Appl. Comput. Harmon. Anal. 62 (2023) 261–309

Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

Data-driven efficient solvers for Langevin dynamics on manifold in high dimensions

Yuan Gao [a], Jian-Guo Liu [b], Nan Wu [c,*]

[a] *Department of Mathematics, Purdue University, West Lafayette, IN, United States of America*
[b] *Department of Mathematics and Department of Physics, Duke University, Durham, NC, United States of America*
[c] *Department of Mathematical Sciences, The University of Texas at Dallas, Richardson, TX, United States of America*

ARTICLE INFO

Article history:
Received 22 May 2020
Received in revised form 30 May 2022
Accepted 19 September 2022
Available online 26 September 2022
Communicated by Hau-Tieng Wu

Keywords:
Diffusion map
Reaction coordinates
Voronoi tessellation
Unconditionally stable explicit finite volume scheme
Random walk on point clouds
Exponential convergence

ABSTRACT

We study the Langevin dynamics of a physical system with manifold structure $\mathcal{M} \subset \mathbb{R}^p$ based on collected sample points $\{\mathbf{x}_i\}_{i=1}^n \subset \mathcal{M}$ that probe the unknown manifold $\mathcal{M}$. Through the diffusion map, we first learn the reaction coordinates $\{\mathbf{y}_i\}_{i=1}^n \subset \mathcal{N}$ corresponding to $\{\mathbf{x}_i\}_{i=1}^n$, where $\mathcal{N}$ is a manifold diffeomorphic to $\mathcal{M}$ and isometrically embedded in $\mathbb{R}^\ell$ with $\ell \ll p$. The induced Langevin dynamics on $\mathcal{N}$ in terms of the reaction coordinates captures the slow time scale dynamics such as conformational changes in biochemical reactions. To construct an efficient and stable approximation for the Langevin dynamics on $\mathcal{N}$, we leverage the corresponding Fokker-Planck equation on the manifold $\mathcal{N}$ in terms of the reaction coordinates $\mathbf{y}$. We propose an implementable, unconditionally stable, data-driven finite volume scheme for this Fokker-Planck equation, which automatically incorporates the manifold structure of $\mathcal{N}$. Furthermore, we provide a weighted L^2 convergence analysis of the finite volume scheme to the Fokker-Planck equation on $\mathcal{N}$. The proposed finite volume scheme leads to a Markov chain on $\{\mathbf{y}_i\}_{i=1}^n$ with an approximated transition probability and jump rate between the nearest neighbor points. After an unconditionally stable explicit time discretization, the data-driven finite volume scheme gives an approximated Markov process for the Langevin dynamics on $\mathcal{N}$ and the approximated Markov process enjoys detailed balance, ergodicity, and other good properties.

* Corresponding author.
E-mail addresses: gao662@purdue.edu (Y. Gao), jliu@math.duke.edu (J.-G. Liu), nan.wu@utdallas.edu (N. Wu).

https://doi.org/10.1016/j.acha.2022.09.003
1063-5203/

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 $\mathbf{x}$ in $\mathbb{R}^p$ with $p \gg 1$. Due to some equality and inequality constraints, we assume the essential structure of the system P is an unknown d dimensional closed smooth Riemannian submanifold $\mathcal{M}$ of $\mathbb{R}^p$ [15,14]. The manifold $\mathcal{M}$ is unknown in the sense that we do not know the charts and the metric of $\mathcal{M}$. 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 [15,14,48]. For instance, a typical one dimensional reaction coordinate is the distance between a carbon center and a nucleophile in an S_N2 reaction (one simple nucleophilic substitution reaction mechanism); see also the conformational transitions of alanine dipeptide representing by two backbone dihedral angles [28]. There are many other collective physical/chemical quantities, such as bond length/angle, dihedral angles, and intermolecular distance, that can be used as the reaction coordinates to represent the whole process of conformational transitions or chemical reactions. Mathematically, the reaction coordinates should be a smooth embedding $\mathbf{y} = \Phi(\mathbf{x}) : \mathcal{M} \hookrightarrow \mathbb{R}^\ell$ with $\ell \ll p$ 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 [45]. Suppose $\{\mathbf{x}_i\}_{i=1}^n$ are n data points well distributed on the unknown manifold $\mathcal{M} \subset \mathbb{R}^p$, while these n points are collected by some sampling methods. A nonlinear dimension reduction algorithm constructs an embedding Φ by using the coordinates of $\{\mathbf{x}_i\}_{i=1}^n$ in $\mathbb{R}^p$ so that we can present the high dimensional data $\{\mathbf{x}_i\}_{i=1}^n \subset \mathcal{M} \subset \mathbb{R}^p$ as $\mathbf{y}_i = \Phi(\mathbf{x}_i) \subset \mathcal{N} \subset \mathbb{R}^\ell$ in the low dimensional space.

We assume the dynamics for the physical system P can be described by a continuous strong Markov process on $\mathcal{M} \subset \mathbb{R}^p$. Particularly, the simplest and widely used physical model is the following over-damped Langevin dynamics with a drift determined by some potentials U on $\mathcal{M}$:

$$\mathrm{d}\mathbf{x}_t = -\nabla_{\mathcal{M}} U(\mathbf{x}_t)\,\mathrm{d}t + \sqrt{2kT}\,\mathrm{d}_{\mathcal{M}} B_t. \tag{1.1}$$

We explain the notations in (1.1) below. Let $\{\tau_i^{\mathcal{M}};\ 1 \le i \le d\}$ be an orthonormal basis of the tangent plane $T_{\mathbf{x}_t}\mathcal{M}$. Denote $\nabla_{\mathcal{M}} := \sum_{i=1}^d \tau_i^{\mathcal{M}} \nabla_{\tau_i^{\mathcal{M}}} = \sum_{i=1}^d \tau_i^{\mathcal{M}} \otimes \tau_i^{\mathcal{M}} \nabla$ as the surface gradient and $\nabla_{\tau_i^{\mathcal{M}}} = \tau_i^{\mathcal{M}} \cdot \nabla$ as the tangential derivative in the direction of $\tau_i^{\mathcal{M}}$. Let k be the Boltzmann constant, T be the temperature [20] and $\mathrm{d}_{\mathcal{M}} B_t$ be a Brownian motion on $\mathcal{M}$. This Brownian motion on manifold can be realized by the projection of the standard Brownian motion B_t in the ambient space $\mathbb{R}^p$:

$$\mathrm{d}_{\mathcal{M}} B_t := \sum_{i=1}^d \tau_i^{\mathcal{M}}(\mathbf{x}_t) \otimes \tau_i^{\mathcal{M}}(\mathbf{x}_t) \circ \mathrm{d}B_t, \tag{1.2}$$

where $\circ$ is understood in the Stratonovich sense in the stochastic integral [35, p. 19, p. 78]. We refer the readers to [3] for the general Langevin SDEs on Riemannian manifolds.

The potential $U(\mathbf{x})$ is also known as the energy landscape for the physical system P, which is usually very complicated and indicates all the possible (meta)stable states of a physical system. For instance, in a simple nucleophilic substitution reaction mechanism, the states of reactants and products are two stable states in the energy landscape [20]. A saddle point state on the energy landscape is called the transition state, whose value determines the energy barrier in a chemical reaction. In this paper, we assume the output of the potential U at each data point $\{\mathbf{x}_i\}_{i=1}^n$ is known. The equilibrium of this system P, also known as the invariant probability density measure, is $\rho_\infty(\mathbf{x}) \propto e^{-\frac{U(\mathbf{x})}{kT}}$, $\mathbf{x} \in \mathcal{M}$.

Suppose we learn the reaction coordinates $\mathbf{y} = \Phi(\mathbf{x}), \mathbf{x} \in \mathcal{M}$. The diffeomorphism $\Phi : \mathcal{M} \to \mathcal{N}$ induces a map Φ_* from the space $\Gamma(T\mathcal{M})$ of the smooth vector fields on $\mathcal{M}$ to the space $\Gamma(T\mathcal{N})$ of the smooth vector fields on $\mathcal{N}$ such that for any $f \in C^\infty(\mathcal{N})$ and $V \in \Gamma(T\mathcal{M})$

$$(\Phi_* V) f(\mathbf{y}) = V(f \circ \Phi)(\mathbf{x}), \quad \mathbf{y} = \Phi(\mathbf{x}). \tag{1.3}$$

The Stratonovich formulation transforms consistently under diffeomorphism Φ [35, p. 20]. Notice $\tau_i^{\mathcal{N}} \in \mathbb{R}^\ell$ is defined by the induced map Φ_* and B_t is the ℓ-dimensional Brownian motion. Therefore, instead of considering (1.1) on $\mathcal{M}$ directly, we consider the SDE on $\mathcal{N}$

$$\mathrm{d}\mathbf{y}_t = -\nabla_{\mathcal{N}} U_{\mathcal{N}}(\mathbf{y}_t)\,\mathrm{d}t + \sqrt{2kT}\sum_{i=1}^{d} \tau_i^{\mathcal{N}}(\mathbf{y}_t) \otimes \tau_i^{\mathcal{N}}(\mathbf{y}_t) \circ \mathrm{d}B_t, \tag{1.4}$$

where $\nabla_{\mathcal{N}} := \sum_{i=1}^d \tau_i^{\mathcal{N}} \nabla_{\tau_i^{\mathcal{N}}} = \sum_{i=1}^d \tau_i^{\mathcal{N}} \otimes \tau_i^{\mathcal{N}} \nabla$ is the surface gradient, $\nabla_{\tau_i^{\mathcal{N}}} = \tau_i^{\mathcal{N}} \cdot \nabla$ is the tangential derivative in the direction of $\tau_i^{\mathcal{N}}$, and $U_{\mathcal{N}}$ is the induced potential on manifold $\mathcal{N}$ by the composition

$$U_{\mathcal{N}}(\mathbf{y}) := U(\mathbf{x}) = U(\Phi^{-1}(\mathbf{y})). \tag{1.5}$$

The main goal of this paper is to accurately simulate the induced Langevin dynamics (1.4) in terms of the reaction coordinates $\mathbf{y}$ and the information of the potential U. To simulate the SDE (1.4) on $\mathcal{N}$ without exact manifold information, one of the most natural ways is to construct an approximated stochastic process on the points $\{\mathbf{y}_i\}_{i=1}^n$. However, the standard Euler–Maruyama method on manifold can not provide a stable simulation. Hence, our strategy for constructing a stochastic process over $\{\mathbf{y}_i\}_{i=1}^n$ is described as follows: (i) we detour to approximate the corresponding Fokker-Planck equation on the manifold with a finite volume scheme; (ii) we construct an approximated Voronoi tessellation associated with $\{\mathbf{y}_i\}_{i=1}^n$; (iii) we construct the transition probability and the jump rate from the finite volume scheme.

By Ito's formula, the SDE (1.4) gives the following Fokker-Planck equation, which is the governing equation for the density $\rho_t^{\mathcal{N}} := \rho^{\mathcal{N}}(\mathbf{y}, t)$,

$$\partial_t \rho_t^{\mathcal{N}} = \mathrm{div}_{\mathcal{N}}(kT\nabla_{\mathcal{N}}\rho_t^{\mathcal{N}} + \rho_t^{\mathcal{N}}\nabla_{\mathcal{N}} U_{\mathcal{N}}) =: \mathrm{FP}^{\mathcal{N}}\rho_t^{\mathcal{N}}, \tag{1.6}$$

where $\mathrm{div}_{\mathcal{N}}$ is the surface divergence defined as $\mathrm{div}_{\mathcal{N}}\xi = \sum_{i=1}^d \tau_i^{\mathcal{N}} \cdot \nabla_{\tau_i^{\mathcal{N}}}\xi$. One equivalent form of (1.6) is the relative entropy formulation

$$\partial_t \rho_t^{\mathcal{N}} = kT\mathrm{div}_{\mathcal{N}}\left(e^{-\frac{U_{\mathcal{N}}}{kT}}\nabla_{\mathcal{N}}(\rho_t^{\mathcal{N}} e^{\frac{U_{\mathcal{N}}}{kT}})\right) = \mathrm{FP}^{\mathcal{N}}\rho_t^{\mathcal{N}}. \tag{1.7}$$

Then the main issue is to simulate the Fokker-Planck equation (1.6) on $\mathcal{N}$ whose solution $\rho_t^{\mathcal{N}}(\mathbf{y})$ drives any initial density $\rho_0^{\mathcal{N}}$ to the invariant measure $\rho_\infty^{\mathcal{N}}(\mathbf{y}) \propto e^{-\frac{U_{\mathcal{N}}(\mathbf{y})}{kT}}$. After designing a finite volume scheme for the Fokker-Planck equation (1.6) on $\mathcal{N}$, we construct an approximated transition probability and jump rate from it. This approximated Markov process on the manifold automatically incorporates both the manifold structure and the equilibrium information. It enables some implementable data-driven computations on the manifold such as finding the optimal cluster-cluster coarse-grained network, cf. [16,19,37,40,50,47] and finding the transition path and energy landscape of chemical reactions, cf. [23,21,43,22,28,33,31,32].

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 assume that we can sample n points $\{\mathbf{x}_i\}_{i=1}^n$ from $\mathcal{M}$ based on a density function on $\mathcal{M}$ with lower and

upper bounds so that the data points are well distributed on $\mathcal{M}$. In Section 2, we first show that the diffusion map can approximate an embedding Φ of the manifold $\mathcal{M}$. Then, we apply the diffusion map algorithm [15] on $\{\mathbf{x}_i\}_{i=1}^n$ to find the reaction coordinates so that we have $\{\mathbf{y}_i = \Phi(\mathbf{x}_i)\}_{i=1}^n \subset \mathcal{N} = \Phi(\mathcal{M}) \subset \mathbb{R}^\ell$. Note that $\{\mathbf{y}_i\}_{i=1}^n$ can also be regarded as the samples based on a density function on $\mathcal{N}$ with lower and upper bounds.

Next, we focus on simulating the Fokker-Planck equation (1.6) with a given equilibrium potential $U_{\mathcal{N}}(\mathbf{y})$. To find the trajectory $\rho_t^{\mathcal{N}}$, we need to solve the Fokker-Planck equation on the manifold $\mathcal{N} \subset \mathbb{R}^\ell$. Our method uses the data points $\{\mathbf{y}_i\}_{i=1}^n \subset \mathcal{N}$ to construct a discrete approximation of the Fokker-Planck equation (1.6). It is proved that the data points are well-distributed on $\mathcal{N}$ whenever the points are sampled based on a density function with lower and upper bounds [55,41]. Hence, we can construct a "regularly shaped" Voronoi tessellation on $\mathcal{N}$ from $\{\mathbf{y}_i\}_{i=1}^n \subset \mathcal{N}$. With the help of such Voronoi tessellation, we introduce a finite volume scheme by applying the relative entropy formulation and finite volume method to (1.6). The finite volume scheme assigns a transition probability and a jump rate for an approximated Markov process on $\{\mathbf{y}_i\}_{i=1}^n$, i.e., random walk between the nearest neighbor points. In Section 3.2, we prove all the good properties of the approximated Markov process on $\{\mathbf{y}_i\}_{i=1}^n$ including detailed balance, ergodicity, L^1-contraction and χ^2-divergence dissipation law.

To obtain an implementable finite volume scheme, an approximated Voronoi tessellation associated with $\{\mathbf{y}_i\}_{i=1}^n$ needs to be constructed with high accuracy. By using the Euclidean coordinates of $\{\mathbf{y}_i\}_{i=1}^n$, each Voronoi cell can be approximated by a polygon in a tangent space of $\mathcal{N}$; see Section 3.4 for detailed error estimates for the approximated cell volume and face area. Therefore, an approximated transition probability based on the volume of each polygon and the areas of its faces can be assigned over $\{\mathbf{y}_i\}_{i=1}^n$ and leads to an implementable finite volume scheme (3.78) for the Fokker-Planck equation (1.6); see Section 3 and Theorem 3.14 for consistence and convergence analysis for this implementable finite volume scheme. We also provide an unconditionally stable explicit time descretization for the finite volume scheme based on the detailed balance property of the Markov process in Section 3.5. This explicit scheme is very efficient and enjoys a mass conservation law, unconditional maximum principle and exponential convergence to equilibrium. At last, to show the accuracy of the finite volume scheme, we simulate challenging numerical examples including datasets on a dumbbell, a Klein bottle and a sphere in Section 4.

The approximated transition probability between the nearest neighbor points for the Markov process on $\{\mathbf{y}_i\}_{i=1}^n \subset \mathcal{N}$ reveals the manifold structure and enables us to efficiently conduct computations such as clustering, coarse-graining and finding the minimal energy path on the 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.

The remaining part of the paper will be organized as follows. In Section 2, we use diffusion map to learn the reaction coordinates $\mathbf{y}$. In Section 3, we propose a data-driven solver for the Fokker-Planck equation on manifold $\mathcal{N}$, which assigns an approximated transition probability and a jump rate for an approximated Markov process on $\{\mathbf{y}_i\}_{i=1}^n$. In Section 4, we also provide several simulation results for the Fokker-Planck dynamics on manifolds learned from point clouds. All the technical lemmas are provided in Appendix for completeness. 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 learning the reaction coordinates $\mathbf{y}$ for the d-dimensional manifold $\mathcal{N} \subset \mathbb{R}^\ell$ 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.

Assumption 2.1. Let $\mathcal{M}$ be a d dimensional smooth closed Riemannian submanifold of $\mathbb{R}^p$. Suppose that ρ^* is a smooth probability density function on the manifold $\mathcal{M}$. We assume that ρ^* is bounded from below and from above, i.e. $\rho_m \leq \rho* \leq \rho_M$. Let $\{\mathbf{x}_1 \cdots, \mathbf{x}_n\} \subset \mathcal{M} \overset{i.i.d.}{\sim} \rho^*$.

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
$\mathbf{x}$, $\mathbf{y}$	Points on $\mathcal{M}$ and $\mathcal{N}$ respectively
$dV_{\mathcal{M}}$, $dV_{\mathcal{N}}$	Volume forms on $\mathcal{M}$ and $\mathcal{N}$ respectively
Δ	Laplace Beltrami operator of a manifold
λ_i, ψ_i	The eigenvalues and the corresponding orthonormal (in L^2) eigenfunctions of Δ
Φ	Reaction coordinates (Smooth embedding of a manifold)
X, Y	Random variables with the range $\mathcal{M}$ and $\mathcal{N}$ respectively
ρ^*, $\rho_t^{\mathcal{M}}$	Probability density functions on $\mathcal{M}$
ρ^{**}, $\rho_t^{\mathcal{N}}$	Probability density functions on $\mathcal{N}$
$n \in \mathbb{N}$	Number of data points sampled from $\mathcal{M}$ based on ρ
$\{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$	Data points sampled from $\mathcal{M}$ based on ρ
ϵ	The bandwidth in the diffusion map
K_ϵ	Kernel in the diffusion map
$W_{\epsilon,\alpha}$	Affinity matrix in diffusion map with α 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 $\frac{I-L_{\epsilon,1}}{\epsilon^2}$
C_i	the Voronoi cell around the point $\mathbf{y}_i$ on the manifold $\mathcal{N}$
Γ_{ij}	the Voronoi face between $\mathbf{y}_i$ and $\mathbf{y}_j$ on the manifold $\mathcal{N}$
r	bandwidth in the Voronoi cell approximation algorithm
s	threshold in the Voronoi cell approximation algorithm
ι_k	the projection map in the Voronoi cell approximation algorithm
P_{ij}, η_i	the transition probability and jump rate of constructed Markov chain

Nonlinear dimension reduction algorithms construct maps which map $\{\mathbf{x}_1 \cdots, \mathbf{x}_n\}$ to some low dimensional space $\mathbb{R}^\ell$ while preserving the topological or geometric structure of the underlying manifold. There are a lot of well known dimension reduction algorithms, for instance, ISOMAP [53], eigenmap [6], locally linear embedding (LLE) [49] and its variations like Hessian LLE [17], vector diffusion map [51,52]. In this work, we focus on the algorithm diffusion map which is introduced by Coifman and Lafon [15]. The algorithm of the diffusion map can be described in the following steps:

(i) For $\mathbf{x}, \mathbf{x}' \in \mathcal{M}$, we define $K_\epsilon(\mathbf{x}, \mathbf{x}') = \exp(-\frac{\|\mathbf{x}-\mathbf{x}'\|^2_{\mathbb{R}^p}}{4\epsilon^2})$, where $\epsilon > 0$ is the bandwidth.
(ii) Define $q_\epsilon(\mathbf{x}) := \sum_{i=1}^n K_\epsilon(\mathbf{x}, \mathbf{x}_i)$. We define the affinity matrix which is the $n \times n$ matrix $W_{\epsilon,\alpha}$: $W_{\epsilon,\alpha,ij} := \frac{K_\epsilon(\mathbf{x}_i,\mathbf{x}_j)}{q_\epsilon^\alpha(\mathbf{x}_i) q_\epsilon^\alpha(\mathbf{x}_j)}$. This step is called the α-normalization.
(iii) Define the $n \times n$ diagonal matrix D with diagonal entries $D_{\epsilon,\alpha,ii} = \sum_{j=1}^n W_{\epsilon,\alpha,ij}$. Let

$$L_{\epsilon,\alpha} = D_{\epsilon,\alpha}^{-1} W_{\epsilon,\alpha}. \tag{2.1}$$

(iv) To reduce the dimension of the dataset $\{\mathbf{x}_1 \cdots, \mathbf{x}_n\}$. We choose $\alpha = 1$. Denote

$$\lambda_{0,n,\epsilon} \leq \cdots, \leq \lambda_{n-1,n,\epsilon} \tag{2.2}$$

to be the eigenvalues of $\frac{I-L_{\epsilon,1}}{\epsilon^2}$. We find the first ℓ corresponding eigenvectors of $\frac{I-L_{\epsilon,1}}{\epsilon^2}$, namely, $\{v_{j,n,\epsilon}\}_{j=1}^\ell$. Then the map

$$\mathbf{x}_i \to (v_{1,n,\epsilon}(i), \cdots, v_{\ell,n,\epsilon}(i)) \tag{2.3}$$

reduces the dimension of the dataset into the Euclidean space $\mathbb{R}^\ell$.

Remark 2.2. Note that the matrix $L_{\epsilon,1}$ in (2.1) 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 $D_{\epsilon,1}^{-1/2}$.

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

$$\Delta\psi_i = -\lambda_i\psi_i, \tag{2.4}$$

where ψ_i is the corresponding eigenfunction normalized in $L^2(\mathcal{M})$. We have $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \cdots$. Note that $\psi_0 = \frac{1}{\sqrt{\mathcal{M}}}$ is a constant.

In the rest of this section, we will provide a justification that the diffusion map

$$\mathbf{x}_i \to (v_{1,n,\epsilon}(i), \cdots, v_{\ell,n,\epsilon}(i)) \tag{2.5}$$

approximates an embedding of $\mathcal{M}$ into a Euclidean space. The justification consists of two steps. First, we review the results about the spectral convergence from $\frac{I-L_{\epsilon,1}}{\epsilon^2}$ to $-\Delta$. Intuitively, these results show that the eigenpairs of $\frac{I-L_{\epsilon,1}}{\epsilon^2}$ approximate the corresponding eigenpair of $-\Delta$. Second, we discuss the result that shows the eigenfunctions of $-\Delta$ can be used to construct an embedding of $\mathcal{M}$. Since ψ_0 is a constant, based on the justification, the first eigenvector $v_{0,n,\epsilon}$ of $\frac{I-L_{\epsilon,1}}{\epsilon^2}$ is not used in the construction of the diffusion map.

We start from the theoretical results that relate the diffusion map to the Laplace Beltrami operator when the samples are from a submanifold. It is proved in [15] and [51] that $\frac{I-L_{\epsilon,1}}{\epsilon^2}$ converges pointwisely to $-\Delta$ in the following sense.

Theorem 2.3. *(Coifman-Lafon, [15], Singer-Wu, [51]) Suppose $\alpha = 1$. Under Assumption 2.1, for $f \in C^3(\mathcal{M})$, if $\frac{\sqrt{\log n}}{\sqrt{n}\epsilon^{\frac{d}{2}+2}} \to 0$ and $\epsilon \to 0$ as $n \to \infty$, then with probability greater than $1 - \frac{1}{n^2}$, for all $i = 1, \cdots, n$, we have*

$$\frac{f(\mathbf{x}_i) - \sum_{j=1}^{n} L_{\epsilon,1}(i,j)f(\mathbf{x}_j)}{\epsilon^2} = -\Delta f(\mathbf{x}_i) + O(\epsilon) + O(\frac{\sqrt{\log n}}{\sqrt{n}\epsilon^{\frac{d}{2}+2}}). \tag{2.6}$$

The $\alpha = 1$ normalization in the diffusion map comes from the idea of density estimation. When α is chosen to be 1, the impact of the nonuniform density ρ^* is removed. Hence, the Laplace-Beltrami operator in the previous theorem is not contaminated by the probability density function ρ^*.

A stronger version of the convergence theorem in [52] shows the spectral convergence of the diffusion map in L^2 sense. At last, in [18], it shows the L^∞ spectral convergence result based on the following definition.

Definition 2.4. Under Assumption 2.1, suppose $v_{j,n,\epsilon}$ is an eigenvector of $\frac{I-L_{\epsilon,1}}{\epsilon^2}$ which is normalized in the l^2 norm. Let $N_k = |B_\epsilon^{\mathbb{R}^p}(\mathbf{x}_k) \cap \{\mathbf{x}_1, \cdots, \mathbf{x}_n\}|$, the number of points in the ϵ 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:

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

where $|S^{d-1}|$ is the volume of the $d-1$ dimensional standard sphere. 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:

$$V_{j,n,\epsilon} := \frac{v_{j,n,\epsilon}}{\|v_{j,n,\epsilon}\|_{l^2(1/\hat{\rho^*})}}. \tag{2.7}$$

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 $\vec{\psi}_j = (\psi_j(\mathbf{x}_1), \cdots, \psi_j(\mathbf{x}_n))^\top$ is a discretization of ψ_j which is also normalized in the $L^2(\mathcal{M})$. Therefore, it is reasonable to compare $V_{j,n,\epsilon}$ and $\vec{\psi}_j$ rather than

$v_{j,n,\epsilon}$ and $\vec{\psi}_j$. In the following theorem, it shows that, on $\mathcal{M}$, if we fix K, the bandwidth ϵ is small enough based on K and the number of data points n is large enough based on ϵ, then for all $0 \leq j < K$, with high probability, $\lambda_{j,n,\epsilon}$ is an approximation of the j-th eigenvalue λ_j of $-\Delta$ and $V_{j,n,\epsilon}$ is an approximation of $\vec{\psi}_j$.

Theorem 2.5. *(Dunson-Wu-Wu, [18]) Under Assumption 2.1, suppose all the eigenvalues of Δ are simple. Let (λ_j, ψ_j) be the j-th eigenpair of $-\Delta$ with ψ_j normalized in $L^2(\mathcal{M})$. Let $L_{\epsilon,1}$ be the matrix in (2.1). Let $(\lambda_{j,n,\epsilon}, V_{j,n,\epsilon})$ be the j-th eigenpair of $\frac{I-L_{\epsilon,1}}{\epsilon^2}$ with $V_{j,n,\epsilon}$ normalized as in Definition 2.4. Fix a positive integer K, let $\Gamma_K := \min_{1\leq j\leq K} \mathrm{dist}(\lambda_j, \sigma(-\Delta) \setminus \{\lambda_j\})$, where $\sigma(-\Delta)$ is the spectrum of $-\Delta$. Suppose*

$$\epsilon \leq \mathcal{K}_1 \min\left(\left(\frac{\min(\Gamma_K, 1)}{\mathcal{K}_2 + \lambda_K^{d/2+5}}\right)^2, \frac{1}{(\mathcal{K}_3 + \lambda_K^{(5d+7)/4})^2}\right), \tag{2.8}$$

where $\mathcal{K}_1$ and $\mathcal{K}_2, \mathcal{K}_3 > 1$ are constants depending on d, the lower bound of the p.d.f. ρ_m, the C^2 norm of p.d.f., the volume, the injectivity radius, the curvature, and the second fundamental form of the manifold.

