DATA-DRIVEN EFFICIENT SOLVERS AND PREDICTIONS OF
CONFORMATIONAL TRANSITIONS FOR LANGEVIN DYNAMICS ON
MANIFOLD IN HIGH DIMENSIONS

YUAN GAO, JIAN-GUO LIU, AND NAN WU

ABSTRACT. We work on dynamic problems with collected data \{x_i\} that distributed on a manifold \(\mathcal{M} \subset \mathbb{R}^p\). Through the diffusion map, we first learn the reaction coordinates \{y_i\} \subset \mathcal{N} where \(\mathcal{N}\) is a manifold isometrically embedded into an Euclidean space \(\mathbb{R}^\ell\) for \(\ell \ll p\). The reaction coordinates enable us to obtain an efficient approximation for the dynamics described by a Fokker-Planck equation on the manifold \(\mathcal{N}\). By using the reaction coordinates, we propose an implementable, unconditionally stable, data-driven upwind scheme which automatically incorporates the manifold structure of \(\mathcal{N}\). Furthermore, we provide a weighted \(L^2\) convergence analysis of the upwind scheme to the Fokker-Planck equation. The proposed upwind scheme leads to a Markov chain with transition probability between the nearest neighbor points. We can benefit from such property to directly conduct manifold-related computations such as finding the optimal coarse-grained network and the minimal energy path that represents chemical reactions or conformational changes. To establish the Fokker-Planck equation, we need to acquire information about the equilibrium potential of the physical system on \(\mathcal{N}\). Hence, we apply a Gaussian Process regression algorithm to generate equilibrium potential for a new physical system with new parameters. Combining with the proposed upwind scheme, we can calculate the trajectory of the Fokker-Planck equation on \(\mathcal{N}\) based on the generated equilibrium potential. Finally, we develop an algorithm to pullback the trajectory to the original high dimensional space as a generative data for the new physical system.

1. Introduction

1.1. Problem set up and goals. We study a complex chemical, biological or physical system \(P\) which can be described by \(p\)-dimensional variables in \(\mathbb{R}^p\) with \(p \gg 1\). We assume the underlying structure of the system \(P\) is an unknown \(d\) dimensional closed Riemannian submanifold \(\mathcal{M}\) of \(\mathbb{R}^p\) \[10, 9\]. The essential physical motions in the system \(P\) are those slow time scale structural changes or conformational changes rather than the fast time scale motions such as vibrations. Therefore, despite the high dimensionality of \(P\) in practice, we can find some intrinsic low dimensional variables, called reaction coordinates, to represent those essential structural or conformational changes in a low dimensional space \[10, 9, 36\]. Formally, the reaction coordinates should be a smooth embedding \(y = \Phi(x) : \mathcal{M} \hookrightarrow \mathbb{R}^\ell\) with \(\ell \ll p\) in order to preserve the topological structure of the underlying manifold. Then, \(\mathcal{N} = \Phi(\mathcal{M})\) is a submanifold of \(\mathbb{R}^\ell\) with the metric induced by the Euclidean metric of \(\mathbb{R}^\ell\). The reaction coordinates can be realized through the nonlinear dimension reduction algorithms. Suppose \(\{x_i\}_{i=1}^n\) are \(n\) data points sampled from \(\mathcal{M} \subset \mathbb{R}^p\). Without knowing

Date: May 27, 2020.

Key words and phrases. Diffusion map, reaction coordinates, Voronoi tessellation, unconditionally stable explicit upwind scheme, Kullback-Leibler divergence, exponential convergence, generative distribution with hidden parameters.
all the information of the system $P$ (i.e., without knowing the manifold $\mathcal{M}$), a nonlinear dimension reduction algorithm constructs an embedding $\Phi$ by using the coordinates of $\{x_i\}$ in $\mathbb{R}^p$, so that we can present the high dimensional data $\{x_i\} \subset \mathcal{M} \subset \mathbb{R}^p$ as $y_i = \Phi(x_i) \subset \mathcal{N} \subset \mathbb{R}^\ell$ in the low dimensional space.

When we study the dynamics on a submanifold of an Euclidean space, the measurement errors of the information on a submanifold in a high dimensional ambient space has much larger impact on the accuracy of the solution than those on a submanifold in a low dimensional ambient space. Thus, proposing a calculation for those slow time scale dynamics based on the dataset $\{y_i\} \subset \mathcal{N}$ is important and enables us to handle the examples such as finding the optimal coarse-grained network and finding the minimal energy path for chemical reactions on $\mathcal{N}$. In contrast, solving dynamic systems on the original manifold $\mathcal{M} \subset \mathbb{R}^p$ may lead to inaccuracy, for instance, for the case of solving PDE on the well-known Swiss Roll manifold. Another advantage of describing slow time scale dynamics using reaction coordinates is that the system converges to its equilibrium. Indeed, the resulting low dimensional reaction coordinates are not sensitive to the invariant measure so that they can be used to reproduce new physical system with new invariant measure [10, 14]; see Theorem 2.6.

We will focus on the data collected from a physical system $P$ which is assumed to be described by a continuous strong Markov process on $\mathcal{M} \subset \mathbb{R}^p$. Therefore, the generator of this process must be a second order elliptic operator [24]. Especially, if the noise is additive and is described by Brownian motion, the leading term of the generator is a Laplace operator. We consider Brownian motion on manifold $\mathcal{M}$ (resp. $\mathcal{N}$). To study the stochastic differential equation (SDE) on manifold, we first write down the Brownian motion on $\mathcal{M}$ symbolically

\begin{equation}
\frac{dx_t}{dt} = \sum_{i=1}^d \tau_i^M(x_t) \otimes \tau_i^M(x_t) \circ dB_t,
\end{equation}

where $\circ$ is understood in the Stratonovich sense in the stochastic integral, $B_t$ is $p$-dimensional Brownian motion and $\{\tau_i^M; 1 \leq i \leq d\}$ are orthonormal basis of tangent plane $T_{x_t} \mathcal{M}$. Besides the Brownian motion on $\mathcal{M}$, we are particularly interested in physical systems with a drift determined by some potentials $U$ on $\mathcal{M}$ depending on the system. Let $k$ be the Boltzmann constant and $T$ be the temperature. Then the SDE on $\mathcal{M}$ is

\begin{equation}
\frac{dx_t}{dt} = -\nabla_M U(x_t) dt + \sqrt{2kT} \sum_{i=1}^d \tau_i^M(x_t) \otimes \tau_i^M(x_t) \circ dB_t,
\end{equation}

where $\nabla_M := \sum_{i=1}^d \tau_i^M \nabla_M \tau_i^M = \sum_{i=1}^d \tau_i^M \otimes \tau_i^M \nabla$ is surface gradient and $\nabla_{\tau_i^M} = \tau_i^M \cdot \nabla$ is the tangential derivative in the direction of $\tau_i^M$. The equilibrium state of this system, also known as invariant probability density measure, is $p_\infty(x) \propto e^{-\frac{U(x)}{kT}}$, $x \in \mathcal{M}$. Notice the diffeomorphism $\Phi$ from $\mathcal{M}$ to $\mathcal{N}$ induce a map $\Phi_*$ from the space, denoted as $\Gamma(T\mathcal{M})$, of smooth vector fields on $\mathcal{M}$ to the space, denoted as $\Gamma(T\mathcal{N})$, of smooth vector fields on $\mathcal{N}$ such that for any $f \in C^\infty(\mathcal{N})$ and $V \in \Gamma(T\mathcal{M})$

\begin{equation}
(\Phi_* V) f(y) = V(f \circ \Phi)(x), \quad y = \Phi(x).
\end{equation}
Thus the Stratonovich formulation transform consistently under diffeomorphism $\Phi$. Notice $\tau^N_t \in \mathbb{R}^\ell$ are defined by the induced map $\Phi_*$ and $B_t$ is $\ell$-dimensional Brownian motion. Therefore, instead of considering $U$ on $\mathcal{M}$ directly, we consider the SDE on $\mathcal{N}$

$$\begin{equation}
\text{d}y_t = -\nabla_N U_N(y_t) \text{d}t + \sqrt{2kT} \sum_{i=1}^d \tau^N_i(y_t) \otimes \tau^N_i(y_t) \circ \text{d}B_t, \tag{1.4}
\end{equation}$$

where $\nabla_N := \sum_{i=1}^d \tau^N_i \nabla_{\tau^N_i} = \sum_{i=1}^d \tau^N_i \otimes \tau^N_i \nabla$ is surface gradient and $\nabla_{\tau^N_i} = \tau^N_i \cdot \nabla$ is the tangential derivative in the direction of $\tau^N_i$.

By Ito’s formula, the SDE \([1.4]\) gives the following Fokker-Planck equation, which is the governing equation for the density $\rho^N(y)$ of $\mathcal{Y}$,

$$\begin{equation}
\partial_t \rho^N_t = \text{div}_N(kT \nabla_N \rho^N_t + \rho^N_t \nabla_N U_N) =: F\rho^N_t, \tag{1.5}
\end{equation}$$

where $\text{div}_N$ is the surface divergence defined as $\text{div}_N \xi = \sum_{i=1}^d \tau^N_i \cdot \nabla_{\tau^N_i} \xi$. One equivalent form of \([1.5]\) is the relative entropy formulation

$$\begin{equation}
\partial_t \rho^N_t = kT \text{div}_N \left( e^{-\frac{U_N}{kT}} \nabla_N (\rho^N_t e^{\frac{U_N}{kT}}) \right) = F\rho^N_t. \tag{1.6}
\end{equation}$$

Now, we are ready to explain the goals and the procedures of this paper. Since the conformational changes with slow time scale are more essential and more stable than the vibrations with fast time scale, the first goal is to find the reaction coordinates $y = \Phi(x), x \in \mathcal{M}$. After learning the reaction coordinates $y$, one can assume that the equilibrium potential $U_N(y)$ is known in the physical system so that we can use it to establish \([1.5]\). However, in a more realistic case, the equilibrium potential $U_N(y)$ may also depends on some other parameters $\theta$ of the physical system such as the temperature, the time window, or the parameters indicating electronic, magnetic, gravitational, drug or catalysis effect. So the second goal is to predict the equilibrium potential $U_N(y; \theta^{(2)})$ for the unlabeled parameters $\theta^{(2)}$ from the gathered information about $U_N(y; \theta^{(1)})$ over the labeled parameters $\theta^{(1)}$. Next, given $U_N(y; \theta^{(2)})$, we need to recover the dynamics on $\mathcal{M}$ in the original space. More precisely, we want to first find out the trajectory $\rho^N_t(y)$ driving any initial density $\rho^N_0$ to the new invariant measure $\rho^N_{\infty}(y) \propto e^{-\frac{U_N(y)}{kT}} := e^{-\frac{U_N(y; \theta^{(2)})}{kT}}$ by solving the Fokker-Planck equation \([1.5]\) on $\mathcal{N}$. Then, we can pull back the density $\rho^N_t(y)$ to a density $\rho^M_t(x)$ on $\mathcal{M}$. Last but not least, based on the dataset $\{y_i\} \subset \mathcal{N}$ reflecting the manifold structure and an approximated transition probability associated with the Fokker-Planck equation, we can efficiently conduct implementable computations on the manifold for finding the optimal cluster-cluster coarse-grained network, c.f. \([11, 15, 26, 29]\) and for finding the minimal energy path of chemical/biological reactions, c.f. \([19, 17, 33, 18, 36]\). In summary, we want to

(I) Learn the reaction coordinate $y$ for the $d$-dimensional manifold $\mathcal{N} \subset \mathbb{R}^\ell$;

(II) Use the information for the equilibrium potential $U_N(y, \theta^{(1)})$ suggested by collected data to reproduce new equilibrium potential $U^*_N(y) = U_N(y, \theta^{(2)})$;

(III) For the new equilibrium potential $U^*_N(y)$, find out the trajectory $\rho^N_t$ driving any initial density $\rho^N_0$, which may come from the invariant measure for the old physical system, to the new invariant measure $\rho^N_{\infty}(y) \propto e^{-\frac{U^*_N(y)}{kT}}$;

(IV) With the trajectory $\rho^N_t$ in lower dimension, we want to recover the dynamics (trajectory) of $x$ in the original real space $\mathcal{M} \subset \mathbb{R}^p$;
(V) Find the minimal energy path (committor function), between two local minima of the equilibrium potential $U_N(y)$, which represents the slow time scale transition from one stable state to another stable state.

1.2. Practical difficulties and mathematical implementations. The first difficulty is that we are not able to acquire all the information about the system $P$. Hence, we sample $n$ points $\{x_i\}_{i=1}^n$ from $M$ based on a density function on $M$ with lower and upper bounds. To achieve goal (I), in Section 2, we apply the Diffusion map algorithm [10] on $\{x_i\}_{i=1}^n$ to find the reaction coordinates so that we have $\{y_i = \Phi(x_i)\}_{i=1}^n \subset N = \Phi(M) \subset \mathbb{R}^\ell$. Note that $\{y_i\}$ can also be regarded as the samples based on a density function on $N$ with lower and upper bounds.

Next, we focus on the goal (III) with a given equilibrium potential $W(y) := U_N^*(y)$. To find the trajectory $\rho^N_t$, we need to solve the Fokker-Planck equation (1.5) on manifold $N \subset \mathbb{R}^\ell$ with potential $W(y)$. Our method uses the data points $\{y_i\} \subset N$ to construct a discrete approximation of the Fokker-Planck equation (1.5). It is proved that the data points are well-distributed on $N$ whenever the points are sampled based on a density function with lower and upper bounds [42, 30]. Hence, we can construct a “regularly shaped” Voronoi tessellation on $N$ from $\{y_i\} \subset N$. With the help of such Voronoi tessellation, we introduce an upwind scheme by applying relative entropy formulation and finite volume method to (1.5). The upwind scheme is associated with an assignment of the transition probability for an approximated Markov process on $\{y_i\}$ between the nearest neighbor points. Due to the manifold structure of $N$, each Voronoi cell can be approximated by a polygon in a tangent space of $N$ with high accuracy. (see Section 3.3 for details.) Therefore, an approximated transition probability based on the volume of each polygon and the areas of its faces can be assigned over $\{y_i\}$ and leads to an implementable upwind scheme for the Fokker-Planck equation (1.5). (See Section 3 and Theorem 3.14 for consistence and convergence analysis for this upwind scheme.) We also provide an unconditionally stable explicit time discretization for the upwind scheme based on the detailed balance property of the Markov process in Section 3.5. This explicit scheme is very efficient and enjoys mass conservation law and the exponential convergence to the equilibrium. At last, to show the accuracy of the upwind scheme, we simulate challenging numerical examples including dataset on dumbbell, torus and sphere in Section 3.6.

The approximated transition probability between the nearest neighbor points for the Markov process on $\{y_i\} \subset N$ shows the manifold structure and enables us to efficiently do computations such as cluster, coarse-graining and minimal energy path on manifold. Notice this transition probability between the nearest neighbor points not only incorporates the manifold information but also gives an adapted graph network on the manifold. Based on the above observation, we conduct some examples in Section 5 including optimal coarse-graining partition (cluster) in Section 5.1 and finding the minimal energy path in Section 5.2. The later one is a fundamental question in finding the activation energy in chemical reaction or protein folding.

In practice, the equilibrium potential $U_N(y)$ not only depends on the configuration of $N$, but may also depend on some other parameters of the physical system, for instance, the temperature, the time window or the parameters indicating electronic, magnetic, gravitational, drug or catalysis effect. Consequently, we generalize the equilibrium potential to a function $U_N(y; \theta)$, where $\theta$ is a $q$-dimensional independent parameter. In other words, $U_N(y; \theta)$ can be considered as a function on
the product manifold $\mathcal{N} \times \mathbb{R}^q$. Although we are able to obtain the reaction coordinates through the Diffusion map, it is very expensive to collect the information of the equilibrium potential directly for all parameters $\theta$ over all points $\{y_i\}$. Hence, to achieve goal (II), we apply to a Gaussian Process regression based on the intrinsic geometry of $\mathcal{N} \times \mathbb{R}^q$. Specifically, suppose we can observe the potential $U_N(y, \theta)$ with noise under some labeled parameters $\theta^{(1)} = \{\theta_i\}_{i=1}^r \subset \mathbb{R}^q$ over a small amount of labeled points $\{y_1, y_2, \cdots, y_m\}$ on the manifold $\mathcal{N}$, i.e. we observe $r$ vectors

\begin{equation}
\tilde{z}_j(i) = U_N(y_i, \theta_j) + \tilde{\sigma}_j(i),
\end{equation}

where $\tilde{z}_j = (z_j(1), \cdots, z_j(m)), \tilde{\sigma}_j = (\sigma_j(1), \cdots, \sigma_j(m)) \in \mathbb{R}^m$ and $1 \leq j \leq r$. Here, $\tilde{\sigma}_j$ is a Gaussian noise vector and we assume that $\sigma_j(i) \in \mathcal{N}(0, \sigma_{\text{noise}}^2)$ for all $i, j$. Then we want to reproduce the potential $U_N(y, \theta)$ with unlabeled parameters $\theta^{(2)} = \{\theta_i\}_{i=r+1}^{r+s} \subset \mathbb{R}^q$ over a large number of unlabeled points $\{y_{m+1}, y_{m+2}, \cdots, y_{m+n}\}$ on $\mathcal{N}$. In Section 4, we choose a Gaussian Process prior for the unknown function as $U_N \sim \text{GP}(0, C)$, where $C$ is the covariance function which is a function on $\mathcal{N} \times \mathbb{R}^q \times \mathcal{N} \times \mathbb{R}^q$. A natural and canonical choice for the covariance function is the heat kernel of the manifold $\mathcal{N} \times \mathbb{R}^q$. Unfortunately, without knowing the metric of $\mathcal{N}$, the heat kernel on $\mathcal{N} \times \mathbb{R}^q$ is intractable to calculate. Hence, we apply the Diffusion based Gaussian Process (DBGP) proposed in [14] to recover the heat kernel on $\mathcal{N}$ and on $\mathcal{N} \times \mathbb{R}^q$. The simulation on dumbbell is provided in Section 4.1. It is worth to mention that this method can be generalized to the case when the parameter $\theta$ is an independent variable from a submanifold $\mathcal{N}' \subset \mathbb{R}^q$, i.e. $U_N(y, \theta)$ is a function on $\mathcal{N} \times \mathcal{N}'$ with the product metric.

Due to the curse of high dimensionality, we collect the partial information about the equilibrium potential on the submanifold $\mathcal{N}$ in a low dimensional space and reproduce it on the whole manifold in Section 4. Then, we can process the upwind scheme in Section 3 to find the trajectory $\rho_N^t$ based on the reproduced equilibrium potential. To achieve goal (IV), we can pull back the density function $\rho_N^t$ to a density function $\rho_M^t$ on $\mathcal{M}$ through an estimation of the Jacobian of the reaction coordinates $\Phi$. Detailed pullback algorithm is developed based on the results in [32]. We provide and justify this algorithm in Section 6.

The remaining part of the paper will be organized as follows. In Section 2, we use diffusion map to achieve goal (I). In Section 3, we focus on goal (III) and propose the solver for Fokker-Planck on manifold with several simulations. In Section 4, we deal with goal (IV) to generate new potential with hidden parameters. Then goal (V) is discussed in Section 5. Finally, we recover the dynamics in the original high dimensional space in Section 6. All the commonly used notations are listed in Table 1 for the sake of clarity.

2. Review of nonlinear dimension reduction and diffusion map

In this section, we focus on the goal (I), i.e. learn the reaction coordinate $y$ for the $d$-dimensional manifold $\mathcal{N} \subset \mathbb{R}^d$ to extract the conformational changes with slow time scale from other fast time scale vibrations. We first introduce the basic idea about the nonlinear dimension reduction under the following assumption.
Table 1. Commonly used notations in this paper.

| Symbols | Meaning |
|---------|---------|
| \(\mathbb{R}^p, \mathbb{R}^\ell\) | High (low) dimensional ambient spaces |
| \(d\) | Dimension of the Riemannian manifolds |
| \(\mathcal{M}, \mathcal{N}\) | \(d\)-dimensional smooth closed Riemannian submanifolds of the Euclidean spaces |
| \(x, y\) | Points on \(\mathcal{M}\) and \(\mathcal{N}\) respectively |
| \(dV_\mathcal{M}, dV_\mathcal{N}\) | Volume forms on \(\mathcal{M}\) and \(\mathcal{N}\) respectively |
| \(\Delta\) | Laplace Beltrami operator of a manifold |
| \(\lambda_i, \psi_i\) | The eigenvalues and the corresponding orthonormal (in \(L^2\)) eigenfunctions of \(\Delta\) |
| \(\Phi\) | Reaction coordinates (Smooth embedding of a manifold) |
| \(X, Y\) | Random variables with the range \(\mathcal{M}\) and \(\mathcal{N}\) respectively |
| \(\rho, \rho_\mathcal{M}, \rho_\mathcal{N}\) | Probability density functions on \(\mathcal{M}\) |
| \(\rho_\mathcal{N}, \rho_\mathcal{N}^t\) | Probability density functions on \(\mathcal{N}\) |
| \(n \in \mathbb{N}\) | Number of data points sampled from \(\mathcal{M}\) based on \(\rho\) |
| \(\{x_1, \cdots, x_n\}\) | Data points sampled from \(\mathcal{M}\) based on \(\rho\) |
| \(\epsilon\) | The bandwidth in the Diffusion map |
| \(K_\epsilon\) | Kernel in the Diffusion map |
| \(W_{\epsilon, \alpha}\) | Affinity matrix in Diffusion map with \(\alpha\) normalization |
| \(L_{\epsilon, \alpha}\) | Diffusion map matrix |
| \(\lambda_{i,n,\epsilon}, v_{i,n,\epsilon}\) | The eigenvalues and the corresponding orthonormal eigenvectors in \(l^2\) of \(L_{\epsilon, \alpha} - I\) |
| \(K, \mathbf{t}\) | Parameters in the DBGP algorithm |
| \(C_i\) | the Voronoi cell around the point \(y_i\) on the manifold \(\mathcal{N}\) |
| \(\Gamma_{ij}\) | the Voronoi face between \(y_i\) and \(y_j\) on the manifold \(\mathcal{N}\) |
| \(r\) | bandwidth in the Voronoi cell approximation algorithm |
| \(s\) | threshold in the Voronoi cell approximation algorithm |
| \(\iota_k\) | the projection map in the Voronoi cell approximation algorithm |
| \(\delta\) | the bandwidth in the pullback algorithm |

**Assumption 2.1.** Let \(\mathcal{M}\) be a \(d\) dimensional smooth closed Riemannian submanifold of \(\mathbb{R}^p\). Suppose that \(\rho\) is a smooth probability density function on the manifold \(\mathcal{M}\). We assume that \(\rho\) is bounded from below and from above, i.e. \(\rho_m \leq \rho \leq \rho_M\). Let \(\{x_1, \cdots, x_n\} \subset \mathcal{M}^{i.i.d.}\). Let \(\rho\).

Nonlinear dimension reduction algorithms construct maps which map \(\{x_1, \cdots, x_n\}\) to some low dimensional space \(\mathbb{R}^\ell\) while preserve the topological or geometric structure of the underlying manifold. There are a lot of well known dimension reduction algorithms, for instance, ISOMAP \[41\], eigenmap \[5\], locally linear embedding (LLE) \[37\] and its variations like Hessian LLE \[13\], vector diffusion map \[39, 40\]. In this work, we focus on the algorithm Diffusion map which is introduced by Coifman and Lafon \[10\]. The algorithm of the Diffusion map can be described in the following steps:

(i) For \(x, x' \in \mathcal{M}\), we define \(K_\epsilon(x, x') = \exp(-\frac{\|x - x'\|_{\mathbb{R}^p}^2}{4\epsilon^2})\), where \(\epsilon > 0\) is the bandwidth.
(ii) Define
\[
q_\epsilon(x) := \sum_{i=1}^{n} K_\epsilon(x, x_i).
\]
We define the affinity matrix which is a \(n \times n\) matrix \(W_{\epsilon,\alpha}\):
\[
W_{\epsilon,\alpha,ij} := \frac{K_\epsilon(x_i, x_j)}{q_\alpha^\epsilon(x_i) q_\alpha^\epsilon(x_j)}.
\]
This step is called the \(\alpha\)-normalization.

(iii) Define a \(n \times n\) diagonal matrix \(D\) with diagonal entry
\[
D_{\epsilon,\alpha,ii} = \sum_{j=1}^{n} W_{\epsilon,\alpha,ij}.
\]
Let
\[
L_{\epsilon,\alpha} = D_{\epsilon,\alpha}^{-1} W_{\epsilon,\alpha}.
\]

(iv) To reduce the dimension of the dataset \(\{x_1, \ldots, x_n\}\). We choose \(\alpha = 1\). Denote
\[
\lambda_{0,n,\epsilon} \leq \cdots \leq \lambda_{n-1,n,\epsilon}
\]
to be the eigenvalues of \(L_{\epsilon,1}^{-1/2}\). We find the first \(\ell\) corresponding eigenvectors of \(L_{\epsilon,1}^{-1/2}\), namely, \(\{v_{j,n,\epsilon}\}_{j=1}^{\ell}\). Then the map
\[
x_i \rightarrow (v_{1,n,\epsilon}(i), \ldots, v_{\ell,n,\epsilon}(i))
\]
reduces the dimension of the dataset into the Euclidean space \(\mathbb{R}^\ell\).

Remark 2.2. Note that the matrix \(L_{\epsilon,1}\) in (2.4) may not be symmetric in general. Therefore, in the implementation, we use the matrix \(\tilde{L}_{\epsilon,1} = D_{\epsilon,1}^{-1/2} W_{\epsilon,1} D_{\epsilon,1}^{-1/2}\). \(\tilde{L}_{\epsilon,1}\) is similar to \(L_{\epsilon,1}\) and is symmetric. Therefore, they share the same eigenvalues and the corresponding eigenvectors differ by \(D_{\epsilon,1}^{-1/2}\).

The Diffusion map can be understood from the spectral geometry point of view. Let \(M\) be a \(d\) dimensional smooth closed Riemannian manifold. Let \(\Delta\) be the Laplace-Beltrami operator of \(M\). Let \(\{\lambda_i\}_{i=0}^{\infty}\) be the eigenvalues of \(-\Delta\), and
\[
\Delta \psi_i = -\lambda_i \psi_i,
\]
where \(\psi_i\) is the corresponding eigenfunction normalized in \(L^2(M)\). We have \(0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \cdots\).

