Computational Optimal Transport and Filtering on Riemannian Manifolds

Daniel Grange^{*} Mohammad Al-Jarrah[†] Ricardo Baptista[‡] Amirhossein Taghvaei[†] Tryphon T. Georgiou^{*} Sean Phillips[§] Allen Tannenbaum^{*}

Abstract— In this paper we extend recent developments in computational optimal transport to the setting of Riemannian manifolds. In particular, we show how to learn optimal transport maps from samples that relate probability distributions defined on manifolds. Specializing these maps for sampling conditional probability distributions provides an ensemble approach for solving nonlinear filtering problems defined on such geometries. The proposed computational methodology is illustrated with examples of transport and nonlinear filtering on Lie groups, including the circle S¹, the special Euclidean group SE(2), and the special orthogonal group SO(3).

Index Terms—Optimal Transportation, Optimal Control, Nonlinear Filtering, Riemannian manifolds.

I. INTRODUCTION

THE theory of optimal transport (OT) has emerged as a powerful mathematical tool in a wide range of engineering and control applications [2]. This is largely due to the fact that it induces a natural and computationally tractable geometry on the space of probability distributions [3], [4]. The metric that the theory provides to quantify distance between distributions, the *Wasserstein metric*, gives rise to natural geodesic flows and transport maps that can be used to interpolate, average, and correspond distributions in a physically meaningful sense. For these reasons, OT has proven enabling for an ever expanding range of applications in machine learning [4]–[6], and image processing [7]–[10], besides ones in control and estimation [2], [11]–[13].

In this rapidly developing landscape of OT techniques and applications, neural networks and stochastic optimization have

Approved for public release; distribution is unlimited. Public Affairs approval # AFRL-2023-4936. The views expressed are those of the authors and do not necessarily reflect official policy or position of the Department of the Air Force, the Department of Defense or the U.S. Government.

*Department of Computer Science, Stony Brook University, NY; {daniel.grange,allen.tannenbaum}@stonybrook.edu.

[†]Department of Aeronautics and Astronautics, University of Washington, Seattle, WA; {mohd9485,amirtag}@uw.edu.

[‡]Department of Computing and Mathematical Sciences, California Insitute of Technology, Pasadena, CA; rsb@caltech.edu.

*Department of Mechanical and Aerospace Engineering, University of California, Irvine, CA; tryphon@uci.edu.

 $^{\$}$ Space Vehicles Directorate, Air Force Research Laboratory, Kirtland AFB, NM

DG and MAJ led the computational effort of the project, A. Taghvaei led the methodological effort, largely built upon [1], and the remaining authors supervised the development of the work. The research was supported in part by the NSF under grant EPCN-2318977, AFOSR under FA9550-23-1-0096 and ARO under W911NF-22-1-0292.

come to provide a potentially transformative framework for the development of efficient and scalable numerical algorithms [14]–[18]. The focus so far on utilizing such techniques however has been limited to applications of OT on Euclidean spaces. Yet, optimal transport can equally well be considered on manifolds, that are especially relevant in control and robotic applications. A manifold structure is naturally imposed by geometric constraints, as in attitude estimation of aircraft [19], [20], localization of mobile robots [21], [22], and visual tracking of humans and objects [23], [24].

Thus, one of the goals of the present paper is to develop a computational framework for OT in the setting of Riemannian manifolds [25], [26] with special attention to matrix Liegroups, as these encompass the majority of the motivating applications. A second goal of the paper is to use the framework for sampling conditional distributions, in order to perform nonlinear filtering on Riemannian manifolds.

Specifically, we make the following key contributions:

- We propose a sample-based computational methodology for computing OT maps on Riemannian manifolds. Our proposed methodology extends the min-max formulation of [17] by combining it with McCann's characterization of optimal transport maps [25].
- ii) We propose a sample-based and likelihood-free method to sample conditional distributions on Riemannian manifolds. In order to do so, we use the recently introduced framework of block-triangular transport maps, that is used in the context of conditional generative models [27], [28] and nonlinear filtering [29], [30].
- iii) We illustrate our proposed algorithms on several numerical examples on the circle, special Euclidean group SE(2), and the special orthogonal group SO(3).

II. PROBLEM FORMULATION AND BACKGROUND

Let \mathcal{M} be a smooth connected manifold without boundary that is equipped with the Riemannian metric $\langle \cdot, \cdot \rangle_g$. Let d(z, z')denote the geodesic distance for any $z, z' \in \mathcal{M}$. We are interested in solving the following two problems.

