OPTIMAL TRANSPORT ON LIE GROUP ORBITS

BAHAR TAŞKESEN

Abstract. In its most general form, the optimal transport problem is an infinite-dimensional optimization problem, yet certain notable instances admit closed-form solutions. We identify the common source of this tractability as *symmetry* and formalize it using Lie group theory. Fixing a Lie group action on the outcome space and a reference distribution, we study optimal transport between measures lying on the same Lie group orbit of the reference distribution. In this setting, the Monge problem admits an explicit upper bound given by an optimization over the stabilizer subgroup of the reference distribution. The reduced problem's dimension scales with that of the stabilizing subgroup and, in the tractable cases we study, is either zero or finite. Under mild regularity conditions, a simple algebraic certificate, verified at an optimizer of the reduced problem, ensures tightness of the upper bound for both the Monge and Kantorovich formulations, with the optimal map realized by a single group element. This orbit-based viewpoint unifies known closed-form solutions, such as those for one-dimensional and elliptical distributions, and yields a new one for optimal transport between Wishart distributions.

1. Introduction

Originating in the seminal work of Monge (1781) and later given its analytic form by Kantorovich (1942), the optimal transport problem seeks the minimal total cost of transforming one probability measure into another with respect to a prescribed transportation cost function defined on the underlying space. When both probability measures are discrete and explicitly specified by enumerating their atoms and associated probabilities, the optimal transport problem reduces to a finite-dimensional linear program. Thus, classical polynomial-time algorithms (e.g., interior-point methods (Karmarkar, 1984)) provide efficient solutions whose complexity scales polynomially with the input size. In sharp contrast, when discrete measures have implicitly defined supports, such as when distributions factorize across multiple dimensions, the number of atoms can grow exponentially with the dimension, resulting in a problem that remains polynomially describable yet is provably #P-hard (Taşkesen et al., 2023b). Likewise, the optimal transport problem between a generic (possibly continuous) measure and a discrete measure is also known to be #P-hard (Taşkesen et al., 2023a). When both probability measures are *continuous*, the optimal transport problem manifests as an infinite-dimensional linear program. Following the previously observed complexity trend, one might naturally conjecture that the complexity of optimal transport between two continuous measures, each having infinitely many atoms, would be equally formidable. Curiously, however, certain instances of the optimal transport problem between continuous probability measures admit explicit closed-form solutions, rendering these problems uniquely tractable despite their inherently infinite-dimensional structure.

University of Chicago, Booth School of Business, bahar.taskesen@chicagobooth.edu.

Early breakthroughs showed that infinite-dimensionality need not preclude analytic tractability. In one dimension, Dall'Aglio (1956) showed that the optimal transport problem when induced by any power of absolute difference cost is available in closed form in terms of the quantile functions of the marginals. For multivariate Gaussian distributions, Dowson and Landau (1982) and Olkin and Pukelsheim (1982) derived an explicit solution for the optimal transport problem induced by the quadratic cost function in terms of the means and covariance matrices. Gelbrich (1990) then generalized this to the full class of elliptically contoured distributions, yielding a single closed-form formula for the optimal transport problem and the associated optimal solution. For a comprehensive survey of closed-form solutions of optimal transport, see (Rachev and Rüschendorf, 1998, §3).

Taken together, these instances appear disparate; it is not evident a priori why an inherently infinite-dimensional problem should, in select cases, admit explicit closed-form solutions. We show that the unifying reason is symmetry, and we use Lie group theory to identify and exploit such symmetries between distributions. Specifically, we fix a Lie group action on the outcome space and a reference distribution, and study optimal transport between distributions lying on the same Lie group orbit. This orbit perspective induces a natural equivalence among distributions, allowing familiar families to be recognized as members of a single Lie group orbit. Our main result shows that, in this setting, the seemingly infinite-dimensional optimal transport problem reduces to an upper bound given by an optimization over the stabilizing subgroup of the reference law. The dimension of this reduced problem scales with that of the stabilizing subgroup and, in the tractable cases we study, is either zero or finite. Under mild conditions, a brief algebraic check turns this bound into exact equality for both the Monge and Kantorovich formulations, with the optimal transport realized by a single group element. This Lie-theoretic viewpoint recovers classical formulas for one-dimensional monotone rearrangement, elliptical families, and product distributions with coordinate-wise symmetries, and yields a new closed-form solution for transport between two Wishart distributions.

Notation. For $n \in \mathbb{N}_{>0}$, the ambient n-dimensional Euclidean space is denoted by $\mathbb{R}^n$, and it is endowed with its Borel σ-algebra $\mathcal{B}(\mathbb{R}^n)$, the Lebesgue measure $\mathcal{L}^n$, the Euclidean norm $\|\cdot\|$ and the standard inner product $\langle \cdot, \cdot \rangle$. Throughout, subsets of $\mathbb{R}^n$ are equipped with the Borel σ-algebra inherited from $\mathbb{R}^n$. We write $\mathcal{X}$ for a Polish subset of a finite-dimensional Euclidean space, in examples (with $d \in \mathbb{N}_{>0}$), $\mathcal{X}$ will be $\mathbb{R}^d$, the positive orthant $\mathbb{R}^d_+$, or the cone of positive-definite matrices $\mathbb{S}^d_{++}$. The identity map on $\mathcal{X}$ is denoted by $\mathrm{id}_\mathcal{X} : \mathcal{X} \to \mathcal{X}$, $\mathrm{id}_\mathcal{X}(\boldsymbol{x}) = \boldsymbol{x}$ for every $\boldsymbol{x} \in \mathcal{X}$. We write $\boldsymbol{I}_d$ for the $d \times d$ identity matrix and $\mathrm{GL}(d) = \{\boldsymbol{A} \in \mathbb{R}^{d \times d} \mid \det(\boldsymbol{A}) \neq 0\}$ for the real general linear group. $\mathcal{O}(d)$ denotes the orthogonal group in dimension d. We denote the set of real symmetric $d \times d$ matrices by $\mathrm{Sym}(d) = \{\mathbf{H} \in \mathbb{R}^{d \times d} \mid \mathbf{H}^\top = \mathbf{H}\}$. The Hadamard (element-wise) product is denoted by $\odot$. We denote by $\mathcal{B}(\mathcal{X})$ the Borel σ-algebra of $\mathcal{X}$. $\mathcal{P}(\mathcal{X})$ denotes the set of Borel probability measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. For a measurable cost function $c : \mathcal{X} \times \mathcal{X} \to [0, \infty]$ define

$$\mathcal{P}_c(\mathcal{X}) = \left\{ \mu_0 \in \mathcal{P}(\mathcal{X}) \;\middle|\; \exists \boldsymbol{x}_0 \in \mathcal{X} \text{ with } \int_\mathcal{X} c(\boldsymbol{x}, \boldsymbol{x}_0) d\mu_0(\boldsymbol{x}) < \infty \right\},$$

the class of c-integrable probability measures. For a Borel map $T : \mathcal{X} \to \mathcal{X}$ and $\mu \in \mathcal{P}(\mathcal{X})$, the push-forward $T_\# \mu \in \mathcal{P}(\mathcal{X})$ is defined by $T_\# \mu(\mathcal{A}) = \mu(T^{-1}(\mathcal{A}))$ for all $\mathcal{A} \in \mathcal{B}(\mathcal{X})$.

2. Preliminaries

This section introduces mass transportation problems and develops the Lie group background needed for our main results.

2.1. **Mass transportation problems.** In his memoir, Monge (1781) formulated the problem of transporting one distribution of mass into another at minimal cost. Formally, given a Borel-measurable cost function $c : \mathcal{X} \times \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ and two measures $\mu_0, \mu_1 \in \mathcal{P}_c(\mathcal{X})$, the *Monge problem* induced by cost function c is defined as

$$\text{(MP)} \qquad \mathbb{M}_c(\mu_0, \mu_1) = \inf_{T \in \mathcal{T}(\mu_0,\mu_1)} \int_{\mathcal{X}} c\,(\boldsymbol{x}, T(\boldsymbol{x}))\, d\mu_0(\boldsymbol{x}),$$

where

$$\mathcal{T}(\mu_0, \mu_1) = \{T : \mathcal{X} \to \mathcal{X} \text{ Borel-measurable } \mid T_\# \mu_0 = \mu_1\}$$

is the set of admissible transport maps. If a map $T^\star \in \mathcal{T}(\mu_0, \mu_1)$ attains the infimum in (MP), then it is called an optimal Monge map between μ_0 and μ_1.

The admissible set $\mathcal{T}(\mu_0, \mu_1)$ can be empty. For instance, if μ_0 has an atom while μ_1 is atomless, no measurable (deterministic) map can push μ_0 to μ_1 because a map cannot split mass. Even when $\mathcal{T}(\mu_0, \mu_1)$ is nonempty, the constraint $T_\# \mu_0 = \mu_1$ of (MP) is non-convex. To see this suppose that $\mu_0 = \mu_1 = \mathcal{U}([0, 1])$. Now, let $T_1(x) = x$ and $T_2(x) = 1 - x$. Then, a simple calculation shows that $T_1, T_2 \in \mathcal{T}(\mu_0, \mu_1)$. Their midpoint $\bar{T} = 0.5(T_1 + T_2)$ is the constant map sending every x to $1/2$; that is, it pushes μ_0 to a Dirac measure $\delta_{1/2}$, $\bar{T}_\# \mu_0 = \delta_{1/2} \neq \mu_1$ implying that $\bar{T} \notin \mathcal{T}(\mu_0, \mu_1)$. Thus the Monge problem is an infinite-dimensional optimization over a generally nonconvex feasible set.

In 1942, Kantorovich (1942) introduced a convex relaxation of (MP) by replacing maps with probability couplings. Given $\mu_0, \mu_1 \in \mathcal{P}_c(\mathcal{X})$, the *Kantorovich problem* induced by cost function c is defined as

$$\text{(KTP)} \qquad \mathbb{K}_c(\mu_0, \mu_1) = \inf_{\Gamma \in \Pi(\mu_0,\mu_1)} \int_{\mathcal{X} \times \mathcal{X}} c(\boldsymbol{x}, \boldsymbol{y}) \mathrm{d}\Gamma(\boldsymbol{x}, \boldsymbol{y}),$$

where $\Pi(\mu_0, \mu_1)$ denotes the set of all probability measures on $\mathcal{X} \times \mathcal{X}$ with first marginal μ_0 and second marginal μ_1. If $\Gamma^\star \in \Pi(\mu_0, \mu_1)$ solves (KTP), then it is called an optimal transportation plan between μ_0 and μ_1. The optimization in (KTP) is commonly called the optimal transport problem; to avoid ambiguity we will refer to (KTP) as the *Kantorovich problem* and to (MP) as the *Monge problem.*

The deterministic nature of the Monge problem is appealing as it yields an explicit transport map that relocates mass without splitting it, in contrast to the probabilistic transportation plans of the Kantorovich formulation in the form of couplings. This brings interpretability and aligns with settings where splitting is not physically meaningful or desirable (e.g., moving a pile of soil or routing indivisible items). There is a tight connection between the Monge problem (MP) and the Kantorovich problem (KTP): whenever a Monge solution exists, and under suitable regularity, an optimal Kantorovich plan concentrates on the graph of the Monge map. Existence of Monge solutions for the quadratic cost on Euclidean space was first developed in (Brenier, 1987; Rüschendorf and Rachev, 1990); for non-quadratic costs, existence was investigated in Rüschendorf (1991); Smith and Knott (1992); McCann (1995). One of the most general results linking Monge solutions to gradients of c-convex functions is due to Villani (2008). For ease of reference, we now adapt (Villani, 2008, Theorem 10.28) to our notation and restate it here. Before doing so, we record the regularity hypotheses on cost function c and recall the definition of c-convexity required by that theorem.

