GEOMETRY OF PROBABILITY SIMPLEX VIA OPTIMAL TRANSPORT

WUCHEN LI

Abstract. We study the Riemannian structures of the probability simplex on a weighted graph introduced by L^2-Wasserstein metric. The main idea is to embed the probability simplex as a submanifold of the positive orthant. From this embedding, we establish the geometry formulas of the probability simplex in Euclidean coordinates. The geometry computations on discrete simplex guide us to introduce the ones in the Fréchet manifold of densities supported on a finite dimensional base manifold. Following the steps of Nelson, Bakery-Émery, Lott-Villani-Strum and the geometry of density manifold, we demonstrate an identity that connects the Bakery-Émery Γ_2 operator (carré du champ itéré) and Yano's formula on the base manifold. Several examples of differential equations in probability simplex are demonstrated.

1. Introduction

In recent years, optimal transport, a.k.a. the Monge-Kantorovich problem has attracted a lot of attention in various fields, such as partial differential equations [5, 27], functional inequalities [28], differential geometry [20, 22, 30] and evolutionary dynamics [18]. Given a finite dimensional manifold M, it studies metrics in the set of probability densities $\mathcal{P}(M)$. In particular, the L^2-Wasserstein metric W endows the Riemannian metric tensor g_W in the set of densities. In literature [17], the infinite dimensional density space $(\mathcal{P}(M), g_W)$, in the sense of Fréchet manifold [16], is named density manifold. It gives the other viewpoint on many classical equations. For example, the gradient flow in density manifold connects with the Fokker-Planck equation [15], while the Hamiltonian flow in density manifold relates to the Schrödinger equation [4, 17, 25, 26].

The Riemannian formulas in $(P(M), g_W)$ play critical roles in analyzing dynamical systems in $P(M)$, understanding the geometry of base manifold M and designing related numerical schemes. Three examples are as follows. First, the Hessian operator in density manifold has been used to study the long-time behaviors of Fokker-Planck equations [6, 28]; Second, the differential structures between $\mathcal{P}(M)$ and its base manifold M have deep relations. For instance, the Hessian of negative Boltzmann-Shannon entropy in density manifold relates to the Ricci curvature by the Bochner's formula [3] and connects to the Yano's formula [11]; Lastly, the Riemannian gradient operator in density manifold has been used to design the evolutionary dynamics in population games. For more details see [9, 18]. However, the Riemannian structures of density manifold are not completely clear. For example, Villani asks the following questions on page 444-445 of his famous book [31].

Key words and phrases. Optimal transport; Probability manifold; Linear weighted Laplacian operator; Information geometry; Graph.

"Can one define the Christopher symbol, Laplace operator, volume form and divergence operator on density manifold, at least formally?" In addition, "Open problem 15.11: Find a nice formula for the Hessian of the functional in density manifold."

In this paper, we study the Riemannian structure of the probability simplex supported on finite states with L^2-Wasserstein metric tensor, a topic founded by [8, 21, 23]. And we further extend the derivations into the infinite dimensional density manifold. We pay special attention to the finite states for two reasons. On one hand, many realistic models, such as population genetics [29] and statistical models [1, 2], are built on finite states. Moreover, numerical schemes of Fokker-Planck equations [7] and Schrödinger equations [10] are usually designed on the finite discrete grids. On the other hand, the finite states will guide us the intuition for the study of infinite dimensional settings.

The approach mainly follows the study of linear weighted Laplacian operator (a concept defined in section 2) in [10, 11, 18]. The result is sketched as follows. We construct a Riemannian metric tensor in the discrete positive measure space $(\mathcal{M}_+, g)$, where g is a positive definite metric tensor in positive orthant. In Theorem 1, we show that the probability simplex endowed with the L^2-Wasserstein metric tensor $(\mathcal{P}_+, g_W)$ is a submanifold of $(\mathcal{M}_+, g)$. By the geometry structures in $(\mathcal{M}_+, g)$, we derive the corresponding ones in $(\mathcal{P}_+, g_W)$, including Christopher symbol, volume form, Laplacian-Beltrami and Hessian operators by [Proposition 2-9]. Similar derivations are also provided in the infinite dimensional density manifold by [Proposition 10-18]. The Hessian operator in density manifold is provided by [Proposition 15 and 18], by which the connection among the metric tensor in $\mathcal{P}(M)$, Bakry-Émery Γ_2 operator and Yano formula in M are introduced in [Proposition 19].

In literature, the Riemannian structures of density manifold $(\mathcal{P}(M), g_W)$ have been studied from the differential structures of the base manifold M, by the novel works of Lafferty [17], Otto [27], Lott [19] and Gigli [14]. In [17], guiding by the stochastic mechanics, the Riemannian and symplectic structures of density manifold are introduced by the tangent vectors in the base manifold M. This approach follows Moser's theorem [24]. The density manifold is viewed as the quotient of the group of diffeomorphisms in M which preserve the Riemannian volume of M. The gradient operator in density manifold is derived by Otto [27]. In this approach, the dual coordinate system is used, which is often named Otto calculus [31]. The geometric calculations in this dual coordinate system have been done in [19], in which the curvature tensor in density manifold is established by the ones in M. [14] develops a rigorous second order analysis on the space of probability measures with the finite second moment. In contrast to their works, we study the geometry of density manifold by embedding it to the positive measure space. Thus the geometry formulas are given in Euclidean (L^2) coordinates. They are given directly from the Riemannian metric tensor, the inverse of linear Laplacian operator. This angle lets us derive the geometry formulas of probability simplex supported on both graphs and base manifold M.

We note that the idea of embedding probability simplex into positive octant is motivated by Shahshahani metric [29], also called Fisher-Rao metric in information geometry [1, 2]. And the L^2-Wasserstein metric tensor in probability simplex is closely related to the osmotic diffusion considered in Nelson's stochastic mechanics [17, 26]. In addition, the connection between Γ_2 operator and the metric tensor of density manifold follows the work of Nelson [26], Bakery-Émery [3], Lott-Villani-Strum [20, 30] and the study of Yano's formula [11, 33].

The plan of this paper is as follows. In section 2 and 3, we review the optimal transport on a graph and establish the Riemannian structure of discrete probability simplex. In section 4, we introduce the associated geometry formulas in density manifold. Several examples of differential equations on probability simplex are introduced in section 5.

2. Review of optimal transport on graphs

In this section, we review the dynamical optimal transport on a graph [11]. We introduce the Riemannian manifold structure for a probability simplex supported on a finite set of states, which are vertices of a graph.

Consider a weighted undirected finite graph $G = (V, E, \omega)$, which contains the vertex set $V = \{1, 2, \cdots, n\}$, the edge set E, and the weights set ω on the edges. Here $\omega = (\omega_{ij})_{i,j\in V} \in \mathbb{R}^{n\times n}$ is a symmetric matrix, such that $\omega_{ij} = \begin{cases}\omega_{ji} > 0 & \text{if } (i,j)\in E\\ 0 & \text{otherwise}\end{cases}$. We denote the adjacent set or neighborhood of node i by $N(i) = \{j \in V : (i, j) \in E\}$.

The probability simplex supported on all vertices of G is defined by

$$\mathcal{P}(G) = \{(\rho_1, \cdots, \rho_n) \mid \sum_{i=i}^{n} \rho_i = 1 , \quad \rho_i \geq 0\} \in \mathbb{R}^n ,$$

where ρ_i is the discrete probability distribution at node i, whose interior is denoted by

$$\mathcal{P}_+(G) = \{(\rho_1, \cdots, \rho_n) \mid \sum_{i=1}^{n} \rho_i = 1 , \quad \rho_i > 0\} ,$$

and whose boundary is given by $\partial\mathcal{P}(G) = \mathcal{P}(G) \setminus \mathcal{P}_+(G)$.

We first define the "metric tensor" on graphs. The discrete *vector field* $v = (v_{ij})_{i,j\in V} \in \mathbb{R}^{n\times n}$ is a *skew-symmetric matrix*:

$$v_{ij} = \begin{cases}-v_{ji} & \text{if } (i,j)\in E\\ 0 & \text{otherwise}\end{cases} .$$

Given a potential $\Phi \in \mathbb{R}^n$, a discrete *gradient vector field* $\nabla_G\Phi = (\nabla_{ij}\Phi)_{i,j\in V} \in \mathbb{R}^{n\times n}$ is defined by

$$\nabla_{ij}\Phi = \sqrt{\omega_{ij}}(\Phi(i) - \Phi(j)) .$$

For the simplicity of notation, we vectorize the matrix $v, \nabla_G\Phi \in \mathbb{R}^{n\times n}$. We decompose the indirect edge set by two direct edge sets $E = \overrightarrow{E} \cup \overleftarrow{E}$ with $\overrightarrow{E} = \{(i, j) \in E : i > j\}$ and $\overleftarrow{E} = \{(i, j) \in E : i < j\}$. We denote $v = (v_{ij})_{(i,j)\in\overrightarrow{E}} \in \mathbb{R}^{|E|}$, $\nabla\Phi = (\nabla_{ij}\Phi)_{(i,j)\in\overrightarrow{E}} \in \mathbb{R}^{|E|}$.

Define the *metric tensor* of discrete vectors v, $\tilde{v}$ at each node i by

$$(v, \tilde{v})_i := \frac{1}{2}\sum_{j\in N(i)} v_{ij}\tilde{v}_{ij} .$$

We next define the expectation value of the discrete inner product w.r.t the probability distribution ρ:

$$(v,\tilde{v})_\rho := \sum_{i\in I} \rho_i \ (v,\tilde{v})_i = \sum_{(i,j)\in \vec{E}} v_{ij}\tilde{v}_{ij}\frac{\rho_i+\rho_j}{2} \ .$$

If $\tilde{v} = \nabla_G \Phi$ is a gradient vector field, the above can be rewritten as

$$(v,\nabla_G\Phi)_\rho = \sum_{(i,j)\in\vec{E}} v_{ij}\sqrt{\omega_{ij}}(\Phi_i - \Phi_j)\frac{\rho_i+\rho_j}{2} = \sum_{i=1}^n \Phi_i \sum_{j\in N(i)} \sqrt{\omega_{ij}} v_{ij}\frac{\rho_i+\rho_j}{2} = \Phi^\mathsf{T}\big(-\mathrm{div}_G(\rho v)\big) \ ,$$

where

$$-\mathrm{div}_G(\rho v) := \Big(\sum_{j\in N(i)} v_{ij}\frac{\rho_i+\rho_j}{2}\Big)_{i=1}^n \ . \tag{1}$$

There are two definitions hidden in (1). First, $\mathrm{div}_G \colon \mathbb{R}^{n\times n} \to \mathbb{R}^n$ maps any given vector field m on the graph to

$$\mathrm{div}_G(m) = (\sum_{j\in N(i)} \sqrt{\omega_{ij}} m_{ji})_{i=1}^n \ .$$

Second, the probability weight (flux) $m = \rho v$ of the vector field v is defined by

$$m_{ij} = \begin{cases} \frac{\rho_i+\rho_j}{2} v_{ij} & \text{if } (i,j)\in E \\ 0 & \text{otherwise} \end{cases} \ ,$$

where $\frac{\rho_i+\rho_j}{2}$ represents the probability weight on the edge $(i,j)\in E$.

We are now ready to introduce the L^2-Wasserstein metric on probability set $\mathcal{P}(G)$.

Definition 1 (L^2-Wasserstein metric on a graph)**.** *Given ρ^0, $\rho^1 \in \mathcal{P}(G)$, the metric $W\colon \mathcal{P}(G)\times\mathcal{P}(G)\to\mathbb{R}$ is defined by*

$$\begin{aligned}&\big(W(\rho^0,\rho^1)\big)^2\\ &= \inf_{\rho(t),v(t)} \Big\{ \int_0^1 (v(t),v(t))_{\rho(t)} dt \colon \dot{\rho}(t) + \mathrm{div}_G(\rho(t)v(t)) = 0,\ \rho(0)=\rho^0,\ \rho(1)=\rho^1 \Big\} \ .\end{aligned} \tag{2}$$

Here the infimum is performed over the set of pairs $(\rho(t), v(t))$ such that $\rho\in H^1(0,1;\mathbb{R}^n)$, $v : [0,1]\to S^{n\times n}$ is measurable.

Variational problem (2) has the other equivalent representation, which introduces the probability simplex the Riemannian structure. To show this point, the following matrix functions are needed.

Definition 2 (Linear Weighted Laplacian matrix)**.** *Define the matrix function $L(\cdot) : \mathbb{R}^n \to \mathbb{R}^{n\times n}$ by*

$$L(a) = D^\mathsf{T}\Theta(a)D \ , \quad a = (a_i)_{i=1}^n \in \mathbb{R}^n \ ,$$

where

- $D \in \mathbb{R}^{|E|\times n}$ *is the discrete gradient operator*

$$D_{(i,j)\in E, k\in V} = \begin{cases} \sqrt{\omega_{ij}}, & \text{if } i = k,\ i > j \\ -\sqrt{\omega_{ij}}, & \text{if } j = k,\ i > j \\ 0, & \text{otherwise} \end{cases} ;$$

- $-D^{\mathsf{T}} \in \mathbb{R}^{n\times|E|}$ *is the discrete divergence operator, also named the oriented incidence matrix* [12]*;*
- $\Theta(a) \in \mathbb{R}^{|E|\times|E|}$ *is a weight matrix depending on* a*,*

$$\Theta(a)_{(i,j)\in E,(k,l)\in E} = \begin{cases} \frac{a_i+a_j}{2} & \text{if } (i,j) = (k,l) \in E \\ 0 & \text{otherwise} \end{cases} .$$

Let $a = \rho \in \mathcal{P}_+(G)$, we next study the property of matrix $L(\rho)$, from which we shall build the Riemannian metric tensor of probability simplex.

Lemma 1 (Discrete Hodge decomposition)**.** *Given* $\rho \in \mathcal{P}_+(G)$*, the following properties hold:*

(i) $L(\rho)$ *is a semi-positive matrix with a single zero eigenvalue. Denote the eigenvalue and corresponding orthonormal eigenvectors of* $L(\rho)$ *by* $0 = \lambda_0(\rho) < \lambda_1(\rho) \leq \cdots \leq \lambda_{n-1}(\rho)$*, and* $U(\rho) = (u_0, u_1(\rho), \cdots, u_{n-1}(\rho))$*,*

$$L(\rho) = U(\rho)\begin{pmatrix} 0 & & & \\ & \lambda_1(\rho) & & \\ & & \ddots & \\ & & & \lambda_{n-1}(\rho) \end{pmatrix} U(\rho)^T ,$$

with

$$u_0 = \frac{1}{\sqrt{n}}(1, \cdots, 1)^{\mathsf{T}} . \tag{3}$$

(ii) *For any discrete vector field* v *and* $\rho \in \mathcal{P}_+(G)$*, there exists a unique discrete gradient vector field* $\nabla_G\Phi \in \mathbb{R}^{|E|}$*, such that*

$$v_{ij} = \nabla_{ij}\Phi + \Psi_{ij} , \quad div_G(\rho\Psi) = 0 ,$$

where Ψ *is a discrete divergence free vector field w.r.t.* ρ*. In addition,*

$$(v, v)_\rho = (\nabla_G\Phi, \nabla_G\Phi)_\rho + (\Psi, \Psi)_\rho .$$

Proof. The proof is a direct extension of classical graph Hodge decomposition with the probability weight function $\frac{\rho_i+\rho_j}{2}$. Given a discrete vector field v and $\rho \in \mathcal{P}_+(G)$, we shall show that there exists a unique gradient vector field $\nabla_G\Phi$, such that

$$-\mathrm{div}_G(\rho\nabla_G\Phi) = L(\rho)\Phi = -\mathrm{div}_G(\rho v) .$$

Consider

$$\Phi^{\mathsf{T}} L(\rho)\Phi = \sum_{(i,j)\in\overrightarrow{E}} \omega_{ij}(\Phi(i) - \Phi(j))^2 \frac{\rho_i + \rho_j}{2} = 0 .$$

Since $\rho_i > 0$ for any $i \in V$ and the graph is connected, we find that $\Phi_1 = \cdots = \Phi_n$ is the only solution of above equation. Thus 0 must be the simple eigenvalue of $L(\rho)$

with eigenvector $(1, \cdots, 1)^\mathsf{T}$. Since $\mathrm{div}_G(\rho v) \in \mathrm{Ran}\big(L(\rho)\big)$ and $\mathrm{Ker}\big(L(\rho)\big) = \{u_0\}$. Thus there exists a unique solution of Φ up to a constant shrift, i.e. $\nabla_G\Phi$ is unique. And $\Psi = v - \nabla_G\Phi$ satisfies $\mathrm{div}_G(\rho\Psi) = \mathrm{div}_G(\rho v) - \mathrm{div}_G(\rho\nabla\Phi) = 0$. Let $v_{ij} = \nabla_{ij}\Phi + \Psi_{ij}$, where $\mathrm{div}_G(\rho\Psi) = 0$. Then

$$\begin{aligned}(v, v)_\rho =&(\nabla_G\Phi, \nabla_G\Phi)_\rho + 2(\nabla_G\Phi, \Psi)_\rho + (\Psi, \Psi)_\rho \\ =&(\nabla_G\Phi, \nabla_G\Phi)_\rho + (\Psi, \Psi)_\rho \ ,\end{aligned}$$

which finishes the proof. □

From Lemma 1, for any discrete vector field v, there exists an unique pair $(\nabla_G\Phi, \Psi)$, such that $\mathrm{div}_G(\rho v) = \mathrm{div}_G(\rho\nabla_G\Phi)$ and

$$(v, v)_\rho = (\nabla_G\Phi, \nabla_G\Phi)_\rho + (\Psi, \Psi)_\rho \geq (\nabla_G\Phi, \nabla_G\Phi)_\rho \ .$$

Thus the metric W defined in (2) is equivalent to

$$\begin{aligned}&\big(W(\rho^0, \rho^1)\big)^2 \\ =&\inf_{\Phi(t)}\{\int_0^1 (\nabla_G\Phi(t), \nabla_G\Phi(t))_{\rho(t)}dt \colon \frac{d\rho}{dt} + \mathrm{div}_G(\rho\nabla_G\Phi) = 0,\ \rho(0) = \rho^0,\ \rho(1) = \rho^1\} \\ =&\inf_{\Phi(t)}\{\int_0^1 \Phi^\mathsf{T}(t)L(\rho(t))\Phi(t)dt \colon \frac{d\rho}{dt} = L(\rho)\Phi,\ \rho(0) = \rho^0,\ \rho(1) = \rho^1\} \ .\end{aligned} \tag{4}$$

2.1. **Riemannian manifold of probability simplex.** We demonstrate that (4) will introduce the Riemannian metric tensor of the probability simplex in both primal and dual coordinates. The probability simplex $\mathcal{P}(G)$ is a manifold with boundary. To simplify the discussion, we focus on the interior $\mathcal{P}_+(G)$. For more details of the study of geodesics in the boundary set, see [13].

Denote the tangent space at a point $\rho \in \mathcal{P}_+(G)$ by

$$T_\rho\mathcal{P}_+(G) = \{(\sigma(i))_{i=1}^n \in \mathbb{R}^n \colon \sum_{i=1}^n \sigma(i) = 0\} \ .$$

Denote the space of potential function on the set of vertices set by $\mathcal{F}(G) = \{(\Phi_i)_{i=1}^n \in \mathbb{R}^n\}$. Consider the quotient space

$$\mathcal{F}(G)/\mathbb{R} = \{[\Phi] \mid (\Phi(i))_{i=1}^n \in \mathbb{R}^n\} \ ,$$

where $[\Phi] = \{(\Phi(1) + c, \cdots, \Phi(n) + c) \mid c \in \mathbb{R}\}$ are functions defined up to addition of constants.

We introduce an identification map by linear Laplacian operator

$$\mathbf{V} \colon \mathcal{F}(G)/\mathbb{R} \to T_\rho\mathcal{P}_+(G) \ , \qquad \mathbf{V}_\Phi = L(\rho)\Phi \ .$$

From Lemma 1, $\mathbf{V} \colon \mathcal{F}(G)/\mathbb{R} \to T_\rho\mathcal{P}_+(G)$ is a well defined map, linear, and one to one. I.e., $\mathcal{F}(G)/\mathbb{R} \cong T_\rho^*\mathcal{P}_+(G)$. Here $T_\rho^*\mathcal{P}_+(G)$ is the cotangent space of $\mathcal{P}_+(G)$.