Based on the work of [6], [25], [3] and [35], we know that there is a \(\ell\) such that for \(x \in M\), the map
\[
\Psi(x) = (\psi_1(x), \ldots, \psi_\ell(x)),
\]
is a smooth embedding of \(M\) into \(\mathbb{R}^\ell\). The readers may refer to Appendix A for more detailed discussion about the map \(\Psi\).

Remark 2.3. Note that \(\lambda_0 = 0\) and \(\psi_0 = \frac{1}{\sqrt{\text{Vol}(M)}}\) which is a constant. Hence, we never use \(\psi_0\) in constructing an embedding.
Next, we review some theoretical results in justification of the diffusion map as a nonlinear dimension reduction algorithm. Those results relate the Diffusion map to the Laplace Beltrami operator and the eigenfunctions of it. First, it is proved in [10] and [39] that the matrix \( \frac{I - L_{2\epsilon}}{\epsilon^2} \) converges pointwisely to the Laplace Beltrami operator of the manifold in the following sense.

**Theorem 2.4. (Coifman-Lafon, [10], Singer-Wu, [39])** Suppose \( \alpha = 1 \). Under Assumption 2.1, for \( f \in C^3(\mathcal{M}) \), if \( \frac{\log n}{\sqrt{n\epsilon^2 + 2}} \to 0 \) and \( \epsilon \to 0 \) as \( n \to \infty \), then with probability greater than \( 1 - \frac{1}{n^2} \), for all \( i = 1, \ldots, n \), we have

\[
\frac{\sum_{j=1}^{n} L_{\epsilon,1}(i,j) f(x_j) - f(x_i)}{\epsilon^2} = \Delta f(x_i) + O\left(\frac{\sqrt{\log n}}{\sqrt{n\epsilon^2 + 2}}\right).
\]

The \( \alpha = 1 \) normalization in the Diffusion Map comes from the idea of the density estimation. When \( \alpha \) is chosen to be 1, the impact of the non uniform density \( \rho \) is removed. Hence, the Laplace-Beltrami operator in the previous theorem is not contaminated by the probability density function \( \rho \).

A stronger version of the convergence theorem in [10] shows the spectral convergence of the diffusion map in \( L^2 \) sense. At last, in [14], it shows the \( L^\infty \) spectral convergence result based on the following definition.

**Definition 2.5. Under Assumption 2.1**, suppose \( v_{j,n,\epsilon} \) is an eigenvector of \( \frac{I - L_{2\epsilon}}{\epsilon^2} \) which is normalized in the \( l^2 \) norm. Let \( N_k = |B_\epsilon^3(x_k) \cap \{x_1, \ldots, x_n\}| \), the number of points in the \( \epsilon \) ball in the ambient space. Then, we define the \( l^2 \) norm of \( v_{j,n,\epsilon} \) with respect to the inverse estimated probability density \( 1/\hat{\rho} \) as:

\[
\|v_{j,n,\epsilon}\|_{l^2(1/\hat{\rho})} := \sqrt{\frac{|S^{d-1}|\epsilon^d}{d} \sum_{i=1}^{n} v_{j,n,\epsilon}(k) \cdot N_k}.
\]

Define the renormalization of \( v_{j,n,\epsilon} \) in the \( l^2 \) norm with respect to the inverse estimated probability density \( 1/\hat{\rho} \) as:

\[
V_{j,n,\epsilon} := \frac{v_{j,n,\epsilon}}{\|v_{j,n,\epsilon}\|_{l^2(1/\hat{\rho})}}.
\]

Intuitively, \( v_{j,n,\epsilon} \) is a discretization of some function on \( \mathcal{M} \) while \( \|v_{j,n,\epsilon}\|_{l^2(1/\hat{\rho})} \) is an approximation of the \( L^2(\mathcal{M}) \) norm of the function. Hence, \( V_{j,n,\epsilon} \) can be regarded as a discretization of some function that is normalized in \( L^2(\mathcal{M}) \). On the other hand, the vector \( \tilde{v}_j = (\psi_j(x_1), \ldots, \psi_j(x_n))^\top \) is a discretization of \( \psi_j \) which is also normalized in \( L^2(\mathcal{M}) \). Therefore, it is reasonable to compare \( V_{j,n,\epsilon} \) and \( \tilde{v}_j \) rather than \( v_{j,n,\epsilon} \) and \( \psi_j \). In the following theorem, it shows that, on a closed manifold \( \mathcal{M} \), if we fix \( K \) and we choose the bandwidth \( \epsilon \) based on the number of data points \( n \), then for \( i < K \), with high probability, \( \lambda_{j,n,\epsilon} \) is an approximation of the \( j \)-th eigenvalue \( \lambda_j \) of \( -\Delta \) and \( V_{j,n,\epsilon} \) is an approximation of \( \tilde{v}_j \).

**Theorem 2.6. (Dunson-Wu-Wu, [14])** Under Assumption 2.1, suppose all the eigenvalues of \( \Delta \) are simple. Let \( (\lambda_j, \psi_j) \) be the \( j \)-th eigenpair of \( -\Delta \) with \( \psi_j \) normalized in \( L^2(\mathcal{M}) \). Let \( L_{\epsilon,1} \) be the matrix in (2.4). Let \( (\lambda_{j,n,\epsilon}, V_{j,n,\epsilon}) \) be the \( j \)-th eigenpair of \( \frac{I - L_{2\epsilon}}{\epsilon^2} \) with \( V_{j,n,\epsilon} \) normalized as in Definition 2.5. Let \( \Gamma_K := \min_{1 \leq j \leq K} \text{dist}(\lambda_j, \sigma(-\Delta) \setminus \{\lambda_j\}) \). For a positive integer \( K \), if we
choose $\epsilon = c(n) = n^{-\frac{1}{4d+15}}$ and $\epsilon \leq K_1 \min \left( \frac{\min(F_{K,1})}{K_2 + \lambda_{K}^{d+2+5}}, \frac{1}{(2+\lambda_{K}^{d+1})^2} \right)$, then there is a sequence $a_n \in \{1, -1\}$ such that for all $i < K$ with probability greater than $1 - n^{-2}$,

\begin{equation}
|\lambda_{j,n,\epsilon} - \lambda_j| \leq K_3 \epsilon^{3/2},
\end{equation}

\begin{equation}
\max_{1 \leq i \leq n} |a_n v_{j,n,\epsilon}(i) - \psi_j(x_i)| \leq K_4 \epsilon^{1/2},
\end{equation}

where $K_1$ and $K_2 > 1$ are constants depending on the lower bound of the p.d.f. $\rho_m$, the $C^2$ norm of p.d.f. $\rho$, the volume, the injectivity radius and the curvature of the manifold and $K_3$ and $K_4$ depend on the curvature of $\mathcal{M}$, $\rho_m$ and the $C^1$ norm of $\rho$.

**Remark 2.7.** The above theorem assumes that the eigenvalues of $\Delta$ are simple for notation simplicity. In the case when the eigenvalues are not simple, the same theorem still works by introducing the eigenprojection.

Based on Definition 2.5 and Theorem 2.6, we provide the following definition of the reaction coordinates which we use in this work.

**Definition 2.8 (Reaction coordinates).** Let $(\lambda_i, \psi_i)$ be the $i$-th eigenpair of the Laplace Beltrami operator on $\mathcal{M}$, $-\Delta$, with $\psi_j$ normalized in $L^2(\mathcal{M})$. Suppose for $x \in \mathcal{M}$

\begin{equation}
\Psi(x) = (\psi_1(x), \cdots, \psi_{\ell}(x)),
\end{equation}

is a smooth embedding of $\mathcal{M}$ into $\mathbb{R}^\ell$. Let $A$ be a $\ell \times \ell$ diagonal matrix such that $A_{ii} = \|v_{j,n,\epsilon}\|_{L^2(\mathcal{M})}$ as defined in Definition 2.5. Under Assumption 2.1, we define

\begin{equation}
y_i = \Phi(x_i) := A \circ \Psi(x_i) \in \mathbb{R}^\ell,
\end{equation}

to be the reaction coordinates of $x_i$.

Note that $A \circ \Psi$ is also a smooth embedding of $\mathcal{M}$ into $\mathbb{R}^\ell$. Hence, by Theorem 2.6, we have a justification of the diffusion map. Let $\{v_{1,n,\epsilon}, \cdots, v_{\ell,n,\epsilon}\}$ be the first $\ell$ eigenvectors of $\frac{L_{\mathcal{M}}}{\epsilon^2}$ in Step (iv) of the algorithm of the Diffusion map. Then, the Diffusion map

\begin{equation}
x_i \rightarrow (v_{1,n,\epsilon}(i), \cdots, v_{\ell,n,\epsilon}(i)),
\end{equation}

is an approximation of $y_i = \Phi(x_i) := A \circ \Psi(x_i)$ over the data points $\{x_1, \cdots, x_n\}$.

3. **Solution to the Fokker-Planck equation on $\mathcal{N}$**

Suppose $\mathcal{N}$ is a $d$ dimensional smooth closed Riemannian submanifold of $\mathbb{R}^\ell$ with the coordinates $\mathbf{y}$ obtained in Section 2. In this section, we will focus on the goal (III), i.e. given an equilibrium potential for the new physical system as $W(\mathbf{y}) := U_N'(\mathbf{y})$, we want to solve the Fokker-Planck equation on $\mathcal{N}$ which drives any initial data $\rho_0$ to the equilibrium density on $\mathcal{N}$, $\rho_N^N(\mathbf{y}) \propto e^{-W(\mathbf{y})}$ (after taking $kT = 1$). To study the trajectory of $\rho_t$ driving any initial data $\rho_0$ to the equilibrium density $\rho_N^N(\mathbf{y}) \propto e^{-W(\mathbf{y})}$ (for instance, how the old physical system change to the new physical system), it is sufficient to solve the following Fokker-Planck equation on manifold $\mathcal{N}$

\begin{equation}
\partial_t \rho_t^N = \text{div}_N(\nabla_N \rho_t^N + \rho_t^N \nabla_N W).
\end{equation}
For notation simplicity, in the remaining of this section, we will denote the equilibrium density for the Fokker-Planck equation (1.5) as \( \pi := \rho_\infty^N \).

As mentioned before, since we do not have exact information of \( N \), the only implementable method is to use the data \( \{ y_i \} \subset N \) sampled from invariant density \( \pi \) on \( N \) to construct directly a good discrete approximation to the Fokker-Planck equation (3.1). Assume \( \pi \) has upper and lower bounds, thus it is proved the sampled data is well-distributed on \( N \). In Section 3.1, we will construct a Voronoi tessellation for \( N \) from \( \{ y_i \} \subset N \) and then assign the transition probability for an approximated Markov process on \( \{ y_i \} \) between the nearest neighbor points. This transition probability with detailed balance property also gives an theoretic upwind scheme for Fokker-Planck equation (3.1). We give the stability and convergence analysis for this scheme in Section 3.3. However, without the exact metric on \( N \), to propose an implementable scheme, the Voronoi tessellation needs to be further approximated. In Section 3.4, thanks to the metric on \( N \) induced by low dimensional Euclidean distance, the volumes of each Voronoi cell and the areas of the each Voronoi face can be further approximated by polygons in its tangent plane in \( \mathbb{R}^\ell \) with high order accuracy. Therefore the new transition probability based on polygons can be assigned and leads to an implementable upwind scheme for Fokker-Planck equation (3.1); see Theorem 3.14. In Section 3.5, we provide an unconditionally stable explicit time descretization for the upwind scheme based on the detailed balance property of the Markov process, which satisfies mass conservation law and exponentially converges to the equilibrium. Challenging numerical simulations including dataset on torus, dumbbell and sphere are conducted in Section 3.6.

3.1. Construction of Voronoi tessellation and the upwind scheme on manifold \( N \). In this section, we construct an upwind scheme based on the Voronoi tessellation for manifold \( N \). We will see the advantage is that the Voronoi tessellation automatically gives a positive-preserving upwind scheme for the Markov process with detailed balance; see Lemma 3.4.

Suppose \( ( N, d_N ) \) is a \( d \) dimensional smooth closed submanifold of \( \mathbb{R}^\ell \) and \( d_N \) is induced by the Euclidean metric in \( \mathbb{R}^\ell \). \( Q := \{ y_i \}_{i=1}^n \) are sampled from the equilibrium density \( \pi = \rho_\infty^N \). Define the Voronoi cell as

\[
C_i := \{ y \in N; d_N(y, y_i) \leq d_N(y, y_j) \text{ for all } y_j \in Q \},
\]

with the volume \( |C_i| = \mathcal{H}^d(C_i) \). Then \( N = \bigcup_{i=1}^n C_i \) is a Voronoi tessellation of manifold \( N \). Denote the Voronoi face for cell \( C_i \) as

\[
\Gamma_{ij} := C_i \cap C_j, \text{ and its area as } |\Gamma_{ij}| = \mathcal{H}^{d-1}(\Gamma_{ij})
\]

for any \( j = 1, \cdots, n \). If \( \Gamma_{ij} = \emptyset \) or \( i \neq j \) then we set \( |\Gamma_{ij}| = 0 \). We define the bisector between \( y_i \) and \( y_j \) to be the set

\[
G_{ij} := \{ y \in N; d_N(y, y_i) = d_N(y, y_j) \}.
\]

Obviously, we have \( \Gamma_{ij} \subset G_{ij} \).

Define the associated adjacent sample points as

\[
VF(i) := \{ j; \Gamma_{ij} \neq \emptyset \}.
\]

We have the following basic facts about the Voronoi cells and bisectors on a manifold.
Assumption 3.3. $M$ is a well defined unit outward normal vector field on a dimensional submanifold. Moreover, the minimizing geodesic between $y$ and $y_j$ on the bisector between $y_i$ and $y_j$. Let $y^*$ be the midpoint of the minimizing geodesic between $y_i$ and $y_j$. Then there is a $d$ dimensional geodesic ball $B$ around $y^*$ such that $M_{ij} = B \cap G_{ij}$ is a $d-1$ dimensional submanifold. Moreover, the minimizing geodesic between $y_i$ and $y_j$ is perpendicular to $M_{ij}$ at $y^*$.

Based on the above proposition, we make the following assumption about each Voronoi face $\Gamma_{ij}$.

Assumption 3.3. We assume that each Voronoi face $\Gamma_{ij} \subset M_{ij}$ for the submanifold $M_{ij}$ and there is a well defined unit outward normal vector field $n$ on each $\partial C_i$.

Above assumption enables us to apply the divergence Theorem on each Voronoi cell. Let us first clarify the density on each cell. Let $\chi_{C_i}$ be the characteristic function such that $\chi_{C_i} = 1$ for $y \in C_i$ and 0 otherwise. For $i = 1, \cdots, n$,

$$\rho_{\text{approx}}(y) = \sum_{i=1}^{n} \pi_i \chi_{C_i}(y)$$

is the piecewise constant probability distribution on $\mathcal{N}$ provided $\sum_{i=1}^{n} \pi_i |C_i| = 1$ and $\pi_i \geq 0$. Let $\pi_i$ be the approximated equilibrium density at $y_i$ satisfying $\sum_{i=1}^{n} \pi_i |C_i| = 1$. If $\rho_{\text{approx}}(y) = \sum_i \pi_i \chi_{C_i}(y)$ is an approximation of density $\rho_{\mathcal{N}}(y)$, then $\rho_i$ is an approximation of the density $\rho_{\mathcal{N}}(y_i)$.

Recall the Fokker-Planck on $\mathcal{N}$ (3.1). We first recast (3.1) in the relative entropy form

$$\partial_t \rho_i^\mathcal{N} = \text{div}_{\mathcal{N}} \left( \pi \nabla_{\mathcal{N}} \left( \frac{\rho_i^\mathcal{N}}{\pi} \right) \right).$$

Then using the finite volume method and the divergence theorem on manifold, we have

$$\frac{d}{dt} \rho_i |C_i| = \frac{d}{dt} \int_{C_i} \rho_{\text{approx}} H^d(C_i) = \sum_{j \in V_F(i)} \int_{\Gamma_{ij}} \pi n \cdot \nabla_{\mathcal{N}} \left( \frac{\rho_{\text{approx}}}{\pi} \right) \mathcal{H}^{d-1}(\Gamma_{ij}).$$
where \( \mathbf{n} \) is the unit outward normal vector field on \( \partial C_i \). Based on this, we introduce the following upwind scheme. For \( i = 1, \ldots, n \),

\[
\frac{d}{dt} \rho_i |C_i| = \frac{1}{2} \sum_{j \in VF(i)} \frac{\pi_i + \pi_j}{|y_i - y_j|} |\Gamma_{ij}| \left( \frac{\rho_j}{\pi_j} - \frac{\rho_i}{\pi_i} \right).
\]

We will first formulate upwind scheme (3.8) as the forward equation for a Markov process with basic properties such as ergodicity in Section 3.2. We then show the truncation error analysis and stability analysis and thus convergence of the scheme (3.8) later in Section 3.3.

### 3.2. Associated Markov process, detailed balance and ergodicity

We will first formulate upwind scheme (3.8) as the forward equation for a Markov process and then in Proposition 3.5, we study the generator of the Markov process, which leads to ergodicity of \( \rho_i \).

**Lemma 3.4.** The upwind scheme (3.8) is the forward equation for a Markov Process with transition probability \( P_{ij} \) (from state \( j \) to \( i \)) and jump rate \( \lambda_j \)

\[
\frac{d}{dt} \rho_i |C_i| = \sum_{j \in VF(i)} \lambda_j P_{ij} \rho_j |C_j| - \lambda_i \rho_i |C_i|,
\]

where

\[
\lambda_i := \frac{1}{2 |C_i| \pi_i} \sum_{j \in VF(i)} \frac{\pi_i + \pi_j}{|y_i - y_j|} |\Gamma_{ij}|, \quad i = 1, 2, \ldots, n;
\]

\[
P_{ij} := \frac{\lambda_j}{\lambda_i} \frac{\pi_i + \pi_j}{|\Gamma_{ij}|} |\Gamma_{ij}| |y_i - y_j|, \quad j \in VF(i); \quad P_{ij} = 0, \quad j \notin VF(i).
\]

(i) It satisfies \( \sum_i P_{ij} = 1 \) and the detailed balance property

\[
P_{ij} \lambda_j \pi_j |C_j| = P_{ji} \lambda_i \pi_i |C_i|.
\]

(ii) With \( \{w_i\}_{i=1}^n := \{\rho_i |C_i|\}_{i=1}^n \), we recast the forward equation (3.9) as

\[
\frac{d}{dt} w = A^* w,
\]

where

\[
A^* = (a^*_{ij})_{n \times n}, \quad a^*_{ij} := \begin{cases} -\lambda_i, & j = i; \\ \lambda_j P_{ij}, & j \neq i. \end{cases}
\]

(iii) \( \sum_{i=1}^n a^*_{ij} = 0 \), which gives the conservation law for \( w_i = \rho_i |C_i| \)

\[
\frac{d}{dt} \sum_{i=1}^n w_i = \sum_{i=1}^n \sum_{j=1}^n a^*_{ij} w_j = 0;
\]

(iv) We have the dissipation relation

\[
\frac{d}{dt} \sum_i \frac{\rho_i^2}{\pi_i} |C_i| = - \sum_{i,j} \frac{\pi_i + \pi_j}{2} |\Gamma_{ij}| \left( \frac{\rho_j}{\pi_j} - \frac{\rho_i}{\pi_i} \right)^2.
\]
Proof. First, one can rewrite (3.8) as (3.9) with \( \lambda_i = \frac{1}{\pi_i C_i |\pi_i|} \sum_{j \in V_F(i)} \pi_{i,j} \frac{\pi_{i,j}}{|y_i - y_j|} |\Gamma_{ij}| \) and \( P_{ij} \lambda_j = \frac{\pi_i + \pi_j}{2|y_i - y_j|} |\Gamma_{ij}| \). Then since \( \frac{\pi_i + \pi_j}{|y_i - y_j|} |\Gamma_{ij}| \) is symmetric, we have

\[
\frac{d}{dt} \left( \sum_{i=1}^n |C_i| \rho_i \right) = \sum_{i,j} \frac{1}{2} \frac{\pi_i + \pi_j}{|y_i - y_j|} |\Gamma_{ij}| \left( \frac{\rho_j}{\pi_j} - \frac{\rho_i}{\pi_i} \right) = 0. \tag{3.16}
\]

Second we can check

\[
\sum_i P_{ij} = \sum_{i \in V_F(j)} P_{ij} = \frac{1}{\lambda_j} \sum_{i \in V_F(j)} \frac{\pi_i + \pi_j}{2|y_i - y_j|} |\Gamma_{ij}| = 1. \tag{3.17}
\]

Third the detailed balance property comes from the symmetric property of \( \frac{\pi_i + \pi_j}{|y_i - y_j|} |\Gamma_{ij}| \) and

\[
\lambda_j P_{ij} \pi_j |C_j| = \frac{\pi_i + \pi_j}{2|y_i - y_j|} |\Gamma_{ij}| = \lambda_i P_{ji} \pi_i |C_i|. \tag{3.18}
\]

Next, the conservation law follows directly from \( \sum_{i=1}^n a_{ij}^* = 0 \).

Finally, multiplying

\[
\frac{d}{dt} \rho_i |C_i| = \sum_{j \in V_F(i)} \lambda_j P_{ji} \pi_i |C_i| \frac{\rho_j}{\pi_j} - \lambda_i \pi_i |C_i| \frac{\rho_i}{\pi_i} = \sum_{j \in V_F(i)} \lambda_i \pi_i |C_i| P_{ji} \left( \frac{\rho_j}{\pi_j} - \frac{\rho_i}{\pi_i} \right) = \frac{d}{dt} \sum_{i} \rho_i^2 \pi_i \tag{3.19}
\]

by \( \frac{\rho_j}{\pi_j} \) and detailed balance property (5.14) show that

\[
\frac{d}{dt} \sum_i \rho_i^2 \pi_i = \sum_{i,j} \lambda_i \pi_i |C_i| P_{ji} \left( \frac{\rho_j}{\pi_j} - \frac{\rho_i}{\pi_i} \right)^2 = \sum_{i,j} \frac{\pi_i + \pi_j}{2|y_i - y_j|} |\Gamma_{ij}| \left( \frac{\rho_j}{\pi_j} - \frac{\rho_i}{\pi_i} \right)^2. \tag{3.20}
\]

This gives the dissipation relation (3.15).

\[\square\]

Proposition 3.5. Let \( A := (a_{ij})_{n \times n} \) be the transport matrix of \( A^* \) in (3.13) with \( a_{ij}^* = a_{ji} \). Then \( \{u_i\}_{i=1}^n := \{ \rho_i \pi_i \}_{i=1}^n \) is the solution to the backward equation

\[
\frac{d}{dt} u = Au. \tag{3.21}
\]

Moreover, 0 is the simple, principle eigenvalue of \( A \) with the ground state \( \{1,1, \cdots, 1\} \). We thus have the exponential decay of \( \rho_i \) with respect to time \( t \),

\[
\max_i \left| \frac{\rho_i(t) - \pi_i}{\pi_i} \right| \leq ce^{-\alpha t}, \tag{3.22}
\]

where \( \alpha > 0 \) is the absolute value of the spectral gap of \( A \).
Proof. With the detailed balance property (3.11), we recast the forward equation (3.9) as
\[
\frac{d}{dt} \rho_i | C_i | = \sum_{j \in V_F(i)} \lambda_j P_{ij} \pi_j | C_j | \frac{\rho_j}{\pi_j} - \lambda_i \pi_i | C_i | \frac{\rho_i}{\pi_i},
\]
which gives
\[
\frac{d}{dt} \rho_i \pi_i = \lambda_i \left( \sum_{j \in V_F(i)} P_{ji} \frac{\rho_j}{\pi_j} - \frac{\rho_i}{\pi_i} \right).
\]

Next, we show \( \{u_i\}_{i=1}^n := \{ \frac{\rho_i}{\pi_i} \}_{i=1}^n \) is the solution to this backward equation. Recast (3.24) as
\[
\frac{d}{dt} u_i = \sum_{j=1}^n a_{ij} u_j.
\]
Here \( A = \{a_{ij}\} \) is the generator of the backward equation
\[
\frac{d}{dt} u = Au.
\]
One can check \( \sum_{j=1}^n a_{ij} = 0, a_{ij} \geq 0 \) for \( j \neq i \) and \( a_{ii} < 0 \).

Moreover, due to the detailed balance property (3.11), we know \( A \) is self-adjoint in the weighted \( l^2 \)-space
\[
\langle u, Av \rangle_{\pi | C} := \sum_{i,j} u_i a_{ij} v_j \pi_i | C_i | = \sum_{i,j} a_{ji} u_i v_j \pi_j | C_j | =: \langle Au, v \rangle_{\pi | C}.
\]
Thus we know the eigenvalues of \( A \) are real. For the matrix \( \mu_1 I + A \) with \( \mu_1 > \max_i \lambda_i \), we know each element is non negative and \( \sum_j (\mu_1 + a_{ij}) = \mu_1 > 0 \). Since Voronoi tessellation \( \mathcal{N} = \bigcup_{i=1}^n C_i \), we know if \( \mathcal{N} \) is strongly connect then \( \mu_1 I + A \) is irreducible. So by the Perron-Frobenius theorem for \( \mu_1 I + A \), we know the Perron-Frobenius eigenvalue (i.e. the principal eigenvalue) of \( \mu_1 I + A \) is \( \mu_1 \) and \( \mu_1 > 0 \) is simple eigenvalue with the ground state \( u^* := (1, 1, \cdots, 1) \) and other eigenvalues \( \mu_i \) satisfy \( |\mu_i| < \mu_1 \). Therefore, we have the exponential decay of \( u \) to its ground state \( u^* \)
\[
\|u - u^*\| \leq c e^{(\mu_2 - \mu_1) t}.
\]
Here \( \mu_1 - |\mu_2| > 0 \) is the spectral gap of \( \mu_1 I + A \), which is also the spectral gap of \( A \).

See [27] for the ergodicity of upwind schemes in unbounded space.

3.3. Truncation error estimate, stability and convergence of the upwind scheme (3.8).

In this section we prove the stability of (3.8) in Lemma 3.6. Then we obtain the convergence of the solution to upwind scheme (3.8) to the solution of Fokker-Planck (3.1) in Theorem 3.7.

