Entropic Optimal Transport with Data-Driven Metrics on the Roto-Translation Group

Gautam Pai[1(✉)], Gijs Bellaard[3], Rick Sengers[3], Luc Florack[3], and Remco Duits[2,3]

[1] Mathematics of Imaging and AI Group, Department of Applied Mathematics, University of Twente (UT), Enschede, The Netherlands
g.pai@utwente.nl

[2] Eindhoven Artificial Intelligence Systems Institute (EAISI), Eindhoven, The Netherlands
r.duits@tue.nl

[3] Centre for Analysis, Scientific Computing and Applications - Eindhoven University of Technology (TU/e), Eindhoven, The Netherlands
{g.bellaard,h.j.c.e.sengers,l.m.j.florack}@tue.nl

Abstract. We enumerate a framework for optimal transportation on Lie Groups using costs that depend on geodesic distances derived from spatially varying data-driven metrics. We build on the entropic regularized formulation which can be efficiently solved using Sinkhorn iterations. We estimate local distances on the Lie group using logarithmic distance approximations and formulate their extension to a more general setting of data-driven metric tensors. Our formulation leads to a data-driven approximation of the Gibbs kernel which is essential to the Sinkhorn framework. We demonstrate our method with two experiments: Tractography with Diffusion-Weighted MRI and crossing-preserving interpolations of measures in SE(2).

Keywords: Lie Groups · Optimal Transport · Wasserstein Barycenters · Metric Geometry

1 Introduction

Lie Groups: Lie groups play a crucial role in understanding continuous symmetries and are particularly useful for modeling geometric transformations such as translations, rotations, and scaling which are very common in many applications. A Lie group is both a smooth manifold and a group, where the group operations are compatible with the manifold's smooth structure. Additionally, Lie groups can be endowed with a metric tensor field turning it into a Riemannian manifold.

In image analysis, the roto-translation group SE(2) has been of considerable interest. Images can be *lifted* into a space of positions and orientations on which SE(2) group acts naturally. The lifting enables an orientation-aware processing

T. A. Bubba et al. (Eds.): SSVM 2025, LNCS 15668, pp. 350–363, 2025.
https://doi.org/10.1007/978-3-031-92369-2_27

making it easier to detect important image substructures like lines, crossings, and bifurcations. Processing images by *lifting* them to SE(2) has been shown to have practical value for various applications like crossing-preserving denoising [13] and geometric deep learning [4].

There are compelling advantages of working with methods that include Lie group symmetries. First, it is natural to demand that any operation on input data (e.g. denoising, segmentation, feature extraction, etc.) be *equivariant* to a specific group acting on it. For example, operating on the image after a roto-translation or roto-translating the output after the operation must yield identical results. Secondly, the differential geometric properties of Lie groups like Riemannian and sub-Riemannian geodesics, distances, and kernels allow for very convenient implementations of many non-trivial geometric constraints that are otherwise challenging to implement. For example, anisotropic (or even sub-Riemannian) metrics in SE(2) relate to association fields from neurogeometry [17] and have a line completion behavior in images which is useful in many practical applications.

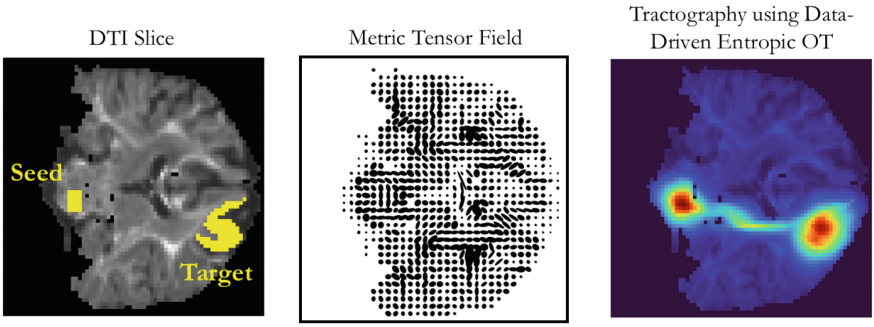


Fig. 1. Data-Driven Entropic OT for DTI Tractography: We estimate a tractography from diffusion tensor imaging (DTI) by applying entropic regularised optimal transport. (Left) A slice image from a DTI scan with designated seed and target regions. (Center) The spatially varying data-driven metric tensor field of 2x2 symmetric positive definite matrix field visualized as ellipses on $\mathbb{R}^2$ (subsampled by 2 for visualization). (Right) Tractography estimated using Entropic OT over the data-driven cost on the Translation group $\mathbb{R}^2$. Our method considers the underlying domain's heterogeneous and anisotropic metric tensor field.

Optimal Transport on Lie Groups with Left-Invariant Costs: Optimal transport (OT) is concerned with finding the most cost-efficient way of transforming one probability measure into another and is applicable to measures defined on very general metric spaces including Riemannian manifolds [1,20]. Recent works like [8,14] explore the theory and applications of OT on Lie groups especially the roto-translation group SE(2). However, here the focus has been on working with costs that correspond to geodesic distances derived from *left-invariant* metrics on the Lie group (see (2)). In particular, [8] shows

that OT using left-invariant costs leads to *invariance* of the Wasserstein distance and *equivariance* of Wasserstein Barycenters and gradient flow iterations for the same group actions on input arguments (see also Remark 2.19 in [19]). Moreover, Entropic OT using left-invariant metrics can be computed efficiently using group convolutions and local geodesic distance estimates that enable a sufficiently accurate approximation of the true Gibbs kernel on the manifold.