If n is sufficiently large so that $\epsilon = \epsilon(n) \geq (\frac{\log n}{n})^{\frac{1}{4d+13}}$, then with probability greater than $1 - n^{-2}$, for all $0 \leq j < K$

$$|\lambda_{j,n,\epsilon} - \lambda_j| \leq \mathcal{K}_4 \epsilon^{3/2}.$$

If n is sufficiently large so that $\epsilon = \epsilon(n) \geq (\frac{\log n}{n})^{\frac{1}{4d+8}}$, then with probability greater than $1 - n^{-2}$, there are $a_j \in \{1, -1\}$ such that for all $0 \leq j < K$

$$\max_{1\leq i\leq n} |a_j V_{j,n,\epsilon}(i) - \psi_j(x_i)| \leq \mathcal{K}_5 \epsilon^{1/2}.$$

$\mathcal{K}_4$ depends on d, the diameter of the manifold and the lower bound and the C^2 norm of the p.d.f. $\mathcal{K}_5$ depends on d, the diameter and the volume of the manifold, and the lower bound and the C^2 norm of the p.d.f.

Remark 2.6. Note that in the above theorem, the coefficients $\mathcal{K}_4$ and $\mathcal{K}_5$ only depend on the geometry of the manifold and the data points distribution on the manifold. They are independent of the eigenvalues and the eigengaps of Δ. In the spectral convergence analysis, the dependence on the eigenvalues and the eigengaps of Δ is reflected from the relation (2.8). The relation implies that ϵ should be smaller when K increases.

Moreover, the above theorem assumes that the eigenvalues of Δ are simple for notational simplicity. In the case when the eigenvalues are not simple, the same theorem still works by introducing an orthogonal transformation on the eigenspace. The readers may refer to Remark 4 in [12] for details.

The matrix $\frac{I-L_{\epsilon,1}}{\epsilon^2}$ can be regarded as the density corrected graph Laplacian on the complete undirected graph with vertices $\{\mathbf{x}_1 \cdots, \mathbf{x}_n\}$ and Gaussian weights on the edges. Hence, the above theorem discusses the spectral convergence of a density corrected graph Laplacian to the Laplace-Beltrami operator in the L^∞ sense. We also refer the readers to [54,10,12] which discuss the spectral convergence rates of the graph Laplacians to the Laplace Beltrami operator in the L^2 sense and [11] which is another work discussing the spectral convergence rate of the graph Laplacian to the Laplace-Beltrami operator in the L^∞ sense.

Next, we review some results in spectral geometry. Based on the work of [7], [36], [4] and [46], we know that the eigenfunctions of Δ can be used to construct an embedding of the manifold into a Euclidean space. More explicitly, we describe the following theorem in [4]. The readers may refer to Appendix A for more detailed discussions about the relevant theorems in [7], [36] and [46].

Theorem 2.7. *(Bates, [4]) Suppose $\mathcal{M}$ is a d dimensional smooth closed Riemannian manifold. Suppose that the Ricci curvature of $\mathcal{M}$ has lower bound $\mathrm{Ric}_{\mathcal{M}} \geq (d-1)k$, the injectivity radius of $\mathcal{M}$ has lower bound*

$i(\mathcal{M}) \geq i_0$ and the volume of $\mathcal{M}$ has upper bound $\mathrm{Vol}(\mathcal{M}) \leq V$. There is a $C(d, k, i_0, V)$ such that if $q \geq C$, for $\mathbf{x} \in \mathcal{M}$

$$\Psi_q(\mathbf{x}) = (\psi_1(\mathbf{x}), \cdots, \psi_q(\mathbf{x})), \tag{2.9}$$

is a smooth embedding of $\mathcal{M}$ into $\mathbb{R}^q$.

Recall that ψ_0 is a constant, so it is not used in the construction of the embedding. Based on the above theorem, let ℓ be the smallest integer q such that $\Psi_q(\mathbf{x})$ is an embedding and we define

$$\Psi(\mathbf{x}) = (\psi_1(\mathbf{x}), \cdots, \psi_\ell(\mathbf{x})). \tag{2.10}$$

Hence, we have $d \leq \ell \leq C(d, k, i_0, V)$. In other words, ℓ can be bounded above by the dimension, Ricci curvature lower bound, the injectivity radius lower bound and the volume upper bound of the manifold $\mathcal{M}$.

Based on Definition 2.4 and Theorem 2.5, we provide the following definition of the reaction coordinates which we use in this work.

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

$$\Psi(\mathbf{x}) = (\psi_1(\mathbf{x}), \cdots, \psi_\ell(\mathbf{x})), \tag{2.11}$$

is a smooth embedding of $\mathcal{M}$ into $\mathbb{R}^\ell$. Let A be a $\ell \times \ell$ diagonal matrix such that $A_{jj} = a_j \|v_{j,n,\epsilon}\|_{l^2(1/\hat{\rho^*})}$ where $\|v_{j,n,\epsilon}\|_{l^2(1/\hat{\rho^*})}$ is defined in Definition 2.4 and a_j is defined in Theorem 2.5. Under Assumption 2.1, we define

$$\mathbf{y}_i = \Phi(\mathbf{x}_i) := A \circ \Psi(\mathbf{x}_i) \in \mathbb{R}^\ell, \tag{2.12}$$

to be the reaction coordinates of $\mathbf{x}_i$

Note that $A \circ \Psi$ is also a smooth embedding of $\mathcal{M}$ into $\mathbb{R}^\ell$. Hence, by Theorem 2.5, we have a justification of the diffusion map. Let $\{v_{1,n,\epsilon}, \cdots, v_{\ell,n,\epsilon}\}$ be the first ℓ eigenvectors of $\frac{I - L_{\epsilon,1}}{\epsilon^2}$ in Step (iv) of the algorithm of the diffusion map. Then, the diffusion map

$$\mathbf{x}_i \to (v_{1,n,\epsilon}(i), \cdots, v_{\ell,n,\epsilon}(i)), \tag{2.13}$$

is an approximation of $\mathbf{y}_i = \Phi(\mathbf{x}_i) := A \circ \Psi(\mathbf{x}_i)$ over the data points $\{\mathbf{x}_1 \cdots, \mathbf{x}_n\}$.

Although the diffusion map is applied to construct the reaction coordinates in this work, it also can be used to solve the Fokker-Planck equations on $\mathcal{M}$. Under Assumption 2.1, for $f \in C^3(\mathcal{M})$, it is shown in [15] that $\frac{I - L_{\epsilon,\alpha}}{\epsilon^2}$ converges pointwisely (in the sense of Theorem 2.3) to the Kolmogrov backward operator

$$\mathcal{L}_\alpha f = -\Delta f + 2(1 - \alpha)\nabla U \cdot \nabla f,$$

where $U = -\log \rho^*$ and ρ^* is the unknown sample density defined in Assumption 2.1. Hence, the eigenpairs of $\frac{I - L_{\epsilon,\alpha}}{\epsilon^2}$ approximate the corresponding eigenpairs of $\mathcal{L}_\alpha$. When $\alpha = 1/2$, let $\mathcal{L} = \mathcal{L}_{1/2}$. Let

$$\mathcal{L}^* f = -\Delta f - \nabla \cdot (f \nabla U)$$

be the Kolmogrov forward (Fokker-Planck) operator. Then, $\mathcal{L}$ and $\mathcal{L}^*$ share the same eigenvalues and their eigenfunctions differ by a factor ρ^*. The solution to the Fokker-Planck equation

$$\partial \rho_t = -\mathcal{L}^* \rho_t$$

can be expressed as a series sum in terms of the eigenpairs of the $\mathcal{L}^*$. The coefficients in the series are determined by the projection of the initial condition onto each eigenspace. Suppose the unknown sample density ρ^* is approximated through a density estimation. Then, the eigenpairs of $\mathcal{L}^*$ over the sample points can be approximated by using the eigenpairs $\frac{I-L_{\epsilon,1/2}}{\epsilon^2}$. In [8,9], the authors construct an approximation of the solution to the Fokker-Planck equation by using the spectral method and they explore the dynamics on the manifold. The solution is constructed by projecting the discretization of the initial condition onto the approximation of the finite dimensional eigenspaces of $\mathcal{L}^*$. It is worth mentioning that the setup and methods applied in our work are different from [15,8,9]. First, we will use the diffusion map to find the reaction coordinates and reconstruct a manifold $\mathcal{N}$ in a low dimensional space. As we describe in the introduction, instead of solving the Fokker-Planck equation on $\mathcal{M}$ in the high dimensional space, we will solve the Fokker-Planck equation on $\mathcal{N}$. Second, we assume that the equilibrium potential $U_{\mathcal{N}}$ in the Fokker-Planck equation on $\mathcal{N}$ is equal to $-\log \rho_\infty^{\mathcal{N}}$, where $\rho_\infty^{\mathcal{N}}$ is a known equilibrium density. However, it is not necessary that $\rho_\infty^{\mathcal{N}}$ is equal to the sample density on $\mathcal{N}$. At last, we will propose a finite volume scheme rather than apply the spectral method to solve the Fokker-Planck equation.

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, given an equilibrium potential $U_{\mathcal{N}}(\mathbf{y})$, we will focus on designing a data-driven solver for the Fokker-Planck equation on $\mathcal{N}$ which drives any initial data ρ_0 to the equilibrium density on $\mathcal{N}$, $\rho_\infty^{\mathcal{N}}(\mathbf{y}) \propto e^{-U_{\mathcal{N}}(\mathbf{y})}$ (after taking $kT = 1$). To study the trajectory of ρ_t driving any initial data ρ_0 to the equilibrium density $\rho_\infty^{\mathcal{N}}(\mathbf{y})$, it is sufficient to solve the following Fokker-Planck equation on manifold $\mathcal{N}$

$$\partial_t \rho_t^{\mathcal{N}} = \mathrm{div}_{\mathcal{N}}(\nabla_{\mathcal{N}} \rho_t^{\mathcal{N}} + \rho_t^{\mathcal{N}} \nabla_{\mathcal{N}} U_{\mathcal{N}}). \tag{3.1}$$

For notational simplicity, in the remainder of this section, we will denote the equilibrium density for the Fokker-Planck equation (1.6) as $\pi(\mathbf{y}) := \rho_\infty^{\mathcal{N}}(\mathbf{y})$.

As mentioned before, since we do not have exact information of $\mathcal{N}$, the only implementable method is to use the data $\{\mathbf{y}_i\}_{i=1}^n \subset \mathcal{N}$ to construct directly a good discrete approximation to the Fokker-Planck equation (3.1). We know that $\mathbf{y}_i = \Phi(\mathbf{x}_i)$, where Φ is the reaction coordinates defined in Definition 2.8 and $\{\mathbf{x}_i\}_{i=1}^n$ are the samples on $\mathcal{M}$ based on the density function ρ^* in Assumption 2.1. Hence, $\{\mathbf{y}_i\}_{i=1}^n$ are the samples on $\mathcal{N}$ based on a density function ρ^{**}, where ρ^{**} is the induced density function of ρ^* by the reaction coordinates Φ. Since ρ^* has an upper bound and a positive lower bound, ρ^{**} also has an upper bound and a positive lower bound. It can be proved that $\{\mathbf{y}_i\}_{i=1}^n$ are well-distributed on $\mathcal{N}$ when they are sampled based on a density function with such bounds [55,41].

In Section 3.1, we will construct a Voronoi tessellation for $\mathcal{N}$ from $\{\mathbf{y}_i\}_{i=1}^n \subset \mathcal{N}$ and then assign the transition probability for an approximated Markov process on $\{\mathbf{y}_i\}_{i=1}^n$ between the nearest neighbor points. This transition probability with detailed balance property also gives a finite volume scheme for solving the 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 $\mathcal{N}$, to propose an implementable scheme, the Voronoi tessellation needs to be further approximated. In Section 3.4, thanks to the metric on $\mathcal{N}$ induced by the low dimensional Euclidean distance in $\mathbb{R}^\ell$, the volumes of the Voronoi cells and the areas of the Voronoi faces 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 finite volume scheme for Fokker-Planck equation (3.1); see Theorem 3.14. In Section 3.5, we provide an unconditionally

stable explicit time discretization for the finite volume scheme based on the detailed balance property of the Markov process, which satisfies a mass conservation law and exponentially converges to the equilibrium. As a consequence, we obtained an approximated Markov process on $\{\mathbf{y}_i\}_{i=1}^n$, i.e., random walk between the nearest neighbor points with an approximated transition probability and jump rate that enjoys good properties such as the conservation of mass, L^1 contraction for the forward equation, L^∞ maximal principle for the backward equation and the L^2 error estimates.

3.1. Construction of the Voronoi tessellation and the finite volume scheme on manifold $\mathcal{N}$

In this section, we construct a finite volume scheme based on the Voronoi tessellation for manifold $\mathcal{N}$. We will see the advantage is that the Voronoi tessellation automatically gives a positivity-preserving finite volume scheme for the Markov process with detailed balance; see Lemma 3.4.

Suppose $(\mathcal{N}, d_{\mathcal{N}})$ is a d dimensional smooth closed submanifold of $\mathbb{R}^\ell$ and $d_{\mathcal{N}}$ is induced by the Euclidean metric in $\mathbb{R}^\ell$. Let $S \subset \mathcal{N}$. We have

$$\mathcal{H}_\delta^k(S) = \inf\{\sum_{i=1}^{\infty} \operatorname{diam}(U_i)^k, S \subset \cup_{i=1}^\infty U_i, \operatorname{diam}(U_i) < \delta\}, \tag{3.2}$$

where the infimum is taken over all countable covers of S in $\mathcal{N}$ and the diameter of the set U_i is measured in metric $d_{\mathcal{N}}$. Then, the k dimensional Hausdorff measure $\mathcal{H}^k(S)$ of S in $\mathcal{N}$ is defined as

$$\mathcal{H}^k(S) = \lim_{\delta \to 0} \mathcal{H}_\delta^k(S). \tag{3.3}$$

For the samples $\{\mathbf{y}_i\}_{i=1}^n \subset \mathcal{N}$, we define the Voronoi cell as

$$C_i := \{\mathbf{y} \in \mathcal{N};\ \mathrm{d}_{\mathcal{N}}(\mathbf{y}, \mathbf{y}_i) \le \mathrm{d}_{\mathcal{N}}(\mathbf{y}, \mathbf{y}_j) \text{ for all } \mathbf{y}_j,\ j = 1, \cdots, n\}, \tag{3.4}$$

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

$$\Gamma_{ij} := C_i \cap C_j, \text{ and its area as } |\Gamma_{ij}| = \mathcal{H}^{d-1}(\Gamma_{ij}) \tag{3.5}$$

for any $j = 1, \cdots, n$. If $\Gamma_{ij} = \emptyset$ or $i = j$ then we set $|\Gamma_{ij}| = 0$. We define the bisector between $\mathbf{y}_i$ and $\mathbf{y}_j$ to be the set

$$G_{ij} := \{\mathbf{y} \in \mathcal{N};\ \mathrm{d}_{\mathcal{N}}(\mathbf{y}, \mathbf{y}_i) = \mathrm{d}_{\mathcal{N}}(\mathbf{y}, \mathbf{y}_j)\}. \tag{3.6}$$

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

Define the associated adjacent sample points as

$$\mathrm{VF}(i) := \{j;\ \Gamma_{ij} \neq \emptyset\}. \tag{3.7}$$

First, we have the following basic facts about the Voronoi cells on a manifold.

Proposition 3.1. *If C_i is the Voronoi cell containing the $\mathbf{y}_i$ and C_i is contained in the geodesic ball centered at $\mathbf{y}_i$ whose radius is equal to the injectivity radius of $\mathcal{N}$ at $\mathbf{y}_i$, then C_i is star shaped.*

Proof. For any $\mathbf{y} \in C_i$, if there is a point $\mathbf{y}'$ on the minimizing geodesic between $\mathbf{y}$ and $\mathbf{y}_i$ such that $\mathbf{y}' \notin C_i$ and $\mathbf{y}' \in C_j$, then $d_{\mathcal{N}}(\mathbf{y}', \mathbf{y}_j) < d_{\mathcal{N}}(\mathbf{y}', \mathbf{y}_i)$. Therefore, we have $d_{\mathcal{N}}(\mathbf{y}, \mathbf{y}_j) \le d_{\mathcal{N}}(\mathbf{y}, \mathbf{y}') + d_{\mathcal{N}}(\mathbf{y}', \mathbf{y}_j) < d_{\mathcal{N}}(\mathbf{y}, \mathbf{y}') + d_{\mathcal{N}}(\mathbf{y}', \mathbf{y}_i) = d_{\mathcal{N}}(\mathbf{y}, \mathbf{y}_i)$. This contradicts to $\mathbf{y} \in C_i$. Hence, any point on the minimizing geodesic between $\mathbf{y}$ and $\mathbf{y}_i$ is in C_i. □

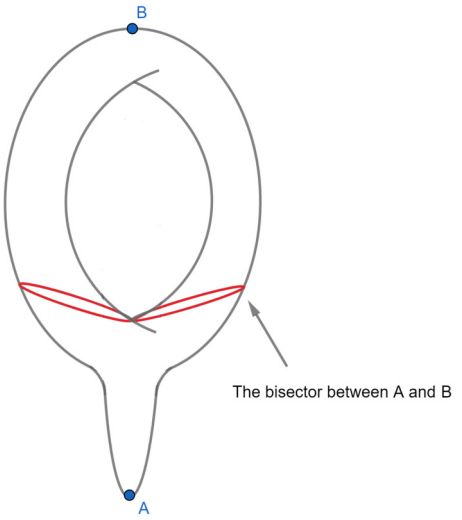


Fig. 1. An example in which the bisector between two points is not a submanifold.

Note that the above fact holds regardless how $\{\mathbf{y}_i\}_{i=1}^n$ are sampled.

Next, we want to discuss the geometric properties of the Voronoi faces. We start from the bisector between two points. A natural question is whether a bisector between two points on a closed d dimensional manifold is a $d-1$ dimensional submanifold. Unfortunately, the answer is negative for two arbitrary points due to the topological and geometrical structure of the manifold. In Fig. 1, we show an example that on a manifold diffeomorphic to a torus, the bisector between two points A and B is a figure "8" curve. There are special cases when the bisector between any two points is a submanifold globally. In [5], the author proves that any bisector between two points is a totally geodesic submanifold if and only if the manifold has constant curvature. An obvious example of this case is the round sphere, where any bisector is a great hypersphere. Hence, on the round sphere, any Voronoi surface is a part of a great hypersphere.

In this work, we prove the following local regularity result for the bisectors on any manifold. In fact, we show that when two points are close enough, then an open neighborhood on the bisector around the midpoint of the minimizing geodesic connecting those two points is a $d-1$ dimensional submanifold. The proof of the proposition with a figure to illustrate the statement of the proposition is in Appendix B.

Proposition 3.2. *Suppose δ is small enough depending on the bounds of the sectional curvatures and the injectivity radius of $\mathcal{N}$. For any $\mathbf{y}_i \in \mathcal{N}$, let $B_\delta(\mathbf{y}_i)$ be an open geodesic ball of radius δ at $\mathbf{y}_i$. Suppose $\mathbf{y}_j \in B_\delta(\mathbf{y}_i)$ and G_{ij} is the bisector between $\mathbf{y}_i$ and $\mathbf{y}_j$. Then, $M_{ij} = B_\delta(\mathbf{y}_i) \cap G_{ij}$ is a $d-1$ dimensional submanifold. Let $\mathbf{y}^*$ be the midpoint of the minimizing geodesic between $\mathbf{y}_i$ and $\mathbf{y}_j$. Then $\mathbf{y}^* \in M_{ij}$ and the minimizing geodesic between $\mathbf{y}_i$ and $\mathbf{y}_j$ is perpendicular to M_{ij} at $\mathbf{y}^*$.*

Since $\{\mathbf{y}_i\}_{i=1}^n$ are sampled based on a density function ρ^{**} with a positive lower bound, when n is large enough, with high probability, there are enough points in a small geodesic ball. Hence, we can assume that each Voronoi cell is small enough so that it is contained in a geodesic ball. We propose the following assumption.

Assumption 3.3. For δ defined in Proposition 3.2, let $B_{\frac{\delta}{2}}(\mathbf{y}_i)$ be an open geodesic ball centered at y_i with radius $\frac{\delta}{2}$. We assume that when n is large enough, we have $C_i \subset B_{\frac{\delta}{2}}(\mathbf{y}_i)$ for $i = 1, \cdots, n$.

Suppose Γ_{ij} is the Voronoi face between $\mathbf{y}_i$ and $\mathbf{y}_j$, Assumption 3.3 implies that $\mathbf{y}_j \in B_\delta(y_i)$. Based on Assumption 3.3 and Proposition 3.2, the interior of the Voronoi face Γ_{ij} is an open subset of a submanifold

M_{ij}. Hence, there is a well defined unit outward normal vector field $\mathbf{n}$ on each ∂C_i and we can apply the divergence Theorem on each Voronoi cell.

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

$$\partial_t \rho_t^{\mathcal{N}} = \mathrm{div}_{\mathcal{N}} \left(\pi \nabla_{\mathcal{N}} \left(\frac{\rho_t^{\mathcal{N}}}{\pi} \right) \right). \tag{3.8}$$

We drop the dependence $t, \mathcal{N}$ in the short hand notation $\rho = \rho_t^{\mathcal{N}}$. We integrate this on C_i and use the divergence theorem on cell C_i to obtain

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{C_i} \rho \mathcal{H}^d(C_i) = \sum_{j \in \mathrm{VF}(i)} \int_{\Gamma_{ij}} \pi \mathbf{n} \cdot \nabla_{\mathcal{N}} \left(\frac{\rho}{\pi} \right) \mathcal{H}^{d-1}(\Gamma_{ij}), \tag{3.9}$$

where $\mathbf{n}$ is the unit outward normal vector field on ∂C_i.

To design the numerical algorithm, first, let us clarify the probability on each cell. Then the probability in C_i can be approximated as

$$\int_{C_i} \rho(\mathbf{y}) \mathcal{H}^d(C_i) \approx \rho(\mathbf{y}_i)(1 + \mathrm{diam}(C_i))|C_i|. \tag{3.10}$$

Second, we use ρ_i to approximate the exact solution $\rho(\mathbf{y}_i)$ on each cell C_i. Let π_i be the approximated equilibrium density at $\mathbf{y}_i$ satisfying $\sum_{i=1}^n \pi_i |C_i| = 1$. Notice $\rho_\infty(\mathbf{y}) \propto e^{-U_{\mathcal{N}}(\mathbf{y})}$, so $\pi_i > 0$ for all i. Then the surface gradient in (3.9) can be approximated by

$$\sum_{j \in \mathrm{VF}(i)} \int_{\Gamma_{ij}} \pi \mathbf{n} \cdot \nabla_{\mathcal{N}} \left(\frac{\rho}{\pi} \right) \mathcal{H}^{d-1}(\Gamma_{ij}) \approx \frac{1}{2} \sum_{j \in \mathrm{VF}(i)} \frac{\pi_i + \pi_j}{|y_i - y_j|} |\Gamma_{ij}| \left(\frac{\rho_j}{\pi_j} - \frac{\rho_i}{\pi_i} \right). \tag{3.11}$$

Therefore, combining (3.10) and (3.11), we give the following finite volume scheme. For $i = 1, \cdots, n$,

$$\frac{\mathrm{d}}{\mathrm{d}t} \rho_i |C_i| = \frac{1}{2} \sum_{j \in \mathrm{VF}(i)} \frac{\pi_i + \pi_j}{|y_i - y_j|} |\Gamma_{ij}| \left(\frac{\rho_j}{\pi_j} - \frac{\rho_i}{\pi_i} \right). \tag{3.12}$$

Let χ_{C_i} be the characteristic function such that $\chi_{C_i} = 1$ for $\mathbf{y} \in C_i$ and 0 otherwise. Then

$$\rho^{\mathrm{approx}}(\mathbf{y}) := \sum_{i=1}^n \rho_i \chi_{C_i}(\mathbf{y})$$

is the piecewise constant probability distribution on $\mathcal{N}$ provided $\sum_{i=1}^n \rho_i |C_i| = 1$ and $\rho_i \geq 0$. We will prove later in the convergence analysis Theorem 3.7 that the exact solution ρ can be approximated by the numerical piecewise constant probability distribution constructed from $\rho_i, i = 1, \cdots, n$.

We will first formulate finite volume scheme (3.12) 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 (3.12) later in Section 3.3.

3.2. Associated Markov process, detailed balance and ergodicity

We will first formulate finite volume scheme (3.12) 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 $\frac{\rho_i(t)}{\pi_i}$.

Roughly speaking, the forward equation leads to the conservation law while the backward equation leads to maximum norm estimates for $\frac{\rho_i}{\pi_i}$.

Lemma 3.4. *Let $\pi_i > 0$ for all $i = 1, \cdots, n$. The finite volume scheme (3.12) is the forward equation for a Markov Process with transition probability P_{ji} (from state j to i) and jump rate η_j*

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho_i|C_i| = \Big(\sum_{j\in VF(i)} \eta_j P_{ji}\rho_j|C_j|\Big) - \eta_i\rho_i|C_i|, \tag{3.13}$$

where

$$\begin{aligned} \eta_i = \sum_{j\neq i} Q_{ij} =: \frac{1}{2|C_i|\pi_i}\sum_{j\in VF(i)} \frac{\pi_i+\pi_j}{|y_i-y_j|}|\Gamma_{ij}| > 0, \quad i = 1, 2, \cdots, n; \\ P_{ij} := \begin{cases} \dfrac{Q_{ij}}{\eta_i} = \dfrac{1}{\eta_i}\dfrac{\pi_i+\pi_j}{2\pi_i|C_i|}\dfrac{|\Gamma_{ij}|}{|y_i-y_j|}, & j \in VF(i); \\ 0, & j \notin VF(i). \end{cases} \end{aligned} \tag{3.14}$$

(i) P is transition probability matrix satisfying $\sum_j P_{ij} = 1$ and the detailed balance property

$$P_{ji}\eta_j\pi_j|C_j| = P_{ij}\eta_i\pi_i|C_i| = \frac{\pi_i+\pi_j}{2}\frac{|\Gamma_{ij}|}{|y_i-y_j|} \quad \forall i,\, j. \tag{3.15}$$

(ii) With $\{w_i\}_{i=1}^n := \{\rho_i|C_i|\}_{i=1}^n$, we recast the forward equation (3.13) as

$$\frac{\mathrm{d}}{\mathrm{d}t}w = Q^*w, \tag{3.16}$$

where Q^* is the transpose of Q-matrix defined as

$$Q = (a_{ij})_{n\times n}, \quad a_{ij} := \begin{cases} -\eta_i, & j = i; \\ \eta_i P_{ij}, & j \neq i. \end{cases} \tag{3.17}$$

(iii) $\sum_{i=j}^n a_{ij} = 0$, which gives the conservation law for $\sum_i w_i$

$$\frac{\mathrm{d}}{\mathrm{d}t}\sum_{i=1}^n w_i = \sum_{i=1}^n\sum_{j=1}^n a_{ji}w_j = 0; \tag{3.18}$$

(iv) We have the dissipation relation for χ^2-divergence

$$\frac{\mathrm{d}}{\mathrm{d}t}\sum_i \frac{\rho_i^2}{\pi_i}|C_i| = -\sum_{i,j}\frac{\pi_i+\pi_j}{2}\frac{|\Gamma_{ij}|}{|y_i-y_j|}\left(\frac{\rho_j}{\pi_j}-\frac{\rho_i}{\pi_i}\right)^2. \tag{3.19}$$

Proof. First, one can rewrite (3.12) as (3.13) with $\eta_i = \frac{1}{2|C_i|\pi_i}\sum_{j\in \mathrm{VF}(i)}\frac{\pi_i+\pi_j}{|y_i-y_j|}|\Gamma_{ij}|$ and $P_{ji}\eta_j = \frac{\pi_i+\pi_j}{2|y_i-y_j|}\frac{|\Gamma_{ij}|}{\pi_j|C_j|}$. Then since $\frac{\pi_i+\pi_j}{|y_i-y_j|}|\Gamma_{ij}|$ is symmetric, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big(\sum_{i=1}^n |C_i|\rho_i\Big) = \sum_{i,j}\frac{1}{2}\frac{\pi_i+\pi_j}{|y_i-y_j|}|\Gamma_{ij}|\left(\frac{\rho_j}{\pi_j}-\frac{\rho_i}{\pi_i}\right) = 0. \tag{3.20}$$

Second we can check

$$\sum_i P_{ji} = \sum_{i\in \mathrm{VF}(j)} P_{ji} = \frac{1}{\eta_j} \sum_{i\in \mathrm{VF}(j)} \frac{\pi_i+\pi_j}{2|y_i-y_j|} \frac{|\Gamma_{ij}|}{\pi_j |C_j|} = 1. \tag{3.21}$$

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

$$\eta_j P_{ji} \pi_j |C_j| = \frac{\pi_i+\pi_j}{2|y_i-y_j|}|\Gamma_{ij}| = \eta_i P_{ij} \pi_i |C_i|. \tag{3.22}$$

Next, the conservation law follows directly from $\sum_{i=1}^n a_{ji} = 0$ by (3.17) and (3.21). Denote the diagonal rate matrix as $R = \mathrm{diag}(\eta_j)$, then we obtain Q-matrix $Q = R(P - I)$.