First, we have the following stability property, which corresponds to the Markov chain version of the Crandall-Tartar lemma for monotone schemes. This lemma is also known as the total variation diminishing for two density solutions.
Lemma 3.6. Any two solutions \( \rho_i \) and \( \tilde{\rho}_i \) to upwind scheme \((3.9)\) have the following stability properties

\[
\frac{d}{dt} \sum_{i=1}^{n} |\rho_i - \tilde{\rho}_i| |C_i| \leq 0; \tag{3.29}
\]

\[
\frac{d}{dt} \sum_{i=1}^{n} |\tilde{\rho}_i| |C_i| \leq 0. \tag{3.30}
\]

Proof. First assume \( \rho_i \) and \( \tilde{\rho}_i \) are two solutions to upwind scheme \((3.9)\). We have

\[
\frac{d}{dt} (\rho_i |C_i| - \tilde{\rho}_i |C_i|) = \sum_{j \in V_F(i)} P_{ij} \lambda_j |C_j| (\rho_j - \tilde{\rho}_j) - \lambda_i |C_i| (\rho_i - \tilde{\rho}_i). \tag{3.31}
\]

Multiply \( \text{sgn}(\rho_i - \tilde{\rho}_i) \) to both sides and then take summation with respect to \( i \)

\[
\frac{d}{dt} \sum_{i=1}^{n} |C_i| |\rho_i - \tilde{\rho}_i| \leq \sum_{i,j} P_{ij} \lambda_j |C_j| |\rho_j - \tilde{\rho}_j| - \sum_{i=1}^{n} \lambda_i |C_i| |\rho_i - \tilde{\rho}_i| = 0,
\]

where we used \( \sum_{i \in V_F(j)} P_{ij} = 1 \). Second, take time derivative in \((3.9)\), then we have

\[
\frac{d^2}{dt^2} \rho_i |C_i| = \sum_{j \in V_F(i)} P_{ij} \lambda_j |C_j| \dot{\rho}_j - \lambda_i |C_i| \dot{\rho}_i. \tag{3.33}
\]

Then similarly we can multiply \( \text{sgn}(\dot{\rho}_i) \) to both sides and obtain

\[
\frac{d}{dt} \sum_{i=1}^{n} |C_i| |\dot{\rho}_i| \leq \sum_{i,j} P_{ij} \lambda_j |C_j| |\dot{\rho}_j| - \sum_{i=1}^{n} \lambda_i |C_i| |\dot{\rho}_i| = 0,
\]

where we used \( \sum_{i \in V_F(j)} P_{ij} = 1 \).

We conclude this section by the following convergence theorem in the weighted \( L^2 \) sense.

Theorem 3.7 (Convergence). Suppose \( \rho(y, t), t \in [0, T] \) is a smooth solution to Fokker-Planck equation \((3.1)\) on manifold \( \mathcal{N} \subset \mathbb{R}^t \) with initial density \( \rho^0(y) \). Let \( \mathcal{N} = \bigcup_{i=1}^{n} C_i \) be the Voronoi tessellation of manifold \( \mathcal{N} \) based on \( \{y_i\} \). Let

\[
h = \max \left( \max_{i=1,...,n} (\text{diam}(C_i)), \max_{i=1,...,n} \frac{d\mathcal{N}(y_i, y_j)}{\pi_i} \right) \tag{3.35}
\]

Let \( \rho_i \) be the solution to the upwind scheme \((3.8)\) with initial data \( \rho_i^0 \) and \( \epsilon_i := \rho(y_i) - \rho_i \). Under Assumption 3.3, we have the following error estimate

\[
\max_{t \in [0, T]} \sum_i \epsilon_i(t)^2 \frac{|C_i|}{\pi_i} \leq \left( \sum_i \epsilon_i(0)^2 \frac{|C_i|}{\pi_i} + O(h^2 (nh \max_i |\partial C_i| + 1)) \right) \exp^T,
\]

where the constant in \( O(h^2 (nh \max_i |\partial C_i| + 1)) \) depends on \( \text{Vol}(\mathcal{N}) \), the minimum of \( \pi \), the \( C^1 \) norm of \( \pi \), the \( L^\infty \) norm of \( \partial_i \mathcal{N} \rho \) and the \( C^2 \) norm of \( \frac{\partial^2}{\partial \pi} \).
Proof. Let \( \rho_i^e := \frac{1}{|C_i|} \int_{C_i} \rho \, dy \) be the cell average. Plug the exact solution into the numerical scheme

\[
\partial_t (\rho_i^e |C_i|) = \sum_{j \in V F(i)} \frac{\pi_i + \pi_j}{2|y_i - y_j|} |\Gamma_{ij}| \left( \frac{\rho(y_j)}{\pi_j} - \frac{\rho(y_i)}{\pi_i} \right) + \sum_{j \in V F(i)} \varepsilon_{ij},
\]

(3.37)

\[
\varepsilon_{ij} := \int_{\Gamma_{ij}} \pi n_{ij} \cdot \nabla_N \frac{\rho}{\pi} d\mathcal{H}^{d-1} - \frac{\pi_i + \pi_j}{2|y_i - y_j|} |\Gamma_{ij}| \left( \frac{\rho(y_j)}{\pi_j} - \frac{\rho(y_i)}{\pi_i} \right),
\]

where \( n_{ij} \) is the restriction of the unit outward normal vector field on \( \Gamma_{ij} \). Exchanging \( i, j \) above, we can see that \( \varepsilon_{ij} \) is anti-symmetric.

Subtracting the numerical scheme (3.38) from (3.37), we have

\[
\frac{d}{dt} \sum_i \frac{e_i}{\pi_i} |C_i| = \sum_{j \in V F(i)} \frac{\pi_i + \pi_j}{2|y_i - y_j|} |\Gamma_{ij}| \left( \frac{e_j}{\pi_j} - \frac{e_i}{\pi_i} \right) + \sum_{j \in V F(i)} \varepsilon_{ij} + \partial_t ((\rho(y_i) - \rho_i^e) e_i |C_i|/\pi_i).
\]

Similar to the dissipation relation (3.15), multiplying \( \varepsilon_{ij}/\pi_i \) shows that

\[
\frac{d}{dt} \sum_i \frac{e_i^2}{\pi_i} |C_i| = - \sum_{j \in V F(i)} \frac{\pi_i + \pi_j}{2|y_i - y_j|} |\Gamma_{ij}| \left( \frac{e_j}{\pi_j} - \frac{e_i}{\pi_i} \right)^2 + \sum_{j \in V F(i)} \sum_i 2\varepsilon_{ij} \frac{e_i}{\pi_i} + \sum_i 2\partial_t ((\rho(y_i) - \rho_i^e) e_i |C_i|/\pi_i).
\]

Since \( \varepsilon_{ij} \) is anti-symmetric,

\[
\frac{d}{dt} \sum_i \frac{e_i^2}{\pi_i} |C_i| = - \sum_{j \in V F(i)} \frac{\pi_i + \pi_j}{2|y_i - y_j|} |\Gamma_{ij}| \left( \frac{e_j}{\pi_j} - \frac{e_i}{\pi_i} \right)^2 + \sum_{j \in V F(i)} \sum_i \varepsilon_{ij} \left( \frac{e_i}{\pi_i} - \frac{e_j}{\pi_j} \right) + \sum_i 2\partial_t ((\rho(y_i) - \rho_i^e) e_i |C_i|/\pi_i).
\]

Applying Young's inequality to the last two terms, we have

\[
\sum_i \sum_{j \in V F(i)} \varepsilon_{ij} \left( \frac{e_i}{\pi_i} - \frac{e_j}{\pi_j} \right)
\]

(3.41)

\[
\leq \frac{1}{2} \sum_i \sum_{j \in V F(i)} \frac{\pi_i + \pi_j}{2|y_i - y_j|} |\Gamma_{ij}| \left( \frac{e_j}{\pi_j} - \frac{e_i}{\pi_i} \right)^2 + \frac{1}{2} \sum_i \sum_{j \in V F(i)} \frac{\varepsilon_{ij}^2}{\pi_i + \pi_j} |\Gamma_{ij}|;
\]

(3.42)

\[
\sum_i 2\partial_t ((\rho(y_i) - \rho_i^e) e_i |C_i|/\pi_i) \leq \sum_i \left[ \partial_t ((\rho(y_i) - \rho_i^e) |C_i|/\pi_i) \right]^2 + \sum_i e_i^2 |C_i|/\pi_i.
\]

Thus we have

\[
\frac{d}{dt} \sum_i \frac{e_i^2}{\pi_i} |C_i| \leq - \frac{1}{2} \sum_i \sum_{j \in V F(i)} \frac{\pi_i + \pi_j}{2|y_i - y_j|} |\Gamma_{ij}| \left( \frac{e_j}{\pi_j} - \frac{e_i}{\pi_i} \right)^2 + \frac{1}{2} \sum_i \sum_{j \in V F(i)} \frac{\varepsilon_{ij}^2}{\pi_i + \pi_j} |\Gamma_{ij}| + \sum_i \frac{\varepsilon_{ij}^2 |C_i|}{\pi_i} + \sum_i e_i^2 |C_i|/\pi_i.
\]

(3.43)

Next, we bound the term \( \sum_i \sum_{j \in V F(i)} \frac{\varepsilon_{ij}^2}{\pi_i + \pi_j} |\Gamma_{ij}| \).

Let \( G_{ij} \) be the bisector between \( y_i \) and \( y_j \). Suppose \( y^* \) is the intersection point of the minimizing geodesic from \( y_i \) to \( y_j \) and \( G_{ij} \). We have \( d_N(y^*, y_i) = d_N(y^*, y_j) \). Suppose \( T \) is the unit tangent
vector of the minimizing geodesic at \( y^* \). From the Taylor expansion of \( \frac{\rho}{\pi} \) along the geodesic, we have

\[
\frac{\rho}{\pi}(y_j) - \frac{\rho}{\pi}(y^*) = T \cdot \nabla_N \frac{\rho}{\pi}(y^*) d_N(y^*, y_j) + O(d_N^2(y^*, y_j)),
\]

(3.44)

\[
\frac{\rho}{\pi}(y^*) - \frac{\rho}{\pi}(y_i) = T \cdot \nabla_N \frac{\rho}{\pi}(y^*) d_N(y^*, y_i) + O(d_N^2(y^*, y_i)).
\]

(3.45)

By Assumption 3.3, \( n_{ij} \) can be extended to a unit normal vector field on the \( d - 1 \) dimensional submanifold \( M_{ij} \subseteq G_{ij} \). We also call the extension to be \( n_{ij} \). By Proposition 3.2 \( T = n_{ij}(y^*) \).

Therefore, if we add the above two equations, we have

\[
\frac{\rho}{\pi}(y_j) - \frac{\rho}{\pi}(y_i) = n_{ij} \cdot \nabla_N \frac{\rho}{\pi}(y^*) d_N(y_i, y_j) + O(d_N^2(y_i, y_j)).
\]

(3.46)

Hence,

\[
n_{ij} \cdot \nabla_N \frac{\rho}{\pi}(y^*) = \frac{\rho(y_j) - \rho(y_i)}{d_N(y_i, y_j)} + O(d_N(y_i, y_j)) = \frac{\rho(y_j) - \rho(y_i)}{|y_i - y_j|} + O(d_N(y_i, y_j)),
\]

(3.47)

where we apply Lemma 3.8 in the last step. Similarly,

\[
\pi(y_j) - \pi(y^*) = O(d_N(y^*, y_j)),
\]

(3.48)

\[
\pi(y^*) - \pi(y_i) = O(d_N(y^*, y_i)).
\]

(3.49)

Hence,

\[
\pi(y^*) = \frac{\pi_i + \pi_j}{2} + O(d_N(y_i, y_j)).
\]

(3.50)

Therefore,

\[
\pi(y^*) n_{ij} \cdot \nabla_N \frac{\rho}{\pi}(y^*) = \frac{\pi_i + \pi_j}{2} \frac{\rho(y_j) - \rho(y_i)}{|y_i - y_j|} + O(d_N(y_i, y_j)).
\]

(3.51)

For any \( y \) on \( \Gamma_{ij} \),

\[
\pi(y) n_{ij} \cdot \nabla_N \frac{\rho}{\pi}(y) = \pi(y^*) n \cdot \nabla_N \frac{\rho}{\pi}(y^*) + O(d_N(y, y^*))
\]

\[
= \pi(y^*) n \cdot \nabla_N \frac{\rho}{\pi}(y^*) + O(d_N(y_i, y) + d_N(y_i, y_j))
\]

(3.52)

\[
= \pi(y^*) n \cdot \nabla_N \frac{\rho}{\pi}(y^*) + O(\text{diam}(C_i) + d_N(y_i, y_j)),
\]

(3.53)

where \( \text{diam}(C_i) \) is the diameter of \( C_i \) measured with respect to the distance in \( N \). Thus,

\[
\pi(y) n_{ij} \cdot \nabla_N \frac{\rho}{\pi}(y) = \frac{\pi_i + \pi_j}{2} \frac{\rho(y_j) - \rho(y_i)}{|y_i - y_j|} + O(\text{diam}(C_i) + d_N(y_i, y_j)).
\]

(3.54)

We conclude that

\[
\epsilon_{ij} = O((\text{diam}(C_i) + d_N(y_i, y_j))|\Gamma_{ij}|).
\]

(3.55)

Therefore,

\[
\frac{\epsilon_{ij}^2}{2|y_i, y_j| |\Gamma_{ij}|} = O(d_N(y_i, y_j)(\text{diam}(C_i) + d_N(y_i, y_j))^2|\Gamma_{ij}|).
\]

(3.56)
If we sum up all \( j \in VF(i) \),
\[
(3.57) \quad \sum_{j \in VF(i)} \frac{e_{ij}^2}{\pi_i + \pi_j |\Gamma_{ij}|} = \max_{j \in VF(i)} d_N(y_i, y_j)(\text{diam}(C_i) + d_N(y_i, y_j))^2O(|\partial C_i|).
\]

Hence,
\[
(3.58) \quad \sum_{i} \sum_{j \in VF(i)} \frac{e_{ij}^2}{\pi_i + \pi_j |\Gamma_{ij}|} = O(nh^3 \max_i |\partial C_i|),
\]
where the constant depends on the minimum of \( \pi \), the \( C^1 \) norm of \( \pi \) and the \( C^2 \) norm of \( \frac{\rho}{\pi} \).

Next, we bound \( \sum_i [\partial_t (\rho(y_i) - \rho_i^t)]^2 \frac{|C_i|}{\pi_i} \)
\[
(3.59) \quad \partial_t (\rho(y_i) - \rho_i^t) = O(\text{diam}(C_i)) = O(h),
\]
where the constant depends on the \( L^\infty \) norm of \( \partial_t \nabla N \rho \). Since \( \sum_i |C_i| = \text{Vol}(N) \),
\[
(3.60) \quad \sum_i [\partial_t (\rho(y_i) - \rho_i^t)]^2 \frac{|C_i|}{\pi_i} = O(h^2),
\]
where the constant depends on the \( L^\infty \) norm of \( \partial_t \nabla N \rho \), \( \text{Vol}(N) \) and minimum of \( \pi \). Hence,
\[
(3.61) \quad \frac{d}{dt} \sum_i \frac{e_i^2 |C_i|}{\pi_i} \leq O(h^2(nh \max_i |\partial C_i| + 1)) + \sum_i e_i^2 \frac{|C_i|}{\pi_i}.
\]

In conclusion,
\[
(3.62) \quad \max_{t \in [0, T]} \sum_i e_i(t)^2 \frac{|C_i|}{\pi_i} \leq \left( \sum_i e_i(0)^2 \frac{|C_i|}{\pi_i} + O(h^2(nh \max_i |\partial C_i| + 1)) \right) \exp^T.
\]

\( \Box \)

4. Approximation of Voronoi cells on manifold. Recall that \( \{y_i\}_{i=1}^n \) are sampled from the equilibrium density \( \rho_N^N \) on a smooth closed submanifold \( N \) in \( \mathbb{R}^\ell \). In this section, we first introduce an algorithm to approximate the volumes of the Voronoi cells and the areas of the Voronoi faces constructed from \( \{y_i\}_{i=1}^n \).

When the geodesic distance between two points on \( N \) is small, the next lemma relates the Euclidean distance and the geodesic distance between them. The proof can be found in Lemma B.3 in [44].

**Lemma 3.8.** Suppose \( y, y' \in N \) such that \( d_N(y, y') \) is small enough. Then
\[
(3.63) \quad \|y' - y\|_{\mathbb{R}^\ell} = d_N(y, y')(1 + O(d_N^2(y, y'))),
\]
where the constant in \( O(d_N^2(y, y')) \) depending on the second fundamental form of \( N \) in \( \mathbb{R}^\ell \) at \( y \).

Above lemma implies that if \( r \) is small enough, then for all \( y_k \) and any \( y \in B_r^\mathbb{R}^\ell(y_k) \cap N \), there is a constant \( D_1 > 1 \) depending on the second fundamental form of \( N \) in \( \mathbb{R}^\ell \), such that
\[
(3.64) \quad d_N(y, y_k) \leq D_1 \|y_k - y\|_{\mathbb{R}^\ell}.
\]

We further make the following assumption about the Voronoi cells and the distribution of \( \{y_i\}_{i=1}^n \) on \( N \).
Assumption 3.9. For $n$ large enough, there exists $r$ depending $n$ such that the following conditions hold for any $y_k$:

1. Suppose $\mathcal{B}^\mathbb{R}^d_r(y_k) \cap \{y_i\}_{i=1}^n = \{y_{k,1}, \cdots, y_{k,N_k}\}$. We have $C_k \subset \mathcal{B}^\mathbb{R}^d_r(y_k)$. Moreover, if $\Gamma_{kj}$ is a Voronoi surface of $C_k$ between $y_k$ and $y_j$, then $y_j \notin \mathcal{B}^\mathbb{R}^d_r(y_k)$. Suppose $y_j = y_{k,m}$, then we introduce the notation $\Gamma_{k,m} = \Gamma_{kj}$.

2. For any $i = 1, \cdots N_k$, there is a constant $D_2 < 1$ such that $d_{\mathcal{N}}(y_{k,i}, y_k) \geq D_2 r$.

The following lemma is a consequence of (2) in the above assumption.

Lemma 3.10. Under Assumption 3.9, $d_{\mathcal{N}}(\partial C_k, y_k) \geq \frac{1}{2} D_2 r$. There are constants $K_1$ and $K_2$ depending on $D_1$, $D_2$ and the Ricci curvature of $\mathcal{N}$, such that

$$
(3.65) \quad K_1 r^d \leq |C_k| \leq K_2 r^d,
$$

Proof. Suppose $G_{k,i}$ is the bisector between $y_k$ and $y_{k,i}$. Then $d_{\mathcal{N}}(\Gamma_{k,i}, y_k) \geq d_{\mathcal{N}}(G_{k,i}, y_k) \geq \frac{1}{2} D_2 r$. Hence, $d_{\mathcal{N}}(\partial C_k, y_k) \geq \frac{1}{2} D_2 r$. Therefore, each $C_i$ contains a geodesic ball of radius $\frac{1}{2} D_2 r$ and is contained in the geodesic ball of radius $D_1 r$. By Lemma B.1 in [44] when $r$ is small enough, the volume geodesic ball of radius $r$ can be bounded from below by $K'_1 r^d$ and from above by $K'_2 r^d$ where $K'_1$ and $K'_2$ depend on the Ricci curvature of $\mathcal{N}$. The conclusion follows. \hfill \Box

In order to introduce the Voronoi cell approximation algorithm, we need the following definition.

Definition 3.11. For any $0 < r < 1$ and $y_k \in \{y_i\}_{i=1}^n$, suppose $\mathcal{B}^\mathbb{R}^d_{\sqrt{r}}(y_k) \cap \{y_i\}_{i=1}^n = \{y_{k,1}, \cdots, y_{k,N_k}\}$. We define the discrete local covariance matrix at $y_k$,

$$
(3.66) \quad C_{n,r}(y_k) := \frac{1}{n} \sum_{i=1}^{N_k} (y_{k,i} - y_k)(y_{k,i} - y_k)^\top \in \mathbb{R}^{\ell \times \ell}.
$$

Suppose $\{\beta_{n,r,1}, \cdots, \beta_{n,r,d}\}$ are the first $d$ orthonormal eigenvectors corresponding to $C_{n,r}(y_k)$’s largest $d$ eigenvalues. Define a map $\iota_k(u) : \mathbb{R}^\ell \to \mathbb{R}^d$ as

$$
(3.67) \quad \iota_k(u) := (u^\top \beta_{n,r,1}, \cdots, u^\top \beta_{n,r,d}).
$$

For any $y \in \mathbb{R}^\ell$, define $\bar{\iota}_k(y) = \iota_k(y - y_k)$.

Based on the above definition, we propose the following algorithm to find the approximated volumes $|\bar{C}_k|$ of the Voronoi cells $C_k$ and the approximated areas $|\bar{\Gamma}_{k\ell}|$ of the Voronoi faces $\Gamma_{k\ell}$. 
Algorithm 1: Approximation of the Voronoi cell

**Parameters:** Algorithm inputs are the bandwidth \( r \) and the threshold \( s \)

1. Choose \( 0 < r < 1 \). For each \( y_k \in \{ y_i \}_{i=1}^n \), find
   \[
   B_{\sqrt{r}}(y_k) \cap \{ y_i \}_{i=1}^n =: \{ y_{k,1}, \cdots, y_{k,N_k} \}, \quad B_{r}(y_k) \cap \{ y_i \}_{i=1}^n =: \{ y_{k,1}, \cdots, y_{k,N_k} \}.
   \]

2. Construct the matrix \( C_{n,r}(y_k) \) as in (3.66) by using the \( \{ y_{k,1}, \cdots, y_{k,N_k} \} \). Find the orthonormal eigenvectors corresponding to \( C_{n,r}(y_k) \)’s largest \( d \) eigenvalues. Denote them as \( \{ \beta_{n,r,1}, \cdots, \beta_{n,r,d} \} \).

3. Use \( \{ \beta_{n,r,1}, \cdots, \beta_{n,r,d} \} \) to construct \( \tilde{\iota}_k \) as in (3.67). Find \( v_{k,i} = \tilde{\iota}_k(y_{k,i}) \), for \( i = 1, \cdots, N_k \).

4. Find the Voronoi cell decomposition of \( \{ 0, v_{k,1}, \cdots, v_{k,N_k} \} \) in \( \mathbb{R}^d \). Denote the Voronoi cell containing 0 to be \( \tilde{C}_{k,0} \) and the Voronoi cell containing \( v_{k,i} \) to be \( \tilde{C}_{k,i} \). Denote the face \( \tilde{F}_{k,i} = \tilde{C}_{k,0} \cup \tilde{C}_{k,i} \).

5. Find the approximation of \( |C_k| \) as
   \[
   |\tilde{C}_k| := |\tilde{C}_{k,0}| := \mathcal{H}^d(\tilde{C}_{k,0}).
   \]

6. Find \( |\tilde{F}_{k,i}| = \mathcal{H}^{d-1}(\tilde{F}_{k,i}) \). Define \( \tilde{\Gamma} \in \mathbb{R}^{n \times n} \) such that
   \[
   A_{k\ell} := \frac{\tilde{A}_{k\ell} + \tilde{A}_{\ell k}}{2}, \quad \tilde{A}_{k\ell} = \begin{cases} |\tilde{F}_{k,i}| & \text{if } y_{k,i} = y_{k,i} \in B_{r}(y_k); \\ 0 & \text{otherwise}. \end{cases}
   \]

7. If \( A_{k\ell} \geq s \), then \( |\tilde{\Gamma}_{k\ell}| = A_{k\ell} \). Otherwise \( |\tilde{\Gamma}_{k\ell}| = s \). Then \( |\tilde{\Gamma}_{k\ell}| \) is an approximation of \( |\Gamma_{k\ell}| \).

The idea of the above algorithm can be summarized as follows. For each \( y_k \), by using the points in a larger ball \( B_{\sqrt{r}}(y_k) \), we construct the matrix \( C_{n,r}(y_k) \). Then, the first \( d \) orthonormal eigenvectors will be an approximation of an orthonormal basis of \( T_{y_k} \mathcal{N} \). Next, we project the points in a smaller ball \( B_{r}(y_k) \) onto this tangent space approximation. Now the points around \( y_k \) are projected into a \( d \) dimensional Euclidean space and \( y_k \) is projected to the origin. If we find the Voronoi cell around the origin in the Euclidean space, then it gives the approximation of the Voronoi cell around \( y_k \) in \( \mathcal{N} \). Obviously, the better estimation of the tangent space we have, there are smaller errors in the approximation of the volumes of the Voronoi cells and the areas of the Voronoi faces. To have better estimation of the tangent space, we need that the number of points in a ball goes to infinity as \( n \) goes to infinity. However, to approximate the Voronoi cells and faces, the number of points in a ball should not be too large. This is reason that we choose two different scales \( r \) and \( \sqrt{r} \) in our algorithm.

In the next proposition, we show that \( |\tilde{C}_k| \) is a good approximation of \( |C_k| \). The proof of the proposition is in the Appendix.

**Proposition 3.12.** Let \( |\tilde{C}_k| \) be the approximated volume of \( C_k \) in (3.68). Under Assumption 3.9, if we choose \( r \to 0 \), \( nr^d \) goes to some constant and \( \frac{nr^{d-1}}{\log n} \to \infty \) as \( n \to \infty \), then with probability greater than \( 1 - \frac{1}{n^r} \), for all \( y_k \), we have \( |\tilde{C}_k| = |\tilde{C}_{k,0}| = |C_k|(1 + o(r)) \).

Since we are approximating the tangent plane of the manifold \( \mathcal{N} \), the error between \( |\Gamma_{k,i}| \) and \( |\tilde{\Gamma}_{k,i}| \) will not be much smaller than \( |\Gamma_{k,i}| \) itself when \( |\Gamma_{k,i}| \) is too small. However, in the next proposition,
we show that if $|\Gamma_{ki}|$ is large enough, then $|\tilde{\Gamma}_{ki}|$ is a good approximation of $|\Gamma_{ki}|$. The proof of the proposition is in the appendix.