Contributions: In this paper, we enumerate a framework for Entropic OT on Lie groups with costs that are *data-driven* and typically not left-invariant to the group action. We leverage the logarithmic distance approximations of the Lie group and take into account the data-driven nature of a certain class of metric tensor fields on the underlying Riemannian manifold. We show that our formulation yields valid approximations of the Gibbs kernel and is hence applicable to a more general setup for Entropic OT on Riemannian manifolds. Our contributions can be listed as follows:

1. We formulate Entropic OT on translation and roto-translation Lie Groups like $\mathbb{R}^3$ and SE(2) using distances/costs that correspond to data-driven metric tensor fields. We extend the Sinkhorn algorithm to this setting with the aid of a data-driven diffusion of the measures with Gibbs kernels that account for spatially varying metric tensor fields on the underlying manifold. An example is shown in Fig. 1 on the translation group $\mathbb{R}^2$.
2. We derive local distance estimates for a certain class of such data-driven metric-tensor fields. We theoretically show that they can be used as a reasonable substitution (as underpinned by Theorem 1) for the true Gibbs kernel of the Riemannian manifold and used in the Sinkhorn algorithm.
3. We show (in Theorem 2) that OT with data-driven costs is also *data-driven equivariant*. i.e. Wasserstein Barycenters in this setting are equivariant to a simultaneous group action on the metric tensor field and input measures.
4. We experimentally demonstrate our approach in two applications: Tractography in Diffusion Tensor Imaging (DTI) and data-driven interpolation of measures in SE(2). With a reduced competitive runtime, our experiments consistently agree with methods based on the expensive numerical computation of geodesic distances.

2 Preliminaries

We start with a brief overview of some mathematical preliminaries for the main goal of data-driven optimal transport on Lie groups.

Lie Groups: We recall that a Lie group is a smooth manifold G equipped with a group structure such that the group multiplication $\texttt{mult} : G \times G \to G$ given by $\texttt{mult}(g, h) = gh$ and the inverse $\texttt{inv} : G \to G$ given by $\texttt{inv}(g) = g^{-1}$ are smooth maps. The operations of left translation by a fixed element $g \in G$ is defined as the mapping: $L_g : G \to G, \;\; L_g(h) = gh$. Every $g \in G$ has an associated tangent space T_gG. The tangent space at the identity element $e \in G$, T_eG, is called the Lie algebra. For a smooth map $\varphi : G \to N$ between G and another smooth

manifold N, the differential of φ at $p \in G$ is a linear map $(\mathrm{d}\varphi)_p : T_pG \to T_{\varphi(p)}N$ from the tangent space of G at p to the tangent space N at $\varphi(p)$. The image of $\mathrm{d}\varphi_p(\dot{p})$ of a tangent vector $\dot{p} \in T_pG$ under $\mathrm{d}\varphi_p$ is called the push-forward of $\dot{p}$ by φ. In case $\varphi = L_g$ for a given $g \in G$ we have $(\mathrm{d}L_g)_p : T_pG \to T_{gp}G$

Group Action, Invariance, and Equivariance: Given a group G and a set X, a *group action* $A_g : X \to X$ is a mapping that satisfies $A_{g_2} \circ A_{g_1} = A_{g_2 g_1}$ for all $g_1, g_2 \in G$ and $A_e = \mathrm{id}_X$. We say that G *acts* on X. Let $A_g : X \to X$ and $B_g : Y \to Y$ be two group actions. A function $\varphi : X \to Y$ is *invariant* under A_g if $\varphi \circ A_g = \varphi$. A function $\varphi : X \to Y$ is *equivariant* (w.r.t. A_g and B_g) if $\varphi \circ A_g = B_g \circ \varphi$. In the specific case where the group action is the left-translation L_g we speak of *left-invariance*. For example: $\mathcal{L}_g : L^2(G) \to L^2(G)$ defined by: $(\mathcal{L}_g U)(h) := U(g^{-1}h)$ is called the left translation of functions $U \in X = L^2(G)$.

Left Invariant Metric Tensor and Distance: As an additional structure, we endow the group G with a metric tensor field $\mathcal{G}$, turning it into a Riemannian manifold which we denote by $(G, \mathcal{G})$. The metric tensor at a point $p \in G$ is a bilinear symmetric positive definite map $\mathcal{G}_p : T_pG \times T_pG \to \mathbb{R}$ which defines an inner product in the linear space T_pG. The metric induces a norm defined by $\|v\|_{\mathcal{G}} = \sqrt{\mathcal{G}_p(v, v)}$ for any tangent vector $v \in T_pM$. The metric tensor induces a natural way of defining a distance $d_{\mathcal{G}}$ on G as

$$d_{\mathcal{G}}(p, h) := \inf_{\substack{\gamma \in PC([0,1];G) \\ \gamma(0)=p,\, \gamma(1)=h}} L(\gamma), \qquad \text{where } L(\gamma) := \int_0^1 \|\dot{\gamma}(t)\| \mathrm{d}t, \tag{1}$$

where $\mathrm{PC}([0, 1], G)$ denotes the family of piece-wise continuously differentiable curves in G. We say that the metric tensor is left-invariant if and only if for all $g, h \in G$ and $u, v \in T_hG$: $\mathcal{G}_h(u, v) = \mathcal{G}_{L_g(h)}((\mathrm{d}L_g)_h u, (\mathrm{d}L_g)_h v)$. When this property holds, it induces a left-invariant distance:

$$d_{\mathcal{G}}(p, h) = d_{\mathcal{G}}(L_g(p), L_g(h)), \quad \forall g, p, h \in G. \tag{2}$$

Distinguishing Left-Invariant and Data-Driven Metrics: By choosing a left-invariant basis for the tangent bundle TG of a Lie group G, left-invariant metric tensor fields $\mathcal{G}$ can be conveniently represented using constant coefficients in this basis as fixed symmetric positive definite matrices throughout the domain G. Data-driven metrics in general do not satisfy Eq. (2) and hence in practice, such metrics (and their coefficients in the left-invariant basis) are generally inhomogeneous along the underlying domain.

Entropic Optimal Transport: Among the many numerical strategies that have been proposed in the literature for Optimal Transportation problems, probably the most popular direction consists in adding an entropic regularization term to the problem [6,9,23], and considering an approximation of the measures by discrete measures via sampling or discretization on a grid. The discretization, and the additional entropic term lead to an approximation of the transport

plan that can be computed using a fixed-point strategy, popularly known as the Sinkhorn algorithm.