This identification induces the following inner product on $T_\rho\mathcal{P}_+(G)$.

Definition 3 (Inner product in dual coordinates)**.** *Given $\rho \in \mathcal{P}_+(G)$, the inner product $g_W : T_\rho\mathcal{P}_+(G) \times T_\rho\mathcal{P}_+(G) \to \mathbb{R}$ takes any two tangent vectors $\sigma_1 = \mathbf{V}_{\Phi_1}$ and $\sigma_2 = \mathbf{V}_{\Phi_2} \in T_\rho\mathcal{P}_+(G)$ to*

$$g_W(\sigma_1, \sigma_2) = \sigma_1^{\mathsf{T}}\Phi_2 = \sigma_2^{\mathsf{T}}\Phi_1 = (\nabla_G\Phi_1, \nabla_G\Phi_2)_\rho \ . \tag{5}$$

The above is written in the dual coordinates of Riemannian manifold. I.e. $\Phi \in \mathcal{F}(G)/\mathbb{R}$. We would like to give the inner product in primal coordinates $\sigma \in T_\rho\mathcal{P}_+(G)$. I.e. $\sigma \in T_\rho\mathcal{P}_+(G)$.

Definition 4 (Inner product in primal coordinates)**.** *Given $\rho \in \mathcal{P}_+(G)$, the inner product $g_W(\cdot,\cdot) : T_\rho\mathcal{P}_+(G) \times T_\rho\mathcal{P}_+(G) \to \mathbb{R}$ is defined by*

$$g_W(\sigma_1, \sigma_2) := {\sigma_1}^{\mathsf{T}} L(\rho)^\dagger \sigma_2 \ , \quad \textit{for any } \sigma_1, \sigma_2 \in T_\rho\mathcal{P}_+(G) \ ,$$

where matrix $L(\rho)^\dagger$ is the pseudo-inverse of $L(\rho)$ defined by

$$L(\rho)^\dagger = U(\rho) \begin{pmatrix} 0 & & & \\ & \frac{1}{\lambda_1(\rho)} & & \\ & & \ddots & \\ & & & \frac{1}{\lambda_{n-1}(\rho)} \end{pmatrix} U(\rho)^{\mathsf{T}} \ .$$

The connection of primal and dual coordinates are as follows. Denote $\sigma_i = L(\rho)\Phi_i$, $i = 1, 2$, then

$${\sigma_1}^{\mathsf{T}} L(\rho)^\dagger \sigma_2 = \Phi_1^{\mathsf{T}} L(\rho) L(\rho)^\dagger L(\rho)\Phi_2 = \Phi_1^{\mathsf{T}} L(\rho)\Phi_2 \ , \tag{6}$$

where the second equality is from the definition of pseudo-inverse, i.e. $L(\rho)L(\rho)^\dagger L(\rho) = L(\rho)$. Denote $\dot\rho = \frac{d\rho(t)}{dt}$, from which the variational problem (4) is rewritten as

$$\big(W(\rho^0, \rho^1)\big)^2 = \inf_{\rho(t) \in \mathcal{P}_+(G)} \Big\{ \int_0^1 \dot\rho(t)^{\mathsf{T}} L(\rho(t))^\dagger \dot\rho(t) dt \ : \rho(0) = \rho^0, \ \rho(1) = \rho^1 \Big\} \ .$$

By the time reparametrization, we have

$$W(\rho^0, \rho^1) = \inf_{\rho(t) \in \mathcal{P}_+(G)} \Big\{ \int_0^1 \sqrt{\dot\rho^{\mathsf{T}}(t) L(\rho(t))^\dagger \dot\rho(t)} dt \ : \rho(0) = \rho^0 \ , \ \rho(1) = \rho^1 \Big\} \ .$$

Since $L(\rho)^\dagger$ is positive definite in $T_\rho\mathcal{P}_+(G)$,.

$$\inf_{\sigma \in \mathbb{R}^n} \Big\{ \sigma^{\mathsf{T}} L(\rho)^\dagger \sigma \colon \sum_{i=1}^n \sigma_i = 0 \ , \quad \sum_{i=1}^n \sigma_i^2 = 1 \Big\} = \frac{1}{\lambda_1(\rho)} > 0 \ ,$$

and $L(\rho)$ is smooth w.r.t. $\rho \in \mathcal{P}_+(G)$, then $(\mathcal{P}_+(G), g_W)$ is a $(n-1)$ dimensional Riemannian manifold. As in [17], we call $(\mathcal{P}_+(G), g_W)$ *probability manifold.*

3. Riemannian structures on probability manifold

In this section, we first construct a Riemannian metric tensor on positive orthant. We then embed $(\mathcal{P}_+(G), g_W)$ as a submanifold of positive orthant. From this angle, we derive many geometry formulas in $(\mathcal{P}_+(G), g_W)$ by the Euclidean coordinates in positive orthant.

Consider the positive orthant (discrete positive measure space) by

$$\mathcal{M}_+(G) = \{(\mu_1, \cdots, \mu_n) \mid \mu_i > 0\} \subset \mathbb{R}^n_+ \ ,$$

and

$$T_\mu \mathcal{M}_+(G) = \mathbb{R}^n \ .$$

It is clear that $\mathcal{P}_+(G) \subset \mathcal{M}_+(G)$. We next define a Riemannian inner product on $\mathcal{M}_+(G)$.

Definition 5 (Inner product on $\mathcal{M}_+(G)$)**.** *Given $\mu \in \mathcal{M}_+(G)$, define the inner product $g_{\mathcal{M}} \colon T_\mu\mathcal{M}_+(G)) \times T_\mu\mathcal{M}_+(G) \to \mathbb{R}$ by*

$$g_{\mathcal{M}}(a_1, a_2) = a_1^{\mathsf{T}} g(\mu) a_2 \ , \quad \textit{for any } a_1,\ a_2 \in T_\mu\mathcal{M}_+(G) \ ,$$

where

$$g(\mu) = L(\mu)^\dagger + u_0 u_0^{\mathsf{T}} \in \mathbb{R}^{n\times n} \ ,$$

and $u_0 = \frac{1}{\sqrt{n}}(1, \cdots, 1)^{\mathsf{T}}$.

We next show that $(\mathcal{M}_+(G), g_{\mathcal{M}})$ is a smooth n dimensional Riemannian manifold, and $(\mathcal{P}_+(G), g_W)$ is a $(n-1)$ dimensional submanifold of $(\mathcal{M}_+(G), g_{\mathcal{M}})$ with the induced metric tensor.

Theorem 1 (Induced metric)**.** *Denote a natural inclusion by*

$$\imath \colon \mathcal{P}_+(G) \to \mathcal{M}_+(G) \ , \qquad \imath(\rho) = \rho \ ,$$

then $\imath$ induces a Riemannian metric tenor g_W on $\mathcal{P}_+(G)$ via pullback:

$$g_W(\sigma_1, \sigma_2) = g_{\mathcal{M}}(\sigma_1, \sigma_2) \ , \quad \textit{for any } \sigma_1, \sigma_2 \in T_\rho\mathcal{P}_+(G) \ .$$

Proof. Since $L(\mu)^\dagger = U(\mu) \begin{pmatrix} 0 & & & \\ & \frac{1}{\lambda_1(\mu)} & & \\ & & \ddots & \\ & & & \frac{1}{\lambda_{n-1}(\mu)} \end{pmatrix} U(\mu)^{\mathsf{T}}$ with $U(\mu) = [u_0, u_1(\mu), \cdots, u_{n-1}(\mu)]$, then

$$g(\mu) = \sum_{i=1}^{n-1} \frac{1}{\lambda_i(\mu)} u_i(\mu) u_i(\mu)^{\mathsf{T}} + u_0 u_0^{\mathsf{T}} = U(\mu) \begin{pmatrix} 1 & & & \\ & \frac{1}{\lambda_1(\mu)} & & \\ & & \ddots & \\ & & & \frac{1}{\lambda_{n-1}(\mu)} \end{pmatrix} U(\mu)^{\mathsf{T}} \ .$$

So $g(\mu)$ is a positive definite matrix and smooth w.r.t. $\mu \in \mathcal{M}_+(G)$, and $(\mathcal{M}_+(G), g_{\mathcal{M}})$ is a smooth Riemannian manifold.

For any $\rho \in \mathcal{P}_+(G)$ and $\sigma_1, \sigma_2 \in T_\rho\mathcal{P}_+(G)$, we shall show $g_W(\sigma_1, \sigma_2) = \sigma_1^{\mathsf{T}} g(\rho) \sigma_2$, i.e.

$$\sigma_1^{\mathsf{T}} (L(\rho)^\dagger + u_0 u_0^{\mathsf{T}}) \sigma_2 = \sigma_1^{\mathsf{T}} L(\rho)^\dagger \sigma_2 \ .$$

Since $u_0 = \frac{1}{\sqrt{n}}(1, \cdots, 1)^{\mathsf{T}}$, then $u_0^{\mathsf{T}} \sigma_2 = \frac{1}{\sqrt{n}} \sum_{i=1}^n \sigma_2(i) = 0$, thus $\sigma_1^{\mathsf{T}} u_0 u_0^{\mathsf{T}} \sigma_2 = 0$, which finishes the proof. ∎

Many geometry formulas of $(\mathcal{P}_+(G), g_W)$ can be derived from the ones in $(\mathcal{M}_+(G), g_{\mathcal{M}})$. We illustrate an example by the gradient operator.

Proposition 1 (Gradient)**.** *Given $\mathcal{F} \in C^\infty(\mathcal{M}_+(G))$, denote its gradient operators in $(\mathcal{M}_+(G), g_{\mathcal{M}})$ and $(\mathcal{P}_+(G), g_W)$ by $\nabla_g \mathcal{F}(\rho) \in T_\rho\mathcal{M}_+(G)$, $\nabla_W \mathcal{F}(\rho) \in T_\rho \mathcal{P}_+(G)$, respectively. Then*

$$\nabla_W \mathcal{F}(\rho) = -div_G(\rho \nabla_G d_\rho \mathcal{F}(\rho)) \ , \quad d_\rho \mathcal{F}(\rho) = (\frac{\partial}{\partial \rho_i}\mathcal{F}(\rho))_{i=1}^n \ ,$$

is the orthogonal projection related to g on $\mathcal{M}_+(G)$, of the restriction of $\nabla_g\mathcal{F}(\rho)$ to $\mathcal{P}_+(G)$, i.e.

$$\nabla_W \mathcal{F}(\rho) = \nabla_g\mathcal{F}(\rho) - g_{\mathcal{M}}(\nabla_g \mathcal{F}(\rho), u_0)u_0 \ .$$

Proof. Since

$$g(\mu)^{-1} = U(\mu)\begin{pmatrix} 1 & & & \\ & \lambda_1(\mu) & & \\ & & \ddots & \\ & & & \lambda_{n-1}(\mu)\end{pmatrix} U(\mu)^{\mathsf{T}} = L(\mu) + u_0u_0^{\mathsf{T}} \ ,$$

then

$$\nabla_g \mathcal{F}(\rho) = \Big(L(\rho) + u_0u_0^{\mathsf{T}}\Big) d_\rho \mathcal{F}(\rho) \ .$$

For any $\sigma \in T_\rho \mathcal{P}_+(G)$, notice $\sigma^T u_0 = 0$ and $u_0 \in \ker(L(\rho))$, then

$$\sigma^{\mathsf{T}} g(\rho) u_0 = \sigma^{\mathsf{T}}(L(\rho)^\dagger + u_0u_0^{\mathsf{T}})u_0 = \sigma^{\mathsf{T}}L(\rho)^\dagger u_0 = 0 \ .$$

Thus the unit norm vector at $\rho \in \mathcal{P}_+(G)$ is a constant vector u_0. The orthogonal projection of $\nabla_g \mathcal{F}(\rho)$ is

$$\begin{aligned}\nabla_g\mathcal{F}(\rho) - \Big(u_0^{\mathsf{T}} g(\rho)\nabla_g \mathcal{F}(\rho)\Big)u_0 =& \Big(L(\rho) + u_0u_0^{\mathsf{T}}\Big)d_\rho\mathcal{F}(\rho) - u_0u_0^{\mathsf{T}} d_\rho \mathcal{F}(\rho) \\ =& L(\rho) d_\rho \mathcal{F}(\rho) \ ,\end{aligned}$$

which finishes the proof. □

We next proceed the geometric computations in $(\mathcal{P}_+(G), g_W)$ by primal coordinates. Given $\rho \in \mathcal{P}_+(G)$, we first compute the commutator $[\cdot,\cdot]_W \colon T_\rho\mathcal{P}_+(G) \times T_\rho \mathcal{P}_+(G) \to T_\rho\mathcal{P}_+(G)$.

Proposition 2 (Commutator)**.** *Given constant vectors σ_1, $\sigma_2 \in T_\rho\mathcal{P}_+(G)$, then*

$$[\sigma_1, \sigma_2]_W = 0 \ .$$

Proof. Consider $\mathcal{F} \in C^\infty(\mathcal{P}_+(G))$. Since $(\sigma_1\mathcal{F})(\rho) := \frac{d}{dt}|_{t=0}\mathcal{F}(\rho + t\sigma_1)$, then

$$\begin{aligned}([\sigma_1,\sigma_2]_W\mathcal{F})(\rho) =& \frac{d}{dt}|_{t=0}\frac{d}{ds}|_{s=0}\mathcal{F}(\rho + t\sigma_1 + s\sigma_2) - \frac{d}{ds}|_{s=0}\frac{d}{dt}|_{t=0}\mathcal{F}(\rho + t\sigma_1 + s\sigma_2) \\ =& \sigma_1^{\mathsf{T}} d^2_\rho \mathcal{F}(\rho)\sigma_2 - \sigma_2^{\mathsf{T}} d^2_\rho\mathcal{F}(\rho)\sigma_1 = 0 \ ,\end{aligned}$$

where $d^2_\rho\mathcal{F}(\rho) = (\frac{\partial^2}{\partial\rho_i\partial\rho_j}\mathcal{F}(\rho))_{1\le i,j\le n}$ is the second differential of $\mathcal{F}(\rho)$. □

We next derive the Levi-Civita connection in probability manifold. For two discrete vector fields v, $\tilde{v}$, we further denote $\cdot\circ\cdot\colon \mathbb{R}^{|E|}\times\mathbb{R}^{|E|}\to\mathbb{R}^n$ by

$$v\circ\tilde{v} := ((v,\tilde{v})_i)_{i=1}^n = (\frac{1}{2}\sum_{j\in N(i)} v_{ij}\tilde{v}_{ij})_{i=1}^n \in\mathbb{R}^n\ .$$

Proposition 3 (Levi-Civita connection)**.** *$\nabla^W_\cdot\cdot\colon T_\rho\mathcal{P}_+(G)\times T_\rho\mathcal{P}_+(G)\to T_\rho\mathcal{P}_+(G)$ is defined by*

$$\nabla^W_{\sigma_1}\sigma_2 = -\frac{1}{2}\big[L(\sigma_1)L(\rho)^\dagger\sigma_2 + L(\sigma_2)L(\rho)^\dagger\sigma_1\big] + \frac{1}{2}L(\rho)\Big(\nabla_G L(\rho)^\dagger\sigma_1\circ\nabla_G L(\rho)^\dagger\sigma_2\Big)\ ,$$

where $\big(\nabla_G L(\rho)^\dagger\sigma_1\circ\nabla_G L(\rho)^\dagger\sigma_2\big) = \frac{1}{2}\Big(\sum_{j\in N(i)}(\nabla_{ij}L(\rho)^\dagger\sigma_1)(\nabla_{ij}L(\rho)^\dagger\sigma_2)\Big)_{i=1}^n\in\mathbb{R}^n$.

Proof. Given $\sigma_i\in\mathcal{T}_\rho\mathcal{P}_+(G)$, $i=1,2,3$, by the Koszul formula and Proposition 2, we have

$$\begin{aligned}
&g_W(\nabla^W_{\sigma_1}\sigma_2,\sigma_3)\\
=&\frac{1}{2}\Big\{\sigma_1(\sigma_2^\mathsf{T}L(\rho)^\dagger\sigma_3) + \sigma_2(\sigma_1^\mathsf{T}L(\rho)^\dagger\sigma_3) - \sigma_3(\sigma_1^\mathsf{T}L(\rho)^\dagger\sigma_2)\Big\}\\
=&\frac{1}{2}\frac{d}{dt}|_{t=0}\Big\{(\sigma_2^\mathsf{T}L(\rho+t\sigma_1)^\dagger\sigma_3 + \sigma_1^\mathsf{T}L(\rho+t\sigma_2)^\dagger\sigma_3 - \sigma_1^\mathsf{T}L(\rho+t\sigma_3)^\dagger\sigma_2\Big\}\\
=&-\frac{1}{2}\Big[\sigma_2^\mathsf{T}L(\rho)^\dagger L(\sigma_1)L(\rho)^\dagger\sigma_3 + \sigma_1^\mathsf{T}L(\rho)^\dagger L(\sigma_2)L(\rho)^\dagger\sigma_3\Big] + \frac{1}{2}\sigma_1^\mathsf{T}L(\rho)^\dagger L(\sigma_3)L(\rho)^\dagger\sigma_2\\
=&\sigma_3^\mathsf{T}L(\rho)^\dagger\Big(-\frac{1}{2}[L(\sigma_1)L(\rho)^\dagger\sigma_2 + L(\sigma_2)L(\rho)^\dagger\sigma_1] + \frac{1}{2}L(\rho)(\nabla_G L(\rho)^\dagger\sigma_1\circ\nabla_G L(\rho)^\dagger\sigma_2)\Big)\ ,
\end{aligned}$$

where the last two equalities are shown by the Claim 1 and 2 proved in below. Since $g_W(\nabla^W_{\sigma_1}\sigma_2,\sigma_3) = \sigma_3^\mathsf{T}L(\rho)^\dagger\nabla^W_{\sigma_1}\sigma_2$, for any $\sigma_3\in\mathcal{P}_+(G)$, the result is proved.