**Proposition 3.13.** Let $|\tilde{\Gamma}_{ki}|$ be the approximated area of $\Gamma_{ki}$ in (3.69). Let $s = a_1 r^d$ in the last step of Algorithm 1, for some constant $a_1$. Under Assumption 3.3 and Assumption 3.9, if we choose $r \to 0$, $n^d$ goes to some constant and $n \log n \to \infty$ as $n \to \infty$, then for $r$ small enough, with probability greater than $1 - \frac{1}{n^2}$, for all $y_k$, we have

\begin{equation}
|\Gamma_{ki}| = |\tilde{\Gamma}_{ki}| + O(r^d).
\end{equation}

Hence, if $|\Gamma_{ki}| \geq a_2 r^{d-1}$ for some constant $a_2$, then

\begin{equation}
|\Gamma_{ki}| = |\tilde{\Gamma}_{ki}|(1 + O(r))
\end{equation}

At last, if we use our approximation of the volumes of the Voronoi cells and the areas of the Voronoi faces in (3.8) we have the following implementable upwind scheme based only on the collected dataset $\{y_i\} \subset N$

\begin{equation}
\frac{d}{dt} \tilde{\rho}_i|\tilde{C}_i| = \frac{1}{2} \sum_{j \in VF(i)} \frac{\pi_i + \pi_j}{|y_i - y_j|} |\tilde{\Gamma}_{ij}| \left( \frac{\tilde{\rho}_j}{\pi_j} - \frac{\tilde{\rho}_i}{\pi_i} \right).
\end{equation}

Moreover, same as Lemma 3.4, we know the upwind scheme (3.72) is the forward equation for a Markov Process with transition probability $\tilde{P}_{ij}$ and jump rate $\tilde{\lambda}_i$

\begin{equation}
\frac{d}{dt} \tilde{\rho}_i|\tilde{C}_i| = \sum_{j \in VF(i)} \tilde{\lambda}_j \tilde{P}_{ij} \tilde{\rho}_j|\tilde{C}_j| - \tilde{\lambda}_i \tilde{\rho}_i|\tilde{C}_i|,
\end{equation}

where for $i = 1, \cdots, n$, $j = 1, \cdots, n$,

\begin{equation}
\tilde{\lambda}_i := \frac{1}{2|\tilde{C}_i|} \sum_{j \in VF(i)} \frac{\pi_i + \pi_j}{|y_i - y_j|} |\tilde{\Gamma}_{ij}|,
\end{equation}

\begin{equation}
\tilde{P}_{ij} := \frac{1}{\tilde{\lambda}_j \pi_j |\tilde{C}_j|} \frac{|\tilde{\Gamma}_{ij}|}{|y_i - y_j|}, \quad j \in VF(i); \quad \tilde{P}_{ij} = 0, \quad j \notin VF(i).
\end{equation}

It also satisfies the detailed balance property

\begin{equation}
\tilde{\lambda}_j \tilde{P}_{ij} \pi_j |\tilde{C}_j| = \tilde{\lambda}_i \tilde{P}_{ji} \pi_i |\tilde{C}_i|,
\end{equation}

conservation laws and the stability analysis in Lemma 3.6.

Now we state and prove the convergence of the implementable upwind scheme (3.72). The bound of the error in the weighted $\ell^2$ norm is summarized in the following theorem. Due to the estimation error in the Voronoi cells and faces, the error in Theorem 3.7 $e^T$ is replaced by $e^{2T}$. Assume for $i = 1, \cdots, n$, $|VF(i)|$, the cardinality of $VF(i)$, is order 1.

**Theorem 3.14.** Suppose $\rho(y, t), t \in [0, T]$ is a smooth solution to the Fokker-Planck equation (3.1) on manifold $N \subset \mathbb{R}^\ell$ with initial density $\rho^0(y)$. Assume the sampled data $\{y_i\}_{i=1}^n \subset N$ satisfies Assumption 3.9. Let $\tilde{\rho}_i(t)$ be the solution of the upwind scheme (3.72). Let $\tilde{\epsilon}_i := \rho(y_i) - \tilde{\rho}_i$. In Algorithm 1, and we choose threshold $s = a_1 r^d$ for some constant $a_1$. Under Assumptions 3.3, if
we choose \( r \to 0 \), \( nr^d \) goes to some constant and \( \frac{nr^d}{\log n} \to \infty \) as \( n \to \infty \), then for \( r \) small enough, with probability greater than \( 1 - \frac{1}{n^2} \), we have

\[
\max_{t \in [0, T]} \sum_{i} \epsilon_i(t) \frac{\partial_C |C_i|}{\pi_i} \leq \left( \sum_i \epsilon_i(t)^2 \right) \frac{|C_i|}{\pi_i} + cr^2T,
\]

where \( c \) is a constant independent of \( r \) and \( n \).

**Proof.** Define \( \rho_i^e := \int |C_i| \rho dy \). Plug the exact solution into the numerical scheme

\[
\partial_t \rho_i^e|C_i| = \sum_{j \in VF(i)} \left( \frac{\pi_i + \pi_j}{2|y_i - y_j|} \right) \left| \hat{\Gamma}_{ij} \right| \left( \frac{\rho(y_j)}{\pi_j} - \frac{\rho(y_i)}{\pi_i} \right)
\]

\[
\sum_{j \in VF(i)} \int \nabla_N \frac{\rho}{\pi} d\mathcal{H}^{d-1} - \sum_{j \in VF(i)} \pi_i + \pi_j \left( \frac{\pi_i}{2|y_i - y_j|} \right) \left( \frac{\rho(y_j)}{\pi_j} - \frac{\rho(y_i)}{\pi_i} \right),
\]

where \( \pi_i \) is the restriction of the unit outward normal vector field on \( \Gamma_{ij} \). Subtracting the numerical scheme (3.72) from (3.77), we have

\[
\frac{d}{dt} \tilde{\epsilon}_i/C_i = \sum_{j \in VF(i)} \left( \frac{\pi_i + \pi_j}{2|y_i - y_j|} \right) \left| \hat{\Gamma}_{ij} \right| \left( \frac{\tilde{\rho}_i}{\pi_i} - \frac{\tilde{\rho}_j}{\pi_j} \right) + \sum_{j \in VF(i)} \epsilon_{ij} + \partial_t ((\rho(y_i) - \rho_i^e)|C_i|) + \frac{d}{dt} \tilde{\rho}_i(|\tilde{C}_i| - |C_i|),
\]

where

\[
\epsilon_{ij} := \int \nabla_N \frac{\rho}{\pi} d\mathcal{H}^{d-1} - \pi_i + \pi_j \left( \frac{\pi_i}{2|y_i - y_j|} \right) \left( \frac{\rho(y_j)}{\pi_j} - \frac{\rho(y_i)}{\pi_i} \right) + \frac{\pi_i + \pi_j}{2|y_i - y_j|} (|\Gamma_{ij}| - |\tilde{\Gamma}_{ij}|) \left( \frac{\rho(y_j)}{\pi_j} - \frac{\rho(y_i)}{\pi_i} \right)
\]

Note that \( \epsilon_{ij} \) is anti-symmetric, hence by the same argument in Theorem 3.7, we have

\[
\frac{d}{dt} \sum_i \epsilon_i^2 |C_i| \pi_i \leq - \frac{1}{2} \sum_i \sum_{j \in VF(i)} \pi_i + \pi_j \left( \frac{\epsilon_i}{\pi_i} - \frac{\epsilon_j}{\pi_j} \right)^2 + \sum_i \sum_{j \in VF(i)} \left( \frac{\rho(y_j) - \rho_i^e}{\pi_j} \right)^2 |C_i| \pi_i + \sum_i 2 \epsilon_i^2 |C_i| \pi_i
\]

\[
\leq \sum_i \sum_{j \in VF(i)} \left( \frac{\epsilon_i^2}{\pi_i + \pi_j |\Gamma_{ij}|} \right) + \sum_i \left( \frac{\rho(y_i) - \rho_i^e}{\pi_i} \right)^2 |C_i| \pi_i + \sum_i 2 \epsilon_i^2 |C_i| \pi_i
\]

\[
= \epsilon_1 + \epsilon_2 + \epsilon_3 + 2 \sum_i \epsilon_i^2 |C_i| \pi_i.
\]

First, we estimate the term \( \epsilon_1 \), in particular, \( \frac{\epsilon_i^2}{\pi_i + \pi_j |\Gamma_{ij}|} \) for \( j \in VF(i) \). Since the exact solution is smooth such that

\[
|\rho(y_i, t) - \rho(y_j, t)| \leq C_{Lip} |y_i - y_j|,
\]
where we use $n_r$.

Hence,

$$
\epsilon_{ij} = O((\text{diam}(C_i) + d_N(y_i, y_j))|\Gamma_{ij}|) + O(|\Gamma_{ij} - |\bar{\Gamma}_{ij}||).
$$

Note that $|\bar{\Gamma}_{ij}| \geq s = a_1 r^d$. Hence, by Proposition 3.13,

$$
\frac{|\Gamma_{ij} - |\bar{\Gamma}_{ij}||^2}{|\Gamma_{ij}|} = O(r^d).
$$

By Assumption 3.9 and Lemma 3.8, $d_N(y_i, y_j)$ and $\text{diam}(C_i)$ are of order $r$. By Assumption 3.3, since $N$ is compact, there is a constant $K$ such that $|\Gamma_{ij}| \leq K r^{d-1}$. Therefore, $\frac{\epsilon_{ij}^2}{2 |y_i - y_j| |\Gamma_{ij}|} = O(r^{d+1})$ and

$$
\epsilon_1 = \sum_i \sum_{j \in VF(i)} \frac{\epsilon_{ij}^2}{2 |y_i - y_j| |\Gamma_{ij}|} = O(n r^{d+1} \max_i |VF(i)|) = O(r \max_i |VF(i)|),
$$

where we use $nr^d$ goes to some constant in the last step.

Second, we estimate $\epsilon_2 + \epsilon_3$. By Proposition 3.12 and (3.65),

$$
\sum_i \left( \frac{d}{dt} \tilde{\rho}_i \left( \frac{|\tilde{C}_i| - |C_i|}{|C_i|} \right)^2 \frac{|C_i|}{\tilde{\pi}_i} \right) = O(r^2).
$$

By (3.60) and Assumption 3.9

$$
\sum_i [\partial_t (\rho(y_i) - \rho_i^c)]^2 \frac{|C_i|}{\tilde{\pi}_i} = O(r^2).
$$

We sum up all the terms,

$$
\frac{d}{dt} \sum_i \tilde{\epsilon}_i(\tilde{C}_i - \tilde{\pi}_i)^2 \frac{|C_i|}{\tilde{\pi}_i} \leq O(r \max_i |VF(i)|) + 2 \sum_i \tilde{\epsilon}_i^2 \frac{|C_i|}{\tilde{\pi}_i}.
$$

In conclusion

$$
\max_{t \in [0,T]} \sum_i \tilde{\epsilon}_i(t)^2 \frac{|C_i|}{\tilde{\pi}_i} \leq \left( \sum_i \tilde{\epsilon}_i(0)^2 \frac{|C_i|}{\tilde{\pi}_i} + O(r \max_i |VF(i)|) \right) e^{2T}.
$$

3.5. Unconditional stable explicit time stepping and exponential convergence. To the end of this section, we show that the detailed balance property (3.75) leads to stability and exponential convergence of a discrete-in-time Markov process.

Let $\rho_i^k$ be the discrete density at the discrete time $k \Delta t$. To achieve both the stability and the efficiency, we introduce the following unconditional stable explicit scheme

$$
\frac{\rho_i^{k+1}}{\pi_i} = \frac{\rho_i^k}{\pi_i} - \tilde{\lambda}_i \Delta t \frac{\rho_i^k}{\pi_i} + \Delta t \sum_{j \in VF(i)} \tilde{\lambda}_i \tilde{P}_{ji} \frac{\rho_j^k}{\pi_j}.
$$
where \( \hat{\lambda}_i \) and \( \hat{P}_{ij} \) are defined in (3.74). The above equation is equivalent to

\[
\frac{\rho_i^{k+1}}{\pi_i} = \frac{\rho_i^k}{\pi_i} + \frac{\hat{\lambda}_i \Delta t}{1 + \hat{\lambda}_i \Delta t} \left( \sum_{j \in VF(i)} \hat{P}_{ij} \frac{\rho_j^k}{\pi_j} - \frac{\rho_i^k}{\pi_i} \right).
\]

For \( u_i^{k+1} := \frac{\rho_i^{k+1}}{\pi_i} \), the matrix formulation of (3.91) is

\[
u^{k+1} = (I + \Delta t \hat{B}) u^k,
\]

where

\[
\hat{B} := \{ \hat{b}_{ij} \} = \begin{cases} -\frac{\hat{\lambda}_j}{1 + \hat{\lambda}_j \Delta t}, & j = i; \\ \frac{\hat{\lambda}_i - \hat{\lambda}_i}{1 + \hat{\lambda}_i \Delta t} \hat{P}_{ji}, & j \neq i \end{cases}
\]

satisfies \( \sum_j \hat{b}_{ij} = 0 \).

Now we show \( \hat{B} \) is the generator of a new Markov process.

For \( u_i^{k+1} := \rho_i^{k+1} | \tilde{C}_i | \), (3.90), together with detailed balance property (3.75), yields

\[
\rho_i^{k+1} | \tilde{C}_i | - \rho_i^k | \tilde{C}_i | = \Delta t \left( \sum_{j \in VF(i)} \hat{\lambda}_j \hat{P}_{ij} \rho_j^k | \tilde{C}_j | - \hat{\lambda}_i \rho_i^{k+1} | \tilde{C}_i | \right),
\]

which can be recast as

\[
(1 + \Delta t \hat{\lambda}_i) \rho_i^{k+1} | \tilde{C}_i | = (1 + \Delta t \hat{\lambda}_i) \rho_i^k | \tilde{C}_i | + \Delta t \left( \sum_{j \in VF(i)} \hat{\lambda}_j \hat{P}_{ij} \rho_j^k | \tilde{C}_j | - \hat{\lambda}_i | \tilde{C}_i | \rho_i^k \right).
\]

Denote \( g_i^{k+1} := (1 + \Delta t \hat{\lambda}_i) \rho_i^{k+1} | \tilde{C}_i | \), (3.95) can be simplified as

\[
g_i^{k+1} = g_i^k + \Delta t \left( \sum_j \frac{\hat{\lambda}_j}{1 + \Delta t \hat{\lambda}_j} \hat{P}_{ij} g_j^k - \frac{\hat{\lambda}_i}{1 + \Delta t \hat{\lambda}_i} g_i^k \right).
\]

This is a new Markov process for \( g_i \) with transition probability \( \hat{P}_{ij} \) and a new jump rate \( s_j = \frac{\hat{\lambda}_j}{1 + \Delta t \hat{\lambda}_j} \).

With \( \hat{B} \) in (3.93), the matrix formulation for \( g \) is

\[
g^{k+1} = (I + \Delta t \hat{B})^* g^k.
\]

One can check \( (1 + \Delta t \hat{\lambda}_i) \pi_i | \tilde{C}_i | \) is a new equilibrium.

**Proposition 3.15.** Let \( \Delta t \) be the time step and consider the explicit scheme (3.90). Assume the initial data satisfies

\[
\sum_i (1 + \lambda_i \Delta t) \rho_i^0 | \tilde{C}_i | = \sum_i (1 + \lambda_i \Delta t) \pi_i | \tilde{C}_i |.
\]

Then we have

(i) the conversational law for \( g_i^{k+1} := (1 + \Delta t \hat{\lambda}_i) \rho_i^{k+1} | \tilde{C}_i | \), i.e.

\[
\sum_i (1 + \lambda_i \Delta t) \rho_i^{k+1} | \tilde{C}_i | = \sum_i (1 + \lambda_i \Delta t) \rho_i^k | \tilde{C}_i |;
\]
(ii) the unconditional maximal principle for $\hat{\rho}_i$

\[
\max_i \rho_j^{k+1} \leq \max_j \frac{\hat{\rho}_j^k}{\pi_j}.
\]  

(iii) the $\ell^\infty$ contraction

\[
\max_i \left| \frac{\hat{\rho}_i^{k+1}}{\pi_i} - 1 \right| \leq \max_i \left| \frac{\hat{\rho}_i^k}{\pi_i} - 1 \right|,
\]

(iv) the exponential convergence

\[
\left| \frac{\hat{\rho}_i^k}{\pi_i} - 1 \right|_{\ell^\infty} \leq c|\mu_2|^k, \quad |\mu_2| < 1,
\]

where $\mu_2$ is the second eigenvalue (in terms of the magnitude) of $I+\Delta t \hat{B}$, i.e. $\mu_2 = 1 - \text{gap}_{\hat{B}} \Delta t$ and $\text{gap}_{\hat{B}}$ is the spectral gap of $\hat{B}$.

Proof. First, recast (3.91) as

\[
\frac{\hat{\rho}_i^{n+1}}{\pi_i} = \frac{1}{1 + \hat{\lambda}_i \Delta t} \frac{\hat{\rho}_i^n}{\pi_i} + \frac{\hat{\lambda}_i \Delta t}{1 + \hat{\lambda}_i \Delta t} \left( \sum_{j \in VF(i)} \tilde{P}_{ij} \frac{\rho_j^n}{\pi_j} \right).
\]

which gives the unconditional maximal principle (3.100).

Second, from (3.103), we have

\[
\frac{\hat{\rho}_i^{k+1}}{\pi_i} - 1 = \frac{1}{1 + \hat{\lambda}_i \Delta t} \left( \frac{\hat{\rho}_i^k}{\pi_i} - 1 \right) + \frac{\hat{\lambda}_i \Delta t}{1 + \hat{\lambda}_i \Delta t} \sum_{j \in VF(i)} \tilde{P}_{ij} \left( \frac{\rho_j^k}{\pi_j} - 1 \right).
\]

Then we have

\[
\left| \frac{\hat{\rho}_i^{k+1}}{\pi_i} - 1 \right| \leq \frac{1}{1 + \hat{\lambda}_i \Delta t} \left| \frac{\hat{\rho}_i^k}{\pi_i} - 1 \right| + \frac{\hat{\lambda}_i \Delta t}{1 + \hat{\lambda}_i \Delta t} \sum_{j \in VF(i)} \tilde{P}_{ij} \left| \frac{\rho_j^k}{\pi_j} - 1 \right| \leq \max_i \left| \frac{\rho_i^n}{\pi_i} - 1 \right|,
\]

which gives (3.101).

Third, recall the matrix formulation (3.92). Every element in $(I + \Delta t \hat{B})^m$ is strictly positive for some $m$. By Perron-Frobenius theorem, $\mu_1 = 1$ is the simple, principal eigenvalue of $I + \Delta t \hat{B}$ with the ground state $u^* \equiv \{1, 1, \cdots, 1\}$ and other eigenvalues $\mu_i$ satisfy $|\mu_i| < \mu_1$. On one hand, the mass conservation for initial data $u^0$ satisfies (3.99), i.e.,

\[
\sum_i (u^0_i - u^*_i) u^*_i (1 + \Delta t \lambda_i) |C|_i = 0.
\]

On the other hand, $I + \Delta t \hat{B}$ is self-adjoint operator in the weighted $l^2((1 + \Delta t \lambda) |C|)$ space, we can express $u^0$ using

\[
u^0 - u^* = \sum_{j=2} c_j u_j, \quad u_j \text{ is the eigenfunction corresponding to } \mu_j.
\]

Therefore, we have

\[
u^k - u^* = (I + \Delta t \hat{B})^k (u^0 - u^*) = \sum_{j=2} c_j \mu_j^k u_j.
\]
which concludes

\begin{equation}
\left\| \frac{\rho_i^k}{\pi_i} - 1 \right\|_{\infty} \leq c|\mu_2|^k \quad \text{with } |\mu_2| < 1.
\end{equation}

Here \( \mu_2 \) is the second eigenvalue (in terms of the magnitude) of \( I + \Delta t \hat{B} \) sitting in the ball with radius \( \mu_1 = 1 \) and thus \( |\mu_2| < 1 \).

Finally, taking summation with respect to \( i \) in (3.94) shows

\begin{equation}
\sum_i \left( \rho_i^{k+1} |\tilde{C}_i| - \rho_i^k |\tilde{C}_i| \right) = \Delta t \left( \sum_{i,j} \tilde{\lambda}_j \tilde{P}_{ij} \rho_j^k |\tilde{C}_j| - \sum_i \tilde{\lambda}_i \rho_i^{k+1} |\tilde{C}_i| \right)
\end{equation}

which gives (3.99).

\[ \square \]

As a comparison, we also give some other standard stability estimates for explicit and implicit schemes.

Lemma 3.16 below gives the maximal principle, and exponential convergence for an explicit scheme under Courant-Friedrichs-Lewy (CFL) condition. Lemma 3.17 below gives the unconditional maximal principle, and exponential convergence for an implicit scheme.

**Lemma 3.16.** Let \( \Delta t \) be the time step and consider the explicit scheme for (3.73)

\begin{equation}
\frac{\rho_i^{n+1} |\tilde{C}_i| - \rho_i^n |\tilde{C}_i|}{\Delta t} = \left( \sum_{j \in VF(i)} \tilde{\lambda}_j \tilde{P}_{ij} \rho_j^n |\tilde{C}_j| - \tilde{\lambda}_i \rho_i^n |\tilde{C}_i| \right).
\end{equation}

With the detailed balance property (3.75), and the CFL condition for \( \Delta t \)

\begin{equation}
\Delta t \leq \frac{1}{\tilde{\lambda}_i} = \frac{2 |\tilde{C}_i|}{\sum_{j \in VF(i)} |\tilde{\lambda}_j| |\tilde{P}_{ij}|},
\end{equation}

we have

(i) the conversational law for \( \rho_i^{k+1} |\tilde{C}_i| \), i.e.

\begin{equation}
\sum_i \rho_i^{k+1} |\tilde{C}_i| = \sum_i \rho_i^k |\tilde{C}_i|;
\end{equation}

(ii) the equivalent updates for \( u_i^{k+1} = \frac{\rho_i^{k+1}}{\pi_i} \)

\begin{equation}
u^{k+1} = (I + \Delta t B)u^k, \quad \text{with } b_{ij} := \begin{cases} -\tilde{\lambda}_i, & j = i; \\ \tilde{\lambda}_i \tilde{P}_{ji}, & j \neq i; \end{cases}
\end{equation}

(iii) the maximal principle for \( \frac{\rho_i}{\pi_i} \)

\begin{equation}
\max_j \left( \frac{\rho_j^{k+1}}{\pi_j} \right) \leq \max_j \left( \frac{\rho_j^k}{\pi_j} \right).
\end{equation}
(iv) the $\ell^\infty$ contraction

\[
\max_i \left| \frac{\rho_i^{k+1}}{\pi_i} - 1 \right| \leq \max_i \left| \frac{\rho_i^k}{\pi_i} - 1 \right|
\]

(v) the exponential convergence

\[
\left\| \frac{\rho_i^k}{\pi_i} - 1 \right\|_{\ell^\infty} \leq c|\mu_2|^k, \quad |\mu_2| < 1,
\]

where $\mu_2$ is the second eigenvalue (in terms of the magnitude) of $(I + \Delta t B)^{-1}$.

**Lemma 3.17.** Let $\Delta t$ be the time step and consider the implicit scheme

\[
\frac{\rho_i^{n+1}}{\pi_i} = \frac{\rho_i^n}{\pi_i} - \tilde{\lambda}_i \Delta t \frac{\rho_i^{n+1}}{\pi_i} + \Delta t \sum_{j \in VF(i)} \tilde{\lambda}_i \tilde{P}_{ji} \frac{\rho_j^{n+1}}{\pi_j}.
\]

We have the following unconditional properties:

(i) the conversational law for $\rho_i^{k+1}|\tilde{C}_i|$, i.e.

\[
\sum_i \rho_i^{k+1}|\tilde{C}_i| = \sum_i \rho_i^k|\tilde{C}_i|,
\]

(ii) the equivalent updates for $u_i^{k+1} = \frac{\rho_i^{k+1}}{\pi_i}$

\[
(I - \Delta t B) u_i^{k+1} = u_i^k, \quad \text{with } b_{ij} = \begin{cases} -\tilde{\lambda}_i, & j = i; \\ \tilde{\lambda}_i \tilde{P}_{ji}, & j \neq i; \end{cases}
\]

(iii) the maximal principle for $\rho_i^k$

\[
\max_i \frac{\rho_j^{k+1}}{\pi_j} \leq \max_j \frac{\rho_j^k}{\pi_j}.
\]

(iv) the $\ell^\infty$ contraction

\[
\max_i \left| \frac{\rho_i^{k+1}}{\pi_i} - 1 \right| \leq \max_i \left| \frac{\rho_i^k}{\pi_i} - 1 \right|
\]

(v) the exponential convergence

\[
\left\| \frac{\rho_i^k}{\pi_i} - 1 \right\|_{\ell^\infty} \leq c|\mu_2|^k, \quad |\mu_2| < 1,
\]

where $\mu_2$ is the second eigenvalue (in terms of the magnitude) of $(I + \Delta t B)^{-1}$.

The proof of the two lemmas are same as Proposition 3.15 and we omit it. Roughly speaking, the forward equation leads to the conservation law while the backward equation leads to the maximal principle. The advantage of the explicit scheme (3.111) is its efficiency but the disadvantage is the requirement of CFL condition on $\Delta t$. Indeed, the convergence rate for the explicit scheme (3.111) is slow since the spectral gap vanishes as $\Delta t \to 0$. On the other hand, the unconditionally stable implicit scheme (3.118) gives the exponential convergence with fast rate when we take $\Delta t$ large enough but it is not efficient. Therefore, the unconditional stable explicit scheme (3.90) achieves both the efficiency and the stability.
We refer to [31] [8] [34] [20] [21] [23] for more discussions on the corresponding generalized gradient flow of the relative entropy with graph Wasserstein distance on discrete space and Benamou-Brenier formula.

3.6. **Simulations for Fokker-Planck solver.** In this section, we use the dataset \{\(y_i\)\}_{i=1}^{2000} with the reaction coordinates on the underlining manifolds which are dumbbell, torus and sphere to solve the Fokker-Planck equation (3.1) following the unconditionally stable explicit scheme (3.90).