Consider the Kantarovich problem: Let $\mu_0, \mu_1 \in \text{Prob}(G)$ be two probability densities over the Lie Group G with a metric tensor field $\mathcal{G}$ and with respect to the left-invariant (Haar) measure μ. We use the following shorthand for integration on G: $\int d\mu(g) = \int dg$. The entropic ε-regularized 2-Wasserstein distance between $\mu_0, \mu_1 \in \text{Prob}(G)$ can be written as:

$$\mathcal{W}^2_{2,\varepsilon}(\mu_0, \mu_1) = \inf_{\pi \in \Pi(\mu_0, \mu_1)} \int_G \int_G d_{\mathcal{G}}(g,h)^2 \pi(g,h)\, dg\, dh - \varepsilon H(\pi) \tag{3}$$

where $\Pi(\mu_0, \mu_1) = \{\pi \in \text{Prob}(G \times G) : \int_G \pi(g,h) dh = \mu_0(g)$ and $\int_G \pi(g,h) dg = \mu_1(h)\}$ is the set of couplings and $H(\pi) = -\int_G \int_G \pi(g,h) \log \pi(g,h)\ dg\ dh$ is the Entropy of a coupling. One can show [8,9] that the minimizer in Eq. (3) can be written as: $\pi^\star(g,h) = v(g) \cdot K_{\varepsilon,\mathcal{G}}(g,h) \cdot w(h)$ where $K_{\varepsilon,\mathcal{G}}(g,h) = \exp(-d_{\mathcal{G}}(g,h)^2/\varepsilon)$ is the Gibbs Kernel and the potentials $v(.)$ and $w(.)$ can be estimated iteratively using Sinkhorn iterations:

$$\begin{aligned} v_{i+1}(g) &= \mu_0(g) / \left(\int_G K_{\varepsilon,\mathcal{G}}(g,h) \cdot w_i(h)\ dh \right) \\ w_{i+1}(h) &= \mu_1(h) / \left(\int_G K_{\varepsilon,\mathcal{G}}(g,h) \cdot v_{i+1}(g)\ dg \right) \end{aligned} \tag{4}$$

for $i = 0, 1, 2, \ldots, N$, till convergence.

For left-invariant metrics, the denominator can be computed conveniently using a group-convolution with a left-invariant Gibbs kernel $K_\varepsilon(g,h) = k_\varepsilon(g^{-1}h) := \exp(-d_{e,\mathcal{G}}(g^{-1}h)^2/\varepsilon)$. This is because for left-invariant metrics $\mathcal{G}$: $d_{\mathcal{G}}(g,h) = d_{\mathcal{G}}(g^{-1}g, g^{-1}h) = d_{\mathcal{G}}(e, g^{-1}h) =: d_{e,\mathcal{G}}(g^{-1}h)$ where e is the unit element of the Lie group G. However, this is not true for general data-driven metrics, and in such scenarios, this cannot be simplified into a straightforward convolution. However, in Sect. 3 we propose a *data-driven* integral computation that is relatively efficient to implement in a GPU using fast matrix-vector operations. The entropic regularized formulation described above has been extended to other OT problems like Wasserstein Barycenters [6,9,23] and Gradient Flows [18]. A particular case relevant in this paper is the computation of displacement or McCann's interpolation between source and target densities $\mu_0, \mu_1 \in \text{Prob}(G)$:

$$\mu_t = \underset{\mu \in \text{Prob}(G)}{\arg\min}\ (1-t)\, W^2_{\mathcal{G},\varepsilon}(\mu_0, \mu) + t\, W^2_{\mathcal{G},\varepsilon}(\mu, \mu_1), \quad t \in [0,1]. \tag{5}$$

In all of these different entropic regularized OT problems, an appropriate integral operation with a Gibbs Kernel like Eq. (4) is an essential requirement. Furthermore, in the case of measures defined on Riemannian manifolds, the choice of the Metric tensor field $\mathcal{G}$ crucially determines the smoothing action and thereby the barycenter μ_t.

3 Entropic Optimal Transport Using Data-Driven Metrics

Motivation: The main focus of this paper is to efficiently and accurately approximate the kernel operation in Eq. (4) over a Lie group with a particular focus on *data-driven* metrics $\mathcal{D}$. In general, the Riemannian distance $d_{\mathcal{D}}$ corresponding to such metrics $\mathcal{D}$ does not have a closed form and one typically resorts to computationally expensive numerical implementations. However, in the same spirit as [23], we choose to apply the Gibbs kernel by replacing true Riemannian distances with an approximation. First, we assume a class of data-driven metrics $\mathcal{D}$ such that the Riemannian distances *locally* around some $g \in G$ can be approximated with a left-invariant metric assumed by $\mathcal{G} = \mathcal{D}_g$. This approach is applicable for metric tensor fields that are slowly varying with respect to some base metric that we choose to be left-invariant. For example, in the metric tensor field shown in Fig. 1 comprising of 2×2 symmetric positive definite matrices (visualized as ellipses in $\mathbb{R}^2$), we assume that *locally* in a neighborhood, the metric tensors do not vary significantly with respect to a base isotropic metric in $\mathbb{R}^2$. We make this formulation more precise in Theorem 1. Secondly, we extend the results from prior work, notably [5,8], and approximate the left-invariant distances with logarithmic distance approximations of the Lie group

$$K_{\varepsilon,\mathcal{D}}(g,h) = e^{-d^2_{\mathcal{D}}(g,h)/\varepsilon} \approx e^{-d^2_{\mathcal{D}_g}(e,h^{-1}g)/\varepsilon} \approx e^{-\rho^2_{\mathcal{D}_g}(h^{-1}g)/\varepsilon} = k_{g,\varepsilon}(h^{-1}g) \tag{6}$$

where $\rho_{\mathcal{D}_g} : G \to \mathbb{R}_{\geq 0}$ logarithmic distance approximation centered at $g \in G$ with a left-invariant metric tensor constructed with $\mathcal{D}_g$. In the entropic setting where the regularization parameter $\varepsilon > 0$ implicitly induces a notion of locality on the underlying domain, we posit that a more tractable logarithmic approximation can substitute the Riemannian distance-based Gibbs kernel.