Optimal control: This is the problem to steer a random process X(t), taking values in \mathcal{M} , from an initial probability distribution P to a terminal probability distribution Q. It is formulated as follows:

$$\min_{u} \mathbb{E}\left[\int_{0}^{1} \|u(t)\|_{g}^{2} dt\right],$$
s.t. $\dot{X}(t) = u(t), \quad X(0) \sim P, \quad X(1) \sim Q,$
(1)

where the control input $u(t) \in T_{X(t)}\mathcal{M}$ for all $t \in [0, 1]$, and the control cost is the square of the Riemannian norm $||u||_g := \sqrt{\langle u, u \rangle_g}$. In practice, such problems arise when controlling an ensemble of agents or a swarm of robots.

Optimal filtering: The second problem we are interested is to compute the conditional distribution of a hidden random variable $X \in \mathcal{M}$ given an observed random variable $Y \in \mathbb{R}^m$. The conditional distribution of X|Y, i.e., the posterior distribution, is given by Bayes' law as

$$P_{X|Y}(x|y) = \frac{P_{Y|X}(y|x)P_X(x)}{P_Y(y)},$$
(2)

where P_X is the prior probability distribution of X, $P_{Y|X}$ is the likelihood of observing Y given X, and $P_Y(y) = \int_{\mathcal{M}} P_{Y|X}(y|x)P_X(x)dx$ is the probability distribution of Y. Sampling the conditional distribution in (2) is an essential step in many nonlinear ensemble filtering algorithms [31]. Classic algorithms include particle filters (or sequential Monte Carlo methods), which suffer from weight degeneracy [32], [33], and Kalman-filter-type algorithms, which often fail to represent multi-modal distributions [21], [34]–[36].

In Section III we will show that, in general, both problems can be formulated as problems of OT on Riemannian manifolds. But before we proceed, we review next some key results of the theory of OT on Riemannian manifolds.

A. Background on OT on Riemannian manifolds

Given two probability distributions P and Q on \mathcal{M} , the Monge optimal transportation problem seeks a map $T: \mathcal{M} \to \mathcal{M}$ that solves the optimization problem

$$\inf_{T \in \mathcal{T}(P,Q)} \mathbb{E}_{Z \sim P}[c(Z,T(Z))],$$
(3)

where $\mathcal{T}(P,Q) := \{T : \mathcal{M} \to \mathcal{M}; T_{\#}P = Q\}$ is the set of all transport maps pushing forward P to Q, and $c : \mathcal{M} \times \mathcal{M} \to \mathbb{R}$ is a lower semi-continuous cost function that is bounded from below. To account for the challenging nonlinear constraint in (3), the Monge problem is relaxed by replacing deterministic transport maps with stochastic couplings and solving

$$\inf_{\pi \in \Pi(P,Q)} \mathbb{E}_{(Z,Z') \sim \pi}[c(Z,Z')], \tag{4}$$

instead, where $\Pi(P,Q)$ denotes the set of all joint distributions on $\mathcal{M} \times \mathcal{M}$ with marginals P and Q. This relaxation, due to Kantorovich, turns the Monge problem into a linear program, whose dual becomes

$$\sup_{(\phi,\psi)\in \operatorname{Lip}_c} \mathbb{E}_{Z\sim P}[\phi(Z)] + E_{Z'\sim Q}[\psi(Z')], \tag{5}$$

where Lip_{c} is the set of pairs of functions ϕ, ψ from $\mathcal{M} \to \mathbb{R}$,

$$\operatorname{Lip}_{c} := \{(\phi, \psi) \mid \phi(z) + \psi(z') \leq c(z, z'), \, \forall z, z' \in \mathcal{M} \}$$

Definition 2.1: Given a function $\phi : \mathcal{M} \to \mathbb{R} \cup \{\pm \infty\}$, the inf-*c* convolution is given by

$$\phi^{c}(z') = \inf_{z \in M} \left[c(z, z') - \phi(z) \right].$$

Moreover, ϕ is said to be *c*-concave if $\exists \psi$ such that $\phi = \psi^c$.

Theorem 2.2 (McCann [25]): Consider the Monge problem (3) with cost $c(z, z') = d(z, z')^2/2$, and assume that P is absolutely continuous with respect to the volume measure on \mathcal{M} . Then, there exists a unique minimizer $T(z) = \exp_z[-\nabla\phi(z)]$ where ϕ is *c*-concave and the pair (ϕ, ϕ^c) maximize the dual Kantorovich problem (5).