### 3.6.1. Example I: Fokker-Planck evolution on dumbbell.

Suppose \((\theta, \phi) \in [0, 2\pi) \times [0, \pi)\), then we have the following dumbbell in the \(\mathbb{R}^{200}\) parametrized as \((x, y, z, 0, \cdots, 0) = f_1(\theta, \phi) \in \mathbb{R}^{200}\), where

\[
\begin{align*}
    r & = \sqrt{1 + 0.95^4 \cos(2\phi)^2 - 1} + 0.95 \cos(2\phi) \\
    x & = r \sin(\phi) \cos(\theta) \\
    y & = r \sin(\phi) \sin(\theta) \\
    z & = r \cos(\phi).
\end{align*}
\]

After composite with a dilation and rotation map \(f_2\) of \(\mathbb{R}^{200}\), we have an embedded dumbbell \(\mathcal{M} \subset \mathbb{R}^{200}\). \(f_2 \circ f_1(\theta, \phi)\) is the parametrization of \(\mathcal{M}\). We sample 4000 points \((\theta_1, \phi_1), \cdots, (\theta_{4000}, \phi_{4000})\) on \([0, 2\pi) \times [0, \pi)\). Let \(x_i = f_2 \circ f_1(\theta_i, \phi_i)\), then we have a non uniform sample \(\{x_i\}_{i=1}^{4000}\) on \(\mathcal{M}\). We apply the Diffusion map to find the reaction coordinates \(\{y_i\}_{i=1}^{4000}\) of \(\{x_i\}_{i=1}^{4000}\) in \(\mathbb{R}^3\), i.e. \(\{y_i\}_{i=1}^{4000}\) can be regarded a non uniform sample on a dumbbell \(\mathcal{N} \subset \mathbb{R}^3\).

Suppose \(\psi_i\) is the \(i\) th eigenfunction of the Laplace-Beltrami operator on \(\mathcal{N}\). Assume the initial density \(\rho^0\) is \(\psi_2\) plus some constant (so that \(\rho^0\) is positive) as shown in Fig 1. Assume the equilibrium density \(\pi\) is \(\psi_8\) plus some constant as shown in Fig 1. We first obtain the approximated Voronoi cell volumes \(|\tilde{C}_i|_{i=1}^{4000}\) and the areas \(\tilde{\Gamma}_{ij}\) from Algorithm 1 by taking the bandwidth \(r = 0.16\) and threshold \(s = 0\). Then we adjust the initial data, i.e., we replace \(\rho^0\) by \(c \rho^0\) such that (3.98) holds.

We set the time step \(\Delta t = 0.05\). Let \(T = k \Delta t\) for the integer \(k\) and \(1 \leq k \leq 2000\), i.e., we iterate the scheme for 20000 times and set the final time to be \(T = 20000 \ast \Delta t = 1000\). We use the unconditional stable explicit scheme (3.90) to solve \(\rho^k\). We compare the numerical relative error in maximum norm with the theoretic relative error, \(|\mu_2|^k = 0.9997^k\) in (3.102), in the semilog-plot in Fig 2. The exponential convergence rate is exactly same. To clearly see the dynamics of the change of the density over the 4000 points, we plot \(\rho^k\) for \(k = 20, 60, 100, 160, 220, 4000\), correspondingly \(T = 1, 3, 5, 8, 11, 200\) in Fig 3.

### 3.6.2. Example II: Fokker-Planck evolution on torus.

Suppose \(\{y_i\}_{i=1}^{2000}\) are 2000 points sampled on a torus \(\mathcal{N} = T^2 \subset \mathbb{R}^3\), i.e. we can regard \(\{y_i\}_{i=1}^{2000}\) as the reaction coordinates of 2000 points sampled on \(\mathcal{M}\) (a manifold diffeomorphic to a torus) in some high dimensional space. Suppose \(\psi_i\) is the \(i\) th eigenfunction of the Laplace-Beltrami operator on \(\mathcal{N}\). Assume the initial density \(\rho^0\) is \(\psi_2\) plus some constant (so that \(\rho^0\) is positive) as shown in Fig 1. Assume the equilibrium density \(\pi\) is \(\psi_7\) plus some constant as shown in Fig 1. We first obtain the approximated Voronoi cell volumes \(|\tilde{C}_i|_{i=1}^{2000}\) and the areas \(\tilde{\Gamma}_{ij}\) from Algorithm 1 by taking the bandwidth \(r = 0.22\) and threshold \(s = 0\). Then we adjust the initial data, i.e., we replace \(\rho^0\) by \(c \rho^0\) such that (3.98) holds. We set the time step \(\Delta t = 0.05\). Let \(T = k \Delta t\) for
**Figure 1.** Left: The initial density is the second eigenfunction of the Laplace Beltrami operator on a dumbbell $\mathcal{N} \subset \mathbb{R}^3$ plus a constant. We plot it over 4000 points $\{y_i\}_{i=1}^{4000} \subset \mathcal{N} \subset \mathbb{R}^3$. Right: The equilibrium density is the eighth eigenfunction of the Laplace Beltrami operator on a dumbbell $\mathcal{N} \subset \mathbb{R}^3$ plus a constant. We plot it over 4000 points $\{y_i\}_{i=1}^{4000} \subset \mathcal{N} \subset \mathbb{R}^3$.

**Figure 2.** The semilog-plot comparison between the numerical relative error with theoretic relative error. The numerical relative error is the error from the unconditional stable explicit scheme (3.90) with $\Delta t = 0.05$ and $1 \leq k \leq 20000$. The theoretic relative error is based on (3.102) with $|\mu_2|^k = 0.9997^k$. 
Figure 3. The density dynamics $\rho^k$ from the unconditional stable explicit scheme (3.90) with $\Delta t = 0.05$. We plot $\rho^k$ for $k = 20, 60, 100, 160, 220, 4000$, correspondingly on time $T = 1, 3, 5, 8, 11, 200$.

3.6.3. Example III: The “breakup” of Pangaea via Fokker-Planck evolution on sphere.
In this example, we use the Fokker-Planck evolution on sphere to simulate the dynamics of the
Figure 4. Left: The initial density is the second eigenfunction of the Laplace Beltrami operator on a torus $\mathcal{N} \subset \mathbb{R}^3$ plus a constant. We plot it over 2000 points $\{y_i\}_{i=1}^{2000} \subset \mathcal{N} \subset \mathbb{R}^3$. Right: The equilibrium density is the seventh eigenfunction of the Laplace Beltrami operator on a torus $\mathcal{N} \subset \mathbb{R}^3$ plus a constant. We plot it over 2000 points $\{y_i\}_{i=1}^{2000}$.

Figure 5. The semilog-plot comparison between the numerical relative error with theoretic relative error. The numerical relative error is the error from the unconditional stable explicit scheme (3.90) with $\Delta t = 0.05$ and $1 \leq k \leq 10000$. The theoretic relative error is base on (3.102) with $|\mu_2|^k = 0.9992^k$. 
Figure 6. The density dynamics $\rho^k$ from the unconditional stable explicit scheme (3.90) with $\Delta t = 0.05$. We plot $\rho^k$ for $k = 20, 60, 100, 160, 220, 2000$, correspondingly on time $T = 1, 3, 5, 8, 11, 100$.

altitude of continents and the depth of oceans for earth based on the dataset for initial distribution of Pangaea and the equilibrium distribution of the current earth.

Suppose $\{y_i\}_{i=1}^{2000}$ are the points on the unit sphere $\mathcal{N} = S^2 \subset \mathbb{R}^3$, i.e., $\{y_i\}_{i=1}^{2000}$ are the reaction coordinates of 2000 points on $\mathcal{M}$ (a manifold diffeomorphic to a sphere) in some high dimensional space. Assume the initial density $\rho_0^i$ at $\{y_i\}$ are extracted from the Pangaea continents map file [1] as shown in Fig.7 (down left). Assume the equilibrium $\{\pi_i\}$ at $\{y_i\}$ are collected from the ETOPO5 topography data [2] expressing the altitude of continents and the depth of oceans for earth.

We first obtain the approximated Voronoi cell volumes $|\tilde{C}_i|_{i=1}^n$ and areas $\tilde{\Gamma}_{ij}$ from Algorithm 1 by taking the bandwidth $r = 0.3$ and threshold $s = 0$. Then after an adjustment to the initial data, i.e., replacing $\rho_0$ by $c\rho_0$ such that (3.98) holds. Then we adjust the initial data, i.e., we replace $\rho_0$ by $c\rho_0$ such that (3.98) holds. We set the time step $\Delta t = 0.05$. Let $T = k\Delta t$ for the integer $k$ and $1 \leq k \leq 10000$, i.e., we iterate the scheme for 10000 times and set the final time to be $T = 10000 \times \Delta t = 500$. We use the unconditional stable explicit scheme (3.90) to solve $\rho^k$. In Fig.7 (up), the numerical relative error in maximum norm is semilog-plotted using circles. Compared with decay of the theoretic relative error $|\mu_2|^k$ in (3.102), blue line in the semilog-plot, the exponential convergence rate is exactly same. The initial 3D plot of Pangaea continents is shown in Fig.7 (down left) while the final 3D plot at $T = 500$ of the simulated altitude and depth...
Figure 7. Simulations for the density dynamics of altitude and depth of continents and oceans starting from pangaea (down left) to the final altitude of land-ocean (down right) with parameters $dt = 0.05$, $T = 500$. (up) The semilog-plot comparison between the numerical relative error in maximum norm (blue circle) with theoretic relative error $|\mu_2|^k = 0.9985^k$ (blue line) in (3.102).

of continents and oceans are shown in Fig 7 (down right). To clearly see the dynamics of altitude and depth of continents and oceans at $n$ points with longitude and latitude, starting from same pangaea continents with time step $\Delta t = 0.05$, four snapshots at $T = 0, 25, 50, 75$ of the dynamics are shown in Fig 8.

4. Reproduce the equilibrium potential $U_N(y; \theta)$ for new physical system

In this section, we focus on the goal (II), i.e. use the collected data for the equilibrium $U_N(y, \theta^{(1)})$ to generate new equilibrium $U_N(y, \theta^{(2)})$ for different parameters $\theta^{(2)}$. Suppose $\mathcal{N}$ is a $d$ dimensional smooth closed Riemannian submanifold of $\mathbb{R}^d$. Suppose $U_N$ is an unknown function on $\mathcal{N}$ such

\[1\] The altitude and depth exceed the range $[-3800m, 3800m]$ is cut off for clarity.
that $U_N$ also depends on $q$ independent parameters, i.e.

$$U_N(y, \theta) : N \times \mathbb{R}^q \to \mathbb{R},$$

Suppose we have observations $\tilde{z}_j(i)$ of $U_N$ with some noise over the labeled points $(y_i, \theta_j) \in N \times \mathbb{R}^q$ for $1 \leq i \leq m$ and $1 \leq j \leq r$, i.e.

$$\tilde{z}_j(i) = U_N(y_i, \theta_j) + \tilde{\sigma}_j(i),$$

where $\tilde{z}_j = (z_j(1), \cdots, z_j(m)), \tilde{\sigma}_j = (\sigma_j(1), \cdots, \sigma_j(m)) \in \mathbb{R}^m$. Here, $\tilde{\sigma}_j$ is a Gaussian noise vector and we assume that $\sigma_j(i) \in \mathcal{N}(0, \sigma^2_{\text{noise}})$ for all $i, j$. The goal of this section is to predict the value of $U_N$ over some other unlabeled points $\{y_i\}$ on $N$ for $m+1 \leq i \leq m+n$ and unlabeled parameters $\theta_j \in \mathbb{R}^q$ for $r+1 \leq j \leq r+s$.

Below we denote $y_1, y_2, \cdots, y_m, y_{m+1}, \cdots, y_{m+n}$ as $m+n$ points on $N$, $\theta_1, \cdots, \theta_r, \theta_{r+1}, \cdots, \theta_{r+s}$ as $r+s$ points in $\mathbb{R}^q$. Suppose we choose a Gaussian Process (GP) prior for the unknown function as $U_N \sim GP(0, C)$, where $C$ is the covariance function which is a function on $N \times \mathbb{R}^q \times N \times \mathbb{R}^q$. We discuss here the choice for the covariance function. Note that since $y$ and $\theta$ are independent, $N \times \mathbb{R}^q$ can be regarded as a product manifold with the product metric. On the other hand, the heat kernel on a manifold is related to the geometric structure of the manifold. For example, Varadhan’s formula [13] relates the heat kernel on a closed manifold to the geodesic distance. Hence, to take the intrinsic geometric structure of the manifold $N \times \mathbb{R}^q$ into account, we choose the covariance...
function $\mathcal{C}$ to be the heat kernel on $\mathcal{N} \times \mathbb{R}^q$. Since $\mathcal{N} \times \mathbb{R}^q$ has the product metric, the Laplace Beltrami operator on $\mathcal{N} \times \mathbb{R}^q$ is $\Delta_\mathcal{N} + \Delta_{\mathbb{R}^q}$, where $\Delta_\mathcal{N}$ is the Laplace Beltrami operator on $\mathcal{N}$ and $\Delta_{\mathbb{R}^q}$ is the Laplace Beltrami operator on $\mathbb{R}^q$. Therefore, if $(y, \theta)$ and $(y', \theta')$ are two points on $\mathcal{N} \times \mathbb{R}^q$, then the heat kernel on $\mathcal{N} \times \mathbb{R}^q$ with diffusion time $t$ is

$$\begin{align*}
H^N_{heat}(t, (y, \theta), (y', \theta')) = \frac{1}{(4\pi t)^{k/2}} e^{-\frac{\|y - y'\|^2}{4t}} H^N_{heat}(t, y, y'),
\end{align*}$$

where $H^N_{heat}(t, y, y')$ is the heat kernel on $\mathcal{N}$. Hence, the GP prior for the unknown function is $U_N \sim GP(0, H^N_{heat}(t))$. The diffusion time $t$ can be regarded as the bandwidth of the covariance function.

Denote $\rho \in \mathbb{R}^{mr}$ to be the discretization of $U_N$ over $(y_i, \theta_j)$ for $1 \leq i \leq m$ and $1 \leq j \leq r$, so that $\rho((j-1)m+i) = U_N(y_i, \theta_j)$. Denote $\mathbf{z} \in \mathbb{R}^{mr}$ with $\mathbf{z}((j-1)m+i) = \tilde{z}_j(i)$ for $1 \leq i \leq m$ and $1 \leq j \leq r$. Denote $\mathbf{\rho}_* \in \mathbb{R}^{mr+ms+ns}$ with

$$\begin{align*}
\mathbf{\rho}_* = \begin{bmatrix}
\mathbf{\rho}_1^1 \\
\mathbf{\rho}_2^2 \\
\mathbf{\rho}_3^3
\end{bmatrix},
\end{align*}$$

Specifically, we have

$$\begin{align*}
\mathbf{\rho}_1^1((j-1)n+i) &= U_N(y_i, \theta_j), \quad \text{for } m+1 \leq i \leq m+n \text{ and } 1 \leq j \leq r, \\
\mathbf{\rho}_2^2((j-1)m+i) &= U_N(y_i, \theta_j), \quad \text{for } 1 \leq i \leq m \text{ and } r+1 \leq j \leq r+s, \\
\mathbf{\rho}_3^3((j-1)n+i) &= U_N(y_i, \theta_j), \quad \text{for } m+1 \leq i \leq m+n \text{ and } r+1 \leq j \leq r+s.
\end{align*}$$

In other words, $\mathbf{\rho}_1^1$ is the discretization of $U_N$ over the labeled parameters but unlabeled points on the manifold $\mathcal{N}$. $\mathbf{\rho}_2^2$ is the discretization of $U_N$ over the labeled points on the manifold $\mathcal{N}$ but the unlabeled parameters. $\mathbf{\rho}_3^3$ is the discretization of $U_N$ over the unlabeled parameters and unlabeled points on the manifold $\mathcal{N}$. Obviously given $\mathbf{z}$, the prediction of $\mathbf{\rho}_1^1$ and $\mathbf{\rho}_2^2$ are simply regression on $\mathcal{N}$ and $\mathbb{R}^q$. Hence, we are more interested in the prediction of $\mathbf{\rho}_3^3$.

Under the GP prior for $U_N$, the joint distribution of $\mathbf{\rho}$ and $\mathbf{\rho}_*$ is still Gaussian with

$$\begin{align*}
p(\mathbf{\rho}, \mathbf{\rho}_*) = \mathcal{N}(0, C),
\end{align*}$$

where $C$ is the covariance matrix induced from the kernel $H^N_{heat}(t, (y, \theta), (y', \theta'))$. Specifically, $C$ can be expressed in the following way. Let $\Sigma_{ij} = H^N_{heat}(t, y_i, \theta_j)$ for $1 \leq i, j \leq m+n$, which is an $(m+n) \times (m+n)$ covariance matrix. We decompose it into block matrices

$$\begin{align*}
\Sigma = \begin{bmatrix}
\Sigma_1 & \Sigma_2 \\
\Sigma_3 & \Sigma_4
\end{bmatrix},
\end{align*}$$

where $\Sigma_1$ is an $m \times m$ block matrix. Let $\Sigma'_{ij} = \frac{1}{(4\pi t)^{k/2}} e^{-\frac{\|y_i - y_j\|^2}{4t}}$ for $1 \leq i, j \leq r+s$, which is a $(r+s) \times (r+s)$ covariance matrix. We decompose it into block matrices

$$\begin{align*}
\Sigma' = \begin{bmatrix}
\Sigma'_1 & \Sigma'_2 \\
\Sigma'_3 & \Sigma'_4
\end{bmatrix},
\end{align*}$$
where $\Sigma_1$ is a $r \times r$ block matrix. Then, the covariance matrix $C$ is an $(m+n)(r+s) \times (m+n)(r+s)$ matrix such that $C = \Sigma' \otimes \Sigma$. Moreover, $C$ can be decomposed as

\begin{equation}
C = \begin{bmatrix}
\Sigma_1' \otimes \Sigma_1 & \Sigma_1' \otimes \Sigma_1 & \Sigma_2' \otimes \Sigma_1 & \Sigma_2' \otimes \Sigma_1 \\
\Sigma_1' \otimes \Sigma_3 & \Sigma_1' \otimes \Sigma_3 & \Sigma_2' \otimes \Sigma_3 & \Sigma_2' \otimes \Sigma_3 \\
\Sigma_3' \otimes \Sigma_1 & \Sigma_3' \otimes \Sigma_1 & \Sigma_4' \otimes \Sigma_1 & \Sigma_4' \otimes \Sigma_1 \\
\Sigma_3' \otimes \Sigma_3 & \Sigma_3' \otimes \Sigma_3 & \Sigma_4' \otimes \Sigma_3 & \Sigma_4' \otimes \Sigma_3 
\end{bmatrix}.
\end{equation}

Denote $C_1 = \Sigma_1' \otimes \Sigma_1$, then $C$ can be decomposed as

\begin{equation}
C = \begin{bmatrix}
C_1 & C_2 \\
C_2^T & C_3
\end{bmatrix}.
\end{equation}

Under the regression model and the GP prior, we have

\begin{equation}
p(\mathbf{z}, \mathbf{\rho}_s) = \mathcal{N}(0, \bar{C}),
\end{equation}

where

\begin{equation}
\bar{C} = C + \begin{bmatrix}
\sigma_{\text{noise}}^2 I_{mr \times mr} & 0 \\
0 & 0
\end{bmatrix} = \begin{bmatrix}
C_1 + \sigma_{\text{noise}}^2 I_{mr \times mr} & C_2 \\
C_2^T & C_3
\end{bmatrix}.
\end{equation}

By a direct calculation, the predictive distribution is

\begin{equation}
p(\mathbf{\rho}_s | \mathbf{z}) = \mathcal{N}(C_2^T (C_1 + \sigma_{\text{noise}}^2 I_{mr \times mr})^{-1} \mathbf{z}, C_3 - C_2^T (C_1 + \sigma_{\text{noise}}^2 I_{mr \times mr})^{-1} C_2).
\end{equation}

Then above formula implies that our prediction for $\mathbf{\rho}_s$ is

\begin{equation}
\Sigma_3' \otimes \Sigma_3 (\Sigma_1' \otimes \Sigma_1 + \sigma_{\text{noise}}^2 I_{mr \times mr})^{-1} \mathbf{z}
\end{equation}

To find an approximation for $\Sigma$, we use the Diffusion-based Gaussian Process(DBGP) proposed in \[14\]. The idea of DBGP comes from the spectral representation of the heat kernel. In fact, let $\{\lambda_i^N\}$ be the eigenvalues of $-\Delta$, and

\begin{equation}
\Delta^N \psi_i^N = -\lambda_i^N \psi_i^N,
\end{equation}

where $\psi_i^N$ is the corresponding eigenfunction normalized in $L^2(\mathcal{N})$. We have $0 = \lambda_0^N \leq \lambda_1^N \leq \lambda_2^N \leq \cdots$. Then the heat kernel of $\mathcal{N}$ can be represented as

\begin{equation}
H_{\text{heat}}^N(t, \mathbf{y}, \mathbf{y}') = \sum_{i=0}^{\infty} e^{-\lambda_i^N t} \psi_i^N(\mathbf{y}) \psi_i^N(\mathbf{y}').
\end{equation}

If we apply the Diffusion map to the points $\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_m, \mathbf{y}_{m+1}, \cdots, \mathbf{y}_{m+n}$, and construct the matrix $L_{\epsilon,1}$ as in (2.4). Denote

\begin{equation}
\lambda_{0,m+n,\epsilon} \leq \cdots \leq \lambda_{m+n-1,m+n,\epsilon},
\end{equation}

to be the eigenvalues of $\frac{L_{\epsilon,1} - L}{\epsilon^2}$. And let $V_{i,m+n,\epsilon}$ be the corresponding eigenvectors normalized as (2.10). Let

\begin{equation}
H^K_{\epsilon,t} = \sum_{i=0}^{K-1} e^{-\lambda_{i,m+n,\epsilon} t} V_{i,m+n,\epsilon} V_{i,m+n,\epsilon}^T \in \mathbb{R}^{n \times n}.
\end{equation}
By Theorem 2.6, we expect that, for suitable choices of $\epsilon$, $K$ and $t$ based on $m + n$, the $i,j$th entry of $H_{\epsilon,t}^K$ is an approximation of $H_{\text{heat}}^N(t,y_i,y_j)$ for all $i$ and $j$ with high probability. Hence, we can use $H_{\epsilon,t}^K$ to approximate $\Sigma$.

Note that for $(y, \theta) \in \mathcal{N} \times \mathbb{R}^q$, $y$ and the components of $\theta$, $\theta(1), \ldots, \theta(q)$, are independent parameters of the function $U_\mathcal{N}$. Therefore, it is reasonable to impose different bandwidths corresponding to different parameters in constructing the covariance function. In fact, this is the motivation that we apply DBGP only on $\mathcal{N}$ rather than $\mathcal{N} \times \mathbb{R}^q$ so that we can make choices of the bandwidths adaptive. Hence, we introduce the bandwidths $t_0, t_1, \ldots, t_q$. For $1 \leq i,j \leq r + s$, we define

$$
\Sigma_{ij}' = \prod_{k=1}^q \frac{1}{(4\pi t_k)^{1/2}} e^{\frac{|\theta_i(k) - \theta_j(k)|^2}{4t_k}},
$$

(4.20)

and $\Sigma = H_{\epsilon,t_0}^K$. Then $C = \Sigma' \otimes \Sigma$ is the covariance matrix. We summarize our algorithm as follows.
Algorithm 2: DBGP on Product manifold ALGORITHM

Parameters:
(1) Bandwidths $\epsilon, t_0, t_1, \cdots t_q$.
(2) Number of eigenpairs in the Diffusion map $K$.
(3) Variance of the error $\sigma_{\text{noise}}$.
(4) $z(j - 1) \mu + i = \bar{z}_j(i)$ for $1 \leq i \leq m$ and $1 \leq j \leq r$, where $\bar{z}_j(i)$ defined in (4.2) is the observations over the labeled points under labeled parameters.

1 Construct the $(m + n) \times (m + n)$ matrices $W_{\epsilon,1}$ and $D_{\epsilon,1}$ as in (2.2) and (2.3) by using the bandwidth $\epsilon$ and data points $\{y_1, y_2, \cdots, y_m, y_{m+1}, \cdots, y_{m+n}\}$. Let $\bar{\tilde{L}}_{\epsilon,1} - I_{\epsilon,2} = \frac{D_{\epsilon,1}^{-1/2}W_{\epsilon,1}D_{\epsilon,1}^{-1/2}}{\epsilon^2}$.

2 Find the first $K$ eigenpairs of $\frac{\bar{\tilde{L}}_{\epsilon,1} - I_{\epsilon,2}}{\epsilon^2}$, namely $\{\lambda_{i,m+n,\epsilon}, U_{i,m+n,\epsilon}\}_{i=0}^{K-1}$. Let $\tilde{v}_{i,m+n,\epsilon}$ be the normalized vector of $D_{\epsilon,1}^{-1/2}U_{i,m+n,\epsilon}$ in the $l^2$ norm.

3 For $j = 1, \cdots, m + n$, find
$$N(j) = |B_{\epsilon}^{\text{gp}}(y_j) \cap \{y_1, \cdots, y_{m+n}\}|.$$ Calculate
$$\|\tilde{v}_{i,m+n,\epsilon}\|_2(1/\rho) = \sqrt{|S_{d-1}| \sum_{j=1}^{n} \tilde{v}_{i,m+n,\epsilon}^2(j) \frac{N(j)}{d}}.$$ For $i = 0, \cdots, K - 1$, construct
$$V_{i,n,\epsilon} = \frac{\tilde{v}_{i,n,\epsilon}}{\|\tilde{v}_{i,n,\epsilon}\|_2(1/\rho)}.$$

4 Construct $H_{\epsilon,t_0}^{K}$ as described in (4.19). Rewrite $H_{\epsilon,t_0}^{K}$ as
$$H_{\epsilon,t_0}^{K} = \left[\begin{array}{cc} \Sigma_1 & \Sigma_2 \\ \Sigma_3 & \Sigma_4 \end{array}\right],$$
where $\Sigma_1$ is an $m \times m$ matrix and $\Sigma_2, \Sigma_3, \Sigma_4$ are block matrices.