In this data-driven setting, the approximated kernel is generally not globally left-invariant. I.e. in general it does not hold that $k_{g_1,\varepsilon}(.) = k_{g_2,\varepsilon}(.)$ for all $g_1, g_2 \in G$. Computationally this does cause an increase in overhead since we now smooth with a different filter for each grid location in the discretized domain G. This is in contrast to a constant kernel in the left-invariant setting that is very efficiently implemented in modern geometric deep-learning pipelines using group convolutions. However, with a sufficient decay enforced by ε, we only store the values of the kernel in a small patch around every point and we efficiently apply the kernel operations on a GPU with straightforward matrix-vector multiplications.

Logarithmic Distance Approximations on Lie Groups: For a Lie group with a surjective Lie group exponential $\exp : T_eG \to G$ we can, after choosing an appropriate subset $S \subset T_eG$, make it injective. This way the Lie group logarithm $\log : U \to T_eG$, the inverse of the exponential, is well-defined on $U = \exp(S)$. Given a left-invariant metric $\mathcal{G}$ with corresponding distance $d(e,g)$, we can use the Lie group logarithm to create an approximation of $d(e,g)$ in the following way. The *logarithmic distance approximation* $\rho : U \to \mathbb{R}_{\geq 0}$ is defined as:

$$\rho(g) = \| \log g \|_e \tag{7}$$

where $\|\cdot\|_e : T_eG \to \mathbb{R}_{\geq 0}$ is the norm induced by $\mathcal{G}_e$. The calculation of ρ is relatively easy and often expressible in closed form for many important Lie groups (for example matrix Lie groups), and is generally applicable. We now enumerate these approximations for the following lie-groups of interest in this paper: $\mathbb{R}^3$ and SE(2).

Roto-Translation Group SE(2): An analytic SE(2) distance approximation is expanded in detail in [5], building upon standard approximation theory in Lie groups [11,21]. Let $\mathrm{SE}(2) = \mathbb{R}^2 \times \mathbb{R}/(2\pi\mathbb{Z})$ be the special Euclidean group of the two-dimensional plane, i.e. the group of two-dimensional roto-translations. For all $g_1 = (x_1, y_1, \theta_1)$, $g_2 = (x_2, y_2, \theta_2) \in \mathrm{SE}(2)$ we have the group product: $g_1 \cdot g_2 := (x_1 + x_2 \cos\theta_1 - y_2 \sin\theta_1, y_1 + x_2 \sin\theta_1 + y_2 \cos\theta_1, \theta_1 + \theta_2 \bmod 2\pi)$ and the unit element is $e = (0, 0, 0)$. The Lie group logarithm is $\log(x, y, \theta) = c^1 \partial_x|_e + c^2 \partial_y|_e + c^3 \partial_\theta|_e \in T_eG$:

$$(c^1, c^2, c^3) = \left(\left(x \cos\tfrac{\theta}{2} + y \sin\tfrac{\theta}{2}\right) / \operatorname{sinc}\tfrac{\theta}{2}, \left(-x \sin\tfrac{\theta}{2} + y \cos\tfrac{\theta}{2}\right) / \operatorname{sinc}\tfrac{\theta}{2}, \theta\right). \quad (8)$$

In the logarithm calculation the small angle identification: $\theta \in [-\pi, \pi)$ is important. A basis of left-invariant vector fields is given by $\mathcal{A}_1 = \cos\theta\ \partial_x + \sin\theta\ \partial_y$, $\mathcal{A}_2 = -\sin\theta\ \partial_x + \cos\theta\ \partial_y$, and $\mathcal{A}_3 = \partial_\theta$. We define left-invariant metric tensor field by: $\mathcal{G}(\mathcal{A}_i, \mathcal{A}_j) = \gamma_{ij}$. The logarithmic distance approximation is given by:

$$\rho(g) = \sqrt{\sum_{i,j=1}^{3} \gamma_{ij}\ c^i(g)\, c^j(g)} \quad (9)$$

Translation Group $\mathbb{R}^3$: For all $g_1 = (x^1, x^2, x^3)$, $g_2 = (y^1, y^2, y^3) \in \mathbb{R}^3$ we have the group product: $g_1 \cdot g_2 := (x^1 + y^1, x^2 + y^2, x^3 + y^3)$ and the unit element is $e = (0, 0, 0)$. The Lie algebra of $\mathbb{R}^n$ as a Lie group is isomorphic to itself, with the Lie bracket being trivial since the group is Abelian and the left-invariant basis is simply $\partial_{x^1}, \partial_{x^2}, \partial_{x^3}$. Furthermore, the exponential and logarithmic maps are both identities. Given a left-invariant symmetric 3×3 positive definite metric tensor $\mathcal{G}(\partial_{x^i}, \partial_{x^j}) = \gamma_{ij}$, the logarithmic distance approximation around $g = (x^1, x^2, x^3)$ is given by: $\rho(g) = \sqrt{\sum_{i,j=1}^{3} \gamma_{ij} x^i x^j}$.

We now report two theoretical arguments to validate our framework. First, we report that distances corresponding to data-driven metrics that have a certain degree of regularity with respect to a left-invariant metric can be approximated locally with left-invariant distances (and hence with logarithmic distance approximations). Secondly, we show that OT is *data-driven equivariant*, i.e. - OT is equivariant to the group action on the input measures and the data-driven metric tensor field simultaneously.