Remark 2.3: For the case where $\mathcal{M} = \mathbb{R}^n$ and $c(z, z') = \frac{1}{2} ||z - z'||_2^2$, the optimal map is given by $T(z) = z - \nabla \phi(z)$, where $||z||_2^2/2 - \phi(z)$ is a convex function; this is a celebrated result due to Brenier [37].

B. Computational methods for OT on manifolds

The majority of existing computational algorithms for OT on manifolds are concerned with constructing normalizing flows and c-concave potential functions [38]-[41]. For instance, [38] proposes a general method to represent c-concave functions by taking the inf-c convolution of a piecewise constant function, reference [39] introduces a set of diffeomorphisms on circles, tori, and spheres that are used as building blocks of a neural network to represent c-concave functions, and reference [42] is concerned with implementation of diffusion models on Riemannian manifolds via a variety of approaches that are based on projection, using a Liealgebra basis and coordinate vector-fields, in order to represent general vector-fields. Lastly, [41] introduces the geodesic distance layer, which generalizes the concept of linear layers to manifolds, used as an input layer to a neural net to represent vector-fields on manifolds.

In the present paper, we also use neural networks to represent transport maps on the given manifold. Our neural network architectures, along with the special details for each example that we present, are explained in Section IV.

III. SOLUTION METHODOLOGY

In this section, we present the OT formulation of the two problems, optimal control and filtering, along with a stochastic optimization formulation for their numerical solution.

A. Solution to the optimal control problem

The optimal control problem (1) is precisely the Benamou-Brenier formulation of the optimal transport problem on a Riemannian manifold [43]. The optimal cost coincides with the optimal cost of the Kantorovich problem (4). By the Cauchy-Schwarz inequality we have

$$\int_0^1 \|u(t)\|_g^2 \, dt \ge (\int_0^1 \|u(t)\|_g \, dt)^2 = d(X(0), X(1))^2,$$

with equality when $||u(t)||_g$ is constant and the trajectory is a geodesic. Upon taking expectation of both sides and applying the initial and terminal constraints, we have that

$$\mathbb{E}\left[\int_{0}^{1} \|u(t)\|_{g}^{2} dt\right] \ge \mathbb{E}_{(X(0),X(1))\sim\pi} \left[d(X(0),X(1))^{2}\right]$$

for any $\pi \in \Pi(P,Q)$. Lastly, by infimizing both sides we obtain a lower-bound for the optimal control cost, with equality when u(t) steers X(t) along geodesics that connect X(0) to T(X(0)) where $T(x) = \exp_x[-\nabla\phi(x)]$ is the optimal transport map from P to Q. As a result, the solution to the optimal control problem is given by

$$X(t) = \exp_{X(0)}[-t\nabla\phi(X(0))], \text{ for } t \in [0,1]$$
 (6)

where ϕ solves the dual Kantorovich problem (5). The exponential exemplifies that the optimal trajectories are geodesics.

In order to numerically solve the Kantorovich dual problem (5), we use the result of Theorem 2.2 to replace ψ with ϕ^c . Using the definition of ϕ^c ,

$$\mathbb{E}_{Z'\sim Q}[\phi^c(Z')] = \mathbb{E}_{Z'\sim Q}[\min_z c(z, Z') - \phi(z)]$$

= $\min_U \mathbb{E}_{Z'\sim Q}[c(\exp_{Z'}[-U(Z')], Z') - \phi(\exp_{Z'}[-U(Z')])]$

where we assumed that

$$\exp_{Z'}[-U(Z')] = \arg\min_{z} c(z, Z') - \phi(z), \quad \text{a.e.}$$

for some vector-field $U : \mathcal{M} \to T\mathcal{M}$. With this assumption, we can express the Kantorovich dual problem as

$$\max_{\phi} \min_{U} \mathbb{E}_{Z \sim P}[\phi(Z)]$$

$$+ \mathbb{E}_{Z' \sim Q}[c(\exp_{Z'}(-U(Z')), Z') - \phi(\exp_{Z'}(-U(Z')))]$$
(7)

This is a max-min optimization problem with the objective function that can be approximated using samples from the two distributions P and Q. If (ϕ, U) is an optimal pair, then $T(z) = \exp_z[-\nabla\phi(z)]$ is the optimal transport map from P to Q and $T^{-1}(z') = \exp_{z'}[-U(z')]$ is the optimal transport map from Q to P.