Assumption 1. *The cost function $c : \mathcal{X} \times \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ is continuous and bounded below. Additionally for every $\boldsymbol{y} \in \mathcal{X}$, the map $\boldsymbol{x} \mapsto c(\boldsymbol{x}, \boldsymbol{y})$ belongs to $\mathcal{C}^1(\mathcal{X})$, and for every $\boldsymbol{x} \in \mathcal{X}$ the map $\boldsymbol{y} \mapsto \nabla_{\boldsymbol{x}} c(\boldsymbol{x}, \boldsymbol{y})$ is injective.*

Assumption 1 contains the twist condition (see (Villani, 2008, §10)): for each fixed $\boldsymbol{x}$, no two distinct targets $\boldsymbol{y} \neq \boldsymbol{y}'$ yield the same $\boldsymbol{x}$-gradient of the cost. Geometrically, different destinations exert different first-order "forces" at $\boldsymbol{x}$. For the quadratic cost $c(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|^2$, we have $\nabla_{\boldsymbol{x}} c(\boldsymbol{x}, \boldsymbol{y}) = 2(\boldsymbol{x} - \boldsymbol{y})$, so injectivity in $\boldsymbol{y}$ is immediate; hence Assumption 1 holds.

Definition 1 (c-convexity)**.** *Suppose $c : \mathcal{X} \times \mathcal{X} \to (-\infty, +\infty]$. A function $\psi : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ is said to be c-convex if it is not identical to $+\infty$ everywhere, and there exists $\phi : \mathcal{X} \to \mathbb{R} \cup \{\pm\infty\}$ such that*

$$\psi(\boldsymbol{x}) = \sup_{\boldsymbol{y} \in \mathcal{X}} \phi(\boldsymbol{y}) - c(\boldsymbol{x}, \boldsymbol{y}) \quad \forall \boldsymbol{x} \in \mathcal{X}.$$

Note that for $c(\boldsymbol{x}, \boldsymbol{y}) = \langle \boldsymbol{x}, \boldsymbol{y} \rangle$ on $\mathbb{R}^d \times \mathbb{R}^d$, a function is c-convex if and only if it is lower semicontinuous and convex on $\mathbb{R}^d$. Moreover, for $c(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|^2$, ψ is c-convex if and only if $\boldsymbol{x} \mapsto \psi(\boldsymbol{x}) + \|\boldsymbol{x}\|^2$ is convex. In particular, if $\psi \in \mathcal{C}^2(\mathcal{X})$ then this holds if and only if $\nabla^2 \psi(\boldsymbol{x}) \succeq -2\boldsymbol{I}_d$ for all $\boldsymbol{x} \in \mathcal{X}$. We are now ready to state (Villani, 2008, Theorem 10.28).

Theorem 2.1. *Suppose $c : \mathcal{X} \times \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ satisfies Assumption 1 and assume that any c-convex function is differentiable μ_0-almost surely on its domain of c-subdifferentiability. Let $\mu_0, \mu_1 \in \mathcal{P}_c(\mathcal{X})$, then*

(i) (KTP) *admits a unique (in law) optimal transportation plan $\Gamma^\star \in \Pi(\mu_0, \mu_1)$.*
(ii) *There exists a unique Monge map $T^\star : \mathcal{X} \to \mathcal{X}$ solving* (MP).
(iii) *There exists a c-convex function φ such that*

$$\nabla_{\boldsymbol{x}} \varphi(\boldsymbol{x}) + \nabla_{\boldsymbol{x}} c(\boldsymbol{x}, T^\star(\boldsymbol{x})) = 0 \quad \mu_0\text{-almost surely.} \tag{$\star$}$$

(iv) *$\Gamma^\star$ is concentrated on the graph $T^\star$, that is, $\Gamma^\star = (\mathrm{id}_{\mathcal{X}}, T^\star)_{\#}\mu_0$.*

For the quadratic cost $c(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|^2$, the hypothesis on c-convex differentiability holds whenever $\mu_0 \ll \mathcal{L}^n$; (Villani, 2008, Example 10.36). Theorem 2.1 delineates conditions under which the Monge map is characterized via the gradient condition ($\star$).

2.2. **Lie group theory.** Originating with Lie (1891), Lie groups formalize continuous symmetry, that is, they are groups that are simultaneously smooth manifolds, with multiplication and inversion smooth. A canonical example is the circle (planar rotations), where composition and inversion vary smoothly with the angle. In what follows, we will review the minimal Lie-theoretic background used throughout; for comprehensive treatments see (Duistermaat and Kolk, 2012; Hall, 2013; Boumal, 2023).

Definition 2 (Lie group)**.** *Let $\mathcal{G}$ be both a smooth manifold and a group with operation $\cdot$. If the product map* $\mathrm{prod} : \mathcal{G} \times \mathcal{G} \to \mathcal{G} : (g, h) \mapsto \mathrm{prod}(g, h) = g \cdot h$ *and* *the inverse map* $\mathrm{inv} : \mathcal{G} \to \mathcal{G} : g \mapsto \mathrm{inv(g)} = \mathrm{g}^{-1}$ *are smooth, then* $\mathcal{G}$ *is a Lie group. Smoothness of* prod *is understood with respect to the product manifold structure on* $\mathcal{G} \times \mathcal{G}$.

We now illustrate Definition 2 with a standard example: the general linear group $\mathrm{GL}(d)$. Note that determinant is a polynomial map, so $\mathrm{GL}(d) = \det^{-1}(\mathbb{R} \setminus \{0\})$ is an open subset of $\mathbb{R}^{d^2}$. Hence, $\mathrm{GL}(d)$ is a smooth manifold of dimension d^2. The group operation is matrix multiplication and inversion. The multiplication map $\mathrm{prod} : \mathrm{GL}(d) \times \mathrm{GL}(d) \to \mathrm{GL}(d)$, with $(\boldsymbol{A}, \boldsymbol{B}) \mapsto \boldsymbol{AB}$ is smooth because each entry of $\boldsymbol{AB}$ is a polynomial in the entries of $\boldsymbol{A}$ and $\boldsymbol{B}$. The inverse map $\mathrm{inv} : \mathrm{GL}(d) \to \mathrm{GL}(d)$, with $\boldsymbol{A} \mapsto \boldsymbol{A}^{-1}$ is smooth on $\mathrm{GL}(d)$ since $\boldsymbol{A}^{-1} = \mathrm{adj}(\boldsymbol{A})/\det(\boldsymbol{A})$, where $\mathrm{adj}(\boldsymbol{A})$ is polynomial in the entries of $\boldsymbol{A}$ and $\det(\boldsymbol{A}) \neq 0$ on $\mathrm{GL}(d)$. Thus both prod and inv are smooth, so $\mathrm{GL}(d)$ is a Lie group. Many other familiar matrix groups are Lie groups including $\mathcal{O}(d)$ and $\mathrm{SL}(d) = \{\boldsymbol{A} \in \mathbb{R}^{d\times d} \mid \det(\boldsymbol{A}) = 1\}$; see (Hall, 2013, Chapter 1).

Definition 3 (Left-group action and orbit)**.** *Given a Lie group* $\mathcal{G}$ *and a manifold* $\mathcal{M}$*, a left-group action is a map* $\phi : \mathcal{G} \times \mathcal{M} \to \mathcal{M}$ *such that:*

(i) *for all* $\rho \in \mathcal{M}$, $\phi(e, \rho) = \rho$

(ii) *for all* $g, h \in \mathcal{G}$ *and* $\rho \in \mathcal{M}$, $\phi(g \cdot h, \rho) = \phi(g, \phi(h, \rho))$,

where e *is the identity element of* $\mathcal{G}$*. The orbit of* $\rho \in \mathcal{M}$ *through the left-group action* ϕ *of* $\mathcal{G}$ *is the set* $\mathcal{G}_\rho = \{\phi(g, \rho) \mid g \in \mathcal{G}\}$.

Orbits induce an equivalence relation on $\mathcal{M}$: declare ρ_1 and ρ_2 equivalent whenever $\rho_2 = \phi(g, \rho_1)$ for some $g \in \mathcal{G}$. In other words, two points are equivalent if and only if they lie on the same $\mathcal{G}$-orbit.

3. Distributions in Lie Group Orbits

We denote the infinite-dimensional group of all smooth diffeomorphisms of $\mathcal{X}$ by $\mathrm{Diff}(\mathcal{X})$, defined as

$$\mathrm{Diff}(\mathcal{X}) = \left\{\Phi : \mathcal{X} \to \mathcal{X} \mid \Phi \text{ is a } \mathcal{C}^\infty \text{ bijection and } \Phi^{-1} \text{ is } \mathcal{C}^\infty\right\}$$

We regard diffeomorphisms as group elements, where the group operation is composition, the identity element is $\mathrm{id}_{\mathcal{X}}$, and inversion is the usual map inverse. Throughout the paper any subgroup $\mathcal{G} \subseteq \mathrm{Diff}(\mathcal{X})$ is endowed with the group operation given by composition of maps: $(g_1, g_2) \mapsto g_1 \circ g_2$.

The goal of this paper is to identify when, and explain why optimal transport admits closed-form solutions. Our guiding principle is symmetry, formalized via Lie group actions. We show that many tractable instances occur when the source and target measures lie in the same orbit of a group acting on the outcome space. Accordingly, we begin by describing the push-forward action of subgroups of diffeomorphisms on probability measures, which lets us speak of orbits exactly as in classical group actions, but now at the level of distributions.

Lemma 1. *Let* $\mathcal{G} \subset \mathrm{Diff}(\mathcal{X})$ *be a Lie group. Define*

$$\phi : \mathcal{G} \times \mathcal{P}(\mathcal{X}) \to \mathcal{P}(\mathcal{X}), \qquad \phi(g, \mu) = g_\# \mu.$$

Then, ϕ *is a left-group action, that is, for all* $g, h \in \mathcal{G}$ *and* $\mu \in \mathcal{P}(\mathcal{X})$,

(i) $e_\# \mu = \mu$ *for every* $\mu \in \mathcal{P}(\mathcal{X})$*, where* e *is the identity element of* $\mathcal{G}$.