Data-Driven Distance Approximations: We consider data-driven metrics $\mathcal{D}$ of the form $\mathcal{D} = e^{2f}\mathcal{G}$ where $\mathcal{G}$ is some "base" metric (e.g. a left-invariant metric in Eq. (2)), and $f : M \to \mathbb{R}$ a smooth scalar field. This is known as a *Weyl transformation*. We are trying to analyze how good a local approximation of the metric such as $\hat{\mathcal{D}} = e^{2f(p_1)}\mathcal{G}$ is with respect to the actual metric $\mathcal{D}$, around any

point p_1. We are mostly interested in how the approximate distance $d_{\hat{\mathcal{D}}}(p_1, p_2)$ compares to the true distance $d_{\mathcal{D}}(p_1, p_2)$ as that is what is used in practice. As we are working with distances, them being nonnegative, their relative difference is the only measure that makes sense. The following theorem shows that if we take f to be slowly varying w.r.t. the base metric $d_{\mathcal{G}}$, and not too big or small, then the approximate distance $d_{\hat{\mathcal{D}}}(p_1, p_2)$ and the true distance $d_{\mathcal{D}}(p_1, p_2)$ are close together locally.

Theorem 1. *Let M be a manifold with a Riemannian metric tensor field $\mathcal{G}$. Let $f : M \to \mathbb{R}$ be K-Lipschitz w.r.t $d_{\mathcal{G}}$, meaning that $|f(p_1) - f(p_2)| \leq K d_{\mathcal{G}}(p_1, p_2)$. Define $m = \max e^f / \min e^f \geq 1$. Pick a point $p_1 \in M$. Define the metrics $\mathcal{D} = e^{2f}\mathcal{G}$ and $\hat{\mathcal{D}} = e^{2f(p_1)}\mathcal{G}$. We have*

$$e^{-Km\, d_{\mathcal{G}}(p_1,p_2)} \leq \frac{d_{\mathcal{D}}(p_1, p_2)}{d_{\hat{\mathcal{D}}}(p_1, p_2)} \leq e^{K d_{\mathcal{G}}(p_1,p_2)} \text{ for all } p_2 \in M. \tag{10}$$

Proof. Let $\gamma : [0, 1] \to M$ be any curve between $\gamma(0) = p_1$ and $\gamma(1) = p_2$. Then

$$L_{\mathcal{D}}(\gamma) = \int_0^1 \|\dot{\gamma}(t)\|_{\mathcal{D}} \, \mathrm{d}t = \int_0^1 e^{f(\gamma(t))} \, \|\dot{\gamma}(t)\|_{\mathcal{G}} \, \mathrm{d}t \tag{11}$$

from which we see the straightforward bounds

$$\left(\min_{0\leq t\leq 1} e^{f(\gamma(t))}\right) L_{\mathcal{G}}(\gamma) \leq L_{\mathcal{D}}(\gamma) \leq \left(\max_{0\leq t\leq 1} e^{f(\gamma(t))}\right) L_{\mathcal{G}}(\gamma). \tag{12}$$

Let $\gamma : [0, 1] \to M$ be a minimizing geodesic between $\gamma(0) = p_1$ and $\gamma(1) = p_2$ w.r.t. the distance $d_{\mathcal{G}}$. Then $L_{\mathcal{G}}(\gamma) = d_{\mathcal{G}}(p_1, p_2)$ and $d_{\mathcal{G}}(p_1, \gamma(t)) \leq d_{\mathcal{G}}(p_1, p_2)$, so

$$\begin{aligned} d_{\mathcal{D}}(p_1, p_2) \leq L_{\mathcal{D}}(\gamma) &\leq \left(\max_{0\leq t\leq 1} e^{f(\gamma(t))}\right) L_{\mathcal{G}}(\gamma) \\ &\leq \left(\max_{0\leq t\leq 1} e^{f(p_1)+K d_{\mathcal{G}}(p_1,\gamma(t))}\right) d_{\mathcal{G}}(p_1, p_2) \\ &= e^{f(p_1)+K d_{\mathcal{G}}(p_1,p_2)} d_{\mathcal{G}}(p_1, p_2). \end{aligned} \tag{13}$$

Now let $\gamma : [0, 1] \to M$ be a minimizing geodesic between $\gamma(0) = p_1$ and $\gamma(1) = p_2$ w.r.t. the metric $\mathcal{D}$, meaning that $L_{\mathcal{D}}(\gamma) = d_{\mathcal{D}}(p_1, p_2)$ and

$$\min e^f d_{\mathcal{G}}(p_1, \gamma(t)) \leq d_{\mathcal{D}}(p_1, \gamma(t)) \leq d_{\mathcal{D}}(p_1, p_2) \leq \max e^f d_{\mathcal{G}}(p_1, p_2), \tag{14}$$

so $d_{\mathcal{G}}(p_1, \gamma(t)) \leq m d_{\mathcal{G}}(p_1, p_2)$, and akin: $d_{\mathcal{D}}(p_1, p_2) \geq e^{f(p_1)-Km d_{\mathcal{G}}(p_1,p_2)} d_{\mathcal{G}}(p_1, p_2)$.

Data-Driven Wasserstein Equivariance: In the next theorem we show that OT using a data-driven metric is *data-driven equivariant*. To this aid, we use a straightforward lemma:

Lemma 1. *Let (X_1, d_1) and (X_2, d_2) be two (complete and separable) metric spaces and $\varphi : X_1 \to X_2$ an isometry between them. Let $\mu_t^i(\mu_0, \mu_1)$ denote (entropic regularized) Wasserstein barycenter interpolation on either space. We have*

$$\varphi_{\#}\mu_t^1(\mu_0, \mu_1) = \mu_t^2(\varphi_{\#}\mu_0, \varphi_{\#}\mu_1) \text{ for all } t \in [0, 1]. \tag{15}$$

Proof. Abstractly, the isometry φ is an isomorphism between the metric spaces under consideration. This means that all things that can be done on either space are in correspondence through φ, including Wasserstein barycenter. To be more specific, in (3) we have that the set of couplings satisfies $\varphi_{\#}\pi \in \Pi(\varphi_{\#}\mu_0, \varphi_{\#}\mu_1) \Leftrightarrow \pi \in \Pi(\mu_0, \mu_1)$ and that everything right of the infimum, that being the associated cost $C(\pi)$ of the coupling π, satisfies $C(\pi) = C(\varphi_{\#}\pi)$. From this one deduces that $\mathcal{W}_\varepsilon^2(\varphi_{\#}\mu_0, \varphi_{\#}\mu_1) = \mathcal{W}_\varepsilon^2(\mu_0, \mu_1)$ as expected. An analogous argument applied to (5) shows (15).