Claim 1:

$$\frac{d}{dt}|_{t=0}\sigma_2^\mathsf{T}L(\rho+t\sigma_1)^\dagger\sigma_3 = -\sigma_2^\mathsf{T}L(\rho)^\dagger L(\sigma_1)L(\rho)^\dagger\sigma_3\ .$$

Proof of Claim 1. From Lemma 1, $\sigma_2^\mathsf{T}L(\rho+t\sigma_1)^\dagger\sigma_3 = \sigma_2^\mathsf{T}g(\rho+t\sigma_1)\sigma_3$. Since $L(\rho) = -D^\mathsf{T}\Theta(\rho)D$ is linear w.r.t. ρ, we shall show that

$$\frac{d}{dt}g(\rho+t\sigma_1) = -g(\rho+t\sigma_1)L(\sigma_1)g(\rho+t\sigma_1)\ . \tag{7}$$

This is true since

$$\begin{aligned}
\frac{d}{dt}g(\rho+t\sigma_1) =& -g(\rho+t\sigma_1)\cdot\frac{d}{dt}g(\rho+t\sigma_1)^\dagger\cdot g(\rho+t\sigma_1)\\
=& -g(\rho+t\sigma_1)\cdot\frac{d}{dt}[L(\rho+t\sigma_1)+u_0u_0^\mathsf{T}]\cdot g(\rho+t\sigma_1)\\
=& -g(\rho+t\sigma_1)\cdot L(\sigma_1)\cdot g(\rho+t\sigma_1)\ .
\end{aligned} \tag{8}$$

In above, the first equality holds by

$$0 = \frac{d}{dt}\mathbb{I} = \frac{d}{dt}(g(\rho+t\sigma_1)\cdot g(\rho+t\sigma_1)^{-1}) = \frac{d}{dt}g(\rho+t\sigma_1)\cdot g(\rho+t\sigma_1)^{-1} + g(\rho+t\sigma_1)\cdot\frac{d}{dt}g(\rho+t\sigma_1)^{-1}\ ,$$

and the second equality is true because the element in matrix function $L(\rho)$ is linear w.r.t. ρ. Hence from (7),

$$\begin{aligned}\frac{d}{dt}\sigma_2^{\mathsf{T}}L(\rho+t\sigma_1)^{\dagger}\sigma_3 =&\sigma_2^{\mathsf{T}}\frac{d}{dt}|_{t=0}g(\rho+t\sigma_1)\sigma_3\\ =&-(g(\rho)\sigma_2)^{\mathsf{T}}\cdot L(\sigma_1)\cdot g(\rho)\sigma_3\ .\end{aligned}$$

For any $\sigma\in T_\rho\mathcal{P}_+(G)$, then

$$g(\rho)\sigma=L(\rho)^{\dagger}\sigma+u_0u_0^{\mathsf{T}}\sigma=L(\rho)^{\dagger}\sigma+(\frac{1}{\sqrt{n}}\sum_{i=1}^n\sigma(i))u_0=L(\rho)^{\dagger}\sigma\ .$$

Putting this in (8), we prove the claim 1. □

Claim 2:

$$\sigma_1^{\mathsf{T}}L(\rho)^{\dagger}L(\sigma_3)L(\rho)^{\dagger}\sigma_2=\sigma_3^{\mathsf{T}}L(\rho)^{\dagger}L(\rho)\big(\nabla_GL(\rho)^{\dagger}\sigma_1\circ\nabla_GL(\rho)^{\dagger}\sigma_2\big)\ .$$

Proof of Claim 2. Since $L(\sigma_3)=D^{\mathsf{T}}\Theta(\sigma_3)D$, then

$$\begin{aligned}&\sigma_1^{\mathsf{T}}L(\rho)^{\dagger}L(\sigma_3)L(\rho)^{\dagger}\sigma_2\\ =&(DL(\rho)^{\dagger}\sigma_1)^{\mathsf{T}}\Theta(\sigma_3)DL(\rho)^{\dagger}\sigma_2\\ =&\frac{1}{2}\sum_{(i,j)\in E}(\nabla_{ij}L(\rho)^{\dagger}\sigma_1)(\nabla_{ij}L(\rho)^{\dagger}\sigma_2)\frac{\sigma_3(i)+\sigma_3(j)}{2}\\ =&\frac{1}{2}\sum_{i=1}^n\sigma_2(i)\frac{1}{2}\sum_{j\in N(i)}(\nabla_{ij}L(\rho)^{\dagger}\sigma_1)(\nabla_{ij}L(\rho)^{\dagger}\sigma_2)+\frac{1}{2}\sum_{j=1}^n\sigma_3(j)\frac{1}{2}\sum_{i\in N(j)}(\nabla_{ij}L(\rho)^{\dagger}\sigma_1)(\nabla_{ij}L(\rho)^{\dagger}\sigma_2)\\ =&\frac{1}{4}\sigma_3^{\mathsf{T}}\Big(\sum_{j\in N(i)}(\nabla_{ij}L(\rho)^{\dagger}\sigma_1)(\nabla_{ij}L(\rho)^{\dagger}\sigma_2)\Big)_{i=1}^n+\frac{1}{4}\sigma_3^{\mathsf{T}}\Big(\sum_{i\in N(j)}(\nabla_{ij}L(\rho)^{\dagger}\sigma_1)(\nabla_{ij}L(\rho)^{\dagger}\sigma_2)\Big)_{j=1}^n\\ =&\frac{1}{2}\sigma_3^{\mathsf{T}}\Big(\sum_{j\in N(i)}(\nabla_{ij}L(\rho)^{\dagger}\sigma_1)(\nabla_{ij}L(\rho)^{\dagger}\sigma_2)\Big)_{i=1}^n\ .\end{aligned}$$

where the second last equality holds by relabeling i and j, i.e.

$$\Big(\sum_{j\in N(i)}(\nabla_{ij}L(\rho)^{\dagger}\sigma_1)(\nabla_{ij}L(\rho)^{\dagger}\sigma_2)\Big)_{i=1}^n=\Big(\sum_{i\in N(j)}(\nabla_{ij}L(\rho)^{\dagger}\sigma_1)(\nabla_{ij}L(\rho)^{\dagger}\sigma_2)\Big)_{j=1}^n\ .$$

Since $(\nabla_GL(\rho)^{\dagger}\sigma_1\circ\nabla_GL(\rho)^{\dagger}\sigma_2)=\frac{1}{2}\Big(\sum_{j\in N(i)}\nabla_{ij}L(\rho)^{\dagger}\sigma_1\cdot\nabla_{ij}L(\rho)^{\dagger}\sigma_2\Big)_{i=1}^n\in\mathbb{R}^n$, then

$$\begin{aligned}&\frac{1}{2}\sigma_3^{\mathsf{T}}\Big(\sum_{j\in N(i)}\nabla_{ij}L(\rho)^{\dagger}\sigma_1\nabla_{ij}L(\rho)^{\dagger}\sigma_2\Big)_{i=1}^n\\ =&\sigma_3^{\mathsf{T}}g(\rho)g(\rho)^{-1}(\nabla_GL(\rho)^{\dagger}\sigma_1\circ\nabla_GL(\rho)^{\dagger}\sigma_2)\\ =&\sigma_3^{\mathsf{T}}(L(\rho)+u_0u_0^{\mathsf{T}})(L(\rho)^{\dagger}+u_0u_0^{\mathsf{T}})(\nabla_GL(\rho)^{\dagger}\sigma_1\circ\nabla_GL(\rho)^{\dagger}\sigma_2)\\ =&\sigma_3^{\mathsf{T}}L(\rho)L(\rho)^{\dagger}(\nabla_GL(\rho)^{\dagger}\sigma_1\circ\nabla_GL(\rho)^{\dagger}\sigma_2)\ ,\end{aligned}$$

where the last equality is from $\sigma_3^{\mathsf{T}}u_0=0$. It finishes the proof. □

□

By the Levi-Civita connection, we introduce the Christopher symbol of $(\mathcal{P}_+(G), g_W)$ in the standard Euclidean basis, i.e. $(\frac{\partial}{\partial\rho_1}, \cdots, \frac{\partial}{\partial\rho_n})$.

Proposition 4 (Christopher symbol)**.** *Denote* $\Gamma^{W,k} = (\Gamma^{W,k}_{ij})_{1\le i,j\le n} \in \mathbb{R}^{n\times n}$, *such that*

$$\sum_{1\le i,j\le n} \Gamma^{W,k}_{ij}\sigma_1(i)\sigma_2(j) = (\nabla^W_{\sigma_1}\sigma_2)(k)\ .$$

Then

$$\begin{aligned}\Gamma^{W,k}_{ij} =&\frac{1}{4}\sum_{k'\in N(k)}\omega_{kk'}\Big\{\frac{1}{2}(L^\dagger_{k'i}-L^\dagger_{ki})(\delta_{jk}+\delta_{jk'})+\frac{1}{2}(L^\dagger_{k'j}-L^\dagger_{kj})(\delta_{ik}+\delta_{ik'})\\ &+\theta_{kk'}\sum_{k''\in N(k)}\omega_{kk''}(L^\dagger_{ki}-L^\dagger_{k''i})(L^\dagger_{kj}-L^\dagger_{k''j})\\ &-\theta_{kk'}\sum_{k'''\in N(k')}\omega_{k'k'''}(L^\dagger_{k'i}-L^\dagger_{k'''i})(L^\dagger_{k'j}-L^\dagger_{k'''j})\Big\}\ ,\end{aligned}$$

where $L(\rho)^\dagger = (L^\dagger_{ij})_{1\le i,j\le n}$, $\delta_{ij} = \begin{cases}1 & \text{if } i=j\\ 0 & \text{if } i\neq j\end{cases}$, $\theta_{kk'} = \frac{\rho_k+\rho_{k'}}{2}$ *with the neighborhood index*

$$k''-k-k'-k'''\ .$$

Proof. Denote $\Phi_i := L(\rho)^\dagger\sigma_i \in \mathbb{R}^n$, $i = 1, 2$, i.e. $\Phi_i(k) = \sum_{j=1}^n L^\dagger_{kj}\sigma_i(j)$. And denote

$$\Phi_{12} := \frac{1}{2}(\sum_{k'\in N(k)}\nabla_{kk'}\Phi_1\nabla_{kk'}\Phi_2)^n_{k=1} \in \mathbb{R}^n\ , \tag{9}$$

From Proposition 3,

$$\begin{aligned}\nabla^W_{\sigma_1}\sigma_2(k) =&-\frac{1}{2}\big[L(\sigma_1)L(\rho)^\dagger\sigma_2+L(\sigma_2)L(\rho)^\dagger\sigma_1-L(\rho)(\nabla_G L(\rho)^\dagger\sigma_1\circ\nabla_G L(\rho)^\dagger\sigma_2)\big](k)\\ =&-\frac{1}{2}\sum_{k'\in N(k)}(\Phi_2(k)-\Phi_2(k'))\frac{\sigma_1(k)+\sigma_1(k')}{2} \qquad (C1)\\ &-\frac{1}{2}\sum_{k'\in N(k)}(\Phi_1(k)-\Phi_1(k'))\frac{\sigma_2(k)+\sigma_2(k')}{2} \qquad (C2)\\ &+\frac{1}{2}\sum_{k'\in N(k)}(\Phi_{12}(k)-\Phi_{12}(k'))\theta_{kk'}\ . \qquad (C3)\end{aligned}$$

We next compute (C1), (C2), (C3). Here

$$\begin{aligned}(C1) =& -\frac{1}{2}\sum_{k'\in N(k)}\omega_{kk'}\sum_{j=1}^{n}(L^{\dagger}_{kj}-L^{\dagger}_{k'j})\sigma_2(j)\frac{\sigma_1(k)+\sigma_1(k')}{2}\\ =&\frac{1}{4}\sum_{k'\in N(k)}\omega_{kk'}\sum_{j=1}^{n}(L^{\dagger}_{k'j}-L^{\dagger}_{kj})\sigma_2(j)\sum_{i=1}^{n}(\delta_{ki}+\delta_{k'i})\sigma_1(i)\\ =&\frac{1}{4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k'\in N(k)}\omega_{kk'}(L^{\dagger}_{k'j}-L^{\dagger}_{kj})(\delta_{ki}+\delta_{k'i})\sigma_1(i)\sigma_2(j)\ .\end{aligned}$$

Similarly,

$$(C2) = \frac{1}{4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k'\in N(k)}\omega_{kk'}(L^{\dagger}_{k'i}-L^{\dagger}_{ki})(\delta_{kj}+\delta_{k'j})\sigma_1(i)\sigma_2(j)\ .$$

By applying (9), we have

$$\begin{aligned}(C3) =&\frac{1}{2}\sum_{k'\in N(k)}\theta_{kk'}\omega_{kk'}(\Phi_{12}(k)-\Phi_{12}(k'))\\ =&\frac{1}{4}\sum_{k'\in N(k)}\theta_{kk'}\omega_{kk'}\Big\{\sum_{k''\in N(k)}\omega_{kk''}(\Phi_1(k)-\Phi_1(k''))(\Phi_2(k)-\Phi_2(k''))\\ &-\sum_{k'''\in N(k')}\omega_{k'k'''}(\Phi_1(k')-\Phi_1(k'''))(\Phi_2(k')-\Phi_2(k'''))\Big\}\\ =&\frac{1}{4}\sum_{k'\in N(k)}\theta_{kk'}\omega_{kk'}\Big\{\sum_{k''\in N(k)}\omega_{kk''}\sum_{j=1}^{n}(L^{\dagger}_{jk}-L^{\dagger}_{jk''})\sigma_1(j)\sum_{i=1}^{n}(L^{\dagger}_{ik}-L^{\dagger}_{ik''})\sigma_2(i)\\ &-\sum_{k'''\in N(k')}\omega_{k'k'''}\sum_{j=1}^{n}(L^{\dagger}_{jk'}-L^{\dagger}_{jk'''})\sigma_1(j)\sum_{i=1}^{n}(L^{\dagger}_{ik'}-L^{\dagger}_{ik'''})\sigma_2(i)\Big\}\\ =&\frac{1}{4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k'\in N(k)}\theta_{kk'}\omega_{kk'}\Big\{\sum_{k''\in N(k)}\omega_{kk''}(L^{\dagger}_{jk}-L^{\dagger}_{jk''})(L^{\dagger}_{ik}-L^{\dagger}_{ik''})\\ &-\sum_{k'''\in N(k')}\omega_{k'k'''}(L^{\dagger}_{jk'}-L^{\dagger}_{jk'''})(L^{\dagger}_{ik'}-L^{\dagger}_{ik'''})\Big\}\sigma_1(j)\sigma_2(i)\ .\end{aligned}$$

From $\sum_{1\le i,j\le n}\Gamma^{W,k}_{ij}\sigma_1(i)\sigma_2(j) = (\nabla^W_{\sigma_1}\sigma_2)(k) = (C1)+(C2)+(C3)$, we prove the result.

□

Proposition 5 (Parallel transport)**.** *Denote $\rho\colon (a,b)\to\mathcal{P}_+(G)$ be a smooth curve. Consider $\sigma(t)\in\mathcal{P}_+(G)$ be a vector field along curve ρ, then the equation for $\sigma(t)$ to be parallel along $\rho(t)$ is*

$$\dot\sigma-\frac{1}{2}\Big(L(\sigma)L(\rho)^{\dagger}\dot\rho+L(\dot\rho)L(\rho)^{\dagger}\sigma\Big)+\frac{1}{2}L(\rho)\big(\nabla_G L(\rho)^{\dagger}\dot\rho\circ\nabla_G L(\rho)^{\dagger}\sigma\big)=0\ .$$

The geodesic equation satisfies

$$\ddot\rho - L(\dot\rho)L(\rho)^{\dagger}\dot\rho + \frac{1}{2}L(\rho)\big(\nabla_G L(\rho)^{\dagger}\dot\rho \circ \nabla_G L(\rho)^{\dagger}\dot\rho\big) = 0 \ . \tag{10}$$

Proof. The parallel equation is derived by computing $\nabla^W_{\dot\rho(t)}\sigma(t) = \Big(\dot\sigma_k + \sum_{1\le i,j\le n}\Gamma^{W,k}_{ij}\sigma_i\dot\rho_j\Big)_{k=1}^n = 0$. Let $\sigma(t) = \dot\rho(t)$, then the geodesic is derived by setting $\nabla^W_{\dot\rho(t)}\dot\rho(t) = \Big(\ddot\rho_k + \sum_{1\le i,j\le n}\Gamma^{W,k}_{ij}\dot\rho_i\dot\rho_j\Big)_{k=1}^n = 0$. □

We are ready to give the curvature tensor, $R_W(\cdot,\cdot)\cdot\colon T_\rho\mathcal{P}_+(G)\times T_\rho\mathcal{P}_+(G)\times T_\rho\mathcal{P}_+(G)\to T_\rho\mathcal{P}_+(G)$.

Proposition 6 (Curvature tensor)**.** *Given σ_1, σ_2, σ_3, $\sigma_4\in T_\rho\mathcal{P}_+(G)$, then*

$$\begin{aligned}
&g_W(R_W(\sigma_1,\sigma_2)\sigma_3,\sigma_4)\\
=\ &\frac{1}{2}\Big\{\sigma_2^{\mathsf T}L(\rho)^{\dagger}L(m(\sigma_1,\sigma_3))L(\rho)^{\dagger}\sigma_4 + \sigma_1^{\mathsf T}L(\rho)^{\dagger}L(m(\sigma_2,\sigma_4))L(\rho)^{\dagger}\sigma_3\\
&-\sigma_2^{\mathsf T}L(\rho)^{\dagger}L(m(\sigma_1,\sigma_4))L(\rho)^{\dagger}\sigma_3 - \sigma_1^{\mathsf T}L(\rho)^{\dagger}L(m(\sigma_2,\sigma_3))L(\rho)^{\dagger}\sigma_4\Big\}\\
&+\frac{1}{4}\Big\{2n(\sigma_1,\sigma_2)^{\mathsf T}L(\rho)^{\dagger}n(\sigma_3,\sigma_4) + n(\sigma_1,\sigma_3)^{\mathsf T}L(\rho)^{\dagger}n(\sigma_2,\sigma_4) - n(\sigma_2,\sigma_3)^{\mathsf T}L(\rho)^{\dagger}n(\sigma_1,\sigma_4)\Big\}\ ,
\end{aligned}$$

where m, $n\colon T_\rho\mathcal{P}_+(G)\times T_\rho\mathcal{P}_+(G)\to T_\rho\mathcal{P}_+(G)$ are symmetric, antisymmetric operators respectively, defined by

$$m(\sigma_a,\sigma_b) := \nabla^W_{\sigma_a}\sigma_b = -\frac{1}{2}[L(\sigma_a)L(\rho)^{\dagger}\sigma_b + L(\sigma_b)L(\rho)^{\dagger}\sigma_a] + \frac{1}{2}L(\rho)(\nabla_G L(\rho)^{\dagger}\sigma_a\circ\nabla_G L(\rho)^{\dagger}\sigma_b)\ ,$$

and

$$n(\sigma_a,\sigma_b) := L(\sigma_a)L(\rho)^{\dagger}\sigma_b - L(\sigma_b)L(\rho)^{\dagger}\sigma_a\ .$$

Proof. Since $[\sigma_1,\sigma_2]_W = 0$, then $R_W(\sigma_1,\sigma_2)\sigma_3 = \nabla^W_{\sigma_1}\nabla^W_{\sigma_2}\sigma_3 - \nabla^W_{\sigma_2}\nabla^W_{\sigma_1}\sigma_3$. Thus

$$\begin{aligned}
g_W(R_W(\sigma_1,\sigma_2)\sigma_3,\sigma_4) =&\sigma_1(g_W(\nabla^W_{\sigma_2}\sigma_3,\sigma_4)) - g_W(\nabla^W_{\sigma_2}\sigma_3,\nabla^W_{\sigma_1}\sigma_4)\\
&-\sigma_2(g_W(\nabla^W_{\sigma_1}\sigma_3,\sigma_4)) + g_W(\nabla^W_{\sigma_1}\sigma_3,\nabla^W_{\sigma_2}\sigma_4)\ .
\end{aligned}\tag{11}$$

We first compute $\sigma_1(g_W(\nabla^W_{\sigma_2}\sigma_3,\sigma_4))$. Denote $\dot\rho(t)=\sigma_1$ and notice Claim 1, then

$$\begin{aligned}
&\sigma_1(g_W(\nabla^W_{\sigma_2}\sigma_3,\sigma_4))\\
=&\frac{d}{dt}|_{t=0}\Big\{-\frac{1}{2}\Big[\sigma_3^{\mathsf T}L(\rho(t))^{\dagger}L(\sigma_2)L(\rho(t))^{\dagger}\sigma_4 + \sigma_2^{\mathsf T}L(\rho(t))^{\dagger}L(\sigma_3)L(\rho(t))^{\dagger}\sigma_4\Big]\\
&\qquad+\frac{1}{2}\sigma_2^{\mathsf T}L(\rho(t))^{\dagger}L(\sigma_4)L(\rho(t))^{\dagger}\sigma_3\Big\}\\
=&\frac{1}{2}\Big[\sigma_3^{\mathsf T}L(\rho)^{\dagger}L(\sigma_1)L(\rho)^{\dagger}L(\sigma_2)L(\rho)^{\dagger}\sigma_4 + \sigma_3^{\mathsf T}L(\rho)^{\dagger}L(\sigma_2)L(\rho)^{\dagger}L(\sigma_1)L(\rho)^{\dagger}\sigma_4\\
&+\sigma_2^{\mathsf T}L(\rho)^{\dagger}L(\sigma_1)L(\rho)^{\dagger}L(\sigma_3)L(\rho)^{\dagger}\sigma_4 + \sigma_2^{\mathsf T}L(\rho)^{\dagger}L(\sigma_3)L(\rho)^{\dagger}L(\sigma_1)L(\rho)^{\dagger}\sigma_4\Big]\\
&-\frac{1}{2}\Big[\sigma_2^{\mathsf T}L(\rho)^{\dagger}L(\sigma_1)L(\rho)^{\dagger}L(\sigma_4)L(\rho)^{\dagger}\sigma_3 + \sigma_2^{\mathsf T}L(\rho)^{\dagger}L(\sigma_4)L(\rho)^{\dagger}L(\sigma_1)L(\rho)^{\dagger}\sigma_3\Big]\ .
\end{aligned}\tag{12}$$

