The existence of geodesics in Wasserstein spaces over path groups and loop groups

Jinghai Shao *

Center for Applied Mathematics, Tianjin University, Tianjin 300072, China

February 28, 2018

In this work we prove the existence and uniqueness of the optimal transport map for L^p-Wasserstein distance with $p > 1$, and particularly present an explicit expression of the optimal transport map for the case $p = 2$. As an application, we show the existence of geodesics connecting probability measures satisfying suitable condition on path groups and loop groups.

Keywords: Monge-Kantorovich problem; Path groups; Loop groups; Heat kernel measure; Pinned Wiener measure

1 Introduction

In the seminal works of K.T. Sturm [27] and Lott-Villani [20], a new concept of curvature-dimension condition has been developed on the abstract metric space to replace the lower bound of Ricci curvature of Riemannian manifold via the convexity of the relative entropy on the Wasserstein space. This convexity is measured by the behavior of the relative entropy along geodesics connecting two probability measures in the Wasserstein space over this metric space. This concept is equivalent to the Ricci curvature lower bound for Riemannian manifold as shown in [29] and possesses the advantage of stability under Gromov-Hausdorff convergence. There are many extensions of this concept in various setting, for example, Finsler space [23], Alexandrov spaces [24, 31], infinitesimally Hilbertian metric measure spaces [12]. The starting point of this concept is that the studied Wasserstein space is a geodesic space, that is, for any two probability measures ν_0 and ν_1 satisfying

*Email: shaojh@bnu.edu.cn

some additional condition, there exists a geodesic under the L^2-Wasserstein metric. The validation of this basic property usually depends on the study of the Monge-Kantorovich problem in respective space. In this work, we shall study the optimal transport on path groups and loop groups and apply the obtained optimal transport maps to show the existence of geodesics in the Wasserstein spaces over path groups and loop groups.

The Monge-Kantorovich problem is to consider how to move the mass from one distribution to another as efficiently as possible. Here the efficiency is measured against a positive cost function $c(x, y)$. Precisely, given two probability measures μ and ν on a measurable space X, define its Wasserstein distance by

$$W_c(\mu, \nu) = \inf \Big\{ \int_{X\times X} c(x, y)\, \pi(\mathrm{d}x, \mathrm{d}y); \quad \pi \in \mathscr{C}(\mu, \nu) \Big\}, \tag{1.1}$$

where $c : X \times X \to [0, +\infty]$ is called the cost function and $\mathscr{C}(\mu, \nu)$ is the set of all probability measures on $X \times X$ with marginals μ and ν respectively. Then the Monge-Kantorovich problem is to find a measurable map $\mathscr{T}$ satisfying $\nu = (\mathscr{T})_*\mu$ such that the probability measure $\pi = (id \times \mathscr{T})_*\mu$ attains the infimum in (1.1). Here the notion $(\mathscr{T})_*\mu$ denote the push forward of measure μ by a measurable map $\mathscr{T}$, i.e. $(\mathscr{T})_*\mu = \mu \circ \mathscr{T}^{-1}$; id denotes the identity map. It is well known that the solving of this problem is very crucially dependent on the cost function. On Euclidean space $\mathbb{R}^d$ and Riemannian manifold, there are many works to solve this problem with respect to different cost functions such as [6] [22] [17] [19]. Refer to [3] for general survey on this respect and to [28] for detail discussions.

When the dimension of the space goes to infinity, Feyel and Üstünel in [16] proved the existence and uniqueness of the optimal transport map on the abstract Wiener space. In [15], together with Fang, we solved the Monge-Kantorovich problem on loop groups. There we use the "Riemannian distance", a kind of Cameron-Martin distance in some sense, to define the L^2-Wasserstein distance. The advantage of this distance is that there exists a sequence of suitable finite dimensional approximations, which makes it possible to use the results in finite dimensional Lie groups. However, the "Riemannian distance" is too large. It behaves like the Cameron-Martin distance in Wiener space in some sense, which equals to infinite almost everywhere with respect to the Wiener measure. This causes great difficulty in ensuring the finiteness of the Wasserstein distance between two probability measures on loop groups. Furthermore, there is no explicit expression of the optimal transport in this case.

In this work, we shall use another important distance, L^2-distance, to define the Wasserstein distance on path or loop groups. Since the L^2-distance is always bounded when the Lie group is compact, the induced Wasserstein distance between any two probability measures is finite. Therefore, the finiteness of Wasserstein distance is no longer a constraint of the existence of optimal transport map in this situation. As an application, there exists an invertible optimal transport map pushing the heat kernel measure forward to the pinned Wiener measure on loop group. These two probability measures play important role in the stochastic analysis of loop groups. Another advantage of using L^2-distance is that an explicit form of optimal transport map can be given, which helps us to show the existence of geodesic connecting two probability measures on path groups or loop groups.

The existence of optimal transport map has a lot of applications. For example, it is applied to construct the solution of Monge-Ampère equation (cf. for instance, [13]), and to establish Prékopa-Leindler inequalities in [8]. In [18], it helps to construct the gradient flow of relative entropy in the space of probability measures, which provides a new method to construct the solution of Fokker-Planck equations. This method has been systemically studied and was developed to deal with more general differential equations in [4].

When studying the Monge-Kantorovich problem on path and loop groups using the L^2-distance, we need to consider the derivative of Riemannian distance on Lie group, which adds some condition on Lie group about the cut locus of its identity element.

Let G be a connected compact Lie group with Lie algebra $\mathscr{G}$ which is endowed with an inner product $\langle\,,\,\rangle_{\mathscr{G}}$, and the associated Riemannian distance is denoted by $\rho(\cdot,\cdot)$. Given a point $x\in G$, point $y\in G$ is called a *cut point* of x if there exists a geodesic $\gamma:[0,\infty)\to G$ parameterized by arc length with $\gamma(0)=x$ such that $\gamma(t_0)=y$ for some $t_0>0$ and for any $t\le t_0$, $\rho(\gamma(0),\gamma(t))=t$ and for any $t>t_0$, $\rho(\gamma(0),\gamma(t))<t$. The union of all cut points of x is called the *cut locus* of x and denoted by $Cut(x)$. A map $V:[a,b]\to\mathscr{G}$ is called a *piecewise continuous curve* if there exists a finite subdivision $a=a_0<a_1<\ldots<a_k=b$ such that $V\big|_{[a_{i-1},a_i]}$ is continuous for $i=1,\ldots,k$.

The condition on the cut locus used in this work is:

(H) If the cut locus $Cut(e)$ of the identity element e of G is not empty, then for any continuous curve $\{x_t\}_{t\in[a,b]}\subset Cut(e)$, there exists a piecewise continuous curve $\{X_t\}_{t\in[a,b]}$ in $\mathscr{G}$ such that $\exp_e X_t=x_t$, $\forall\,t\in[a,b]$, where $\exp_e$ denotes the exponential map determined by the geodesic equations in the setting of Riemannian manifold.

Examples:

- the n-dimensional torus $T_n=S^1\times\cdots\times S^1$ is a connected compact Lie group and satisfies the hypothesis (H).
- The Heisenberg group $\mathbb{H}^n$ endowed with Carnot-Carathéodory distance satisfies the assumption (H) by [5, Theorem 3.4]. Indeed, the Heisenberg group $\mathbb{H}^n$ is a noncommutative stratified nilpotent Lie group. As a set it can be identified with its Lie algebra $\mathbb{R}^{2n+1}\simeq\mathbb{C}^n\times\mathbb{R}$ via exponential coordinates. Denote a point in $\mathbb{H}^n$ by $\boldsymbol{x}=(\xi,\eta,t)=[\zeta,t]$ where $\xi=(\xi_1,\ldots,\xi_n)$, $\eta=(\eta_1,\ldots,\eta_n)\in\mathbb{R}^n$, $t\in\mathbb{R}$ and $\zeta=(\zeta_1,\ldots,\zeta_n)\in\mathbb{C}^n$ with $\zeta_j=\xi_j+i\eta_j$. The group law is given by $[\zeta,t]\cdot[\zeta',t']:=[\zeta+\zeta',t+t'+2\sum_{j=1}^n\mathrm{Im}\zeta_j\bar{\zeta}_j']$. The set $L^*:=\{[0,s]\in\mathbb{H}^n;\ s\in\mathbb{R}\backslash\{0\}\}$ is the cut locus of identity element $[0,0]\in\mathbb{H}^n$. Set $\mathbb{S}=\{a+ib\in\mathbb{C}^n;\ |a+ib|=1\}$.

For any $a+ib \in \mathbb{S}$, $v \in \mathbb{R}$ and $r>0$, we say that a curve $\gamma : [0,r] \to \mathbb{H}^n$ is a curve with parameter $(a+ib, v, r)$ if $\gamma(s) = (\xi(s), \eta(s), t(s))$ where

$$\begin{aligned}
\xi_j(s) &= \frac{r}{v}\Big(b_j\Big(1-\cos\frac{vs}{r}\Big) + a_j \sin\frac{vs}{r}\Big),\\
\eta_j(s) &= \frac{r}{v}\Big(-a_j\Big(1-\cos\frac{vs}{r}\Big) + b_j \sin\frac{vs}{r}\Big),\\
t(s) &= \frac{2r^2}{v^2}\Big(\frac{vs}{r} - \sin\frac{vs}{r}\Big), \qquad j=1,\dots,n,
\end{aligned}$$

when $v \neq 0$ and

$$\gamma(s) = (a_1 s, \dots, a_n s, b_1 s, \dots, b_n s, 0)$$

when $v=0$. Each curve with parameter $(a+ib, 2\pi, \sqrt{\pi|t|})$ for some $a+ib \in \mathbb{S}$ is a sub-unit minimal geodesic from $[0,0]$ to $\boldsymbol{x} = [0,t] \in L^*$, from which one can easily verify $\mathbb{H}^n$ satisfies the assumption (H).

In the following, after introducing some necessary notations on path and loop groups, we present our main results of this paper.

Denote $\mathcal{P}(G)$ the path group, that is,

$$\mathcal{P}(G) = \big\{\ell : [0,1] \to G \text{ continuous; } \ell(0) = e\big\},$$

where e denotes the unit element of Lie group G. Let $\rho(\cdot,\cdot)$ be the Riemannian metric on G, that is,

$$\rho(x,y) = \inf\Big\{L(\gamma) := \Big(\int_0^1 \big|\gamma(t)^{-1}\frac{\mathrm{d}}{\mathrm{d}t}\gamma(t)\big|_{\mathscr{G}}^2\,\mathrm{d}t\Big)^{1/2}\Big\},$$

where the infimum is taken over all absolutely continuous curves connecting x and y. It is easy to see $\rho(x,y) = \rho(e, x^{-1}y)$ by the definition. The topology of $\mathcal{P}(G)$ is determined by the uniform distance $d_\infty(\gamma_1,\gamma_2)$ for γ_1, $\gamma_2 \in \mathcal{P}(G)$, i.e.

$$d_\infty(\gamma_1,\gamma_2) := \max_{t\in[0,1]} \rho(\gamma_1(t),\gamma_2(t)). \tag{1.2}$$

Under this topology, $\mathcal{P}(G)$ becomes a complete separable space. We now introduce another distance on $\mathcal{P}(G)$, the L^2-distance:

$$d_{L^2}(\gamma_1,\gamma_2) = \Big(\int_0^1 \rho(\gamma_1(t),\gamma_2(t))^2\,\mathrm{d}t\Big)^{1/2}. \tag{1.3}$$

It is obvious that $d_{L^2}(\gamma_1,\gamma_2) \le d_\infty(\gamma_1,\gamma_2)$ for any γ_1, $\gamma_2 \in \mathcal{P}(G)$. In this paper, we consider the Wasserstein distance induced by the L^2-distance on $\mathcal{P}(G)$. Given two probability measures ν and σ over $\mathcal{P}(G)$, the L^p-Wasserstein distance between them is defined by:

$$W_p(\nu,\sigma) = \inf\Big\{\int_{\mathcal{P}(G)\times\mathcal{P}(G)} d_{L^2}(\gamma_1,\gamma_2)^p\,\pi(\mathrm{d}\gamma_1,\mathrm{d}\gamma_2);\ \pi \in \mathscr{C}(\nu,\sigma)\Big\}^{1/p},\quad p>1, \tag{1.4}$$

where $\mathscr{C}(\nu,\sigma)$ stands for the set of all probability measures on the product space $\mathcal{P}(G)\times\mathcal{P}(G)$ with marginals ν and σ respectively.