Theorem 2. *Let G be a Lie group with a left-invariant Riemannian metric tensor field $\mathcal{G}$. Let $\mu_t(\mu_0, \mu_1, U)$ be the (entropic regularized) Wasserstein barycenter interpolation with the data-driven $U : M \to \mathbb{R}_{>0}$ Riemannian metric tensor field $\mathcal{D}_U := U\,\mathcal{G}$. The interpolation is* data-driven equivariant *in the sense that*

$$(L_g)_{\#}\mu_t(\mu_0, \mu_1, U) = \mu_t((L_g)_{\#}\mu_0, (L_g)_{\#}\mu_1, \mathcal{L}_g U) \tag{16}$$

Proof. Given Eq. (1) we see that the only thing we need to prove is that L_g is an isometry between $(G, \mathcal{D}_U)$ and $(G, \mathcal{D}_{\mathcal{L}_g U})$, indeed:

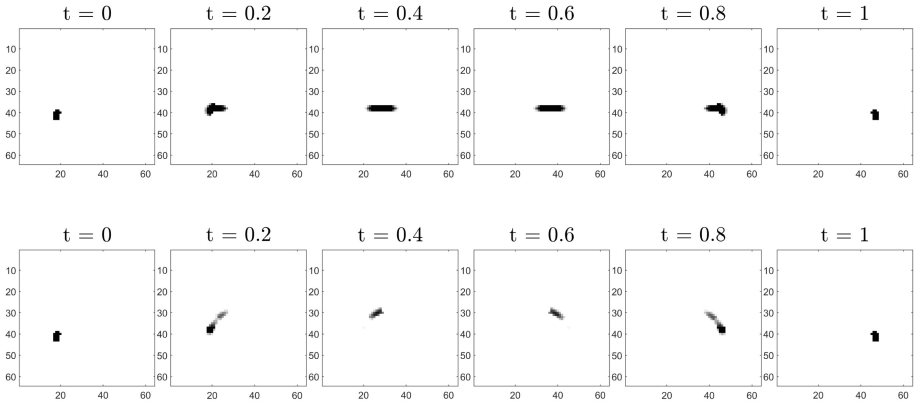
$$\begin{aligned} \mathcal{D}_{\mathcal{L}_g U}|_{gh}((\mathrm{d}L_g)_h u, (\mathrm{d}L_g)_h v) &= (\mathcal{L}_g U)(gh)\; \mathcal{G}|_{gh}((\mathrm{d}L_g)_h u, (\mathrm{d}L_g)_h v) \\ &= U(h)\; \mathcal{G}|_h(u, v) = \mathcal{D}_U|_h(u, v) \end{aligned} \tag{17}$$

The left-invariance of $\mathcal{G}$ is applied at the second equality. $(L_g)_{\#} : \mathrm{Prob}(G) \to \mathrm{Prob}(G)$ is the push forward of measures on G, analogous to $\mathcal{L}_g$.

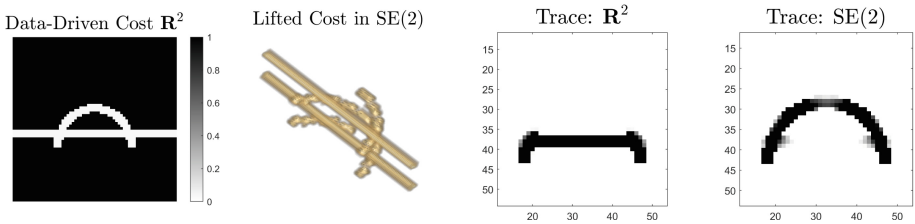
4 Experiments

The focus of our experiments is to trace the Wasserstein interpolation between source and target distributions using a data-driven metric on Lie groups SE(2) and $\mathbb{R}^n$. In particular, we compute the Wasserstein Barycenter μ_t from Eq. (5) between the source and target densities μ_0, μ_1 for N equispaced time steps $\{t_i = \frac{i}{N-1}\}_{i=0}^{i=N-1}$ and integrate the results in time to visualize the complete flow in a single cumulative density: $\tilde{\mu} = \frac{1}{N}\sum_{i=0}^{N-1} \mu_{t_i}$

Data-Driven Displacement Interpolation in $\mathbb{R}^2$ and SE(2): In Fig. 2, we experimentally validate our approach using data-driven metrics on Lie groups $\mathbb{R}^n$ and SE(2). Specifically, we leverage a well-known result in OT: that the barycenter between two measures coincides with the McCann's or displacement interpolation, which transports the mass of the measure along a geodesic on the underlying manifold [25, Section 7]. To this aid, we validate that the path traced



(a) Displacement interpolation in $\mathbb{R}^2$ (top row) and SE(2) (bottom row)



(b) Data-Driven Cost Functions on $\mathbb{R}^2$ and SE(2) and the paths traced from Figure 2a

Fig. 2. Data-Driven Tractography for Lie Groups: We demonstrate an example of data-driven tractography on Lie groups by tracing the path of displacement interpolation between two point-like masses on $\mathbb{R}^2$ and SE(2). Figure 2b visualizes the data-driven cost function leading to the displacement interpolations in (a) which is then integrated yielding a trace in (b). For SE(2) we employ an anisotropic metric of $\zeta = 5$ in conjunction with the lifted cost.

by a small point measure between the source ($t = 0$) and target ($t = 1$) densities correlates with a geodesic on the underlying Riemannian manifold.