(ii) $(g \circ h)_\# \mu = g_\# (h_\# \mu)$.

Proof of Lemma 1. As g is Borel measurable, $g_{\#}\mu$ is a Borel probability measure, and $g_{\#}\mu(\mathcal{X}) = \mu(g^{-1}(\mathcal{X})) = \mu(\mathcal{X}) = 1$. For any $\mathcal{A} \in \mathcal{B}(\mathcal{X})$, we have

$$\begin{aligned}(g \circ h)_{\#}\mu(\mathcal{A}) &= \mu\left((g \circ h)^{-1}(\mathcal{A})\right) \\ &= \mu\left(h^{-1}(g^{-1}(\mathcal{A}))\right) \\ &= h_{\#}\mu\left(g^{-1}(\mathcal{A})\right) = g_{\#}(h_{\#}\mu)(\mathcal{A}),\end{aligned}$$

which proves compatibility. The identity property follows from $e^{-1}(\mathcal{A}) = \mathcal{A}$, hence $e_{\#}\mu(\mathcal{A}) = \mu(\mathcal{A})$. □

With this action in hand, we introduce the basic objects it generates.

Definition 4 (Orbit of a measure)**.** *Let $\mathcal{G} \subset \mathrm{Diff}(\mathcal{X})$. For $\rho \in \mathcal{P}(\mathcal{X})$, we define its $\mathcal{G}$-orbit as $\mathcal{G}_{\#}\rho = \{g_{\#}\rho \mid g \in \mathcal{G}\} \subset \mathcal{P}(\mathcal{X})$.*

Thus two measures are equivalent if one is the push-forward of the other by some $g \in \mathcal{G}$.

Lemma 2 (Orbit measures remain absolutely continuous)**.** *If $\rho \ll \mathcal{L}^n$, then for every $g \in \mathcal{G}$ we have $g_{\#}\rho \ll \mathcal{L}^n$ and, for $\mathcal{L}^n$-a.e. $\boldsymbol{y} \in \mathcal{X}$,*

$$\frac{\mathrm{d}(g_{\#}\rho)}{\mathrm{d}\mathcal{L}^n}(\boldsymbol{y}) = \frac{\mathrm{d}\rho}{\mathrm{d}\mathcal{L}^n}\left(g^{-1}(\boldsymbol{y})\right)\left|\det(Dg^{-1}(\boldsymbol{y}))\right|,$$

where $Dg^{-1}(\boldsymbol{y})$ is the Jacobian matrix of g^{-1} evaluated at $\boldsymbol{y}$.

Proof of Lemma 2. Because g is a $\mathcal{C}^1$-diffeomorphism, the change-of-variables formula results in the claimed density. □

Definition 5 (Stabilizer subgroup)**.** *For a reference measure $\rho \in \mathcal{P}(\mathcal{X})$, the stabilizer of ρ inside $\mathcal{G}$ is $\mathrm{Stab}_{\mathcal{G}}(\rho) = \{h \in \mathcal{G} \mid h_{\#}\rho = \rho\}$.*

Thus $\mathrm{Stab}_{\mathcal{G}}(\rho)$ consists precisely of those group elements that leave the reference measure unchanged under the push-forward action $(h, \rho) \mapsto h_{\#}\rho$. Note that $\mathrm{Stab}_{\mathcal{G}}(\rho)$ is non-empty as the identity element e of $\mathcal{G}$ already falls into $\mathrm{Stab}_{\mathcal{G}}(\rho)$.

With the action, orbits, and stabilizer in place, the subsequent section studies the optimal transport problem between measures lying on a common orbit.

4. Transport of Distributions within Lie Group Orbits

For the remainder of this paper, we fix a Lie group $\mathcal{G} \subseteq \mathrm{Diff}(\mathcal{X})$ and a reference probability measure $\rho \in \mathcal{P}(\mathcal{X})$. We study the Monge transportation problem (MP) induced by a Borel-measurable cost function $c : \mathcal{X} \times \mathcal{X} \to [0, +\infty]$, between two measures in the orbit of ρ with finite c-moment, that is, $\mu_0, \mu_1 \in \mathcal{G}_{\#}(\rho) \cap \mathcal{P}_c(\mathcal{X})$. Equivalently, there exists $g_0, g_1 \in \mathcal{G}$, such that $\mu_0 = g_{0\#}\rho$ and $\mu_1 = g_{1\#}\rho$. We assume $\rho \ll \mathcal{L}^n$, then by Lemma 2, μ_0 and μ_1 are absolutely continuous as well, and we write $r_i = \mathrm{d}\mu_i/\mathrm{d}\mathcal{L}^n$ for their densities.

A crucial observation underpinning our analysis is that, while the Monge problem is typically an infinite-dimensional and notoriously challenging optimization problem, a subtle algebraic structure emerges when the measures reside in a common $\mathcal{G}$-orbit. First, although $\mathcal{T}(\mu_0, \mu_1)$ may be empty for arbitrary distributions, when measures reside on a common orbit, then the group element $T = g_1 \circ g_0^{-1} \in \mathcal{G}$ satisfies $T_{\#}\mu_0 = \mu_1$, so $\mathcal{T}(\mu_0, \mu_1)$ is not empty. Moreover, every admissible transport admits a canonical factorization through the reference law: it can be pulled back to the ρ-coordinates. The next lemma formalizes this observation and serves as our bridge from arbitrary transportation maps to ρ-preserving transformations.

Lemma 3. *If a measurable map $T : \mathcal{X} \to \mathcal{X}$ satisfies $T_\# \mu_0 = \mu_1$, then the composite $H = g_1^{-1} \circ T \circ g_0$ is such that $H_\# \rho = \rho$ and $T = g_1 \circ H \circ g_0^{-1}$.*

Proof of Lemma 3. Because g_0 and g_1 are diffeomorphisms, their inverses are measurable. Then, by definition of the push-forward operator, for any $\mathcal{A} \in \mathcal{B}(\mathcal{X})$, we have

$$\begin{aligned} H_\# \rho(\mathcal{A}) &= \rho\left(H^{-1}(\mathcal{A})\right) = \rho\left((g_0^{-1} \circ T^{-1} \circ g_1)(\mathcal{A})\right) \\ &= g_{0\#}\rho\left((T^{-1} \circ g_1)(\mathcal{A})\right) = \mu_0\left((T^{-1} \circ g_1)(\mathcal{A})\right) \\ &= T_\# \mu_0\left(g_1(\mathcal{A})\right) = \mu_1\left(g_1(\mathcal{A})\right) = g_{1\#}\rho\left(g_1(\mathcal{A})\right) = \rho(\mathcal{A}). \end{aligned}$$

Hence, $H_\# \rho = \rho$, and this proves the first assertion of the lemma. By definition,

$$g_1 \circ H \circ g_0^{-1} = g_1 \circ \left(g_1^{-1} \circ T \circ g_0\right) \circ g_0^{-1} = T.$$

This observation completes our proof. □

Lemma 3 shows that any admissible $T \in \mathcal{T}(\mu_0, \mu_1)$ is obtained by lifting a ρ-preserving transformation H and conjugating by g_0, g_1. Thus the Monge search can be viewed as a search over ρ-preserving maps on the reference space. If, in addition, H lies in $\mathcal{G}$, then $H \in \mathrm{Stab}_{\mathcal{G}}(\rho)$. To systematically leverage this structure, we introduce *Lie group orbit transport problem* between $(g_0)_\# \rho$ and $(g_1)_\# \rho$ induced by the cost function c as in the following:

$$\text{(LGOP)} \qquad \mathbb{J}_c(g_0, g_1) = \inf_{h \in \mathrm{Stab}_{\mathcal{G}}(\rho)} \int_{\mathcal{X}} c(g_0(\boldsymbol{x}), (g_1 \circ h)(\boldsymbol{x})) \mathrm{d}\rho(\boldsymbol{x}).$$

Because (LGOP) searches only among measure preserving maps lying in $\mathcal{G}$, the feasible set of (LGOP) is contained in that of the Monge problem (MP). Consequently, (LGOP) provides an upper bound on (MP); the first statement of Theorem 4.1 formalizes this claim. Moreover, if we denote by $h^\star$ as an optimal solution of (LGOP), then $T^* = g_1 \circ h^\star \circ g_0^{-1}$ belongs to $\mathcal{T}(\mu_0, \mu_1)$ by Lemma 3.

What characterizes a Monge map is the existence of a c-convex function such that ($\star$) holds true. Thus solving (MP) is equivalent to searching for a c-convex function φ solving ($\star$). When the cost function c is smooth, differentiation of ($\star$) results in partial differential equation (PDE) of optimal transport (Villani, 2003, §12). In special case of quadratic cost, solving (MP) is equivalent to searching for a convex φ that solves an instance of the Monge-Ampère equation:

$$\det(\nabla^2_{\boldsymbol{x}} \varphi(\boldsymbol{x})) = \frac{r_0(\boldsymbol{x})}{r_1(\nabla_{\boldsymbol{x}} \varphi(\boldsymbol{x}))},$$

which is a fully-non-linear elliptic PDE, and explicit solutions are rarely available.

In contrast to the computational burden of searching a c-convex function that solves a non-linear PDE, verifying the existence of a c-convex potential that fulfills ($\star$) for $T^* = g_1 \circ h^\star \circ g_0^{-1}$ becomes a direct algebraic check. We leverage this insight to characterize precisely the conditions under which the Lie group orbit formulation (LGOP) and the Monge problem (MP) coincide, thereby reducing solving a non-linear PDE into an algebraic criterion once (LGOP) is solved. The following theorem formalizes this correspondence and provides explicit criteria under which the equivalence holds.

Theorem 4.1. *Problems* (MP), (KTP) *and* (LGOP) *satisfy the following properties.*

(i) $\mathbb{M}_c(\mu_0, \mu_1) \leq \mathbb{J}_c(g_0, g_1)$.
(ii) *If there exists* $h^\star$ *solving* (LGOP)*, then* $T^* = g_1 \circ h^\star \circ g_0^{-1} \in \mathcal{T}(\mu_0, \mu_1)$.
(iii) *If the cost function* c *satisfies Assumption 1, every* c*-convex function is differentiable* μ_0*-almost surely on its domain of* c*-subdifferentiability, and there exists a* c*-convex function* φ *such that* T^* *satisfies* ($\star$) μ_0*-almost surely, then*
   (a) $\mathbb{M}_c(\mu_0, \mu_1) = \mathbb{J}_c(g_0, g_1)$*, and* T^* *solves* (MP).
   (b) $\mathbb{K}_c(\mu_0, \mu_1) = \mathbb{J}_c(g_0, g_1)$*, and* $\Gamma^\star = (\mathrm{id}_\mathcal{X}, T^*)_\# \mu_0$ *solves* (KTP).