Similarly,

$$
\begin{aligned}
&\sigma_2(g_W(\nabla^W_{\sigma_1}\sigma_3,\sigma_4))\\
=&\frac{1}{2}\Big[\sigma_3^{\mathsf T}L(\rho)^{\dagger}L(\sigma_2)L(\rho)^{\dagger}L(\sigma_1)L(\rho)^{\dagger}\sigma_4+\sigma_3^{\mathsf T}L(\rho)^{\dagger}L(\sigma_1)L(\rho)^{\dagger}L(\sigma_2)L(\rho)^{\dagger}\sigma_4\\
&+\sigma_1^{\mathsf T}L(\rho)^{\dagger}L(\sigma_2)L(\rho)^{\dagger}L(\sigma_3)L(\rho)^{\dagger}\sigma_4+\sigma_1^{\mathsf T}L(\rho)^{\dagger}L(\sigma_3)L(\rho)^{\dagger}L(\sigma_2)L(\rho)^{\dagger}\sigma_4\Big]\\
-&\frac{1}{2}\Big[\sigma_1^{\mathsf T}L(\rho)^{\dagger}L(\sigma_2)L(\rho)^{\dagger}L(\sigma_4)L(\rho)^{\dagger}\sigma_3+\sigma_1^{\mathsf T}L(\rho)^{\dagger}L(\sigma_4)L(\rho)^{\dagger}L(\sigma_2)L(\rho)^{\dagger}\sigma_3\Big]\ .
\end{aligned}
\tag{13}
$$

We next derive $g_W(\nabla^W_{\sigma_2}\sigma_3,\nabla^W_{\sigma_1}\sigma_4)$. From proposition 3, we have

$$
\begin{aligned}
&g_W(\nabla^W_{\sigma_2}\sigma_3,\nabla^W_{\sigma_1}\sigma_4)\\
=&\frac{1}{2}\Big\{\sigma_2^{\mathsf T}L(\rho)^{\dagger}L(\nabla^W_{\sigma_1}\sigma_4)L(\rho)^{\dagger}\sigma_3-\sigma_2^{\mathsf T}L(\rho)^{\dagger}L(\sigma_3)L(\rho)^{\dagger}\nabla^W_{\sigma_1}\sigma_4-\sigma_3^{\mathsf T}L(\rho)^{\dagger}L(\sigma_2)L(\rho)^{\dagger}\nabla^W_{\sigma_1}\sigma_4\Big\}\ .
\end{aligned}
$$

Notice

$$
\nabla^W_{\sigma_1}\sigma_4=\frac{1}{2}\{L(\rho)(\nabla_G L(\rho)^{\dagger}\sigma_1\circ\nabla_G L(\rho)^{\dagger}\sigma_4)-L(\sigma_4)L(\rho)^{\dagger}\sigma_1-L(\sigma_1)L(\rho)^{\dagger}\sigma_4\}
$$

and $L(\cdot)$ is a matrix function linear on ρ. Thus

$$
\begin{aligned}
&g_W(\nabla^W_{\sigma_2}\sigma_3,\nabla^W_{\sigma_1}\sigma_4)\\
=&\frac{1}{4}\Big\{\sigma_2^{\mathsf T}L(\rho)^{\dagger}L\Big(L(\rho)(\nabla_G L(\rho)^{\dagger}\sigma_1\circ\nabla_G L(\rho)^{\dagger}\sigma_4)\Big)L(\rho)^{\dagger}\sigma_3\\
&+\sigma_1^{\mathsf T}L(\rho)^{\dagger}L\Big(L(\rho)(\nabla_G L(\rho)^{\dagger}\sigma_2\circ\nabla_G L(\rho)^{\dagger}\sigma_3)\Big)L(\rho)^{\dagger}\sigma_4\\
&-\sigma_2^{\mathsf T}L(\rho)^{\dagger}L\Big(L(\sigma_1)L(\rho)^{\dagger}\sigma_4+L(\sigma_4)L(\rho)^{\dagger}\sigma_1\Big)L(\rho)^{\dagger}\sigma_3\\
&-\sigma_1^{\mathsf T}L(\rho)^{\dagger}L\Big(L(\sigma_2)L(\rho)^{\dagger}\sigma_3+L(\sigma_3)L(\rho)^{\dagger}\sigma_2\Big)L(\rho)^{\dagger}\sigma_4\\
&+\sigma_2^{\mathsf T}L(\rho)^{\dagger}L(\sigma_3)L(\rho)^{\dagger}L(\sigma_4)L(\rho)^{\dagger}\sigma_1+\sigma_2^{\mathsf T}L(\rho)^{\dagger}L(\sigma_3)L(\rho)^{\dagger}L(\sigma_1)L(\rho)^{\dagger}\sigma_4\\
&+\sigma_3^{\mathsf T}L(\rho)^{\dagger}L(\sigma_2)L(\rho)^{\dagger}L(\sigma_4)L(\rho)^{\dagger}\sigma_1+\sigma_3^{\mathsf T}L(\rho)^{\dagger}L(\sigma_2)L(\rho)^{\dagger}L(\sigma_1)L(\rho)^{\dagger}\sigma_4\Big\}\ .
\end{aligned}
\tag{14}
$$

In the derivation of (14), we use the following results, which can be proved similarly as the ones in Claim 2:

$$
\sigma_2^{\mathsf T}L(\rho)^{\dagger}L(\sigma_3)\Big(\nabla_G L(\rho)^{\dagger}\sigma_1\circ\nabla_G L(\rho)^{\dagger}\sigma_4\Big)=\sigma_1^{\mathsf T}L(\rho)^{\dagger}L\Big(L(\sigma_3)L(\rho)^{\dagger}\sigma_2\Big)L(\rho)^{\dagger}\sigma_4\ ,
$$

and

$$
\begin{aligned}
&\sigma_2^{\mathsf T}L(\rho)^{\dagger}L\Big(L(\rho)(\nabla_G L(\rho)^{\dagger}\sigma_1\circ\nabla_G L(\rho)^{\dagger}\sigma_4)\Big)L(\rho)^{\dagger}\sigma_3\\
=&\sigma_1^{\mathsf T}L(\rho)^{\dagger}L\Big(L(\rho)(\nabla_G L(\rho)^{\dagger}\sigma_2\circ\nabla_G L(\rho)^{\dagger}\sigma_3)\Big)L(\rho)^{\dagger}\sigma_4\\
=&(L(\rho)^{\dagger}\sigma_2\circ L(\rho)^{\dagger}\sigma_3)^{\mathsf T}L(\rho)(\nabla_G L(\rho)^{\dagger}\sigma_1\circ\nabla_G L(\rho)^{\dagger}\sigma_4)\ .
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&g_W(\nabla^W_{\sigma_1}\sigma_3, \nabla^W_{\sigma_2}\sigma_4)\\
=&\frac{1}{4}\Big\{\sigma_1^\mathsf{T} L(\rho)^\dagger L\Big(L(\rho)(\nabla_G L(\rho)^\dagger\sigma_2 \circ \nabla_G L(\rho)^\dagger\sigma_4)\Big)L(\rho)^\dagger\sigma_3\\
&+\sigma_2^\mathsf{T} L(\rho)^\dagger L\Big(L(\rho)(\nabla_G L(\rho)^\dagger\sigma_1 \circ \nabla_G L(\rho)^\dagger\sigma_3)\Big)L(\rho)^\dagger\sigma_4\\
&-\sigma_1^\mathsf{T} L(\rho)^\dagger L\Big(L(\sigma_2)L(\rho)^\dagger\sigma_4 + L(\sigma_4)L(\rho)^\dagger\sigma_2\Big)L(\rho)^\dagger\sigma_3\\
&-\sigma_2^\mathsf{T} L(\rho)^\dagger L\Big(L(\sigma_1)L(\rho)^\dagger\sigma_3 + L(\sigma_3)L(\rho)^\dagger\sigma_1\Big)L(\rho)^\dagger\sigma_4\\
&+\sigma_1^\mathsf{T} L(\rho)^\dagger L(\sigma_3)L(\rho)^\dagger L(\sigma_4)L(\rho)^\dagger\sigma_2 + \sigma_1^\mathsf{T} L(\rho)^\dagger L(\sigma_3)L(\rho)^\dagger L(\sigma_2)L(\rho)^\dagger\sigma_4\\
&+\sigma_3^\mathsf{T} L(\rho)^\dagger L(\sigma_1)L(\rho)^\dagger L(\sigma_4)L(\rho)^\dagger\sigma_2 + \sigma_3^\mathsf{T} L(\rho)^\dagger L(\sigma_1)L(\rho)^\dagger L(\sigma_2)L(\rho)^\dagger\sigma_4\Big\}\ .
\end{aligned}
\tag{15}
$$

Substituting (12), (13), (14), (15) into (11), we derive the result. □

3.1. **Jacobi equation.** In this sequel, we establish two formulations of Jacobi fields in $(\mathcal{P}_+(G), g_W)$. As in Riemannian geometry, the Jacobi equation refers $\nabla^W_{\dot\rho}\nabla^W_{\dot\rho}J + R^W(J,\dot\rho)\dot\rho = 0$, with given initial condition $J(0)$, $\dot J(0) \in \mathbb{R}^n$. I.e. denote $\rho(t)$ as the geodesic and $(X_1(t), \cdots, X_{n-1}(t))$ as an orthonormal frame along $\rho(t)$. Consider $J(t) = \sum_{i=1}^{n-1} a_i(t)X_i(t)$, thus

$$\ddot a_i + \sum_{k=1}^{n-1} a_k g_W(R_W(X_i,\dot\rho)\dot\rho, X_k) = 0\ , \quad \text{for } i = 1, 2, \cdots, n-1\ .$$

The other formulation of the Jacobi equation is derived by the calculus of variation.

Lemma 2 (Variations of energy)**.** *Consider an infinitesmall deformation $\rho_\epsilon(t) = \rho(t) + \epsilon h(t) \in \mathcal{P}_+(G)$ with $h(t) = (h_i(t))_{i=1}^n \in \big(C^\infty[0,1]\big)^n$, $\sum_{i=1}^n h_i(t) = 0$, and $h(0) = h(1) = 0$,*

$$\mathcal{E}(\rho_\epsilon) = \int_0^1 \frac{1}{2}\dot\rho_\epsilon(t)^\mathsf{T} L(\rho_\epsilon(t))^\dagger \dot\rho_\epsilon(t)dt = \mathcal{E}(\rho) + \epsilon\delta\mathcal{E}(\rho)(h) + \frac{\epsilon^2}{2}\delta^2\mathcal{E}(\rho)(h) + o(\epsilon^2)\ . \tag{16}$$

The first and second variations defined in (16) *satisfy*

$$\delta\mathcal{E}(\rho)(h) = \int_0^1\ \dot\rho^\mathsf{T} L(\rho)^\dagger\big(\dot h - \frac{1}{2}L(h)L(\rho)^\dagger\dot\rho\big)\ dt\ ,$$

and

$$\delta^2\mathcal{E}(\rho)(h) = \int_0^1\ \big(\dot h - L(h)L(\rho)^\dagger\dot\rho\big)^\mathsf{T} L(\rho)^\dagger\big(\dot h - L(h)L(\rho)^\dagger\dot\rho\big)\ dt\ . \tag{17}$$

Proof. Since $\dot\rho_\epsilon \in T_\rho\mathcal{P}_+(G)$ and Theorem 1 holds, then $\dot\rho_\epsilon^\mathsf{T} L(\rho_\epsilon)^\dagger\dot\rho_\epsilon = \dot\rho_\epsilon^\mathsf{T} g(\rho_\epsilon)\dot\rho_\epsilon$. We then derive the result by the Taylor expansion.

$$\begin{aligned}
\mathcal{E}(\rho_\epsilon) =& \int_0^1 \frac{1}{2}(\dot\rho+\epsilon\dot h)^\mathsf{T} L(\rho+\epsilon h)^\dagger(\dot\rho+\epsilon\dot h)\ dt = \int_0^1 \frac{1}{2}(\dot\rho+\epsilon\dot h)^\mathsf{T} g(\rho+\epsilon h)(\dot\rho+\epsilon\dot h)\ dt \\
=& \int_0^1 \frac{1}{2}\dot\rho^\mathsf{T} g(\rho+\epsilon h)\dot\rho + \epsilon\dot h^\mathsf{T} g(\rho+\epsilon h)\dot\rho + \frac{\epsilon^2}{2}\dot h^\mathsf{T} g(\rho+\epsilon h)\dot h\ dt \\
=& \int_0^1 \Big\{\frac{1}{2}\dot\rho^\mathsf{T} L(\rho)^\dagger\dot\rho + \epsilon\big[\frac{1}{2}\dot\rho^\mathsf{T}\frac{d}{d\epsilon}|_{\epsilon=0} g(\rho+\epsilon h)\dot\rho + \dot h^\mathsf{T} L(\rho)^\dagger\dot\rho\big] \\
&+ \frac{\epsilon^2}{2}\big[\frac{1}{2}\dot\rho^\mathsf{T}\frac{d^2}{d\epsilon^2}|_{\epsilon=0} g(\rho+\epsilon h)\dot\rho + \ 2\dot h^\mathsf{T}\frac{d}{d\epsilon}|_{\epsilon=0} g(\rho+\epsilon h)\dot\rho \ + \ \dot h^\mathsf{T} L(\rho)^\dagger h\big]\Big\}\ dt + o(\epsilon^2)\ .
\end{aligned} \tag{18}$$

To continue the derivation, we need the following claim.

Claim 3:

$$\begin{cases} \dfrac{d}{d\epsilon} g(\rho+\epsilon h) = - g(\rho+\epsilon h)\cdot L(h)\cdot g(\rho+\epsilon h) \\ \dfrac{d^2}{d\epsilon^2} g(\rho+\epsilon h) = 2g(\rho+\epsilon h)\cdot L(h)\cdot g(\rho+\epsilon h)\cdot L(h)\cdot g(\rho+\epsilon h)\ . \end{cases} \tag{19}$$

Proof of Claim 3. We show (19) by the following steps. From (7), we have

$$\begin{aligned}
\frac{d}{d\epsilon} g(\rho+\epsilon h) =& - g(\rho+\epsilon h)\cdot\frac{d}{d\epsilon} g(\rho+\epsilon h)^{-1}\cdot g(\rho+\epsilon h) \\
=& - g(\rho+\epsilon h)\cdot\frac{d}{d\epsilon}(L(\rho+\epsilon h)+u_0u_0^\mathsf{T})\cdot g(\rho+\epsilon h) \\
=& - g(\rho+\epsilon h)\cdot L(h)\cdot g(\rho+\epsilon h)\ .
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
\frac{d^2}{d\epsilon^2} g(\rho+\epsilon h) =& \frac{d}{d\epsilon}\Big(\frac{d}{d\epsilon} g(\rho+\epsilon h)\Big) = -\frac{d}{d\epsilon}\Big(g(\rho+\epsilon h)\cdot L(h)\cdot g(\rho+\epsilon h)\Big) \\
=& 2g(\rho+\epsilon h)\cdot L(h)\cdot g(\rho+\epsilon h)\cdot L(h)\cdot g(\rho+\epsilon h)\ .
\end{aligned}$$

□

Substituting (19) into (18) and using the fact $g(\rho)\dot\rho = L(\rho)^\dagger\dot\rho$ by $\dot\rho \in T_\rho\mathcal{P}_+(G)$, we derive the first and second variations in Lemma 2. □

Theorem 2 (Jacobi equation)**.** *Consider the geodesic $\rho(t) \in \mathcal{P}_+(G)$ connecting ρ^0 and ρ^1. The Jacobi equation of $h(t)$ along the geodesic $\rho(t)$ satisfies*

$$\dot h - L(h)L(\rho)^\dagger\dot\rho = 0\ , \quad h(0) = h(1) = 0\ . \tag{20}$$

Proof. Denote $\delta^2\mathcal{E}(\rho)(h) = \int_0^1 A(h,\dot h)dt$, where

$$A(h,\dot h) = \big(\dot h - L(h)L(\rho)^\dagger\dot\rho\big)^\mathsf{T} L(\rho)^\dagger\big(\dot h - L(h)L(\rho)^\dagger\dot\rho\big)\ .$$

The Euler-Lagrange equation of (17), $\frac{d}{dt}\nabla_{\dot{h}}A(h,\dot{h}) = \nabla_h A(h,\dot{h})$, satisfies

$$\frac{d}{dt}\Big(L(\rho)^\dagger\big(\dot{h}-L(h)L(\rho)^\dagger\dot{\rho}\big)\Big) = \Big(\nabla_G L(\rho)^\dagger\dot{\rho}\circ\nabla_G L(\rho)^\dagger\big(\dot{h}-L(h)L(\rho)^\dagger\dot{\rho}\big)\Big) \ . \tag{21}$$

On one hand, if $h(t)$ satisfies (20), then it is a solution of (21). On the other hand, if $h(t)$ satisfies (21), then $\delta^2\mathcal{E}(h) = \frac{1}{2}\int_0^1 h^T\Big(\frac{d}{dt}\nabla_{\dot{h}}A(h,\dot{h}) - \nabla_h A(h,\dot{h})\Big)dt = 0$. Since $\delta^2\mathcal{E}(h) = \int_0^1 \ \big(\dot{h}-L(h)L(\rho)^\dagger\dot{\rho}\big)^\mathsf{T} L(\rho)^\dagger\big(\dot{h}-L(h)L(\rho)^\dagger\dot{\rho}\big)\ dt = 0$ and $L(\rho)^\dagger$ is postive definite in $T_\rho\mathcal{P}_+(G)$, then $h(t)$ satisfies (20). □

3.2. **Volume form, Divergence, Laplace-Beltrami and Hessian operator.** In this sequel, we study the volume form in $(\mathcal{P}_+(G), g_W)$, based on which we introduce the divergence and Laplace-Beltrami operator.

Theorem 3 (Volume form)**.** *Denote vol as the volume form of $\mathcal{P}_+(G)$ in the Euclidean metric. Then the volume form of $(\mathcal{P}_+(G), g_W)$ satisfies*

$$dvol_W = \Pi(\rho)^{-\frac{1}{2}}dvol \ , \quad with \quad \Pi(\rho) := \Pi_{i=1}^{n-1}\lambda_i(\rho) \ ,$$

where $\lambda_i(\rho)$ are positive eigenvalues of $L(\rho)$.

Proof. Since $\mathcal{M}_+(G)$ is a smooth oriented manifold with dimension n, then $\mathcal{P}_+(G)$ is a submanifold of $\mathcal{M}_+(G)$ with co-dimension 1. The orientation of $\mathcal{P}_+(G)$ is induced by its normal unit vector field u_0 and the orientation on $\mathcal{M}_+(G)$. Then the volume form on $(\mathcal{P}_+(G), g_W)$ can be given as follows. Denote the volume form of $(\mathcal{M}_+(G), g_\mathcal{M})$ by vol_g, then for any $\rho\in\mathcal{P}_+(G)$ and any $u_1,\cdots,u_{n-1}\in T_\rho\mathcal{P}_+(G)$,

$$(d\mathrm{vol}_W)(u_1,\cdots,u_{n-1}) = (d\mathrm{vol}_g)(u_0,u_1,\cdots,u_{n-1}) \ .$$

Since $d\mathrm{vol}_g = \sqrt{\det(g(\rho))}d\mathrm{vol}$ and $g(\rho)u_0 = 1$, it is clear that $d\mathrm{vol}_W = \sqrt{\det(g(\rho))}d\mathrm{vol} = \Pi(\rho)^{-\frac{1}{2}}d\mathrm{vol}$. □

We continue the derivations based on volume forms in $(\mathcal{P}_+(G), g_W)$.

