harmonic Analysis*, vol. 40, no. 1, pp. 68–96, 2016.

[7] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput*, vol. 13, pp. 1373–1397, 2003.