B. Solution to the optimal filtering problem

The problem of computing conditional distributions can be formulated as an OT problem using the idea of block-triangular transport maps. In particular, if a map of the form $S : (x, y) \mapsto$ (T(x, y), y) transports the independent coupling $P_X \otimes P_Y$ to the joint distribution P_{XY} , then the map T(x, y) transports the prior P_X to the conditional distribution $P_{X|Y}(\cdot|y)$ for any value y of the observation; see Theorem 2.4 in [27] for a proof of this result. That is,

if
$$(T, \operatorname{Id})_{\#}(P_X \otimes P_Y) = P_{XY}$$

then $T(\cdot, y)_{\#}P_X = P_{X|Y}(\cdot|y)$, a.e. y .

To seek the optimal transport map T that characterizes posterior distributions, we express the Monge problem (3) with $P = P_X \otimes P_Y$, $Q = P_{XY}$, constrain the maps to have a block-triangular structure, form the Kantorovich dual, and apply the inf-c convolution representation explained above in Section III-A, which yields the max-min optimization problem

$$\max_{\phi} \min_{U} \mathbb{E}_{(X,Y)\sim P_{XY}}[\phi(X,Y)]$$

$$+ \mathbb{E}_{(X,Y)\sim P_X\otimes P_Y}[c(X,\exp_X(-T(X,Y)))$$

$$- \phi(\exp_X(-T(X,Y)),Y)].$$
(8)

Similarly to the optimization problem (7), the objective function in (8) can be approximated using samples of the pair $(X,Y) \sim P_{XY}$. If (ϕ, U) is an optimal pair, then $T(x, y) = \exp_{z}[-U(x, y)]$ is the optimal transport map from P_{X} to $P_{X|Y}(\cdot|y)$ for any value y of the observation.

IV. NUMERICAL RESULTS

In order to numerically solve the proposed optimal control (1) and optimal filtering (2) problems, we use their stochastic optimization formulations (7) and (8), respectively. Both of these problems involve optimizing over a real-valued function $\phi : \mathcal{M} \times \mathbb{R}^m \to \mathbb{R}$, and a vector-field $U : \mathcal{M} \times$ $\mathbb{R}^m \to T\mathcal{M}$ (with the convention that m = 0 for (7)). To this end, we use neural networks to represent ϕ and U. The input layer to the neural network is based on a specific coordinate representation of the manifold that is described within each example. For all examples, the middle layers are fully connected residual blocks (for these examples, we used one or two blocks of size 32). When \mathcal{M} is a *n*-dimensional Lie-group, we model the output of the vector-field as an element of the Lie-algebra, which is isomorphic to \mathbb{R}^n . We used the ADAM optimizer to solve the max-min problem with batch-size 64, learning rate 10^{-3} , 10 inner-loop minimization iterations per one maximization iteration, and a total of 10^3 to 10⁴ maximization iterations. The neural networks are trained on 10^3 to 10^5 samples from the distributions. The details of the numerical codes are available online¹.

A. Optimal transport mapping on S¹

The first example of OT on the circle S^1 uses coordinates $\theta \in [0, 2\pi)$, with the tangent space $T_{\theta}S^1 \cong \mathbb{R}^1$. For all $u \in T_{\theta}S^1$, we use the exponential map $\exp_x(u) = (x+u) \mod 2\pi$, and the geodesic distance

$$d_{S^1}(\theta, \theta') = \min(|\theta - \theta'|, 2\pi - |\theta - \theta'|), \quad \forall \theta, \theta' \in [0, 2\pi).$$

We implement our proposed numerical procedure for two choices of the marginal distribution: (1) P and Q are distributions centered at $\bar{\theta} = 0$ and π , respectively, constructed from standard Gaussians modulo² 2π ; (2) P is the uniform distribution on S^1 and Q is a mixture of two similarly projected Gaussians, centered at 0 and π . We use the map $\theta \mapsto [\cos(\theta), \sin(\theta)]$ as the input layer of the networks. The given initial and final distribution, along with the optimal trajectory $\exp_x[-tU(x)]_{\#}P$ for $t \in [0,1]$ are depicted in Figure 1. The first case highlights the periodic structure of the circle, where half of the initial distribution is transported in a clockwise direction, while the other half moves counterclockwise, as opposed a constant shift by π which is optimal in the Euclidean case. The second example highlights the ability of transporting to multi-modal distributions on the circle.