5 For $\{\theta_1, \cdots, \theta_r, \theta_{r+1}, \cdots, \theta_{r+s}\}$, construct the $(r + s) \times (r + s)$ matrix
$$\Sigma'_{ij} = \prod_{k=1}^{q} \frac{1}{(4\pi t_k)^{1/2}} \exp\left(-\frac{\theta_{i}(k) - \theta_{j}(k)^2}{4t_k}\right).$$ Rewrite $\Sigma'$ as
$$\Sigma' = \left[\begin{array}{cc} \Sigma'_1 & \Sigma'_2 \\ \Sigma'_3 & \Sigma'_4 \end{array}\right],$$
where $\Sigma'_1$ is an $r \times r$ matrix and $\Sigma'_2, \Sigma'_3, \Sigma'_4$ are block matrices.

6 Given $z \in \mathbb{R}^{mr}$,
$$\rho^3 = \Sigma_3 \otimes \Sigma_3(\Sigma'_1 \otimes \Sigma_1 + \sigma^2_{\text{noise}}I_{mr \times mr})^{-1}z$$ is our prediction.
The algorithm involves parameters $\epsilon$, $K$, $t_0, t_1, \cdots, t_q$ and $\sigma_{\text{noise}}$. We propose to estimate these parameters by maximizing the marginal likelihood,

$$p(z|\epsilon, K, t_0, t_1, \cdots, t_q, \sigma_{\text{noise}}) \propto \frac{1}{\det(\Sigma_1^{\prime} \otimes \Sigma_1 + \sigma_{\text{noise}}^2 I_{mr \times mr})} e^{-z^{\top} (\Sigma_1^{\prime} \otimes \Sigma_1 + \sigma_{\text{noise}}^2 I_{mr \times mr}) z}.$$

There is a simple generalization of the above discussion. Suppose $\mathcal{N}$ is a $d$ dimensional smooth closed Riemannian submanifold of $\mathbb{R}^d$ and $\mathcal{N}'$ is a $d'$ dimensional smooth closed Riemannian submanifold of $\mathbb{R}^{d'}$. Suppose $U_{\mathcal{N}}(y, \theta)$ is an unknown function such that

$$U_{\mathcal{N}}(y, \theta): \mathcal{N} \times \mathcal{N}' \to \mathbb{R}.$$

Suppose that $y_1, y_2, \cdots, y_m, y_{m+1}, \cdots, y_{m+n}$ are $m+n$ points on $\mathcal{N}$. $\theta_1, \cdots, \theta_r, \theta_{r+1}, \cdots, \theta_{r+s}$ are $r+s$ points in $\mathcal{N}'$. We make the observations of $U_{\mathcal{N}}$ over the labeled points $(y_i, \theta_j)$ for $1 \leq i \leq m$ and $1 \leq j \leq r$, i.e.

$$z_j(i) = U_{\mathcal{N}}(y_i, \theta_j) + \sigma_j(i),$$

where $z_j = (z_j(1), \cdots, z_j(m))$, $\sigma_j = (\sigma_j(1), \cdots, \sigma_j(m)) \in \mathbb{R}^m$.

Since $\mathcal{N} \times \mathcal{N}'$ has the product metric, if $(y, \theta)$ and $(y', \theta')$ are two points on $\mathcal{N} \times \mathcal{N}'$, then the heat kernel on $\mathcal{N} \times \mathcal{N}'$ with diffusion time $t$ is

$$H_{\text{heat}}^{\mathcal{N} \times \mathcal{N}'}(t, (y, \theta), (y', \theta')) = H_{\text{heat}}^\mathcal{N}(t, \theta, \theta')H_{\text{heat}}^\mathcal{N}'(t, y, y'),$$

where $H_{\text{heat}}^\mathcal{N}(t, y, y')$ and $H_{\text{heat}}^\mathcal{N}'(t, \theta, \theta')$ are the heat kernels on $\mathcal{N}$ and $\mathcal{N}'$ respectively. The diffusion time $t$ can be regarded as the bandwidth of the covariance function. Since $y$ and $\theta$ are independent, we can select the different bandwidths for $y$ and $\theta$ separately in the covariance function. Therefore, we can choose the GP prior for the unknown function to be $U_{\mathcal{N}} \sim GP(0, H_{\text{heat}}^\mathcal{N}(t_1)H_{\text{heat}}^\mathcal{N}'(t_2))$.

Through (4.19), we can construct $H_{\mathcal{N} \times \mathcal{N}'}^{K_1, \mathcal{N}}$ and $H_{\mathcal{N} \times \mathcal{N}'}^{K_2, \mathcal{N}'}$ by $y_1, \cdots, y_{m+n}$ and $\theta_1, \cdots, \theta_{r+s}$ respectively. Then, $H_{\mathcal{N} \times \mathcal{N}'}^{K_1, \mathcal{N}} \otimes H_{\mathcal{N} \times \mathcal{N}'}^{K_2, \mathcal{N}'}$ is an approximation to the covariance matrix induced by $H_{\text{heat}}^\mathcal{N}(t_1)H_{\text{heat}}^\mathcal{N}'(t_2)$ for GP regression.

4.1. Simulations on dumbbell.

Suppose $(\theta, \phi) \in [0, 2\pi) \times [0, \pi)$, then we have the following dumbbell in the $\mathbb{R}^3$ parametrized as $(x, y, z, 0, \cdots, 0) = f_1(\theta, \phi) \in \mathbb{R}^{200}$, where $f_1$ is described in (3.124). After composite with a dilation and rotation map $f_2$ of $\mathbb{R}^{200}$, we have an embedded dumbbell $\mathcal{M} \subset \mathbb{R}^{200}$. $f_2 \circ f_1(\theta, \phi)$ is the parametrization of $\mathcal{M}$. We choose 400 points $(\theta_1, \phi_1), \cdots, (\theta_{400}, \phi_{400})$ on a grid on $[0, 2\pi) \times [0, \pi)$ and we sample 4000 points $(\theta_{401}, \phi_{401}), \cdots, (\theta_{4400}, \phi_{4400})$ on $[0, 2\pi) \times [0, \pi)$. Let $x_i = f_2 \circ f_1(\theta_i, \phi_i)$, then we have a non uniform sample $\{x_i\}_{i=1}^{4400}$ on $\mathcal{M}$. We apply the Diffusion map to find the reaction coordinates $\{y_i\}_{i=1}^{400}$ of $\{x_i\}_{i=1}^{4400}$ in $\mathbb{R}^3$ as shown in Fig[9] i.e. $\{y_i\}_{i=1}^{400}$ can be regarded a non uniform sample on a dumbbell $\mathcal{N} \subset \mathbb{R}^3$.

Suppose $U_{\mathcal{N}}(y, \eta)$ is an unknown function on $\mathcal{N}$ depending on the parameter $\eta$. Here, we explain where the actual values of $U_{\mathcal{N}}(y, \eta)$ come from. $U_{\mathcal{N}}(y, \eta)$ is supposed to be the equilibrium potential that can be used to establish the Fokker-Planck equation (1.6). However, in order to show the effort of the DBGP algorithm, without loss of generality, we assume $U_{\mathcal{N}}(y, \eta)$ is modified from a solution of equation (1.0). More precisely, suppose $\psi_i$ is the $i$ th eigenfunction of the Laplace-Beltrami operator on $\mathcal{N}$. Assume the initial density $\rho^0$ is $\psi_{12} plus some constant (so that $\rho^0$ is positive). Assume the equilibrium density $\pi$ is $\psi_2$ plus some constant. We first obtain the approximated
5. Cluster, coarse-graining and transition state theory on manifold

The upwind scheme we constructed in Section 3 also enables us to do the cluster and coarse-graining on manifold. The approximated transition probability we constructed efficiently shows the dynamics on the manifold and no other information from the manifold is needed. Based on it, to deal with goal (V), we conduct a few examples including cluster, coarse-graining and most probable transition path in this section.

5.1. Cluster and coarse-graining. We study a continuous time Markov process on a finite state space $S = \{1, \ldots, N\}$ with the transition probability (from site $j$ to $i$) $P_{ij}$ and the jumping rate $\lambda_i$.
Figure 9. For \( \{x_i\}_{i=1}^{4400} \) sampled from a dumbbell in \( \mathbb{R}^{200} \), we find the reaction coordinates \( \{y_i\}_{i=1}^{4400} \) of \( \{x_i\}_{i=1}^{4400} \) in \( \mathbb{R}^3 \) after applying the Diffusion map.

Figure 10. Actual values of the equilibrium potential \( U_N(y, \eta) \) for \( \eta = 0, 0.1, 0.2, \cdots, 0.8 \) over the 4400 points \( \{y_i\}_{i=1}^{4400} \) on the chart \( [0, 2\pi) \times [0, \pi) \) of the dumbbell.

obtained in Section 3. To obtain a simpler but efficient representation, also known as coarse-grained model, for the network with finite states, one need to partition the network into important components/clusters. In this section, we follow the idea of [15, 26, 29] that using the transition probability of the Markov chain to efficiently partition the network.

Denote \( f_i := \rho_i |C_i| \) and \( \pi_i := \pi_i |C_i| \) (with a slight abuse of notation) and the forward equation for the Markov process (3.9) reads

\[
\frac{d}{dt} f_i = \sum_{j=1}^{N} P_{ij} \lambda_j f_j - \lambda_i f_i, \quad i = 1, \cdots, N.
\]
Figure 11. We plot the observations of $U_{\mathcal{N}}(y, \eta)$ with noise for $\eta = 0, 0.1, 0.2, \cdots, 0.8$ over the 400 labeled points $\{y_i\}_{i=1}^{400}$ on the chart $[0, 2\pi) \times [0, \pi)$ of the dumbbell.

Figure 12. We compare the actual value of $U_{\mathcal{N}}(y, 0.44)$ and our prediction of $U_{\mathcal{N}}(y, 0.44)$ over the 4000 unlabeled points $\{y_i\}_{i=400}^{4000}$ on the chart $[0, 2\pi) \times [0, \pi)$ of the dumbbell. The root mean square error is 0.516.

Notice that the transition probability satisfies
\begin{equation}
(5.2) \quad P_{ij} \geq 0, \quad \sum_i P_{ij} = 1.
\end{equation}

We call the Markov process satisfies detailed balance property if there exists a distribution $\pi_i$ such that
\begin{equation}
(5.3) \quad P_{ij} \lambda_j \pi_j = P_{ji} \lambda_i \pi_i.
\end{equation}
We describe how to do the $K$-clusters for $S = S_1 \cup \cdots \cup S_K$ and determine an optimal coarse-graining Markov process between clusters with transition probability $\hat{P}_{mn}^*$, $m, n = 1, \cdots, K$. Here $K$ is the Perron index and below, we do the Perron cluster analysis (c.f. [11]).

Notice we do not allow empty sets in the partition $S = S_1 \cup \cdots \cup S_K$. For a fixed $K$-partition, define the coarse-graining distribution $g_m := \sum_{i \in S_m} \lambda_i \pi_i$, and $\hat{\pi}_m := \frac{g_m}{\sum_{m=1}^{K} g_m}$, $\hat{\lambda}_m := \frac{g_m}{\hat{\pi}_m}$. Therefore we have
\begin{equation}
\hat{\lambda}_m \hat{\pi}_m = \sum_{i \in S_m} \lambda_i \pi_i,
\end{equation}
and define the conditional probability on $S_m$
\begin{equation}
\mu_m(i) = \frac{\lambda_i \pi_i}{\hat{\lambda}_m \hat{\pi}_m},
\end{equation}
where $\chi_m(i) = 1$ if $i \in S_m$ and $\chi_m(i) = 0$ otherwise. Obviously, we have $\sum_{i \in S_m} \mu_m(i) = 1$.

For any transition probability between clusters $\hat{P}_{ij}$, $i, j \in S$, define the prolongation transition probability on $\hat{P}_{ij}$, $i, j \in S$ as
\begin{equation}
\hat{P}_{ij} = \sum_{m,n} \chi_n(j) \mu_m(i) \hat{P}_{mn} = \sum_{m,n} \chi_n(j) \chi_m(i) \lambda_i \pi_i \hat{P}_{mn} / \hat{\lambda}_m \hat{\pi}_m.
\end{equation}
It can be directly verified that for any $j$ for some $S_n$,
\begin{equation}
\sum_i \hat{P}_{ij} = \sum_i \sum_m \mu_m(i) \hat{P}_{mn} = \sum_m (\sum_{i \in S_m} \mu_m(i)) \hat{P}_{mn} = 1.
\end{equation}

Now following the idea of [15, 26], we define the discrepancy functional between $P_{ij}$ and $\hat{P}_{ij}$ indicating the efficiency and accuracy of the coarse-graining
\begin{equation}
E := \|P - \hat{P}\|_\pi := \sum_{i,j} \frac{\lambda_i \pi_i}{\hat{\lambda}_m \hat{\pi}_m} (P_{ij} - \hat{P}_{ij})^2.
\end{equation}
By minimizing this discrepancy, we want to find the optimal partition that efficiently represents the original network. To include the case of Markov process with a jumping rate $\lambda$, the procedures below are slight modifications of the methods developed in [15, 28].

5.1.1. The optimal coarse-graining transition probability for fixed $K$-clusters. In this section, for a fixed partition $S = S_1 \cup \cdots \cup S_K$, we find the optimal coarse-graining transition probability $\hat{P}_{mn}^*$.

We describe how to do the
\begin{equation}
E(\hat{P}) = \sum_{i,j} \frac{\lambda_j \pi_j}{\lambda_i \pi_i} P_{ij}^2 - 2 \sum_{m,n} \sum_{i \in S_m, j \in S_n} \frac{\lambda_j \pi_j}{\hat{\lambda}_m \hat{\pi}_m} \hat{P}_{mn} \hat{P}_{ij} + \sum_{m,n=1}^{K} \frac{\hat{\lambda}_n \hat{\pi}_n}{\hat{\lambda}_m \hat{\pi}_m} \hat{P}_{mn}^2.
\end{equation}

for any coarse-graining transition probability $\hat{P}$.
(ii) The optimal coarse graining transition probability $\hat{P}$ is given by

$$\hat{P}_{mn} = \frac{1}{\hat{\lambda}_n \hat{\pi}_m} \sum_{i \in S_m, j \in S_n} \lambda_j \pi_j P_{ij}, \text{ for any } m, n = 1, \cdots, K$$

and satisfies $\sum_m \hat{P}_{mn}^* = 1$;  

(iii) If $P_{ij}$ satisfies the detailed balance condition (5.3) with respect to $\pi_j$, then the optimal coarse-graining transition probability $\hat{P}_{mn}^*$ also satisfies the detailed balance condition with respect to $\hat{\pi}_n$

$$\hat{P}_{mn}^* \hat{\lambda}_m \hat{\pi}_n = \hat{P}_{mn}^* \hat{\lambda}_n \hat{\pi}_n.$$  

(iv) Under the detailed balance assumption, the minimal discrepancy is

$$E(\hat{P}^*) = \sum_{i, j} \frac{\lambda_j \pi_j P_{ij}^2}{\hat{\lambda}_i \hat{\pi}_i} (P_{ij}^2 - 2 \hat{P}_{ij} P_{ij} + \hat{P}_{ij}^2)$$

Proof. For (i), from (5.6), we have

$$E(\hat{P}) = \sum_{i, j} \frac{\lambda_j \pi_j}{\hat{\lambda}_i \hat{\pi}_i} (P_{ij}^2 - 2 \hat{P}_{ij} P_{ij} + \hat{P}_{ij}^2)$$

$$= \sum_{i, j} \frac{\lambda_j \pi_j}{\hat{\lambda}_i \hat{\pi}_i} P_{ij}^2 - 2 \sum_{i, j} \frac{\lambda_j \pi_j}{\hat{\lambda}_i \hat{\pi}_i} \hat{P}_{ij} \hat{P}_{ij} + \sum_{i, j} \frac{\lambda_j \pi_j}{\hat{\lambda}_i \hat{\pi}_i} \hat{P}_{ij} \hat{P}_{ij}$$

$$= \sum_{i, j} \frac{\lambda_j \pi_j}{\hat{\lambda}_i \hat{\pi}_i} P_{ij}^2 - 2 \sum_{i, j} \frac{\lambda_j \pi_j}{\hat{\lambda}_i \hat{\pi}_i} \hat{P}_{ij} \hat{P}_{ij} + \sum_{i, j} \frac{\lambda_j \pi_j}{\hat{\lambda}_i \hat{\pi}_i} \hat{P}_{ij} \hat{P}_{ij}$$

Moreover, one can also recast $E(\hat{P})$ as

$$E(\hat{P}) = \sum_{m, n} \sum_{i \in S_m, j \in S_n} \lambda_i \pi_i \lambda_j \pi_j \left[ \frac{P_{ij}}{\lambda_i \pi_i} - \frac{\hat{P}_{mn}^*}{\hat{\lambda}_m \hat{\pi}_m} \hat{P}_{mn}^* \right]^2.$$

To prove (ii), notice the necessary condition for $\hat{P}$ to minimize $E(\hat{P})$ is $\frac{\delta E(\hat{P})}{\delta \hat{P}_{mn}^*} = 0$ for $m, n = 1, \cdots, K$. Thus we obtain

$$\hat{\lambda}_n \hat{\pi}_n \hat{P}_{mn}^* = \sum_{i \in S_m, j \in S_n} \lambda_j \pi_j P_{ij} \text{ for any } m, n = 1, \cdots, K$$

and conclude (5.10). Besides, one can verify that

$$\sum_m \hat{P}_{mn}^* = \sum_i \left( \sum_{j \in S_n} \frac{\lambda_j \pi_j}{\hat{\lambda}_n \hat{\pi}_n} \right) P_{ij} = \sum_i P_{ij} = 1.$$

To prove (iii), we start from (5.10) by exchanging index $m, n$,

$$\hat{\lambda}_m \hat{\pi}_m \hat{P}_{mn}^* = \sum_{i \in S_n, j \in S_m} \lambda_j \pi_j P_{ij} = \sum_{i \in S_m, j \in S_n} \lambda_i \pi_i P_{ji}.$$
where the second equality follows from exchanging index $i, j$. Combining (5.14) and (5.16), the original detailed balance property for $P_{ij}$ in (5.3) implies the detailed balance property for $\hat{P}_{mn}^*$ in (iii).

For (iv), since $P_{ij}$ and thus $\hat{P}_{mn}^*$ satisfy the detailed balance condition, the minimal discrepancy is

$$E(\hat{P}^*) = \sum_{i,j} \frac{\lambda_i \pi_j}{\lambda_i \pi_i} P_{ij}^2 - 2 \sum_{m,n} \hat{P}_{mn}^* \hat{P}_{mn}^* + \sum_{m,n} \frac{\hat{\lambda}_n \hat{\pi}_n}{\lambda_m \hat{\pi}_m} (\hat{P}_{mn}^*)^2$$

$$= \sum_{i,j} \frac{\lambda_i \pi_j}{\lambda_i \pi_i} P_{ij}^2 - \sum_{m,n} \frac{\hat{\lambda}_n \hat{\pi}_n}{\lambda_m \hat{\pi}_m} (\hat{P}_{mn}^*)^2$$

$$= \sum_{i,j} P_{ij} P_{ji} - \sum_{m,n} \hat{P}_{mn}^* \hat{P}_{mn}^*.$$

5.1.2. Using K-means to find the optimal partition. From Lemma 5.1 in the last section, to find the best partition which minimizes the discrepancy functional $E$ is equivalent to

$$\min_{\{S_1, \cdots, S_K\}} E(S_1, \cdots, S_K, \hat{P}^*)$$

$$= \min_{\{S_1, \cdots, S_K\}} \sum_{m,n=1}^{K} \sum_{i \in S_m, j \in S_n} \lambda_i \pi_i \lambda_j \pi_j (\frac{P_{ij}}{\lambda_i \pi_i} - \frac{\hat{P}_{mn}^*}{\hat{\lambda}_m \hat{\pi}_m})^2$$

$$= \sum_{i,j} \frac{\lambda_i \pi_j}{\lambda_i \pi_i} P_{ij}^2 - \max_{\{S_1, \cdots, S_K\}} \sum_{m,n} \frac{\hat{\lambda}_n \hat{\pi}_n}{\lambda_m \hat{\pi}_m} (\hat{P}_{mn}^*)^2.$$

Notice we do not allow empty sets in the partition $S = S_1 \cup \cdots \cup S_K$. Inspired by the last equivalent formula the discrepancy functional, to find the optimal partition using K-means, [15, 28] define the following distance function between $j$ and the cluster $S_n$ for any $j = 1, \cdots, N, n = 1, \cdots, K$

$$d(j, S_n) := \sum_{m} \sum_{i \in S_m} \lambda_i \pi_i \lambda_j \pi_j (\frac{P_{ij}}{\lambda_i \pi_i} - \frac{\hat{P}_{mn}^*}{\hat{\lambda}_m \hat{\pi}_m})^2.$$

Then standard Kmeans schemes can be applied to iteratively update the partitions; see [15, 28] for simulations about the well known Zacharys Karate Club Network.

5.2. Transition state theory on manifold. Computing the minimal energy path on a manifold $\mathcal{N}$ is important in the molecular dynamics for chemical reaction and protein folding for biomolecular. After the dimension reduction in Section 2, we learned the reaction coordinates $\mathbf{y}$ for $\mathcal{N} \subset \mathbb{R}^\ell$ with $\ell \sim O(1)$. The minimal energy path from one stable state to another stable state on $\mathcal{N}$ represents the process of the chemical reaction. It also indicates the energy barrier (a.k.a. activation energy) for this chemical reaction in different situations, for instance, the activation energy is lower for a catalyzed reaction.

More precisely, a chemical reaction from reactants $a \in \mathcal{N}$ through a transition state $c \in \mathcal{N}$ to the products $b \in \mathcal{N}$ can be described by the reaction coordinate $\mathbf{y}$, which can be obtained from manifold learning described above, and a path on the reaction coordinate $\mathbf{y}(t) \in \mathcal{N}$ with a pseudo-time $t \in [0, T]$ and $\mathbf{y}(0) = a, \mathbf{y}(T) = b$. This chemical reaction can be characterized by
an underlining potential $V(y)$ in terms of the reaction coordinate $y \in \mathcal{N}$, which has a few deep wells separated by high barriers. Assume $a$ and $b$ are two local minimums with attractor basins $A, B \subset \mathcal{N}$ such that $\max(V(a), V(b)) < V(\partial A \cup \partial B)$. The energy barrier to achieve the chemical reaction from $a$ to $b$ is $V(c) - V(a)$. The minimal energy path is a path with minimal energy barrier given by

\begin{equation}
(5.20) \quad y^*(t) = \arg \min_{y(t)} \max_{t \in [0,1]} V(y(t)),
\end{equation}

and the transition state is given by $c = \arg \max_{t \in [0,1]} V(y^*(t))$. The existence of the minimal energy path is guaranteed by the mountain pass theorem \[22\], which is the foundation of calculating the transition state $c$, and we know $c$ is the saddle point of $V$. There are some recently developed effective optimization methods in finding the minimal energy path $y^*(t)$ and transition state $c$, for instance, the string method \[18, 17\], Ekeland variational method \[19\] and some recently developed primal-dual methods, c.f. \[7\]. However, these methods can not be directly applied since we do not have exact information for the manifold $\mathcal{N}$ and potential $V(y)$. Below, we will develop an implementable method based on only on dataset $\{y_i\} \subset \mathcal{N}$ and the approximated transition probability between the nearest neighbor points obtained in Section 3.

To model this chemical reaction, we use an over-damped Langevin equation on $\mathcal{N}$ in \[1.4\]. The transition rate is given by the Kramers reaction rate formula \[16, 38\]

\begin{equation}
(5.21) \quad k = \frac{|\lambda_c|}{2\pi} \sqrt{\frac{\det \nabla_{\mathcal{N}}^2 V(a)}{\det \nabla_{\mathcal{N}}^2 V(c)}} e^{\frac{V(c) - V(a)}{kT}}, \quad \text{as } kT \to 0,
\end{equation}

where $\lambda_c$ is the unique negative eigenvalue of the Hessian at $c$. This is also known as Arrhenius’s law of reaction rates. As long as we obtain the minimal energy path, which gives the transition state $c$, then we can compute the reaction rate using the approximated formula for \[5.21\].

By the Freidlin-Wentzell theory, $\{y_t\}$ satisfies the large deviation principle with the rate function

\begin{equation}
(5.22) \quad I(y) := \frac{1}{2} \int_0^T |\dot{y}(s) + \nabla_{\mathcal{N}} V(y(s))|^2 \, ds.
\end{equation}

Then the maximal probable path in some set $D \subset \{y(\cdot) \in C[0,T]; y(0) = a, y(T) = b\}$ is defined as the optimal path $y^*(t)$ minimizing the action $I(y)$ in $D$ (also called the minimum action path). The maximal probable path $y^*(t)$ satisfies

\begin{equation}
(5.23) \quad \frac{d}{ds} y(s) = -\nabla_{\mathcal{N}} V(y(s)),
\end{equation}

and one can check the minimal energy path also satisfies \[5.23\].

In practice, we consider the continuous time Markov chain \[3.73\] on finite states $\{y_i\} \subset \mathcal{N}$. One of effective methods is to calculate the probability current of reactive trajectories and find the optimal reaction pathway with the maximal min-current is developed in \[33, 18, 16\]. Another strategy is to use the optimal the cluster-cluster transition probability obtained in Section 5.1. This idea is already successfully used in single-cell cellular state inference by incorporating molecular network features \[12\].
6. Pullback

With the trajectory $\rho^N$ on $\mathcal{N}$ obtained in Section 3, the last goal is to recover the dynamics (trajectory) of $x$ in the original real space $\mathcal{M} \subset \mathbb{R}^p$. Suppose $\{x_1, \cdots, x_n\}$ are $n$ data points sampled on $\mathcal{M}$ in a high dimensional space $\mathbb{R}^p$ based on some probability density function $\rho$ that satisfy Assumption 2.1. In Section 2, we know that we can find the reaction coordinates $\Phi$ of the data points $\{x_1, \cdots, x_n\}$ which is a smooth embedding of $\mathcal{M}$ into $\mathbb{R}^\ell$. We also know that the diffusion map is an approximation of the map $\Phi$. Suppose $\mathcal{N} = \Phi(\mathcal{M})$ and $y_i = \Phi(x_i)$ are the reaction coordinates of $x_i$. By using the reaction coordinates, in Sections 3 and 4, we are able to find the probability density functions $\rho^\mathcal{N}$ on $\mathcal{N}$. In this section we propose an algorithm to find the pullback density function $\rho^\mathcal{M}$ on $\mathcal{M}$ which is defined as follows.