Set μ be the Wiener measure on $\mathcal{P}(G)$, which is the diffusion measure corresponding to the left invariant Laplace operator $\frac{1}{2}\sum_{i=1}^d \tilde{\xi}_i^2$ on G, where $\{\xi_1,\dots,\xi_d\}$ denotes an orthonormal basis of $\mathscr{G}$ and $\tilde{\xi}$ denotes the associated left invariant vector field on G.

Our first main results are the following two theorems on the existence and uniqueness of optimal transport maps on path groups and loop groups.

Theorem 1.1 *Let G be a connected compact Lie group and satisfy assumption (H). Let ν and σ be two probability measures on $\mathcal{P}(G)$, and assume ν is absolutely continuous with respect to the Wiener measure μ on $\mathcal{P}(G)$. Then for each $p>1$, there exists a unique measurable map $\mathscr{T}_p:\mathcal{P}(G)\to\mathcal{P}(G)$ such that it pushes ν forward to σ and*

$$W_p(\nu,\sigma)^p=\int_{\mathcal{P}(G)} d_{L^2}(\gamma,\mathscr{T}_p(\gamma))^p\,\mathrm{d}\nu(\gamma).$$

Furthermore, there exists some function ϕ in the Sobolev space $\mathbf{D}_1^2(\mu)$ such that the map $\mathscr{T}_2$ can be expressed as

$$\mathscr{T}_2(\gamma)(t)=\exp_{\gamma(t)}\Big(\frac{1}{2}\ell_{\gamma(t)}\frac{\mathrm{d}^2}{\mathrm{d}t^2}\big(\nabla\phi(\gamma)\big)(t)\Big),\quad a.e.\, t\in[0,1], \tag{1.5}$$

for almost every $\gamma\in\mathcal{P}(G)$. Here $\exp_\gamma$ denotes the geodesic exponential map on Lie group.

Theorem 1.2 *Let G be a connected compact Lie group and satisfy assumption (H). Let $\mathcal{L}_eG=\{\ell:[0,1]\to G$ continuous; $\ell(0)=\ell(1)=e\}$. Let σ_1 and σ_2 be two probability measures on $\mathcal{L}_eG$. Assume σ_1 is absolutely continuous with respect to the heat kernel measure ν on $\mathcal{L}_eG$. Then for each $p>1$ there exists a unique measurable map $\mathscr{T}_p:\mathcal{L}_eG\to\mathcal{L}_eG$ such that $(\mathscr{T}_p)_*\sigma_1=\sigma_2$ and*

$$W_p(\sigma_1,\sigma_2)^p=\int_{\mathcal{L}_eG} d_{L^2}(\ell,\mathscr{T}_p(\ell))^p\,\mathrm{d}\sigma_1(\ell),$$

where

$$W_p(\sigma_1,\sigma_2)^p:=\inf\Big\{\int_{\mathcal{L}_eG\times\mathcal{L}_eG} d_{L^2}(\ell_1,\ell_2)^p\,\pi(\mathrm{d}\ell_1,\mathrm{d}\ell_2);\ \pi\in\mathscr{C}(\sigma_1,\sigma_2)\Big\}.$$

In particular, for each $p>1$, there exists a unique measurable map $\mathscr{T}_p:\mathcal{L}_eG\to\mathcal{L}_eG$ such that $\mathscr{T}_p$ pushes heat kernel measure ν forward to pinned Wiener measure μ_0 on $\mathcal{L}_eG$, and its inverse $\mathscr{T}_p^{-1}$ pushes μ_0 forward to ν.

Moreover, for $p=2$ there exists some ϕ in the Sobolev space $\mathbf{D}_1^2(\nu)$ such that the map $\mathscr{T}_2$ can be expressed as

$$\mathscr{T}_2(\gamma)(\theta)=\exp_{\gamma(\theta)}\Big(\frac{1}{2}\ell_{\gamma(\theta)}\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\big(\nabla\phi(\gamma)\big)(\theta)\Big),\quad a.e.\, \theta\in[0,1], \tag{1.6}$$

for almost every $\gamma\in\mathcal{L}_eG$. Here $\exp_\gamma$ denotes the geodesic exponential map on Lie group.

The basic idea to prove Theorem 1.1 and Theorem 1.2 is similar to that of [15, 22] based on the solution of dual Kantorovich problem. The solution of dual Kantorovich problem gives us a pair of functions (ϕ, ϕ^c), where

$$\phi^c(y) = \inf_{x\in X}\{c(x,y) - \phi(x)\},$$

for some cost function $c(\cdot,\cdot)$ on some metric space X. Then the key point is to show that there is a uniquely determined measurable map $y = \mathscr{T}(x)$ such that

$$\phi(x) + \phi^c(\mathscr{T}(x)) = c(x, \mathscr{T}(x))$$

holds for suitable choice of x. In the language of c-convexity (cf. [28, Chapter 5]), it is equivalent to show that the subdifferential $\partial_c\phi(x)$ contains only one element for suitable choice of x.

Due to the explicit expression of the optimal transport map for L^2-Wasserstein distance, we applied previous results to show the existence of geodesics in the Wasserstein spaces over path groups and loop groups.

Theorem 1.3 *Assume the conditions of Theorem 1.1 hold. Then for any two probability measures ν_0, ν_1 on $\mathcal{P}(G)$ with ν_0 being absolutely continuous w.r.t. Wiener measure μ, there exists a curve of probability measures $(\nu_r)_{r\in[0,1]}$ connecting ν_0 and ν_1 satisfying*

$$W_2(\nu_0, \nu_r) = rW_2(\nu_0, \nu_1), \quad r \in [0,1].$$

Similarly, under the conditions of Theorem 1.2, for any two probability measures σ_0, σ_1 on $\mathcal{L}_eG$ with σ_0 absolutely continuous w.r.t. the heat kernel measure, there exists a geodesic $(\sigma_r)_{r\in[0,1]}$ in $(\mathscr{P}(\mathcal{L}_eG), W_2)$ connecting σ_0 to σ_1.

This paper is organized as follows: in the next section, we introduce some notations and basic results on Lie group. In section 3, we give the proofs of Theorem 1.1 in the case $p = 2$ and Theorem 1.3 in order to explain the idea of the argument. For the general case $p > 1$, the proof of Theorem 1.1 is stated in section 4. In the last section, we investigate the Monge-Kantorovich problem on loop groups. Some basic notations on loop group and the argument of Theorem 1.2 are stated there.

2 Preliminaries

We first review some basic notions and results on the Lie group and its Lie algebra. The proofs of these results will be omitted, and refer to Warner's book [30] for details.

A Lie group G is a differentiable manifold which is also endowed with a group structure such that the map $G \times G \to G$ defined by $(\sigma, \tau) \mapsto \sigma\tau^{-1}$ is smooth. Let $\sigma \in G$, left translation by σ and right translation by σ are respectively the diffeomorphisms ℓ_σ and r_σ of G defined by

$$\ell_\sigma(\tau) = \sigma\tau, \quad r_\sigma(\tau) = \tau\sigma \quad \text{for all } \tau \in G.$$

A vector field X on G is called left invariant if for each $\sigma \in G$,

$$d\ell_\sigma \circ X = X \circ \ell_\sigma.$$

A Lie algebra of the Lie group G is defined to be the Lie algebra $\mathscr{G}$ of left invariant vector fields on G. The map $\alpha : \mathscr{G} \to T_eG$ defined by $\alpha(X) = X(e)$ is an isomorphism from the Lie algebra $\mathscr{G}$ to the tangent space of G at the identity. α is injective and surjective. It will be convenient at times to look on the Lie algebra as the tangent space of G at the identity. We consider the left invariant vector fields on G, then the tangent space T_gG at every point $g \in G$ can be viewed as $g\mathscr{G}$, and the inner product $\langle\, ,\,\rangle_{\mathscr{G}}$ induces a inner product on T_gG by

$$\langle d\ell_g \circ X, d\ell_g \circ Y\rangle = \langle X, Y\rangle_{\mathscr{G}}, \quad X, Y \in \mathscr{G}.$$

Examples of Lie group and its Lie algebra:

a) The set $gl(n, \mathbb{R})$ of all $n \times n$ real matrices is a real vector space. Matrices are added and multiplied by scalars componentwise. $gl(n, \mathbb{R})$ becomes a Lie algebra if we set $[A, B] = AB - BA$.

   The general linear group $Gl(n, \mathbb{R})$ is the set of all $n \times n$ non-singular real matrices. Then $Gl(n, \mathbb{R})$ becomes a Lie group under matrix multiplication, and $gl(n, \mathbb{R})$ can be considered as the Lie algebra of $Gl(n, \mathbb{R})$.

b) *Special linear group* $Sl(n, \mathbb{R}) = \{A \in Gl(n, \mathbb{R}) : \det A = 1\}$ is a Lie group. Its Lie algebra will be matrices of trace 0, $sl(n, \mathbb{R}) = \{A \in gl(n, \mathbb{R}) : \operatorname{trace} A = 0\}$.

Definition 2.1 *Let G and H be Lie groups. A map $\phi : G \to H$ is a (Lie group) homomorphism if ϕ is both C^∞ and a group homomorphism of the abstract groups.*

Let $\mathscr{G}$ and $\mathscr{H}$ be Lie algebra, a map $\psi : \mathscr{G} \to \mathscr{H}$ is a (Lie algebra) homomorphism if it is linear and preserves Lie brackets, i.e. $\psi([X, Y]) = [\psi(X), \psi(Y)]$ for all $X, Y \in \mathscr{G}$.

A homomorphism $\phi : \mathbb{R} \to G$ is called a 1-parameter subgroup of G. For each $X \in \mathscr{G}$, there exists a unique 1-parameter subgroup $t \mapsto \sigma_X(t)$ such that its tangent vector at 0 is $X(e)$. This induces a definition of exponential map on Lie group by $\exp X = \sigma_X(1)$. This definition of exponential map does not depend on the metric on $\mathscr{G}$. In matrix Lie groups, the exponential map $\exp A$ coincides with the usual exponential of matrices

$$\exp A = \sum_{k=0}^{\infty} \frac{A^k}{k!}.$$

Definition 2.2 *Let $\sigma \in G$. We define the action* Ad_σ *on* $\mathscr{G}$ *by*

$$\mathrm{Ad}_\sigma X = \left\{\frac{\mathrm{d}}{\mathrm{d}t}\sigma\exp(tX)\sigma^{-1}\right\}_{t=0} \quad \textit{for each } X \textit{ in } \mathscr{G}. \tag{2.1}$$

$\mathrm{Ad} : G \to \mathrm{Aut}(\mathscr{G})$ *is called the adjoint representation of Lie group G, where* $\mathrm{Aut}(\mathscr{G})$ *denotes the set of automorphisms on $\mathscr{G}$.*

By the definition, it is easy to obtain that for all $\sigma,\,\tau \in G$ $\mathrm{Ad}_\sigma X = d\ell_\sigma dr_{\sigma^{-1}}X$. Let $X,\,Y \in \mathscr{G}$, define

$$\mathrm{ad}_X Y = \left\{\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Ad}_{\exp tX}Y\right\}_{t=0}. \tag{2.2}$$

Then $\mathrm{ad}_X Y = [X, Y]$.