Finally, by detailed balance property (3.15) and $\sum_j P_{ij} = 1$, we recast (3.13) as

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\rho_i |C_i| &= \sum_{j\in \mathrm{VF}(i)} \eta_i P_{ij} \pi_i |C_i| \frac{\rho_j}{\pi_j} - \eta_i \pi_i |C_i| \frac{\rho_i}{\pi_i} \\ &= \sum_{j\in \mathrm{VF}(i)} \eta_i \pi_i |C_i| P_{ij} \left(\frac{\rho_j}{\pi_j} - \frac{\rho_i}{\pi_i} \right) = \sum_{j\in \mathrm{VF}(i)} \frac{\pi_i+\pi_j}{2|y_i-y_j|} |\Gamma_{ij}| \left(\frac{\rho_j}{\pi_j} - \frac{\rho_i}{\pi_i} \right) \end{aligned} \tag{3.23}$$

Multiplying this by $\frac{\rho_i}{\pi_i}$ and show that

$$\frac{\mathrm{d}}{\mathrm{d}t} \sum_i \frac{\rho_i^2}{\pi_i} |C_i| = -\sum_{i,j} \frac{\pi_i+\pi_j}{2} \frac{|\Gamma_{ij}|}{|y_i-y_j|} \left(\frac{\rho_j}{\pi_j} - \frac{\rho_i}{\pi_i} \right)^2. \tag{3.24}$$

This gives the dissipation relation (3.19). □

Proposition 3.5. *Let $\pi_i > 0$ for all $i = 1, \cdots, n$. Let $Q := (a_{ij})_{n\times n}$ be the Q-matrix defined in (3.17). Then $\{u_i\}_{i=1}^n := \{\frac{\rho_i}{\pi_i}\}_{i=1}^n$ is the solution to the backward equation*

$$\frac{\mathrm{d}}{\mathrm{d}t} u = Qu. \tag{3.25}$$

Moreover, let η_i be the jump rate defined in (3.14). We conclude 0 is the simple, principle eigenvalue of Q with the ground state $\{1, 1, \cdots, 1\}$. We thus have the exponential decay of $\rho_i(t)$ with respect to time t,

$$\max_i \frac{|\rho_i(t) - \pi_i|}{\pi_i} \le c e^{-(\lambda_1 - |\lambda_2|)t}, \tag{3.26}$$

where $\lambda_1 > \max_i \eta_i$ is the principle eigenvalue of $\lambda I + Q$, $|\lambda_2| < \lambda_1$ is the second largest eigenvalue of $\lambda_1 I + Q$, and $\lambda_1 - |\lambda_2| > 0$ is the spectral gap of $\lambda_1 I + Q$.

Proof. Recall (3.23). We recast the forward equation (3.13) as

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho_i |C_i| = \sum_{j\in \mathrm{VF}(i)} \eta_i P_{ij} \pi_i |C_i| \frac{\rho_j}{\pi_j} - \eta_i \pi_i |C_i| \frac{\rho_i}{\pi_i}, \tag{3.27}$$

which gives

$$\frac{\mathrm{d}}{\mathrm{d}t} \frac{\rho_i}{\pi_i} = \sum_{j\in \mathrm{VF}(i)} \eta_i P_{ij} \frac{\rho_j}{\pi_j} - \eta_i \frac{\rho_i}{\pi_i}. \tag{3.28}$$

Next, we show $\{u_i\}_{i=1}^n := \{\frac{\rho_i}{\pi_i}\}_{i=1}^n$ is the solution to this backward equation. Recast (3.28) as

$$\frac{\mathrm{d}}{\mathrm{d}t}u_i = \sum_{j=1}^n a_{ij}u_j. \tag{3.29}$$

Here $Q = \{a_{ij}\}$ is the generator of the backward equation

$$\frac{\mathrm{d}}{\mathrm{d}t}u = Qu. \tag{3.30}$$

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 (3.15), we know Q is self-adjoint in the weighted l^2-space

$$\langle u, Qv\rangle_{\pi|C|} := \sum_{i,j} u_i a_{ij} v_j \pi_i |C_i| = \sum_{i,j} a_{ji} u_i v_j \pi_j |C_j| =: \langle Qu, v\rangle_{\pi|C|}. \tag{3.31}$$

Thus we know the eigenvalues of Q are real. For the matrix $\lambda_1 I + Q$ with $\lambda_1 > \max_i \eta_i$, we know each element is non-negative and $\lambda_1 + \sum_{j=1}^n a_{ij} = \lambda_1 > 0$. Since the Voronoi tessellation $\mathcal{N} = \cup_{i=1}^n C_i$, when $\mathcal{N}$ is strongly connected, $\lambda_1 I + Q$ is irreducible. By the Perron-Frobenius theorem for $\lambda_1 I + Q$, we know the Perron–Frobenius eigenvalue (i.e. the principal eigenvalue) of $\lambda_1 I + Q$ is λ_1 and $\lambda_1 > 0$ is a simple eigenvalue with the ground state $u^* := (1, 1, \cdots, 1)$ and other eigenvalues λ_i of $\lambda_1 I + Q$ satisfy $|\lambda_i| < \lambda_1$. Therefore, we have

$$\lambda_1(u - u^*) + \frac{\mathrm{d}}{\mathrm{d}t}(u - u^*) = (\lambda_1 I + Q)(u - u^*). \tag{3.32}$$

Let $\|\cdot\|$ be the weighted l^2 norm defined as $\|v\|^2 := \sum_i v_i^2 \pi_i |C_i|$. Taking weighted inner product of (3.32) with $u - u^*$ and from $\langle u - u^*, u^*\rangle_{\pi|C|} = 0$, we have

$$\lambda_1 \|u - u^*\|^2 + \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}(\|u - u^*\|^2) = \langle (\lambda_1 I + Q)(u - u^*),\ u - u^*\rangle_{\pi|C|} \leq |\lambda_2| \|u - u^*\|^2, \tag{3.33}$$

where we used $|\lambda_2| < \lambda_1$ is the second largest eigenvalue of $\lambda_1 I + Q$. This gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\|u - u^*\| \leq (|\lambda_2| - \lambda_1)\|u - u^*\|. \tag{3.34}$$

Therefore we obtain the exponential decay of u to its ground state u^*

$$\|u - u^*\| \leq c e^{(|\lambda_2| - \lambda_1)t}. \tag{3.35}$$

Here $\lambda_1 - |\lambda_2| > 0$ is the spectral gap of $\lambda_1 I + Q$. □

We refer to [39] for the ergodicity of finite volume schemes in an unbounded space. We refer to [42,13,44, 24,25,29] for more discussions on the corresponding generalized gradient flow of the relative entropy with graph Wasserstein distance on discrete space and Benamou-Brenier formula. See also [38,57,30] for some related data-driven algorithms when solving equations on a network graph and for irreversible processes.

3.3. Truncation error estimate, stability and convergence of the finite volume scheme (3.12)

In this section we prove the stability of (3.12) in Lemma 3.6. Then we obtain the convergence of the solution to finite volume scheme (3.12) to the solution of Fokker-Planck equation (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 ρ_i and $\tilde{\rho}_i$ to finite volume scheme (3.13) have the following stability properties*

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\sum_{i=1}^{n}|\rho_i-\tilde{\rho}_i||C_i| \leq 0; \\ \frac{\mathrm{d}}{\mathrm{d}t}\sum_{i=1}^{n}|\dot{\rho}_i||C_i| \leq 0, \end{aligned} \tag{3.36}$$

where $\dot{\rho}_i$ is the time derivative of $\rho_i(t)$.

Proof. First assume ρ_i and $\tilde{\rho}_i$ are two solutions to finite volume scheme (3.13). We have

$$\frac{\mathrm{d}}{\mathrm{d}t}(\rho_i|C_i|-\tilde{\rho}_i|C_i|) = \sum_{j\in \mathrm{VF}(i)} P_{ji}\eta_j|C_j|(\rho_j-\tilde{\rho}_j)-\eta_i|C_i|(\rho_i-\tilde{\rho}_i). \tag{3.37}$$

Notice that for any function u, multiplying u by its sign gives an absolute value $|u|$. From [34, Lemma 7.6], the derivative of $|u|$ equals the derivative of u multiplied by the sign of u, i.e., $D|u| = \mathrm{sgn}(u)Du$. Multiply $\mathrm{sgn}(\rho_i-\tilde{\rho}_i)$ to both sides and then take summation with respect to i

$$\frac{\mathrm{d}}{\mathrm{d}t}\sum_{i=1}^{n}|C_i||\rho_i-\tilde{\rho}_i| \leq \sum_{i,j}P_{ji}\eta_j|C_j||\rho_j-\tilde{\rho}_j|-\sum_{i=1}^{n}\eta_i|C_i||\rho_i-\tilde{\rho}_i| = 0, \tag{3.38}$$

where we used $\sum_{i\in\mathrm{VF}(j)}P_{ji}=1$. Second, take time derivative in (3.13), then we have

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\rho_i|C_i| = \sum_{j\in\mathrm{VF}(i)}P_{ji}\eta_j|C_j|\dot{\rho}_j-\eta_i|C_i|\dot{\rho}_i. \tag{3.39}$$

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

$$\frac{\mathrm{d}}{\mathrm{d}t}\sum_{i=1}^{n}|C_i||\dot{\rho}_i| \leq \sum_{i,j}P_{ji}\eta_j|C_j||\dot{\rho}_j|-\sum_{i=1}^{n}\eta_i|C_i||\dot{\rho}_i| = 0, \tag{3.40}$$

where we used $\sum_{i\in\mathrm{VF}(j)}P_{ji}=1$. □

We conclude this section by the following convergence theorem in the weighted L^2 sense. Although the proof of the convergence theorem is rather standard, cf. [26], however the error estimation on manifold requires some careful treatments for the symmetric cancellation and some approximation lemmas for the Voronoi tessellation.

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

$$h = \max\Big(\max_{i=1,\dots,n}(\mathrm{diam}(C_i)), \max_{i=1,\dots,n}\big(\max_{j\in VF(i)}d_{\mathcal{N}}(\mathbf{y}_i,\mathbf{y}_j)\big)\Big) \tag{3.41}$$

Let $\{\rho_i\}_{i=1}^n$ be the solution to the finite volume scheme (3.12) with initial data $\{\rho_i^0\}_{i=1}^n$ and $e_i := \rho(\mathbf{y}_i) - \rho_i$. Under Assumption 3.3, we have the following error estimate

$$\max_{t\in[0,T]} \sum_i e_i(t)^2 \frac{|C_i|}{\pi_i} \leq \Big(\sum_i e_i(0)^2 \frac{|C_i|}{\pi_i} + O(h^2(nh \max_i |\partial C_i| + 1)) \Big) e^T, \tag{3.42}$$

where the constant in $O(h^2(nh \max_i |\partial C_i| + 1))$ depends on $\mathrm{Vol}(\mathcal{N})$, the minimum of π, the C^1 norm of π, the L^∞ norm of $\partial_t \nabla_{\mathcal{N}} \rho$ and the C^2 norm of $\frac{\rho}{\pi}$.

Proof. Let $\rho_i^e := \frac{1}{|C_i|} \int_{C_i} \rho \, \mathrm{d}y$ be the cell average. Plug the exact solution into the numerical scheme

$$\begin{aligned} \partial_t(\rho_i^e |C_i|) &= \sum_{j\in \mathrm{VF}(i)} \frac{\pi_i + \pi_j}{2|\mathbf{y}_i - \mathbf{y}_j|} |\Gamma_{ij}| \left(\frac{\rho(\mathbf{y}_j)}{\pi_j} - \frac{\rho(\mathbf{y}_i)}{\pi_i} \right) + \sum_{j\in \mathrm{VF}(i)} \varepsilon_{ij}, \\ \varepsilon_{ij} &:= \int_{\Gamma_{ij}} \pi \mathbf{n}_{ij} \cdot \nabla_{\mathcal{N}} \frac{\rho}{\pi} \, \mathrm{d}\mathcal{H}^{d-1} \; - \; \frac{\pi_i + \pi_j}{2|\mathbf{y}_i - \mathbf{y}_j|} |\Gamma_{ij}| \left(\frac{\rho(\mathbf{y}_j)}{\pi_j} - \frac{\rho(\mathbf{y}_i)}{\pi_i} \right), \end{aligned} \tag{3.43}$$

where $\mathbf{n}_{ij}$ is the restriction of the unit outward normal vector field on Γ_{ij}. Exchanging i, j above, we can see that ε_{ij} is anti-symmetric.

Subtracting the numerical scheme (3.12) from (3.43), we have

$$\frac{\mathrm{d}}{\mathrm{d}t}(e_i|C_i|) = \sum_{j\in \mathrm{VF}(i)} \frac{\pi_i + \pi_j}{2|\mathbf{y}_i - \mathbf{y}_j|} |\Gamma_{ij}| \left(\frac{e_j}{\pi_j} - \frac{e_i}{\pi_i} \right) + \sum_{j\in \mathrm{VF}(i)} \varepsilon_{ij} + \partial_t((\rho(\mathbf{y}_i) - \rho_i^e)|C_i|). \tag{3.44}$$

Similar to the dissipation relation (3.19), multiplying $\frac{e_i}{\pi_i}$ shows that

$$\frac{\mathrm{d}}{\mathrm{d}t} \sum_i e_i^2 \frac{|C_i|}{\pi_i} = -\sum_i \sum_{j\in \mathrm{VF}(i)} \frac{\pi_i + \pi_j}{2|\mathbf{y}_i - \mathbf{y}_j|} |\Gamma_{ij}| \left(\frac{e_j}{\pi_j} - \frac{e_i}{\pi_i} \right)^2 + \sum_i \sum_{j\in \mathrm{VF}(i)} 2\varepsilon_{ij} \frac{e_i}{\pi_i} + \sum_i 2\partial_t(\rho(\mathbf{y}_i) - \rho_i^e) e_i \frac{|C_i|}{\pi_i}. \tag{3.45}$$

Since ε_{ij} is anti-symmetric,

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} \sum_i e_i^2 \frac{|C_i|}{\pi_i} = &-\sum_i \sum_{j\in \mathrm{VF}(i)} \frac{\pi_i + \pi_j}{2|\mathbf{y}_i - \mathbf{y}_j|} |\Gamma_{ij}| \left(\frac{e_j}{\pi_j} - \frac{e_i}{\pi_i} \right)^2 \\ &+ \sum_i \sum_{j\in \mathrm{VF}(i)} \varepsilon_{ij} \left(\frac{e_i}{\pi_i} - \frac{e_j}{\pi_j} \right) + \sum_i 2\partial_t(\rho(\mathbf{y}_i) - \rho_i^e) e_i \frac{|C_i|}{\pi_i}. \end{aligned} \tag{3.46}$$

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

$$\begin{aligned} &\sum_i \sum_{j\in \mathrm{VF}(i)} \varepsilon_{ij} \left(\frac{e_i}{\pi_i} - \frac{e_j}{\pi_j} \right) \\ &\leq \frac{1}{2} \sum_i \sum_{j\in \mathrm{VF}(i)} \frac{\pi_i + \pi_j}{2|\mathbf{y}_i - \mathbf{y}_j|} |\Gamma_{ij}| \left(\frac{e_j}{\pi_j} - \frac{e_i}{\pi_i} \right)^2 + \frac{1}{2} \sum_i \sum_{j\in \mathrm{VF}(i)} \frac{\varepsilon_{ij}^2}{\frac{\pi_i + \pi_j}{2|\mathbf{y}_i - \mathbf{y}_j|} |\Gamma_{ij}|}; \end{aligned} \tag{3.47}$$

$$\sum_i 2\partial_t(\rho(\mathbf{y}_i) - \rho_i^e) e_i \frac{|C_i|}{\pi_i} \leq \sum_i [\partial_t(\rho(\mathbf{y}_i) - \rho_i^e)]^2 \frac{|C_i|}{\pi_i} + \sum_i e_i^2 \frac{|C_i|}{\pi_i}. \tag{3.48}$$

Thus we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\sum_i e_i^2\frac{|C_i|}{\pi_i} \leq & -\frac{1}{2}\sum_i\sum_{j\in \mathrm{VF}(i)}\frac{\pi_i+\pi_j}{2|\mathbf{y}_i-\mathbf{y}_j|}|\Gamma_{ij}|\left(\frac{e_j}{\pi_j}-\frac{e_i}{\pi_i}\right)^2+\frac{1}{2}\sum_i\sum_{j\in \mathrm{VF}(i)}\frac{\varepsilon_{ij}^2}{\frac{\pi_i+\pi_j}{2|\mathbf{y}_i-\mathbf{y}_j|}|\Gamma_{ij}|} \\ & +\sum_i[\partial_t(\rho(\mathbf{y}_i)-\rho_i^e)]^2\frac{|C_i|}{\pi_i}+\sum_i e_i^2\frac{|C_i|}{\pi_i} \\ \leq & \sum_i\sum_{j\in \mathrm{VF}(i)}\frac{\varepsilon_{ij}^2}{\frac{\pi_i+\pi_j}{2|\mathbf{y}_i-\mathbf{y}_j|}|\Gamma_{ij}|}+\sum_i[\partial_t(\rho(\mathbf{y}_i)-\rho_i^e)]^2\frac{|C_i|}{\pi_i}+\sum_i e_i^2\frac{|C_i|}{\pi_i}.\end{aligned} \tag{3.49}$$

Next, we bound the term $\sum_i\sum_{j\in \mathrm{VF}(i)}\frac{\varepsilon_{ij}^2}{\frac{\pi_i+\pi_j}{2|\mathbf{y}_i-\mathbf{y}_j|}|\Gamma_{ij}|}$.

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

$$\frac{\rho}{\pi}(\mathbf{y}_j)-\frac{\rho}{\pi}(\mathbf{y}^*)=T\cdot\nabla_{\mathcal{N}}\frac{\rho}{\pi}(\mathbf{y}^*)d_{\mathcal{N}}(\mathbf{y}^*,\mathbf{y}_j)+O(d_{\mathcal{N}}^2(\mathbf{y}^*,\mathbf{y}_j)), \tag{3.50}$$

$$\frac{\rho}{\pi}(\mathbf{y}^*)-\frac{\rho}{\pi}(\mathbf{y}_i)=T\cdot\nabla_{\mathcal{N}}\frac{\rho}{\pi}(\mathbf{y}^*)d_{\mathcal{N}}(\mathbf{y}^*,\mathbf{y}_i)+O(d_{\mathcal{N}}^2(\mathbf{y}^*,\mathbf{y}_i)). \tag{3.51}$$

By Assumption 3.3 and Proposition 3.2, $\mathbf{n}_{ij}$ can be extended to a unit normal vector field on the $d-1$ dimensional submanifold $M_{ij}\subset G_{ij}$. We also call the extension to be $\mathbf{n}_{ij}$. We have $T=\mathbf{n}_{ij}(\mathbf{y}^*)$. Therefore, if we add the above two equations, we have

$$\frac{\rho}{\pi}(\mathbf{y}_j)-\frac{\rho}{\pi}(\mathbf{y}_i)=\mathbf{n}_{ij}\cdot\nabla_{\mathcal{N}}\frac{\rho}{\pi}(\mathbf{y}^*)d_{\mathcal{N}}(\mathbf{y}_i,\mathbf{y}_j)+O(d_{\mathcal{N}}^2(\mathbf{y}_i,\mathbf{y}_j)). \tag{3.52}$$

Hence,

$$\mathbf{n}_{ij}\cdot\nabla_{\mathcal{N}}\frac{\rho}{\pi}(\mathbf{y}^*)=\frac{\frac{\rho(\mathbf{y}_j)}{\pi_j}-\frac{\rho(\mathbf{y}_i)}{\pi_i}}{d_{\mathcal{N}}(\mathbf{y}_i,\mathbf{y}_j)}+O(d_{\mathcal{N}}(\mathbf{y}_i,\mathbf{y}_j))=\frac{\frac{\rho(\mathbf{y}_j)}{\pi_j}-\frac{\rho(\mathbf{y}_i)}{\pi_i}}{|\mathbf{y}_i-\mathbf{y}_j|}+O(d_{\mathcal{N}}(\mathbf{y}_i,\mathbf{y}_j)), \tag{3.53}$$

where we apply Lemma 3.9 in the last step. Similarly,

$$\pi(\mathbf{y}_j)-\pi(\mathbf{y}^*)=O(d_{\mathcal{N}}(\mathbf{y}^*,\mathbf{y}_j)), \tag{3.54}$$

$$\pi(\mathbf{y}^*)-\pi(\mathbf{y}_i)=O(d_{\mathcal{N}}(\mathbf{y}^*,\mathbf{y}_i)). \tag{3.55}$$

Hence,

$$\pi(\mathbf{y}^*)=\frac{\pi_i+\pi_j}{2}+O(d_{\mathcal{N}}(\mathbf{y}_i,\mathbf{y}_j)). \tag{3.56}$$

Therefore,

$$\pi(\mathbf{y}^*)\mathbf{n}_{ij}\cdot\nabla_{\mathcal{N}}\frac{\rho}{\pi}(\mathbf{y}^*)=\frac{\pi_i+\pi_j}{2}\frac{\frac{\rho(\mathbf{y}_j)}{\pi_j}-\frac{\rho(\mathbf{y}_i)}{\pi_i}}{|\mathbf{y}_i-\mathbf{y}_j|}+O(d_{\mathcal{N}}(\mathbf{y}_i,\mathbf{y}_j)). \tag{3.57}$$

For any $\mathbf{y}$ on Γ_{ij},

$$\begin{aligned}\pi(\mathbf{y})\mathbf{n}_{ij}\cdot\nabla_{\mathcal{N}}\frac{\rho}{\pi}(\mathbf{y}) =& \pi(\mathbf{y}^*)\mathbf{n}\cdot\nabla_{\mathcal{N}}\frac{\rho}{\pi}(\mathbf{y}^*)+O(d_{\mathcal{N}}(\mathbf{y},\mathbf{y}^*)) \qquad (3.58)\\ =& \pi(\mathbf{y}^*)\mathbf{n}\cdot\nabla_{\mathcal{N}}\frac{\rho}{\pi}(\mathbf{y}^*)+O(d_{\mathcal{N}}(\mathbf{y}_i,\mathbf{y})+d_{\mathcal{N}}(\mathbf{y}_i,\mathbf{y}_j)) \\ =& \pi(\mathbf{y}^*)\mathbf{n}\cdot\nabla_{\mathcal{N}}\frac{\rho}{\pi}(\mathbf{y}^*)+O(\mathrm{diam}(C_i)+d_{\mathcal{N}}(\mathbf{y}_i,\mathbf{y}_j)), \qquad (3.59)\end{aligned}$$

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

$$\pi(\mathbf{y})\mathbf{n}_{ij}\cdot\nabla_{\mathcal{N}}\frac{\rho}{\pi}(\mathbf{y})=\frac{\pi_i+\pi_j}{2}\frac{\frac{\rho(\mathbf{y}_j)}{\pi_j}-\frac{\rho(\mathbf{y}_i)}{\pi_i}}{|\mathbf{y}_i-\mathbf{y}_j|}+O(\operatorname{diam}(C_i)+d_{\mathcal{N}}(\mathbf{y}_i,\mathbf{y}_j)). \tag{3.60}$$

We conclude that

$$\varepsilon_{ij}=O((\operatorname{diam}(C_i)+d_{\mathcal{N}}(\mathbf{y}_i,\mathbf{y}_j))|\Gamma_{ij}|). \tag{3.61}$$

Therefore,

$$\frac{\varepsilon_{ij}^2}{\frac{\pi_i+\pi_j}{2|\mathbf{y}_i-\mathbf{y}_j|}|\Gamma_{ij}|}=O(d_{\mathcal{N}}(\mathbf{y}_i,\mathbf{y}_j)(\operatorname{diam}(C_i)+d_{\mathcal{N}}(\mathbf{y}_i,\mathbf{y}_j))^2|\Gamma_{ij}|). \tag{3.62}$$

If we sum up all $j\in\mathrm{VF}(i)$,

$$\sum_{j\in\mathrm{VF}(i)}\frac{\varepsilon_{ij}^2}{\frac{\pi_i+\pi_j}{2|\mathbf{y}_i-\mathbf{y}_j|}|\Gamma_{ij}|}=\max_{j\in\mathrm{VF}(i)}d_{\mathcal{N}}(\mathbf{y}_i,\mathbf{y}_j)(\operatorname{diam}(C_i)+d_{\mathcal{N}}(\mathbf{y}_i,\mathbf{y}_j))^2O(|\partial C_i|). \tag{3.63}$$

Hence,

$$\sum_i\sum_{j\in\mathrm{VF}(i)}\frac{\varepsilon_{ij}^2}{\frac{\pi_i+\pi_j}{2|\mathbf{y}_i-\mathbf{y}_j|}|\Gamma_{ij}|}=O(nh^3\max_i|\partial C_i|), \tag{3.64}$$

where the constant depends on the minimum of π, the C^1 norm of π and the C^2 norm of $\frac{\rho}{\pi}$.

Next, we bound $\sum_i[\partial_t(\rho(\mathbf{y}_i)-\rho_i^e)]^2\frac{|C_i|}{\pi_i}$. Notice that

$$\partial_t(\rho(\mathbf{y}_i)-\rho_i^e)=O(\operatorname{diam}(C_i))=O(h), \tag{3.65}$$

where the constant depends on the L^∞ norm of $\partial_t\nabla_{\mathcal{N}}\rho$. Since $\sum_i|C_i|=\mathrm{Vol}(\mathcal{N})$,

$$\sum_i[\partial_t(\rho(\mathbf{y}_i)-\rho_i^e)]^2\frac{|C_i|}{\pi_i}=O(h^2), \tag{3.66}$$

where the constant depends on the L^∞ norm of $\partial_t\nabla_{\mathcal{N}}\rho$, $\mathrm{Vol}(\mathcal{N})$ and minimum of π. Hence,

$$\frac{\mathrm{d}}{\mathrm{d}t}\sum_i e_i^2\frac{|C_i|}{\pi_i}\le O(h^2(nh\max_i|\partial C_i|+1))+\sum_i e_i^2\frac{|C_i|}{\pi_i}. \tag{3.67}$$

In conclusion,

$$\max_{t\in[0,T]}\sum_i e_i(t)^2\frac{|C_i|}{\pi_i}\le\Big(\sum_i e_i(0)^2\frac{|C_i|}{\pi_i}+O(h^2(nh\max_i|\partial C_i|+1))\Big)e^T. \quad\square \tag{3.68}$$

3.4. Approximation of Voronoi cells on manifold

Recall that $\{\mathbf{y}_i\}_{i=1}^n$ are samples on the smooth closed submanifold $\mathcal{N}$ in $\mathbb{R}^\ell$ based on the density function ρ^{**}. In this section, we introduce an algorithm to approximate the volumes of the Voronoi cells and the areas of the Voronoi faces constructed from $\{\mathbf{y}_i\}_{i=1}^n$.