Proposition 7 (Divergence, Laplace-Beltrami operators)**.** *Denote $\mathcal{G}(\rho) = (\mathcal{G}_i(\rho))_{i=1}^n$, $\mathcal{G}_i(\rho)\in C^\infty(\mathcal{P}_+(G))$. The divergence operator $div_W(\cdot)\colon (C^\infty(\mathcal{P}_+(G)))^n \to C^\infty(\mathcal{P}_+(G))$ satisfies*

$$div_W\mathcal{G}(\rho) = \Pi(\rho)^{\frac{1}{2}}\nabla_\rho\cdot(\mathcal{G}(\rho)\Pi(\rho)^{-\frac{1}{2}}) = \nabla_\rho\cdot\mathcal{G}(\rho) - \frac{1}{2}\mathcal{G}(\rho)^\mathsf{T}d_\rho\log\Pi(\rho) \ ,$$

where $\nabla_\rho\cdot = \sum_{i=1}^n\frac{\partial}{\partial\rho_i}\cdot$.

Denote $\mathcal{F}(\rho)\in C^\infty(\mathcal{P}_+(G))$. The Laplace-Beltrami operator $\Delta_W\colon C^\infty(\mathcal{P}_+(G)) \to C^\infty(\mathcal{P}_+(G))$ satisfies

$$\begin{aligned}\Delta_W\mathcal{F}(\rho) =& \Pi(\rho)^{\frac{1}{2}}\nabla_\rho\cdot(\Pi(\rho)^{-\frac{1}{2}}L(\rho)d_\rho\mathcal{F}(\rho))\\ =& tr(L(\rho)\cdot d_\rho^2\mathcal{F}(\rho)) - \frac{1}{2}(d_\rho\mathcal{F}(\rho))^\mathsf{T}L(\rho)\big(d_\rho\log\Pi(\rho)\big)\\ =& \sum_{(i,j)\in\vec{E}}\omega_{ij}\frac{\rho_i+\rho_j}{2}(\frac{\partial^2}{\partial\rho_i^2} - 2\frac{\partial^2}{\partial\rho_i\partial\rho_j} + \frac{\partial^2}{\partial\rho_j^2})\mathcal{F}(\rho) - \frac{1}{2}\sum_{(i,j)\in\vec{E}}\nabla_{ij}d_\rho\log\Pi(\rho)\cdot\nabla_{ij}d_\rho\mathcal{F}(\rho)\frac{\rho_i+\rho_j}{2} \ .\end{aligned}$$

Proof. Consider a test function $\mathcal{F}(\rho) \in C^\infty(\mathcal{P}_+(G))$ with compact support in $\mathcal{P}_+(G)$. Then

$$\begin{aligned}\int_{\mathcal{P}_+(G)} g_W(\nabla_W \mathcal{F}(\rho), \mathcal{G}(\rho)) d\mathrm{vol}_W &= \int_{\mathcal{P}_+(G)} d_\rho \mathcal{F}(\rho) \cdot \mathcal{G}(\rho) \Pi(\rho)^{-\frac{1}{2}} d\mathrm{vol} \\ &= -\int_{\mathcal{P}_+(G)} \mathcal{F}(\rho) \nabla_\rho \cdot (\mathcal{G}(\rho)\Pi(\rho)^{-\frac{1}{2}}) d\mathrm{vol} \\ &= -\int_{\mathcal{P}_+(G)} \mathcal{F}(\rho)\Pi(\rho)^{\frac{1}{2}} \nabla_\rho \cdot (\mathcal{G}(\rho)\Pi(\rho)^{-\frac{1}{2}}) d\mathrm{vol}_W \ ,\end{aligned}$$

which finishes the proof.

From the divergence operator and noticing $\nabla_W \mathcal{F}(\rho) = L(\rho) d_\rho \mathcal{F}(\rho)$, we have

$$\Delta_W \mathcal{F}(\rho) = \mathrm{div}_W(\nabla_W \mathcal{F}(\rho)) = \nabla_\rho \cdot (L(\rho) d_\rho \mathcal{F}(\rho)) + (d_\rho \mathcal{F}(\rho))^\mathsf{T} L(\rho)(d_\rho \log \Pi(\rho)^{-\frac{1}{2}}) \ .$$

Since

$$\begin{aligned}&\nabla_\rho \cdot (L(\rho) d_\rho \mathcal{F}(\rho)) \\ =&\sum_{i=1}^n \frac{\partial}{\partial \rho_i} \Big(\sum_{j\in N(i)} (\frac{\partial}{\partial \rho_i} - \frac{\partial}{\partial \rho_j}) \mathcal{F}(\rho) \frac{\rho_i+\rho_j}{2} \omega_{ij} \Big) \\ =&\sum_{i=1}^n \sum_{j\in N(i)} (\frac{\partial^2}{\partial \rho_i^2} - \frac{\partial^2}{\partial \rho_j \partial \rho_i}) \mathcal{F}(\rho) \frac{\rho_i+\rho_j}{2}\omega_{ij} - \frac{1}{2}\sum_{i=1}^n \sum_{j\in N(i)} (\frac{\partial}{\partial \rho_i} - \frac{\partial}{\partial \rho_j}) \mathcal{F}(\rho)\omega_{ij} \\ =& \sum_{(i,j)\in \vec{E}} (\frac{\partial^2}{\partial \rho_i^2} - \frac{\partial^2}{\partial \rho_j \partial \rho_i}) \mathcal{F}(\rho) \frac{\rho_i+\rho_j}{2}\omega_{ij} - \frac{1}{2}\sum_{(i,j)\in \vec{E}} (\frac{\partial}{\partial \rho_i} - \frac{\partial}{\partial \rho_j}) \mathcal{F}(\rho)\omega_{ij} \\ +& \sum_{(j,i)\in \vec{E}} (\frac{\partial^2}{\partial \rho_i^2} - \frac{\partial^2}{\partial \rho_j \partial \rho_i}) \mathcal{F}(\rho) \frac{\rho_i+\rho_j}{2}\omega_{ij} - \frac{1}{2}\sum_{(j,i)\in \vec{E}} (\frac{\partial}{\partial \rho_i} - \frac{\partial}{\partial \rho_j}) \mathcal{F}(\rho)\omega_{ij} \qquad \text{Relabel } i \text{ by } j. \\ =& \sum_{(i,j)\in \vec{E}} (\frac{\partial^2}{\partial \rho_i^2} - 2\frac{\partial^2}{\partial \rho_j \partial \rho_i} + \frac{\partial^2}{\partial \rho_j^2}) \mathcal{F}(\rho) \frac{\rho_i+\rho_j}{2}\omega_{ij} \\ =&\mathrm{tr}(L(\rho)\cdot d_\rho^2 \mathcal{F}(\rho)) \ ,\end{aligned}$$

and $(d_\rho \mathcal{F}(\rho))^\mathsf{T} L(\rho)(d_\rho \log \Pi(\rho)^{-\frac{1}{2}}) = \sum_{(i,j)\in \vec{E}} \nabla_{ij} d_\rho \log \Pi(\rho)^{-\frac{1}{2}} \cdot \nabla_{ij} d_\rho \mathcal{F}(\rho) \frac{\rho_i+\rho_j}{2}$, we prove the result. □

In the last, we define the Hessian operator on $(\mathcal{P}_+(G), g_W)$, i.e. $\mathrm{Hess}_W(\cdot,\cdot) \colon T_\rho \mathcal{P}_+(G) \times T_\rho \mathcal{P}_+(G) \to \mathbb{R}$.

Proposition 8 (Hessian operator)**.** *Given $\sigma_1, \sigma_2 \in T_\rho \mathcal{P}_+(G)$, then*

$$\begin{aligned}\mathit{Hess}_W \mathcal{F}(\rho)(\sigma_1,\sigma_2) =&\sigma_1^\mathsf{T} d_\rho^2 \mathcal{F}(\rho)\sigma_2 + \frac{1}{2}\Big\{ d_\rho \mathcal{F}(\rho)^\mathsf{T} L(\sigma_1) L(\rho)^\dagger \sigma_2 + d_\rho \mathcal{F}(\rho)^\mathsf{T} L(\sigma_2) L(\rho)^\dagger \sigma_1 \\ &- d_\rho \mathcal{F}(\rho)^\mathsf{T} L(\rho) \Big(\nabla_G L(\rho)^\dagger \sigma_1 \circ \nabla_G L(\rho)^\dagger \sigma_2 \Big) \Big\} \ .\end{aligned} \tag{22}$$

Proof. From the definition of Hessian on a Riemannian manifold, we have

$$\begin{aligned}\mathrm{Hess}_W\mathcal{F}(\rho)(\sigma_1,\sigma_2) =& g_W(\sigma_1, \nabla^W_{\sigma_2}\nabla_W\mathcal{F}(\rho))\\ =&\sigma_2\big(g_W(\sigma_1,\nabla_W\mathcal{F}(\rho))\big) - g_W(\nabla^W_{\sigma_2}\sigma_1, \nabla_W\mathcal{F}(\rho))\ .\end{aligned} \tag{23}$$

Denote $\frac{d\rho(t)}{dt}|_{t=0} = \sigma_2$. Then

$$\sigma_2\big(g_W(\sigma_1,\nabla_W\mathcal{F}(\rho))\big) = \frac{d}{dt}|_{t=0}\big(\sigma_1^{\mathsf{T}} d_\rho\mathcal{F}(\rho)\big) = \sigma_1^{\mathsf{T}} d^2_\rho\mathcal{F}(\rho)\sigma_2\ .$$

Substituting the above formula and Proposition 3 into (23), we finish the proof. □

Remark 1. As in a Riemannian manifold $(\mathcal{P}_+(G), g_W)$, given the orthonormal basis $X_i = \sqrt{\lambda_i(\rho)}u_i(\rho)$, $i = 1, \cdots, n-1$, it is clear that

$$\Delta_W\mathcal{F}(\rho) = \sum_{i=1}^{n-1}\mathrm{Hess}_W\mathcal{F}(\rho)(X_i, X_i)\ .$$

4. Riemannian structures on density manifold

In this section, the approach in previous sections guides us to derive all geometry formulas in infinite dimensional density space.

Suppose (M, g_M) is a smooth, compact, connected, d-dimensional Riemannian manifold without boundary. g_M is its metric tensor and d_M is its Riemannian metric. A volume element is a positive d-form represented by dx. The total volume of manifold is denoted by $\mathrm{vol}(M)$. The gradient, divergence operator in M is denoted by ∇, $\nabla\cdot$ respectively. $\Delta = \nabla\cdot(\nabla)$ is the Laplace-Beltrami operator in M.

4.1. **Review of density manifold.** We briefly review the L^2-Wassertein metric for the space of densities supported on M. Denote $\mathcal{P}_2(M)$ as the set of probability density functions with finite second moment. Given $\rho^0(x)$ and $\rho^1(x) \in \mathcal{P}_2(M)$, the L^2-Wasserstein metric between ρ^0 and ρ^1 is given by

$$(W(\rho^0,\rho^1))^2 = \inf_\pi \int_M\int_M d_M(x,y)^2\pi(x,y)dxdy\ ,$$

where π is among all joint measures supported on $M\times M$ with martingales ρ^0 and ρ^1. Equivalently, the L^2-Wasserstein metric can be written as the dynamical variational problem. Denote the path $\rho_t = \rho(t,x)$ connecting $\rho^0(x)$ and $\rho^1(x)\in\mathcal{P}_+(M)$, then

$$(W(\rho^0,\rho^1))^2 = \inf_{v_t,\rho_t}\ \Big\{\int_0^1\int_M v_t^2\rho_t dxdt\colon \frac{\partial\rho_t}{\partial t} + \nabla\cdot(\rho_t v_t) = 0\ ,\ \rho_0 = \rho^0\ ,\ \rho_1 = \rho^1\Big\}\ , \tag{24}$$

where the infimum is taken among all Borel vector field functions $v_t = v(t,x)\in T_xM$ and density path ρ_t. The equivalence between (24) and the above linear programming problem can be shown by the duality argument and the Hopf-Lax formula in (M, d_M). For many details see [31].

We now focus on (24). Notice that $\mathcal{P}_2(M)$ is an infinite dimensional manifold with boundary. Here the boundary refers the set in which the density function is zero at certain point. For better illustration, consider the space of positive smooth density functions supported on M,

$$\mathcal{P}_+(M) = \{\rho(x) \in C^\infty(M) \colon \rho(x) > 0 \ , \ \int_M \rho(x)dx = 1\} \subset \mathcal{P}_2(M) \ .$$

In literature [17], $\mathcal{P}_+(M)$, with metric tensor introduced by (24), is called density manifold. The following linear operator is needed for describing the geometry of density manifold.

Definition 6. *Given $a(x) \in C^\infty(M)$, define the weighted Laplacian operator $\Delta_a \colon C^\infty(M) \to C^\infty(M)$,*

$$\Delta_a \Phi(x) = \nabla \cdot (a(x) \nabla \Phi(x)) \ , \quad \Phi(x) \in C^\infty(M) \ .$$

Let $a(x) = \rho(x) \in \mathcal{P}_+(M)$, we can apply the Hodge decomposition on M by the elliptic operator Δ_ρ. For any smooth vector field $v(x) \in T_xM$, there exists a potential function $\Phi(x) \in C^\infty(M)$ module constant shrift, and a divergence free vector field $\Psi(x) \in T_xM$, such that

$$v(x) = \nabla\Phi(x) + \Psi(x) \ , \quad \nabla \cdot (\rho(x)\Psi(x)) = 0 \ .$$

In other words,

$$\int_M v(x)^2 \rho(x) dx = \int_M [(\nabla\Phi(x))^2 + \Psi(x)^2]\rho(x)dx \geq \int_M (\nabla\Phi(x))^2 \rho(x) dx \ .$$

Thus the metric W defined in (24) is equivalent to

$$\begin{aligned} \big(W(\rho^0, \rho^1)\big)^2 &= \inf_{\Phi_t \colon \rho_0 = \rho^0, \ \rho_1 = \rho^1} \{ \int_0^1 \int_M (\nabla\Phi_t)^2 \rho_t dx dt \ : \ \partial_t \rho_t + \nabla \cdot (\rho_t \nabla \Phi_t) = 0 \} \\ &= \inf_{\Phi_t \colon \rho_0 = \rho^0, \ \rho_1 = \rho^1} \{ \int_0^1 \int_M \Phi_t (-\Delta_{\rho_t} \Phi_t) dx dt \ : \ \partial_t \rho_t = -\Delta_{\rho_t} \Phi_t \} \ . \end{aligned}$$

Definition 7. *Denote the tangent space at $\rho \in \mathcal{P}_+(M)$ by*

$$T_\rho \mathcal{P}_+(M) = \{\sigma(x) \in C^\infty(M) \colon \int_M \sigma(x) dx = 0\} \ .$$

Given $\sigma_1, \sigma_2 \in T_\rho\mathcal{P}_+(M)$, the inner product $g_W(\cdot,\cdot) : T_\rho\mathcal{P}_+(M) \times T_\rho\mathcal{P}_+(M) \to \mathbb{R}$ is defined by

$$g_W(\sigma_1, \sigma_2) := \int_M \sigma_1(x)(-\Delta_\rho)^\dagger \sigma_2(x) dx \ ,$$

where $(-\Delta_\rho)^\dagger \colon T_\rho\mathcal{P}_+(M) \to T_\rho\mathcal{P}_+(M)$ is the pseudo inverse operator of $-\nabla \cdot (\rho(x)\nabla)$.

Denote $\Phi_i(x) \in C^\infty(M)$ modulo additive constants, such that $(-\Delta_\rho \Phi_i)(x) = \sigma_i(x)$, $i = 1, 2$. Then

$$\begin{aligned} g_W(\sigma_1, \sigma_2) &= \int_M \Phi_1(x)(-\Delta_\rho)(-\Delta_\rho)^\dagger(-\Delta_\rho)\Phi_2(x) dx \\ &= \int_M \Phi_1(x)(-\nabla \cdot (\rho(x)\nabla\Phi_2(x))) dx \\ &= \int_M \nabla\Phi_1(x)\nabla\Phi_2(x)\rho(x) dx \ . \end{aligned}$$

Denote the path $\rho_t = \rho(t,x) \in \mathcal{P}_+(M)$ connecting $\rho^0(x)$ and $\rho^1(x)$, and let $\partial_t\rho_t = -\Delta_\rho\Phi_t = -\nabla\cdot(\rho_t\nabla\Phi_t)$. Then the Wasserstein metric in (24) can be represented by

$$(W(\rho^0,\rho^1))^2 = \inf_{\rho_t\in\mathcal{P}_+(M)} \Big\{ \int_0^1 g_W(\partial_t\rho_t,\partial_t\rho_t)dt\colon \rho_0=\rho^0 \ , \rho_1=\rho^1\Big\} \ .$$

By the arc-length time reparameterization [19], one can simply denote the L^2-Wasserstein metric by

$$W(\rho^0,\rho^1) = \inf_{\rho_t\in\mathcal{P}_+(M)} \Big\{ \int_0^1 \sqrt{g_W(\partial_t\rho_t,\partial_t\rho_t)}dt\colon \rho_0=\rho^0 \ , \rho_1=\rho^1\Big\} \ .$$

4.2. **Riemannian structures.** Motivated by section 3, we first construct a Riemannian metric in the positive measure space, and embed the density manifold as its submanifold. The geometry structures of density manifold is then introduced by the L^2 coordinates in the positive measure space.

Consider

$$\mathcal{M}_+(M) = \{\mu(x)\in C^\infty(M)\colon \mu(x)>0\} \ .$$

Thus $\mathcal{P}_+(M)\subset\mathcal{M}_+(M)$. Denote the tangent space at $\mu\in\mathcal{M}_+(M)$,

$$T_\mu\mathcal{M}_+(M) = \{A(x)\in C^\infty(M)\} \ .$$

We define a Riemannian inner product on the infinite dimensional manifold $\mathcal{M}_+(M)$.

Given $\mu\in\mathcal{M}_+(M)$ and $A(x)\in C^\infty(M)$, denote a positive definite operator $g(\mu)\colon C^\infty(M)\to C^\infty(M)$ by

$$(g(\mu)A)(x) = \big(-\Delta_\mu\big)^\dagger(A(x)-\int_M A(y)dy) + \int_M A(y)dy \ . \tag{25}$$

Definition 8 (Inner product on $\mathcal{M}_+(M)$)**.** *Given $\mu\in\mathcal{M}_+(M)$, define the inner product $g_{\mathcal{M}}\colon T\mathcal{M}_+(M)\times T\mathcal{M}_+(M)\to\mathbb{R}$ by*

$$\begin{aligned} g_{\mathcal{M}}(A_1(x),A_2(x)) &= \int_M A_1(x)(g(\mu)A_2)(x)dx \\ &= \int_M (A_1(x)-\int_M A_1(y)dy)(-\Delta_\mu)^\dagger(A_2(x)-\int_M A_2(y)dy)dx \\ &\quad + \int_M A_1(y)dy\int_M A_2(y)dy \ , \end{aligned}$$

for any $A_1(x)$, $A_2(x)\in T_\mu\mathcal{M}_+(M)$.

We next show that $(\mathcal{M}_+(M), g_{\mathcal{M}})$ induces a metric in its submanifold $(\mathcal{P}_+(M), g_W)$. Denote $\imath\colon \mathcal{P}_+(M)\to\mathcal{M}_+(M)$ a natural inclusion by $\imath(\rho)=\rho$, then $\imath$ induces a Riemannian metric tensor g_W on $\mathcal{P}_+(M)$ via pullback. In other words, for any $\sigma_1(x),\sigma_2(x)\in T_\rho\mathcal{P}_+(M)$,

$$\begin{aligned} g_{\mathcal{M}}(\sigma_1,\sigma_2) &= \int_M (\sigma_1(x)-\int_M \sigma_1(y)dy)(-\Delta_\mu)^\dagger(\sigma_2(x)-\int_M \sigma_2(y)dy)dx \\ &\quad + \int_M \sigma_1(y)dy\int_M \sigma_2(y)dy \\ &= \int_M \sigma_1(x)(-\Delta_\rho)^\dagger\sigma_2(x)dx = g_W(\sigma_1,\sigma_2) \ . \end{aligned}$$

We are ready to show the Riemannian structure of $(\mathcal{P}_+(M), g_W)$ by the one in $(\mathcal{M}_+(M), g_{\mathcal{M}})$.