For each $A \in \mathscr{G}$, let $\tilde{A}$ denote the unique left invariant vector field on G determined by A. Given a metric $\langle\ ,\ \rangle$ on Lie algebra $\mathscr{G}$. It can induce a left invariant Riemannian metric on G. The Levi-Civita connection on G induced by this metric is given by

$$\langle\nabla_{\tilde{A}}\tilde{B}, \tilde{C}\rangle = \frac{1}{2}\big\{\langle[A,B],C\rangle - \langle[A,C],B\rangle - \langle[B,C],A\rangle\big\}, \quad \text{for } A,\ B,\ C \in \mathscr{G}. \tag{2.3}$$

Let ad^*_A be the adjoint operator of ad_A w.r.t $\langle\ ,\ \rangle$. Then

$$\nabla_{\tilde{A}}\tilde{B} = \widetilde{\nabla_A B}, \quad \nabla_A B = \frac{1}{2}\big(\mathrm{ad}_A B - \mathrm{ad}^*_A B - \mathrm{ad}^*_B A\big). \tag{2.4}$$

Given an orthonormal basis $\{e_i\}_{i=1}^d$ of $\mathscr{G}$, since $\langle\tilde{e}_i, \tilde{e}_j\rangle_\sigma = \langle e_i, e_j\rangle$, $\{\tilde{e}_i\}_{i=1}^d$ is a family of orthonormal vector fields on G. Let

$$\Gamma^k_{ij} = \langle\nabla_{e_i}e_j, e_k\rangle = \langle\nabla_{\tilde{e}_i}\tilde{e}_j, \tilde{e}_k\rangle. \tag{2.5}$$

Then with respect to the Levi-Civita connection a C^1 curve $(\gamma_t,\ a < t < b)$ on G is called a *geodesic* if $\dot\gamma(t) := \mathrm{d}\gamma(t)/\mathrm{d}t$ is parallel along γ, i.e.

$$\nabla_{\dot\gamma(t)}\dot\gamma(t) = 0. \tag{2.6}$$

Setting $\dot\gamma_k(t) = \langle\dot\gamma(t), \tilde{e}_k\rangle$, then the equation (2.6) turns into

$$\frac{\mathrm{d}\dot\gamma_k(t)}{\mathrm{d}t} + \sum_{i,j=1}^d \Gamma^k_{ij}\dot\gamma_i(t)\dot\gamma_j(t) = 0, \quad \text{for } k = 1, \ldots, d. \tag{2.7}$$

Note that in this equation, the Christoffel coefficients Γ^k_{ij} are independent of the curve $\gamma(t)$, which is different to general geodesic equations on manifolds. This geodesic equation induces another definition of exponential map on Lie group when being viewed as a Riemannian manifold, and this exponential map depends on the inner product defined on Lie algebra $\mathscr{G}$. But when $\mathscr{G}$ is endowed with an Ad-invariant metric $\langle\ ,\ \rangle$, namely, $\langle \mathrm{Ad}_g X, \mathrm{Ad}_g Y\rangle = \langle X, Y\rangle$, for all $g \in G$ and X, $Y \in \mathscr{G} = T_e G$, then 1-parameter subgroups are geodesics (see [7, Corollary 3.19]) and every geodesic is coincident with a translation of a segment of 1-parameter subgroup (see J. F. Price [25, Theorem 4.3.3]). This enable us to know that for compact connected Lie groups the exponential maps induced by 1-parameter subgroup are surjective. It is known (cf. [7, proposition 5.4]) that the cut locus of each point g on G is closed, and contains two kinds of points, i.e. if g' is in the cut locus of g, then g' is either the first conjugate point of g along some geodesic connecting g with g', or there exists at least two minimizing geodesics joining g to g'. When G is a simply connected Lie group with Ad-invariant metric, then all geodesics minimize up to the first conjugate point (cf. [7, Corollary 5.12]).

There are lots of work about infinite dimensional stochastic analysis on path groups and loop groups. We refer to [9, 10] and the book [14] for some basic facts and results.

3 Proof of Theorem 1.1: the case $p = 2$

Let us recall a well known result on the solving of Kantorovich dual problem. Refer to [28, Theorem 5.10] for the argument.

Theorem 3.1 *Let X and Y be two Polish spaces and μ, ν be two probability measures on X and Y respectively. Let $c : X \times Y \to \mathbb{R}$ be a lower semicontinuous cost function such that*

$$\forall\, (x, y) \in X \times Y, \quad c(x, y) \geq a(x) + b(y)$$

for some real-valued upper semicontinuous functions $a \in L^1(\mu)$ and $b \in L^1(\nu)$. Then if

$$C(\mu, \nu) := \inf_{\pi \in \mathscr{C}(\mu,\nu)} \int c \,\mathrm{d}\pi$$

is finite, and one has the pointwise upper bound

$$c(x, y) \leq c_X(x) + c_Y(y), \quad (c_X, c_Y) \in L^1(\mu) \times L^1(\nu), \tag{3.1}$$

then both the primal and dual Kantorovich problems have solutions, so

$$\begin{aligned} &\min_{\pi \in \mathscr{C}(\mu,\nu)} \int_{X\times Y} c(x, y) \,\mathrm{d}\pi(x, y) \\ &= \max_{(\phi,\psi)\in L^1(\mu)\times L^1(\nu):\, \phi+\psi\leq c} \Big(\int_X \phi(x) \,\mathrm{d}\mu(x) + \int_Y \psi(y) \,\mathrm{d}\nu(y)\Big) \\ &= \max_{\phi\in L^1(\mu)} \Big(\int_X \phi(x) \,\mathrm{d}\mu(x) + \int_Y \phi^c(y) \,\mathrm{d}\nu(y)\Big), \end{aligned} \tag{3.2}$$

where

$$\phi^c(y) := \inf_{x\in X} \big\{c(x, y) - \phi(x)\big\}. \tag{3.3}$$

In our situation, the diameter D of Lie group G is finite as G is assumed to be compact. Then $d_{L^2}(x,y) \le D < +\infty$, and hence the condition (3.1) is satisfied for any probability measures ν and σ on $\mathcal{P}(G)$. $d_{L^2}(x,y)$ is also continuous on $\mathcal{P}(G)$. According to Theorem 3.1, it holds that for any two probability measures ν and σ on $\mathcal{P}(G)$,

$$W_2(\nu,\sigma)^2 = \sup\Big\{ \int_{\mathcal{P}(G)} \phi(x)\,\nu(\mathrm{d}x) + \int_{\mathcal{P}(G)} \psi(y)\,\sigma(\mathrm{d}y)\Big\},$$

where the supremum runs among all pairs of measurable functions (ϕ,ψ) such that $\phi(x) + \psi(y) \le d_{L^2}(x,y)^2$. Moreover, there exists a pair of functions (ψ^c,ψ) such that

$$W_2(\nu,\sigma)^2 = \int_{\mathcal{P}(G)} \psi^c(x)\,\nu(\mathrm{d}x) + \int_{\mathcal{P}(G)} \psi(y)\,\sigma(\mathrm{d}y), \tag{3.4}$$

where $\psi^c(x) = \inf_{x\in\mathcal{P}(G)} \big\{ d_{L^2}(x,y)^2 - \psi(y)\big\}$. In the rest of this section, we will fix such pair of functions (ψ^c,ψ) and denote by $\phi(x) = \psi^c(x)$. Then ϕ is Lipschitz continuous with respect to the distance $d_{L^2}(x,y)$. In fact, for any fixed x, $z \in \mathcal{P}(G)$, for any $\varepsilon > 0$, there exists $y_\varepsilon \in \mathcal{P}(G)$ such that $\psi^c(z) \ge d_{L^2}(z,y_\varepsilon)^2 - \psi(y_\varepsilon) - \varepsilon$. Then

$$\begin{aligned}
\phi(x) - \phi(z) &= \psi^c(x) - \psi^c(z)\\
&\le d_{L^2}(x,y_\varepsilon)^2 - \psi(y_\varepsilon) - d_{L^2}(z,y_\varepsilon)^2 + \psi(y_\varepsilon) + \varepsilon\\
&\le \big(d_{L^2}(x,y_\varepsilon) + d_{L^2}(z,y_\varepsilon)\big)\big(d_{L^2}(x,y_\varepsilon) - d_{L^2}(z,y_\varepsilon)\big) + \varepsilon\\
&\le 2D d_{L^2}(x,z) + \varepsilon.
\end{aligned}$$

Letting $\varepsilon \to 0^+$, we get $\phi(x) - \phi(z) \le 2D d_{L^2}(x,z)$. Changing the place of x and z, we get ϕ is Lipschitz continuous.

Next, we shall use the Rademacher's theorem on path group $\mathcal{P}(G)$ to show that ϕ is in the Sobolev space. Before this, we introduce some basic notions. Let

$$H(\mathscr{G}) = \Big\{ h : [0,1] \to \mathscr{G};\ h(0) = 0,\ |h|_H^2 = \int_0^1 \big|\dot h(t)\big|_{\mathscr{G}}^2\,\mathrm{d}t < +\infty\Big\}, \tag{3.5}$$

where dot $\cdot$ stands for the derivative with respect to t. Let $F : \mathcal{P}(G) \to \mathbb{R}$ be a measurable function. We set

$$D_h F(\gamma) = \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\Big|_{\varepsilon=0} F(\gamma e^{\varepsilon h}), \quad h \in H(\mathscr{G}),\ \gamma \in \mathcal{P}(G).$$

A function $F \in L^2(\mu)$ is said to be in the Sobolev space $\mathbf{D}_1^2(\mu)$ if there exists $\nabla F \in L^2(\mu; H(\mathscr{G}))$ such that for each $h \in H(\mathscr{G})$, it holds $D_h F = \langle \nabla F, h\rangle_H$ in $L^{2-}(\mu)$, where $L^{2-}(\mu) = \bigcap_{p<2} L^p(\mu)$ and

$$\langle h_1, h_2\rangle_H = \int_0^1 \langle \dot h_1(t), \dot h_2(t)\rangle_{\mathscr{G}}\,\mathrm{d}t, \quad h_1,\ h_2 \in H(\mathscr{G}).$$

A function $F:\mathcal{P}(G)\to\mathbb{R}$ is said to be cylindrical if

$$F(\gamma)=f(\gamma(t_1),\ldots,\gamma(t_n)),\quad f\in C^\infty(G^n),\ 0<t_1<\ldots<t_n\le 1,\ n\in\mathbb{N}.$$

Let $\mathbf{Cyln}(\mathcal{P}(G))$ denote the space of all cylindrical functions. Due to [1], the space $\mathbf{Cyln}(\mathcal{P}(G))$ is dense in $\mathbf{D}_1^2(\mu)$. Now we introduce the third distance, Cameron-Martin distance $d_{\mathcal{P}}$, on $\mathcal{P}(G)$, that is, for $\gamma_1,\ \gamma_2\in\mathcal{P}(G)$,

$$d_{\mathcal{P}}(\gamma_1,\gamma_2)=\Big(\int_0^1\big|v(t)^{-1}\dot v(t)\big|_{\mathscr{G}}^2\,\mathrm{d}t\Big)^{1/2},\quad \text{if } v=\gamma_1^{-1}\gamma_2 \text{ absolutely continuous;} \tag{3.6}$$

otherwise, set $d_{\mathcal{P}}(\gamma_1,\gamma_2)=+\infty$. It's easy to check that $d_\infty(\gamma_1,\gamma_2)\le d_{\mathcal{P}}(\gamma_1,\gamma_2)$ for all $\gamma_1,\ \gamma_2\in\mathcal{P}(G)$. According to the Rademacher's theorem [26, Theorem 1.5] and discussions in subsection 2.1 therein, we obtain that

Lemma 3.2 *Any bounded $d_{\mathcal{P}}$-Lipschitz continuous function F on $\mathcal{P}(G)$ belongs to $\mathbf{D}_1^2(\mu)$.*

Here and in the sequel, a function F on a metric space (X,d) is said to be **d-Lipschitz continuous**, where d is a metric on X, if there exists some constant $C>0$ such that

$$|F(x)-F(y)|\le Cd(x,y),\quad \forall\, x,\ y\in X.$$

Due to the fact

$$d_{L^2}(\gamma_1,\gamma_2)\le d_\infty(\gamma_1,\gamma_2)\le d_{\mathcal{P}}(\gamma_1,\gamma_2),$$

it is clear that a d_{L^2}-Lipschitz continuous function F is also d_∞-Lipschitz and $d_{\mathcal{P}}$-Lipschitz continuous. Combining d_{L^2}-Lipschitz continuity of ϕ with Lemma 3.2, we get ϕ is in the Sobolev space $\mathbf{D}_1^2(\mu)$.