First, we need the following definition.

Definition 3.8. For any $0 < r < 1$ and $\mathbf{y}_k \in \{\mathbf{y}_i\}_{i=1}^n$, suppose $B^{\mathbb{R}^\ell}_{\sqrt{r}}(\mathbf{y}_k) \cap \{\mathbf{y}_i\}_{i=1}^n = \{\mathbf{y}_{k,1}, \cdots, \mathbf{y}_{k,\bar{N}_k}\}$. We define the discrete local covariance matrix at $\mathbf{y}_k$,

$$C_{n,r}(\mathbf{y}_k) := \frac{1}{n}\sum_{i=1}^{\bar{N}_k}(\mathbf{y}_{k,i} - \mathbf{y}_k)(\mathbf{y}_{k,i} - \mathbf{y}_k)^\top \in \mathbb{R}^{\ell\times\ell}. \tag{3.69}$$

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

$$\iota_k(u) := (u^\top\beta_{n,r,1}, \cdots, u^\top\beta_{n,r,d}). \tag{3.70}$$

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

Based on the above definition, we propose the following algorithm to find the approximated volumes $|\tilde{C}_k|$ of the Voronoi cells C_k and the approximated areas $|\tilde{\Gamma}_{k\ell}|$ 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 $\mathbf{y}_k \in \{\mathbf{y}_i\}_{i=1}^n$, find

$$B^{\mathbb{R}^\ell}_{\sqrt{r}}(\mathbf{y}_k) \cap \{\mathbf{y}_i\}_{i=1}^n =: \{\mathbf{y}_{k,1}, \cdots, \mathbf{y}_{k,\bar{N}_k}\}, \quad B^{\mathbb{R}^\ell}_{r}(\mathbf{y}_k) \cap \{\mathbf{y}_i\}_{i=1}^n =: \{\mathbf{y}_{k,1}, \cdots, \mathbf{y}_{k,N_k}\}.$$

2 Construct the matrix $C_{n,r}(\mathbf{y}_k)$ as in (3.69) by using the $\{\mathbf{y}_{k,1}, \cdots, \mathbf{y}_{k,\bar{N}_k}\}$. Find the orthonormal eigenvectors corresponding to $C_{n,r}(\mathbf{y}_k)$'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 (3.70). Find $v_{k,i} = \tilde{\iota}_k(\mathbf{y}_{k,i})$, 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

$$|\tilde{C}_k| := |\tilde{C}_{k,0}| := \mathcal{H}^d(\tilde{C}_{k,0}). \tag{3.71}$$

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

$$A_{k\ell} := \frac{\tilde{A}_{k\ell} + \tilde{A}_{\ell k}}{2}, \quad \tilde{A}_{k\ell} = \begin{cases} |\tilde{F}_{k,i}| & \text{if } \mathbf{y}_\ell = \mathbf{y}_{k,i} \in B^{\mathbb{R}^\ell}_r(\mathbf{y}_k)\ ; \\ 0 & \text{otherwise.} \end{cases} \tag{3.72}$$

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 $\mathbf{y}_k$, by using the points in a larger ball $B^{\mathbb{R}^\ell}_{\sqrt{r}}(\mathbf{y}_k)$, we construct the matrix $C_{n,r}(\mathbf{y}_k)$. Then, the first d orthonormal eigenvectors will be an approximation of an orthonormal basis of $T_{\mathbf{y}_k}\mathcal{N}$. Next, we project the points in a smaller ball $B^{\mathbb{R}^\ell}_r(\mathbf{y}_k)$ onto this tangent space approximation. Now the points around $\mathbf{y}_k$ are projected into a d dimensional Euclidean space and $\mathbf{y}_k$ 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 $\mathbf{y}_k$ 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.

Next, we provide a justification of the above algorithm. When the geodesic distance between two points on $\mathcal{N}$ 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 [56].

Lemma 3.9. *Suppose $\mathbf{y}, \mathbf{y}' \in \mathcal{N}$ such that $d_{\mathcal{N}}(y, y')$ is small enough. Then*

$$\|\mathbf{y}' - \mathbf{y}\|_{\mathbb{R}^\ell} = d_{\mathcal{N}}(\mathbf{y}, \mathbf{y}')(1 + O(d^2_{\mathcal{N}}(\mathbf{y}, \mathbf{y}'))), \tag{3.73}$$

where the constant in $O(d^2_{\mathcal{N}}(\mathbf{y}, \mathbf{y}'))$ depending on the second fundamental form of $\mathcal{N}$ in $\mathbb{R}^\ell$ at $\mathbf{y}$.

The above lemma implies that if r is small enough, then for all $\mathbf{y}_k$ and any $\mathbf{y} \in B_r^{\mathbb{R}^\ell}(\mathbf{y}_k) \cap \mathcal{N}$, there is a constant $D_1 > 1$ depending on the second fundamental form of $\mathcal{N}$ in $\mathbb{R}^\ell$, such that

$$d_{\mathcal{N}}(\mathbf{y}, \mathbf{y}_k) \leq D_1 \|\mathbf{y}_k - \mathbf{y}\|_{\mathbb{R}^\ell}. \tag{3.74}$$

We further make the following assumption about the Voronoi cells and the distribution of $\{\mathbf{y}_i\}_{i=1}^n$ on $\mathcal{N}$.

Assumption 3.10. For n large enough, there exists r depending on n such that nr^d is bounded from above and has a positive lower bound for all n and $\frac{nr^{\frac{d}{2}}}{\log n} \to \infty$ as $n \to \infty$. Moreover, when n is large enough, the following conditions about r hold for any $\mathbf{y}_k$:

(1) Suppose $B_r^{\mathbb{R}^\ell}(\mathbf{y}_k) \cap \{\mathbf{y}_i\}_{i=1}^n = \{\mathbf{y}_{k,1}, \cdots, \mathbf{y}_{k,N_k}\}$. We have $C_k \subset B_r^{\mathbb{R}^\ell}(\mathbf{y}_k)$. Moreover, if Γ_{kj} is a Voronoi surface of C_k between $\mathbf{y}_k$ and $\mathbf{y}_j$, then $\mathbf{y}_j \in B_r^{\mathbb{R}^\ell}(\mathbf{y}_k)$. Suppose $\mathbf{y}_j = \mathbf{y}_{k,m}$, 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 $d_{\mathcal{N}}(\mathbf{y}_{k,i}, \mathbf{y}_k) \geq D_2 r$.

Next, we intuitively explain the relation between Assumption 3.10 and Algorithm 1. Recall that $\{\mathbf{y}_i\}_{i=1}^n$ are sampled based on a density function ρ^{**} with a positive lower bound and upper bound. In Algorithm 1, we use the points in a larger ball $B_{\sqrt{r}}^{\mathbb{R}^\ell}(\mathbf{y}_k)$ to approximate the tangent space $T_{\mathbf{y}_k}\mathcal{N}$. Since ρ^{**} has a positive lower bound, the condition $\frac{nr^{\frac{d}{2}}}{\log n} \to \infty$ as $n \to \infty$ implies that the number of points in $B_{\sqrt{r}}^{\mathbb{R}^\ell}(\mathbf{y}_k)$ goes to infinity as n goes to infinity. Hence, we can have a good estimation of the tangent space. The condition that nr^d is bounded from above and has a positive lower bound for all n implies $r \to 0$ as $n \to \infty$. Since ρ^{**} has an upper bound and a positive lower bound, it also implies that we will have enough but not too many points in the smaller ball $B_r^{\mathbb{R}^\ell}(\mathbf{y}_k)$. Hence, (1) and (2) become mild assumption with this relation between r and n. In fact, since $r \to 0$ as $n \to \infty$, (1) says that the Voronoi cell is in a small ball $B_r^{\mathbb{R}^\ell}(\mathbf{y}_k)$. In (2), since there are not too many points in $B_r^{\mathbb{R}^\ell}(\mathbf{y}_k)$, it is reasonable to assume the distance between the points in $B_r^{\mathbb{R}^\ell}(\mathbf{y}_k)$ and $\mathbf{y}_k$ has a lower bound $D_2 r$. With (1) and (2), we can show that the approximation to Voronoi cell in the tangent space is accurate enough for our analysis.

Consider the geodesic ball $B_{\frac{\delta}{2}}(y_k)$ in Assumption 3.3. By Lemma 3.9, when r is small enough, we have $B_r^{\mathbb{R}^\ell}(\mathbf{y}_k) \cap \mathcal{N} \subset B_{\frac{\delta}{2}}(y_i)$. Since $r \to 0$ as $n \to \infty$, we know that when n is large enough, (1) in Assumption 3.10 implies Assumption 3.3. Hence, when n is large enough, Assumption 3.10 with Proposition 3.2 implies that the interior of each Voronoi face of C_k is an open subset of a $d-1$ dimensional submanifold.

The following lemma is a consequence of (2) in Assumption 3.10.

Lemma 3.11. *Under Assumption 3.10, $d_{\mathcal{N}}(\partial C_k, y_k) \geq \frac{1}{2} D_2 r$. There are constants K_1 and K_2 depending on D_1, D_2 and the Ricci curvature of $\mathcal{N}$, such that*

$$K_1 r^d \leq |C_k| \leq K_2 r^d, \tag{3.75}$$

Proof. Suppose $G_{k,i}$ is the bisector between $\mathbf{y}_k$ and $\mathbf{y}_{k,i}$. 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 [56] when r is small enough, the volume of a 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. □

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 (3.71). If n is large enough, for r satisfying Assumption 3.10, with probability greater than $1 - \frac{1}{n^2}$, for all $\mathbf{y}_k$, 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 $|\Gamma_{ki}|$ and $|\tilde{\Gamma}_{ki}|$ will not be much smaller than $|\Gamma_{ki}|$ itself when $|\Gamma_{ki}|$ 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 Γ_{ki} in (3.72). If n is large enough, for r satisfying Assumption 3.10, let $s = a_1 r^d$ in the last step of Algorithm 1 for some constant a_1, with probability greater than $1 - \frac{1}{n^2}$, for all $\mathbf{y}_k$, we have*

$$|\Gamma_{ki}| = |\tilde{\Gamma}_{ki}| + O(r^d). \tag{3.76}$$

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

$$|\Gamma_{ki}| = |\tilde{\Gamma}_{ki}|(1 + O(r)). \tag{3.77}$$

At last, if we use our approximation of the volumes of the Voronoi cells and the areas of the Voronoi faces in (3.12) we have the following implementable finite volume scheme based only on the collected dataset $\{\mathbf{y}_i\} \subset \mathcal{N}$

$$\frac{\mathrm{d}}{\mathrm{d}t}\tilde{\rho}_i|\tilde{C}_i| = \frac{1}{2}\sum_{j\in \mathrm{VF}(i)} \frac{\pi_i + \pi_j}{|y_i - y_j|}|\tilde{\Gamma}_{ij}|\left(\frac{\tilde{\rho}_j}{\pi_j} - \frac{\tilde{\rho}_i}{\pi_i}\right). \tag{3.78}$$

Moreover, same as Lemma 3.4, we know the finite volume scheme (3.78) is the forward equation for a Markov Process with transition probability $\tilde{P}_{ji}$ and jump rate $\tilde{\eta}_i$

$$\frac{\mathrm{d}}{\mathrm{d}t}\tilde{\rho}_i|\tilde{C}_i| = \sum_{j\in \mathrm{VF}(i)} \tilde{\eta}_j \tilde{P}_{ji}\tilde{\rho}_j|\tilde{C}_j| - \tilde{\eta}_i\tilde{\rho}_i|\tilde{C}_i|, \tag{3.79}$$

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

$$\begin{aligned} \tilde{\eta}_i &:= \frac{1}{2|\tilde{C}_i|\pi_i}\sum_{j\in \mathrm{VF}(i)} \frac{\pi_i + \pi_j}{|y_i - y_j|}|\tilde{\Gamma}_{ij}|, \\ \tilde{P}_{ji} &:= \frac{1}{\tilde{\eta}_j}\frac{\pi_i + \pi_j}{2\pi_j|\tilde{C}_j|}\frac{|\tilde{\Gamma}_{ij}|}{|y_i - y_j|}, \quad j \in \mathrm{VF}(i); \quad \tilde{P}_{ji} = 0, \quad j \notin \mathrm{VF}(i). \end{aligned} \tag{3.80}$$

Similar to Lemma 3.4, we know $\tilde{P}$ is the transition probability matrix with row sum 1. Denote the diagonal rate matrix as $\tilde{R} = \mathrm{diag}(\tilde{\eta}_j)$, then we also obtain an approximated Q-matrix $\tilde{Q} = \tilde{R}(\tilde{P} - I)$. Notice $\pi_i > 0$ for all $i = 1, \cdots, n$, so we always have $\tilde{\eta}_i > 0$ for all i. It also satisfies the detailed balance property

$$\tilde{\eta}_j\tilde{P}_{ji}\pi_j|\tilde{C}_j| = \tilde{\eta}_i\tilde{P}_{ij}\pi_i|\tilde{C}_i|, \tag{3.81}$$

conservation laws and the stability analysis in Lemma 3.6.

Now we state and prove the convergence of the implementable finite volume scheme (3.78). The bound of the error in the weighted ℓ^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 $i = 1, \cdots, n$, $|\mathrm{VF}(i)|$, the cardinality of $\mathrm{VF}(i)$, is order 1.

Theorem 3.14. *Suppose $\rho(\mathbf{y}, t)$, $t \in [0, T]$ is a smooth solution to the Fokker-Planck equation (3.1) on manifold $\mathcal{N} \subset \mathbb{R}^\ell$ with initial density $\rho^0(\mathbf{y})$. Let $\{\tilde{\rho}_i(t)\}_{i=1}^n$ be the solution of the finite volume scheme (3.78). Let $\tilde{e}_i := \rho(\mathbf{y}_i) - \tilde{\rho}_i$. If n is large enough, for r satisfying Assumption 3.10, we choose threshold $s = a_1 r^d$ for some constant a_1 in Algorithm (1), with probability greater than $1 - \frac{1}{n^2}$, we have*

$$\max_{t\in[0,T]} \sum_i \tilde{e}_i(t)^2 \frac{|C_i|}{\pi_i} \leq \big(\sum_i \tilde{e}_i(0)^2 \frac{|C_i|}{\pi_i} + cr\big) e^{2T}, \tag{3.82}$$

where c is a constant independent of r and n.

Proof. Define $\rho_i^e := \frac{1}{|C_i|}\int_{C_i} \rho \, \mathrm{d}y$. Plug the exact solution into the numerical scheme

$$\begin{aligned} \partial_t(\rho_i^e |C_i|) = &\sum_{j\in\mathrm{VF}(i)} \frac{\pi_i + \pi_j}{2|\mathbf{y}_i - \mathbf{y}_j|} |\tilde{\Gamma}_{ij}| \left(\frac{\rho(\mathbf{y}_j)}{\pi_j} - \frac{\rho(\mathbf{y}_i)}{\pi_i}\right) \\ &+ \sum_{j\in\mathrm{VF}(i)} \int_{\Gamma_{ij}} \pi \mathbf{n}_{ij} \cdot \nabla_{\mathcal{N}} \frac{\rho}{\pi} \, \mathrm{d}\mathcal{H}^{d-1} - \sum_{j\in\mathrm{VF}(i)} \frac{\pi_i + \pi_j}{2|\mathbf{y}_i - \mathbf{y}_j|} |\tilde{\Gamma}_{ij}| \left(\frac{\rho(\mathbf{y}_j)}{\pi_j} - \frac{\rho(\mathbf{y}_i)}{\pi_i}\right), \end{aligned} \tag{3.83}$$

where $\mathbf{n}_{ij}$ is the restriction of the unit outward normal vector field on Γ_{ij}. Subtracting the numerical scheme (3.78) from (3.83), we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\tilde{e}_i |C_i| = \sum_{j\in\mathrm{VF}(i)} \frac{\pi_i + \pi_j}{2|y_i - y_j|} |\tilde{\Gamma}_{ij}| \left(\frac{\tilde{e}_j}{\pi_j} - \frac{\tilde{e}_i}{\pi_i}\right) + \sum_{j\in\mathrm{VF}(i)} \varepsilon_{ij} + \partial_t((\rho(\mathbf{y}_i) - \rho_i^e)|C_i|) + \frac{\mathrm{d}}{\mathrm{d}t}\tilde{\rho}_i(|\tilde{C}_i| - |C_i|), \tag{3.84}$$

where

$$\begin{aligned} \varepsilon_{ij} := &\int_{\Gamma_{ij}} \pi \mathbf{n}_{ij} \cdot \nabla_{\mathcal{N}} \frac{\rho}{\pi} \, \mathrm{d}\mathcal{H}^{d-1} - \frac{\pi_i + \pi_j}{2|\mathbf{y}_i - \mathbf{y}_j|} |\Gamma_{ij}| \left(\frac{\rho(\mathbf{y}_j)}{\pi_j} - \frac{\rho(\mathbf{y}_i)}{\pi_i}\right) \\ &+ \frac{\pi_i + \pi_j}{2|y_i - y_j|} (|\Gamma_{ij}| - |\tilde{\Gamma}_{ij}|) \left(\frac{\rho(\mathbf{y}_j)}{\pi_j} - \frac{\rho(\mathbf{y}_i)}{\pi_i}\right). \end{aligned} \tag{3.85}$$

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

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\sum_i \tilde{e}_i^2 \frac{|C_i|}{\pi_i} \leq &-\frac{1}{2}\sum_i \sum_{j\in\mathrm{VF}(i)} \frac{\pi_i + \pi_j}{2|y_i - y_j|} |\tilde{\Gamma}_{ij}| \left(\frac{\tilde{e}_j}{\pi_j} - \frac{\tilde{e}_i}{\pi_i}\right)^2 + \frac{1}{2}\sum_i \sum_{j\in\mathrm{VF}(i)} \frac{\varepsilon_{ij}^2}{\frac{\pi_i+\pi_j}{2|y_i-y_j|}|\tilde{\Gamma}_{ij}|} \\ &+ \sum_i [\partial_t(\rho(\mathbf{y}_i) - \rho_i^e)]^2 \frac{|C_i|}{\pi_i} + \sum_i \big(\frac{\mathrm{d}}{\mathrm{d}t}\tilde{\rho}_i(\frac{|\tilde{C}_i| - |C_i|}{|C_i|})\big)^2 \frac{|C_i|}{\pi_i} + 2\sum_i e_i^2 \frac{|C_i|}{\pi_i} \\ \leq &\sum_i \sum_{j\in\mathrm{VF}(i)} \frac{\varepsilon_{ij}^2}{\frac{\pi_i+\pi_j}{2|y_i-y_j|}|\tilde{\Gamma}_{ij}|} + \sum_i [\partial_t(\rho(\mathbf{y}_i) - \rho_i^e)]^2 \frac{|C_i|}{\pi_i} \\ &+ \sum_i \big(\frac{\mathrm{d}}{\mathrm{d}t}\tilde{\rho}_i(\frac{|\tilde{C}_i| - |C_i|}{|C_i|})\big)^2 \frac{|C_i|}{\pi_i} + 2\sum_i e_i^2 \frac{|C_i|}{\pi_i} \end{aligned} \tag{3.86}$$

$$=: \epsilon_1 + \epsilon_2 + \epsilon_3 + 2\sum_i e_i^2 \frac{|C_i|}{\pi_i}.$$

First, we estimate the term ϵ_1, in particular, $\frac{\varepsilon_{ij}^2}{\frac{\pi_i+\pi_j}{2|y_i-y_j|}|\tilde{\Gamma}_{ij}|}$ for $j \in \mathrm{VF}(i)$. Since the exact solution is smooth such that

$$|\rho(\mathbf{y}_i, t) - \rho(\mathbf{y}_j, t)| \leq C_{Lip}|\mathbf{y}_i - \mathbf{y}_j|, \tag{3.87}$$

by (3.61),

$$\varepsilon_{ij} = O((\mathrm{diam}(C_i) + d_{\mathcal{N}}(\mathbf{y}_i, \mathbf{y}_j))|\Gamma_{ij}|) + O(|\Gamma_{ij}| - |\tilde{\Gamma}_{ij}|). \tag{3.88}$$

Hence,

$$\frac{\varepsilon_{ij}^2}{\frac{\pi_i+\pi_j}{2|y_i-y_j|}|\tilde{\Gamma}_{ij}|} = O(d_{\mathcal{N}}(\mathbf{y}_i, \mathbf{y}_j)(\mathrm{diam}(C_i) + d_{\mathcal{N}}(\mathbf{y}_i, \mathbf{y}_j))^2 \frac{|\Gamma_{ij}|^2}{|\tilde{\Gamma}_{ij}|}) + O(d_{\mathcal{N}}(\mathbf{y}_i, \mathbf{y}_j)\frac{(|\Gamma_{ij}| - |\tilde{\Gamma}_{ij}|)^2}{|\tilde{\Gamma}_{ij}|}). \tag{3.89}$$

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

$$\frac{(|\Gamma_{ij}| - |\tilde{\Gamma}_{ij}|)^2}{|\tilde{\Gamma}_{ij}|} = O(r^d). \tag{3.90}$$

By Assumption 3.10 and Lemma 3.9, $d_{\mathcal{N}}(\mathbf{y}_i, \mathbf{y}_j)$ and $\mathrm{diam}(C_i)$ are of order r. By Assumption 3.10 and Proposition 3.2, since $\mathcal{N}$ is compact, there is a constant K such that $|\Gamma_{ij}| \leq K r^{d-1}$. Therefore, $\frac{\varepsilon_{ij}^2}{\frac{\pi_i+\pi_j}{2|y_i-y_j|}|\tilde{\Gamma}_{ij}|} = O(r^{d+1})$ and

$$\epsilon_1 = \sum_i \sum_{j\in \mathrm{VF}(i)} \frac{\varepsilon_{ij}^2}{\frac{\pi_i+\pi_j}{2|y_i-y_j|}|\tilde{\Gamma}_{ij}|} = O(nr^{d+1} \max_i |\mathrm{VF}(i)|) = O(r \max_i |\mathrm{VF}(i)|), \tag{3.91}$$

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 (3.75),

$$\sum_i \Big(\frac{\mathrm{d}}{\mathrm{d}t}\tilde{\rho}_i\big(\frac{|\tilde{C}_i| - |C_i|}{|C_i|}\big)\Big)^2 \frac{|C_i|}{\pi_i} = O(r^2). \tag{3.92}$$

By (3.66) and Assumption 3.10,

$$\sum_i [\partial_t(\rho(\mathbf{y}_i) - \rho_i^e)]^2 \frac{|C_i|}{\pi_i} = O(r^2). \tag{3.93}$$

We sum up all the terms,

$$\frac{\mathrm{d}}{\mathrm{d}t}\sum_i \tilde{e}_i^2 \frac{|C_i|}{\pi_i} \leq O(r \max_i |\mathrm{VF}(i)|) + 2\sum_i e_i^2 \frac{|C_i|}{\pi_i}. \tag{3.94}$$

In conclusion

$$\max_{t\in[0,T]} \sum_i \tilde{e}_i(t)^2 \frac{|C_i|}{\pi_i} \leq \Big(\sum_i \tilde{e}_i(0)^2 \frac{|C_i|}{\pi_i} + O(r \max_i |\mathrm{VF}(i)|)\Big) e^{2T}. \quad \square \tag{3.95}$$

3.5. Unconditionally stable explicit time stepping and exponential convergence

To the end of this section, we show that the detailed balance property (3.81) leads to stability and exponential convergence of a discrete-in-time Markov process.

Under the detailed balance condition (3.81), we recast (3.79) to

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\tilde{\rho}_i}{\pi_i} = \sum_{j\in \mathrm{VF}(i)} \tilde{\eta}_i \tilde{P}_{ij}\frac{\tilde{\rho}_j}{\pi_j} - \tilde{\eta}_i\frac{\tilde{\rho}_i}{\pi_i}. \tag{3.96}$$

Let ρ_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

$$\frac{\rho_i^{k+1}}{\pi_i} = \frac{\rho_i^k}{\pi_i} - \tilde{\eta}_i\Delta t\frac{\rho_i^{k+1}}{\pi_i} + \Delta t\sum_{j\in \mathrm{VF}(i)} \tilde{\eta}_i\tilde{P}_{ij}\frac{\rho_j^k}{\pi_j}, \tag{3.97}$$

where $\tilde{\eta}_i$ and $\tilde{P}_{ij}$ are defined in (3.80). The above equation is equivalent to

$$\frac{\rho_i^{k+1}}{\pi_i} = \frac{\rho_i^k}{\pi_i} + \frac{\tilde{\eta}_i\Delta t}{1+\tilde{\eta}_i\Delta t}\left(\sum_{j\in \mathrm{VF}(i)} \tilde{P}_{ij}\frac{\rho_j^k}{\pi_j} - \frac{\rho_i^k}{\pi_i}\right). \tag{3.98}$$

For $u_i^{k+1} := \frac{\rho_i^{k+1}}{\pi_i}$, the matrix formulation of (3.98) is

$$u^{k+1} = (I + \Delta t Q)u^k, \tag{3.99}$$

where

$$Q := \{\hat{b}_{ij}\} = \begin{cases} -\frac{\tilde{\eta}_i}{1+\tilde{\eta}_i\Delta t}, & j = i; \\ \frac{\tilde{\eta}_i}{1+\tilde{\eta}_i\Delta t}\tilde{P}_{ij}, & j \neq i \end{cases} \tag{3.100}$$

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

Below, we first summarize the explicit time stepping (3.97) as an algorithm and then prove the unconditionally stability.

Algorithm 2: Explicit time stepping for Markov process.

Parameters: Algorithm inputs: error tolerance ϵ, time step Δt, the initial distribution (ρ_i^0), the target invariant measure (π_i), the approximated volume $|\tilde{C}_k|$ of Voronoi cell and ares $|\tilde{\Gamma_{k\ell}}|$

1 Compute transition probability matrix $\tilde{P}_{ij}$ and $\tilde{\eta}$ defined in (3.80).

2 Compute discrete time transition probability matrix Q defined in (3.100).

3 $k \to k+1$ iteration: $\frac{\rho^{k+1}}{\pi} = (I + \Delta t Q)\frac{\rho^k}{\pi}$. Repeat until $\|\frac{\rho^{k+1}}{\pi} - 1\|_\infty < \epsilon$.

Now we show Q defined in (3.100) is the generator of a new Markov process.