We discretize the spatial X-Y plane $[0, 1] \times [0, 1]$ into a 64×64 grid and place two small point-like distributions at $(0.2, 0.25)$ and $(0.8, 0.25)$ in Fig. 2a. We use the setup of Theorem 1 to construct an inhomogeneous isotropic data-driven metric tensor field in $\mathbb{R}^2$ using the cost function visualized as a scalar field: $\mathbb{R}^2 \to \mathbb{R}$ in Fig. 2b. The data-driven cost has a very high value in the dark non-tubular regions and hence the *shortest path* between the seed and target points would follow the curve that overlaps with the density $\tilde{\mu}$ shown in the 3$^{\text{rd}}$ column in Fig. 2b. We extend the same setup for SE(2) where we discretize the domain into a volume of 64×64 grid and $\theta \in [0, 2\pi)$ into 12 orientations. We *lift* the source and target densities and the cost function into SE(2) [10] with the appropriate local orientations. We analogously construct an inhomogeneous and *anisotropic* data-driven metric tensor field using the cost visualized as a scalar

field: $\mathrm{SE}(2) \to \mathbb{R}$ in the 2nd column in Fig. 2b. Note that Fig. 2b shows only the regions of low values of the data-driven cost on SE(2). The 4th column shows the output after cumulating and projecting the measures from Fig. 2a back onto the image plane. Our output agrees with previous well-known results on data-driven geodesics in SE(2) [3,7] that yield curves with minimal abrupt turns at all points along the path.

Tractography in Diffusion-Weighted Imaging: Diffusion-weighted magnetic resonance imaging (DWI) is a non-invasive imaging technique based on modeling the average displacement of water molecules in biological tissue [16]. Due to the non-alignment of fibrous tissue on a micrometer level, water molecules displace anisotropically. Clinical MRI scanners sensitize these water molecules to a magnetic gradient field along preferred directions, by which diffusion information can be obtained only at millimeter scale [2]. This poses an inverse problem called tractography: which is the task of recovering the underlying fiber tracts given an input DWI image. The underlying premise is that the local propagation of water molecules is promoted along fiber directions and impeded in the transversal directions. One particular image modality obtained from DWI is Diffusion Tensor Imaging (DTI), capturing local information of water propagation in six degrees of freedom as a symmetric positive definite tensor D, called the DTI tensor [24]. The DTI tensor models three cardinal directions of diffusion of water molecules (top-left in Fig. 3).

In this setting, a Riemannian manifold $G = \mathbb{R}^3$ then can be modeled by introducing a data-driven metric tensor $\mathcal{D}(\partial_i, \partial_j) = (D^{-1})_{ij}$ and geodesics are derived as length-minimizing curves, cf. Eq. (1). Indeed we can construct the Riemannian metric such that smaller Riemannian lengths amount to directions of large diffusion and postulate that geodesics in the Riemannian space are tentative fiber tracts [12]. We compute a geodesic tractography on a DWI scan from the Human Connectome Data Project (dataset "WU-Minn HCP Data—1200 Subjects": subject 100307; TE/TR/echo spacing 89.5/5520.0/0.78 ms; b=2000 s/mm2) in Fig. 3. We delineate source and target regions and compute 5000 geodesics by numerically solving Eq. (1) with start and endpoints distributed randomly. The geodesics are then accumulated as an empirical density discretized into a grid of dimension $91 \times 109 \times 91$ by counting the number of curves passing through each voxel as visualized in the bottom left of Fig. 3.

We use the 3×3 data-driven metric tensor field $\gamma_{ij} = (D^{-1})_{ij}$ to compute the data-driven Entropic Wasserstein Barycenters μ_t on $\mathbb{R}^3$ from Eq. (5) and integrate it into a cumulative density $\tilde{\mu}$ explained earlier. In this experiment we highlight that we implemented the *de-biased* Sinkhorn barycenters algorithm from [15] that aided in reducing the smoothing bias. Additionally, we approximate the Gibbs kernel in Eq. (6) for every $5 \times 5 \times 5$ patch for entropic smoothing with $\varepsilon = 0.8$. The results from the bottom right in Fig. 3 show a strong agreement with the pure geodesic-based approach and serve as a practical validation of our method. We remark that our approach does not include any computation of geodesics or geodesic distances as part of our method to obtain a meaningful tractography.

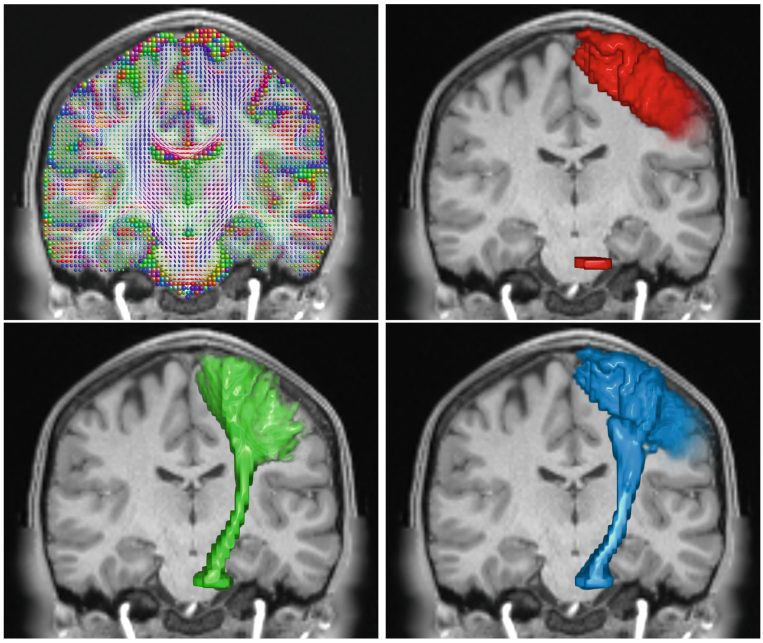


Fig. 3. Entropic OT for Geodesic Tractography: (Top Left) DTI tensor field D^{ij}, per voxel represented by quadratic form levelset $\sum_{i,j=1}^{3} D^{ij}(x)y_i y_j = 1$ visualized as ellipsoids. Although a 2D coronal slice of the brain is shown and used as a reference for context, it should be noted that the experiment is performed in $\mathbb{R}^3$. **(Top Right)** Source and target densities given by two brain regions, viz. part of the brain stem and the precentral gyrus. The Cortico Spinal Tract connects these regions. **(Bottom Left)** Results from geodesic tracking [22]. The geodesics are accumulated as an empirical density by counting the number of geodesics passing through each voxel. **(Bottom Right)** Accumulated density resulting from the Entropic OT. We observe that our method strongly agrees with the geodesic-based approach and does not compute any geodesics or large distances on the manifold.