Proof of Theorem 4.1. We first show that (MP) provides a lower bound on (LGOP), and this proves the first claim in the theorem statement. To this end, choose any h feasible in (LGOP), and define $T = g_1 \circ h \circ g_0^{-1}$. By Lemma 3, $T \in \mathcal{T}(\mu_0, \mu_1)$. We next verify that objective value attained by h in (LGOP) is equivalent to the one attained by T in (MP):

$$\begin{aligned}
\int_\mathcal{X} c(g_0(\boldsymbol{x}), (g_1 \circ h)(\boldsymbol{x})) \mathrm{d}\rho(\boldsymbol{x}) &= \int_\mathcal{X} c(g_0(\boldsymbol{x}), (g_1 \circ g_1^{-1} \circ T \circ g_0)(\boldsymbol{x})) \mathrm{d}\rho(\boldsymbol{x}) \\
&= \int_\mathcal{X} c(g_0(\boldsymbol{x}), (T \circ g_0)(\boldsymbol{x})) \mathrm{d}\rho(\boldsymbol{x}) \\
&= \int_\mathcal{X} c(\boldsymbol{x}, T(\boldsymbol{x})) \mathrm{d}\mu_0(\boldsymbol{x}).
\end{aligned}$$

Taking the infimum over h in both sides of the equality implies that (MP) provides a lower bound on (LGOP). This observation proves assertion (i).

As any $h \in \mathrm{Stab}_\mathcal{G}(\rho)$ satisfies $h_\# \rho = \rho$, assertion (ii) follows by Lemma 3.

In the remainder of the proof, assume that c satisfies Assumption 1, and there exists a c-convex function φ such that $T^\star = g_1 \circ h^\star \circ g_0^{-1}$ satisfies ($\star$) μ_0-almost surely. By Theorem 2.1, the Monge map is unique μ_0-a.e. and is characterized by ($\star$). Since $T^\star$ is admissible and satisfies the same first-order condition with a c-convex potential, it coincides μ_0-a.e. with the Monge map and hence solves (MP). Next, we relate its value to (LGOP):

$$\begin{aligned}
\mathbb{M}_c(\mu_0, \mu_1) &= \int_\mathcal{X} c(\boldsymbol{x}, T^\star(\boldsymbol{x})) \mathrm{d}\mu_0(\boldsymbol{x}) \\
&= \int_\mathcal{X} c(g_0(\boldsymbol{x}), (T^\star \circ g_0)(\boldsymbol{x})) \mathrm{d}\rho(\boldsymbol{x}) \\
&= \int_\mathcal{X} c(g_0(\boldsymbol{x}), (g_1 \circ h^\star)(\boldsymbol{x})) \mathrm{d}\rho(\boldsymbol{x}) = \mathbb{J}_c(g_0, g_1),
\end{aligned}$$

where the second line uses $\mu_0 = g_{0\#}\rho$, and the third uses the definition of $T^\star$ and that $h^\star$ solves (LGOP). This proves (iii) (a). Finally, Theorem 2.1 (iv) yields that the optimal coupling is concentrated on the graph of $T^\star$, i.e., $\Gamma^\star = (\mathrm{id}_\mathcal{X}, T^\star)_\# \mu_0$, proving (iii) (b). □

5. Examples

We illustrate how common distribution families arise as Lie group orbits and when Theorem 4.1 yields an explicit Monge map. For each family of distributions, we proceed in two steps:

(i) exhibit a reference measure ρ,

(ii) identify a finite-dimensional Lie group $\mathcal{G} \subset \mathrm{Diff}(\mathbb{R}^d)$.

We then show that the orbit $\mathcal{G}_{\#}\rho$ coincides with the family and identify the stabilizer subgroup $\mathrm{Stab}_{\mathcal{G}}(\rho)$. Throughout this section the ground cost is the squared Euclidean norm. For every concrete orbit we therefore solve the reduced problem (LGOP). Whenever this optimization admits an analytic minimizer, we construct a candidate transportation map feasible for (MP), verify the existence of a c-convex potential satisfying ($\star$), and conclude that (MP) and (LGOP) are equivalent, obtaining a closed-form solutions for the Monge problem (MP) and Kantorovich problem (KTP).

5.1. **Non-degenerate multivariate Gaussian distributions:** We denote the multivariate Gaussian distribution with mean $\boldsymbol{m} \in \mathbb{R}^d$ and covariance $\boldsymbol{\Sigma} \in \mathbb{S}^d_{++}$ by $\mathcal{N}(\boldsymbol{m}, \boldsymbol{\Sigma})$.

(i) We let the reference distribution to be the standard Normal distribution, that is, $\rho = \mathcal{N}(\mathbf{0}_d, \boldsymbol{I}_d)$ with probability density function

$$r(\boldsymbol{x}) = (2\pi)^{-d/2} \exp\left(-1/2\|\boldsymbol{x}\|^2\right).$$

(ii) The acting group is the affine group:

$$\mathcal{G}_{\text{aff}} = \left\{ f_{(\boldsymbol{m},\boldsymbol{A})} : \mathbb{R}^d \to \mathbb{R}^d \mid f_{(\boldsymbol{m},\boldsymbol{A})}(\boldsymbol{z}) = \boldsymbol{m} + \boldsymbol{A}\boldsymbol{z}, \boldsymbol{m} \in \mathbb{R}^d, \boldsymbol{A} \in \mathrm{GL}(d) \right\},$$

with composition

$$(f_{(\boldsymbol{m},\boldsymbol{A})} \circ f_{(\tilde{\boldsymbol{m}},\tilde{\boldsymbol{A}})})(\boldsymbol{z}) = f_{(\boldsymbol{m},\boldsymbol{A})}(\tilde{\boldsymbol{m}} + \tilde{\boldsymbol{A}}\boldsymbol{z}) = \boldsymbol{m} + \boldsymbol{A}(\tilde{\boldsymbol{m}} + \tilde{\boldsymbol{A}}\boldsymbol{z}) = f_{(\boldsymbol{m}+\boldsymbol{A}\tilde{\boldsymbol{m}}, \boldsymbol{A}\tilde{\boldsymbol{A}})}(\boldsymbol{z}).$$

Lemma 4. *For any* $\boldsymbol{\Sigma} \in \mathbb{S}^d_{++}$, $f_{(\boldsymbol{m},\boldsymbol{\Sigma}^{1/2})}{}_{\#}\rho = \mathcal{N}(\boldsymbol{m}, \boldsymbol{\Sigma})$ *and* $\{g_{\#}\rho \mid g \in \mathcal{G}_{\text{aff}}\} = \{\mathcal{N}(\boldsymbol{m}, \boldsymbol{\Sigma}) \mid \boldsymbol{m} \in \mathbb{R}^d, \boldsymbol{\Sigma} \in \mathbb{S}^d_{++}\}$.

Proof of Lemma 4. For some $\boldsymbol{m} \in \mathbb{R}^d$ and $\boldsymbol{A} \in \mathrm{GL}(d)$, we choose $f_{(\boldsymbol{m},\boldsymbol{A})} \in \mathcal{G}_{\text{aff}}$, then for every $\mathcal{A} \in \mathcal{B}(\mathbb{R}^d)$, we have

$$\begin{aligned}
f_{(\boldsymbol{m},\boldsymbol{A})}{}_{\#}\rho(\mathcal{A}) &= \rho(f^{-1}_{(\boldsymbol{m},\boldsymbol{A})}(\mathcal{A})) \\
&= \int_{\mathcal{A}} r\left(f^{-1}_{(\boldsymbol{m},\boldsymbol{A})}(\boldsymbol{x})\right) \left|\det(Df^{-1}_{(\boldsymbol{m},\boldsymbol{A})}(\boldsymbol{x}))\right| \mathrm{d}\boldsymbol{x} \\
&= \int_{\mathcal{A}} r\left(\boldsymbol{A}^{-1}(\boldsymbol{x} - \boldsymbol{m})\right) |\det(\boldsymbol{A}^{-1})| \mathrm{d}\boldsymbol{x} \\
&= \int_{\mathcal{A}} (2\pi)^{-d/2} \exp\left(-1/2(\boldsymbol{x} - \boldsymbol{m})^\top (\boldsymbol{A}\boldsymbol{A}^\top)^{-1}(\boldsymbol{x} - \boldsymbol{m})\right) (\det(\boldsymbol{A}\boldsymbol{A}^\top))^{-1/2} \mathrm{d}\boldsymbol{x},
\end{aligned}$$

where the first equality follows by the definition of the push-forward operator, the second equality follows by the multivariate change-of-variables formula, the third equality follows because $f^{-1}_{(\boldsymbol{m},\boldsymbol{A})}(\boldsymbol{x}) = \boldsymbol{A}^{-1}(\boldsymbol{x} - \boldsymbol{m})$ and $Df^{-1}_{(\boldsymbol{m},\boldsymbol{A})}(\boldsymbol{x}) = \boldsymbol{A}^{-1}$. The final equality is due to definition of r. The final representation above is the probability density function of a Gaussian distribution with mean $\boldsymbol{m}$ and covariance $\boldsymbol{A}\boldsymbol{A}^\top$ evaluated at the point $\boldsymbol{x}$, and thus we may conclude that $f_{(\boldsymbol{m},\boldsymbol{A})}{}_{\#}\rho = \mathcal{N}(\boldsymbol{m}, \boldsymbol{A}\boldsymbol{A}^\top)$. This implies that $\{g_{\#}\rho \mid g \in \mathcal{G}_{\text{aff}}\} \subseteq \mathcal{N}(\boldsymbol{m}, \boldsymbol{\Sigma}) \mid \boldsymbol{m} \in \mathbb{R}^d, \boldsymbol{\Sigma} \in \mathbb{S}^d_{++}\}$. Conversely, for any $\mathcal{N}(\boldsymbol{m}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} \in \mathbb{S}^d_{++}$, choose $\boldsymbol{\Sigma}^{1/2} \in \mathrm{GL}(d)$; then $f_{(\boldsymbol{m},\boldsymbol{\Sigma}^{1/2})}{}_{\#}\rho = \mathcal{N}(\boldsymbol{m}, \boldsymbol{\Sigma})$, so the reverse inclusion holds. □

Lemma 5. *The stabilizer of* $\rho = \mathcal{N}(\mathbf{0}_d, \boldsymbol{I}_d)$ *under* $\mathcal{G}_{\text{aff}}$ *is* $\mathrm{Stab}_{\mathcal{G}_{\text{aff}}}(\rho) = \{f_{(\mathbf{0}_d,\boldsymbol{Q})} \mid \boldsymbol{Q} \in \mathcal{O}(d)\}$.