B. Optimal transport mapping on SE(2)

Next, we extend the previous example to the special Euclidean group SE(2). Elements of SE(2) are represented by the coordinate $z = (x, y, \theta) \in \mathbb{R}^2 \times [0, 2\pi)$. For all $u \in T_z SE(2) \cong \mathbb{R}^3$, we consider the exponential map

$$\exp_z(u) = (x + u_x, y + u_y, (\theta + u_\theta) \mod 2\pi),$$

and the geodesic distance

$$d_{SE(2)}^{2}(z,z')^{2} = d_{S^{1}}(\theta,\theta')^{2} + |x-x'|^{2} + |y-y'|^{2}$$

¹https://github.com/Dan-Grange/OTManifold ²We set P, Q as having density $e^{-((\theta - \bar{\theta}) \mod 2\pi)^2/2}/\sqrt{2\pi}$.



Fig. 1. OT mapping on the circle. The left panels in (a) and (b) show the given initial and final distributions. The right panels show the trajectory of optimal transportation evaluated by our numerical procedure. The curves are obtained with a kernel density estimation on samples.



Fig. 2. OT mapping on SE(2). The elements of SE(2) are depicted as arrows. Samples from the marginal distributions P and Q are shown in the left panel, and the optimal flow, obtained as a result of our numerical procedure, is depicted in the right panel.

for $z = (x, y, \theta)$ and $z' = (x', y', \theta')$. We use $z \rightarrow [x, y, \cos(\theta), \sin(\theta)]$ as the input layer to the neural networks. Figure 2 qualitatively demonstrates the capability of our proposed approach to optimally transport distributions on SE(2).

C. Optimal filtering on SO(2)

We now consider the problem of estimating the orientation $\theta \in [0, 2\pi)$ of a ground robot in a circular room as depicted in Figure 3. The robot is equipped with a sensor that measures its distance to the wall of the room along the direction that the robot is facing, i.e. $Y = h(\theta) + W$, where

$$h(\theta) = \ell \cos(\theta) + \sqrt{1 - \ell^2 \sin(\theta)^2},$$

 ℓ is the distance of the robot from the center, and $W \sim N(0, 10^{-2})$ is additive Gaussian noise. We assume a uniform prior distribution for the orientation θ and use our proposed numerical procedure to characterize the conditional distribution of the orientation θ given the observation Y. We use the input layer $(\theta, y) \mapsto (\cos(\theta), \sin(\theta), y)$ in our networks for this example. Figure 3 depicts the resulting posterior distribution. Due to the symmetric geometry of the circle, there are two angles that are consistent with an observation. As a result, the conditional distribution is bimodal, which is approximated well by our numerical procedure.

In order to test the algorithm in a dynamic setting, we let the orientation change with a constant velocity u using the model: $\theta_{t+1} = \theta_t + u + \xi_t$ where $\xi_t \sim N(0, 10^{-2})$ is the process noise. Our goal is now to compute the conditional (filtering) distribution of the orientation θ_t given the history of sensory observation Y_1, Y_2, \ldots, Y_t , generated by $Y_t = h(\theta_t) + W_t$.



Fig. 3. Optimal filtering on a circle. The left panel shows the problem setup. The right panel shows the prior, exact and estimated posterior conditional distribution of the orientation given the observed signal. As expected, the bimodal posterior is captured by our algorithm.

To do so, we use the numerical framework in Section III to update the conditional distributions for θ_t sequentially as new observations arrive, as in [1]. We consider two setups for the orientation's velocity u in this example. In the first setup, we assume the velocity is known and non-zero. In this case, due to the known direction of the motion, the problem becomes observable, leading to a unimodal filtering distribution. In the second setup, we assume zero-velocity in our algorithm, while the actual velocity is non-zero. Since the direction of motion is unknown, the conditional distribution remains bimodal. The results are depicted in Figure 4. For comparison, we have shown results from the Ensemble Kalman filter (EnKF) [44] and the sequential importance sampling and resampling (SIR) particle filter [31]. As a quantitative comparison, we evaluated the mean-squared-error (MSE) in estimating $f(\theta) = d_{S^1}(\theta, 0)$. The result shows consistency of our algorithm with SIR. We expect the OT approach to outperform SIR in highdimensional problems, as shown in the Euclidean case in [1].

D. Optimal filtering on SE(2)

Building on the previous example, we assume the position of the robot is also unknown, and consider the goal of estimating the position and orientation of the robot simultaneously. The state of the robot is now represented by $z = (x, y, \theta)$, an element of SE(2). We assume y = 0, and consider independent uniform distributions on [-1, 1] for x and $[0, 2\pi)$ for θ . We consider the same distance measuring sensor as in Section IV-C. Figure 5 depicts the simulation results. We



Fig. 4. Simulation results of EnKF, SIR, and OT for the (dynamic) optimal filtering example on SO(2). Left panel: the trajectory of the samples/particles along with the true state for the cases of known and unknown velocity. Right panel: The mean-squared-error (MSE) for estimating the true state averaged over 10 independent simulations.

observe that the algorithm is able to capture the complicated posterior distribution. For verification, we show the predictive distribution for the observation Y prior to receiving the realized observation, and after receiving the observation. We note that the posterior predictive distribution is concentrated around the exact distance.