Proposition 3.3 (Key proposition) *If there exist γ_1 and γ_2 such that*

$$\phi(\gamma_1)+\psi(\gamma_2)=d_{L^2}(\gamma_1,\gamma_2)^2, \tag{3.7}$$

and ϕ is differentiable at γ_1, then γ_2 is uniquely determined by γ_1 and ϕ.

Proof. For $h\in H(\mathscr{G})$ and $\varepsilon>0$, by the fact $\phi=\psi^c$, we get

$$\phi(\gamma_1 e^{\varepsilon h})+\psi(\gamma_2)\le d_{L^2}(\gamma_1 e^{\varepsilon h},\gamma_2)^2.$$

Subtracting (3.7) from both sides of this inequality yields

$$\begin{aligned}\phi(\gamma_1 e^{\varepsilon h})-\phi(\gamma_1)&\le d_{L^2}(\gamma_1 e^{\varepsilon h},\gamma_2)^2-d_{L^2}(\gamma_1,\gamma_2)^2\\ &=\int_0^1\rho\big(\gamma_1(t)e^{\varepsilon h(t)},\gamma_2(t)\big)^2\,\mathrm{d}t-\int_0^1\rho\big(\gamma_1(t),\gamma_2(t)\big)^2\,\mathrm{d}t.\end{aligned} \tag{3.8}$$

For each fixed $t \in [0,1]$, there exits a constant speed geodesic $v_t : [0,1] \to G$ such that $v_t(0) = \gamma_2(t)^{-1}\gamma_1(t)$, $v_t(1) = e$ and

$$L(v_t)^2 = \int_0^1 \left|v_t^{-1}(s)\frac{\mathrm{d}}{\mathrm{d}s}v_t(s)\right|_{\mathscr{G}}^2 \mathrm{d}s = \rho(\gamma_1(t), \gamma_2(t))^2.$$

Set $\tilde{v}_t(s) = v_t(s)e^{(1-s)\varepsilon h(t)}$, $s \in [0,1]$. Then $\tilde{v}_t(0) = \gamma_2(t)^{-1}\gamma_1(t)e^{\varepsilon h(t)}$ and $\tilde{v}_t(1) = e$. Hence,

$$\rho\big(\gamma_1(t)e^{\varepsilon h(t)}, \gamma_2(t)\big)^2 \le L(\tilde{v}_t)^2. \tag{3.9}$$

As

$$d_s\tilde{v}_t(s) = \big(\dot{v}_t(s)e^{(1-s)\varepsilon h(t)} - \varepsilon\tilde{v}_t(s)h(t)\big)ds,$$

where d_s stands for the derivative with respect to s, we get

$$\begin{aligned} L(\tilde{v}_t)^2 &= \int_0^1 \left|\tilde{v}_t(s)^{-1}\dot{\tilde{v}}_t(s)\right|_{\mathscr{G}}^2 \mathrm{d}s \\ &= \int_0^1 \left|\mathrm{Ad}_{e^{-(1-s)\varepsilon h(t)}}v_t(s)^{-1}\dot{v}_t(s) - \varepsilon h(t)\right|_{\mathscr{G}}^2 \mathrm{d}s \\ &= \int_0^1 \left|v_t(s)^{-1}\dot{v}_t(s)\right|_{\mathscr{G}}^2 - 2\varepsilon\langle \mathrm{Ad}_{e^{-(1-s)\varepsilon h(t)}}v_t(s)^{-1}\dot{v}_t(s), h(t)\rangle_{\mathscr{G}} + \varepsilon^2|h(t)|_{\mathscr{G}}^2 \mathrm{d}s \\ &= \rho(\gamma_1(t), \gamma_2(t))^2 - 2\varepsilon\int_0^1 \langle \mathrm{Ad}_{e^{-(1-s)\varepsilon h(t)}}v_t(s)^{-1}\dot{v}_t(s), h(t)\rangle_{\mathscr{G}} \mathrm{d}s + \varepsilon^2|h(t)|_{\mathscr{G}}^2. \end{aligned}$$

Invoking (3.8) and (3.9), we obtain

$$\phi(\gamma_1 e^{\varepsilon h}) - \phi(\gamma_1) \le -2\varepsilon\int_0^1\int_0^1 \langle \mathrm{Ad}_{e^{-(1-s)\varepsilon h(t)}}v_t(s)^{-1}\dot{v}_t(s), h(t)\rangle_{\mathscr{G}} \mathrm{d}s\mathrm{d}t + \varepsilon^2\int_0^1 |h(t)|_{\mathscr{G}}^2 \mathrm{d}t.$$

Dividing both sides by ε, letting $\varepsilon \to 0^+$ and $\varepsilon \to 0^-$ respectively, it follows

$$\langle \nabla\phi(\gamma_1),\, h\rangle_H \le -2\int_0^1 \Big\langle \int_0^1 v_t(s)^{-1}\dot{v}_t(s)\,\mathrm{d}s,\, h(t)\Big\rangle_{\mathscr{G}} \mathrm{d}t, \tag{3.10}$$

$$\langle \nabla\phi(\gamma_1),\, h\rangle_H \ge -2\int_0^1 \Big\langle \int_0^1 v_t(s)^{-1}\dot{v}_t(s)\,\mathrm{d}s,\, h(t)\Big\rangle_{\mathscr{G}} \mathrm{d}t. \tag{3.11}$$

Set

$$V_t(u) = \int_0^u v_t(s)^{-1}\dot{v}_t(s)\,\mathrm{d}s, \quad u \in [0,1], \tag{3.12}$$

then we have shown by (3.10) (3.11) that

$$\langle \nabla\phi(\gamma_1),\, h\rangle_H = -2\int_0^1 \langle V_t(1),\, h(t)\rangle_{\mathscr{G}} \mathrm{d}t, \tag{3.13}$$

which implies that if $V_t(1)$ as a function of t is continuous at some $t_0 \in [0,1]$ then $V_{t_0}(1)$ is uniquely determined. In fact, take a sequence of smooth functions h_ε such that $0 \le h_\varepsilon \le 1$, and

$$h_\varepsilon(t) = \begin{cases} 1 & t \in [t_0-\varepsilon, t_0+\varepsilon] \cap [0,1], \\ 0 & t \notin [t_0-2\varepsilon, t_0+2\varepsilon] \cap [0,1]. \end{cases}$$

Set $\{e_1, \dots, e_d\}$ be an orthonormal basis of $\mathscr{G}$. We have

$$\langle V_{t_0}(1), e_i \rangle_{\mathscr{G}} = \lim_{\varepsilon \to 0} \int_0^1 \langle V_t(1), h_\varepsilon(t) e_i \rangle_{\mathscr{G}} \, \mathrm{d}t = \lim_{\varepsilon \to 0} -\frac{1}{2} \langle \nabla \phi(\gamma_1), h_\varepsilon e_i \rangle_H . \tag{3.14}$$

Moreover, as $\langle \nabla\phi(\gamma_1), h_\varepsilon e_i\rangle_H$ is measurable from $\mathcal{P}(G)$ to $\mathbb{R}$, the limitation $\langle V_{t_0}(1), e_i\rangle_{\mathscr{G}}$ is also measurable. Then $V_{t_0}(1) = \sum_i \langle V_{t_0}(1), e_i\rangle_{\mathscr{G}} e_i$ is measurable with respect to the variable γ_1.

We shall show below the following assertion: according to our assumption (H), we can always choose a family of minimizing geodesics $(v_t)_{t\in[0,1]}$ such that $v_t(0) = \gamma_2^{-1}(t)\gamma_1(t)$, $v_t(1) = e$, and there exists an at most countable subset $\Omega \subset [0,1]$ such that $t \mapsto \dot v_t(1)$ is continuous on $[0,1]\backslash\Omega$. Hence, $t \mapsto V_t(1)$ is continuous on $[0,1]\backslash\Omega$. If this assertion is correct, then $V_t(1)$ is uniquely determined by $\nabla\phi(\gamma_1)$ at all $t \in [0,1]\backslash\Omega$. Then, due to Lemma 3.4 below, we know that $V_t(1)$ determines uniquely a geodesic $v_t : [0,1] \to G$ so that $v_t(1) = e$. Therefore, $\gamma_2(t)$ is uniquely determined by $\nabla\phi(\gamma_1)$ at $t \in [0,1]\backslash\Omega$. Since $t \mapsto \gamma_2(t)$ is continuous, then we get the desired result that γ_2 is uniquely determined by ϕ and γ_1. Now we proceed to the proof of previous assertion.

Case 1: If $\{\gamma_2(t)^{-1}\gamma_1(t),\ t \in [0,1]\}$ does not go across the cut locus of e in G, then there exists a unique family of minimizing geodesics $\big(v_t\big)_{t\in[0,1]}$ such that $v_t(0) = \gamma_2^{-1}(t)\gamma_1(t)$, $v_t(1) = e$, and $t \mapsto \dot v_t(1)$ is continuous. The geodesic equation guarantees that $t \mapsto v_t(s)^{-1}\dot v_t(s)$ for $s \in [0,1]$ is also continuous, which implies the continuity of $t \mapsto V_t(1)$ for $t \in [0,1]$.

Case 2: If $\gamma_2^{-1}(t)\gamma_1(t)$ goes across the cut locus of e. Then the continuity of $\gamma_1(t)$, $\gamma_2(t)$ and the closeness of the cut locus of e yield that the set Ω containing all t such that $\gamma_2^{-1}(t)\gamma_1(t)$ enters or leaves the cut locus of e is not empty and at most countable. So the set $I_1 := \{t \in [0,1]\backslash\Omega;\ \gamma_2^{-1}(t)\gamma_1(t) \notin \mathit{Cut}(e)\}$ and the set $I_2 := \{t \in [0,1]\backslash\Omega;\ \gamma_2^{-1}(t)\gamma_1(t) \in \mathit{Cut}(e)\}$ can both be represented as the union of at most countable open intervals. For each open interval $(s_1, s_2) \subset I_1$, above discussion in case 1 show that there exist a curve $t \mapsto V_t(1)$ for $t \in (s_1, s_2)$. For each open interval $(s_1', s_2') \subset I_2$, our assumption (H) may guarantee that we can choose a family of geodesics $(v_t)_{t\in(s_1',s_2')}$ such that $v_t(0) = \gamma_2^{-1}\gamma_1(t)$, $v_t(1) = e$, and $t \mapsto \dot v_t(0)$ is continuous on (s_1', s_2'). This yields $t \mapsto V_t(1)$ is continuous on (s_1', s_2').