For $w_i^{k+1} := \rho_i^{k+1}|\tilde{C}_i|$, (3.97), together with detailed balance property (3.81), yields

$$\rho_i^{k+1}|\tilde{C}_i| - \rho_i^k|\tilde{C}_i| = \Delta t\left(\sum_{j\in \mathrm{VF}(i)} \tilde{\eta}_j\tilde{P}_{ji}\rho_j^k|\tilde{C}_j| - \tilde{\eta}_i\rho_i^{k+1}|\tilde{C}_i|\right), \tag{3.101}$$

which can be recast as

$$(1+\Delta t\tilde{\eta}_i)\rho_i^{k+1}|\tilde{C}_i| = (1+\Delta t\tilde{\eta}_i)\rho_i^{k}|\tilde{C}_i| + \Delta t\left(\sum_{j\in \mathrm{VF}(i)} \tilde{\eta}_j\tilde{P}_{ji}\rho_j^k|\tilde{C}_j| - \tilde{\eta}_i|\tilde{C}_i|\rho_i^k\right). \tag{3.102}$$

Denote $g_i^{k+1} := (1+\Delta t\tilde{\eta}_i)\rho_i^{k+1}|\tilde{C}_i|$. (3.102) can be simplified as

$$g_i^{k+1} = g_i^k + \Delta t\left(\sum_j \frac{\tilde{\eta}_j}{1+\Delta t\tilde{\eta}_j}\tilde{P}_{ji}g_j^k - \frac{\tilde{\eta}_i}{1+\Delta t\tilde{\eta}_i}g_i^k\right). \tag{3.103}$$

This is a new Markov process for g_i with transition probability $\tilde{P}_{ji}$ and a new jump rate $s_j = \frac{\tilde{\eta}_j}{1+\Delta t\tilde{\eta}_j}$. With Q in (3.100), the matrix formulation for g is

$$g^{k+1} = (I+\Delta tQ)^* g^k. \tag{3.104}$$

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

Proposition 3.15. *Assume $\pi_i > 0$ for all $i = 1, \cdots, n$. Let $\tilde{\eta}_i$ be the approximated jump rate and $\tilde{P}_{ij}$ be the approximated transition probability defined in (3.80). Let Δt be the time step and consider the explicit time stepping (3.97), i.e., Algorithm 2. Assume the initial data satisfies*

$$\sum_i (1+\tilde{\eta}_i\Delta t)\rho_i^0|\tilde{C}_i| = \sum_i (1+\tilde{\eta}_i\Delta t)\pi_i|\tilde{C}_i|. \tag{3.105}$$

Then we have

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

$$\sum_i (1+\eta_i\Delta t)\rho_i^{k+1}|\tilde{C}_i| = \sum_i (1+\eta_i\Delta t)\rho_i^{k}|\tilde{C}_i|; \tag{3.106}$$

(ii) the unconditional maximum principle for $\frac{\rho_i}{\pi_i}$

$$\max_i \frac{\rho_j^{k+1}}{\pi_j} \le \max_j \frac{\rho_j^k}{\pi_j}. \tag{3.107}$$

(iii) the ℓ^∞ contraction

$$\max_i \left|\frac{\rho_i^{k+1}}{\pi_i} - 1\right| \le \max_i \left|\frac{\rho_i^{k}}{\pi_i} - 1\right|; \tag{3.108}$$

(iv) the exponential convergence

$$\left\|\frac{\rho_i^k}{\pi_i} - 1\right\|_{\ell^\infty} \le c|\lambda_2|^k, \quad |\lambda_2| < 1, \tag{3.109}$$

where λ_2 is the second eigenvalue (in terms of the magnitude) of $I+\Delta tQ$, i.e. $\lambda_2 = 1 - \mathrm{gap}_Q\,\Delta t$ and gap_Q is the spectral gap of Q.

Proof. First, recast (3.98) as

$$\frac{\rho_i^{n+1}}{\pi_i} = \frac{1}{1+\tilde{\eta}_i\Delta t}\frac{\rho_i^n}{\pi_i} + \frac{\tilde{\eta}_i\Delta t}{1+\tilde{\eta}_i\Delta t}\left(\sum_{j\in \mathrm{VF}(i)} \tilde{P}_{ij}\frac{\rho_j^n}{\pi_j}\right), \tag{3.110}$$

which gives the unconditional maximum principle (3.107).

Second, from (3.110), we have

$$\frac{\rho_i^{k+1}}{\pi_i} - 1 = \frac{1}{1+\tilde{\eta}_i\Delta t}\left(\frac{\rho_i^k}{\pi_i} - 1\right) + \frac{\tilde{\eta}_i\Delta t}{1+\tilde{\eta}_i\Delta t}\sum_{j\in \mathrm{VF}(i)} \tilde{P}_{ij}\left(\frac{\rho_j^k}{\pi_j} - 1\right). \tag{3.111}$$

Then we have

$$\left|\frac{\rho_i^{k+1}}{\pi_i} - 1\right| \le \frac{1}{1+\tilde{\eta}_i\Delta t}\left|\frac{\rho_i^k}{\pi_i} - 1\right| + \frac{\tilde{\eta}_i\Delta t}{1+\tilde{\eta}_i\Delta t}\sum_{j\in \mathrm{VF}(i)} \tilde{P}_{ij}\left|\frac{\rho_j^k}{\pi_j} - 1\right| \le \max_i \left|\frac{\rho_j^n}{\pi_j} - 1\right|, \tag{3.112}$$

which gives (3.108).

Third, recall the matrix formulation (3.99). Every element in $(I+\Delta t Q)^m$ is strictly positive for some m. By Perron-Frobenius theorem, $\lambda_1 = 1$ is the simple, principal eigenvalue of $I+\Delta t Q$ with the ground state $u^* \equiv \{1,1,\cdots,1\}$ and other eigenvalues λ_i satisfy $|\lambda_i| < \lambda_1$. On one hand, the mass conservation for initial data $u^0 = \frac{\rho^0}{\pi}$ satisfies (3.106), i.e.,

$$\sum_i (u_i^0 - u_i^*)u_i^*(1+\Delta t\eta_i)\pi|\tilde{C}|_i = 0. \tag{3.113}$$

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

$$u^0 - u^* = \sum_{j=2} c_j u_j, \quad u_j \text{ is the eigenfunction corresponding to } \lambda_j. \tag{3.114}$$

Therefore, we have

$$u^k - u^* = (I+\Delta t Q)^k(u^0 - u^*) = \sum_{j=2} c_j\lambda_j^k u_j, \tag{3.115}$$

which concludes

$$\left\|\frac{\rho_i^k}{\pi_i} - 1\right\|_{\ell^\infty} \le c|\lambda_2|^k \quad \text{with } |\lambda_2| < 1. \tag{3.116}$$

Here λ_2 is the second eigenvalue (in terms of the magnitude) of $I+\Delta t Q$ sitting in the ball with radius $\lambda_1 = 1$ and thus $|\lambda_2| < 1$.

Finally, taking summation with respect to i in (3.101) shows

$$\begin{aligned}\sum_i \left(\rho_i^{k+1}|\tilde{C}_i| - \rho_i^k|\tilde{C}_i|\right) =& \Delta t\left(\sum_{i,j}\tilde{\eta}_j\tilde{P}_{ji}\rho_j^k|\tilde{C}_j| - \sum_i \tilde{\eta}_i\rho_i^{k+1}|\tilde{C}_i|\right)\\ =& \Delta t\left(\sum_j \tilde{\eta}_j\rho_j^k|\tilde{C}_j| - \sum_i \tilde{\eta}_i\rho_i^{k+1}|\tilde{C}_i|\right),\end{aligned} \tag{3.117}$$

which gives (3.106). □

As a comparison, we also give some other standard stability estimates for both explicit and implicit schemes in Appendix D and show that only the unconditional stable explicit scheme (3.97) achieves both the efficiency and the stability. We refer to [27] for successful applications of Algorithm 2 to image morphing problems with 2D structured spacial grids. With structured grids, instead of Voronoi cell approximations obtained from sample points, the computations of explicit time stepping for Markov process using Algorithm 2 are more accurate. [27] also combines Algorithm 2 with a thresholding dynamics to simulate mass-conserved shape dynamics for distribution with binary values $1, 2$.

4. Simulations for Fokker-Planck solver

In this section, we conduct some challenging numerical simulations with reaction coordinates for the dumbbell, the Klein bottle and sphere. We use the dataset $\{\mathbf{y}_i\}_{i=1}^{2000}$ with the reaction coordinates on the underlying manifolds including dumbbell, Klein bottle and sphere to solve the Fokker-Planck equation (3.1) following the unconditionally stable explicit scheme (3.97).

4.1. Comparison with a ground-truth dynamics on sphere

In this section, we construct a ground-truth exact solution given by an oscillated von Mises-Fisher distribution on the 2-sphere in $\mathbb{R}^3$. This distribution is a commonly used distribution in physics and bioinformatics, for instance, to model the electric field-induced dipole interaction. For other complicated applications, it is hard to construct a ground-truth exact solution with an exact source term. So we refer to [8,9] for other comparison methods without knowing an exact solution.

We choose the spherical coordinates as

$$\theta \in [0, \pi], \quad \varphi \in [0, 2\pi], \quad \text{with} \quad x = \cos\varphi \sin\theta, \quad y = \sin\varphi \sin\theta, \quad z = \cos\theta. \tag{4.1}$$

For $t \in [0, 2]$, define three parameters

$$\kappa(t) = 1 + 0.2\sin(t), \quad a(t) = \pi/2 + 0.2\sin(3t), \quad b(t) = 5t. \tag{4.2}$$

Define the polar angle

$$\eta(\theta, \varphi, t) = \cos a(t)\cos\theta + \sin a(t)\sin\theta\cos(\varphi - b(t)). \tag{4.3}$$

Then we choose the exact solution as the von Mises–Fisher distribution

$$\rho_v(\theta, \varphi, t) = C(\kappa(t))e^{\kappa(t)\eta(\theta,\varphi,t)}, \tag{4.4}$$

where $C(\kappa) = \frac{\kappa}{4\pi \sinh\kappa}$.

Based on the surface gradient and surface divergence on sphere

$$\nabla_{\mathcal{N}} f = \frac{\partial f}{\partial\theta}\hat{\theta} + \frac{1}{\sin\theta}\frac{\partial f}{\partial\varphi}\hat{\varphi}, \quad \nabla_{\mathcal{N}} \cdot \vec{F} = \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}[\sin\theta F_\theta] + \frac{1}{\sin\theta}\frac{\partial F_\varphi}{\partial\varphi}, \tag{4.5}$$

then it satisfies Fokker-Planck equation (3.1) with source term

Table 2
The root mean square error e.

Time	0	0.4	0.8	1.2	1.6	2
RMSE	0	0.0151	0.0138	0.0126	0.0149	0.0140

$$\begin{aligned} g(\theta,\varphi,t) &= \mathrm{div}_{\mathcal{N}}(\nabla_{\mathcal{N}}\rho^{\mathcal{N}} + \rho_t^{\mathcal{N}}\nabla_{\mathcal{N}}U_{\mathcal{N}}) - \partial_t\rho^{\mathcal{N}} \\ &= \frac{\partial^2\rho^{\mathcal{N}}}{\partial\theta^2} + \frac{\partial\rho^{\mathcal{N}}}{\partial\theta}\frac{\partial U_{\mathcal{N}}}{\partial\theta} + \rho^{\mathcal{N}}\frac{\partial^2 U_{\mathcal{N}}}{\partial\theta^2} + \cot\theta[\frac{\partial\rho^{\mathcal{N}}}{\partial\theta} + \rho^{\mathcal{N}}\frac{\partial U_{\mathcal{N}}}{\partial\theta}] \\ &\quad + \frac{1}{\sin^2\theta}[\frac{\partial^2\rho^{\mathcal{N}}}{\partial\varphi^2} + \frac{\partial\rho^{\mathcal{N}}}{\partial\varphi}\frac{\partial U_{\mathcal{N}}}{\partial\varphi} + \rho^{\mathcal{N}}\frac{\partial^2 U_{\mathcal{N}}}{\partial\varphi^2}] - \partial_t\rho^{\mathcal{N}}. \end{aligned} \tag{4.6}$$

Take $U = 1$, plugging (4.4) into the RHS of (4.6), we obtain

$$g(\theta,\varphi,t) = \rho\left[\kappa^2\left[\eta_\theta^2 + \sin^2 a\sin^2(\varphi - b)\right] - 2\kappa\eta - \frac{C'}{C}\kappa' - (\kappa'\eta + \kappa\eta_t)\right]; \tag{4.7}$$

see details in Appendix E.

With $\pi = e^{-U}$, and the source term g computed from the exact solution (4.7). Then Algorithm 2, i.e., the explicit scheme (3.98), becomes

$$\frac{\rho_i^{k+1}}{\pi_i} = \frac{\rho_i^k}{\pi_i} + \frac{\tilde{\eta}_i\Delta t}{1 + \tilde{\eta}_i\Delta t}\left(\sum_{j\in \mathrm{VF}(i)}\tilde{P}_{ij}\frac{\rho_j^k}{\pi_j} - \frac{\rho_i^k}{\pi_i}\right) - \Delta t\frac{g_i^k}{\pi_i}. \tag{4.8}$$

Here for the discrete source term g_i^k, we use continuous time derivatives at time step k and discrete spacial derivative on grid i. For $u_i^{k+1} := \frac{\rho_i^{k+1}}{\pi_i}$, with the additional source term $g(\theta,\varphi,t)$, the matrix formulation with Q defined in (3.100) is $u^{k+1} = (I + \Delta t Q)u^k - \Delta t\frac{g^k}{\pi}$.

To compare the numerical solution and the exact solution with a long time validation. We take $\Delta t = 0.001$ and final time as $T = 2000 * \Delta t$ with iteration number 2000. We first sample 2000 data points on a unit sphere $\mathcal{N} = S^2 \subset \mathbb{R}^3$, then we compute the approximated Voronoi cell volumes $|\tilde{C}_i|_{i=1}^n$ and areas $\tilde{\Gamma}_{ij}$ from Algorithm 1 by taking the bandwidth $r = 0.3$. The equilibrium $\{\pi_i\}$ is taking to be constant, which is normalized so that the total mass condition (3.105) is satisfied. In Fig. 2. We plot 6 snapshots at $t = 0, 0.4, 0.8, 1.2, 1.6, 2.0$ for both numerical solution ρ_i and exact solution $\rho_v(\theta,\varphi,t)$ in (4.4) starting from the same initial data given by $\rho_v(\theta,\varphi,0)$. We also list Table 2 to show the root mean square error (RMSE) $e := \sqrt{\frac{1}{2000}\sum_{i=1}^{2000}|\rho_i - \rho_v(i)|^2}$ at these 6 times. A video showing the dynamics of both numerical solution ρ_i and exact solution ρ_v is provided in https://youtu.be/x98J8CSYBq8.

4.2. Example I: Fokker-Planck evolution on dumbbell

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

$$\begin{aligned} r &= \sqrt{\sqrt{1 + 0.95^4(\cos(2\phi)^2 - 1)} + 0.95^2\cos(2\phi)} \\ x &= r\sin(\phi)\cos(\theta), \quad y = r\sin(\phi)\sin(\theta), \quad z = r\cos(\phi). \end{aligned} \tag{4.9}$$

After composition with a dilation and rotation map f_2 of $\mathbb{R}^{200}$, we have an embedded dumbbell $\mathcal{M} \subset \mathbb{R}^{200}$. Suppose $f_2 \circ f_1(\theta,\phi)$ 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 $\mathbf{x}_i = f_2 \circ f_1(\theta_i,\phi_i)$, then we have a non uniform sample $\{\mathbf{x}_i\}_{i=1}^{4000}$ on $\mathcal{M}$. We apply the

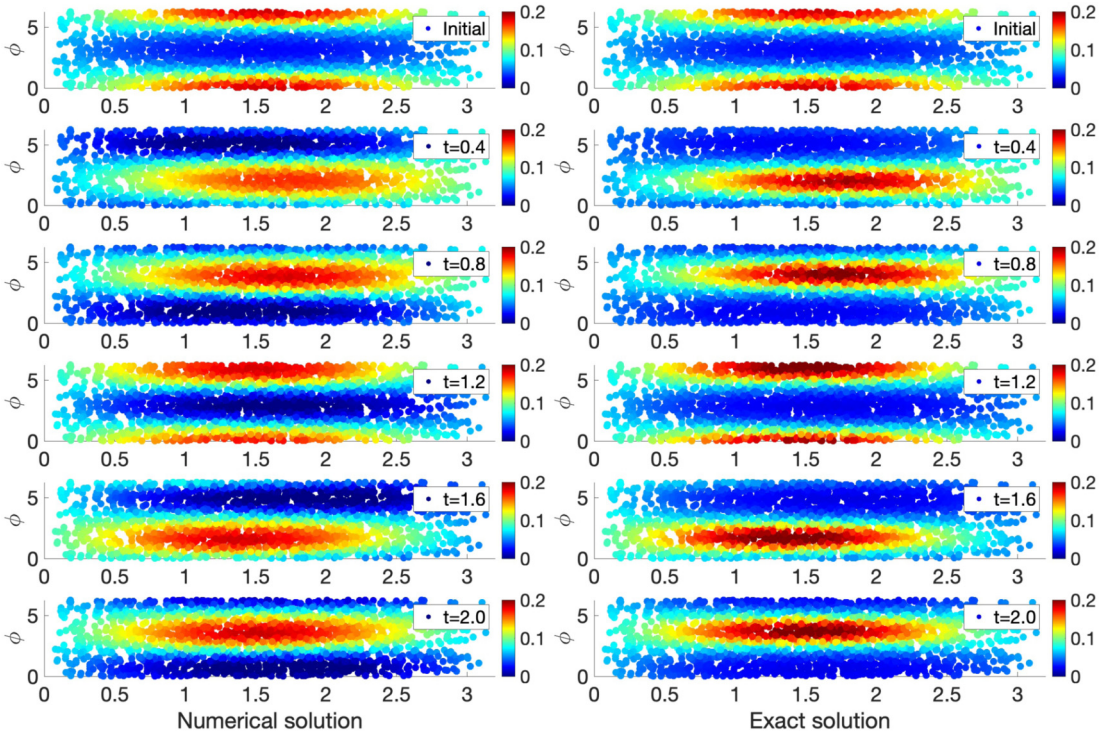


Fig. 2. Left: The numerical solution ρ_i^k in terms of θ, φ over 2000 points $\{\mathbf{y}_i\}_{i=1}^{2000} \subset S^2 \subset \mathbb{R}^3$. Right: The exact solution given by the oscillated von Mises–Fisher distribution $\rho_v(\theta, \varphi, t)$ in (4.4). We plot 6 snapshots at $t = 0, 0.4, 0.8, 1.2, 1.6, 2.0$.

diffusion map to find the reaction coordinates $\{\mathbf{y}_i\}_{i=1}^{4000}$ of $\{\mathbf{x}_i\}_{i=1}^{4000}$ in $\mathbb{R}^3$, i.e. $\{\mathbf{y}_i\}_{i=1}^{4000}$ can be regarded as a non uniform sample on a dumbbell $\mathcal{N} \subset \mathbb{R}^3$.

Suppose ψ_i is the i th eigenfunction of the Laplace-Beltrami operator on $\mathcal{N}$. Assume the initial density ρ^0 is ψ_2 plus some constant (so that ρ^0 is positive) as shown in Fig. 3. Assume the equilibrium density π is ψ_8 plus some constant as shown in Fig. 3. 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 ρ^0 by $c\rho^0$ such that (3.105) holds. We set the time step $\Delta t = 0.05$. Let $T = k\Delta t$ for the integer k and $1 \leq k \leq 20000$, i.e., we iterate the scheme for 20000 times and set the final time to be $T = 20000 * \Delta t = 1000$. We use the unconditional stable explicit scheme (3.97) to solve ρ^k. We compare the numerical relative error in maximum norm with the theoretic relative error, $|\lambda_2|^k = 0.9997^k$ in (3.109), in the semilog-plot in Fig. 4. The exponential convergence rate is exactly same. To clearly see the dynamics of the change of the density over the 4000 points, we plot ρ^k for $k = 20, 60, 100, 160, 220, 4000$, correspondingly $T = 1, 3, 5, 8, 11, 200$ in Fig. 5.

4.3. Example II: Fokker-Planck evolution on Klein bottle

Suppose $(\theta, \phi) \in [0, 2\pi) \times [0, 2\pi)$, then we have the following Klein bottle in $\mathcal{N} \subset \mathbb{R}^4$ parametrized as $(x, y, z, w) = f(\theta, \phi) \in \mathbb{R}^4$, where

$$\begin{aligned} x &= (1 + 0.3\cos(\theta))\cos(\phi) \\ y &= (1 + 0.3\cos(\theta))\sin(\phi) \\ z &= 0.3\sin(\theta)\cos(\frac{\phi}{2}) \\ w &= 0.3\sin(\theta)\sin(\frac{\phi}{2}) \end{aligned} \tag{4.10}$$

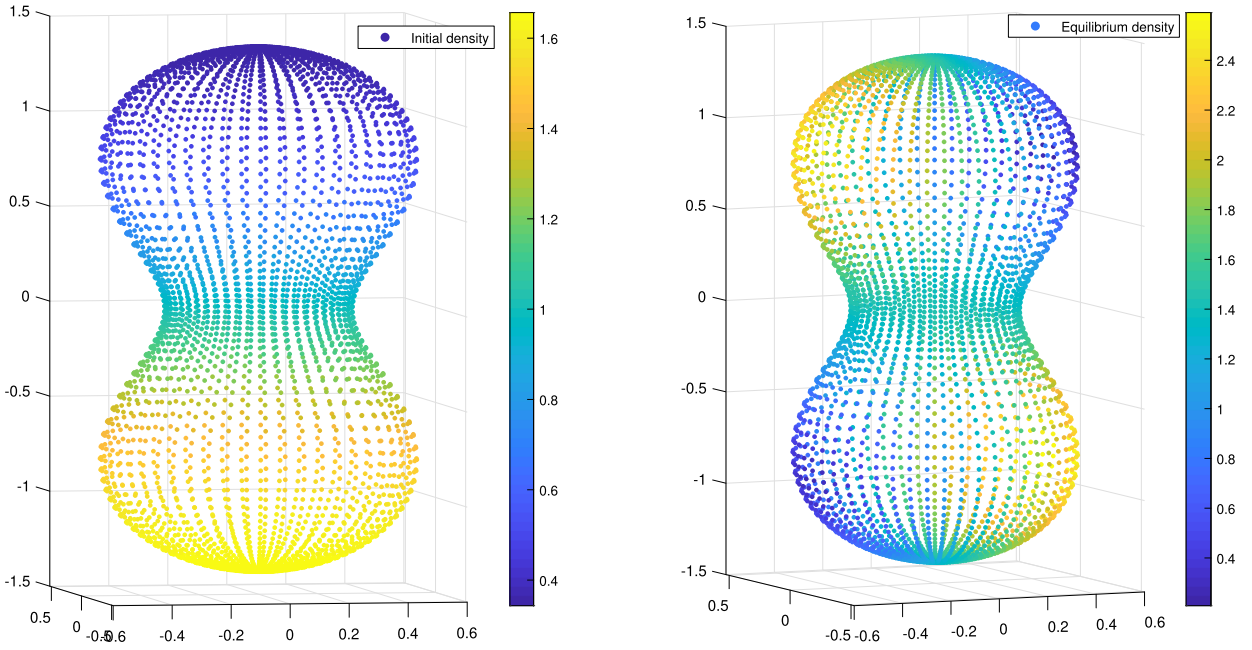


Fig. 3. Left: The initial density is the second eigenfunction of the Laplace Beltrami operator on a dumbbell $\mathcal{N} \subset \mathbb{R}^3$ plus a constant. We plot it over 4000 points $\{\mathbf{y}_i\}_{i=1}^{4000} \subset \mathcal{N} \subset \mathbb{R}^3$. Right: The equilibrium density is the 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 $\{\mathbf{y}_i\}_{i=1}^{4000} \subset \mathcal{N} \subset \mathbb{R}^3$.

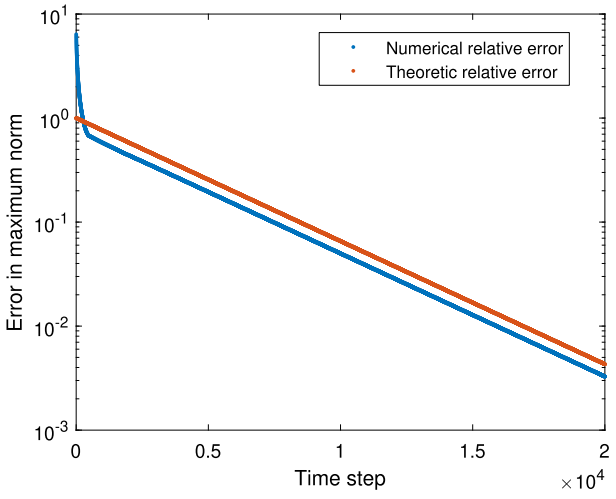


Fig. 4. The semilog-plot comparison between the numerical relative error with theoretic relative error. The numerical relative error is the error from the unconditional stable explicit scheme (3.97) with $\Delta t = 0.05$ and $1 \leq k \leq 20000$. The theoretic relative error is based on (3.109) with $|\lambda_2|^k = 0.9997^k$.

We sample 2000 points $(\theta_1, \phi_1), \cdots, (\theta_{2000}, \phi_{2000})$ on $[0, 2\pi) \times [0, 2\pi)$. Let $\mathbf{y}_i = f(\theta_i, \phi_i)$, then we have non uniform samples $\{\mathbf{y}_i\}_{i=1}^{2000}$ on $\mathcal{N}$. We can regard them as the reaction coordinates of 2000 points sampled on $\mathcal{M}$ (a manifold diffeomorphic to a Klein bottle) in some high dimensional space. In this example, we will visualize the functions on the Klein bottle by two methods. First, consider the projection from $\mathbb{R}^4$ to $\mathbb{R}^3$ by $(x, y, z, w) \to (x, y, z)$. The restriction of the projection on $\mathcal{N}$ maps the Klein bottle to a pinched torus in $\mathbb{R}^3$. Second, consider the projection from $\mathbb{R}^4$ to $\mathbb{R}^3$ by $(x, y, z, w) \to (y, z, w)$. The restriction of the projection on $\mathcal{N}$ maps the Klein bottle to a Roman surface in $\mathbb{R}^3$. For any function on the Klein bottle, we will visualize it by plotting it on both the pinched torus and the Roman surface.

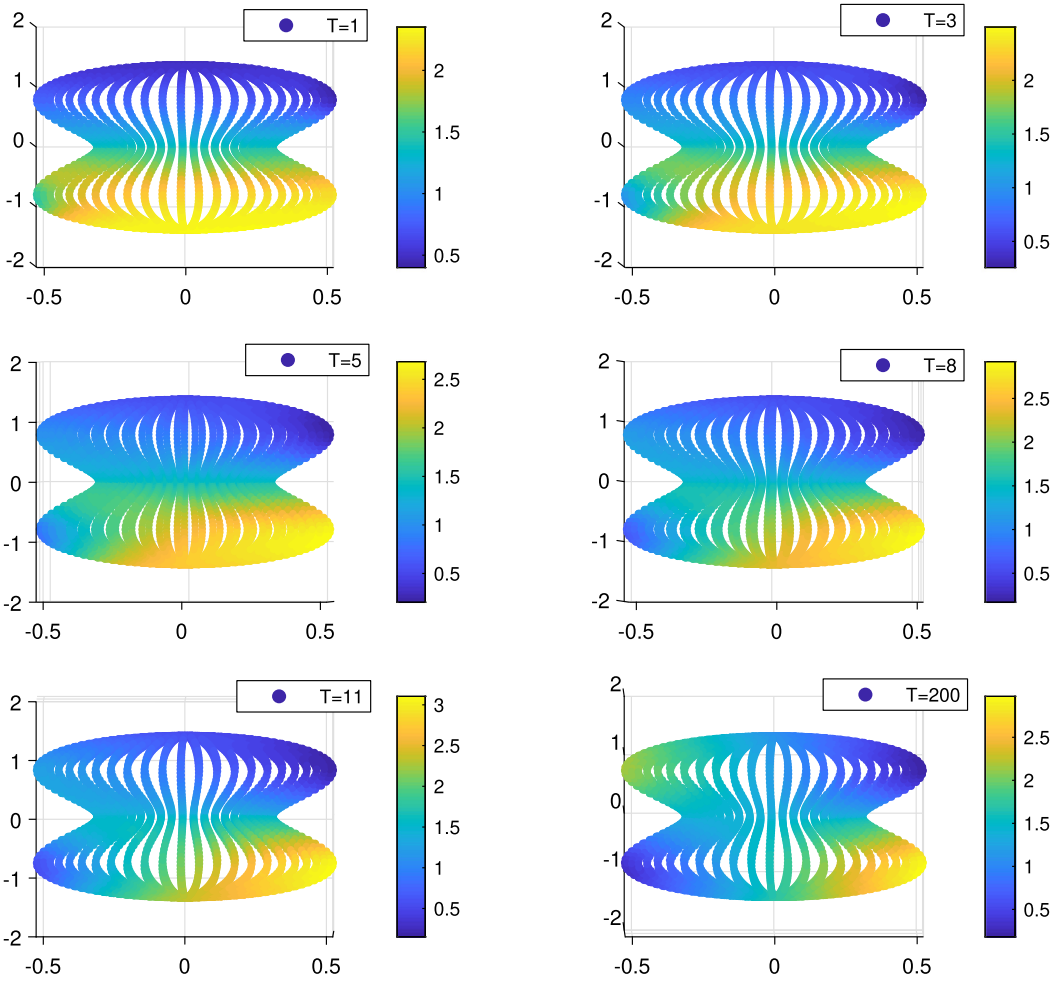


Fig. 5. The density dynamics ρ^k from the unconditional stable explicit scheme (3.97) with $\Delta t = 0.05$. We plot ρ^k for $k = 20, 60, 100, 160, 220, 4000$, correspondingly on time $T = 1, 3, 5, 8, 11, 200$.