Proposition 9 (Gradient)**.** *Consider $\mathcal{F} \in C^\infty(\mathcal{M}_+(M))$, denote its gradient operators in $(\mathcal{M}_+(M), g_{\mathcal{M}})$ and $(\mathcal{P}_+(M), g_W)$ by $\nabla_g\mathcal{F}(\rho) \in T_\rho\mathcal{M}_+(M)$, $\nabla_W\mathcal{F}(\rho) \in T_\rho\mathcal{P}_+(M)$, respectively. Then*

$$\nabla_W\mathcal{F}(\rho) = \nabla_g\mathcal{F}(\rho) - g_{\mathcal{M}}(\nabla_g\mathcal{F}(\rho), u_0)u_0 = -\nabla\cdot\left(\rho(x)\nabla\frac{\delta}{\delta\rho(x)}\mathcal{F}(\rho)\right) ,$$

where $\frac{\delta}{\delta\rho(x)}$ represents the L^2 first variation and $u_0(x) \equiv \frac{1}{\sqrt{vol(M)}}$.

Proof. Notice that

$$\nabla_g\mathcal{F}(\rho)(x) = (g(\rho)^{-1}\delta_\rho\mathcal{F}(\rho))(x)$$

We show that $g(\rho)^{-1}\colon C^\infty(M) \to C^\infty(M)$ is defined by

$$(g(\rho)^{-1}A)(x) = -\Delta_\rho A(x) + \int_M A(y)dy\ , \quad \text{for any } A(x) \in C^\infty(M)\ .$$

Since

$$\begin{aligned} g(\rho)^{-1}(g(\rho)A)(x) =& (-\Delta_\rho)[(-\Delta_\rho)^\dagger\big(A(x) - \int_M A(y)dy\big) + \int_M A(y)dy] + \int_M A(y)dy \\ =& A(x) - \int_M A(y)dy + \int_M A(y)dy = A(x)\ . \end{aligned}$$

Thus

$$\nabla_g\mathcal{F}(\rho) = g(\rho)^{-1}\delta_\rho\mathcal{F}(\rho) = -\Delta_\rho\delta_\rho\mathcal{F}(\rho)(x) + \int_M \frac{\delta}{\delta\rho(y)}\mathcal{F}(\rho)dy\ .$$

Denote $u_0(x) = \frac{1}{\sqrt{\mathrm{vol}(M)}}$. For any $\sigma \in T_\rho\mathcal{P}_+(M)$, then

$$\int_M \sigma(x)(g(\rho)u_0)(x)dx = \int_M \sigma(x)\big[(-\Delta_\rho^\dagger)(u_0(x) - \int_M u_0(y)dy) + \int_M u_0(y)dy\big]dx = 0\ .$$

Notice

$$g(\nabla_g\mathcal{F}(\rho), u_0) = \int_M (-\Delta_\rho\delta_\rho\mathcal{F}(\rho)(x) + \int_M \frac{\delta}{\delta\rho(y)}\mathcal{F}(\rho)dy)\frac{1}{\sqrt{\mathrm{vol}(M)}}dx = \sqrt{\mathrm{vol}(M)}\int_M \frac{\delta}{\delta\rho(y)}\mathcal{F}(\rho)dy\ ,$$

then the orthogonal projection of $\nabla_g\mathcal{F}(\rho)$ is

$$\begin{aligned} &\nabla_g\mathcal{F}(\rho) - g(\nabla_g\mathcal{F}(\rho), u_0)u_0 \\ =& -\Delta_\rho\delta_\rho\mathcal{F}(\rho)(x) + \int_M \frac{\delta}{\delta\rho(y)}\mathcal{F}(\rho)dy - \int_M \frac{\delta}{\delta\rho(y)}\mathcal{F}(\rho)dy\cdot\sqrt{\mathrm{vol}(M)}\cdot\frac{1}{\sqrt{\mathrm{vol}(M)}} \\ =& -\Delta_\rho\delta_\rho\mathcal{F}(\rho)(x) = -\nabla\cdot\left(\rho(x)\nabla\frac{\delta}{\delta\rho(x)}\mathcal{F}(\rho)\right) , \end{aligned}$$

which finishes the proof. □

We derive the commutator on density manifold, i.e. $[\cdot,\cdot]_W\colon T_\rho\mathcal{P}_+(M)\times T_\rho\mathcal{P}_+(M) \to T_\rho\mathcal{P}_+(M)$.

Proposition 10 (Commutator)**.** *Given $\sigma_1(x)$, $\sigma_2(x) \in T_\rho\mathcal{P}_+(M)$, then*

$$[\sigma_1, \sigma_2]_W = 0\ .$$

Proof. Consider $\mathcal{F} \in C^\infty(\mathcal{P}_+(M))$. Denote $\sigma_1(\mathcal{F}(\rho)) := \frac{d}{dt}|_{t=0}\mathcal{F}(\rho + t\sigma_1)$, then

$$\begin{aligned}[\sigma_1, \sigma_2]_W(\mathcal{F}(\rho)) =& \frac{d}{dt}|_{t=0}\frac{d}{ds}|_{s=0}\mathcal{F}(\rho + t\sigma_1 + s\sigma_2) - \frac{d}{ds}|_{s=0}\frac{d}{dt}|_{t=0}\mathcal{F}(\rho + t\sigma_1 + s\sigma_2)\\ =& \int_M\int_M \frac{\delta^2}{\delta\rho(x)\delta\rho(y)}\mathcal{F}(\rho)\sigma_1(x)\sigma_2(y) - \frac{\delta^2}{\delta\rho(x)\delta\rho(y)}\mathcal{F}(\rho)\sigma_1(y)\sigma_2(x)dxdy = 0 \ ,\end{aligned}$$

where $\frac{\delta^2}{\delta\rho(x)\delta\rho(y)}$ is the L^2 second variation of $\mathcal{F}(\rho)$. $\square$

We derive the Levi-Civita connection by the commutator on density manifold.

Proposition 11 (Levi-Civita connection)**.** *The operator* $\nabla^W_\cdot\cdot : T_\rho\mathcal{P}_+(M)) \times T_\rho\mathcal{P}_+(M) \to T_\rho\mathcal{P}_+(M)$ *is defined by*

$$\nabla^W_{\sigma_1(x)}\sigma_2(x) = -\frac{1}{2}\Big(\Delta_{\sigma_1}\Delta_\rho^\dagger\sigma_2 + \Delta_{\sigma_2}\Delta_\rho^\dagger\sigma_1 + \Delta_\rho(\nabla\Delta_\rho^\dagger\sigma_1\cdot\nabla\Delta_\rho^\dagger\sigma_2)\Big)(x) \ .$$

Proof. We apply the Koszul formula for $\sigma_i \in \mathcal{T}_\rho\mathcal{P}_+(M)$, $i = 1, 2, 3$. Notice $[\sigma_i, \sigma_j]_W = 0$, then

$$\begin{aligned}g_W(\nabla^W_{\sigma_1}\sigma_2, \sigma_3) =& \frac{1}{2}\Big\{\sigma_1(g_W(\sigma_2, \sigma_3)) + \sigma_2(g_W(\sigma_1, \sigma_3)) - \sigma_3(g_W(\sigma_1, \sigma_2))\Big\}\\ =&\frac{1}{2}\frac{d}{dt}|_{t=0}\int_M\Big\{\sigma_2(x)(-\Delta_{\rho+t\sigma_1})^\dagger\sigma_3(x) + \sigma_1(x)(-\Delta_{\rho+t\sigma_2})^\dagger\sigma_3(x) - \sigma_1(x)(-\Delta_{\rho+t\sigma_3})^\dagger\sigma_2(x)\Big\}dx \ .\end{aligned} \tag{26}$$

The next two claims are needed to further calculate (26).

Claim 3:

$$\frac{d}{dt}|_{t=0}\int_M\sigma_2(x)(-\Delta^\dagger_{\rho+t\sigma_1})\sigma_3(x)dx = \int_M\sigma_2(x)\Delta_\rho^\dagger\Delta_{\sigma_1}\Delta_\rho^\dagger\sigma_3(x)dx \ .$$

Proof of Claim 3. Since

$$0 = \frac{d}{dt}(g(\rho+t\sigma_1)g(\rho+t\sigma_1)^{-1}) = \frac{d}{dt}g(\rho+t\sigma_1)g(\rho+t\sigma_1)^{-1} + g(\rho+t\sigma_1)\frac{d}{dt}g(\rho+t\sigma_1)^{-1} \ ,$$

then

$$\begin{aligned}\frac{d}{dt}g(\rho+t\sigma_1) =& -g(\rho+t\sigma_1)\frac{d}{dt}g(\rho+t\sigma_1)^\dagger g(\rho+t\sigma_1)\\ =& -g(\rho+t\sigma_1)\frac{d}{dt}(-\Delta_{\rho+t\sigma_1})g(\rho+t\sigma_1)\\ =& g(\rho+t\sigma_1)\Delta_{\sigma_1}g(\rho+t\sigma_1) \ ,\end{aligned} \tag{27}$$

where the second equality is true since $\Delta_\rho = \nabla\cdot(\rho(x)\nabla)$ is linear w.r.t. ρ. From (6), we obtain

$$\int_M\sigma_2(x)(-\Delta_\rho^\dagger)\sigma_3(x)dx = \int_M\sigma_2(x)\big(g(\rho+t\sigma_1)\sigma_3\big)(x)dx \ . \tag{28}$$

From (34) and (28), we have

$$\begin{aligned}\frac{d}{dt}|_{t=0}\int_M \sigma_2(x)(-\Delta_{\rho+t\sigma_1})^\dagger\sigma_3(x)dx &= \int_M \sigma_2(x)\frac{d}{dt}|_{t=0}g(\rho+t\sigma_1)\sigma_3(x)dx\\ &= \int_M \sigma_2(x)g(\rho)(-\Delta_{\sigma_1})g(\rho)\sigma_3(x)dx\\ &= \int_M \sigma_2(x)\Delta_\rho^\dagger\Delta_{\sigma_1}\Delta_\rho^\dagger\sigma_3(x)dx\ .\end{aligned}$$

where the last equality follows from the fact that, for any $\sigma(x)\in T_\rho\mathcal{P}_+(M)$,

$$g(\rho)\sigma(x) = (-\Delta_\rho)^\dagger\sigma(x) + \int_M \sigma(x)dx = (-\Delta_\rho)^\dagger\sigma(x)\ .$$

□

Claim 4:

$$\int_M \sigma_1(x)\Delta_\rho^\dagger\Delta_{\sigma_3}\Delta_\rho^\dagger\sigma_2(x)dx = -\int_M \sigma_3(x)\Delta_\rho^\dagger\Delta_\rho\Big(\nabla\Delta_\rho^\dagger\sigma_1(x)\cdot\nabla\Delta_\rho^\dagger\sigma_2(x)\Big)dx\ .$$

Proof of Claim 4.

$$\begin{aligned}\int_M \sigma_1(x)\Delta_\rho^\dagger\Delta_{\sigma_3}\Delta_\rho^\dagger\sigma_2(x)dx &= \int_M \Delta_\rho^\dagger\sigma_1(x)\nabla\cdot(\sigma_3(x)\nabla\Delta_\rho^\dagger\sigma_2(x))dx\\ &= -\int_M \sigma_3(x)\big(\nabla\Delta_\rho^\dagger\sigma_1(x)\cdot\nabla\Delta_\rho^\dagger\sigma_2(x)\big)dx\\ &= -\int_M \sigma_3(x)\Delta_\rho^\dagger\Delta_\rho\Big(\nabla\Delta_\rho^\dagger\sigma_1(x)\cdot\nabla\Delta_\rho^\dagger\sigma_2(x)\Big)dx\ ,\end{aligned}$$

where the second equality is given by integration by parts and the third one is introduced by $\Delta^\dagger\Delta\sigma = \sigma$ if $\sigma\in T_\rho\mathcal{P}_+(M)$. □

Applying Claim 3 and 4, we have

$$\begin{aligned}&g_W(\nabla^W_{\sigma_1}\sigma_2,\sigma_3)\\ =&\frac{1}{2}\int_M \sigma_3(x)\Delta_\rho^\dagger\Delta_{\sigma_1}\Delta_\rho^\dagger\sigma_2(x) + \sigma_3(x)\Delta_\rho^\dagger\Delta_{\sigma_2}\Delta_\rho^\dagger\sigma_1(x) - \sigma_1(x)\Delta_\rho^\dagger\Delta_{\sigma_3}\Delta_\rho^\dagger\sigma_2(x)dx\\ =&\int_M \sigma_3(x)(-\Delta_\rho)^\dagger(-\frac{1}{2})\big\{\Delta_{\sigma_1}\Delta_\rho^\dagger\sigma_2(x) + \sigma_3(x)\Delta_\rho^\dagger\Delta_{\sigma_2}\Delta_\rho^\dagger\sigma_1(x) - \sigma_1(x)\Delta_\rho^\dagger\Delta_{\sigma_3}\Delta_\rho^\dagger\sigma_2(x)\big\}dx\ .\end{aligned}$$

From the definition of inner product g_W, we finish the proof. □

By the Levi-Civita connection, we define the Christopher symbol in density manifold.

Definition 9. *Denote the Christopher symbol operator at $\rho\in\mathcal{P}_+(M)$ at $x\in M$ by $\Gamma^{W,x}\colon T_\rho\mathcal{P}_+(M)\times T_\rho\mathcal{P}_+(M)\to\mathbb{R}$,*

$$\Gamma^{W,x}(\sigma_1,\sigma_2) = -\frac{1}{2}\Big(\Delta_{\sigma_1}\Delta_\rho^\dagger\sigma_2 + \Delta_{\sigma_2}\Delta_\rho^\dagger\sigma_1 + \Delta_\rho(\nabla\Delta_\rho^\dagger\sigma_1\cdot\nabla\Delta_\rho^\dagger\sigma_2)\Big)(x)\ .$$

It is clear $\nabla^W_{\sigma_1}\sigma_2(x) = \nabla^W_{\sigma_2}\sigma_1(x) = \Gamma^{W,x}(\sigma_1,\sigma_2)$. By this Christopher symbol, the parallel transport and geodesic equation in density manifold are derived.

Proposition 12 (Parallel transport)**.** *Denote $\rho\colon (a,b)\to\mathcal{P}_+(M)$. Consider $\sigma_t=\sigma(t,x)\in T_\rho\mathcal{P}_+(G)$ be a vector field along curve $\rho_t=\rho(t,x)$, then the equation for σ_t to be parallel along ρ_t is*

$$\partial_t\sigma_t=\frac{1}{2}\big(\Delta_{\sigma_t}\Delta_{\rho_t}^{\dagger}\partial_t\rho_t+\Delta_{\partial_t\rho_t}\Delta_{\rho_t}^{\dagger}\sigma_t+\Delta_{\rho_t}(\nabla\Delta_{\rho}^{\dagger}\partial_t\rho_t\cdot\nabla\Delta_{\rho}^{\dagger}\sigma_t)\big)\ .$$

Let $\sigma_t=\partial_t\rho_t$, then the geodesic equation satisfies

$$\partial_{tt}\rho_t=\Delta_{\partial_t\rho_t}\Delta_{\rho_t}^{\dagger}\partial_t\rho_t+\frac{1}{2}\Delta_{\rho_t}(\nabla\Delta_{\rho_t}^{\dagger}\partial_t\rho_t)^2\ . \tag{29}$$

Proof. The parallel transport equation is derived by

$$\partial_t\sigma(t,x)=-\Gamma^{W,x}(\sigma_t,\partial_t\rho_t)=\frac{1}{2}\big(\Delta_{\sigma_t}\Delta_{\rho_t}^{\dagger}\partial_t\rho_t+\Delta_{\partial_t\rho_t}\Delta_{\rho_t}^{\dagger}\sigma_t+\Delta_{\rho_t}(\nabla\Delta_{\rho}^{\dagger}\partial_t\rho_t\cdot\nabla\Delta_{\rho}^{\dagger}\sigma_t)\big)\ .$$

And the geodesic equation is introduced by

$$\partial_{tt}\rho_t=-\Gamma^{W,x}(\partial_t\rho_t,\partial_t\rho_t)=\Delta_{\partial_t\rho_t}\Delta_{\rho_t}^{\dagger}\partial_t\rho_t+\frac{1}{2}\Delta_{\rho_t}(\nabla\Delta_{\rho_t}^{\dagger}\partial_t\rho_t)^2\ .$$

□

We introduce the curvature formulas in $(\mathcal{P}_+(M),g_W)$ by the similar derivation in (11). Denote $R_W(\cdot,\cdot)\cdot\colon T_\rho\mathcal{P}_+(M)\times T_\rho\mathcal{P}_+(M)\times T_\rho\mathcal{P}_+(M)\to T_\rho\mathcal{P}_+(M)$.

Proposition 13 (Curvature tensor)**.** *Given σ_1, σ_2, σ_3, $\sigma_4\in T_\rho\mathcal{P}_+(M)$, then*

$$\begin{aligned}&g_W(R_W(\sigma_1,\sigma_2)\sigma_3,\sigma_4)\\ =&\int_M\frac{1}{2}\Big\{\sigma_2(x)\Delta_\rho^{\dagger}\Delta_{m(\sigma_1,\sigma_4)}\Delta_\rho^{\dagger}\sigma_3(x)+\sigma_1(x)\Delta_\rho^{\dagger}\Delta_{m(\sigma_2,\sigma_4)}\Delta_\rho^{\dagger}\sigma_3(x)\Big\}\\ &\quad-\sigma_2(x)\Delta_\rho^{\dagger}\Delta_{m(\sigma_1,\sigma_3)}\Delta_\rho^{\dagger}\sigma_4(x)-\sigma_1(x)\Delta_\rho^{\dagger}\Delta_{m(\sigma_2,\sigma_3)}\Delta_\rho^{\dagger}\sigma_4(x)\Big\}\\ &-\frac{1}{4}\Big\{2n(\sigma_1,\sigma_2)\Delta_\rho^{\dagger}n(\sigma_3,\sigma_4)-n(\sigma_1,\sigma_3)\Delta_\rho^{\dagger}n(\sigma_2,\sigma_4)+n(\sigma_2,\sigma_3)\Delta_\rho^{\dagger}n(\sigma_1,\sigma_4)\Big\}dx\ ,\end{aligned} \tag{30}$$

where operators m, $n\colon T_\rho\mathcal{P}_+(M)\times T_\rho\mathcal{P}_+(M)\to T_\rho\mathcal{P}_+(M)$ are defined by

$$m(\sigma_a,\sigma_b):=-\frac{1}{2}\Big(\Delta_{\sigma_a}\Delta_\rho^{\dagger}\sigma_b+\Delta_{\sigma_b}\Delta_\rho^{\dagger}\sigma_a+\Delta_\rho(\nabla\Delta_\rho^{\dagger}\sigma_a\cdot\nabla\Delta_\rho^{\dagger}\sigma_b)\Big)\ ,$$

and

$$n(\sigma_a,\sigma_b):=\Delta_{\sigma_a}\Delta_\rho^{\dagger}\sigma_b-\Delta_{\sigma_b}\Delta_\rho^{\dagger}\sigma_a\ .$$

We compute the Hessian operator in density manifold by (23). Denote $\mathrm{Hess}_W(\cdot,\cdot)\colon T_\rho\mathcal{P}_+(M)\times T_\rho\mathcal{P}_+(M)\to T_\rho\mathcal{P}_+(M)$.