In all, we can choose a $(V_t(1))_{t\in[0,1]}$ such that $t \mapsto V_t(1)$ is continuous on $[0,1]\backslash\Omega$. Therefore, we have proved the assertion and complete the proof of this proposition. ∎

Lemma 3.4 *Using the notations as above. Then $V_t(1)$ determines uniquely a minimizing geodesic $v_t : [0,1] \to G$ such that $v_t(1) = e$.*

Proof. Let $a \in \mathscr{G}$, $\varepsilon \in \mathbb{R}$ and $c \in C^2([0,1],\mathbb{R})$ such that $c(0) = c(1) = 0$. Consider $v_{t,\varepsilon}(s) = v_t(s)e^{\varepsilon c(s)a}$, $s \in [0,1]$. Then $v_{t,\varepsilon}(0) = v_t(0)$ and $v_{t,\varepsilon}(1) = v_t(1)$.

$$d_s v_{t,\varepsilon}(s) = v_{t,\varepsilon}(s)\Big(\mathrm{Ad}_{e^{-\varepsilon c(s)a}} v_t(s)^{-1}\dot{v}_t(s) + \varepsilon c'(s)a\Big)ds,$$

and

$$L(v_{t,\varepsilon})^2 = \int_0^1 \Big|\mathrm{Ad}_{e^{-\varepsilon c(s)a}} v_t(s)^{-1}\dot{v}_t(s) + \varepsilon c'(s)a\Big|_{\mathscr{G}}^2 \,\mathrm{d}s.$$

Since $\varepsilon \mapsto L(v_{t,\varepsilon})^2$ arrives its minimum at $\varepsilon = 0$, we get

$$\begin{aligned} 0 =& \frac{\mathrm{d}}{\mathrm{d}\varepsilon} L(v_{t,\varepsilon})^2 = 2\int_0^1 \langle v_t(s)^{-1}\dot{v}_t(s), c'(s)a\rangle_{\mathscr{G}} \,\mathrm{d}s \\ =& 2\int_0^1 \langle V_t'(s), c'(s)a\rangle_{\mathscr{G}} \,\mathrm{d}s \\ =& 2\langle V_t(1), c'(1)a\rangle_{\mathscr{G}} - 2\langle V_t(0), c'(0)a\rangle_{\mathscr{G}} - 2\int_0^1 \langle V_t(s), c''(s)a\rangle_{\mathscr{G}} \,\mathrm{d}s. \end{aligned}$$

This yields

$$\langle V_t(1), c'(1)a\rangle_{\mathscr{G}} = \int_0^1 \langle V_t(s), c''(s)a\rangle_{\mathscr{G}} \,\mathrm{d}s. \tag{3.15}$$

Assume $\tilde{v}_t$ be another minimizing geodesic such that $\tilde{v}_t(1) = e$ and $\tilde{V}_t(1) = V_t(1)$, where $\tilde{V}_t(u) = \int_0^u \tilde{v}_t(s)^{-1}\dot{\tilde{v}}_t(s)\,\mathrm{d}s$, $u \in [0,1]$. Then analogous deduction yields

$$\int_0^1 \langle \tilde{V}_t(s), c''(s)a\rangle_{\mathscr{G}} \,\mathrm{d}s = \int_0^1 \langle V_t(s), c''(s)a\rangle_{\mathscr{G}} \,\mathrm{d}s.$$

Since $s \mapsto c''(s)a$ is dense in $L^2(\mu;\mathscr{G})$,

$$V_t(s) = \tilde{V}_t(s), \quad \text{for almost every } s \in [0,1].$$

The continuity of $s \mapsto V_t(s)$ and $s \mapsto \tilde{V}_t(s)$ yields

$$V_t(s) = \tilde{V}_t(s), \quad \text{for all } s \in [0,1].$$

Therefore,

$$\frac{\mathrm{d}}{\mathrm{d}s} V_t(s) = \frac{\mathrm{d}}{\mathrm{d}s}\tilde{V}_t(s), \ \textit{i.e.}\ v_t(s)^{-1}\dot{v}_t(s) = \tilde{v}_t(s)^{-1}\dot{\tilde{v}}_t(s) =: k_t(s).$$

Since the solution of

$$d_s v_t(s) = v_t(s)k_t(s)ds, \quad v_t(1) = e$$

is unique, it follows that $\tilde{v}_t(s) = v_t(s)$ for all $s \in [0,1]$. In particular, $\tilde{v}_t(0) = v_t(0)$. The proof is complete. ∎

Proof of Theorem 1.1 for $p = 2$. We have shown in (3.4) that

$$W_2(\nu, \sigma)^2 = \int_{\mathcal{P}(G)} \phi(\gamma_1)\,\mathrm{d}\nu + \int_{\mathcal{P}(G)} \psi(\gamma_2)\,\mathrm{d}\sigma.$$

By Lemma 3.2, ϕ is in $\mathbf{D}_1^2(\mu)$. So ϕ is μ-almost everywhere differentiable, so does also with respect to ν by the absolute continuity of ν relative to μ. Since $d_{L^2}(\cdot,\cdot)$ is continuous from $\mathcal{P}(G) \times \mathcal{P}(G)$ to $\mathbb{R}$, and $\mathscr{C}(\nu, \sigma)$ is tight, there exists an optimal transport plan $\pi \in \mathscr{C}(\nu, \sigma)$ such that

$$W_2(\nu, \sigma)^2 = \int_{\mathcal{P}(G)\times\mathcal{P}(G)} d_{L^2}(\gamma_1, \gamma_2)^2\,\pi(\mathrm{d}\gamma_1, \mathrm{d}\gamma_2).$$

Hence,

$$\int_{\mathcal{P}(G)\times\mathcal{P}(G)} \phi(\gamma_1) + \psi(\gamma_2)\,\pi(\mathrm{d}\gamma_1, \mathrm{d}\gamma_2) = \int_{\mathcal{P}(G)\times\mathcal{P}(G)} d_{L^2}(\gamma_1, \gamma_2)^2\,\pi(\mathrm{d}\gamma_1, \mathrm{d}\gamma_2).$$

As $\phi = \psi^c$, there exists a measurable set $\Omega_1 \subset \mathcal{P}(G) \times \mathcal{P}(G)$ such that $\pi(\Omega_1) = 1$, and

$$\phi(\gamma_1) + \psi(\gamma_2) = d_{L^2}(\gamma_1, \gamma_2)^2, \quad \forall\,(\gamma_1, \gamma_2) \in \Omega_1.$$

Since ϕ is ν-a.e. differentiable, there exists a measurable set $A \subset \mathcal{P}(G)$ with $\nu(A) = 1$ on which ϕ is differentiable everywhere. Let $\Omega = \Omega_1 \cap (A \times \mathcal{P}(G))$, then $\pi(\Omega) = 1$.

For a point $(\gamma_1, \gamma_2) \in \Omega$, Proposition 3.3 yields that $\gamma_2 \in \mathcal{P}(G)$ is uniquely determined by γ_1 and ϕ such that

$$\phi(\gamma_1) + \psi(\gamma_2) = d_{L^2}(\gamma_1, \gamma_2)^2.$$

Denote this map by $\gamma_2 = \mathscr{T}(\gamma_1)$. Assume $\mathscr{T}$ is measurable, then for any measurable function F on $\mathcal{P}(G) \times \mathcal{P}(G)$,

$$\begin{aligned}\int_{\mathcal{P}(G)\times\mathcal{P}(G)} F(\gamma_1, \gamma_2)\,\pi(\mathrm{d}\gamma_1, \mathrm{d}\gamma_2) &= \int_{\mathcal{P}(G)\times\mathcal{P}(G)} F(\gamma_1, \mathscr{T}(\gamma_1))\,\pi(\mathrm{d}\gamma_1, \mathrm{d}\gamma_2)\\ &= \int_{\mathcal{P}(G)} F(\gamma_1, \mathscr{T}(\gamma_1))\,\nu(\mathrm{d}\gamma_1).\end{aligned}$$

This implies that

$$\pi = (id \times \mathscr{T})_*\nu \quad \text{and } (\mathscr{T})_*\nu = \sigma. \tag{3.16}$$

If there exists another measurable map $\mathscr{S} : \mathcal{P}(G) \to \mathcal{P}(G)$ such that $(\mathscr{S})_*\nu = \sigma$ and

$$W_2(\nu, \sigma)^2 = \int_{\mathcal{P}(G)} d_{L^2}(\gamma, \mathscr{S}(\gamma))^2\,\nu(\mathrm{d}\gamma).$$

Then the measure $\tilde{\pi} := (id \times \mathscr{S})_* \nu$ is an optimal transport map. Since in above discussion π is arbitrary optimal transport plan in $\mathscr{C}(\nu, \sigma)$, applying (3.16) to $\tilde{\pi}$, we obtain

$$\tilde{\pi} = (id \times \mathscr{T})_* \nu, \quad \text{and } \mathscr{S} = \mathscr{T}, \ \nu\text{-a.e.}.$$

This proves the uniqueness of $\mathscr{T}$.

Now we proceed to the measurability of $\mathscr{T}$.

Let $\{\beta_n, \, n \geq 1\} \subset C^\infty([0,1], \mathbb{R})$ be an orthonormal basis of the space $H(\mathbb{R}) = \big\{ f : [0,1] \to \mathbb{R}; \ f(0) = 0, \ \int_0^1 |f'(s)|^2 \mathrm{d}s < +\infty \big\}$. Define

$$c_n(t) = \int_0^t \beta_n(s) \, \mathrm{d}s - t \int_0^1 \beta_n(s) \, \mathrm{d}s.$$

Let $\{e_1, \ldots, e_d\}$ be an orthonormal basis of $\mathscr{G}$. Then $\{\beta_n e_i, \, n \geq 1, \, i = 1, \ldots, d\}$ be an orthonormal basis of $H(\mathscr{G})$. Let $U_t(u) = \int_0^u V_t(s) \, \mathrm{d}s$. Then (3.15) can be rewritten as

$$\langle V_t(1), c_n'(1) e_i \rangle_{\mathscr{G}} = \int_0^1 \langle \dot{U}_t(s), \beta_n'(s) e_i \rangle_{\mathscr{G}} \, \mathrm{d}s = \langle U_t, \beta_n e_i \rangle_{H(\mathbb{R})}.$$

It follows that

$$U_t(s) = \sum_{n \geq 1} \sum_{i=1}^{d} \langle V_t(1), c_n'(1) e_i \rangle_{\mathscr{G}} \beta_n(s) e_i. \tag{3.17}$$

We have shown in the proof of Proposition 3.3 that $V_t(1)$ is measurable with respect to γ_1 if $V_t(1)$ is continuous at t. So for $t \notin \Omega$, $U_t(s)$ is also measurable with respect to γ_1 for each $s \in [0,1]$, so does $V_t(s)$. Then by the definition (3.12),

$$d_s v_t(s) = v_t(s) dV_t(s), \quad v_t(1) = e.$$

Therefore for each $t \in [0,1] \backslash \Omega$, $v_t(s)$ is a measurable mapping of γ_1 for each $s \in [0,1]$. Then we obtain the measurability of $\gamma_1 \mapsto \gamma_2(t) = \gamma_1(t) v_t(0)^{-1}$ for $t \in [0,1] \backslash \Omega$.