Suppose ψ_i is the i th eigenfunction of the Laplace-Beltrami operator on $\mathcal{N}$. Assume the initial density ρ^0 is ψ_2 plus some constant (so that ρ^0 is positive) as shown in Fig. 6. Assume the equilibrium density π is ψ_7 plus some constant as shown in Fig. 6. 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 $r = 0.23$ and threshold $s = 0$. Then we adjust the initial data, i.e., we replace ρ^0 by $c\rho^0$ such that (3.105) 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 $T = 10000 * \Delta t = 500$. We use the unconditional stable explicit scheme (3.97) to solve ρ^k. We compare the numerical relative error in maximum norm with the theoretic relative error, $|\lambda_2|^k = 0.9993^k$ in (3.109), in the semilog-plot in Fig. 7. The exponential convergence rate is exactly the same. To clearly see the dynamics of the change of the density over the 2000 points, we plot ρ^k for $k = 50, 1000, 2000, 10000$, correspondingly $T = 2.5, 50, 100, 500$ in Fig. 8.

4.4. 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 altitude of continents and the depth of oceans for earth based on the dataset for initial distribution of Pangaea supercontinent (250 million years ago) and the equilibrium distribution of the current earth.

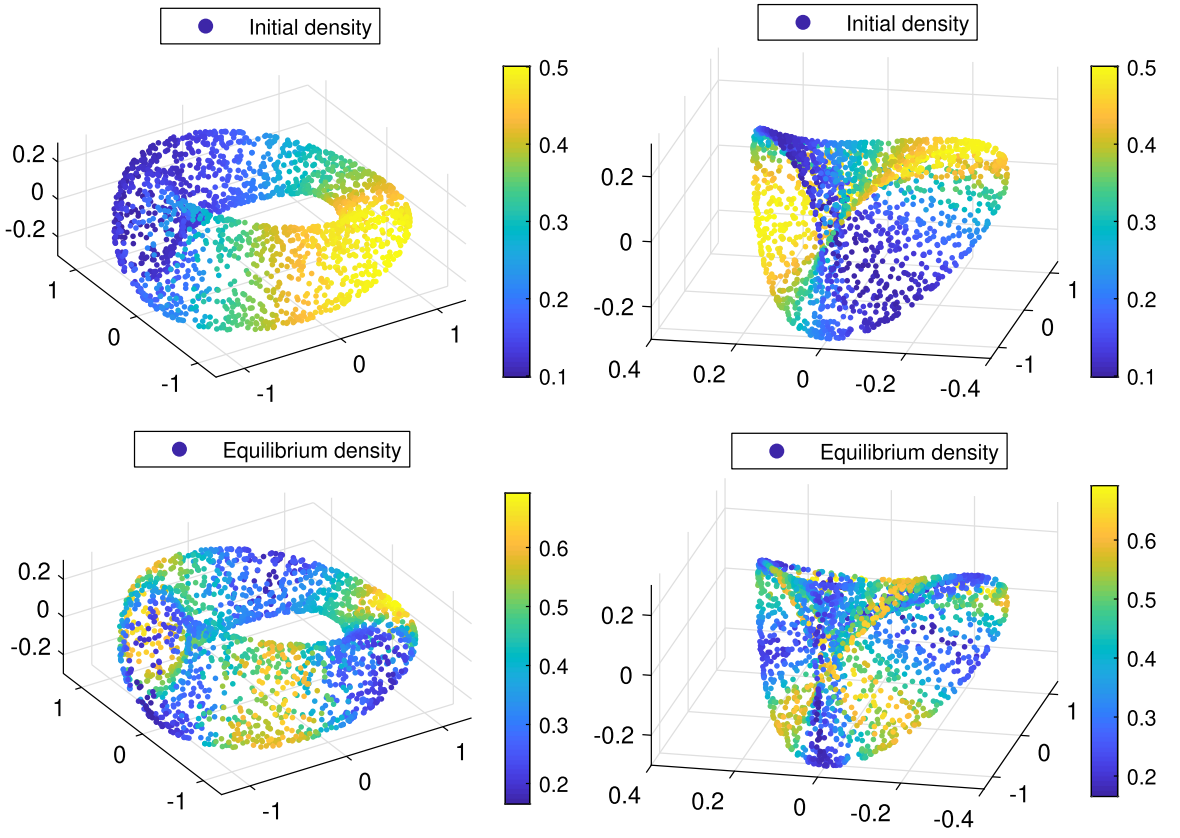


Fig. 6. Top two panels: The initial density is the second eigenfunction of the Laplace Beltrami operator on a Klein bottle $\mathcal{N} \subset \mathbb{R}^4$ plus a constant. We plot it over the pinched torus and the roman surface in $\mathbb{R}^3$ respectively. Bottom two panels: The equilibrium density is the seventh eigenfunction of the Laplace Beltrami operator on a Klein bottle $\mathcal{N} \subset \mathbb{R}^4$ plus a constant. We plot it over the pinched torus and the roman surface in $\mathbb{R}^3$ respectively.

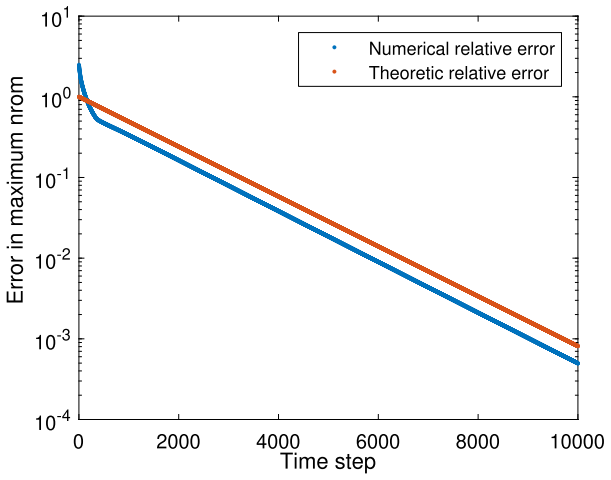


Fig. 7. The semilog-plot comparison between the numerical relative error with theoretic relative error in the Klein bottle example. The numerical relative error is the error from the unconditional stable explicit scheme (3.97) with $\Delta t = 0.05$ and $1 \leq k \leq 10000$. The theoretic relative error is base on (3.109) with $|\lambda_2|^k = 0.9993^k$.

Suppose $\{\mathbf{y}_i\}_{i=1}^{2000}$ are the points on the unit sphere $\mathcal{N} = S^2 \subset \mathbb{R}^3$, i.e., $\{\mathbf{y}_i\}_{i=1}^{2000}$ 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 ρ_i^0 at $\{\mathbf{y}_i\}$, $i = 1, \cdots, 2000$ are extracted from the Pangaea continents map

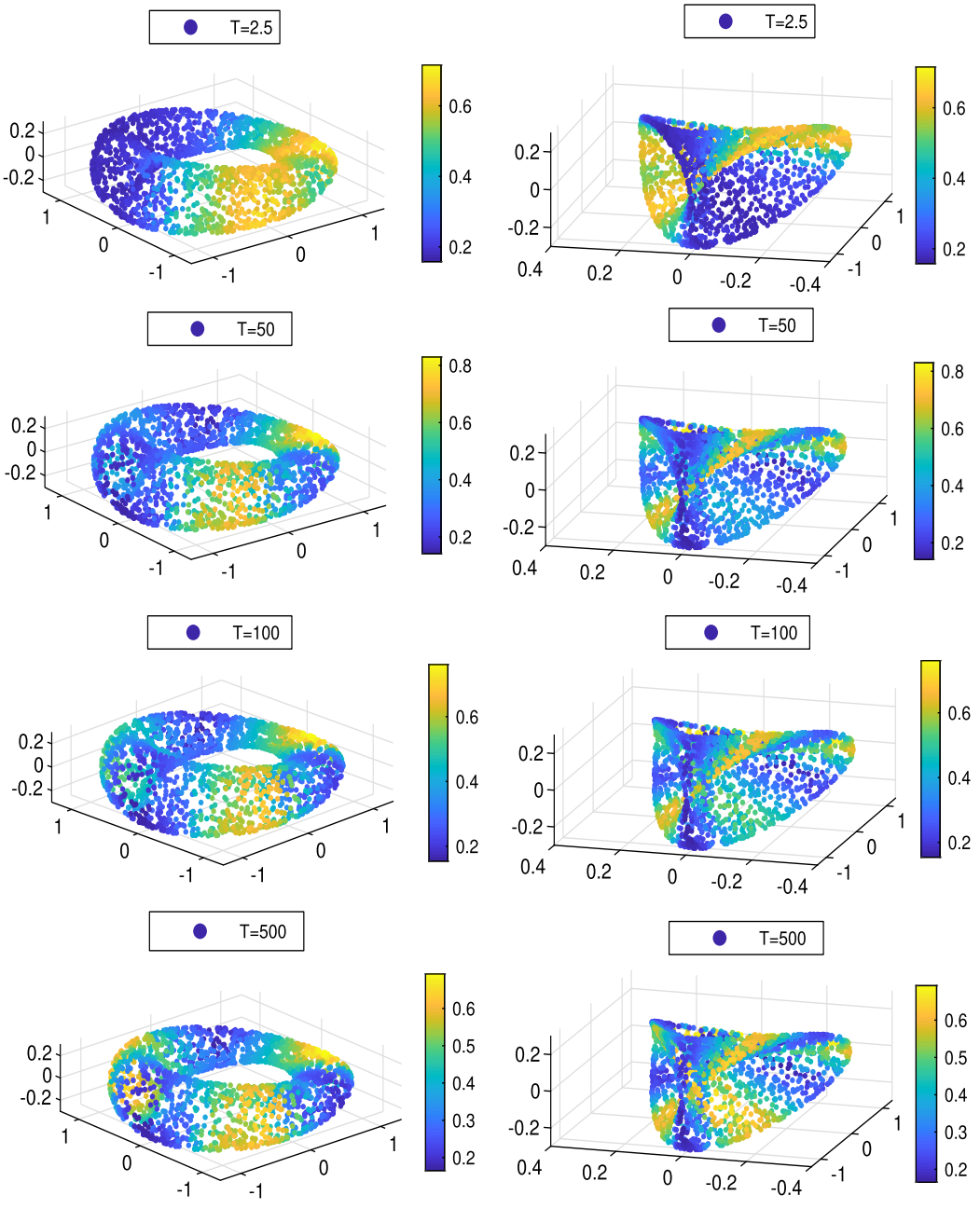


Fig. 8. The density dynamics ρ^k from the unconditional stable explicit scheme (3.97) with $\Delta t = 0.05$. On the left four panels, we plot ρ^k for $k = 50, 1000, 2000, 10000$ corresponding to time $T = 2.5, 50, 100, 500$ on the pinched torus. On the right four panels, we plot ρ^k for $k = 50, 1000, 2000, 10000$ corresponding to time $T = 2.5, 50, 100, 500$ on the Roman surface.

file [1] as shown in Fig. 9 (down left). The value of the initial density $\rho_i^0 \in \{1, 2\}$ where 1 represents oceans and 2 represents continents. Assume the equilibrium $\{\pi_i\}$ at $\{\mathbf{y}_i\}$ are collected from the ETOPO5 topography data [2] expressing the altitude of continents and the depth of oceans for earth. The value of the equilibrium $\{\pi_i\}$ ranges from -7000 to 7000 where the positive values represent the altitude of continents, negative values represent the depth of oceans and 0 represents sea level. Before plugging into the Fokker-Planck equation, we add a constant c_p to π_i such that $\pi_i > 0$ for all i. However, when showing the evolution of continents in figures, we subtract this constant and present the true physical altitudes.

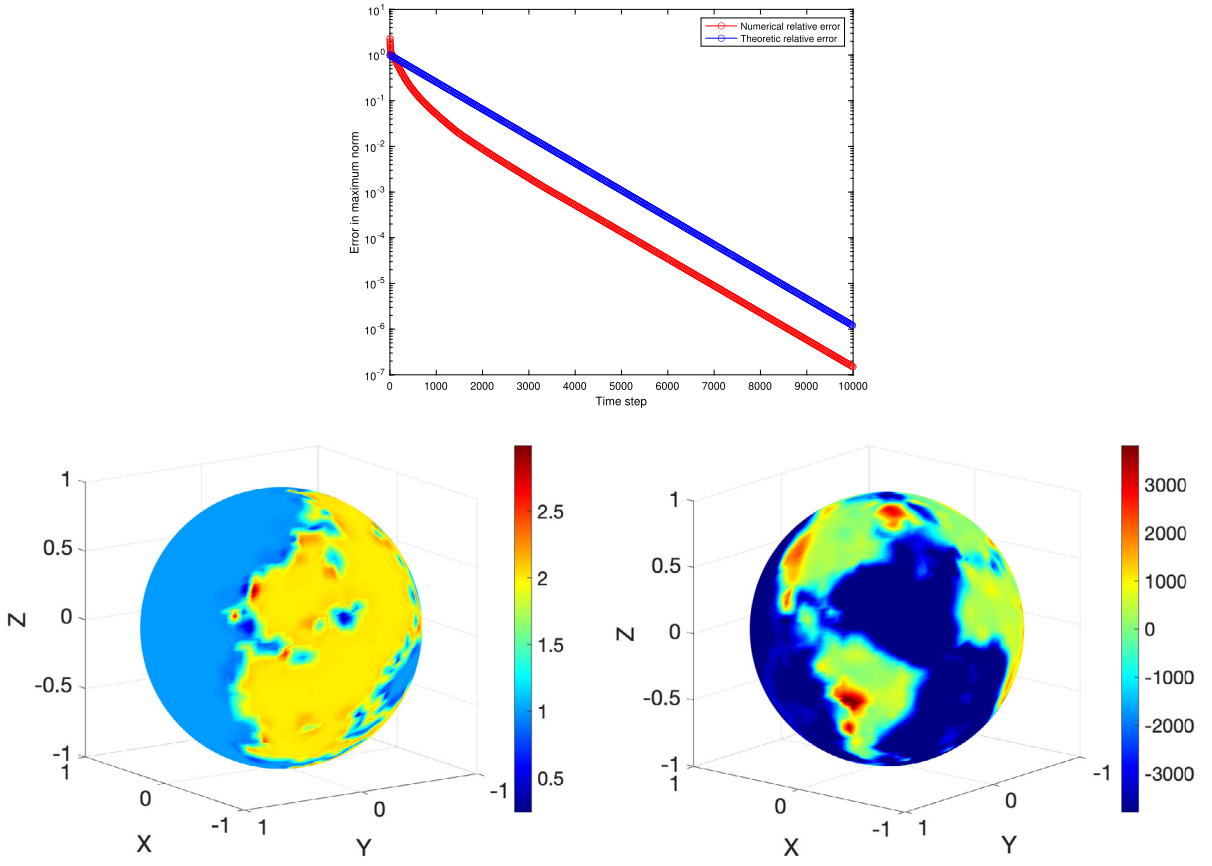


Fig. 9. Simulations for the density dynamics of altitude and depth of continents and oceans starting from Pangaea (down left) to the final altitude of land-ocean (down right) with parameters $dt = 0.05$, $T = 500$. (up) The semilog-plot comparison between the numerical relative error in maximum norm (blue circle) with theoretic relative error $|\lambda_2|^k = 0.9985^k$ (blue line) in (3.109). (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

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 we adjust the initial data, i.e., we replace ρ^0 by $c\rho^0$ such that the total mass condition (3.105) 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 $T = 10000 * \Delta t = 500$. We use the unconditionally stable explicit scheme (3.97) to solve ρ^k. In Fig. 9 (up), the numerical relative error in maximum norm is semilog-plotted using circles. Compared with decay of the theoretic relative error $|\lambda_2|^k$ in (3.109), blue line in the semilog-plot, the exponential convergence rate is exactly same. The initial 3D plot of Pangaea continents is shown in Fig. 9 (down left) while the final 3D plot at $T = 500$ of the simulated altitude and depth of continents and oceans are shown in Fig. 9 (down right).[1] To clearly see the dynamics of altitude and depth of continents and oceans at n points with longitude and latitude, starting from the same Pangaea continents with time step $\Delta t = 0.05$, four snapshots at $T = 0, 1, 10, 75$ of the dynamics are shown in Fig. 10. A video is also provided to show the dynamics of the density https://youtu.be/j5XBPdQhEEs. Here we used nonuniform time intervals since the shapes of continentals (the region with positive altitudes) quickly move from the initial Pangaea supercontinents towards the equilibrium shape of current continents. If we only care about the shape of the continents and keep the binary-valued density during the shape evolution, we refer to [27] for the thresholding adjustment method.

[1] The altitude and depth exceed the range $[-3800\text{ m}, 3800\text{ m}]$ is cut off for clarity.

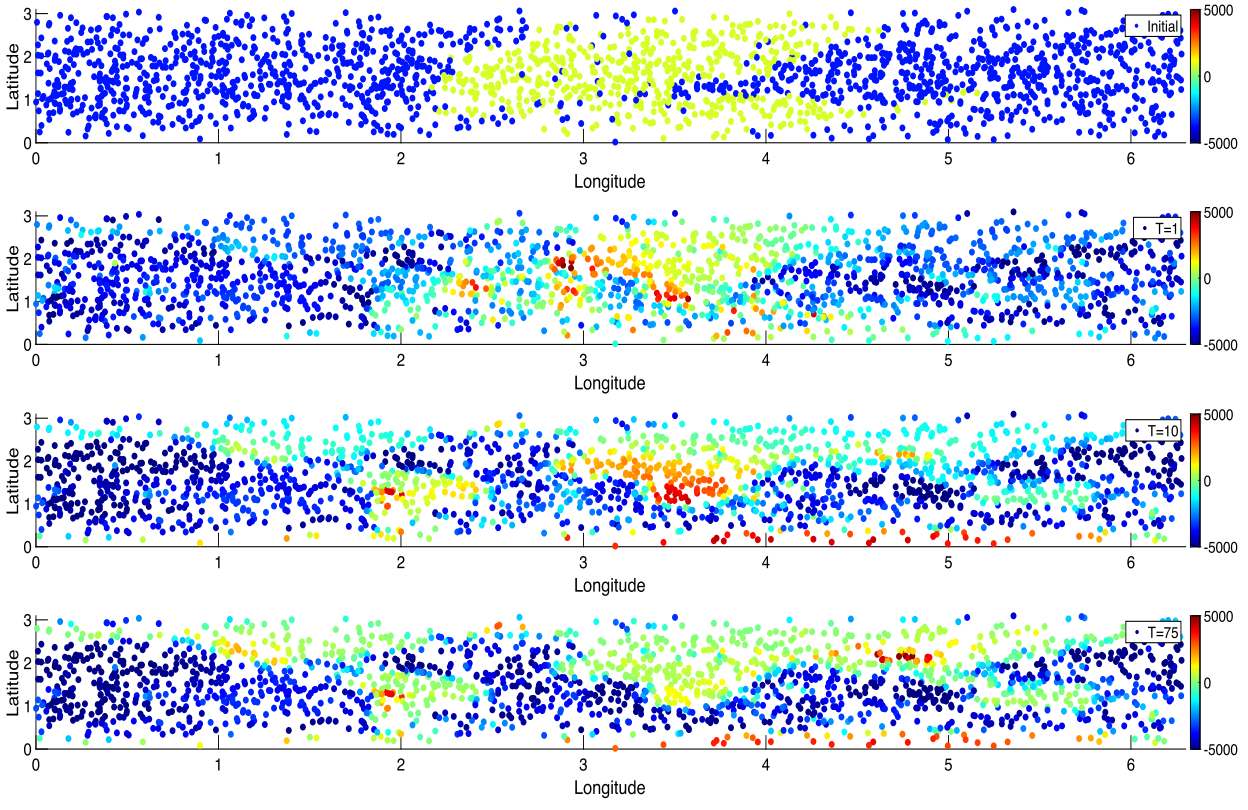


Fig. 10. 2D Snapshots for the density dynamics of altitude and depth of continents and oceans at 2000 points with longitude and latitude starting from Pangaea with parameters $\mathrm{d}t = 0.05$, $T = 0, 1, 10, 75$.

5. 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 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, finite volume scheme for a Fokker-Planck equation which incorporates both the structure of the manifold in the low dimensional space and the equilibrium information. The finite volume scheme defines an approximated Markov process (random walk) on the point cloud with an approximated transition probability and jump rate. We also provide the weighted L^2 convergence analysis of the finite volume scheme to the Fokker-Planck equation on the manifold in the low dimensional space. The efficiency, unconditional stability, and accuracy of the data-driven solver proposed in this paper are justified theoretically. Although we construct several numerical examples to illustrate our data-driven solver, there are still many interesting directions issued from practical problems for future works. An important direction is the manifold-related applications such as the optimal network partitions and the transition path in chemical reactions, especially on the high dimensional practical datasets.

Acknowledgment

Nan Wu thanks the valuable discussion with Professor Hau-Tieng Wu and Chao Shen. Yuan Gao was supported by the National Science Foundation (NSF) under award DMS-2204288. Jian-Guo Liu was supported in part by the National Science Foundation (NSF) under award DMS-2106988.

Appendix A. Theorems about embedding by eigenfunctions of Laplacian

Let Δ 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

$$\Delta\psi_i = -\lambda_i\psi_i, \tag{A.1}$$

where ψ_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 Δ. In [7], the authors provide a theorem about spectral embedding by using all the eigenvalues and eigenfunctions of Δ into the Hilbert space ℓ^2.

Theorem A.1. *(Bérard-Besson-Gallot, [7]) Let M be a d dimensional smooth closed Riemannian manifold. Then, for $\mathbf{x} \in \mathcal{M}$*

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

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

[36] 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, [36]) Let $\mathcal{M}$ be a d dimensional smooth closed Riemannian manifold, for each $\mathbf{x} \in M$, there are $j_1 \leq \cdots \leq j_d$ and the constants $C_1, \cdots, C_d$ such that*

$$\Psi(\mathbf{x}) = (C_1\psi_{j_1}(\mathbf{x}), \cdots, C_d\psi_{j_d}(\mathbf{x})), \tag{A.3}$$

is locally a bi-Lipschitz chart.

Moreover, the next theorem [46] 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.3. *(Portegies, [46]) Let $\mathcal{M}$ be a d dimensional smooth closed Riemannian manifold. Suppose $\mathrm{Ric}_{\mathcal{M}} \geq (d-1)k$, the injectivity radius of $\mathcal{M}$, $i(\mathcal{M}) \geq i_0$ and the volume of $\mathcal{M}$, $\mathrm{Vol}(\mathcal{M}) \leq V$. For any $\epsilon > 0$, there is a $\mathbf{t}_0(\epsilon, d, k, i_0)$ such that for $t < \mathbf{t}_0$, there is $C(\epsilon, d, k, i_0, V, t)$, if $q > C$, then for $\mathbf{x} \in \mathcal{M}$*

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

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 Theorem 2.7, the smallest q that

$$\Psi_1(\mathbf{x}) = (\psi_1(\mathbf{x}), \cdots, \psi_q(\mathbf{x})), \tag{A.5}$$

is a smooth embedding of $\mathcal{M}$ is called the embedding dimension of $\mathcal{M}$. Based on Theorem A.3, the smallest q that

$$\Psi_2(\mathbf{x}) = (2t)^{\frac{d+2}{4}}\sqrt{2}(4\pi)^{\frac{d}{4}}(e^{-\lambda_1 t}\psi_1(\mathbf{x}), \cdots, e^{-\lambda_q t}\psi_q(\mathbf{x})), \tag{A.6}$$

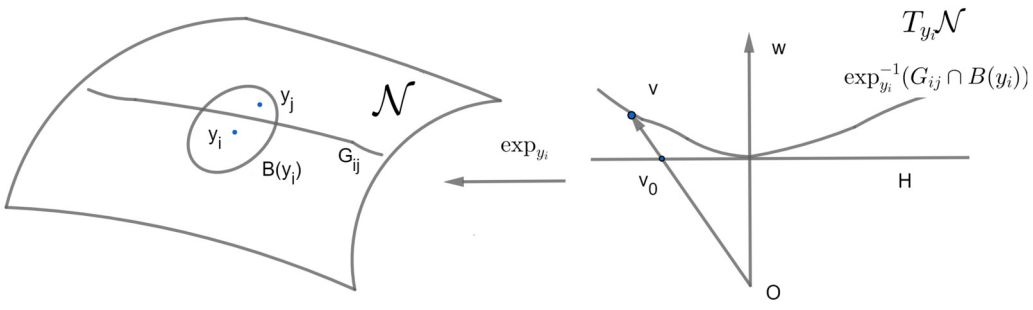


Fig. 11. An illustration to the proof of Proposition 3.2.