Proposition 14 (Hessian operator)**.** *Given $\sigma_1,\sigma_2\in T_\rho\mathcal{P}_+(M)$, then*

$$\begin{aligned}Hess_W\mathcal{F}(\rho)(\sigma_1,\sigma_2)=&\int_M\int_M\frac{\delta^2}{\delta\rho(x)\delta\rho(y)}\mathcal{F}(\rho)\sigma_1(x)\sigma_2(y)dxdy\\ &+\frac{1}{2}\int_M\frac{\delta}{\delta\rho(x)}\mathcal{F}(\rho)\Big(\Delta_{\sigma_1}\Delta_\rho^{\dagger}\sigma_2+\Delta_{\sigma_2}\Delta_\rho^{\dagger}\sigma_1+\Delta_\rho(\nabla\Delta_\rho^{\dagger}\sigma_1\cdot\nabla\Delta_\rho^{\dagger}\sigma_2)\Big)(x)dx\ .\end{aligned} \tag{31}$$

We next provide the formulation of the Laplacian-Beltrami operator in density manifold, i.e. $\Delta_W\colon C^\infty(\mathcal{P}_+(M)) \to C^\infty(\mathcal{P}_+(M))$, by the technique used in proposition 7. The following definitions are needed. Denote $\lambda_i(\rho)>0$, $i=1,2,\cdots$, be positive eigenvalues of $-\Delta_\rho$. In other words, there exists $u_i(x)\in C^\infty(M)$, such that

$$-\nabla\cdot(\rho(x)\nabla u_i(x)) = \lambda_i(\rho)u_i(x)\ .$$

Let $\lambda_{\alpha,\mathcal{F}}(\rho)$, $\alpha\in I$ with the index set I, be eigenvalues of operator $-\Delta_\rho\delta^2\mathcal{F}(\rho)$. In other words, there exists $v_\alpha(x)\in C^\infty(M)$, such that

$$-\nabla_x\cdot\big(\rho(x)\nabla_x(\int_M \frac{\delta^2}{\delta\rho(x)\delta\rho(y)}\mathcal{F}(\rho)v_\alpha(y)dy)\big) = \lambda_{\alpha,\mathcal{F}}v_\alpha(x)\ .$$

Proposition 15 (Laplacian-Beltrami operator)**.** *Given $\mathcal{F}(\rho)\in C^\infty(\mathcal{P}_+(M))$, then*

$$\Delta_W\mathcal{F}(\rho) := tr_{L^2}((-\Delta_\rho)\delta^2\mathcal{F}(\rho)) - \frac{1}{2}\int_M \nabla\log(det(-\Delta_\rho))(x)\nabla\frac{\delta}{\delta\rho(x)}\mathcal{F}(\rho)\rho(x)dx\ ,$$

where $tr_{L^2}((-\Delta_\rho)\delta^2\mathcal{F}(\rho)) = \sum_{\alpha\in I}\lambda_{\alpha,\mathcal{F}}(\rho)$, and $det(-\Delta_\rho)=\Pi_{i=1}^\infty\lambda_i(\rho)$.

Similar as the proof of Lemma 2, we introduce the Jacobi equation on density manifold.

Proposition 16 (Variations of energy and Jacobi equation)**.** *Consider an infinitesmall deformation $\rho_\epsilon(t,x)=\rho(t,x)+\epsilon h(t,x)\in\mathcal{P}_+(M)$ with $h(t,x)\in C^\infty(M)$, $\int_M h(t,x)dx=0$, and $h(0,x)=h(1,x)=0$,*

$$\mathcal{E}(\rho_\epsilon)=\int_0^1\frac{1}{2}\partial_t\rho_t^\epsilon(-\Delta_{\rho_t^\epsilon})^\dagger\partial_t\rho_t^\epsilon dt = \mathcal{E}(\rho)+\epsilon\delta\mathcal{E}(\rho)(h)+\frac{\epsilon^2}{2}\delta^2\mathcal{E}(\rho)(h)+o(\epsilon^2)\ .$$

Then the first and second variations satisfy

$$\delta\mathcal{E}(\rho)(h)=\int_0^1\int_M\partial_t\rho_t(-\Delta_{\rho_t})^\dagger\big(\partial_t h_t-\frac{1}{2}\Delta_{h_t}\Delta_{\rho_t}^\dagger\partial_t\rho_t\big)dxdt\ ,$$

and

$$\delta^2\mathcal{E}(\rho)(h)=\int_0^1\int_M\big(\partial_t h_t-\Delta_{h_t}\Delta_{\rho_t}^\dagger\partial_t\rho_t\big)(-\Delta_{\rho_t})^\dagger\big(\partial_t h_t-\Delta_{h_t}\Delta_{\rho_t}^\dagger\partial_t\rho_t\big)dxdt\ .$$

The Jacobi equation along the geodesic ρ_t satisfies

$$\partial_t h_t-\Delta_{h_t}\Delta_{\rho_t}^\dagger\partial_t\rho_t=0\ ,\quad h(0,x)=h(1,x)=0\ .$$

4.3. **Connections with Dual coordinates.** The dual coordinates in density manifold has been considered in [19] and Chapter 3 of [17]. In this sequel, we illustrate the connection between Otto calculus and the primal coordinate system introduced in this paper.

Denote the smooth cotangent space at $\rho\in\mathcal{P}_+(M)$ by

$$T_\rho^*\mathcal{P}_+(M)=\{F_\Phi,\ \Phi\in C^\infty(M)\colon F_\Phi(\sigma)=\int_M\sigma(x)\Phi(x)dx\ ,\ \text{for any }\sigma\in T_\rho\mathcal{P}_+(M)\}\ .$$

For any constant $c\in\mathbb{R}$ and any $\sigma\in T_\rho\mathcal{P}_+(M)$,

$$F_{(\Phi+c)}(\sigma)=\int_M\Phi(x)\sigma(x)dx+c\int_M\sigma(x)dx=\int_M\Phi(x)\sigma(x)dx=F_\Phi(\sigma)\ .$$

In other words, $C^\infty(M)/\mathbb{R}\cong T_\rho^*\mathcal{P}_+(M)$.

The cotangent space and tangent space of density manifold can be identified by the map $\Phi \to \mathbf{V}_\Phi = -\Delta_\rho \Phi(x)$. In other words, for any tangent vector $\sigma(x) \in T_\rho \mathcal{P}_+(M)$, there exists a unique $\Phi(x) \in C^\infty(M)/\mathbb{R}$, such that

$$\sigma(x) = \mathbf{V}_\Phi(x) = -\nabla \cdot (\rho(x) \nabla \Phi(x)) \ . \tag{32}$$

By the property of elliptical operator Δ_ρ, one can show $C^\infty(M)/\mathbb{R} \cong T_\rho \mathcal{P}_+(M)$. Thus $T_\rho^* \mathcal{P}_+(M) \cong T_\rho \mathcal{P}_+(M)$.

The Riemannian inner product in density manifold can be represented by the cotangent vectors. We apply the potential function $\Phi_1(x)$, $\Phi_2(x)$ to represent the tangent vector $\sigma_1(x), \sigma_2(x) \in T_\rho \mathcal{P}_+(M)$ using (32), i.e. $\mathbf{V}_{\Phi_1} = \sigma_1$, $\mathbf{V}_{\Phi_2} = \sigma_2$. Thus the inner product in density manifold is formulated by

$$\begin{aligned} g_W(\sigma_1, \sigma_2) =& \int_M \sigma_1(x) (-\Delta_\rho)^\dagger \sigma_2(x) dx \\ =& \int_M \Phi_1(x) (-\Delta_\rho) \Phi_2(x) dx \\ =& \int_M \nabla \Phi_1(x) \cdot \nabla \Phi_2(x) \rho(x) dx \\ =& g_W(\mathbf{V}_{\Phi_1}, \mathbf{V}_{\Phi_2}) \ . \end{aligned}$$

As is known in the Fréchet manifold, all geometric formulas in density manifold, including gradient, Hessian, geodesic are equivalent between primal and dual coordinates. We next illustrate this equivalence by the geodesic and Hessian operators.

Proposition 17 (Geodesic)**.** *Denote*

$$\Phi_t = \Phi(t, x) = -(\nabla \cdot \rho(t, x) \nabla)^\dagger \frac{\partial \rho}{\partial t}(t, x) \ . \tag{33}$$

The geodesic equation (29) *is equivalent to the following two equations. One is the compressible Euler equation*

$$\begin{cases} \partial_t \rho_t + \nabla \cdot (\rho_t \nabla \Phi_t) = 0 \\ \partial_t \Phi_t + \frac{1}{2} (\nabla \Phi_t)^2 = 0 \end{cases} ;$$

The other satisfies

$$\partial_{tt} \rho_t = \nabla \cdot \big(\rho_t (\frac{\nabla \cdot (\rho_t v_t)}{\rho_t} v_t + \nabla_{v_t} v_t) \big) \ . \tag{34}$$

where $v_t = v(t, x) = \nabla \Phi_t$.

Remark 2. Equation (34) is (4.12) of [17], which relates to the stochastic mechanics [25].

Proof. On the one hand, substituting (33) into (29), we have

$$\begin{aligned} 0 =& \partial_t(\partial_t \rho_t) - \Delta_{\partial_t \rho_t} \Delta_{\rho_t}^\dagger \partial_t \rho_t - \frac{1}{2} \Delta_{\rho_t} (\nabla \Delta_{\rho_t}^\dagger \partial_t \rho_t)^2 \\ =& -\partial_t (\Delta_{\rho_t} \Phi_t) + \Delta_{\partial_t \rho_t} \Phi_t - \frac{1}{2} \Delta_{\rho_t} (\nabla \Phi_t)^2 \\ =& -\Delta_{\rho_t} (\partial_t \Phi_t + \frac{1}{2} (\nabla \Phi_t)^2) \ . \end{aligned}$$

Recall $\partial_t\rho_t = -\Delta_{\rho_t}\Phi_t$. We derive the compressible Euler equation, where Φ_t is unique up to a shrift of constant function w.r.t. t.

On the other hand, denote $v_t = v(t,x) = \nabla\Phi(t,x)$ in (33), then $\nabla_{v_t}v_t = \frac{1}{2}\nabla(\nabla\Phi_t)^2$. Thus the geodesic equation (29) forms

$$\begin{aligned}\partial_t(\partial_t\rho_t) =&\Delta_{\partial_t\rho_t}\Delta^{\dagger}_{\rho_t}\partial_t\rho_t + \frac{1}{2}\Delta_{\rho_t}(\nabla\Delta^{\dagger}_{\rho_t}\partial_t\rho_t)^2\\ =&-\nabla\cdot(\partial_t\rho_t\nabla\Phi_t) + \frac{1}{2}\nabla\cdot(\rho_t\nabla(-\nabla\Phi_t)^2)\\ =&\nabla\cdot(\nabla\cdot(\rho_t\nabla\Phi_t)\nabla\Phi_t) + \nabla\cdot(\rho_t\nabla_{v_t}v_t)\\ =&\nabla\cdot\big(\rho_t(\frac{\nabla\cdot(\rho_t v_t)}{\rho_t}v_t + \nabla_{v_t}v_t)\big)\ .\end{aligned}$$

□

The second example is the Hessian operator in the density manifold.

Proposition 18 (Hessian operator by cotangent vectors)**.**

$$\begin{aligned}Hess_W\mathcal{F}(\rho)(\mathbf{V}_{\Phi_1},\mathbf{V}_{\Phi_2}) =& \int_M\int_M \nabla_x\nabla_y\frac{\delta^2}{\delta\rho(x)\delta\rho(y)}\mathcal{F}(\rho)\nabla\Phi_1(x)\nabla\Phi_2(y)\rho(x)\rho(y)dxdy\\ &+\int_M(\nabla\nabla\frac{\delta}{\delta\rho(x)}\mathcal{F}(\rho)\nabla\Phi_1(x),\nabla\Phi_2(x))\rho(x)dx\ ,\end{aligned}$$

where $\nabla\nabla$ denotes the Hessian operator on M.

Proof. Denote $\sigma_1(x) = \mathbf{V}_{\Phi_1}$, $\sigma_2(x) = \mathbf{V}_{\Phi_2}$. From (31), we have

$$\begin{aligned}&\mathrm{Hess}_W\mathcal{F}(\rho)\langle\mathbf{V}_{\Phi_1},\mathbf{V}_{\Phi_2}\rangle\\ =&\int_M\int_M\frac{\delta^2}{\delta\rho(x)\delta\rho(y)}\mathcal{F}(\rho)\nabla_x\cdot(\rho(x)\nabla_x\Phi_1(x))\nabla_y\cdot(\rho(y)\nabla_y\Phi_2(y))dxdy \qquad (H1)\\ &+\frac{1}{2}\int_M\frac{\delta}{\delta\rho(x)}\mathcal{F}(\rho)\big\{-\Delta_{\sigma_1}\Phi_2(x) - \Delta_{\sigma_2}\Phi_1(x) + \Delta_\rho(\nabla\Phi_1(x)\cdot\nabla\Phi_2(x))\big\}dx\ . \qquad (H2)\end{aligned}$$

We apply the next two steps to estimate (H1) and (H2). First, by integration by parts w.r.t. x and y twice, we derive

$$(H1) = \int_M\int_M\nabla_x\nabla_y\frac{\delta^2}{\delta\rho(x)\delta\rho(y)}\mathcal{F}(\rho)\nabla_x\Phi_1(x)\nabla_y\Phi_2(y)\rho(x)\rho(y)dxdy\ .$$

Second, denote $F(x) = \frac{\delta}{\delta\rho(x)}\mathcal{F}(\rho)$ in (T2), then

$$\begin{aligned}(H2) =& -\frac{1}{2}\int_M F(x)\nabla\cdot\big\{\sigma_1(x)\nabla\Phi_2(x) + \sigma_2(x)\nabla\Phi_1(x) - \rho(x)\nabla(\nabla\Phi_1(x)\cdot\nabla\Phi_2(x))\big\}dx\\ =& \ \frac{1}{2}\int_M\nabla F(x)\big\{\sigma_1(x)\nabla\Phi_2(x) + \sigma_2(x)\nabla\Phi_1(x) - \rho(x)\nabla(\nabla\Phi_1(x)\cdot\nabla\Phi_2(x))\big\}dx\\ =& \ \frac{1}{2}\int_M\nabla F(x)\big(\sigma_1(x)\nabla\Phi_2(x) + \sigma_2(x)\nabla\Phi_1(x)\big)dx \qquad (H21)\\ &-\frac{1}{2}\int_M\rho(x)\nabla F(x)\cdot\nabla(\nabla\Phi_1(x)\cdot\nabla\Phi_2(x))\big\}dx \qquad (H22)\end{aligned}$$

We next derive (H21). Substitute $\sigma_1(x) = -\nabla\cdot(\rho(x)\nabla\Phi_1(x))$ and $\sigma_2(x) = -\nabla\cdot(\rho(x)\nabla\Phi_2(x))$ into the above formula, then

$$\begin{aligned}(H21) =& -\frac{1}{2}\int_M \nabla F(x)\big(\nabla\cdot(\rho(x)\nabla\Phi_1(x))\nabla\Phi_2(x) + \nabla\cdot(\rho(x)\nabla\Phi_2(x))\nabla\Phi_1(x)\big)dx \\ =&\frac{1}{2}\int_M \Big(\nabla\Phi_1(x)\nabla(\nabla F(x)\cdot\nabla\Phi_2(x)) + \nabla\Phi_2(x)\nabla(\nabla F(x)\cdot\nabla\Phi_1(x))\Big)\rho(x)dx\ .\end{aligned}$$

We last prove the following claim.

Claim 5:

$$\begin{aligned}&2(\nabla\nabla F(x)\nabla\Phi_1(x), \nabla\Phi_2(x)) \\ =&\nabla\Phi_1(x)\nabla(\nabla F(x)\cdot\nabla\Phi_2(x)) + \nabla\Phi_2(x)\nabla(\nabla F(x)\cdot\nabla\Phi_1(x)) - \nabla F(x)\nabla(\nabla\Phi_1(x)\cdot\nabla\Phi_2(x))\ .\end{aligned} \tag{35}$$

Proof of Claim.

$$\begin{aligned}&\sum_{1\le a,b\le d} \nabla_a(\nabla_b F\cdot\nabla^b\Phi_1)\nabla^a\Phi_2 + \nabla_a(\nabla_b F\cdot\nabla^b\Phi_2)\nabla^a\Phi_1 - \nabla_a(\nabla_b\Phi_1\cdot\nabla^b\Phi_2)\nabla^a F \\ =&\sum_{1\le a,b\le d} \nabla_a\nabla_b F\nabla^a\Phi_1\nabla^b\Phi_2 + \nabla_b F\nabla_a\nabla^b\Phi_1\nabla^a\Phi_2 + \nabla_a\nabla_b F\nabla^a\Phi_1\nabla^b\Phi_2 \\ &\quad + \nabla_a\nabla^b\Phi_2\nabla^a\Phi_1 - \nabla^b(\nabla_a\Phi_1\cdot\nabla^a\Phi_2) \\ =&\sum_{1\le a,b\le d} 2\nabla_a\nabla_b F\nabla^a\Phi_1\nabla^b\Phi_2 + \nabla_b F\{\nabla_a\nabla^b\Phi_1\nabla^a\Phi_2 + \nabla_a\nabla^b\Phi_2\nabla^a\Phi_1 - \nabla^b(\nabla_a\Phi_1\cdot\nabla^a\Phi_2)\} \\ =&2\sum_{1\le a,b\le d} \nabla_a\nabla_b F\nabla^a\Phi_1\nabla^b\Phi_2\ .\end{aligned}$$

□

Using the fact $\mathrm{Hess}_W\mathcal{F}(\rho)(\mathbf{V}_{\Phi_1}, \mathbf{V}_{\Phi_2}) = (H1) + (H2) = (H1) + (H21) + (H22)$ and the claim, we finish the proof. □

The Hessian operator reminders us the connection between density manifold and Bakry-Émery [3]. In particular, (35) is exact formula (6) in [3].

4.4. **Density manifold and Bakry-Émery Γ_2 operator.** There is the other connection between the primal and dual coordinates. We next illustrate this by an identity between theBakry-Émery Γ_2 operator in M and geometry of density manifold. In literature [31], this connection is known in dual coordinates. In this sequel, we demonstrate it from the primal coordinates, especially the weighted Laplacian operator Δ_ρ.

Denote $h(x)\in C^\infty(M)$ and the operator $L_h\colon C^\infty(M)\to C^\infty(M)$ by

$$L_h\Phi = \Delta\Phi(x) - \nabla h(x)\cdot\nabla\Phi(x)\ .$$

Let $\Phi_1(x)$, $\Phi_2(x)\in C^\infty(M)$. The Γ operator : $C^\infty(M)\to C^\infty(M)$ is given by

$$\Gamma(\Phi_1, \Phi_2) = \frac{1}{2}[L_h(\Phi_1\Phi_2) - \Phi_1 L_h\Phi_2 - \Phi_2 L_h\Phi_1] = \nabla\Phi_1\cdot\nabla\Phi_2\ ,$$

and the Γ_2 operator : $C^\infty(M) \to C^\infty(M)$ is introduced by

$$\Gamma_2(\Phi_1, \Phi_2) = \frac{1}{2}[L_h\Gamma(\Phi_1, \Phi_2) - \Gamma(L_h\Phi_1, \Phi_2) - \Gamma(L_h\Phi_2, \Phi_1)] \ .$$

Denote the Gibbs measure $\rho^*(x) = \frac{1}{K}e^{-h(x)}$, where $K = \int_M e^{-h(x)}dx$. Consider the relative entropy

$$\mathcal{H}(\rho|\rho^*) = \int_M \rho(x)\log\frac{\rho(x)}{\rho^*(x)}dx = \int_M \rho(x)\log\rho(x)dx + \int_M \rho(x)h(x)dx + K \ .$$

We shall present that the Hessian formula of $\mathcal{H}(\rho|\rho^*)$ in $(\mathcal{P}_+(M), g_W)$ introduces the following identity in Riemannian manifold (M, g_M).