Take a subdivision $\mathcal{P} = \{0 < 1/N < \cdots < (N-1)/N < 1\}$ of [0,1]. We can take N large enough so that for each $i = 0, \ldots, N-1$, $\gamma_2(i/N)$ and $\gamma_2((i+1)/N)$ are not in the cut locus of each other. Define a continuous curve $\gamma^{(N)}(t) = \gamma_2(t)$ for $t \in \mathcal{P}$ and connect $\gamma^{(N)}(i/N)$ with $\gamma^{(N)}((i+1)/N)$ by the unique minimizing geodesic. $\gamma^{(N)}$ is continuous, and $\gamma_1 \mapsto \gamma^{(N)}$ is measurable due to the measurability of the solution of geodesic equation. Letting N tend to $+\infty$, $\gamma^{(N)}$ converges uniformly to γ_2, so $\gamma_1 \mapsto \gamma_2 = \mathscr{T}(\gamma_1)$ is also measurable. Therefore, we have shown the measurability of the map $\mathscr{T}$, which shows the existence and uniqueness of optimal transport map in Theorem 1.1.

To complete the proof of this theorem, it remains to prove the explicit expression of the optimal transport map $\mathscr{T}$. As in Proposition 3.3, for each $t \in [0,1]$, there exists a constant geodesic $(v_t(s))_{s \in [0,1]}$ on G connecting $\gamma_2(t)^{-1} \gamma_1(t)$ to e. For any given continuously differentiable function $c : [0,1] \to \mathbb{R}$ with $c(0) = c(1) = 0$, define $\hat{v}_{t,\varepsilon}(s) = v_t(s)e^{c(s)\varepsilon h(t)}$ for $h \in H$ and $\varepsilon \in \mathbb{R}$. Then $\hat{v}_{t,\varepsilon}(0) = v_t(0)$, $\hat{v}_{t,\varepsilon}(1) = v_t(1)$. Thus the function $\varepsilon \mapsto L(\hat{v}_{t,\varepsilon})^2$ attains its minimum value at $\varepsilon = 0$, which yields that

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\Big|_{\varepsilon=0} L(\hat{v}_{t,\varepsilon})^2 = \int_0^1 \langle v_t^{-1}(s)\dot{v}_t(s), c'(s)h(t)\rangle_{\mathscr{G}} \mathrm{d}s = 0.$$

Taking $c(s) = \sin(k\pi s)$ for $k \in \mathbb{Z}$ and $h(t) = a \in \mathscr{G}$ in previous equation, we obtain

$$\int_0^1 \langle v_t^{-1}(s)\dot{v}_t(s), \cos(k\pi s)a\rangle_{\mathscr{G}} \mathrm{d}s = 0. \tag{3.18}$$

The arbitrariness of $k \in \mathbb{Z}$ and $a \in \mathscr{G}$ means that $s \mapsto v_t^{-1}(s)\dot{v}_t(s)$ is a constant function over $[0,1]$. According to (3.12) and (3.13), if ϕ is differentiable at γ_1, then

$$V_t(1) = \int_0^1 v_t^{-1}(s)\dot{v}_t(s)\mathrm{d}s = v_t^{-1}(0)\dot{v}_t(0),$$

and

$$\int_s^1 V_t(1)\mathrm{d}t = -\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}s}\big(\nabla\phi(\gamma_1)\big)(s). \tag{3.19}$$

Since $t \mapsto V_t(1)$ is continuous on $[0,1]\backslash\Omega$ as shown in Lemma 3.3, (3.19) yields that

$$V_t(1) = \frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}t^2}\big(\nabla\phi(\gamma_1)\big)(t), \quad t \in [0,1]\backslash\Omega. \tag{3.20}$$

As $s \mapsto \gamma_2(t)v_t(s)$ is a geodesic connecting $\gamma_1(t)$ to $\gamma_2(t)$, it can be expressed in terms of geodesic exponential map as

$$\gamma_2(t) = \exp_{\gamma_1(t)}\Big(\ell_{\gamma_1(t)}v_t(0)^{-1}\dot{v}_t(0)\Big) = \exp_{\gamma_1(t)}\Big(\frac{1}{2}\ell_{\gamma_1(t)}\frac{\mathrm{d}^2}{\mathrm{d}t^2}\big(\nabla\phi(\gamma_1)\big)(t)\Big). \tag{3.21}$$

By Lemma 3.2, ϕ is in $\mathbf{D}_1^2(\mu)$ and is μ-almost surely differentiable. The expression (3.21) means that for μ-a.e. $\gamma \in \mathcal{P}(G)$

$$\mathscr{T}(\gamma)(t) = \exp_{\gamma(t)}\Big(\frac{1}{2}\ell_{\gamma(t)}\frac{\mathrm{d}^2}{\mathrm{d}t^2}\big(\nabla\phi(\gamma)\big)(t)\Big), \quad t \in [0,1]\backslash\Omega. \tag{3.22}$$

We have completed the proof till now. ∎

4 Proof of main result: general case $p > 1$, $p \neq 2$

Now we shall prove Theorem 1.1 for general $p > 1$, $p \neq 2$. Recall that for two probability measures ν and σ on $\mathcal{P}(G)$, define the L^p-Wasserstein distance between them by:

$$W_p(\nu,\sigma) = \inf_{\pi}\Big\{\int_{\mathcal{P}(G)\times\mathcal{P}(G)} d_{L^2}(\gamma_1,\gamma_2)^p\,\pi(\mathrm{d}\gamma_1,\mathrm{d}\gamma_2)\Big\}^{1/p}, \tag{4.1}$$

where the infimum is taken over $\mathscr{C}(\nu,\sigma)$. The difficulty in the case $p>1$ and $p\neq 2$ is to prove the uniqueness of γ_2 by the equation

$$\phi(\gamma_1)+\psi(\gamma_2)=d_{L^2}(\gamma_1,\gamma_2)^p.$$

We get around this difficulty by using a more delicate variational method than the method used in the proof of Proposition 3.3.

Proof of Theorem 1.1 for $p>1$ and $p\neq 2$. According to Theorem 3.1, there exists a couple of functions ϕ and ψ on $\mathcal{P}(G)$ such that $\phi=\psi^c$, where the function $c(\gamma_1,\gamma_2)=d_{L^2}(\gamma_1,\gamma_2)^p$ now. The boundedness of d_{L^2} yields easily that $\phi=\psi^c$ is d_{L^2}-Lipschitz continuous, and hence belongs to $\mathbf{D}_1^2(\mu)$ thanks to Lemma 3.2.

To prove Theorem 1.1 for $p>1$, we can get along with the same lines as the proof for $p=2$. We omit similar steps in the argument, and only prove the main different part, which is to prove that: if it holds

$$\phi(\gamma_1)+\psi(\gamma_2)=d_{L^2}(\gamma_1,\gamma_2)^p,$$

where $p>1$, then γ_2 is uniquely determined by γ_1 and ϕ. In fact, let $v_t:[0,1]\to G$ be a constant speed geodesic such that $v_t(0)=\gamma_2(t)^{-1}\gamma_1(t)$, $v_t(1)=e$ and $L(v_t)^2=\rho(\gamma_1(t),\gamma_2(t))^2$. Using the same variation as in the argument of Proposition 3.3 again, we can obtain

$$\langle\nabla\phi(\gamma_1),h\rangle_H=-p\,d_{L^2}(\gamma_1,\gamma_2)^{p-2}\int_0^1\langle V_t(1),h(t)\rangle_{\mathscr{G}}\,\mathrm{d}t \tag{4.2}$$

instead of formula (3.13). This yields that $d_{L^2}(\gamma_1,\gamma_2)^{p-2}V_t(1)$ is uniquely determined by $\nabla\phi(\gamma_1)$ for almost every $t\in[0,1]$. Using the same variation as in the argument of Lemma 3.4, we get formula (3.15) again

$$\langle V_t(1),c'(1)a\rangle_{\mathscr{G}}=\int_0^1\langle V_t(s),c''(s)a\rangle_{\mathscr{G}}\,\mathrm{d}s,$$

for any $c\in C^2([0,1],\mathbb{R})$ with $c(0)=c(1)=0$, any $a\in\mathscr{G}$, and each $t\in[0,1]$. Taking $a=d_{L^2}(\gamma_1,\gamma_2)^{p-2}b$ for $b\in\mathscr{G}$, we get

$$\langle d_{L^2}(\gamma_1,\gamma_2)^{p-2}V_t(1),c'(1)b\rangle_{\mathscr{G}}=\int_0^1\langle d_{L^2}(\gamma_1,\gamma_2)^{p-2}V_t(s),c''(s)b\rangle_{\mathscr{G}}\,\mathrm{d}s. \tag{4.3}$$

Assume $\tilde\gamma_2\in\mathcal{P}(G)$ such that

$$\phi(\gamma_1)+\psi(\tilde\gamma_2)=d_{L^2}(\gamma_1,\gamma_2)^p.$$

Let $\tilde v_t:[0,1]\to G$ be a minimizing geodesic such that $\tilde v_t(1)=e$ and $\tilde v_t(0)=\tilde\gamma_2(t)^{-1}\gamma_1(t)$. Analogously, define $\tilde V_t(u)=\int_0^u\tilde v_t(s)^{-1}\dot{\tilde v}_t(s)\,\mathrm{d}s$ and it holds that

$$\langle d_{L^2}(\gamma_1,\tilde\gamma_2)^{p-2}\tilde V_t(1),c'(1)b\rangle_{\mathscr{G}}=\int_0^1\langle d_{L^2}(\gamma_1,\tilde\gamma_2)^{p-2}\tilde V_t(s),c''(s)b\rangle_{\mathscr{G}}\,\mathrm{d}s. \tag{4.4}$$

Since $d_{L^2}(\gamma_1, \tilde{\gamma}_2)^{p-2}\tilde{V}_t(1)$ is also determined by $\nabla\phi(\gamma_1)$ for almost everywhere $t \in [0,1]$, there exists a subset $\bar{\Omega} \subset [0,1]$ with full Lebesgue measure in $[0,1]$ such that

$$d_{L^2}(\gamma_1, \gamma_2)^{p-2}V_t(1) = d_{L^2}(\gamma_1, \tilde{\gamma}_2)^{p-2}\tilde{V}_t(1), \quad \forall\, t \in \bar{\Omega}. \tag{4.5}$$

Due to the denseness of functions in the form $s \mapsto c''(s)b$ in $L^2(\mu; \mathscr{G})$, and the continuity of $s \mapsto V_t(s)$ and $s \mapsto \tilde{V}_t(s)$, we get

$$d_{L^2}(\gamma_1, \gamma_2)^{p-2}V_t(s) = d_{L^2}(\gamma_1, \tilde{\gamma}_2)^{p-2}\tilde{V}_t(s), \quad \forall\, s \in [0,1],\ t \in \bar{\Omega}. \tag{4.6}$$

It follows then

$$d_{L^2}(\gamma_1, \gamma_2)^{p-2}\dot{V}_t(s) = d_{L^2}(\gamma_1, \tilde{\gamma}_2)^{p-2}\dot{\tilde{V}}_t(s), \quad \text{for a.e. } s \in [0,1],\ t \in \bar{\Omega},$$

where dot $\cdot$ denotes the derivative relative to s. Since $v_t(s)$ and $\tilde{v}_t(s)$ are both minimizing geodesics, integrating both sides of previous equation over s from 0 to 1 yields that

$$d_{L^2}(\gamma_1, \gamma_2)^{p-2}\rho(\gamma_1(t), \gamma_2(t)) = d_{L^2}(\gamma_1, \tilde{\gamma}_2)^{p-2}\rho(\gamma_1(t), \tilde{\gamma}_2(t)), \quad \forall\, t \in \bar{\Omega}. \tag{4.7}$$

Then, integrating the square of both sides over t from 0 to 1 yields

$$d_{L^2}(\gamma_1, \gamma_2)^{2(p-1)} = d_{L^2}(\gamma_1, \tilde{\gamma}_2)^{2(p-1)}. \tag{4.8}$$

Combining this with (4.5), we obtain

$$V_t(1) = \tilde{V}_t(1), \quad \forall\, t \in \bar{\Omega}.$$

Using Lemma 3.4, we have $\gamma_2(t) = \tilde{\gamma}_2(t)$ for $t \in \bar{\Omega}$. The continuity of $\gamma_2(t)$ and $\tilde{\gamma}_2(t)$ yields $\gamma_2(t) \equiv \tilde{\gamma}_2(t)$ for $t \in [0,1]$, and hence $\gamma_2 \in \mathcal{P}(G)$ is uniquely determined. ∎

Proof of Theorem 1.3. We only prove the assertion of this theorem for path groups, and the corresponding assertion for loop groups can be proved in the same way.