Suppose that $\mathcal{M}$ is a $d$ dimensional smooth closed Riemannian submanifold of $\mathbb{R}^p$. $\Phi : \mathcal{M} \hookrightarrow \mathbb{R}^\ell$ is a smooth embedding. Moreover, suppose that $\mathcal{N} = \Phi(\mathcal{M})$, then $\mathcal{N} \subset \mathbb{R}^\ell$ is another $d$ dimensional smooth closed Riemannian manifold. Let $X$ be the random variable defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with the range $\mathcal{M}$, where $\mathbb{P}$ is the probability measure defined on the sigma algebra $\mathcal{F}$ in the event space $\Omega$. Define $\mathcal{P} = X_\ast \mathbb{P}$ to be the induced measure defined on the Borel sigma algebra on $\mathcal{M}$. By the Radon-Nikodym theorem, for any $x \in \mathcal{M}$, $\mathcal{P}(x) = \rho^{\mathcal{M}}(x)dV_M(x)$, where $\rho^\mathcal{M}$ is the probability density function of $X$ associated with the Riemannian volume measure, $dV_M$, defined on $\mathcal{M}$. Furthermore, define $\mathcal{Q}$ to be the induced measure by $\Phi$ defined on the Borel sigma algebra on $\mathcal{N}$; that is, $\mathcal{Q} = \Phi_\ast \mathcal{P} = (\Phi \circ X)_\ast \mathbb{P}$. By Radon-Nikodym chain rule, for $y = \Phi(x)$,

\begin{equation}
\rho^\mathcal{N}(y) = \frac{d\mathcal{Q}(y)}{dV_{\mathcal{N}}(y)} = \frac{d\mathcal{P}(\Phi^{-1}(y))}{dV_M(\Phi^{-1}(y))} \frac{dV_M(\Phi^{-1}(y))}{dV_{\mathcal{N}}(y)} = \frac{\rho^\mathcal{M}(x)}{|\nabla \Phi(x)|}
\end{equation}

is the induced probability density function of the random variable $Y := \Phi \circ X$ associated with the Riemannian volume measure defined on $\mathcal{N}$. Here, $\nabla \Phi \in \mathbb{R}^{d \times d}$ is the Jacobian of $\Phi$ calculated in the charts of $\mathcal{M}$ and $\mathcal{N}$. In contrast, we call $\rho^\mathcal{M}$ to be the pullback probability density function of $\rho^\mathcal{N}$ on $\mathcal{M}$ through $\Phi$.

We introduce the normalized local covariance matrix for a point $y \in \mathcal{N} \subset \mathbb{R}^\ell$ below. For more discussion about the normalized local covariance matrix, the readers may refer to [32].

**Definition 6.1.** For a measurable set $\mathcal{O} \subset \mathcal{N} \subset \mathbb{R}^\ell$ and $y \in \mathcal{O}$, the local covariance matrix at $y$ associated with $\mathcal{O}$ is defined as

\begin{equation}
C_{y, \mathcal{O}} := \mathbb{E}[(Y - y)(Y - y)^\top \chi_\mathcal{O}(Y)] = \int_\mathcal{O} (y' - y)(y' - y)^\top \rho^\mathcal{N}(y)dV_{\mathcal{N}}(y') \in \mathbb{R}^{\ell \times \ell}.
\end{equation}

For $x \in \mathcal{M}$, $y = \Phi(x) \in \mathcal{N}$ and $\delta > 0$, we construct a subset of $\mathcal{N}$:

\begin{equation}
E_\delta(y) = \Phi(B_\delta^{\mathbb{R}^p}(x) \cap \mathcal{M}).
\end{equation}

We define the normalized local covariance matrix as

\begin{equation}
\bar{C}_{y, E_\delta(y)} := \frac{C_{y, E_\delta(y)}}{\delta^d \mathbb{E}[\chi_{E_\delta(y)}(Y)]}
\end{equation}

The normalized local covariance matrix $\bar{C}_{y, E_\delta(y)}$ is related to $|\nabla \Phi|$ by the following theorem.
Theorem 6.2. Suppose δ is small enough. Let μ₁(y) ≥ μ₂(y) ≥ · · · ≥ μ₄(y) be the eigenvalues of \( \tilde{C}_{Y,E₃}(y) \). Then μ₁(y), · · · , μ₄(y) are of order 1, and the rest of the eigenvalues are of order \( O(δ²) \). Moreover,

(6.5) \[ |∇Φ(x)|² = (d + 2)μ₁(y) · · · μ₄(y) + O(δ²). \]

Proof. The proof is a modification of Lemma 3 and Lemma 5 in [32]. We point out the main differences between this work and [32]. First, there is a difference in the notations. In this work, \( Φ \) is a map from \( M \) to \( ℝ^ℓ \), while in [32], \( Φ \) is a map from \( M \) to \( N \) and \( N \) is embedded in \( ℝ^ℓ \) by a map called \( ℱ \). Second, in [32], the normalized local covariance matrix is defined over \( E₃(y) = Φ(B₃(x) ∩ M) \), where \( B₃(x) \) is a geodesic ball in \( M \) ((39) in [32]). In this work, the normalized local covariance matrix is defined over \( E₃(y) = Φ(B₃(x) ∩ M) \) and \( B₃(x) \) is of order of \( δ^{d+1} \), which is a higher order term. Therefore, the conclusion of Lemma 3 and Lemma 5 in [32] remains the same. In other words, there is matrix \( U ∈ O(ℓ) \), such that

(6.6) \[ U\tilde{C}_{Y,E₃}(y)Uᵀ = \left[ \begin{array}{cc} \frac{1}{d+2}∇Φ(x)∇Φ(x)ᵀ & 0 \\ 0 & 0 \end{array} \right] + O(δ²). \]

Since \( \tilde{C}_{Y,E₃}(y) \) is a symmetric matrix, the conclusion of this theorem follows from applying the standard perturbation theory to the right hand side of the above equation. □

Combining the above theorem and (6.1), we have

(6.7) \[ ρ^M(x) = ρ^N(y)\sqrt{(d + 2)μ₁(y) · · · μ₄(y) + O(δ²)}. \]

Hence, we propose the following algorithm.

Algorithm 3: PULLBACK ALGORITHM

Parameters: Algorithm inputs are the bandwidth δ, \( \{x_1, · · · , x_n\} \) and their reaction coordinates \( \{y_1, · · · , y_n\} \).

1 Fix \( x_k ∈ \{x_1, · · · , x_n\} \) and δ > 0, find the points \( \{x_k₁, · · · , x_kN_k\} ⊂ \{x_1, · · · , x_n\} \) such that \( x_ki ∈ B_{δ}^{RP}(x_k) \).

2 Suppose \( \{y_k₁, · · · , y_kN_k\} \) are the corresponding reaction coordinates of \( \{x_k₁, · · · , x_kN_k\} \).

Construct the matrix

(6.8) \[ \tilde{C}_{n,k} = \frac{1}{δ²N_k} \sum_{i=1}^{N_k} (y_{k,i} - y_k)(y_{k,i} - y_k)ᵀ ∈ ℝ^{ℓ×ℓ}. \]

3 Suppose \( μ₁,n,k, · · · , μ₄,n,k \) are the d largest eigenvalues of \( \tilde{C}_{n,k} \). Let

(6.9) \[ J_k = \sqrt{(d + 2)μ₁,n,k · · · μ₄,n,k}. \]

Then \( ρ^N(y)J_k \) is an approximation of \( ρ^M(x_k) \).

We know that \( \frac{1}{n} \sum_{i=1}^{N_k} (y_{k,i} - y_k)(y_{k,i} - y_k)ᵀ \) is an approximation to \( C_{Y,E₃}(y) \) and \( \frac{N_k}{n} \) is an approximation of \( E[χ_{E₃}(y)Y] \). Therefore, the entrywise variance analysis between the \( \tilde{C}_{n,k} \) and \( \tilde{C}_{Y,E₃}(y) \) can be constructed by developing the method of Lemma E.1. and Lemma E.4. in [44]. The error between \( J_k \) and \( |∇Φ(x)| \) can be bounded in the probabilistic sense if we apply the perturbation theory on the variance analysis as Lemma E.4. in [44].
7. Discussion

We focus on the analysis of the dynamics of a physical system with a manifold structure. The underlying manifold structure of the system is reflected through a point cloud in a high dimensional space. By applying the diffusion map, we are able to find the reaction coordinates so that those data points are reduced onto a manifold in a low dimensional space. Based on the reaction coordinates, we propose an implementable, unconditionally stable, upwind scheme for a Fokker-Planck equation which incorporates the structure of the manifold in the low dimensional space. Since the equilibrium potential on the manifold is crucial in establishing the upwind scheme, a Gaussian process regression which respects the geometry of the manifold is applied to recover the equilibrium potential. At last, the trajectory of the density corresponding to the solution of the Fokker-Planck equation is reconstructed in the original high-dimensional space through the pullback algorithm.

Motivated by the upwind scheme, we study the Markov process associated with the Fokker-Planck equation. We also provide the weighted $L^2$ convergence analysis of the upwind scheme to the Fokker-Planck equation. The efficiency and the accuracy of the data-driven approaches proposed in this paper are justified theoretically. However, there are still many interesting directions for future work. An important direction is the manifold-related applications such as the optimal network partitions. Another challenging task is to explore the transition path in chemical reactions, especially on the high dimensional practical dataset.

Acknowledgement

Nan Wu thanks the valuable discussion with Professor Hau-Tieng Wu and Chao Shen. Jian-Guo Liu was supported in part by the National Science Foundation (NSF) under award DMS-1812573.

References

[1] https://commons.wikimedia.org/wiki/file:pangaea_continents.png.
[2] Christopher Amante and Barry W Eakins. ETOPO1 1 arc-minute global relief model: procedures, data sources and analysis. US Department of Commerce, National Oceanic and Atmospheric Administration, National Environmental Satellite, Data, and Information Service, National Geophysical Data Center, Marine Geology and Geophysics Division, 2009.
[3] Jonathan Bates. The embedding dimension of laplacian eigenfunction maps. Applied and Computational Harmonic Analysis, 37(3):516–530, 2014.
[4] John K Beem. Pseudo-riemannian manifolds with totally geodesic bisectors. Proceedings of the American Mathematical Society, 49(1):212–215, 1975.
[5] M. Belkin and P. Niyogi. Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. Neural. Comput., 15(6):1373–1396, 2003.
[6] Pierre Bérard, Gérard Besson, and Sylvain Gallot. Embedding riemannian manifolds by their heat kernel. Geometric & Functional Analysis GAF, 4(4):373–398, 1994.
[7] Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. Journal of Mathematical Imaging and Vision, 40(1):120-145, May 2011.
[8] Shui-Nee Chow, Wen Huang, Yao Li, and Haomin Zhou. Fokkerplanck equations for a free energy functional or markov process on a graph. Archive for Rational Mechanics and Analysis, 203(3):969-1008, Mar 2012.
[9] R. R. Coifman, I. G. Kevrekidis, S. Lafon, M. Maggioni, and B. Nadler. Diffusion maps, reduction coordinates, and low dimensional representation of stochastic systems. Multiscale Modeling & Simulation, 7(2):842–864, 2008.
[10] Ronald R Coifman and Stéphane Lafon. Diffusion maps. \textit{Applied and computational harmonic analysis}, 21(1):5–30, 2006.

[11] Peter Deuflhard and Marcus Weber. Robust perron cluster analysis in conformation dynamics. \textit{Linear Algebra and its Applications}, 398:161184, Mar 2005.

[12] Ji Dong, Peijie Zhou, Yichong Wu, Wendong Wang, Yidong Chen, Xin Zhou, Haoling Xie, Yuan Gao, Jiansen Lu, Jingwei Yang, et al. Enhancing single-cell cellular state inference by incorporating molecular network features. \textit{bioRxiv}, page 699959, 2019.

[13] D. L. Donoho and C. Grimes. Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. 100(10):5591–5596, 2003.

[14] David B Dunson, Hau-Tieng Wu, and Nan Wu. Diffusion based gaussian process regression via heat kernel reconstruction. \textit{arXiv preprint arXiv:1912.05680}, 2019.

[15] W. E, T. Li, and E. Vanden-Eijnden. Optimal partition and effective dynamics of complex networks. \textit{Proceedings of the National Academy of Sciences}, 105(23):79077912, Jun 2008.

[16] Weinan E, Tiejun Li, and Eric Vanden-Eijnden. \textit{Applied stochastic analysis}. Graduate studies in mathematics. American Mathematical Society, 2019.

[17] Weinan E, Weiqing Ren, and Eric Vanden-Eijnden. String method for the study of rare events. \textit{Physical Review B}, 66(5):052301, Aug 2002.

[18] Weinan E and Eric Vanden-Eijnden. Transition-path theory and path-finding algorithms for the study of rare events. \textit{Annual Review of Physical Chemistry}, 61(1):391420, Mar 2010.

[19] Ivar Ekeland and Roger Temam. \textit{Convex Analysis and Variational Problems}, volume 28. SIAM, 1999.

[20] Matthias Erbar. Gradient flows of the entropy for jump processes. \textit{Annales de l'Institut Henri Poincar, Probabilités et Statistiques}, 50(3):920945, Aug 2014.

[21] Antonio Esposito, Francesco S Patacchini, Andre Schlichting, and Dejan Slepcev Ev. Nonlocal-interaction equation on graphs: Gradient flow structure and continuum limit. page 45.

[22] Lawrence C Evans. \textit{Partial differential equations}, volume 19. American Mathematical Soc., 2010.

[23] Yuan Gao and Jian-Guo Liu. A note on parametric bayesian inference via gradient flows. \textit{to appear in Annals of Mathematical Sciences and Applications}, 2020.

[24] Elton P Hsu. \textit{Stochastic analysis on manifolds}, volume 38. American Mathematical Soc., 2002.

[25] Peter W Jones, Mauro Maggioni, and Raanan Schul. Manifold parametrizations by eigenfunctions of the laplacian and heat kernels. \textit{Proceedings of the National Academy of Sciences}, 105(6):1803–1808, 2008.

[26] Stephane Lafon and Ann B Lee. Diffusion maps and coarse-graining: A unified framework for dimensionality reduction, graph partitioning, and data set parameterization. \textit{IEEE transactions on pattern analysis and machine intelligence}, 28(9):1303–1403, 2006.

[27] Lei Li and Jian-Guo Liu. Large time behaviors of upwind schemes by jump processes. \textit{to appear in Math. Comp.}, 2020.

[28] Tiejun Li, Jian Liu, and Weinan E. Probabilistic framework for network partition. \textit{Physical Review E}, 80(2):026106, Aug 2009.

[29] Tiejun Li, Jian Liu, and E Weinan. Probabilistic framework for network partition. \textit{Physical Review E}, 80(2):026106, 2009.

[30] Anning Liu, Jian-Guo Liu, and Yulong Lu. On the rate of convergence of empirical measure in $\infty$-wasserstein distance for unbounded density function. \textit{Quarterly of Applied Mathematics}, 77(4):811–829, 2019.

[31] Jan Maas. Gradient flows of the entropy for finite markov chains. \textit{Journal of Functional Analysis}, 261(8):22502292, Oct 2011.

[32] John Malik, Chao Shen, Hau-Tieng Wu, and Nan Wu. Connecting dots: from local covariance to empirical intrinsic geometry and locally linear embedding. \textit{Pure and Applied Analysis}, 1(4):515–542, 2019.

[33] Philipp Metzner, Christof Schtte, and Eric Vanden-Eijnden. Transition path theory for markov jump processes. \textit{Multiscale Modeling \& Simulation}, 7(3):11921219, Jan 2009.
Appendix A. Theorems about embedding by eigenfunctions of Laplacian

Let $\Delta$ be the Laplace-Beltrami operator of a closed smooth Riemannian manifold $\mathcal{M}$. Let $\{\lambda_i\}_{i=0}^{\infty}$ be the eigenvalues of $-\Delta$, and

$$(A.1) \quad \Delta \psi_i = -\lambda_i \psi_i,$$

where $\psi_i$ is the corresponding eigenfunction normalized in $L^2(\mathcal{M})$. We have $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \cdots$.

In this section, we review the theorems about embedding the manifold $\mathcal{M}$ by using the eigenfunctions of $\Delta$. In [6], the authors provide a theorem about spectral embedding by using all the eigenvalues and eigenfunctions of $\Delta$ into the Hilbert space $\ell^2$.

**Theorem A.1.** (Bérard-Besson-Gallot, [6]) Let $M$ be a $d$ dimensional smooth closed Riemannian manifold with $\text{Ric}(\mathcal{M}) \geq (d - 1)k$ and $\text{diam}(\mathcal{M}) \leq D$. Then, for $x \in \mathcal{M}$

$$(A.2) \quad \Psi(x) = (2t)^{\frac{d+2}{4}} \sqrt{2} (4\pi)^{\frac{d}{4}} (e^{-\lambda_1 t} \psi_1(x), \cdots, e^{-\lambda_q t} \psi_q(x), \cdots),$$

is an embedding of $M$ into $\ell^2$ for all $t > 0$.

[25] improves the above result locally. They show that one can use finite eigenfunctions of Laplace-Beltrami operator to embed the manifold locally. The result can be briefly summarized as follows.
Theorem A.2. (Jones-Maggioni-Schul, [25]) Let $\mathcal{M}$ be a $d$ dimensional smooth closed Riemannian manifold, for each $x \in M$, there are $j_1 \leq \cdots \leq j_d$ and the constants $C_1, \cdots, C_d$ such that
\begin{equation}
\Psi(x) = (C_1 \psi_{j_1}(x), \cdots, C_d \psi_{j_d}(x)),
\end{equation}
is locally a bi-Lipschitz chart.

[3] provides a global result about embedding a manifold into some Euclidean space via the eigenfunctions of the Laplace-Beltrami operator.

Theorem A.3. (Bates, [3]) There is a $q$ such that for $x \in \mathcal{M}$
\begin{equation}
\Psi(x) = (\psi_1(x), \cdots, \psi_q(x)),
\end{equation}
is a smooth embedding of $\mathcal{M}$ into $\mathbb{R}^q$.

Moreover, the next theorem [35] says that we can use the eigenvalues and eigenfunctions of the Laplace-Beltrami operator to construct an almost isometric embedding of the manifold into some Euclidean space.

Theorem A.4. (Portegies, [35]) For any $\epsilon > 0$, there is a $t_0(\epsilon)$ and $K(\epsilon)$, such that if $t < t_0$ and $q > K$, then for $x \in M$
\begin{equation}
\Psi(x) = (2t)^{d+2} \sqrt{2(4\pi)^d} (e^{-\lambda_1 t} \psi_1(x), \cdots, e^{-\lambda_q t} \psi_q(x)),
\end{equation}
is an embedding of $\mathcal{M}$ into $\mathbb{R}^q$ such that $1 - \epsilon < \|\nabla \Psi\|_{op} < 1 + \epsilon$. Here $\|\cdot\|_{op}$ is the operator norm.

Based on the above theorems, the smallest $q$ that
\begin{equation}
\Psi_1(x) = (\psi_1(x), \cdots, \psi_q(x)),
\end{equation}
is a smooth embedding of $\mathcal{M}$ is called the embedding dimension of $\mathcal{M}$. The smallest $q$ that
\begin{equation}
\Psi_2(x) = (2t)^{d+2} \sqrt{2(4\pi)^d} (e^{-\lambda_1 t} \psi_1(x), \cdots, e^{-\lambda_q t} \psi_q(x)),
\end{equation}
is an almost isometric embedding of $\mathcal{M}$ is called the almost isometric embedding dimension of $\mathcal{M}$. We expect the embedding dimension is much smaller than the almost isometric embedding dimension. Hence, for the dimension reduction purpose, we are looking for an embedding of the manifold rather than an almost isometric embedding.

**Appendix B. Proof of Proposition 3.12 and Proposition 3.13**

We start from a study of the matrix $C_{n,r}(y_k)$ in Definition 3.11 and relate it to its continuous form. Consider the local covariance matrix $C_{y_k, B_{\sqrt{\tau}}(y_k) \cap \mathcal{N}}$ defined as in Definition 6.1. Suppose $C_{y_k, B_{\sqrt{\tau}}(y_k) \cap \mathcal{N}}$ has the following eigendecomposition:
\begin{equation}
C_{y_k, B_{\sqrt{\tau}}(y_k) \cap \mathcal{N}} = U(y_k) \Lambda(y_k) U(y_k)^{\top} \in O(\ell),
\end{equation}
where $\Lambda(y_k)$ is a diagonal matrix with the diagonal entries to be eigenvalues of $C_{y_k, B_{\sqrt{\tau}}(y_k) \cap \mathcal{N}}$. Moreover, we have $\Lambda_{11}(y_k) \geq \Lambda_{22}(y_k) \geq \cdots \geq \Lambda_{\ell \ell}(y_k)$. $U(y_k) \in O(\ell)$ consists of the corresponding orthonormal eigenvectors of $C_{y_k, B_{\sqrt{\tau}}(y_k) \cap \mathcal{N}}$. Intuitively, $C_{y_k, B_{\sqrt{\tau}}(y_k) \cap \mathcal{N}}$ is the continuous form of the matrix $C_{n,r}(y_k)$. 
By setting $\epsilon = \sqrt{\tau}$ in Proposition 3.2 in [44], we have the following lemma.

**Lemma B.1.** Assume that $T_{y_k}N$ is generated by the first $d$ standard basis of $\mathbb{R}^\ell$.

\begin{equation}
\Lambda(y_k) = \frac{|S^{d-1}|P(y_k)r^{d+2}}{d(d+2)} \left[ I_{d\times d} 0 \right] + O(r), \tag{B.2} \end{equation}

\begin{equation}
U(y_k) = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} + O(r), \tag{B.3} \end{equation}

where $X_1 \in O(d)$ and $X_2 \in O(\ell - d)$.

Above lemma says that the first $d$ eigenvectors of $C_{y_k,B^{1\ell}_\tau(y_k)\cap N}$ form an orthonormal basis of $T_{y_k}N$ up to an error of order $O(r)$. Note that, for simplicity, we assume $T_{y_k}N$ is generated by the first $d$ standard basis of $\mathbb{R}^\ell$ so that $U(y_k)$ can be expressed in the above block form. Suppose $C_{n,r}(y_k)$ has the following eigendecomposition:

\begin{equation}
C_{n,r}(y_k) = U_n(y_k)\Lambda_n(y_k)U_n(y_k)^T. \tag{B.4} \end{equation}

$\Lambda_n(y_k)$ is a diagonal matrix with the diagonal entries to be eigenvalues of $C_{n,r}(y_k)$. Moreover, we have $\Lambda_{n,11}(y_k) \geq \Lambda_{n,22}(y_k) \geq \cdots \geq \Lambda_{n,\ell\ell}(y_k)$. $U_n(y_k) \in O(\ell)$ consists of the corresponding orthonormal eigenvectors of $C_{n,r}(y_k)$.

The relation between the eigenstructure of $C_{y_k,B^{1\ell}_\tau(y_k)\cap N}$ and $C_{n,r}(y_k)$ is discussed in Lemma E.4 in [44].

**Lemma B.2.** Assume that $T_{y_k}N$ is generated by the first $d$ standard basis of $\mathbb{R}^\ell$. When $n$ is large enough, with probability greater than $1 - \frac{1}{n^2}$, for all $y_k$,

\begin{equation}
\Lambda_n(y_k) = \Lambda(y_k) + O(\sqrt{\frac{\log n}{nr^{-\frac{d}{2}-2}}}), \tag{B.5} \end{equation}

\begin{equation}
U_n(y_k) = \begin{bmatrix} X_1' & 0 \\ 0 & X_2' \end{bmatrix} U(y_k) + O(\sqrt{\frac{\log n}{nr^{-\frac{d}{2}-2}}}), \tag{B.6} \end{equation}

where $X_1' \in O(d)$ and $X_2' \in O(\ell - d)$.

**Remark B.3.** Above lemma follows from Lemma E.4 in [44] if we choose $\epsilon = \sqrt{\tau}$ and $\rho \to \infty$ in Case 0 of Lemma E.4 in [44]. In fact, Case 0 of Lemma E.4 in [44] focuses on the first $d$ eigenpairs of the matrix $C_{y_k,B^{0\ell}_\tau(y_k)\cap N}$ of which we need to recover.

If we combine Lemma B.1 and Lemma B.2, we have

\begin{equation}
\Lambda_n(y_k) = \frac{|S^{d-1}|P(y_k)r^{d+2}}{d(d+2)} \left[ I_{d\times d} 0 \right] + O(r^{\frac{d}{2}+2}) + O(\sqrt{\frac{\log n}{nr^{-\frac{d}{2}-2}}}), \tag{B.7} \end{equation}

\begin{equation}
U_n(y_k) = \begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix} + O(r) + O(\sqrt{\frac{\log n}{nr^{-\frac{d}{2}-2}}}), \tag{B.8} \end{equation}

where $U_1 \in O(d)$ and $U_2 \in O(\ell - d)$. If $\frac{nr^{\frac{d}{2}}}{\log n} \to \infty$ as $n \to \infty$, then $\sqrt{\frac{\log n}{nr^{-\frac{d}{2}-2}}} \leq r$. If $\frac{nr^{\frac{d}{2}+2}}{\log n} \to \infty$ as $n \to \infty$, then $\sqrt{\frac{\log n}{nr^{-\frac{d}{2}-2}}} \leq r$ and $\sqrt{\frac{\log n}{nr^{-\frac{d}{2}-2}}} \leq r^{\frac{d}{2}+2}$. Hence, we have the following proposition.
Proposition B.4. Assume that $T_{Y_k}N$ is generated by the first $d$ standard basis of $\mathbb{R}^\ell$. If $\frac{nr^2}{\log n} \to \infty$ as $n \to \infty$, then with probability greater than $1 - \frac{1}{n^2}$, for all $y_k$,

$$U_n(y_k) = \begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix} + O(r),$$

(B.9)

where $U_1 \in O(d)$ and $U_2 \in O(\ell - d)$.

If $\frac{nr^2 + 2}{\log n} \to \infty$ as $n \to \infty$, then with probability greater than $1 - \frac{1}{n^2}$, for all $y_k$,

$$\Lambda_n(y_k) = \frac{|S^{d-1}|P(y_k)r^{d+2}}{d(d+2)} \begin{bmatrix} I_{d \times d} & 0 \\ 0 & 0 \end{bmatrix} + O(r^{d+2}),$$

(B.10)

$$U_n(y_k) = \begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix} + O(r),$$

(B.11)

where $U_1 \in O(d)$ and $U_2 \in O(\ell - d)$.