Proof of Lemma 5. Next, fix an orthogonal matrix $\boldsymbol{Q} \in \mathcal{O}(d)$ and consider the group element $f_{(\boldsymbol{0}_d,\boldsymbol{Q})}\colon \boldsymbol{z} \mapsto \boldsymbol{Q}\boldsymbol{z}$. Its inverse is $f^{-1}_{(\boldsymbol{0}_d,\boldsymbol{Q})}(\boldsymbol{x}) = \boldsymbol{Q}^\top \boldsymbol{x}$ and $\det(\boldsymbol{Q}) = \pm 1$ as $\boldsymbol{Q} \in \mathcal{O}(d)$. Applying the multivariate change-of-variables formula, for every $\mathcal{A} \in \mathcal{B}(\mathbb{R}^d)$, we obtain

$$\begin{aligned}
\big(f_{(\boldsymbol{0}_d,\boldsymbol{Q})}\big)_{\#}\,\rho(\mathcal{A}) &= \rho(f^{-1}_{(\boldsymbol{0}_d,\boldsymbol{Q})}(\mathcal{A})) \\
&= \int_{\mathcal{A}} r\left(\boldsymbol{Q}^\top\boldsymbol{x}\right)|\det(\boldsymbol{Q}^\top)|\mathrm{d}\boldsymbol{x} \\
&= \int_{\mathcal{A}} (2\pi)^{-d/2}\exp\left(-\tfrac{1}{2}\|\boldsymbol{Q}^\top\boldsymbol{x}\|^2\right)\mathrm{d}\boldsymbol{x} \\
&= \int_{\mathcal{A}} (2\pi)^{-d/2}\exp\left(-\tfrac{1}{2}\|\boldsymbol{x}\|^2\right)\mathrm{d}\boldsymbol{x} = \rho(\mathcal{A}),
\end{aligned}$$

where the third equality follows because $\boldsymbol{Q}^\top\boldsymbol{Q} = \boldsymbol{I}_d$. Hence, every $f_{(\boldsymbol{0}_d,\boldsymbol{Q})}$ leaves ρ invariant, i.e., $f_{(\boldsymbol{0}_d,\boldsymbol{Q})} \in \mathrm{Stab}_{\mathcal{G}_{\mathrm{aff}}}(\rho)$. Conversely, suppose $f_{(\boldsymbol{m},\boldsymbol{A})} \in \mathrm{Stab}_{\mathcal{G}_{\mathrm{aff}}}(\rho)$. Following the same reasoning in the proof of Lemma 4, $f_{(\boldsymbol{m},\boldsymbol{A})\#}\rho = \mathcal{N}(\boldsymbol{m}, \boldsymbol{A}\boldsymbol{A}^\top)$. As $\rho = \mathcal{N}(\boldsymbol{0}_d, \boldsymbol{I}_d)$, uniqueness of Gaussian parameters forces $\boldsymbol{m} = \boldsymbol{0}_d$ and $\boldsymbol{A}\boldsymbol{A}^\top = \boldsymbol{I}_d$, i.e., $\boldsymbol{A} \in \mathcal{O}(d)$. This proves the claim. □

Lemma 6. *Suppose $g_i = f_{(\boldsymbol{m}_i,\boldsymbol{\Sigma}_i^{1/2})} \in \mathcal{G}_{\mathrm{aff}}$ for $i = 0, 1$. Then, (LGOP) induced by the cost function $c(\boldsymbol{x},\boldsymbol{y}) = \|\boldsymbol{x}-\boldsymbol{y}\|^2$ is solved by $h^\star = f_{(\boldsymbol{0},\boldsymbol{V}\boldsymbol{U}^\top)}$ and admits the following closed form expression*

$$\mathbb{J}_c(g_0,g_1) = \|\boldsymbol{m}_0 - \boldsymbol{m}_1\|^2 + \mathrm{Tr}\left(\boldsymbol{\Sigma}_0 + \boldsymbol{\Sigma}_1 - 2(\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}_1^{1/2})^{1/2}\right), \tag{1}$$

where $\boldsymbol{U}$ and $\boldsymbol{V}$ are left and right singular-vector matrices of $\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_1^{1/2}$.

Proof of Lemma 6. By Lemma 5, the stabilizer of $\rho = \mathcal{N}(\boldsymbol{0}_d, \boldsymbol{I}_d)$ in the affine group is $\mathrm{Stab}_{\mathcal{G}_{\mathrm{aff}}}(\rho) = \{h_{\boldsymbol{Q}} : \boldsymbol{z} \mapsto \boldsymbol{Q}\boldsymbol{z} \mid \boldsymbol{Q} \in \mathcal{O}(d)\}$. With $g_i(\boldsymbol{z}) = \boldsymbol{m}_i + \boldsymbol{\Sigma}_i^{1/2}\boldsymbol{z}$, $i = 0, 1$, and $c(\boldsymbol{x},\boldsymbol{y}) = \|\boldsymbol{x}-\boldsymbol{y}\|^2$, we have

$$\begin{aligned}
\mathbb{J}_c(g_0,g_1) &= \inf_{h\in\mathrm{Stab}_{\mathcal{G}_{\mathrm{aff}}}(\rho)} \int_{\mathbb{R}^d} \big\|g_0(\boldsymbol{z}) - (g_1\circ h)(\boldsymbol{z})\big\|^2 \mathrm{d}\rho(\boldsymbol{z}) \\
&= \min_{\boldsymbol{Q}\in\mathcal{O}(d)} \mathbb{E}_{\boldsymbol{z}\sim\rho}\left[\big\|\boldsymbol{m}_0 + \boldsymbol{\Sigma}_0^{1/2}\boldsymbol{z} - (\boldsymbol{m}_1 + \boldsymbol{\Sigma}_1^{1/2}\boldsymbol{Q}\boldsymbol{z})\big\|^2\right].
\end{aligned}$$

Next, set $\boldsymbol{a} = \boldsymbol{m}_0 - \boldsymbol{m}_1$ and $\boldsymbol{A} = \boldsymbol{\Sigma}_0^{1/2} - \boldsymbol{\Sigma}_1^{1/2}\boldsymbol{Q}$. Since $\mathbb{E}_{\boldsymbol{z}\sim\rho}[\boldsymbol{z}] = \boldsymbol{0}$ and $\mathbb{E}_{\boldsymbol{z}\sim\rho}[\boldsymbol{z}\boldsymbol{z}^\top] = \boldsymbol{I}_d$, we have

$$\mathbb{E}_{\boldsymbol{z}\sim\rho}\big[\|\boldsymbol{a} + \boldsymbol{A}\boldsymbol{z}\|^2\big] = \|\boldsymbol{a}\|^2 + \mathbb{E}_{\boldsymbol{z}\sim\rho}\big[\|\boldsymbol{A}\boldsymbol{z}\|^2\big] = \|\boldsymbol{a}\|^2 + \mathrm{Tr}(\boldsymbol{A}^\top\boldsymbol{A}).$$

Hence

$$\mathbb{J}_c(g_0,g_1) = \min_{\boldsymbol{Q}\in\mathcal{O}(d)} \|\boldsymbol{m}_0 - \boldsymbol{m}_1\|^2 + \mathrm{Tr}\left((\boldsymbol{\Sigma}_0^{1/2} - \boldsymbol{\Sigma}_1^{1/2}\boldsymbol{Q})^\top(\boldsymbol{\Sigma}_0^{1/2} - \boldsymbol{\Sigma}_1^{1/2}\boldsymbol{Q})\right).$$

Using the cyclic property of the trace,

$$\begin{aligned}
&\mathrm{Tr}\left((\boldsymbol{\Sigma}_0^{1/2} - \boldsymbol{\Sigma}_1^{1/2}\boldsymbol{Q})^\top(\boldsymbol{\Sigma}_0^{1/2} - \boldsymbol{\Sigma}_1^{1/2}\boldsymbol{Q})\right) \\
&\quad = \mathrm{Tr}(\boldsymbol{\Sigma}_0) - \mathrm{Tr}(\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{Q}) - \mathrm{Tr}(\boldsymbol{Q}^\top\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{\Sigma}_0^{1/2}) + \mathrm{Tr}(\boldsymbol{Q}^\top\boldsymbol{\Sigma}_1\boldsymbol{Q}) \\
&\quad = \mathrm{Tr}(\boldsymbol{\Sigma}_0) + \mathrm{Tr}(\boldsymbol{\Sigma}_1) - 2\,\mathrm{Tr}(\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{Q}\boldsymbol{\Sigma}_0^{1/2}),
\end{aligned}$$

where the second equality holds because $\mathrm{Tr}(\boldsymbol{Q}^\top \boldsymbol{\Sigma}_1 \boldsymbol{Q}) = \mathrm{Tr}(\boldsymbol{\Sigma}_1)$ and $\mathrm{Tr}(\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{Q}) = \mathrm{Tr}(\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{Q}\boldsymbol{\Sigma}_0^{1/2}) = \mathrm{Tr}\left((\boldsymbol{Q}^\top\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{\Sigma}_0^{1/2})^\top\right)$. Therefore, we arrive at

$$\mathbb{J}_c(g_0, g_1) = \min_{\boldsymbol{Q}\in\mathcal{O}(d)} \Psi(\boldsymbol{Q}) = \|\boldsymbol{m}_0 - \boldsymbol{m}_1\|^2 + \mathrm{Tr}\,(\boldsymbol{\Sigma}_0 + \boldsymbol{\Sigma}_1) - 2\,\mathrm{Tr}\left(\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{Q}\boldsymbol{\Sigma}_0^{1/2}\right). \tag{2}$$

Next, we write the singular value decomposition of $\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_1^{1/2} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^\top$, where $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$ is a d-dimensional diagonal matrix and $\boldsymbol{U}, \boldsymbol{V} \in \mathcal{O}(d)$. By the von Neumann's trace inequality, we have

$$\mathrm{Tr}(\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{Q}\boldsymbol{\Sigma}_0^{1/2}) \le \left|\mathrm{Tr}(\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{Q}\boldsymbol{\Sigma}_0^{1/2})\right| \le \sum_{i=1}^{d} \sigma_i(\boldsymbol{Q})\lambda_i = \mathrm{Tr}(\boldsymbol{\Lambda}), \tag{3}$$

where the last inequality follows because singular values of a square orthogonal matrix are all equivalent to 1. Now, we set $\boldsymbol{Q}^\star = \boldsymbol{V}\boldsymbol{U}^\top$. Next, evaluate

$$\mathrm{Tr}(\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{Q}^\star\boldsymbol{\Sigma}_0^{1/2}) = \mathrm{Tr}(\boldsymbol{V}\boldsymbol{U}^\top\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^\top) = \mathrm{Tr}(\boldsymbol{\Lambda}).$$

As $\boldsymbol{Q}^\star$ attains the upper bound in (3), and $\boldsymbol{Q}^\star \in \mathcal{O}(d)$, it solves (2).