Conclusion: We derive entropic OT on Lie groups like $\mathbb{R}^n$ and SE(2) with an emphasis on *Data-Driven* metrics. In Theorem 1 we theoretically analyze our distance approximations and in Theorem 2 show that the overall setup is *data-driven equivariant*. We experimentally validate our method in Fig. 2 and report promising application for tractography in DTI in Figs. 1 and 3. In future work, we aim to expand our framework to other Lie groups. Furthermore, we will explore the possibility of making uncertainty quantification of tractography using barycenters μ_t that could indicate an uncertainty measure that is otherwise non-trivial to obtain for this problem.

Acknowledgements. The authors sincerely thank the anonymous reviewers for their thoughtful and insightful feedback. GP acknowledges support by Sectorplan Bèta (the

Netherlands) under the focus area *Mathematics of Computational Science*. GB & RD gratefully acknowledge the Dutch Foundation of Science (NWO) for funding the VICI 2020 Exact Sciences project (Duits, *Geometric Learning for Image Analysis*, VI.C.202-031). RS and LF are supported by the project *Bringing Tractography into Daily Neurosurgical Practice*, KICH1.ST03.21.004 under the research program *Key Enabling Technologies for Minimally Invasive Interventions in Healthcare*, which is (partially) funded by NWO.

References

1. Ambrosio, L., Rigot, S.: Optimal mass transportation in the heisenberg group. J. Funct. Anal. **208**(2) (2004)
2. Basser, P.J.: Relationships between diffusion tensor and q-space MRI. Magn. Reson. Med. **47**(2), 392–397 (2002)
3. Bekkers, E., Duits, R., Mashtakov, A., Sanguinetti, G.: A PDE approach to data-driven sub-Riemannian geodesics in SE(2). SIAM SIIMs (2015)
4. Bekkers, E., Lafarge, M., Veta, M., Eppenhof, A., Pluim, J., Duits, R.: Roto-translation covariant convolutional networks for medical image analysis. In: MICCAI. Springer (2018)
5. Bellaard, G., Bon, D.L., Pai, G., Smets, B.M., Duits, R.: Analysis of (sub-) Riemannian PDE-G-CNNs. JMIV (2023)
6. Benamou, J.D., Carlier, G., Cuturi, M., Nenna, L., Peyré, G.: Iterative bregman projections for regularized transportation problems. SIAM J. Sci. Comput. (2015)
7. van den Berg, N.J., Smets, B., Pai, G., Mirebeau, J.M., Duits, R.: Geodesic tracking via new data-driven connections of cartan type for vascular tree tracking. JMIV 1–33 (2024)
8. Bon, D., Pai, G., Bellaard, G., Mula, O., Duits, R.: Optimal transport on the lie group of roto-translations. SIAM J. Imaging Sci. **18**(2), 789–821 (2025)
9. Cuturi, M.: Sinkhorn distances: lightspeed computation of optimal transport. In: Advances in Neural Information Processing Systems, vol. 26 (2013)
10. Duits, R.: Perceptual organization in image analysis [ph. d. thesis]. Department of Biomedical Engineering, TU/e (2005)
11. ter Elst, A., Robinson, D.W.: Weighted subcoercive operators on Lie groups. J. Funct. Anal. **157**, 88–163 (1998)
12. Florack, L., Sengers, R., Meesters, S., Smolders, L., Fuster, A.: Riemann-DTI geodesic tractography revisited. In: Anisotropy Across Fields and Scales, pp. 225–243. Springer (2021)
13. Franken, E.M.: Enhancement of crossing elongated structures in images. Ph.D. thesis, Department of Biomedical Engineering, TU/e (2008)
14. Galeotti, M., Citti, G., Sarti, A.: Cortically based optimal transport. JMIV (2022)
15. Janati, H., Cuturi, M., Gramfort, A.: Debiased sinkhorn barycenters. In: International Conference on Machine Learning, pp. 4692–4701. PMLR (2020)
16. Moritani, T., Ekholm, S., Westesson, P.L.: Diffusion-Weighted MR Imaging of the Brain. Springer (2009). https://doi.org/10.1007/978-3-3540-78785-3
17. Petitot, J.: The neurogeometry of pinwheels as a sub-Riemannian contact structure. J. Physiol.-Paris **97**(2–3), 265–309 (2003)
18. Peyré, G.: Entropic approximation of wasserstein gradient flows. SIAM J. Imag. Sci. **8**(4), 2323–2351 (2015)

19. Peyre, G., Cuturi, M.: Computational optimal transport. Found. Trends Mach. Learn. **11**(5–6), 355–607 (2019)
20. Quellmalz, M., Buecher, L., Steidl, G.: Parallelly sliced optimal transport on spheres and on the rotation group. JMIV 1–26 (2024)
21. Rothschild, L.P., Stein, E.M.: Hypoelliptic differential operators and nilpotent groups. Acta Math. **137**, 247 – 320 (1976)
22. Sengers, H.: Riemannian geometry in diffusion weighted imaging. Ph.D. thesis, Department of Mathematics and Computer Science, TU/e (2024)
23. Solomon, J., et al.: Convolutional wasserstein distances: efficient optimal transportation on geometric domains. ACM Trans. Graph. **34**(4) (2015)
24. Stejskal, E.O., Tanner, J.E.: Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. J. Chem. Phys. **42**(1), 288–292 (1965)
25. Villani, C.: Optimal Transport: Old and New. Grundlehren der mathematischen Wissenschaften. Springer, Heidelberg (2008)