Above proposition should be understood in the following way. If $n$ and $r$ satisfy $\frac{nr^2}{\log n} \to \infty$ as $n \to \infty$, then we have an approximation of the tangent space of $N$ at $y_k$, i.e. the first $d$ eigenvectors of $C_{n,r}(y_k)$ are the basis of $T_{y_k}N$ up to an error of order $O(r)$. If $n$ and $r$ satisfy $\frac{nr^2 + 2}{\log n} \to \infty$ as $n \to \infty$, the first $d$ eigenvectors of $C_{n,r}(y_k)$ are the basis of $T_{y_k}N$ up to an error of order $O(r)$. Moreover, there are $d$ significantly large eigenvalues of $C_{n,r}(y_k)$ which are close to the first $d$ eigenvalues of $C_{y_k, B_{\sqrt r} (y_k)} \cap N$.

Next, we show that the map $\tilde{\iota}_k$ in the definition 3.11 restricted on $B_{\sqrt r} (y_k) \cap N$ is a $1 + O(r)$ bi-Lipschitz homeomorphism.

Lemma B.5. Choose $r \to 0$ and $\frac{nr^2}{\log n} \to \infty$ as $n \to \infty$. Suppose $r$ is small enough, then with probability greater than $1 - \frac{1}{n^2}$, for all $y_k$ and any $y, y' \in B_{\sqrt r} (y_k) \cap N$, we have

$$\|\tilde{\iota}_k(y') - \tilde{\iota}_k(y)\|_{\mathbb{R}^d} = \|\iota_k(y' - y)\|_{\mathbb{R}^d} = d_N(y, y')(1 + O(r)).$$

(B.12)

Proof. $\|\tilde{\iota}_k(y') - \tilde{\iota}_k(y)\|_{\mathbb{R}^d} = \|\iota_k(y' - y)\|_{\mathbb{R}^d}$ follows from the definition. Next, we prove $\|\iota_k(y' - y)\|_{\mathbb{R}^d} = d_N(y, y')(1 + O(r))$. For simplicity, we assume $y_k = 0$ and $T_{y_k}N$ is generated by the first $d$ standard basis of $\mathbb{R}^\ell$. For any $y \in \mathbb{R}^\ell$, we use the following notation to simplify the proof:

$$y = [v, v^\perp] \in \mathbb{R}^\ell,$$

(B.13)

where $v \in T_{y_k}N$ forms the first $d$ coordinates of $y$ and $v^\perp \in T_{y_k}^\perp N$ forms the last $\ell - d$ coordinates of $y$. For any $y, y' \in B_{\sqrt r} (y_k) \cap N$, suppose $y = [v_1, v_1^\perp]$ and $y' = [v_2, v_2^\perp]$. Due to the manifold structure of $N$, we have

$$\|v_1^\perp - v_2^\perp\|_{\mathbb{R}^{\ell - d}} \leq C_1 r \|v_1 - v_2\|_{\mathbb{R}^d},$$

for some constant $C_1$ depending on the curvature of $N$. Hence,

$$\|v_1 - v_2\|_{\mathbb{R}^d} \leq \|y' - y\|_{\mathbb{R}^\ell} \leq \|v_1 - v_2\|_{\mathbb{R}^d} \sqrt{1 + C_1^2 r^2},$$

(B.15)

which is equivalent to

$$\|v_1 - v_2\|_{\mathbb{R}^d} = \|y' - y\|_{\mathbb{R}^\ell}(1 + O(r^2)).$$

(B.16)
Moreover, suppose \( \{\beta_{n,r,1}, \cdots, \beta_{n,r,d}\} \) are orthonormal eigenvectors corresponding to \( C_{n,r}(y_k) \)'s largest \( d \) eigenvalues. Then, by Proposition \ref{prop:orthonormal}
\begin{equation}
\beta_{n,r,1} = [\beta_i, 0] + O(r),
\end{equation}
where \( \{\beta_i\}_{i=1}^d \) form an orthonormal basis of \( T_{y_k} \mathcal{N} \approx \mathbb{R}^d \).
\begin{equation}
\| t_k(y' - y) \|_{\mathbb{R}^d} = \| v_1 - v_2 \|_{\mathbb{R}^d} + \| y' - y \|_{\mathbb{R}^d} O(r) = \| y' - y \|_{\mathbb{R}^d} (1 + O(r^2)) + \| y' - y \|_{\mathbb{R}^d} O(r) = \| y' - y \|_{\mathbb{R}^d} (1 + O(r)),
\end{equation}
where we apply \ref{eq:inner_product} in the second last step.

By equation \ref{eq:inner_product}, we know that \( d_N(y, y') \leq 2D_1r \). Hence, by Lemma \ref{lemma:hausdorff_distance}
\begin{equation}
\| t_k(y' - y) \|_{\mathbb{R}^d} = \| y' - y \|_{\mathbb{R}^d} (1 + O(r^2)) = d_N(y, y')(1 + O(d_N^2(y, y')))(1 + O(r))
= d_N(y, y')(1 + O(r^2))(1 + O(r)) = d_N(y, y')(1 + O(r)).
\end{equation}

We introduce the following notations to prove the following lemma and proposition. Denote the boundary of \( C_k \) by \( \partial C_k \). Denote the boundary of \( i_k(C_k) \) by \( \partial i_k(C_k) = i_k(\partial C_k) \). Let \( \tilde{C}_{k,0} \) be the Voronoi cell in \( \mathbb{R}^d \) containing \( 0 \) constructed in the Step 4 in Algorithm \ref{alg:voronoi}. Denote the boundary of \( \tilde{C}_{k,0} \) by \( \partial \tilde{C}_{k} \). Denote \( d_{\mathbb{R}^d}^n(S_1, S_2) \) be the Hausdorff distance between two sets \( S_1 \) and \( S_2 \) in \( \mathbb{R}^d \) with respect to the Euclidean metric.

**Lemma B.6.** Under Assumption \ref{assumption:convex}, if we choose \( r \to 0 \), \( nr^d \) goes to some constant and \( \frac{nr^2}{\log n} \to \infty \) as \( n \to \infty \), then with probability greater than \( 1 - \frac{1}{n^2} \), for all \( y_k \), \( d_{\mathbb{R}^d}^n(\partial i_k(C_k), \partial \tilde{C}_k) = O(r^2) \).

**Proof.** We are going to use Lemma \ref{lemma:boundary} hence we require \( r \to 0 \) and \( \frac{nr^2}{\log n} \to \infty \) as \( n \to \infty \). We expect the number of points in \( B^\ell_r(y_k) \), hence we require \( nr^d \) goes to some constant. For simplicity, in this proof, we use \( | \cdot | \) to denote \( \| \cdot \|_{\mathbb{R}^d} \).

Recall that in Assumption \ref{assumption:convex}, we assume \( B^\ell_r(y_k) \cap \{ y_i \}_{i=1}^n = \{ y_{k,1}, \cdots, y_{k,N_k} \} \). We have \( C_k \subset B^\ell_r(y_k) \). Moreover, if \( \Gamma_{k,j} \) is a Voronoi surface of \( C_k \) between \( y_k \) and \( y_j \), then \( y_j \in B^\ell_r(y_k) \). We denote \( \Gamma_{k,i} \) to be the Voronoi face between \( y_k \) and \( y_{k,i} \).

The proof has two steps, first we show that for any \( v \in \partial i_k(C_k) \), \( d_{\mathbb{R}^d}(v, \partial \tilde{C}_k) = O(r^2) \). We need to consider two cases in this step.

**Case 1:** \( v \in \tilde{C}_{k,0} \)
Suppose \( v = i_k(y) \) for some \( y \in \partial C_k \). Moreover \( y \in \Gamma_{k,i} \). In other word, \( y \) is on the Voronoi face between \( y_k \) and \( y_{k,i} \in B^\ell_r(y_k) \). As shown in figure \ref{fig:voronoi}(a), let \( O \) be the origin in \( \mathbb{R}^d \). Let \( A = v \) and \( B = i_k(y_{k,i}) \). Let \( H \) be the hyperplane in \( \mathbb{R}^d \) which perpendicularly bisects \( OB \). \( M \) is the intersection of \( H \) and \( OB \). Let \( C \) be the point on \( OB \) so that \( AC \) is perpendicular to \( OB \). Since we assume \( A \in \tilde{C}_{k,0}, C \in OM \). We have
\begin{equation}
\|C - M\| = \frac{|CB| - |CO|}{2} = \frac{|CB|^2 - |CO|^2}{2(|CB| + |CO|)}
= \frac{|AB|^2 - |AC|^2 - (|AO|^2 - |AC|^2)}{2|BO|} = \frac{(|AB| + |AO|)(|AB| - |AO|)}{2|BO|}
\end{equation}

By Lemma B.5, (B.25) $|AB| = a(1 + O(r))$ and $|AO| = a(1 + O(r))$, hence $|AB| + |AO|(|AB| - |AO|) = a^2O(r) \leq 2D_1O(r^3)$. By Lemma B.5, (B.26) $|BO| = d_N(y_k, y_k,i)(1 + O(r))$. By (2) in Assumption 3.9, (B.23) $d_N(y_k, y_k,i) \geq D_2r$. Hence,

\[
|AB| + |AO|(|AB| - |AO|) \leq \frac{D_1^2O(r^2)}{D_2^2} = O(r^2).
\]

Since $\tilde{C}_{k,0}$ is convex, $d_{\mathbb{R}^d}(A, \partial \tilde{C}_k) \leq d_{\mathbb{R}^d}(A, H)$. The conclusion follows. Note that if $A \notin \tilde{C}_{k,0}$, we still have $d_{\mathbb{R}^d}(A, H) = O(r^2)$. However, it is not true that $d_{\mathbb{R}^d}(A, \partial \tilde{C}) \leq d_{\mathbb{R}^d}(A, H)$.

**Case 2:** $v \notin \tilde{C}_{k,0}$

Suppose $v = \tilde{t}_k(y)$ for some $y \in \partial C_k$. As shown in figure 13(b), let $O$ be the origin in $\mathbb{R}^d$. Let $A = v$. Suppose $OA$ intersects with $\partial \tilde{C}_k$ at $D$. $D \in F_{k,j}$, where $F_{k,j}$ is Voronoi face in $\mathbb{R}^d$ between $O$ and $B = i_k(y_k)$. $H$ is the hyperplane that perpendicularly bisects $OB$. $M$ is the intersection between $H$ and $OB$. Note that $F_{k,j} \subset H$. Let $C$ be the point on $OB$ so that $AC$ is perpendicular to $OB$. Since we assume $A \notin \tilde{C}_{k,0}$, $C \in BM$. We have

\[
|CM| = \frac{|CO| - |CB|}{2} = \frac{|CO|^2 - |CB|^2}{2(|CB| + |CO|)}
\]

\[
= \frac{|AO|^2 - |AC|^2 - (|AB|^2 - |AC|^2)}{2|BO|} = \frac{(|AB| + |AO|)(|AO| - |AB|)}{2|BO|}.
\]

$y \in \partial C_k$ but we may not have $y \in \Gamma_{k,j}$, therefore, $a = d_N(y, y_{k,j}) \geq d_N(y, y_k) = b$. $d_N(y, y_{k,j}) \leq d_N(y, y_k) + d_N(y_{k,j}, y_k)$, hence by equation (3.64), $b \leq D_1r$ and $a \leq 2D_1r$. By Lemma B.5, (B.25) $|AB| = a(1 + O(r))$ and $|AO| = b(1 + O(r))$. Since $a \geq b$ and $|AB| \leq |AO|$, we have $0 \leq a - b = D_1O(r^2)$. Hence, (B.26) $|AO| - |AB| = D_1O(r^2)$. By Lemma B.5 and Assumption 3.9,

\[
|BO| = d_N(y_k, y_{k,j})(1 + O(r)) \geq D_2r(1 + O(r)),
\]

\[
|AO| = d_N(y, y_k)(1 + O(r)) \leq D_1r(1 + O(r)).
\]
Hence, $(|AB| + |AO|)(|AB| - |AO|) \leq 2|AO|(|AB| - |AO|) \leq 2D_2^2O(r^3)$. Moreover,

$$(B.27) \quad |CM| = \frac{(|AB| + |AO|)(|AB| - |AO|)}{2|BO|} \leq \frac{D_1^2}{D_2}O(r^2) = O(r^2).$$

At last,

$$(B.28) \quad |AD| = \frac{|CM||AO|}{|OC|} \leq \frac{|CM||AO|}{|OM|} = \frac{2|CM||AO|}{|OB|} \leq \frac{2|CM|D_1r(1 + O(r))}{D_2r(1 + O(r))} = O(r^2).$$

Since $d_{\mathbb{R}^d}(A, \partial \tilde{C}) \leq |AD|$, the conclusion follows.

In the second step, we show that for any $v \in \tilde{C}_k$, $d_{\mathbb{R}^d}(v, \partial \tilde{k}_k(C_k)) = O(r^2)$. The proof follows the similar argument as the first step, so we omit it.

Now we prove the first main proposition.

**Proof of Proposition 3.12**

Proof. By Assumption 3.9, for any $y \in \partial C_k$, $\frac{1}{2}D_2r \leq d_N(y, y_k) \leq D_1r$. By Lemma B.6 any for $v \in \partial \tilde{k}_k(C_k)$, $\frac{1}{2}D_2r(1 + O(r)) \leq \|v\|_{\mathbb{R}^d} \leq D_1r(1 + O(r))$. Hence, $\frac{1}{2}D_2r + O(r^2) \leq \|v\|_{\mathbb{R}^d} \leq D_1r + O(r^2)$.

By Lemma B.6 and the triangle inequality, for any $v' \in \partial \tilde{C}_k$, $\frac{1}{2}D_2r + O(r^2) \leq \|v'\|_{\mathbb{R}^d} \leq D_1r + O(r^2)$.

Since $\tilde{C}_{k,0}$ is convex, we conclude that there is a constant $\Omega$ such that $|\tilde{C}_{k,0}| = \Omega d + O(r^{d+1})$. By Lemma B.5 and the fact that $\tilde{C}_{k,0}$ is convex, $|\tilde{k}_k(C_k)| = \Omega d + O(r^{d+1}) = |\tilde{C}_{k,0}|(1 + O(r))$. By Lemma B.5 $|C_k| = |\tilde{k}_k(C_k)|(1 + O(r))^d = |\tilde{k}_k(C_k)|(1 + O(r))$. Therefore, $|\tilde{C}_k| = |\tilde{C}_{k,0}| = |C_k|(1 + O(r))$. 

**Proof of Proposition 3.13**

Proof. We provide a sketch of the proof. Use $|\cdot|$ to denote the $d-1$ dimensional Hausdorff measure. $\partial A$ denotes the topological boundary of a set $A$. Suppose $B^\mathbb{R}^d_i(y_k) \cap \{y_i\}_{i=1}^n = \{y_{k,1}, \cdots, y_{k,N_k}\}$.

Step 1 We approximate the Voronoi face $\Gamma_{k,i}$ by a region in a $d-1$ dimensional affine subspace in $\mathbb{R}^\ell$.

Suppose the minimizing geodesic intersects the bisector $G$ between $y$ and $y_{k,i}$ at $y_{k,i}^*$. Then, by proposition 3.2 there is a $d-1$ dimensional subspace $S_{k,i}$ of $T_{y_{k,i}}N$ which is perpendicular to the tangent vector of the minimizing geodesic at $y_{k,i}^*$. If we realize $T_{y_{k,i}}N$ as a subspace of $\mathbb{R}^\ell$, then the affine subspace $y_{k,i}^* + S_{k,i}$ is tangent to $G$ at $y_{k,i}^*$. Without loss of generality, we rotate and translate the manifold $N$ so that $y_{k,i}^* = 0$ and $S_{k,i}$ is identified with the subspace of $\mathbb{R}^\ell$ generated by the first $d-1$ standard basis. By Assumption 3.3, there is an open subset of $S_{k,i}$ and denote $L_{k,i}$ to be its closure such that for any $y \in \Gamma_{k,i}$, we have

$$y = (u, g_1(u), \cdots, g_{\ell-d+1}(u)),$$

where $u \in L_{k,i} \subset \mathbb{R}^{d-1}$ and $g_i : \mathbb{R}^{d-1} \rightarrow \mathbb{R}$. Moreover, $g_j(0) = 0$ and $\nabla g_j(0) = 0$. The second order derivative of $g_i$ can be bounded by the curvature of $N$ at $y_k$ and $y_{k,i}$. By (1) in Assumption 3.9 $\Gamma_{k,i} \subset C_k \subset B^\mathbb{R}^d_i(y_k)$, hence for any $y \in \Gamma_{k,i}$, $\|y - y_k\|_{\mathbb{R}^\ell} \leq r$. $\|y_{k,i} - y_k\|_{\mathbb{R}^\ell} \leq d_N(y_{k,i}^*, y_k) = \frac{1}{2}d_N(y_{k,i}, y_k) \leq \frac{1}{2}D_1r$. Since $y_{k,i} = 0$,

$$\|y\|_{\mathbb{R}^\ell} = \|y - y_{k,i}^*\|_{\mathbb{R}^\ell} \leq \|y - y_k\|_{\mathbb{R}^\ell} + \|y_{k,i}^* - y_k\|_{\mathbb{R}^\ell} \leq (1 + \frac{1}{2}D_1)r.$$
By (B.29), for any \( u \in L_{k,i} \), \( \|u\|_{\mathbb{R}^{d-1}} \leq \|y\|_{\mathbb{R}^d} \leq (1 + \frac{1}{2}D_1)r \). Thus, \( L_{k,i} \) is contained in a \( d - 1 \) dimensional ball of radius \( (1 + \frac{1}{2}D_1)r \) in \( \mathbb{R}^{d-1} \). Hence

(B.31) \[ |L_{k,i}| \leq |S^{d-1}|(1 + \frac{1}{2}D_1)d^{-1}r^{d-1} \]

(B.29) implies that

(B.32) \[ |\Gamma_{k,i}| = |L_{k,i}| + O(|L_{k,i}|^{\frac{d}{d-1}}) = |L_{k,i}| + O(r^d), \]

where we use \( |L_{k,i}| \leq |S^{d-1}|(1 + \frac{1}{2}D_1)d^{-1}r^{d-1} \) in the last step. Moreover,

(B.33) \[ d_H^d(\partial \Gamma_{k,i}, \partial L_{k,i}) = \max_{u \in \partial L_{k,i}} \sqrt{g_1^2(u) + \cdots + g_{d-1}^2(u)} = O(r^2), \]

where \( d_H^d \) is the Hausdorff distance with respect to the Euclidean metric of \( \mathbb{R}^d \).

Step 2

This step is an analogue of Lemma B.5 when we apply \( \tilde{i}_k \) to the affine subspace \( y_{k,i}^* + T_{y_{k,i}}^* \mathcal{N} \). If we identify both \( T_{y_{k,i}}^* \mathcal{N} \) and \( T_{y_{k,i}} \mathcal{N} \) as the subspaces of \( \mathbb{R}^\ell \), then we show that \( T_{y_{k,i}}^* \mathcal{N} \) is a small perturbation of \( T_{y_{k,i}} \mathcal{N} \) when \( r \) is small. For simplicity, we rotate and translate the manifold so that \( y_k = 0 \) and \( T_{y_{k,i}} \mathcal{N} \) is generated by the first standard orthonormal basis \( \{e_1, \cdots, e_d\} \) of \( \mathbb{R}^d \). By the manifold structure of \( \mathcal{N} \), there is an orthonormal basis \( \{e'_1, \cdots, e'_d\} \) of \( T_{y_{k,i}}^* \mathcal{N} \) with \( e'_i = e_i + O(r^2) \). By Proposition B.4 and the similar argument in Lemma B.5, we can show that \( \tilde{i}_k \) restricted on the affine subspace \( y_{k,i}^* + T_{y_{k,i}}^* \mathcal{N} \) is a \( 1 + O(r) \) bi-Lipschitz homeomorphism.

Step 3

For simplicity, denote \( \tilde{i}_k(y_{k,i}^* + L_{k,i}) \) by \( \tilde{i}_k(L_{k,i}) \), denote \( \tilde{i}_k(y_{k,i}^* + \partial L_{k,i}) \) by \( \tilde{i}_k(\partial L_{k,i}) \) and denote \( \partial \tilde{i}_k(y_{k,i}^* + L_{k,i}) \) by \( \partial \tilde{i}_k(\partial L_{k,i}) \). Since \( \tilde{i}_k \) restricted on the affine subspace \( y_{k,i}^* + T_{y_{k,i}}^* \mathcal{N} \) is homeomorphism, \( \partial \tilde{i}_k(L_{k,i}) = \tilde{i}_k(\partial L_{k,i}) \). Moreover, Lemma B.5 shows that \( \tilde{i}_k \) restricted on \( B_{y_{k,i}^*} \cap \mathcal{N} \) is a homeomorphism. Hence, \( \partial \tilde{i}_k(\Gamma_{k,i}) = \tilde{i}_k(\partial \Gamma_{k,i}) \). Since \( \tilde{i}_k \) is a projection,

(B.34) \[ d_H^d(\partial \tilde{i}_k(\Gamma_{k,i}), \partial \tilde{i}_k(\partial L_{k,i})) = d_H^d(\tilde{i}_k(\partial \Gamma_{k,i}), \tilde{i}_k(\partial L_{k,i})) \leq d_H^d(\partial \Gamma_{k,i}, \partial L_{k,i}) = O(r^2), \]

where we use (B.33) in the last step. Since \( \tilde{i}_k \) is a projection, \( \tilde{i}_k(L_{k,i}) \) is a subset of a \( d - 1 \) dimensional affine subspace of \( \mathbb{R}^d \). Since \( \tilde{i}_k \) restricted on the affine subspace \( y_{k,i}^* + T_{y_{k,i}}^* \mathcal{N} \) is a \( 1 + O(r) \) bi-Lipschitz homeomorphism,

(B.35) \[ |L_{k,i}| = |\tilde{i}_k(L_{k,i})| (1 + O(r)). \]

Step 4

Recall in the step (4) in Algorithm 1 we find the Voronoi cell decomposition of \( \{0, \tilde{i}_k(y_{k,1}), \cdots, \tilde{i}_k(y_{k,N_k})\} \) in \( \mathbb{R}^d \). The Voronoi cell containing 0 is \( \tilde{C}_{k,0} \). The Voronoi face between 0 and \( \tilde{i}_k(y_{k,i}) \) is denoted as \( \tilde{F}_{k,i} \). If \( y \in \partial \Gamma_{k,i} \), then there is a third point \( y_{k,j} \) such that \( d_{\mathcal{N}}(y, y_{k,i}) = d_{\mathcal{N}}(y, y_{k,j}) = d_{\mathcal{N}}(y, y_{k,j}) \).

By using the similar argument in Lemma B.6 we can show that

(B.36) \[ d_H^d(\partial \tilde{i}_k(\Gamma_{k,i}), \partial \tilde{F}_{k,i}) = O(r^2). \]

By (B.34), we have

(B.37) \[ d_H^d(\partial \tilde{i}_k(L_{k,i}), \partial \tilde{F}_{k,i}) = O(r^2). \]

Step 5
By (1) in Assumption 3.9, $C_k \subset B_{r^2}^\mathbb{R}^d(y_k) \cap \mathcal{N}$. Since $\tilde{i}_k$ is a projection, $\tilde{i}_k(C_k)$ is in the ball of radius $r$ centered at 0 in $\mathbb{R}^d$. By Lemma B.6, $\tilde{C}_k$ is in the ball of radius $2r$ centered at 0 in $\mathbb{R}^d$, when $r$ is small enough. $\tilde{F}_{k,i}$ is a convex polygon and is in a $d-1$ dimensional affine subspace $H_{k,i}$ in $\mathbb{R}^d$. We know that $\partial \tilde{F}_{k,i} = \cup_j C_j$, where each $C_j$ is a $d-2$ dimensional convex polygon. Each $C_j$ is a ball of radius $2r$. Hence, we have $\mathcal{H}^{d-2}(\partial \tilde{F}_{k,i}) = O(r^{d-2})$ and any $O(r^2)$ neighborhood of $\partial \tilde{F}_{k,i}$ in $H_{k,i}$ has $d-1$ Hausdorff measure $O(r^d)$. Since $d_{H}^2(\partial \tilde{i}_k(L_{k,i}), \partial \tilde{\Gamma}_{k,i}) = O(r^2)$ and $\tilde{i}_k(L_{k,i})$ is a subset of a $d-1$ dimensional affine subspace of $\mathbb{R}^d$, we rotate and translate $\tilde{i}_k(L_{k,i})$ so that $\partial \tilde{i}_k(L_{k,i})$ is in a $O(r^2)$ neighborhood of $\partial \tilde{\Gamma}_{k,i}$ in $H_{k,i}$. Therefore,

$$|\tilde{i}_k(L_{k,i})| = |\tilde{F}_{k,i}| + O(r^d)$$  \hspace{1cm} (B.38)

Combine (B.31), (B.32), (B.35) and (B.38), we have

$$|\Gamma_{k,i}| = |\tilde{F}_{k,i}| + O(r^d).$$  \hspace{1cm} (B.39)

If $y_\ell = y_{k,i} \in B_{r^2}^\mathbb{R}^d(y_k)$, then $\tilde{A}_{k\ell} = |\tilde{F}_{k,i}|$. So, $|\Gamma_{k\ell}| = \tilde{A}_{k\ell} + O(r^d)$. Similarly, $|\Gamma_{\ell k}| = |\Gamma_{k\ell}| = \tilde{A}_{k\ell} + O(r^d)$. Hence, $|\Gamma_{k\ell}| = |\Gamma_{\ell k}| = \tilde{A}_{k\ell} + O(r^d) = A_{k\ell} + O(r^d)$. If $A_{k\ell} \geq a_1 r^d$, we automatically have the conclusion. If $A_{k\ell} < a_1 r^d$, then $|\Gamma_{k\ell}| = O(r^d)$ and $|\tilde{\Gamma}_{k\ell}| = a_1 r^d$. So, we also have $|\Gamma_{k\ell}| = |\tilde{\Gamma}_{k\ell}| + O(r^d)$.

\[\square\]