For ν_0, $\nu_1 \in \mathscr{P}_0(\mathcal{P}(G))$ with ν_0 being absolutely continuous w.r.t. the Wiener measure μ, according to Theorem 1.1, there exists a unique optimal map $\mathscr{T} : \mathcal{P}(G) \to \mathcal{P}(G)$ such that

$$\pi_0 := (id \times \mathscr{T})_*\nu_0$$

attains the L^2-Wasserstein distance between ν_0 and ν_1, i.e.

$$W_2(\nu_0, \nu_1)^2 = \int_{\mathcal{P}(G)\times\mathcal{P}(G)} d_{L^2}(\gamma_1, \gamma_2)^2\pi_0(\mathrm{d}\gamma_1, \gamma_2).$$

Let ϕ and ψ be the Kantorovich potentials, then

$$W_2(\nu_0, \nu_1)^2 = \int_{\mathcal{P}(G)\times\mathcal{P}(G)} \big(\phi(\gamma_1) + \psi(\gamma_2)\big)\pi_0(\mathrm{d}\gamma_1, \mathrm{d}\gamma_2).$$

Then the support of π_0 is clearly located in the set

$$A = \{(\gamma_1, \gamma_2) \in \mathcal{P}(G) \times \mathcal{P}(G);\ \phi(\gamma_1) + \psi(\gamma_2) = d_{L^2}(\gamma_1, \gamma_2)^2\}.$$

For γ_1, γ_2 satisfying

$$\phi(\gamma_1) + \psi(\gamma_2) = d_{L^2}(\gamma_1, \gamma_2)^2,$$

in the proof of Proposition 3.3, we have shown that for each $t \in [0,1]$ there exists a constant speed geodesic $v_t : [0,1] \to G$ such that $v_t(0) = \gamma_2(t)^{-1}\gamma_1(t)$, $v_t(1) = e$, and

$$L(v_t)^2 = \int_0^1 \left|v_t^{-1}(s)\frac{\mathrm{d}}{\mathrm{d}s}v_t(s)\right|_{\mathscr{G}}^2 \mathrm{d}s = \rho(\gamma_1(t), \gamma_2(t))^2.$$

Set $u_t(s) = \gamma_2(t)v_t(s)$ for $s \in [0,1]$, then $u_t(0) = \gamma_1(t)$ and $u_t(1) = \gamma_2(t)$. For any given $\lambda \in [0,1]$, let $u_\cdot^\lambda$ be in $\mathcal{P}(G)$ defined by $u_t^\lambda = u_t(\lambda)$ for $t \in [0,1]$. The distance between γ_1 and u^λ is

$$d_{L^2}(\gamma_1, u^\lambda) = \lambda d_{L^2}(\gamma_1, \gamma_2). \tag{4.9}$$

Indeed, the curves $s \mapsto u_t(\lambda s)$ and $s \mapsto u_t(\lambda + (1-\lambda)s)$ connect respectively $\gamma_1(t)$ to u_t^λ and u_t^λ to $\gamma_2(t)$. Then

$$\rho(\gamma_1(t), u_t^\lambda) \le \Big(\int_0^1 \lambda^2 \left|v_t^{-1}(\lambda s)\frac{\mathrm{d}}{\mathrm{d}s}v_t(\lambda s)\right|^2 \mathrm{d}s\Big)^{1/2} = \lambda\rho(\gamma_1(t), \gamma_2(t)).$$

Similarly, $\rho(u_t^\lambda, \gamma_2(t)) \le (1-\lambda)\rho(\gamma_1(t), \gamma_2(t))$. Together with the triangle inequality, we can get $\rho(\gamma_1(t), u_t^\lambda) = \lambda\rho(\gamma_1(t), \gamma_2(t))$ for all $\lambda \in [0,1]$, and further (4.9) holds. Consequently, $\lambda \mapsto u^\lambda$ is a geodesic in $(\mathcal{P}(G), d_{L^2})$ connecting γ_1 to γ_2.

Set $\Phi_\lambda(\gamma_1) = u_\cdot^\lambda$ and $\nu_\lambda = (\Phi_\lambda)_*\nu_0$ for $\lambda \in [0,1]$. Then

$$W_2(\nu_0, \nu_\lambda) \le \Big(\int_{\mathcal{P}(G)\times\mathcal{P}(G)} d_{L^2}(\gamma_1, u^\lambda)^2 \mathrm{d}\nu_0(\gamma_1)\Big)^{1/2} = \lambda W_2(\nu_0, \nu_1),$$

and

$$W_2(\nu_\lambda, \nu_1) \le \Big(\int_{\mathcal{P}(G)\times\mathcal{P}(G)} d_{L^2}(u^\lambda, \gamma_2)^2 \mathrm{d}\nu_0(\gamma_1)\Big)^{1/2} = (1-\lambda)W_2(\nu_0, \nu_1).$$

By the triangle inequality, it holds

$$W_2(\nu_0, \nu_\lambda) = \lambda(\nu_0, \nu_1), \quad W_2(\nu_\lambda, \nu_1) = (1-\lambda)W_2(\nu_0, \nu_1).$$

Hence, ν_λ for $\lambda \in [0,1]$ is a geodesic in $\mathscr{P}(\mathcal{P}(G))$ w.r.t. the Wasserstein distance W_2 connecting ν_0 to ν_1. The proof is complete. ∎

5 Optimal transport map on loop groups

Let G be a connected compact Lie group and its Lie algebra $\mathscr{G}$ is endowed with an Ad-invariant metric $\langle\, ,\rangle_{\mathscr{G}}$. Let

$$\mathcal{L}_e G = \{\ell : [0,1] \to G \,\text{continuous};\ \ell(0) = \ell(1) = e\}.$$

The product in $\mathcal{L}_e G$ is defined pointwisely by $(\ell_1 \cdot \ell_2)(\theta) = \ell_1(\theta) \cdot \ell_2(\theta)$, $\theta \in [0,1]$. With the uniform topology

$$d_\infty(\ell_1, \ell_2) = \sup_{\theta \in [0,1]} \rho(\ell_1(\theta), \ell_2(\theta)),$$

where ρ is Riemannian distance on G, $\mathcal{L}_e G$ becomes a topological group. Recall that

$$H(\mathscr{G}) = \Big\{ h : [0,1] \to \mathscr{G};\ h(0) = 0,\ |h|_H^2 = \int_0^1 \big|\dot{h}(t)\big|_{\mathscr{G}}^2 \,\mathrm{d}t < +\infty \Big\}.$$

Let

$$H_0(\mathscr{G}) = \big\{ h \in H(\mathscr{G});\ h(0) = h(1) = 0 \big\}.$$

For $h \in H_0(\mathscr{G})$, set $|h|_{H_0} = \Big(\int_0^1 |\dot{h}(\theta)|_{\mathscr{G}}^2 \,\mathrm{d}\theta \Big)^{1/2}$. It has been shown in [21] that there is a Brownian motion $(g(t))$ on $\mathcal{L}_e G$. In order to be consistent in notations as convention, in the sequel, we shall fix ν to be the law of Brownian motion $g(1)$ on $\mathcal{L}_e G$, which is called heat kernel measure. Let μ_0 denote the pinned Wiener measure on $\mathcal{L}_e G$. Due to [11], ν is absolutely continuous with respect to the pinned Wiener measure μ_0. According to [2], μ_0 is also absolutely continuous with respect to heat kernel measure ν.

For a cylindrical function $F : \mathcal{L}_e G \to \mathbb{R}$ in the form

$$F(\ell) = f(\ell(\theta_1), \ldots, \ell(\theta_n)), \quad f \in C^\infty(G^n),$$

and $h \in H_0(\mathscr{G})$, define

$$(D_h F)(\ell) = \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\Big|_{\varepsilon=0} F(\ell e^{\varepsilon h}) = \sum_{i=1}^n \langle \partial_i f, \ell(\theta_i) h(\theta_i) \rangle_{T_{\ell(\theta_i)} G},$$

where $\partial_i f$ denotes the ith partial derivative. The gradient operator $\nabla^{\mathcal{L}}$ on $\mathcal{L}_e G$ is defined as

$$\big(\nabla^{\mathcal{L}} F\big)(\ell) = \sum_{i=1}^n \ell^{-1}(\theta_i)(\partial_i f) G(\theta_i, \cdot),$$

where $G(\theta_i, \theta) := \theta_i \wedge \theta - \theta_i \theta$. Consider

$$\mathscr{E}(F, F) := \int_{\mathcal{L}_e G} |\nabla^{\mathcal{L}} F|_{H_0}^2 \,\mathrm{d}\nu.$$

Then $\mathscr{E}$ defined on the set of cylindrical functions is closable, and let $\mathbf{D}_1^2(\nu)$ be the domain of the associated Dirichlet form.

Now, we introduce several distance on $\mathcal{L}_eG$. Firstly, the L^2-distance is defined by:

$$d_{L^2}(\ell_1, \ell_2) = \Big(\int_0^1 \rho(\ell_1(\theta), \ell_2(\theta))^2 \, \mathrm{d}\theta \Big)^{1/2}, \quad \ell_1, \ell_2 \in \mathcal{L}_eG. \tag{5.1}$$

Secondly, we shall recall the definition of Riemannian distance on $\mathcal{L}_eG$. In [15], it has shown the existence and uniqueness of optimal transport map for the Monge-Kantorovich problem with the Wasserstein distance defined by the square of Riemannian distance on $\mathcal{L}_eG$.