Fig. 5. Optimal filtering on SE(2). The left panel shows samples from the prior and posterior distribution of the location and orientation of the robot. The right panel shows the prior and posterior prediction for the observation signal, in comparison to the value of the exact distance.

E. Optimal transport mapping on SO(3)

Next, we consider the OT problem on the special orthogonal group SO(3). An element of SO(3) is represented by a rotation matrix R. The elements of the tangent space $T_RSO(3)$ are represented by $R\Omega$ where Ω is a skew-symmetric matrix. A three dimensional vector $\omega \in \mathbb{R}^3$ is uniquely mapped to a skew-symmetric matrix according to

$$[\omega]_{\times} := \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix}.$$

We have the exponential map $\exp_R(R[\omega]_{\times}) = Re^{[\omega]_{\times}}$, where

$$e^{[\omega]_{\times}} = I + \frac{\sin(\|\omega\|)}{\|\omega\|} [\omega]_{\times} + \frac{1 - \cos(\|\omega\|)}{\|\omega\|^2} [\omega]_{\times}^2.$$

Then, the geodesic distance is given by

$$d_{SO(3)}(R, \exp_R(R[\omega]_{\times})) = \sqrt{2} \|\omega\|.$$

The input to the neural net is the matrix representation of the rotation matrix. The output of the vector-field is designed to be a three dimensional vector ω , which is then mapped to



Fig. 6. OT mapping on SO(3). Samples from the optimal transport trajectory from P (blue) to Q (red) are shown in the figure. The two figures show four components of the rotation matrix.



Fig. 7. Optimal filtering on SO(3). The histograms show the approximate posterior distributions for three components of R using samples computed by the estimated transport map. The dashed lines correspond to two rotation matrices that are consistent with the measurement.

an element of the tangent space by $R[\omega]_{\times}$. Figure 6 depicts the result for optimally transporting an initial distribution of rotation matrices P to a final distribution Q. The distributions P and Q are chosen to be

$$\begin{bmatrix} \cos U & \sin U & 0 \\ 0 & 0 & 1 \\ \sin U & -\cos U & 0 \end{bmatrix}, \begin{bmatrix} -\cos U & -\sin U & 0 \\ -\sin U & \cos U & 0 \\ 0 & 0 & -1 \end{bmatrix},$$

respectively, with U being uniformly distributed on $[0, 2\pi]$.

F. Optimal filtering on SO(3)

Lastly, we consider the problem of characterizing the conditional distribution on SO(3). We consider noisy measurement Y = h(R) + W where $h(R) = (R_{1,1}, R_{2,1}, R_{3,1}, R_{2,2})$. and $W \sim N(0, 10^{-2})$. The selected observation function is degenerate because there are two rotation matrices that correspond to any exact measurement. For example,

$$R = \begin{bmatrix} 1 & 0 & 0\\ 0 & \frac{1}{\sqrt{2}} & \mp \frac{1}{\sqrt{2}}\\ 0 & \pm \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

are both consistent with $h(R) = (1, 0, 0, \frac{1}{\sqrt{2}})$. In our experiment, we consider a uniform prior probability distribution on SO(3) and use our proposed numerical procedure to sample from the conditional distribution. The numerical result is depicted in Figure 7. It is observed that the algorithm is able to capture the bimodal distribution.

V. CONCLUDING REMARKS

This work proposes a computational methodology for finding optimal transport maps and sampling conditional probability distributions on several instances of Riemannian manifolds. Our planned future research will focus on: (1) theoretical stability and sample-complexity analysis of the proposed approach, and (2) application of the approach to problems of attitude estimation using the publicly available Multi-Vehicle Stereo Event Camera (MVSEC) dataset [45].