Next, we evaluate $\mathbb{J}_c(g_0, g_1)$:

$$\begin{aligned}\mathbb{J}_c(g_0, g_1) &= \min_{\boldsymbol{Q}\in\mathcal{O}(d)} \Psi(\boldsymbol{Q}) \\ &= \|\boldsymbol{m}_0 - \boldsymbol{m}_1\|^2 + \mathrm{Tr}\left(\boldsymbol{\Sigma}_0 + \boldsymbol{\Sigma}_1 - 2\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{Q}^\star\boldsymbol{\Sigma}_0^{1/2}\right) \\ &= \|\boldsymbol{m}_0 - \boldsymbol{m}_1\|^2 + \mathrm{Tr}\left(\boldsymbol{\Sigma}_0 + \boldsymbol{\Sigma}_1 - 2(\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}_1^{1/2})^{1/2}\right),\end{aligned}$$

where the last equality follows because

$$\begin{aligned}\mathrm{Tr}(\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{Q}^\star\boldsymbol{\Sigma}_0^{1/2}) &= \mathrm{Tr}(\boldsymbol{\Lambda}) = \mathrm{Tr}\left(((\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{U}^\top)(\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^\top))^{1/2}\right) \\ &= \mathrm{Tr}\left(\left((\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{\Sigma}_0^{1/2})^\top(\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{\Sigma}_0^{1/2})\right)^{1/2}\right) = \mathrm{Tr}\left((\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{1/2})^{1/2}\right).\end{aligned}$$

Finally, by definition of (LGOP), $h^\star = f_{(\boldsymbol{0},\boldsymbol{V}\boldsymbol{U}^\top)}$ solves (LGOP). □

Lemma 7. *Suppose that* $g_i = f_{(\boldsymbol{m}_i, \boldsymbol{\Sigma}_i^{1/2})} \in \mathcal{G}_{\mathrm{aff}}$ *for* $i = 0, 1$ *and* $c(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|^2$. *Then,* $T^\star = g_1 \circ h^\star \circ g_0^{-1}$, *where* $h^\star$ *is as defined in Lemma 6, admits the following form*

$$T^\star(\boldsymbol{x}) = \boldsymbol{m}_1 + \boldsymbol{T}(\boldsymbol{x} - \boldsymbol{m}_0), \quad \text{where} \quad \boldsymbol{T} = \boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{1/2})^{1/2}\boldsymbol{\Sigma}_0^{-1/2}. \tag{4}$$

Additionally, $T^\star$ *satisfies* ($\star$) *for some* c*-convex function* φ.

Proof of Lemma 7. We write the singular value decomposition of $\boldsymbol{M} = \boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_1^{1/2}$ as $\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^\top$, where $\boldsymbol{\Lambda}$ is a d-dimensional diagonal matrix. By Lemma 6, $h^\star = f_{(\boldsymbol{0},\boldsymbol{V}\boldsymbol{U}^\top)}$ and as $T^\star = g_1 \circ h^\star \circ g_0^{-1}$, we have

$$T^\star(\boldsymbol{x}) = \boldsymbol{m}_1 + \boldsymbol{T}(\boldsymbol{x} - \boldsymbol{m}_0),$$

where $\boldsymbol{T} = \boldsymbol{\Sigma}_1^{1/2}\boldsymbol{V}\boldsymbol{U}^\top\boldsymbol{\Sigma}_0^{-1/2}$. Multiply both sides of $\boldsymbol{T}$ by $\boldsymbol{\Sigma}_0^{1/2}$ and refer to the resulting matrix as $\boldsymbol{C}$:

$$\boldsymbol{C} = \boldsymbol{\Sigma}_0^{1/2}\boldsymbol{T}\boldsymbol{\Sigma}_0^{1/2} = \boldsymbol{M}\boldsymbol{V}\boldsymbol{U}^\top = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top.$$

Noting that $\boldsymbol{M}\boldsymbol{M}^\top = \boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{1/2} = \boldsymbol{U}\boldsymbol{\Lambda}^2\boldsymbol{U}^\top$ and $\boldsymbol{C}^2 = \boldsymbol{U}\boldsymbol{\Lambda}^2\boldsymbol{U}^\top$, we may conclude that $\boldsymbol{C} = (\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{1/2})^{1/2}$. All in all, we can express $\boldsymbol{T}$ as

$$\boldsymbol{T} = \boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{C}\boldsymbol{\Sigma}_0^{-1/2} = \boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{1/2})^{1/2}\boldsymbol{\Sigma}_0^{-1/2}.$$

Recall that $\boldsymbol{\Sigma}_1 \succ 0$, and thus $\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{1/2} \succ 0$ as similarity transformations preserve eigenvalues. Additionally, $(\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{1/2})^{1/2} \succ 0$ as matrix square root preserves positive definiteness. Finally, because $\boldsymbol{\Sigma}_0^{-1/2} \succ 0$, $\boldsymbol{T} \succ 0$.

Next, define $\varphi(\boldsymbol{x}) = -\|\boldsymbol{x}\|^2 + 2\boldsymbol{m}_1^\top\boldsymbol{x} + (\boldsymbol{x}-\boldsymbol{m}_0)^\top\boldsymbol{T}(\boldsymbol{x}-\boldsymbol{m}_0)$ and evaluate its gradient

$$\nabla_{\boldsymbol{x}}\varphi(\boldsymbol{x}) = -2\boldsymbol{x} + 2\boldsymbol{m}_1 + 2\boldsymbol{T}(\boldsymbol{x}-\boldsymbol{m}_0).$$

Because $\boldsymbol{T} \succ 0$, $\nabla^2_{\boldsymbol{x}}\varphi(\boldsymbol{x}) = -2\boldsymbol{I}_d + 2\boldsymbol{T} \succeq -2\boldsymbol{I}_d$. Consequently, by (Villani, 2008, Example 13.6) φ is a c-convex function. Next, we evaluate ($\star$):

$$-2\boldsymbol{x} + 2\boldsymbol{m}_1 + 2\boldsymbol{T}(\boldsymbol{x}-\boldsymbol{m}_0) + 2(\boldsymbol{x} - T^\star(\boldsymbol{x})) = 0$$

Hence, the criteria in ($\star$) is satisfied, and this observation completes our proof. □

Theorem 5.1. *Suppose that* $\mu_i = \mathcal{N}(\boldsymbol{m}_i, \boldsymbol{\Sigma}_i)$ *for some* $\boldsymbol{m}_i \in \mathbb{R}^d$ *and* $\boldsymbol{\Sigma}_i \in \mathbb{S}^d_{++}$, $i = 0, 1$, *and* $c(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x}-\boldsymbol{y}\|^2$. *Then, we have*

(i) $\mathbb{M}_c(\mu_0, \mu_1) = \mathbb{J}_c(g_0, g_1)$, *and* $T^\star$ *as expressed in* (4) *solves* (MP),

(ii) $\mathbb{K}_c(\mu_0, \mu_1) = \mathbb{J}_c(g_0, g_1)$, *and* $(\mathrm{id}_{\mathcal{X}}, T^\star)_{\#}\mu_0$ *solves* (KTP),

where $g_i = f_{(\boldsymbol{m}_i, \boldsymbol{\Sigma}_i^{1/2})} \in \mathcal{G}_{\mathrm{aff}}$.

Proof of Theorem 5.1. By Lemma 4, we have $f_{(\boldsymbol{m}_i, \boldsymbol{\Sigma}_i^{1/2})}{}_{\#}\rho = \mu_i$ for $i = 0, 1$; hence $\mu_0, \mu_1 \in \mathcal{G}_{\mathrm{aff}\,\#}\rho$. Define $T^\star = g_1 \circ h^\star \circ g_0^{-1}$, where $h^\star$ is the map in Lemma 6. Then, by Lemma 7, $T^\star$ satisfies the requirement in Theorem 4.1 (iii). Consequently, the assertions (i) and (ii) follow directly from Theorem 4.1 (iii)-(a) and (iii)-(b). □

Theorem 5.1 recovers the closed-form solutions provided in (Olkin and Pukelsheim, 1982; Dowson and Landau, 1982; Gelbrich, 1990) from the Lie group viewpoint.

Remark 1 (Degenerate Gaussian distributions)**.** *The affine-orbit argument in §5.1 uses* $\mathrm{GL}(d)$, *hence it assumes* $\boldsymbol{\Sigma}_i \succ 0$. *If* $\boldsymbol{\Sigma}_i \succeq 0$, *then* μ_i *is singular (supported on a proper affine subspace), so a diffeomorphism-based action cannot produce such measures, and the Monge-map existence result we invoke, which requires absolute continuity, need not apply. Nevertheless, the closed-form value in* (1) *extends continuously to* $\boldsymbol{\Sigma}_i \succeq 0$*: the the principal square-root map* $\boldsymbol{A} \mapsto \boldsymbol{A}^{1/2}$ *is continuous on* $\mathbb{S}^d_+$, *and trace and matrix products are continuous. Thus, writing* $\boldsymbol{\Sigma}_i^{(\varepsilon)} = \boldsymbol{\Sigma}_i + \varepsilon\boldsymbol{I} \succ 0$, *the value at* $\boldsymbol{\Sigma}_i^{(\varepsilon)}$ *converges to the same expression with* $\boldsymbol{\Sigma}_i$ *as* $\varepsilon \downarrow 0$. *In particular,* (1) *remains valid for semidefinite covariances, although a Monge map in* $\mathrm{Diff}(\mathbb{R}^d)$ *may fail to exist.*

Remark 2 (Elliptical families as an affine orbit)**.** *A random vector* $\boldsymbol{x} \in \mathbb{R}^d$ *has an elliptical distribution with characteristic generator* $\varsigma : \mathbb{R}_+ \to \mathbb{R}$ *if its characteristic function is* $\exp(i\boldsymbol{t}^\top\boldsymbol{m})\varsigma(\boldsymbol{t}^\top\boldsymbol{S}\boldsymbol{t})$, *for some location* $\boldsymbol{m} \in \mathbb{R}^d$ *and dispersion* $\boldsymbol{S} \in \mathbb{S}^d_{++}$*; and we write* $\boldsymbol{x} \sim \mathcal{E}_\varsigma(\boldsymbol{m}, \boldsymbol{S})$. *When* $\varsigma'(0)$ *exists and finite, then the covariance matrix of* $\boldsymbol{x} \sim \mathcal{E}_\varsigma(\boldsymbol{m}, \boldsymbol{S})$, *satisfies* $\mathrm{Cov}(\mathcal{E}_\varsigma(\boldsymbol{m}, \boldsymbol{S})) = (-2\varsigma'(0))\boldsymbol{S}$. *Reparameterizing the generator* ς *as* $\tilde{\varsigma}(s) = \varsigma(s/(-2\varsigma'(0)))$ *yields* $\mathrm{Cov}(\mathcal{E}_{\tilde{\varsigma}}(\boldsymbol{m}, \boldsymbol{S})) = \boldsymbol{S}$. *With* $\rho = \mathcal{E}_{\tilde{\varsigma}}(\boldsymbol{0}_d, \boldsymbol{I}_d)$, *the affine orbit equals the entire elliptical family (with positive definite* *dispersion matrices) sharing the generator:*

$$\{g_{\#}\rho \mid g \in \mathcal{G}_{\text{aff}}\} = \{\mathcal{E}_{\tilde{\varsigma}}(\boldsymbol{m}, \boldsymbol{S}) \mid \boldsymbol{m} \in \mathbb{R}^d, \boldsymbol{S} \in \mathbb{S}^d_{++}\}.$$

Consequently, the Gaussian orbit calculation in Section 5.1 carries over verbatim, yielding the same closed-form value and optimal map in Theorem 5.1 for any two members of the elliptical family with generator $\tilde{\varsigma}$ that has finite c-moments. This recovers (Gelbrich, 1990, Theorem 2.1) from the Lie group perspective.