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 smalled 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 3.2

Since δ is less than the injectivity radius, there is a Euclidean ball $B_\delta^{T_{y_i}\mathcal{N}}(0)$ of radius δ in the tangent space $T_{y_i}\mathcal{N}$ of $\mathcal{N}$ at $\mathbf{y}_i$ such that the exponential map $\exp_{y_i} : B_\delta^{T_{y_i}\mathcal{N}}(0) \to B_\delta(\mathbf{y}_i)$ is a diffeomorphism. Suppose $y_j = \exp_{y_i}(w)$. We illustrate this setup in Fig. 11. It is sufficient to prove that $\exp_{y_i}^{-1}(B_\delta(\mathbf{y}_i) \cap G_{ij})$ is a $d-1$ dimensional submanifold of $T_{y_i}\mathcal{N}$. For any $v \in \exp_{y_i}^{-1}(B_\delta(\mathbf{y}_i) \cap G_{ij})$, by the definition of the bisector, we have

$$d_{\mathcal{N}}^2(\exp_{y_i}(v), \exp_{y_i}(w)) = d_{\mathcal{N}}^2(\exp_{y_i}(v), y_j) = d_{\mathcal{N}}^2(\exp_{y_i}(v), y_i) = |v|^2. \tag{B.1}$$

Note that

$$d_{\mathcal{N}}^2(\exp_{y_i}(v), \exp_{y_i}(w)) = |v-w|^2 + f(v,w). \tag{B.2}$$

$f(v,w)$ is a smooth function on $B_\delta^{T_{y_i}\mathcal{N}}(0) \times B_\delta^{T_{y_i}\mathcal{N}}(0)$. In particular,

$$f(v,w) = -\frac{1}{3}R_{y_i}(v,w,v,w) + O((|v|^2+|w|^2)^{\frac{5}{2}}) \tag{B.3}$$

for $|v|$ and $|w|$ small, where R_{y_i} is the curvature tensor at y_i. Combine (B.1) and (B.2), we have that

$$|w|^2 - 2v\cdot w + f(v,w) = 0. \tag{B.4}$$

As y_j is a fixed point, we treat w as a fixed vector. Therefore, we use the notation $f_w(x) = f(x,w)$ to indicate f_w is a function of $x \in B_\delta^{T_{y_i}\mathcal{N}}(0)$. Then, we can define a smooth function $F(t,x)$ on $\mathbb{R} \times B_\delta^{T_{y_i}\mathcal{N}}(0)$ as

$$F(t,x) = |w|^2 - 2(1+t)x\cdot w + f_w((1+t)x). \tag{B.5}$$

Let H be the $d-1$ dimensional hyperplane that perpendicularly bisects w in $T_{y_i}\mathcal{N}$. Let v_0 be the vector on H so that $v = (1+t_0)v_0$. By (B.4), $F(t_0,v_0) = 0$. If we can show that $\frac{\partial F(t_0,v_0)}{\partial t} \neq 0$, then by the Implicit Function Theorem, there is a ball B centered at v_0 so that $t = g(x)$ for $x \in B$ and g is differentiable. Hence, $(1+g(x))x$ for $x \in B \cap H$ is a chart for $\exp_{y_i}^{-1}(B_\delta(\mathbf{y}_i) \cap G_{ij})$ around v. We calculate $\frac{\partial F(t_0,v_0)}{\partial t}$:

$$\frac{\partial F(t_0, v_0)}{\partial t} = -2v_0 \cdot w + \nabla f_w(v) \cdot v_0. \tag{B.6}$$

By (B.3), $|\nabla f_w(v)| = O(|v||w|^2)$ where the constant depends on the sectional curvatures at $\mathbf{y}_i$. Since the manifold is compact, the sectional curvatures have upper and lower bounds. Hence,

$$\frac{\partial F(t_0, v_0)}{\partial t} = -2v_0 \cdot w + \nabla f_w(v) \cdot v_0 < -2|v_0||w|\cos(\theta) + |v_0|O(|v||w|^2), \tag{B.7}$$

where θ is the angle between v_0 and w. Since $v_0 \in B_\delta^{T_{y_i}\mathcal{N}}(0) \cap H$, $\cos(\theta) > \frac{|w|}{2\delta}$. $v \in B_\delta^{T_{y_i}\mathcal{N}}(0)$, so $|v| < \delta$. Therefore, when δ is small enough,

$$\frac{\partial F(t_0, v_0)}{\partial t} < |v_0||w|^2(-\frac{1}{\delta} + O(\delta)) < 0. \tag{B.8}$$

Next, we prove the second part of the proposition. $\mathbf{y}^* \in M_{ij}$ follows from the construction. Note that by the triangle inequality and the definition of the bisector, the geodesic sphere centered at $\mathbf{y}_i$ through $\mathbf{y}^*$ is tangent to M_{ij} at $\mathbf{y}^*$. Hence, by Gauss's Lemma, the minimizing geodesic between y_i and y_j is perpendicular to M_{ij} at $\mathbf{y}^*$.

Appendix C. Proof of Proposition 3.12 and Proposition 3.13

We start from a study of the matrix $C_{n,r}(\mathbf{y}_k)$ in Definition 3.8 and relate it to its continuous form. Consider the local covariance matrix $C_{\mathbf{y}_k, B^{\mathbb{R}^\ell}_{\sqrt{r}}(\mathbf{y}_k)\cap\mathcal{N}}$ defined as follows.

$$C_{\mathbf{y}_k, B^{\mathbb{R}^\ell}_{\sqrt{r}}(\mathbf{y}_k)\cap\mathcal{N}} = \int_{B^{\mathbb{R}^\ell}_{\sqrt{r}}(\mathbf{y}_k)\cap\mathcal{N}} (\mathbf{y}-\mathbf{y}_k)(\mathbf{y}-\mathbf{y}_k)^\top \rho^{**}(\mathbf{y}) dV_{\mathcal{N}}(\mathbf{y}) \in \mathbb{R}^{\ell\times\ell}. \tag{C.1}$$

Suppose $C_{\mathbf{y}_k, B^{\mathbb{R}^\ell}_{\sqrt{r}}(\mathbf{y}_k)\cap\mathcal{N}}$ has the following eigendecomposition:

$$C_{\mathbf{y}_k, B^{\mathbb{R}^\ell}_{\sqrt{r}}(\mathbf{y}_k)\cap\mathcal{N}} = U(\mathbf{y}_k)\Lambda(\mathbf{y}_k)U(\mathbf{y}_k)^\top \in O(\ell), \tag{C.2}$$

where $\Lambda(\mathbf{y}_k)$ is a diagonal matrix with the diagonal entries to be eigenvalues of $C_{\mathbf{y}_k, B^{\mathbb{R}^\ell}_{\sqrt{r}}(\mathbf{y}_k)\cap\mathcal{N}}$. Moreover, we have $\Lambda_{11}(\mathbf{y}_k) \geq \Lambda_{22}(\mathbf{y}_k) \geq \cdots \geq \Lambda_{\ell\ell}(\mathbf{y}_k)$. $U(\mathbf{y}_k) \in O(\ell)$ consists of the corresponding orthonormal eigenvectors of $C_{\mathbf{y}_k, B^{\mathbb{R}^\ell}_{\sqrt{r}}(\mathbf{y}_k)\cap\mathcal{N}}$. Intuitively, $C_{\mathbf{y}_k, B^{\mathbb{R}^\ell}_{\sqrt{r}}(\mathbf{y}_k)\cap\mathcal{N}}$ is the continuous form of the matrix $C_{n,r}(\mathbf{y}_k)$.

By setting $\epsilon = \sqrt{r}$ in Proposition 3.2 in [56], we have the following lemma.

Lemma C.1. *Assume that $T_{\mathbf{y}_k}\mathcal{N}$ is generated by the first d standard basis of $\mathbb{R}^\ell$.*

$$\Lambda(\mathbf{y}_k) = \frac{|S^{d-1}|P(\mathbf{y}_k)r^{\frac{d+2}{2}}}{d(d+2)}\Big(\begin{bmatrix} I_{d\times d} & 0 \\ 0 & 0 \end{bmatrix} + O(r)\Big), \tag{C.3}$$

$$U(\mathbf{y}_k) = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} + O(r), \tag{C.4}$$

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

Above lemma says that the first d eigenvectors of $C_{\mathbf{y}_k, B^{\mathbb{R}^\ell}_{\sqrt{r}}(\mathbf{y}_k)\cap\mathcal{N}}$ form an orthonormal basis of $T_{\mathbf{y}_k}\mathcal{N}$ up to an error of order $O(r)$. Note that, for simplicity, we assume $T_{\mathbf{y}_k}\mathcal{N}$ is generated by the first d standard

basis of $\mathbb{R}^\ell$ so that $U(\mathbf{y}_k)$ can be expressed in the above block form. Suppose $C_{n,r}(\mathbf{y}_k)$ has the following eigendecomposition:

$$C_{n,r}(\mathbf{y}_k) = U_n(\mathbf{y}_k)\Lambda_n(\mathbf{y}_k)U_n(\mathbf{y}_k)^\top. \tag{C.5}$$

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

The relation between the eigenstructure of $C_{\mathbf{y}_k, B^{\mathbb{R}^\ell}_{\sqrt{r}}(\mathbf{y}_k)\cap\mathcal{N}}$ and $C_{n,r}(\mathbf{y}_k)$ is discussed in Lemma E.4 in [56].

Lemma C.2. *Assume that $T_{\mathbf{y}_k}\mathcal{N}$ 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 $\mathbf{y}_k$,*

$$\Lambda_n(\mathbf{y}_k) = \Lambda(\mathbf{y}_k) + O(\sqrt{\frac{\log n}{nr^{-\frac{d}{2}-2}}}), \tag{C.6}$$

$$U_n(\mathbf{y}_k) = \begin{bmatrix} X_1' & 0 \\ 0 & X_2' \end{bmatrix} U(\mathbf{y}_k) + O(\sqrt{\frac{\log n}{nr^{\frac{d}{2}-2}}}), \tag{C.7}$$

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

Remark C.3. Above lemma follows from Lemma E.4 in [56] if we choose $\epsilon = \sqrt{r}$ and $\rho \to \infty$ in Case 0 of Lemma E.4 in [56]. In fact, Case 0 of Lemma E.4 in [56] focuses on the first d eigenpairs of the matrix $C_{\mathbf{y}_k, B^{\mathbb{R}^\ell}_{\sqrt{r}}(\mathbf{y}_k)\cap\mathcal{N}}$ of which we need to recover.

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

$$\Lambda_n(\mathbf{y}_k) = \frac{|S^{d-1}|P(\mathbf{y}_k)r^{\frac{d+2}{2}}}{d(d+2)}\begin{bmatrix} I_{d\times d} & 0 \\ 0 & 0 \end{bmatrix} + O(r^{\frac{d}{2}+2}) + O(\sqrt{\frac{\log n}{nr^{-\frac{d}{2}-2}}}), \tag{C.8}$$

$$U_n(\mathbf{y}_k) = \begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix} + O(r) + O(\sqrt{\frac{\log n}{nr^{\frac{d}{2}-2}}}), \tag{C.9}$$

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 C.4. *Assume that $T_{\mathbf{y}_k}\mathcal{N}$ is generated by the first d standard basis of $\mathbb{R}^\ell$. If $\frac{nr^{\frac{d}{2}}}{\log n} \to \infty$ as $n \to \infty$, then with probability greater than $1-\frac{1}{n^2}$, for all $\mathbf{y}_k$,*

$$U_n(\mathbf{y}_k) = \begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix} + O(r), \tag{C.10}$$

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

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

$$\Lambda_n(\mathbf{y}_k) = \frac{|S^{d-1}|P(\mathbf{y}_k)r^{\frac{d+2}{2}}}{d(d+2)}\begin{bmatrix} I_{d\times d} & 0 \\ 0 & 0 \end{bmatrix} + O(r^{\frac{d}{2}+2}), \tag{C.11}$$

$$U_n(\mathbf{y}_k) = \begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix} + O(r), \tag{C.12}$$

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 $\frac{nr^{\frac{d}{2}}}{\log n} \to \infty$ as $n \to \infty$, then we have an approximation of the tangent space of $\mathcal{N}$ at $\mathbf{y}_k$, i.e. the first d eigenvectors of $C_{n,r}(\mathbf{y}_k)$ are the basis of $T_{\mathbf{y}_k}\mathcal{N}$ up to an error of order $O(r)$. If n and r satisfy $\frac{nr^{\frac{d}{2}+2}}{\log n} \to \infty$ as $n \to \infty$, the first d eigenvectors of $C_{n,r}(\mathbf{y}_k)$ are the basis of $T_{\mathbf{y}_k}\mathcal{N}$ up to an error of order $O(r)$. Moreover, there are d significantly large eigenvalues of $C_{n,r}(\mathbf{y}_k)$ which are close to the first d eigenvalues of $C_{\mathbf{y}_k, B^{\mathbb{R}^\ell}_{\sqrt{r}}(\mathbf{y}_k)\cap\mathcal{N}}$.

Next, we show that the map $\tilde{\iota}_k$ in the Definition 3.8 restricted on $B^{\mathbb{R}^\ell}_r(\mathbf{y}_k)\cap\mathcal{N}$ is a $1+O(r)$ bi-Lipschitz homeomorphism.

Lemma C.5. *Suppose* $r \to 0$ *and* $\frac{nr^{\frac{d}{2}}}{\log n} \to \infty$ *as* $n \to \infty$. *Suppose* r *is small enough, then with probability greater than* $1 - \frac{1}{n^2}$, *for all* $\mathbf{y}_k$ *and any* $\mathbf{y}, \mathbf{y}' \in B^{\mathbb{R}^\ell}_r(\mathbf{y}_k)\cap\mathcal{N}$, *we have*

$$\|\tilde{\iota}_k(\mathbf{y}') - \tilde{\iota}_k(\mathbf{y})\|_{\mathbb{R}^d} = \|\iota_k(\mathbf{y}' - \mathbf{y})\|_{\mathbb{R}^d} = d_{\mathcal{N}}(\mathbf{y}, \mathbf{y}')(1 + O(r)). \tag{C.13}$$

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

$$\mathbf{y} = [\![v, v^\perp]\!] \in \mathbb{R}^\ell, \tag{C.14}$$

where $v \in T_{\mathbf{y}_k}\mathcal{N}$ forms the first d coordinates of y and $v^\perp \in T^\perp_{\mathbf{y}_k}\mathcal{N}$ forms the last $\ell - d$ coordinates of $\mathbf{y}$. For any $\mathbf{y}, \mathbf{y}' \in B^{\mathbb{R}^\ell}_r(\mathbf{y}_k)\cap\mathcal{N}$, suppose $\mathbf{y} = [\![v_1, v_1^\perp]\!]$ and $\mathbf{y}' = [\![v_2, v_2^\perp]\!]$. Due to the manifold structure of $\mathcal{N}$, we have

$$\|v_1^\perp - v_2^\perp\|_{\mathbb{R}^{\ell-d}} \le \mathcal{C}_1 r \|v_1 - v_2\|_{\mathbb{R}^d}, \tag{C.15}$$

for some constant $\mathcal{C}_1$ depending on the curvature of $\mathcal{N}$. Hence,

$$\|v_1 - v_2\|_{\mathbb{R}^d} \le \|\mathbf{y}' - \mathbf{y}\|_{\mathbb{R}^\ell} \le \|v_1 - v_2\|_{\mathbb{R}^d}\sqrt{1 + \mathcal{C}_1^2 r^2}, \tag{C.16}$$

which is equivalent to

$$\|v_1 - v_2\|_{\mathbb{R}^d} = \|\mathbf{y}' - \mathbf{y}\|_{\mathbb{R}^\ell}(1 + O(r^2)). \tag{C.17}$$

Moreover, suppose $\{\beta_{n,r,1}, \cdots, \beta_{n,r,d}\}$ are orthonormal eigenvectors corresponding to $C_{n,r}(\mathbf{y}_k)$'s largest d eigenvalues. Then, by Proposition C.4

$$\beta_{n,r,1} = [\![\beta_i, 0]\!] + O(r), \tag{C.18}$$

where $\{\beta_i\}_{i=1}^d$ form an orthonormal basis of $T_{\mathbf{y}_k}\mathcal{N} \approx \mathbb{R}^d$.

$$\begin{aligned} \|\iota_k(\mathbf{y}' - \mathbf{y})\|_{\mathbb{R}^d} =& \|v_1 - v_2\|_{\mathbb{R}^d} + \|\mathbf{y}' - \mathbf{y}\|_{\mathbb{R}^\ell} O(r) \\ =& \|\mathbf{y}' - \mathbf{y}\|_{\mathbb{R}^\ell}(1 + O(r^2)) + \|\mathbf{y}' - \mathbf{y}\|_{\mathbb{R}^\ell} O(r) = \|\mathbf{y}' - \mathbf{y}\|_{\mathbb{R}^\ell}(1 + O(r)), \end{aligned} \tag{C.19}$$

where we apply (C.17) in the second last step.

By equation (3.74), we know that $d_{\mathcal{N}}(y, \mathbf{y}') \leq 2D_1 r$. Hence, by Lemma 3.9,

$$\begin{aligned}\|\iota_k(\mathbf{y}' - \mathbf{y})\|_{\mathbb{R}^d} =& \|\mathbf{y}' - \mathbf{y}\|_{\mathbb{R}^\ell}(1 + O(r)) = d_{\mathcal{N}}(\mathbf{y}, \mathbf{y}')(1 + O(d^2_{\mathcal{N}}(\mathbf{y}, \mathbf{y}')))(1 + O(r)) \\ =& d_{\mathcal{N}}(\mathbf{y}, \mathbf{y}')(1 + O(r^2))(1 + O(r)) = d_{\mathcal{N}}(\mathbf{y}, \mathbf{y}')(1 + O(r)). \quad \square\end{aligned} \tag{C.20}$$

We introduce the following notations to prove the following lemma and proposition. Denote the boundary of C_k by ∂C_k. Denote the boundary of $\tilde{\iota}_k(C_k)$ by $\partial\tilde{\iota}_k(C_k) = \tilde{\iota}_k(\partial C_k)$. Let $\tilde{C}_{k,0}$ be the Voronoi cell in $\mathbb{R}^d$ containing 0 constructed in the Step 4 in Algorithm 1. Denote the boundary of $\tilde{C}_{k,0}$ by $\partial\tilde{C}_k$. Denote $d^{\mathbb{R}^d}_{\mathcal{H}}(S_1, S_2)$ be the Hausdorff distance between two sets S_1 and S_2 in $\mathbb{R}^d$ with respect to the Euclidean metric.

Lemma C.6. *If n is large enough, for r satisfying Assumption 3.10, with probability greater than $1 - \frac{1}{n^2}$, for all $\mathbf{y}_k$, $d^{\mathbb{R}^d}_{\mathcal{H}}(\partial\tilde{\iota}_k(C_k), \partial\tilde{C}_k) = O(r^2)$.*

Proof. For simplicity, in this proof, we use $|\cdot|$ to denote $\|\cdot\|_{\mathbb{R}^d}$. Based on Assumption 3.10, the requirement of Lemma C.5 holds. Recall that in Assumption 3.10, we assume $B^{\mathbb{R}^\ell}_r(\mathbf{y}_k) \cap \{\mathbf{y}_i\}_{i=1}^n = \{\mathbf{y}_{k,1}, \cdots, \mathbf{y}_{k,N_k}\}$. We have $C_k \subset B^{\mathbb{R}^\ell}_r(\mathbf{y}_k)$. Moreover, if Γ_{kj} is a Voronoi surface of C_k between $\mathbf{y}_k$ and $\mathbf{y}_j$, then $\mathbf{y}_j \in B^{\mathbb{R}^\ell}_r(\mathbf{y}_k)$. We denote $\Gamma_{k,i}$ to be the Voronoi face between $\mathbf{y}_k$ and $\mathbf{y}_{k,i}$.

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

Case 1: $v \in \tilde{C}_{k,0}$

Suppose $v = \tilde{\iota}_k(\mathbf{y})$ for some $\mathbf{y} \in \partial C_k$. Moreover $\mathbf{y} \in \Gamma_{k,i}$. In other word, $\mathbf{y}$ is on the Voronoi face between $\mathbf{y}_k$ and $\mathbf{y}_{k,i} \in B^{\mathbb{R}^\ell}_r(\mathbf{y}_k)$. As shown in Fig. 12(a), let O be the origin in $\mathbb{R}^d$. Let $A = v$ and $B = \tilde{\iota}_k(\mathbf{y}_{k,i})$. 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

$$d_{\mathbb{R}^d}(A, H) = |CM| = \frac{|CB| - |CO|}{2} = \frac{|CB|^2 - |CO|^2}{2(|CB| + |CO|)} \tag{C.21}$$

$$= \frac{|AB|^2 - |AC|^2 - (|AO|^2 - |AC|^2)}{2|BO|} = \frac{(|AB| + |AO|)(|AB| - |AO|)}{2|BO|} \tag{C.22}$$

Since $\mathbf{y} \in \Gamma_{k,i}$, by equation (3.74), $d_{\mathcal{N}}(\mathbf{y}, \mathbf{y}_{k,i}) = d_{\mathcal{N}}(\mathbf{y}, \mathbf{y}_k) = a \leq D_1 r$. By Lemma C.5, $|AB| = a(1 + O(r))$ and $|AO| = a(1 + O(r))$, hence $(|AB| + |AO|)(|AB| - |AO|) = 2a^2 O(r) \leq 2D_1^2 O(r^3)$. By Lemma C.5, $|BO| = d_{\mathcal{N}}(\mathbf{y}_k, \mathbf{y}_{k,i})(1 + O(r))$. By (2) in Assumption 3.10, $d_{\mathcal{N}}(\mathbf{y}_k, \mathbf{y}_{k,i}) \geq D_2 r$. Hence,

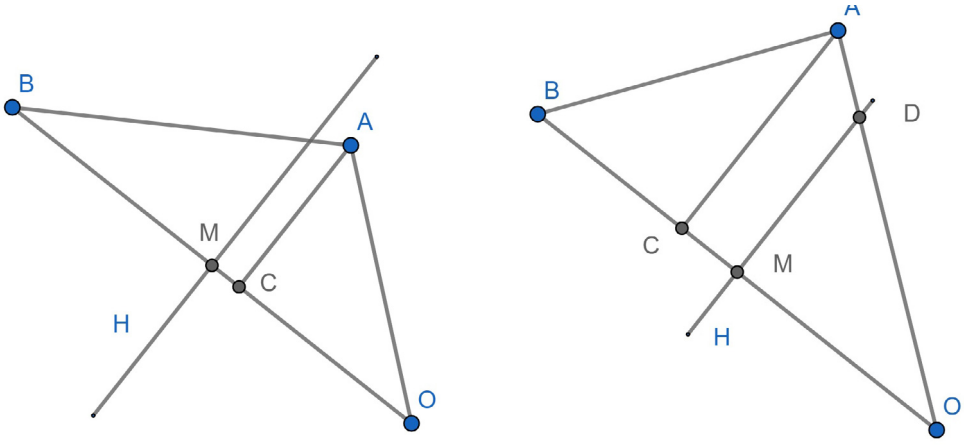
$$d_{\mathbb{R}^d}(A, H) = \frac{(|AB| + |AO|)(|AB| - |AO|)}{2|BO|} \leq \frac{D_1^2}{D_2} O(r^2) = O(r^2). \tag{C.23}$$

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 $v = \tilde{\iota}_k(\mathbf{y})$ for some $\mathbf{y} \in \partial C_k$. As shown in Fig. 12(b), let O be the origin in $\mathbb{R}^d$. Let $A = v$. Suppose OA intersects with $\partial\tilde{C}_k$ at D. $D \in \tilde{F}_{k,j}$, where $\tilde{F}_{k,j}$ is Voronoi face in $\mathbb{R}^d$ between O and $B = \tilde{\iota}_k(\mathbf{y}_{k,j})$. H is the hyperplane that perpendicularly bisects OB. M is the intersection between H and OB. Note that $\tilde{F}_{k,j} \subset H$. 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

$$|CM| = \frac{|CO| - |CB|}{2} = \frac{|CO|^2 - |CB|^2}{2(|CB| + |CO|)} \tag{C.24}$$



(a) The Case when $v = A \in \tilde{C}_{k,0}$ (b) The Case when $v = A \notin \tilde{C}_{k,0}$

Fig. 12. Illustrations of Case 1 and Case 2 in the proof of Lemma C.6.

$$= \frac{|AO|^2 - |AC|^2 - (|AB|^2 - |AC|^2)}{2|BO|} = \frac{(|AB| + |AO|)(|AO| - |AB|)}{2|BO|}. \tag{C.25}$$

$\mathbf{y} \in \partial C_k$ but we may not have $\mathbf{y} \in \Gamma_{k,j}$, therefore, $a = d_{\mathcal{N}}(\mathbf{y}, \mathbf{y}_{k,j}) \geq d_{\mathcal{N}}(\mathbf{y}, \mathbf{y}_k) = b$. $d_{\mathcal{N}}(\mathbf{y}, \mathbf{y}_{k,j}) \leq d_{\mathcal{N}}(\mathbf{y}, \mathbf{y}_k) + d_{\mathcal{N}}(\mathbf{y}_{k,j}, \mathbf{y}_k)$, hence by equation (3.74), $b \leq D_1 r$ and $a \leq 2D_1 r$. By Lemma C.5, $|AB| = a(1 + O(r))$ and $|AO| = b(1 + O(r))$. Since $a \geq b$ and $|AB| \leq |AO|$, we have $0 \leq a - b = D_1 O(r^2)$. Hence, $|AO| - |AB| = D_1 O(r^2)$. By Lemma C.5 and Assumption 3.10,

$$|BO| = d_{\mathcal{N}}(\mathbf{y}_k, \mathbf{y}_{k,j})(1 + O(r)) \geq D_2 r(1 + O(r)), \tag{C.26}$$

$$|AO| = d_{\mathcal{N}}(\mathbf{y}, \mathbf{y}_k)(1 + O(r)) \leq D_1 r(1 + O(r)). \tag{C.27}$$

Hence, $(|AB| + |AO|)(|AB| - |AO|) \leq 2|AO|(|AB| - |AO|) \leq 2D_1^2 O(r^3)$. Moreover,

$$|CM| = \frac{(|AB| + |AO|)(|AB| - |AO|)}{2|BO|} \leq \frac{D_1^2}{D_2} O(r^2) = O(r^2). \tag{C.28}$$

At last,

$$|AD| = \frac{|CM||AO|}{|OC|} \leq \frac{|CM||AO|}{|OM|} = \frac{2|CM||AO|}{|OB|} \leq \frac{2|CM|D_1 r(1 + O(r))}{D_2 r(1 + O(r))} = O(r^2). \tag{C.29}$$

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

In the second step, we show that for any $v \in \tilde{C}_k$, $d_{\mathbb{R}^d}(v, \partial\tilde{\iota}_k(C_k)) = O(r^2)$. 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. By Assumption 3.10, for any $\mathbf{y} \in \partial C_k$, $\frac{1}{2} D_2 r \leq d_{\mathcal{N}}(\mathbf{y}, \mathbf{y}_k) \leq D_1 r$. By Lemma C.5, any for $v \in \partial\tilde{\iota}_k(C_k)$, $\frac{1}{2} D_2 r(1 + O(r)) \leq \|v\|_{\mathbb{R}^d} \leq D_1 r(1 + O(r))$. Hence, $\frac{1}{2} D_2 r + O(r^2) \leq \|v\|_{\mathbb{R}^d} \leq D_1 r + O(r^2)$. By Lemma C.6 and the triangle inequality, for any $v' \in \partial\tilde{C}_k$, $\frac{1}{2} D_2 r + O(r^2) \leq \|v'\|_{\mathbb{R}^d} \leq D_1 r + O(r^2)$. Since $\tilde{C}_{k,0}$ is convex, we conclude that there is a constant Ω such that $|\tilde{C}_{k,0}| = \Omega r^d + O(r^{d+1})$. By Lemma C.6 and the fact that $\tilde{C}_{k,0}$ is convex, $|\tilde{\iota}_k(C_k)| = \Omega r^d + O(r^{d+1}) = |\tilde{C}_{k,0}|(1 + O(r))$. By Lemma C.5, $|C_k| = |\tilde{\iota}_k(C_k)|(1 + O(r))^d = |\tilde{\iota}_k(C_k)|(1 + O(r))$. Therefore, $|\tilde{C}_k| = |\tilde{C}_{k,0}| = |C_k|(1 + O(r))$. □

Proof of Proposition 3.13. We provide a sketch of the proof. Use $|\cdot|$ to denote the $d-1$ dimensional Hausdorff measure. ∂A denotes the topological boundary of a set A. Suppose $B_r^{\mathbb{R}^\ell}(\mathbf{y}_k) \cap \{\mathbf{y}_i\}_{i=1}^n = \{\mathbf{y}_{k,1}, \cdots, \mathbf{y}_{k,N_k}\}$. Suppose $\Gamma_{k,i}$ is the Voronoi face between $\mathbf{y}_k$ and $\mathbf{y}_{k,i}$.