REFERENCES

- M. Al-Jarrah, B. Hosseini, and A. Taghvaei, "Optimal transport particle filters," arXiv:2304.00392, 2023.
- [2] R. Sepulchre, "Optimal transport in estimation and control [about this issue]," *IEEE Control Systems Magazine*, vol. 41, no. 4, pp. 6–8, 2021.
- [3] C. Villani, Optimal Transport: Old and New. Springer, 2008, vol. 338.
- [4] G. Peyré, M. Cuturi *et al.*, "Computational optimal transport: With applications to data science," *Foundations and Trends® in Machine Learning*, vol. 11, no. 5-6, pp. 355–607, 2019.
- [5] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *International conference on machine learning*. PMLR, 2017, pp. 214–223.
- [6] N. Courty, R. Flamary, D. Tuia, and A. Rakotomamonjy, "Optimal transport for domain adaptation," arXiv:1507.00504, 2015.
- [7] S. Kolouri, S. R. Park, M. Thorpe, D. Slepcev, and G. K. Rohde, "Optimal mass transport: Signal processing and machine-learning applications," *IEEE Signal Proc. Magazine*, vol. 34, no. 4, pp. 43–59, 2017.
- [8] A. Dominitz and A. Tannenbaum, "Texture mapping via optimal mass transport," *IEEE transactions on visualization and computer graphics*, vol. 16, no. 3, pp. 419–433, 2010.
- [9] J. Rabin, G. Peyré, J. Delon, and M. Bernot, "Wasserstein barycenter and its application to texture mixing," in *International Conference on Scale Space and Variational Methods in Computer Vision*. Springer, 2011, pp. 435–446.
- [10] Z. Su, Y. Wang, R. Shi, W. Zeng, J. Sun, F. Luo, and X. Gu, "Optimal mass transport for shape matching and comparison," *IEEE transactions* on pattern analysis and machine intelligence, vol. 37, no. 11, pp. 2246– 2259, 2015.
- [11] Y. Chen, J. Karlsson, and A. Ringh, "Optimal transport for applications in control and estimation," *IEEE Control Systems Magazine*, vol. 41, no. 4, pp. 28–33, 2021.
- [12] Y. Chen, T. T. Georgiou, and M. Pavon, "Optimal transport in systems and control," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 4, pp. 89–113, 2021.
- [13] A. Taghvaei and P. G. Mehta, "Optimal transportation methods in nonlinear filtering," *IEEE Control Systems Magazine*, vol. 41, no. 4, pp. 34–49, 2021.
- [14] Y. Chen, T. T. Georgiou, and A. Tannenbaum, "Optimal transport for gaussian mixture models," *IEEE Access*, vol. 7, pp. 6269–6278, 2018.
- [15] J. Leygonie, J. She, A. Almahairi, S. Rajeswar, and A. Courville, "Adversarial computation of optimal transport maps," *arXiv*:1906.09691, 2019.
- [16] Y. Xie, M. Chen, H. Jiang, T. Zhao, and H. Zha, "On scalable and efficient computation of large scale optimal transport," *arXiv*:1905.00158, 2019.
- [17] A. Makkuva, A. Taghvaei, S. Oh, and J. Lee, "Optimal transport mapping via input convex neural networks," in *International Conference* on Machine Learning. PMLR, 2020, pp. 6672–6681.
- [18] A. Korotin, V. Egiazarian, A. Asadulaev, A. Safin, and E. Burnaev, "Wasserstein-2 generative networks," in *International Conference on Learning Representations*, 2021.
- [19] M.-D. Hua, G. Ducard, T. Hamel, R. Mahony, and K. Rudin, "Implementation of a nonlinear attitude estimator for aerial robotic vehicles," *IEEE Transactions on Control Systems Technology*, vol. 22, no. 1, pp. 201–213, 2013.
- [20] A. Barrau and S. Bonnabel, "Intrinsic filtering on Lie groups with applications to attitude estimation," *IEEE Transactions on Automatic Control*, vol. 60, no. 2, pp. 436–449, 2014.
- [21] M. Barczyk, S. Bonnabel, J. Deschaud, and F. Goulette, "Invariant EKF design for scan matching-aided localization," *IEEE Transactions* on Control Systems Technology, vol. 23, no. 6, pp. 2440–2448, 2015.
- [22] J. A. Hesch, D. G. Kottas, S. L. Bowman, and S. I. Roumeliotis, "Camera-imu-based localization: Observability analysis and consistency improvement," *The International Journal of Robotics Research*, vol. 33, no. 1, pp. 182–201, 2014.