A continuous curve $\gamma : [0, 1] \to \mathcal{L}_eG$ is said to be admissible if there exists $z \in H(H_0)$ such that

$$\frac{\partial}{\partial t}\gamma(t, \theta) = \gamma(t, \theta)\frac{\partial}{\partial t}z(t, \theta), \quad \gamma(0, \theta) = e. \tag{5.2}$$

Here

$$H(H_0) = \Big\{ z : [0, 1] \to H_0(\mathscr{G});\ z_t = \int_0^t \frac{\partial}{\partial s}z(s)\, \mathrm{d}s, \|z\|^2 := \int_0^1 |\frac{\partial}{\partial s}z(s)|_{H_0}^2 \, \mathrm{d}s < +\infty \Big\}.$$

For a continuous curve γ on $\mathcal{L}_eG$, if it is admissible, its length is defined by

$$L(\gamma) = \Big(\int_0^1 |\frac{\partial}{\partial s}z(s)|_{H_0}^2 \, \mathrm{d}s \Big)^{1/2};$$

otherwise, its length $L(\gamma) = +\infty$. The Riemannian distance d_L on $\mathcal{L}_eG$ is defined by

$$d_L(\ell_1, \ell_2) = \inf \big\{ L(\gamma);\ \gamma(0) = \ell_1,\ \gamma(1) = \ell_2 \big\}, \tag{5.3}$$

where γ runs over the set of all continuous curves on $\mathcal{L}_eG$. It is clear that d_L is left invariant: $d_L(\ell\ell_1, \ell\ell_2) = d_L(\ell_1, \ell_2)$, ℓ, ℓ_1, $\ell_2 \in \mathcal{L}_eG$. It has been shown in [26, Proposition 3.4] that for ℓ_1, $\ell_2 \in \mathcal{L}_eG$, $d_{\mathcal{P}}(\ell_1, \ell_2) \le d_L(\ell_1, \ell_2)$. Therefore, it holds

$$d_{L^2}(\ell_1, \ell_2) \le d_\infty(\ell_1, \ell_2) \le d_{\mathcal{P}}(\ell_1, \ell_2) \le d_L(\ell_1, \ell_2). \tag{5.4}$$

According to the Rademacher's theorem [26, Theorem 1.5], we get

Lemma 5.1 *Every d_L-Lipschitz (hence, d_{L^2}-Lipschitz) continuous function F is in $\mathbf{D}_1^2(\nu)$.*

After these preparation, we are in a position to state our results. The results in Theorem 1.2 are parts of the results in the following two theorems.

Theorem 5.2 *For every probability measures σ_1 and σ_2 on $\mathcal{L}_eG$. Assume σ_1 is absolutely continuous with respect to the heat kernel measure ν on $\mathcal{L}_eG$. Then for each $p>1$, there exists a unique measurable map $\mathscr{T}_p : \mathcal{L}_eG \to \mathcal{L}_eG$ such that it pushes σ_1 forward to σ_2 and*

$$W_{p,d_{L^2}}(\sigma_1,\sigma_2)^p = \int_{\mathcal{L}_eG} d_{L^2}(\ell, \mathscr{T}_p(\ell))^p\,\sigma_1(\mathrm{d}\ell),$$

where

$$W_{p,d_{L^2}}(\sigma_1,\sigma_2)^p := \inf\Big\{\int_{\mathcal{L}_eG\times\mathcal{L}_eG} d_{L^2}(\ell_1,\ell_2)^p\,\pi(\mathrm{d}\ell_1,\mathrm{d}\ell_2)\Big\},$$

where the infimum runs over the set of all probability measures on $\mathcal{L}_eG\times\mathcal{L}_eG$ with marginals σ_1 and σ_2 respectively.

Proof.(Sketched) The proof of this theorem gets along the same lines as the proof of Theorem 1.1. First, Theorem 3.1 guarantees the existence of the Kantorovich potential ϕ and ψ such that

$$\phi(\ell) = \psi^c(\ell) := \inf_{\ell'\in\mathcal{L}_eG}\big\{d_{L^2}(\ell,\ell')^p - \psi(\ell')\big\}.$$

Then ϕ is d_{L^2}-Lipschitz continuous. By the Rademacher's theorem, Lemma 5.1, ϕ belongs to $\mathbf{D}_1^2(\nu)$. Then using the variational method to show the uniqueness of $\ell_2\in\mathcal{L}_eG$ such that

$$\phi(\ell_1)+\psi(\ell_2) = d_{L^2}(\ell_1,\ell_2)^p,$$

if ϕ is differentiable at ℓ_1. This progress is completely similar to the proof of Proposition 3.3 for case $p=2$ and the discussion in section 3 for case $p>1$. In this step, the different point is just to replace $h\in H(\mathscr{G})$ with $h\in H_0(\mathscr{G})$. Then the desired map is the map defined by $\mathscr{T}_p(\ell_1)=\ell_2$ such that above equation holds. The measurability of this map comes from the construction as in the argument of Theorem 1.1. ∎

Theorem 5.3 *On $\mathcal{L}_eG$, for each $p>1$, there exists a unique measurable map $\mathscr{T}_p : \mathcal{L}_eG \to \mathcal{L}_eG$ such that $\mathscr{T}_p$ pushes heat kernel measure ν forward to pinned Wiener measure μ_0 such that*

$$W_{p,d_{L^2}}(\nu,\mu_0)^p = \int_{\mathcal{L}_eG} d_{L^2}(\ell,\mathscr{T}_p(\ell))^p\,\mathrm{d}\nu(\ell). \tag{5.5}$$

Moreover, $\mathscr{T}_p$ is ν-a.e. reversible, and its inverse $\mathscr{T}_p^{-1}$ pushes μ_0 forward to ν.

Proof. Noting that μ_0 and ν is mutually absolutely continuous with respect to each other, applying Theorem 5.2 yields that there exists a measurable map $\mathscr{T}_p : \mathcal{L}_eG\to\mathcal{L}_eG$ which pushes ν forward to μ_0 and a measurable map $\mathscr{S}_p : \mathcal{L}_eG\to\mathcal{L}_eG$ which pushes μ_0 forward to ν. Furthermore,

$$W_{p,d_{L^2}}(\nu,\mu_0)^p = \int_{\mathcal{L}_eG} d_{L^2}(\ell,\mathscr{T}_p(\ell))\,\mathrm{d}\nu(\ell) = \int_{\mathcal{L}_eG} d_{L^2}(\mathscr{S}_p(\ell),\ell)^p\,\mathrm{d}\mu_0(\ell). \tag{5.6}$$

For any measurable function F on $\mathcal{L}_eG$, we have

$$\int_{\mathcal{L}_eG} F(\ell)\,\mathrm{d}\nu(\ell) = \int_{\mathcal{L}_eG} F(\mathscr{S}_p(\ell))\,\mathrm{d}\mu_0(\ell) = \int_{\mathcal{L}_eG} F(\mathscr{S}_p(\mathscr{T}_p(\ell)))\,\mathrm{d}\nu(\ell). \tag{5.7}$$

Therefore,

$$\mathscr{S}_p \circ \mathscr{T}_p = id, \quad \nu\text{-}a.e.,$$

where *id* denotes the identity map. We conclude the argument immediately. ∎

Acknowledgement This work is supported by NSFC (No. 11431014, 11771327). The author is grateful to Professor S.Z. Fang for the invitation of visiting to University of Bourgogne in July 2017 and for valuable discussion on this topic.

References

[1] S. Aida, *Sobolev spaces over loop groups*, J. Funct. Anal., **127** (1995), 155-172.

[2] S. Aida and B. Driver, *Equivalence of heat kernel measure and pinned Wiener measure on loop groups*, C. R. Math. Acad. Sci. Paris. I, **331**(9) (2000), 709-712.

[3] L. Ambrosio *Optimal transport maps in Monge-Kantorovich problem*, Lecture Note in ICM, Vol. III, 2002, 131-140.

[4] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, Lectures in Mathematics ETH Zurich, Birkhauser Verlag, Basel, 2005.

[5] L. Ambrosio, S. Rigot, *Optimal mass transportation in the Heisenberg group*, J. Funct. Analysis, Vol. **208** (2004), 261-301.

[6] Y. Brenier, *Polar factorization and monotone rearrangement of vector valued functions*, Comm. Pure Appl. Math., **44** (1991), 375-417.

[7] J. Cheeger, D.G. Ebin, *Comparison theorems in Riemannian geometry*, North-Holland Mathematical Library, Vol. 9, North-Holland, Amsterdam, 1975.

[8] D. Cordero-Erausquin, R.J. McCann and M. Schmuckenschläger, *A Riemannian interpolation inequality à la Borell Brascamp and Lieb*, Invent. math. **146** (2001), 219-257.

[9] B. Driver, *Integration by parts and quasi-invariance for heat kernel measure on loop groups*, J. Funct. Analysis, **149**, (1997), 470-547.

[10] B. Driver and T. Lohrenz, *Logarithmic Sobolev inequalities for pinned loop groups*, J. Funct. Analysis, **140** (1996), No. 2 381-448.

[11] B. Driver and V.K. Srimurthy, *Absolute continuity of heat kernel measure with pinned Wiener measure on loop groups*, Ann. Probab. **29** (2001), 691-723.

[12] M. Erbar, K. Kuwada, K. Sturm, *On the equivalence of the entropic curvature-dimension condition and Bochner's inequality on metric measure spaces*, Invent. math. **201** (2015), 993-1071.

[13] L.C. Evans and W. Gangbo, *Differential equations methods for the Monge-Kantorovich mass transfer problem*, Mem. Amer. Math. Soc. **137** (1999), 1-66.

[14] S. Fang, *Introduction of Malliavin calculus*, Tsinghua University press and Springer, 2005.

[15] S. Fang and J. Shao, *Optimal transport maps for Monge-Kantorovich problem on loop groups*, J. Funct. Anal. **248** (2007), 225-257.

[16] D. Feyel and A.S. Üstünel, *Monge-Kantorovitch measure transportation and Monge-Ampère equation on Wiener space*, Probab. Theory Related Fields, **128** (2004), 347-385.

[17] W. Gangbo and R.J. McCann, *The geometry of optimal transportation*, Acta Math., **177** (1996), 113-161.

[18] R. Jordan, D. Kinderlehrer and F. Otto, *The variational formulation of the Fokker- Planck equation*, SIAM J. Math. Anal. **29** (1998), 1-17.

[19] M. Knott and C.S. Smith, *On the optimal mapping of distributions*, J. Optim. Theory Appl. **43**(1) (1984), 39-49.

[20] J. Lott and C. Villani, *Ricci curvature for metric-measure spaces via optimal transport*, Ann. of Math. **169**(3) (2009), 903-991.

[21] P. Malliavin, *Hypoellipticity in infinite dimension.* In Diffusion processes and related problem in Analysis (M. Pinsky ed.), p.17-33. Birkhaüser, Boston.

[22] R.J. McCann, *Polar factorization of maps on Riemannian manifolds*, Geom. Funct. Anal. **11** (3) (2001), 589-608.

[23] S.T. Ohta, *Finsler interpolation inequalities*, Calc. Var. Partial Differ. Equ. **36** (2009), 211-249.

[24] A. Petrunin, *Alexandrov meets Lott-Villani-Sturm*, Münst. J. Math. **4** (2011), 53-64.

[25] J.F. Price, *Lie groups and compact Lie groups*, London Math. Society Lecture note series 25, 1977.

[26] J. Shao, *Hamilton-Jacobi semi-groups in infinite dimensional spaces*, Bull. Sci. Math. **130** (2006), 720-738.

[27] K.T. Sturm, *On the geometry of metric measure spaces I and II*, Acta Math. **196** (2006), 65-131.

[28] C. Villani, *Optimal transport, old and new*, Grundlehren der mathematischen Wissenschaften, vol. 338, Springer Berlin Heidelberg, 2009.

[29] M.-K. von Renesse and K.T. Sturm, *Transport inequalities, gradient estimates, entropy and Ricci curvature*, Comm. Pure Appl. Math. **58** (2005), 923-940.

[30] F.W. Warner, *Foundations of differentiable manifolds and Lie groups*, Graduate text in Mathematics, Vol. 94, Springer, 1983.

[31] H.-C. Zhang, X.-P. Zhu, *Ricci curvature on Alexandrov spaces and rigidity theorems*, Commun. Anal. Geom. **18**(3) (2010), 503-553.