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 $\mathbf{y}$ and $\mathbf{y}_{k,i}$ at $\mathbf{y}_{k,i}^*$. Then, by Assumption 3.10 and Proposition 3.2, there is a $d-1$ dimensional subspace $S_{k,i}$ of $T_{\mathbf{y}_{k,i}^*}\mathcal{N}$ which is perpendicular to the tangent vector of the minimizing geodesic at $\mathbf{y}_{k,i}^*$. If we realize $T_{\mathbf{y}_{k,i}^*}\mathcal{N}$ as a subspace of $\mathbb{R}^\ell$, then the affine subspace $\mathbf{y}_{k,i}^* + S_{k,i}$ is tangent to G at $\mathbf{y}_{k,i}^*$. Without loss of generality, we rotate and translate the manifold $\mathcal{N}$ so that $\mathbf{y}_{k,i}^* = 0$ and $S_{k,i}$ is identified with the subspace of $\mathbb{R}^\ell$ generated by the first $d-1$ standard basis. By Assumption 3.10 and Proposition 3.2, there is an open subset of $S_{k,i}$ and denote $L_{k,i}$ to be its closure such that for any $\mathbf{y} \in \Gamma_{k,i}$, we have

$$\mathbf{y} = (u, g_1(u), \cdots, g_{\ell-d+1}(u)), \tag{C.30}$$

where $u \in L_{k,i} \subset \mathbb{R}^{d-1}$ and $g_i : \mathbb{R}^{d-1} \to \mathbb{R}$. Moreover, $g_j(u)$ is smooth and $g_j(0) = 0$ and $\nabla g_j(0) = 0$. The second order derivative of g_i can be bounded by the curvature of $\mathcal{N}$ at $\mathbf{y}_k$ and $\mathbf{y}_{k,i}$. By (1) in Assumption 3.10, $\Gamma_{k,i} \subset C_k \subset B_r^{\mathbb{R}^\ell}(\mathbf{y}_k)$, hence for any $\mathbf{y} \in \Gamma_{k,i}$, $\|\mathbf{y} - \mathbf{y}_k\|_{\mathbb{R}^\ell} \leq r$. $\|\mathbf{y}_{k,i}^* - \mathbf{y}_k\|_{\mathbb{R}^\ell} \leq d_{\mathcal{N}}(\mathbf{y}_{k,i}^*, \mathbf{y}_k) = \frac{1}{2} d_{\mathcal{N}}(\mathbf{y}_{k,i}, \mathbf{y}_k)) \leq \frac{1}{2} D_1 r$. Since $\mathbf{y}_{k,i} = 0$,

$$\|\mathbf{y}\|_{\mathbb{R}^\ell} = \|\mathbf{y} - \mathbf{y}_{k,i}^*\|_{\mathbb{R}^\ell} \leq \|\mathbf{y} - \mathbf{y}_k\|_{\mathbb{R}^\ell} + \|\mathbf{y}_{k,i}^* - \mathbf{y}_k\|_{\mathbb{R}^\ell} \leq (1 + \frac{1}{2} D_1) r. \tag{C.31}$$

By (C.30), for any $u \in L_{k,i}$, $\|u\|_{\mathbb{R}^{d-1}} \leq \|\mathbf{y}\|_{\mathbb{R}^\ell} \leq (1 + \frac{1}{2} D_1) r$. 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

$$|L_{k,i}| \leq |S^{d-1}| (1 + \frac{1}{2} D_1)^{d-1} r^{d-1} \tag{C.32}$$

(C.30) implies that

$$|\Gamma_{k,i}| = |L_{k,i}| + O(|L_{k,i}|^{\frac{d}{d-1}}) = |L_{k,i}| + O(r^d), \tag{C.33}$$

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,

$$d_{\mathcal{H}}^{\mathbb{R}^\ell}(\partial \Gamma_{k,i}, \partial L_{k,i}) = \max_{u \in \partial L_{k,i}} \sqrt{g_1^2(u) + \cdots, g_{\ell-d+1}^2(u)} = O(r^2), \tag{C.34}$$

where $d_{\mathcal{H}}^{\mathbb{R}^\ell}$ is the Hausdorff distance with respect to the Euclidean metric of $\mathbb{R}^\ell$.

Step 2

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

Step 3

For simplicity, denote $\tilde{\iota}_k(\mathbf{y}_{k,i}^* + L_{k,i})$ by $\tilde{\iota}_k(L_{k,i})$, denote $\tilde{\iota}_k(\mathbf{y}_{k,i}^* + \partial L_{k,i})$ by $\tilde{\iota}_k(\partial L_{k,i})$ and denote $\partial \tilde{\iota}_k(\mathbf{y}_{k,i}^* + L_{k,i})$ by $\partial \tilde{\iota}_k(L_{k,i})$. Since $\tilde{\iota}_k$ restricted on the affine subspace $\mathbf{y}_{k,i}^* + T_{\mathbf{y}_{k,i}^*}\mathcal{N}$ is homeomorphism, $\partial \tilde{\iota}_k(L_{k,i}) =$

$\tilde{\iota}_k(\partial L_{k,i})$. Moreover, Lemma C.5 shows that $\tilde{\iota}_k$ restricted on $B_r^{\mathbb{R}^\ell}(\mathbf{y}_k) \cap \mathcal{N}$ is a homeomorphism. Hence, $\partial\tilde{\iota}_k(\Gamma_{k,i}) = \tilde{\iota}_k(\partial\Gamma_{k,i})$. Since $\tilde{\iota}_k$ is a projection,

$$d_{\mathcal{H}}^{\mathbb{R}^d}(\partial\tilde{\iota}_k(\Gamma_{k,i}), \partial\tilde{\iota}_k(L_{k,i})) = d_{\mathcal{H}}^{\mathbb{R}^d}(\tilde{\iota}_k(\partial\Gamma_{k,i}), \tilde{\iota}_k(\partial L_{k,i})) \leq d_{\mathcal{H}}^{\mathbb{R}^\ell}(\partial\Gamma_{k,i}, \partial L_{k,i}) = O(r^2), \tag{C.35}$$

where we use (C.34) in the last step. Since $\tilde{\iota}_k$ is a projection, $\tilde{\iota}_k(L_{k,i})$ is a subset of a $d-1$ dimensional affine subspace of $\mathbb{R}^d$. Since $\tilde{\iota}_k$ restricted on the affine subspace $\mathbf{y}_{k,i}^* + T_{\mathbf{y}_{k,i}^*}\mathcal{N}$ is a $1+O(r)$ bi-Lipschitz homeomorphism,

$$|L_{k,i}| = |\tilde{\iota}_k(L_{k,i})|(1+O(r)). \tag{C.36}$$

Step 4

Recall in the step (4) in Algorithm 1, we find the Voronoi cell decomposition of $\{0, \tilde{\iota}_k(\mathbf{y}_{k,1}), \cdots, \tilde{\iota}_k(\mathbf{y}_{k,N_k})\}$ in $\mathbb{R}^d$. The Voronoi cell containing 0 is $\tilde{C}_{k,0}$. The Voronoi face between 0 and $\tilde{\iota}_k(\mathbf{y}_{k,i})$ is denoted as $\tilde{F}_{k,i}$. If $\mathbf{y} \in \partial\Gamma_{k,i}$, then there is a third point $\mathbf{y}_{k,j}$ such that $d_{\mathcal{N}}(\mathbf{y}, \mathbf{y}_k) = d_{\mathcal{N}}(\mathbf{y}, \mathbf{y}_{k,i}) = d_{\mathcal{N}}(\mathbf{y}, \mathbf{y}_{k,j})$. By using the similar argument in Lemma C.6, we can show that

$$d_{\mathcal{H}}^{\mathbb{R}^d}(\partial\tilde{\iota}_k(\Gamma_{k,i}), \partial\tilde{F}_{k,i}) = O(r^2). \tag{C.37}$$

By (C.35), we have

$$d_{\mathcal{H}}^{\mathbb{R}^d}(\partial\tilde{\iota}_k(L_{k,i}), \partial\tilde{F}_{k,i}) = O(r^2). \tag{C.38}$$

Step 5

By (1) in Assumption 3.10, $C_k \subset B_r^{\mathbb{R}^\ell}(\mathbf{y}_k) \cap \mathcal{N}$. Since $\tilde{\iota}_k$ is a projection, $\tilde{\iota}_k(C_k)$ is in the ball of radius r centered at 0 in $\mathbb{R}^d$. By Lemma C.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 $\partial\tilde{F}_{k,i} = \cup_j \mathcal{C}_j$, where each $\mathcal{C}_j$ is a $d-2$ dimensional convex polygon. Each $\mathcal{C}_j$ 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 $d_{\mathcal{H}}^{\mathbb{R}^d}(\partial\tilde{\iota}_k(L_{k,i}), \partial\tilde{\Gamma}_{k,i}) = O(r^2)$ and $\tilde{\iota}_k(L_{k,i})$ is a subset of a $d-1$ dimensional affine subspace of $\mathbb{R}^d$, we rotate and translate $\tilde{\iota}_k(L_{k,i})$ so that $\partial\tilde{\iota}_k(L_{k,i})$ is in a $O(r^2)$ neighborhood of $\partial\tilde{\Gamma}_{k,i}$ in $H_{k,i}$. Therefore,

$$|\tilde{\iota}_k(L_{k,i})| = |\tilde{F}_{k,i}| + O(r^d) \tag{C.39}$$

Combine (C.32), (C.33), (C.36) and (C.39), we have

$$|\Gamma_{k,i}| = |\tilde{F}_{k,i}| + O(r^d). \tag{C.40}$$

If $\mathbf{y}_\ell = \mathbf{y}_{k,i} \in B_r^{\mathbb{R}^\ell}(\mathbf{y}_k)$, then $\tilde{A}_{kl} = |\tilde{F}_{k,i}|$. So, $|\Gamma_{k\ell}| = \tilde{A}_{kl} + O(r^d)$. Similarly, $|\Gamma_{\ell k}| = |\Gamma_{k\ell}| = \tilde{A}_{kl} + O(r^d)$. Hence, $|\Gamma_{k\ell}| = \frac{\tilde{A}_{kl}+\tilde{A}_{kl}}{2} + O(r^d) = A_{k\ell} + O(r^d)$. 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 $|\Gamma_{k\ell}| = |\tilde{\Gamma}_{k\ell}| + O(r^d)$. □

Appendix D. Other standard time discretizations

Lemma D.1 below gives the maximum principle, and exponential convergence for an explicit scheme under Courant-Friedrichs-Lewy (CFL) condition. Lemma D.2 below gives the unconditional maximum principle, and exponential convergence for an implicit scheme.

Lemma D.1. *Let $\tilde{\eta}_i$ be the approximated jump rate and $\tilde{P}_{ij}$ be the approximated transition probability defined in (3.80). Let Δt be the time step and consider the explicit scheme for (3.79)*

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

With the detailed balance property (3.81), and the CFL condition for Δt

$$\Delta t \le \frac{1}{\tilde{\eta}_i} = \frac{2|\tilde{C}_i|\pi_i}{\sum_{j\in VF(i)} \frac{\pi_i+\pi_j}{|y_i-y_j|}|\tilde{\Gamma}_{ij}|}, \tag{D.2}$$

we have

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

$$\sum_i \rho_i^{k+1}|\tilde{C}_i| = \sum_i \rho_i^k|\tilde{C}_i|; \tag{D.3}$$

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

$$u^{k+1} = (I + \Delta t Q)u^k, \quad Q := \{b_{ij}\} \text{ with } b_{ij} := \begin{cases} -\tilde{\eta}_i, & j = i; \\ \tilde{\eta}_i\tilde{P}_{ji}, & j \neq i; \end{cases} \tag{D.4}$$

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

$$\max_i j\frac{\rho_j^{k+1}}{\pi_j} \le \max_j \frac{\rho_j^k}{\pi_j}. \tag{D.5}$$

(iv) the ℓ^∞ contraction

$$\max_i \left| \frac{\rho_i^{k+1}}{\pi_i} - 1 \right| \le \max_i \left| \frac{\rho_i^k}{\pi_i} - 1 \right|; \tag{D.6}$$

(v) the exponential convergence

$$\left\| \frac{\rho_i^k}{\pi_i} - 1 \right\|_{\ell^\infty} \le c|\lambda_2|^k, \quad |\lambda_2| < 1, \tag{D.7}$$

where λ_2 is the second eigenvalue (in terms of the magnitude) of $(I + \Delta t Q)$.

Lemma D.2. *Let $\tilde{\eta}_i$ be the approximated jump rate and $\tilde{P}_{ij}$ be the approximated transition probability defined in (3.80). Let Δt be the time step and consider the implicit scheme*

$$\frac{\rho_i^{n+1}}{\pi_i} = \frac{\rho_i^n}{\pi_i} - \tilde{\eta}_i\Delta t\frac{\rho_i^{n+1}}{\pi_i} + \Delta t \sum_{j\in VF(i)} \tilde{\eta}_i\tilde{P}_{ji}\frac{\rho_j^{n+1}}{\pi_j}. \tag{D.8}$$

We have the following unconditional properties:

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

$$\sum_i \rho_i^{k+1}|\tilde{C}_i| = \sum_i \rho_i^k|\tilde{C}_i|; \tag{D.9}$$

(ii) the equivalent updates for $u_i^{k+1} = \frac{\rho_i^{k+1}}{\pi_i}$ *with same* Q *in* (D.4)

$$(I - \Delta t Q)u^{k+1} = u^k; \tag{D.10}$$

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

$$\max_i \frac{\rho_j^{k+1}}{\pi_j} \le \max_j \frac{\rho_j^k}{\pi_j}. \tag{D.11}$$

(iv) the ℓ^∞ *contraction*

$$\max_i \left|\frac{\rho_i^{k+1}}{\pi_i} - 1\right| \le \max_i \left|\frac{\rho_i^k}{\pi_i} - 1\right|; \tag{D.12}$$

(v) the exponential convergence

$$\left\|\frac{\rho_i^k}{\pi_i} - 1\right\|_{\ell^\infty} \le c|\lambda_2|^k, \quad |\lambda_2| < 1, \tag{D.13}$$

where λ_2 *is the second eigenvalue (in terms of the magnitude) of* $(I - \Delta t Q)^{-1}$.

The proof of the two lemmas is same as Proposition 3.15 and we omit it. The advantage of the explicit scheme (D.1) is its efficiency but the disadvantage is the requirement of CFL condition on Δt; see (D.2). Indeed, the convergence rate for the explicit scheme (D.1) is very slow since the spectral gap vanishes as $\Delta t \to 0$. On the other hand, the unconditionally stable implicit scheme (D.8) gives the exponential convergence (D.13) with fast rate when we take Δt large enough but it is time-consuming to solve the inverse matrix in practice.

Appendix E. Computations of source term in von Mises-Fisher's ground-truth distribution

Recall the definition of von Mises-Fisher's distribution with oscillated parameters. We compute the source term in (4.6)

$$\begin{aligned} g(\theta,\varphi,t) =& \rho\Big[\kappa^2\eta_\theta^2 - \kappa\eta + \kappa\eta_\theta \cot\theta + \frac{1}{\sin^2\theta}(\kappa^2\eta_\varphi^2 + \kappa\eta_{\varphi\varphi}) - \frac{C'}{C}\kappa' - (\kappa'\eta + \kappa\eta_t)\Big], \\ C'/C =& \frac{\sinh\kappa - \kappa\cosh\kappa}{\kappa\sinh\kappa}. \end{aligned} \tag{E.1}$$

Using

$$\begin{aligned} \eta_\theta &= -\cos a\sin\theta + \sin a\cos\theta\cos(\varphi - b), \quad \eta_{\theta\theta} = -\cos a\cos\theta - \sin a\sin\theta\cos(\varphi - b) = -\eta, \\ \eta_\varphi &= -\sin a\sin\theta\sin(\varphi - b), \quad \eta_{\varphi\varphi} = -\sin a\sin\theta\cos(\varphi - b) = \cos a\cos\theta - \eta, \\ \eta_t &= -a'\sin a\cos\theta + a'\cos a\sin\theta\cos(\varphi - b) + b'\sin a\sin\theta\sin(\varphi - b). \end{aligned}$$

We obtain

$$\begin{aligned}\frac{g(\theta,\varphi,t)}{\rho} =&\kappa^2[\eta_\theta^2+\frac{\eta_\varphi^2}{\sin^2\theta}]+\kappa[-\eta+\eta_\theta\cot\theta-\frac{\sin a\cos(\varphi-b)}{\sin\theta}]-\frac{C'}{C}\kappa'-(\kappa'\eta+\kappa\eta_t),\\ =&\kappa^2[\eta_\theta^2+\sin^2 a\sin^2(\varphi-b)]+\kappa[-\eta+\eta_\theta\cot\theta-\frac{\sin a\cos(\varphi-b)}{\sin\theta}]-\frac{C'}{C}\kappa'-(\kappa'\eta+\kappa\eta_t),\\ =&\kappa^2[\eta_\theta^2+\sin^2 a\sin^2(\varphi-b)]+\kappa[-\eta-\cos a\cos\theta-\sin a\sin\theta\cos(\varphi-b)]-\frac{C'}{C}\kappa'-(\kappa'\eta+\kappa\eta_t),\\ =&\kappa^2[\eta_\theta^2+\sin^2 a\sin^2(\varphi-b)]-2\kappa\eta-\frac{C'}{C}\kappa'-(\kappa'\eta+\kappa\eta_t),\end{aligned}$$

which gives the source term (4.7).

References

[1] https://commons.wikimedia.org/wiki/file:pangaea_continents.png.
[2] Christopher Amante, 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] Dominique Bakry, Ivan Gentil, Michel Ledoux, et al., Analysis and Geometry of Markov Diffusion Operators, vol. 103, Springer, 2014.
[4] Jonathan Bates, The embedding dimension of Laplacian eigenfunction maps, Appl. Comput. Harmon. Anal. 37 (3) (2014) 516–530.
[5] John K. Beem, Pseudo-Riemannian manifolds with totally geodesic bisectors, Proc. Am. Math. Soc. 49 (1) (1975) 212–215.
[6] M. Belkin, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, Neural Comput. 15 (6) (2003) 1373–1396.
[7] Pierre Bérard, Gérard Besson, Sylvain Gallot, Embedding Riemannian manifolds by their heat kernel, Geom. Funct. Anal. GAFA 4 (4) (1994) 373–398.
[8] Tyrus Berry, Dimitrios Giannakis, John Harlim, Nonparametric forecasting of low-dimensional dynamical systems, Phys. Rev. E 91 (3) (2015) 032915.
[9] Tyrus Berry, John Harlim, Nonparametric uncertainty quantification for stochastic gradient flows, SIAM/ASA J. Uncertain. Quantificat. 3 (1) (2015) 484–508.
[10] Jeff Calder, Nicolas Garcia Trillos, Improved spectral convergence rates for graph Laplacians on ε-graphs and k-NN graphs, Appl. Comput. Harmon. Anal. 60 (2022) 123–175.
[11] Jeff Calder, Nicolas Garcia Trillos, Marta Lewicka, Lipschitz regularity of graph Laplacians on random data clouds, SIAM J. Math. Anal. 54 (1) (2022) 1169–1222.
[12] Xiuyuan Cheng, Nan Wu, Eigen-convergence of Gaussian kernelized graph Laplacian by manifold heat interpolation, Appl. Comput. Harmon. Anal. 61 (2022) 132–190.
[13] Shui-Nee Chow, Wen Huang, Yao Li, Haomin Zhou, Fokker-Planck equations for a free energy functional or Markov process on a graph, Arch. Ration. Mech. Anal. 203 (3) (2012) 969–1008.
[14] R.R. Coifman, I.G. Kevrekidis, S. Lafon, M. Maggioni, B. Nadler, Diffusion maps, reduction coordinates, and low dimensional representation of stochastic systems, Multiscale Model. Simul. 7 (2) (2008) 842–864.
[15] Ronald R. Coifman, Stéphane Lafon, Diffusion maps, Appl. Comput. Harmon. Anal. 21 (1) (2006) 5–30.
[16] Peter Deuflhard, Marcus Weber, Robust Perron cluster analysis in conformation dynamics, Linear Algebra Appl. 398 (Mar 2005) 161–184.
[17] D.L. Donoho, C. Grimes, Hessian eigenmaps: locally linear embedding techniques for high-dimensional data, Proc. Natl. Acad. Sci. 100 (10) (2003) 5591–5596.
[18] David B. Dunson, Hau-Tieng Wu, Nan Wu, Spectral convergence of graph Laplacian and heat kernel reconstruction in L^∞ from random samples, Appl. Comput. Harmon. Anal. 55 (2021) 282–336.
[19] W. E, T. Li, E. Vanden-Eijnden, Optimal partition and effective dynamics of complex networks, Proc. Natl. Acad. Sci. 105 (23) (Jun 2008) 7907–7912.
[20] E. Weinan, Tiejun Li, Eric Vanden-Eijnden, Applied Stochastic Analysis, Graduate Studies in Mathematics, American Mathematical Society, 2019.
[21] E. Weinan, Eric Vanden-Eijnden, Towards a theory of transition paths, J. Stat. Phys. 123 (3) (2006) 503.
[22] E. Weinan, Eric Vanden-Eijnden, Transition-path theory and path-finding algorithms for the study of rare events, Annu. Rev. Phys. Chem. 61 (1) (Mar 2010) 391–420.
[23] Ivar Ekeland, Roger Temam, Convex Analysis and Variational Problems, vol. 28, SIAM, 1999.
[24] Matthias Erbar, Gradient flows of the entropy for jump processes, Ann. Inst. Henri Poincaré Probab. Stat. 50 (3) (Aug 2014) 920–945.
[25] Antonio Esposito, Francesco S. Patacchini, André Schlichting, Dejan Slepčev, Nonlocal-interaction equation on graphs: gradient flow structure and continuum limit, Arch. Ration. Mech. Anal. 240 (2) (2021) 699–760.

[26] Robert Eymard, Thierry Gallouët, Raphaèle Herbin, Finite volume methods, in: Handbook of Numerical Analysis, vol. 7, 2000, pp. 713–1018.
[27] Yuan Gao, Guangzhen Jin, Jian-Guo Liu, Inbetweening auto-animation via Fokker-Planck dynamics and thresholding, Inverse Probl. Imaging 15 (5) (2021) 843.
[28] Yuan Gao, Tiejun Li, Xiaoguang Li, Jian-Guo Liu, Transition path theory for Langevin dynamics on manifold: optimal control and data-driven solver, Multiscale Model. Simul. (2022), in press, arXiv:2010.09988, 2022.
[29] Yuan Gao, Jian-Guo Liu, A note on parametric Bayesian inference via gradient flows, Ann. Math. Sci. Appl. 2 (2020) 261–282.
[30] Yuan Gao, Jian-Guo Liu, Random walk approximation for irreversible drift-diffusion process on manifold: ergodicity, unconditional stability and convergence, arXiv preprint, arXiv:2106.01344, 2021.
[31] Yuan Gao, Jian-Guo Liu, Revisit of macroscopic dynamics for some non-equilibrium chemical reactions from a Hamiltonian viewpoint, J. Stat. Phys. 189 (2) (2022) 1–57.
[32] Yuan Gao, Jian-Guo Liu, A selection principle for weak KAM solutions via Freidlin-Wentzell large deviation principle of invariant measures, arXiv preprint, arXiv:2208.11860, 2022.
[33] Yuan Gao, Jian-Guo Liu, Thermodynamic limit of chemical master equation via nonlinear semigroup, arXiv preprint, arXiv:2205.09313, 2022.
[34] David Gilbarg, Neil S. Trudinger, Elliptic Partial Differential Equations of Second Order, vol. 224, Springer, 2015.
[35] Elton P. Hsu, Stochastic Analysis on Manifolds, vol. 38, American Mathematical Soc., 2002.
[36] Peter W. Jones, Mauro Maggioni, Raanan Schul, Manifold parametrizations by eigenfunctions of the Laplacian and heat kernels, Proc. Natl. Acad. Sci. 105 (6) (2008) 1803–1808.
[37] Stephane Lafon, Ann B. Lee, Diffusion maps and coarse-graining: a unified framework for dimensionality reduction, graph partitioning, and data set parameterization, IEEE Trans. Pattern Anal. Mach. Intell. 28 (9) (2006) 1393–1403.
[38] Rongjie Lai, Jianfeng Lu, Point cloud discretization of Fokker–Planck operators for committor functions, Multiscale Model. Simul. 16 (2) (2018) 710–726.
[39] Lei Li, Jian-Guo Liu, Large time behaviors of upwind schemes by jump processes, Math. Comput. 89 (2020) 2283–2320.
[40] Tiejun Li, Jian Liu, E. Weinan, Probabilistic framework for network partition, Phys. Rev. E 80 (2) (2009) 026106.
[41] Anning Liu, Jian-Guo Liu, Yulong Lu, On the rate of convergence of empirical measure in ∞-Wasserstein distance for unbounded density function, Q. Appl. Math. 77 (4) (2019) 811–829.
[42] Jan Maas, Gradient flows of the entropy for finite Markov chains, J. Funct. Anal. 261 (8) (Oct 2011) 2250–2292.
[43] Philipp Metzner, Christof Schütte, Eric Vanden-Eijnden, Transition path theory for Markov jump processes, Multiscale Model. Simul. 7 (3) (Jan 2009) 1192–1219.
[44] Alexander Mielke, D.R. Michiel Renger, Mark A. Peletier, On the relation between gradient flows and the large-deviation principle, with applications to Markov chains and diffusion, Potential Anal. 41 (4) (Nov 2014) 1293–1327.
[45] Boaz Nadler, Stéphane Lafon, Ronald R. Coifman, Ioannis G. Kevrekidis, Diffusion maps, spectral clustering and reaction coordinates of dynamical systems, Appl. Comput. Harmon. Anal. 21 (1) (2006) 113–127.
[46] Jacobus W. Portegies, Embeddings of Riemannian manifolds with heat kernels and eigenfunctions, Commun. Pure Appl. Math. 69 (3) (2016) 478–518.
[47] Jan-Hendrik Prinz, Hao Wu, Marco Sarich, Bettina Keller, Martin Senne, Martin Held, John D. Chodera, Christof Schütte, Frank Noé, Markov models of molecular kinetics: generation and validation, J. Chem. Phys. 134 (17) (May 2011) 174105.
[48] Mary A. Rohrdanz, Wenwei Zheng, Mauro Maggioni, Cecilia Clementi, Determination of reaction coordinates via locally scaled diffusion map, J. Chem. Phys. 134 (12) (2011) 124116.
[49] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500) (2000) 2323–2326.
[50] Christof Schütte, Frank Noé, Jianfeng Lu, Marco Sarich, Eric Vanden-Eijnden, Markov state models based on milestoning, J. Chem. Phys. 134 (20) (May 2011) 204105.
[51] Amit Singer, H.-T. Wu, Vector diffusion maps and the connection Laplacian, Commun. Pure Appl. Math. 65 (8) (2012) 1067–1144.
[52] Amit Singer, Hau-Tieng Wu, Spectral convergence of the connection Laplacian from random samples, Inf. Inference, J. IMA 6 (1) (2016) 58–123.
[53] J.B. Tenenbaum, V. de Silva, J.C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (5500) (2000) 2319–2323.
[54] Nicolás García Trillos, Moritz Gerlach, Matthias Hein, Dejan Slepčev, Error estimates for spectral convergence of the graph Laplacian on random geometric graphs toward the Laplace–Beltrami operator, Found. Comput. Math. 20 (4) (2020) 827–887.
[55] Nicolás Garcia Trillos, Dejan Slepčev, On the rate of convergence of empirical measures in ∞-transportation distance, Can. J. Math. 67 (6) (2015) 1358–1383.
[56] Hau-Tieng Wu, Nan Wu, Think globally, fit locally under the manifold setup: asymptotic analysis of locally linear embedding, Ann. Stat. 46 (6B) (2018) 3805–3837.
[57] Amber Yuan, Jeff Calder, Braxton Osting, A continuum limit for the pagerank algorithm, Eur. J. Appl. Math. 33 (3) (2022) 472–504.