5.2. **Wishart distributions.** The d-dimensional Wishart family is parameterized by a scale matrix $\boldsymbol{\Sigma} \in \mathbb{S}^d_{++}$ and degrees of freedom p. We denote this distribution by $\mathcal{W}_d(\boldsymbol{\Sigma}, p)$. The support of the Wishart distribution consists of positive semidefinite matrices. The probability density function of a random matrix $\boldsymbol{X} \sim \mathcal{W}_d(\boldsymbol{\Sigma}, p)$ is given by:

$$\boldsymbol{X} \mapsto \frac{\det(\boldsymbol{X})^{\frac{p-d-1}{2}} \exp\left(-\operatorname{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{X})/2\right)}{2^{\frac{dp}{2}} \det(\boldsymbol{\Sigma})^{\frac{p}{2}} \Gamma_d\left(\frac{p}{2}\right)},$$

where $\Gamma_d(\cdot)$ is the multivariate gamma function. When $d = 1$, the Wishart distribution is a scaled χ^2-distribution with p degrees of freedom and scale parameter $\boldsymbol{\Sigma} > 0$. We assume that $p \geq d$, and thus by (Muirhead, 2009, Theorem 3.1.4), support of $\mathcal{W}_d(\boldsymbol{\Sigma}, p)$ is $\mathbb{S}^d_{++}$.

(i) We select the reference measure as $\rho = \mathcal{W}_d(\boldsymbol{I}_d, p)$ with density r.
(ii) The acting group is the congruence action:

$$\mathcal{G}_{\text{cong}} = \{f_{\boldsymbol{G}} : \mathbb{S}^d_{++} \to \mathbb{S}^d_{++} \mid f_{\boldsymbol{G}}(\boldsymbol{Z}) = \boldsymbol{G}\boldsymbol{Z}\boldsymbol{G}^\top, \boldsymbol{G} \in \mathrm{GL}(d)\}$$

with composition

$$(f_{\boldsymbol{G}} \circ f_{\boldsymbol{H}})(\boldsymbol{Z}) = f_{\boldsymbol{G}}(\boldsymbol{H}\boldsymbol{Z}\boldsymbol{H}^\top) = \boldsymbol{G}\boldsymbol{H}\boldsymbol{Z}\boldsymbol{H}^\top\boldsymbol{G}^\top = f_{\boldsymbol{G}\boldsymbol{H}}(\boldsymbol{Z}).$$

The identity element of the group is $f_{\boldsymbol{I}_d}$.

Lemma 8. *For any $\boldsymbol{\Sigma} \in \mathbb{S}^d_{++}$, $f_{\boldsymbol{\Sigma}^{1/2}\#}\rho = \mathcal{W}_d(\boldsymbol{\Sigma}, p)$ and $\{g_{\#}\rho \mid g \in \mathcal{G}_{\text{cong}}\} = \{\mathcal{W}_d(\boldsymbol{\Sigma}, p) \mid \boldsymbol{\Sigma} \in \mathbb{S}^d_{++}\}$.*

Proof. Now, let $\boldsymbol{G} \in \mathrm{GL}(d)$, then for every $\mathcal{A} \in \mathcal{B}(\mathbb{S}^d_{++})$, we have

$$\begin{aligned}
\left(f_{\boldsymbol{G}\#}\rho\right)(\mathcal{A}) &= \rho(f_{\boldsymbol{G}}^{-1}(\mathcal{A})) \\
&= \int_{\mathcal{A}} r\left(f_{\boldsymbol{G}}^{-1}(\boldsymbol{X})\right) \left|\det(Df_{\boldsymbol{G}}^{-1}(\boldsymbol{X}))\right| \mathrm{d}\boldsymbol{X} \\
&= \int_{\mathcal{A}} \frac{\det\left(\boldsymbol{G}^{-1}\boldsymbol{X}\boldsymbol{G}^{-\top}\right)^{\frac{p-d-1}{2}} \exp(-\operatorname{Tr}((\boldsymbol{G}\boldsymbol{G}^\top)^{-1}\boldsymbol{X})/2)}{2^{\frac{dp}{2}}\Gamma_d\left(\frac{p}{2}\right)} |\det(\boldsymbol{G})|^{-(d+1)}\mathrm{d}\boldsymbol{X} \\
&= \int_{\mathcal{A}} \frac{\det(\boldsymbol{X})^{\frac{p-d-1}{2}} \exp\left(-\frac{1}{2}\operatorname{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{X})\right)}{2^{\frac{dp}{2}} \det(\boldsymbol{\Sigma})^{\frac{p}{2}} \Gamma_d\left(\frac{p}{2}\right)} \mathrm{d}\boldsymbol{X},
\end{aligned}$$

where $\boldsymbol{\Sigma} = \boldsymbol{G}\boldsymbol{G}^\top$ and the first equality follows by definition of the push-forward operator. The second equality follows by the change-of-variables formula. The third equality follows because $f_{\boldsymbol{G}}^{-1}(\boldsymbol{X}) = \boldsymbol{G}^{-1}\boldsymbol{X}\boldsymbol{G}^{-\top}$ and its Jacobian determinant on $\mathbb{S}^d_{++}$ is $|\det(Df_{\boldsymbol{G}}^{-1}(\boldsymbol{X}))| = |\det(\boldsymbol{G})|^{-(d+1)}$. Thus, we may conclude that $f_{\boldsymbol{G}\#}\rho = \mathcal{W}_d(\boldsymbol{\Sigma}, p)$, implying that

$$\{g_{\#}\rho \mid g \in \mathcal{G}_{\text{cong}}\} \subseteq \{\mathcal{W}_d(\boldsymbol{\Sigma}, p) \mid \boldsymbol{\Sigma} \in \mathbb{S}^d_{++}\},$$

so every Wishart distribution with degree of freedom p is the push-forward of the reference density ρ by a congruence map. Conversely, for every $\mathbf{\Sigma} \in \mathbb{S}^d_{++}$, we have $f_{\mathbf{\Sigma}^{1/2}\#}\rho = \mathcal{W}_d(\mathbf{\Sigma}, p)$. Thus, every Wishart distribution $\mathcal{W}_d(\mathbf{\Sigma}, p)$ is exactly represented by a congruence push-forward of the reference measure ρ. Therefore, the reverse inclusion holds:

$$\{\mathcal{W}_d(\mathbf{\Sigma}, p) \mid \mathbf{\Sigma} \in \mathbb{S}^d_{++}\} \subseteq \{g_{\#}\rho \mid g \in \mathcal{G}_{\text{cong}}\}.$$

Hence, the whole Wishart family with degree of freedom p is an orbit of ρ. □

Lemma 9. *The stabilizer of* $\rho = \mathcal{W}_d(\boldsymbol{I}_d, p)$ *under* $\mathcal{G}_{\text{cong}}$ *is* $\text{Stab}_{\mathcal{G}_{\text{cong}}}(\rho) = \{f_{\boldsymbol{G}} \mid \boldsymbol{G} \in \mathcal{O}(d)\}$.

Proof of Lemma 9. Let $\boldsymbol{G} \in \mathcal{O}(d)$, so $\boldsymbol{G}^\top \boldsymbol{G} = \boldsymbol{I}_d$ and $\det(\boldsymbol{G}) = \pm 1$. For any Borel set $\mathcal{A} \subset \mathbb{S}^d_{++}$, the change-of-variables formula for the congruence map yields

$$(f_{\boldsymbol{G}})_{\#}\,\rho(\mathcal{A}) = \int_{\mathcal{A}} r\left(f_{\boldsymbol{G}}^{-1}(\boldsymbol{X})\right) |\det(Df_{\boldsymbol{G}}^{-1}(\boldsymbol{X}))| \mathrm{d}\boldsymbol{X},$$

where r is the density of ρ. Because $f_{\boldsymbol{G}}^{-1}(\boldsymbol{X}) = \boldsymbol{G}^\top \boldsymbol{X} \boldsymbol{G}$, a simple calculation reveals $r(f_{\boldsymbol{G}}^{-1}(\boldsymbol{X})) = r(\boldsymbol{X})$. Noting that $|\det(Df_{\boldsymbol{G}}^{-1}(\boldsymbol{X}))| = |\det(\boldsymbol{G})|^{-(d+1)} = 1$ since $|\det(\boldsymbol{G})| = 1$, we conclude that $(f_{\boldsymbol{G}})_{\#}\rho = \rho$, so $f_{\boldsymbol{G}} \in \text{Stab}(\rho)$, which implies

$$\{f_{\boldsymbol{G}} : \boldsymbol{G} \in \mathcal{O}(d)\} \subseteq \text{Stab}_{\mathcal{G}_{\text{cong}}}(\rho).$$