Proposition 19.

$$\begin{aligned} Hess_W\mathcal{H}(\rho|\rho^*)(\mathbf{V}_{\Phi_1}, \mathbf{V}_{\Phi_2}) =& \int_M \frac{\Delta_\rho\Phi_1\Delta_\rho\Phi_2}{\rho} + \big(\nabla\nabla\log\frac{\rho}{e^{-h}}\nabla\Phi_1, \nabla\Phi_2\big)\rho dx \\ =& \int_M \Gamma_2(\Phi_1, \Phi_2)\rho dx \ , \end{aligned}$$

In particular, let $\Phi(x) = \Phi_1(x) = \Phi_2(x)$ *and* $h(x) \equiv 0$, *then*

$$\begin{aligned} &\int_M \frac{1}{\rho(x)}\big(\nabla\cdot(\rho(x)\nabla\Phi(x))\big)^2 + (\nabla\nabla\log\rho(x)\nabla\Phi(x), \nabla\Phi(x))\rho(x)dx \\ =& \int_M \big[Ric_M(\nabla\Phi(x), \nabla\Phi(x)) + tr(\nabla\nabla\Phi(x)\nabla\nabla\Phi(x))\big]\rho(x)dx \ , \end{aligned}$$

where Ric_M *denotes the Ricci tensor on* M.

Proof. The proof is straightforward by the Hessian operator of relative entropy in density manifold. Since

$$\mathrm{Hess}_W\mathcal{H}(\rho|\rho^*)(\mathbf{V}_{\Phi_1}, \mathbf{V}_{\Phi_2}) = \int_M \frac{\Delta_\rho\Phi_1\Delta_\rho\Phi_2}{\rho} + \big(\nabla\nabla\log\rho\nabla\Phi_1, \nabla\Phi_2\big)\rho + \big(\nabla\nabla h\nabla\Phi_1, \nabla\Phi_2\big)\rho dx \ . \tag{36}$$

Rewriting (35) with Γ operator and doing integration by parts, we have

$$\begin{aligned} &\int_M (\nabla\nabla\log\rho\nabla\Phi_1, \nabla\Phi_2)\rho dx \\ =&\frac{1}{2}\int_M \big[\Gamma(\Gamma(\log\rho, \Phi_1), \Phi_2) + \Gamma(\Gamma(\log\rho, \Phi_2), \Phi_1) - \Gamma(\Gamma(\Phi_1, \Phi_2), \log\rho)\big]\rho dx \qquad (37) \\ =& -\frac{1}{2}\int_M \Gamma(\log\rho, \Phi_1)\Delta_\rho\Phi_2 + \Gamma(\log\rho, \Phi_2)\Delta_\rho\Phi_1 + \Gamma(\log\rho, \Gamma(\Phi_1, \Phi_2))\rho dx \ . \end{aligned}$$

Notice

$$\nabla\log\rho = \frac{1}{\rho}\nabla\rho \ , \tag{38}$$

then

$$\begin{aligned} \frac{\Delta_\rho\Phi}{\rho} =& \frac{\nabla\cdot(\rho\nabla\Phi)}{\rho} = \frac{(\nabla\rho, \nabla\Phi) + \rho\Delta\Phi}{\rho} \\ =&(\nabla\log\rho, \nabla\Phi) + \Delta\Phi = \Gamma(\log\rho, \Phi) + \Delta\Phi \ . \end{aligned} \tag{39}$$

Substituting (35), (37), (38), (39) into (36), we have

$$\begin{aligned}
&\mathrm{Hess}_W\mathcal{H}(\rho|\rho^*)(\mathbf{V}_{\Phi_1},\mathbf{V}_{\Phi_2})\\
=&\frac{1}{2}\int_M (\frac{\Delta_\rho\Phi_1}{\rho}-\Gamma(\log\rho,\Phi_1))\Delta_\rho\Phi_2+(\frac{\Delta_\rho\Phi_1}{\rho}-\Gamma(\log\rho,\Phi_2))\Delta_\rho\Phi_1-\Gamma(\log\rho,\Gamma(\Phi_1,\Phi_2))\rho dx\\
&+\int_M \big(\nabla\nabla h\nabla\Phi_1,\nabla\Phi_2\big)\rho dx\\
=&\frac{1}{2}\int_M \Big\{\Delta\Phi_1\Delta_\rho\Phi_2+\Delta\Phi_2\Delta_\rho\Phi_1-\Gamma(\rho,\Gamma(\Phi_1,\Phi_2))\Big\}+\Big\{(\nabla\nabla h\nabla\Phi_1,\nabla\Phi_2)\Big\}\rho dx\\
=&\frac{1}{2}\int_M \Big\{\Delta\Gamma(\Phi_1,\Phi_2)-\Gamma(\Delta\Phi_1,\Phi_2)-\Gamma(\Delta\Phi_2,\Phi_1)\Big\}\rho\\
&\quad+\Big\{\Gamma(\Phi_1,\Gamma(h,\Phi_2))+\Gamma(\Phi_2,\Gamma(h,\Phi_1))-\Gamma(h,\Gamma(\Phi_1,\Phi_2))\Big\}\rho dx\\
=&\frac{1}{2}\int_M \Big\{(\Delta-\nabla h\cdot\nabla)\Gamma(\Phi_1,\Phi_2)-\Gamma((\Delta-\nabla h\cdot\nabla)\Phi_1,\Phi_2)-\Gamma((\Delta-\nabla h\cdot\nabla)\Phi_2,\Phi_1)\Big\}\rho dx\\
=&\int_M \Gamma_2(\Phi_1,\Phi_2)\rho dx\ ,
\end{aligned}$$

where the last equality is from the definition of Γ_2 operator. As in (4a) of [3], from the Bochner's formula, we have

$$\Gamma_2(\Phi_1,\Phi_2)=(\nabla\nabla h\nabla\Phi_1,\nabla\Phi_2)+\mathrm{Ric}_M(\nabla\Phi_1,\nabla\Phi_2)+tr(\nabla\nabla\Phi_1\nabla\nabla\Phi_2)\ ,$$

which finishes the proof. □

Remark 3. We note that formulas (38) and (39) are widely used in Nelson's stochastic mechanics [25]. In particular, (39) relates to the osmotic diffusion operator in [17, 25].

In particular, if $\rho(x)\equiv 1$ and $h(x)\equiv 0$, then Proposition 19 shows the standard Yano's formula [33], which means

$$\int_M \big(\nabla\cdot(\nabla\Phi(x))\big)^2dx=\int_M \mathrm{Ric}_M(\nabla\Phi(x),\nabla\Phi(x))+\mathrm{tr}(\nabla\nabla\Phi(x)\nabla\nabla\Phi(x))dx\ .$$

The derivation of Yano's formula by the Hessian operator in density manifold is reported in [11, 18]. It is one of the motivations for this paper.

5. Differential equations in probability manifold

In this section, we use several examples to illustrate the geometry formulas in probability manifold.

Example 1 (Ordinary differential equations in probability manifold). *Consider $\mathcal{F}(\rho)\in C^\infty(\mathcal{P}(G))$. On one hand, the gradient flow of $\mathcal{F}(\rho)$ satisfies*

$$\dot\rho=-grad_W\mathcal{F}(\rho)\ ,$$

i.e.

$$\dot\rho=-L(\rho)d_\rho\mathcal{F}(\rho)=div_G(\rho\nabla_G d_\rho\mathcal{F}(\rho))\ . \tag{40}$$

On the other hand, the Hamiltonian flow of $\mathcal{F}(\rho)$ satisfies

$$\nabla_{\dot\rho}\dot\rho = \ddot\rho + (\dot\rho^{\mathsf{T}}\Gamma^k\dot\rho)_{k=1}^n = -\nabla_W\mathcal{F}(\rho)\ ,$$

i.e.

$$\ddot\rho - L(\dot\rho)L(\rho)^{\dagger}\dot\rho + \frac{1}{2}L(\rho)(\nabla_G L(\rho)^{\dagger}\dot\rho\circ\nabla_G L(\rho)^{\dagger}\dot\rho) = -L(\rho)d_\rho\mathcal{F}(\rho)\ . \tag{41}$$

Equation (41) *can be rewrite as the first order ODE system. Consider the dual coordinate via Legendre transform, i.e.* $\dot\rho = L(\rho)\Phi$*, then* (41) *forms*

$$\dot\rho + div_G(\rho\nabla_G\Phi) = 0\ ,\quad \dot\Phi + \frac{1}{2}\nabla_G\Phi\circ\nabla_G\Phi = -d_\rho\mathcal{F}(\rho)\ .$$

In other words,

$$\dot\rho = \frac{\partial}{\partial\Phi}\mathcal{H}(\rho,\Phi)\ ,\quad \dot\Phi = -\frac{\partial}{\partial\rho}\mathcal{H}(\rho,\Phi)\ ,$$

where

$$\mathcal{H}(\rho,\Phi) := \frac{1}{2}(\nabla_G\Phi,\nabla_G\Phi)_\rho + \mathcal{F}(\rho)\ .$$

Several examples of (40), (41) *have been studied in* [10, 11]*, which are Fokker-Planck equations, Schrödinger equations on graphs respectively.*

In addition, we are curious about the drift diffusion process associated with the canonical volume form in $(\mathcal{P}_+(G), g_W)$.

Example 2 (Stochastic differential equations in probability manifold)**.** *Consider the Fokker-Planck equation in* $(\mathcal{P}_+(G), g_W)$ *with drift vector* $\nabla_W\mathcal{F}(\rho)$*,* $\mathcal{F}(\rho)\in C^\infty(\mathcal{P}(G))$*, diffusion constant* $\beta > 0$ *on a compact set* $\mathcal{B}\subset\mathcal{P}_+(G)$*. The zero-flux condition is proposed on* $\partial\mathcal{B}$*. Then the Fokker-Planck equation*

$$\frac{\partial\mathbb{P}(t,\rho)}{\partial t} = div_W(\mathbb{P}(t,\rho)\nabla_W\mathcal{F}(\rho)) + \beta\Delta_W\mathbb{P}(t,\rho)$$

satisfies

$$\frac{\partial\mathbb{P}(t,\rho)}{\partial t} = \Pi(\rho)^{\frac{1}{2}}\nabla_\rho\cdot\Big\{\Pi(\rho)^{-\frac{1}{2}}L(\rho)\big(\mathbb{P}(t,\rho)d_\rho\mathcal{F}(\rho) + \beta d_\rho\mathbb{P}(t,\rho)\big)\Big\}\ . \tag{42}$$

An associated drift-diffusion process of (42) *on* $(\mathcal{P}_+(G), g_W)$ *satisfies*

$$d\rho_t = -L(\rho_t)(d_{\rho_t}\mathcal{F}(\rho_t) + \frac{\beta}{2}d_{\rho_t}\log\Pi(\rho_t))dt + \sqrt{2\beta}L(\rho_t)^{\frac{1}{2}}dB_t\ . \tag{43}$$

The stationary solution of (42) *is a Gibbs measure in the space of probability densities supported on* $(\mathcal{P}_+(G), g_W)$*,*

$$\mathbb{P}^*(\rho) = \frac{1}{K}e^{-\frac{\mathcal{F}(\rho)}{\beta}},\quad \textit{where } K = \int_{\mathcal{B}}\Pi(\rho)^{-\frac{1}{2}}e^{-\frac{\mathcal{F}(\rho)}{\beta}}dvol\ . \tag{44}$$

Proof. Using the fact that

$$d_\rho\mathbb{P}(t,\rho) = \mathbb{P}(t,\rho)d_\rho\log\mathbb{P}(t,\rho)\ ,$$

then (42) forms

$$\frac{\partial\mathbb{P}(t,\rho)}{\partial t} = \Pi(\rho)^{\frac{1}{2}}\nabla_\rho\cdot(L(\rho)\mathbb{P}(t,\rho)\big(d_\rho\mathcal{F}(\rho) + \beta d_\rho\log\mathbb{P}(t,\rho)\big)\Pi(\rho)^{-\frac{1}{2}})\ . \tag{45}$$

Denote the density function in Euclidean volume form, $f(t,\rho)=\mathbb{P}(t,\rho)\Pi(\rho)^{-\frac{1}{2}}$, then (45) forms

$$\frac{\partial f(t,\rho)}{\partial t}=\nabla_\rho\cdot[f(t,\rho)L(\rho)d_\rho\mathcal{F}(\rho)]+\beta\nabla_\rho\cdot[f(t,\rho)L(\rho)d_\rho\log\frac{f(t,\rho)}{\Pi(\rho)^{-\frac{1}{2}}}]\ .$$

Thus the infinitesimal generator of (42) satisfies

$$A\hat{f}(\rho)=-\sum_{i=1}^n\frac{\partial\hat{f}(\rho)}{\partial\rho_i}\big(L(\rho)(d_\rho\mathcal{F}(\rho)+\frac{\beta}{2}d_\rho\log\Pi(\rho)\big)(i)+\beta\sum_{1\le i,j\le n}(L(\rho)^{\frac{1}{2}}L(\rho)^{\frac{1}{2}})_{ij}\frac{\partial^2\hat{f}(\rho)}{\partial\rho_i\partial\rho_j}\ ,$$

for any compactly-supported C^2 function $\hat{f}(\rho)$ with $\rho\in\mathcal{B}$. Then the stochastic differential equation (43) is derived. By solving $\frac{\partial}{\partial t}f(t,\rho)=0$, we have

$$f^*(\rho)=\frac{1}{K}\Pi(\rho)^{-\frac{1}{2}}e^{-\frac{\mathcal{F}(\rho)}{\beta}}\ ,\quad\text{where } K=\int_{\mathcal{B}}\Pi(\rho)^{-\frac{1}{2}}e^{-\frac{\mathcal{F}(\rho)}{\beta}}d\mathrm{vol}\ .$$

In other words, $\mathbb{P}^*(\rho)=f^*(\rho)\Pi(\rho)^{\frac{1}{2}}$ satisfies (44). □

Remark 4. Formula (43) suggests that a drift-diffusion process in $(\mathcal{P}_+(M),g_W)$, forms

$$d\rho(t,x)=\nabla\cdot\Big(\rho(t,x)\nabla\frac{\delta}{\delta\rho}(\mathcal{F}+\frac{\beta}{2}\log\det(-\Delta_\rho))(t,x)\Big)dt+\sqrt{2\beta}(-\Delta_\rho)^{\frac{1}{2}}dB(t,x)\ ,$$

where $B(t,x)$ is the standard space-time Brownian motion. See [32] for more details.

6. Discussion

In this paper, we introduce the geometry formulas in probability manifold on graphs with L^2-Wasserstein metric. The main idea is to propose a Riemannian metric in the positive measure space, and study the probability manifold as its submanifold. Similar derivations are introduced in the infinite dimensional settings. We hope that this study would help to design numerical schemes and analyze the related models in statistics and population games. We will continue to work in this direction.

Acknowledgement The author would thank Prof. Shui-Nee Chow, Prof. Stanley Osher, Prof. Haomin Zhou and Prof. Wilfrid Gangbo for many fruitful discussions.

References

[1] S. Amari. *Information Geometry and Its Applications.* Number volume 194 in Applied mathematical sciences. Springer, Japan, 2016.
[2] N. Ay, J. Jost, H. V. Lê, and L. J. Schwachhöfer. *Information Geometry.* Ergebnisse der Mathematik und ihrer Grenzgebiete A @series of modern surveys in mathematics$l3. Folge, volume 64. Springer, Cham, 2017.
[3] D. Bakry and M. Émery. Diffusions Hypercontractives. In J. Azéma and M. Yor, editors, *Séminaire de Probabilités XIX 1983/84*, volume 1123, pages 177–206. Springer Berlin Heidelberg, Berlin, Heidelberg, 1985.
[4] E. A. Carlen. Conservative Diffusions. *Communications in Mathematical Physics*, 94(3):293–315, 1984.
[5] E. A. Carlen and W. Gangbo. Constrained Steepest Descent in the 2-Wasserstein Metric. *Annals of Mathematics*, 157(3):807–846, 2003.
[6] J. A. Carrillo, R. J. McCann, and C. Villani. Contractions in the 2-Wasserstein Length Space and Thermalization of Granular Media. *Archive for Rational Mechanics and Analysis*, 179(2):217–263, 2006.

[7] S.-N. Chow, L. Dieci, W. Li, and H. Zhou. Entropy dissipation semi-discretization schemes for Fokker-Planck equations. *arXiv:1608.02628 [math]*, 2016.
[8] S.-N. Chow, W. Huang, Y. Li, and H. Zhou. Fokker–Planck Equations for a Free Energy Functional or Markov Process on a Graph. *Archive for Rational Mechanics and Analysis*, 203(3):969–1008, 2012.
[9] S.-N. Chow, W. Li, J. Lu, and H. Zhou. Population games and Discrete optimal transport. *arXiv:1704.00855 [math]*, 2017.
[10] S.-N. Chow, W. Li, and H. Zhou. A Discrete Schrodinger Equation via Optimal Transport on Graphs. *arXiv:1705.07583 [math]*, 2017.
[11] S.-N. Chow, W. Li, and H. Zhou. Entropy Dissipation of Fokker-Planck Equations on Graphs. *arXiv:1701.04841 [math]*, 2017.
[12] F. R. K. Chung. *Spectral Graph Theory*. Number no. 92 in Regional conference series in mathematics. Published for the Conference Board of the mathematical sciences by the American Mathematical Society, Providence, R.I, 1997.
[13] W. Gangbo, W. Li, and C. Mou. Geodesic of minimal length in the set of probability measures on graphs. *arXiv:1712.09266 [math]*, 2017.
[14] N. Gigli. *Second Order Analysis on (P2(M),W2)*. American Mathematical Soc.
[15] R. Jordan, D. Kinderlehrer, and F. Otto. The Variational Formulation of the Fokker–Planck Equation. *SIAM Journal on Mathematical Analysis*, 29(1):1–17, 1998.
[16] A. Kriegl and P. W. Michor. *The Convenient Setting of Global Analysis*. Number v. 53 in Mathematical surveys and monographs. American Mathematical Society, Providence, R.I, 1997.
[17] J. D. Lafferty. The Density Manifold and Configuration Space Quantization. *Transactions of the American Mathematical Society*, 305(2):699–741, 1988.
[18] W. Li. A Study of Stochastic Differential Equations and Fokker-Planck Equations with Applications. 2016.
[19] J. Lott. Some Geometric Calculations on Wasserstein Space. *Communications in Mathematical Physics*, 277(2):423–437, 2008.
[20] J. Lott and C. Villani. Ricci Curvature for Metric-Measure Spaces via Optimal Transport. *Annals of Mathematics*, 169(3):903–991, 2009.
[21] J. Maas. Gradient Flows of the Entropy for Finite Markov Chains. *Journal of Functional Analysis*, 261(8):2250–2292, 2011.
[22] R. J. McCann. Polar Factorization of Maps on Riemannian Manifolds. *Geometric & Functional Analysis GAFA*, 11(3):589–608, 2001.
[23] A. Mielke. A Gradient Structure for Reaction–diffusion Systems and for Energy-Drift-Diffusion Systems. *Nonlinearity*, 24(4):1329, 2011.
[24] J. Moser. On the Volume Elements on a Manifold. *Transactions of the American Mathematical Society*, 120(2):286, 1965.
[25] E. Nelson. Derivation of the Schrödinger Equation from Newtonian Mechanics. *Physical Review*, 150(4):1079–1085, 1966.
[26] E. Nelson. *Quantum Fluctuations*. Princeton series in physics. Princeton University Press, Princeton, N.J, 1985.
[27] F. Otto. The Geometry of Dissipative Evolution Equations: The Porous Medium Equation. *Communications in Partial Differential Equations*, 26(1-2):101–174, 2001.
[28] F. Otto and C. Villani. Generalization of an Inequality by Talagrand and Links with the Logarithmic Sobolev Inequality. *Journal of Functional Analysis*, 173(2):361–400, 2000.
[29] S. Shahshahani. A new mathematical framework for the study of linkage and selection. *Memoirs of the American Mathematical Society*, 17(211):0–0, 1979.
[30] K.-T. Sturm. On the Geometry of Metric Measure Spaces. *Acta Mathematica*, 196(1):65–131, 2006.
[31] C. Villani. *Optimal Transport: Old and New*. Number 338 in Grundlehren der mathematischen Wissenschaften. Springer, Berlin, 2009.
[32] M.-K. von Renesse and K.-T. Sturm. Entropic measure and Wasserstein diffusion. *The Annals of Probability*, 37(3):1114–1191, 2009.
[33] K. Yano. On Harmonic and Killing Vector Fields. *Annals of Mathematics*, 55(1):38–45, 1952.

E-mail address: wcli@math.ucla.edu.