- [23] J. Kwon, H. S. Lee, F. C. Park, and K. M. Lee, "A geometric particle filter for template-based visual tracking," *IEEE transactions on pattern analysis and machine intelligence*, vol. 36, no. 4, pp. 625–643, 2013.
- [24] C. Choi and H. I. Christensen, "Robust 3d visual tracking using particle filtering on the special euclidean group: A combined approach of keypoint and edge features," *The International Journal of Robotics Research*, vol. 31, no. 4, pp. 498–519, 2012.
- [25] R. J. McCann, "Polar factorization of maps on Riemannian manifolds," *Geometric & Functional Analysis GAFA*, vol. 11, no. 3, pp. 589–608, 2001.
- [26] L. Ambrosio, N. Gigli, and G. Savaré, Gradient flows: in metric spaces and in the space of probability measures. Springer Science & Business Media, 2005.
- [27] R. Baptista, B. Hosseini, N. B. Kovachki, and Y. Marzouk, "Conditional sampling with monotone gans: from generative models to likelihood-free inference," *arXiv*:2006.06755, 2023.
- [28] Y. Shi, V. De Bortoli, G. Deligiannidis, and A. Doucet, "Conditional simulation using diffusion Schrödinger bridges," in *Uncertainty in Artificial Intelligence*. PMLR, 2022, pp. 1792–1802.
- [29] A. Spantini, R. Baptista, and Y. Marzouk, "Coupling techniques for nonlinear ensemble filtering," *SIAM Review*, vol. 64, no. 4, pp. 921– 953, 2022.
- [30] A. Taghvaei and B. Hosseini, "An optimal transport formulation of Bayes' law for nonlinear filtering algorithms," in 2022 IEEE 61st Conference on Decision and Control (CDC). IEEE, 2022, pp. 6608– 6613.
- [31] A. Doucet and A. M. Johansen, "A tutorial on particle filtering and smoothing: Fifteen years later," *Handbook of nonlinear filtering*, vol. 12, no. 656-704, p. 3, 2009.
- [32] Y. Cheng and J. L. Crassidis, "Particle filtering for attitude estimation using a minimal local-error representation," *Journal of Guidance, Control, and Dynamics*, vol. 33, no. 4, pp. 1305–1310, 2010.
- [33] P. Vernaza and D. D. Lee, "Rao-Blackwellized particle filtering for 6-DOF estimation of attitude and position via GPS and inertial sensors," in *Proceedings of the IEEE International Conference on Robotics and Automation*, 2006, pp. 1571–1578.
- [34] J. L. Crassidis and F. L. Markley, "Unscented filtering for spacecraft attitude estimation," *Journal of Guidance, Control, and Dynamics*, vol. 26, no. 4, pp. 536–542, 2003.
- [35] S. Bonnabel, P. Martin, and E. Salaün, "Invariant extended Kalman filter: theory and application to a velocity-aided attitude estimation problem," in *Proceedings of the 48th IEEE Conference on Decision and Control*, 2009, pp. 1297–1304.
- [36] A. Barrau and S. Bonnabel, "Intrinsic filtering on Lie groups with applications to attitude estimation," *IEEE Transactions on Automatic Control*, vol. 60, no. 2, pp. 436–449, 2015.
- [37] Y. Brenier, "Polar factorization and monotone rearrangement of vectorvalued functions," *Communications on pure and applied mathematics*, vol. 44, no. 4, pp. 375–417, 1991.
- [38] S. Cohen, B. Amos, and Y. Lipman, "Riemannian convex potential maps," in *International Conference on Machine Learning*. PMLR, 2021, pp. 2028–2038.
- [39] D. J. Rezende, G. Papamakarios, S. Racaniere, M. Albergo, G. Kanwar, P. Shanahan, and K. Cranmer, "Normalizing flows on tori and spheres," in *International Conference on Machine Learning*. PMLR, 2020, pp. 8083–8092.
- [40] A. Lou, D. Lim, I. Katsman, L. Huang, Q. Jiang, S. N. Lim, and C. M. De Sa, "Neural manifold ordinary differential equations," *Advances in Neural Information Processing Systems*, vol. 33, pp. 17548–17558, 2020.
- [41] E. Mathieu and M. Nickel, "Riemannian continuous normalizing flows," Advances in Neural Information Processing Systems, vol. 33, pp. 2503– 2515, 2020.
- [42] V. De Bortoli, E. Mathieu, M. Hutchinson, J. Thornton, Y. W. Teh, and A. Doucet, "Riemannian score-based generative modeling," arXiv:2202.02763, 2022.
- [43] J.-D. Benamou and Y. Brenier, "A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem," *Numerische Mathematik*, vol. 84, no. 3, pp. 375–393, 2000.
- [44] G. Evensen, Data Assimilation: The Ensemble Kalman Filter. Springer Science & Business Media, 2006.
- [45] A. Z. Zhu, D. Thakur, T. Özaslan, B. Pfrommer, V. Kumar, and K. Daniilidis, "The multivehicle stereo event camera dataset: An event camera dataset for 3d perception," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 2032–2039, 2018.