Conversely, suppose $f_{\boldsymbol{G}} \in \text{Stab}_{\mathcal{G}_{\text{cong}}}(\rho)$ for some $\boldsymbol{G} \in \text{GL}(d)$. Writing λ for Lebesgue measure on $\text{Sym}(d)$ restricted to $\mathbb{S}^d_{++}$, the change-of-variables identity gives

$$r\left(\boldsymbol{G}^{-1}\boldsymbol{X}\boldsymbol{G}^{-\top}\right) \left|\det(Df_{\boldsymbol{G}}^{-1}(\boldsymbol{X}))\right| = r(\boldsymbol{X}) \qquad \lambda\text{-a.e. } \boldsymbol{X} \in \mathbb{S}^d_{++},$$

where $f_{\boldsymbol{G}}^{-1}(\boldsymbol{X}) = \boldsymbol{G}^{-1}\boldsymbol{X}\boldsymbol{G}^{-\top}$ and $\left|\det(Df_{\boldsymbol{G}}^{-1}(\boldsymbol{X}))\right| = |\det(\boldsymbol{G})|^{-(d+1)}$. Using

$$\det(\boldsymbol{G}^{-1}\boldsymbol{X}\boldsymbol{G}^{-\top}) = \det(\boldsymbol{X})|\det(\boldsymbol{G})|^{-2}$$

and cancelling the common normalizing factor $2^{-dp/2}\Gamma_d(p/2)^{-1}$, we obtain

$$|\det(\boldsymbol{G})|^{-p} \exp\Big(-\tfrac{1}{2}\operatorname{Tr}\big((\boldsymbol{G}\boldsymbol{G}^\top)^{-1} - \boldsymbol{I}_d\big)\boldsymbol{X}\Big) = 1 \qquad \lambda\text{-a.e. } \boldsymbol{X} \in \mathbb{S}^d_{++}.$$

Both sides are continuous in $\boldsymbol{X}$, hence the identity holds for all $\boldsymbol{X} \in \mathbb{S}^d_{++}$. Taking $\boldsymbol{X} = t\boldsymbol{I}_d$ with $t > 0$ gives

$$|\det(\boldsymbol{G})|^{-p} \exp\Big(-\tfrac{t}{2}\big(\operatorname{Tr}((\boldsymbol{G}\boldsymbol{G}^\top)^{-1}) - d\big)\Big) = 1 \quad \text{for all } t > 0. \tag{5}$$

Letting $t \to 0^+$ in (5) yields $|\det(\boldsymbol{G})|^{-p} = 1$, so $|\det(\boldsymbol{G})| = 1$. Differentiating (5) at $t = 0$ gives $\operatorname{Tr}((\boldsymbol{G}\boldsymbol{G}^\top)^{-1}) = d$. Set $\mathbf{\Sigma} = \boldsymbol{G}\boldsymbol{G}^\top \succ 0$; then $\det(\mathbf{\Sigma}) = 1$ and $\operatorname{Tr}(\mathbf{\Sigma}^{-1}) = d$. By AM–GM on the eigenvalues of $\mathbf{\Sigma}^{-1}$, equality forces $\mathbf{\Sigma}^{-1} = \boldsymbol{I}_d$, hence $\mathbf{\Sigma} = \boldsymbol{I}_d$. Therefore $\boldsymbol{G}\boldsymbol{G}^\top = \boldsymbol{I}_d$, i.e., $\boldsymbol{G} \in \mathcal{O}(d)$. Combining both directions concludes our proof. □

Lemma 10. *For some* $\boldsymbol{U}, \boldsymbol{V} \in \mathbb{R}^{d\times d}$*, we have*

$$\mathbb{E}_{\boldsymbol{X}\sim\rho}[\operatorname{Tr}(\boldsymbol{U}\boldsymbol{X}\boldsymbol{V}\boldsymbol{X})] = p\operatorname{Tr}(\boldsymbol{U})\operatorname{Tr}(\boldsymbol{V}) + p\operatorname{Tr}(\boldsymbol{U}\boldsymbol{V}^\top) + p^2\operatorname{Tr}(\boldsymbol{U}\boldsymbol{V}),$$

where $\rho = \mathcal{W}_d(\boldsymbol{I}_d, p)$.

Proof of Lemma 10. Suppose that $\boldsymbol{X} \sim \rho$, then $\boldsymbol{X}$ is in the following form

$$\boldsymbol{X} = \boldsymbol{Z}^\top\boldsymbol{Z}, \text{ where } \boldsymbol{Z} = [\boldsymbol{z}_1, \ldots, \boldsymbol{z}_p]^\top \in \mathbb{R}^{p\times d}, \qquad \boldsymbol{z}_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_d).$$

Next, we insert $\boldsymbol{X} = \boldsymbol{Z}^\top \boldsymbol{Z}$ twice,

$$\mathrm{Tr}(\boldsymbol{U}\boldsymbol{X}\boldsymbol{V}\boldsymbol{X}) = \mathrm{Tr}(\boldsymbol{U}\boldsymbol{Z}^\top \boldsymbol{Z}\boldsymbol{V}\boldsymbol{Z}^\top \boldsymbol{Z}).$$

Because the trace is invariant under cyclic permutations,

$$\mathrm{Tr}(\boldsymbol{U}\boldsymbol{Z}^\top \boldsymbol{Z}\boldsymbol{V}\boldsymbol{Z}^\top \boldsymbol{Z}) = \mathrm{Tr}(\boldsymbol{Z}\boldsymbol{U}\boldsymbol{Z}^\top \boldsymbol{Z}\boldsymbol{V}\boldsymbol{Z}^\top).$$

Then

$$(\boldsymbol{Z}\boldsymbol{U}\boldsymbol{Z}^\top)_{k\ell} = \boldsymbol{z}_k^\top \boldsymbol{U}\boldsymbol{z}_\ell, \qquad (\boldsymbol{Z}\boldsymbol{V}\boldsymbol{Z}^\top)_{\ell k} = \boldsymbol{z}_\ell^\top \boldsymbol{V}\boldsymbol{z}_k.$$

Hence the matrix product inside the trace has (k, ℓ) entry $(\boldsymbol{z}_k^\top \boldsymbol{U}\boldsymbol{z}_\ell)(\boldsymbol{z}_\ell^\top \boldsymbol{V}\boldsymbol{z}_k)$. Summing these diagonal entries yields

$$\mathrm{Tr}(\boldsymbol{U}\boldsymbol{X}\boldsymbol{V}\boldsymbol{X}) = \sum_{k=1}^{p}\sum_{\ell=1}^{p}(\boldsymbol{z}_k^\top \boldsymbol{V}\boldsymbol{z}_\ell)(\boldsymbol{z}_\ell^\top \boldsymbol{U}\boldsymbol{z}_k) = \sum_{k,\ell=1}^{p} T_{k\ell},$$

where we set

$$T_{k\ell} = (\boldsymbol{z}_k^\top \boldsymbol{V}\boldsymbol{z}_\ell)(\boldsymbol{z}_\ell^\top \boldsymbol{U}\boldsymbol{z}_k), \qquad k, \ell \in \{1, \dots, p\}.$$

Note that $T_{kk} = \boldsymbol{z}_k^\top \boldsymbol{V}\boldsymbol{z}_k \boldsymbol{z}_k^\top \boldsymbol{U}\boldsymbol{z}_k$, then we have

$$T_{kk} = \sum_{\substack{a=1,b=1,\\ c=1,e=1}}^{d} z_{k,a}V_{ab}z_{k,b}z_{k,c}U_{ce}z_{k,e}.$$

By Isserlis's theorem (Isserlis, 1918), for $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}_d, \boldsymbol{I}_d)$, we have

$$\begin{aligned}\mathbb{E}[z_a z_b z_c z_e] &= \mathbb{E}[z_a z_b]\mathbb{E}[z_c z_e] + \mathbb{E}[z_a z_c]\mathbb{E}[z_b z_e] + \mathbb{E}[z_a z_e]\mathbb{E}[z_b z_c] \\ &= \delta_{ab}\delta_{ce} + \delta_{ac}\delta_{be} + \delta_{ae}\delta_{bc},\end{aligned}$$

where

$$\delta_{ab} = \begin{cases} 1 & \text{if } a = b, \\ 0 & \text{otherwise,} \end{cases}$$

and the second equality above holds because $\mathbb{E}[z_a z_b]$ is equivalent to 1 if $a = b$ and 0, otherwise. Equipped with this formulation, we evaluate the expectation of T_{kk}:

$$\begin{aligned}\mathbb{E}[T_{kk}] &= \sum_{\substack{a=1,b=1,\\ c=1,e=1}}^{d} V_{ab}U_{ce}\left(\delta_{ab}\delta_{ce} + \delta_{ac}\delta_{be} + \delta_{ae}\delta_{bc}\right) \\ &= \sum_{a=1}^{d}\sum_{c=1}^{d} V_{aa}U_{cc} + \sum_{a=1}^{d}\sum_{b=1}^{d} V_{ab}U_{ab} + \sum_{a=1}^{d}\sum_{b=1}^{d} V_{ab}U_{ba},\end{aligned}$$

where the first term above is equivalent to $\mathrm{Tr}(\boldsymbol{V})\,\mathrm{Tr}(\boldsymbol{U})$, the second term $\mathrm{Tr}(\boldsymbol{U}\boldsymbol{V}^\top)$ and the last term is equivalent to $\mathrm{Tr}(\boldsymbol{U}\boldsymbol{V})$, implying

$$\mathbb{E}[T_{kk}] = \mathrm{Tr}(\boldsymbol{U})\,\mathrm{Tr}(\boldsymbol{V}) + \mathrm{Tr}(\boldsymbol{U}\boldsymbol{V}^\top) + \mathrm{Tr}(\boldsymbol{U}\boldsymbol{V}).$$

Because there are p diagonal indices $k = 1, \dots, p$, the total diagonal contribution of T_{kk} to $\mathbb{E}[\mathrm{Tr}(\boldsymbol{U}\boldsymbol{X}\boldsymbol{V}\boldsymbol{X})]$ is

$$p\left(\mathrm{Tr}(\boldsymbol{U})\,\mathrm{Tr}(\boldsymbol{V}) + \mathrm{Tr}(\boldsymbol{U}\boldsymbol{V}^\top) + \mathrm{Tr}(\boldsymbol{U}\boldsymbol{V})\right).$$

Now, we will evaluate $\mathbb{E}[T_{k\ell}]$ when $k \neq \ell$. Independence of $\boldsymbol{z}_k$ and $\boldsymbol{z}_\ell$ yields

$$\mathbb{E}[T_{k\ell}] = \mathrm{Tr}(\boldsymbol{U}\boldsymbol{V}),$$

and there are $p(p-1)$ such ordered pairs, contributing $p(p-1)\,\mathrm{Tr}(\boldsymbol{U}\boldsymbol{V})$.

Finally, adding diagonal and off-diagonal parts,

$$\begin{aligned}\mathbb{E}_{\boldsymbol{X}\sim\rho}\left[\mathrm{Tr}(\boldsymbol{UXVX})\right] &= p\,\mathrm{Tr}(\boldsymbol{U})\,\mathrm{Tr}(\boldsymbol{V}) + p\,\mathrm{Tr}(\boldsymbol{UV}^\top) + p\,\mathrm{Tr}(\boldsymbol{UV}) \\ &\quad + p(p-1)\,\mathrm{Tr}(\boldsymbol{UV}) \\ &= p\,\mathrm{Tr}(\boldsymbol{U})\,\mathrm{Tr}(\boldsymbol{V}) + p\,\mathrm{Tr}(\boldsymbol{UV}^\top) + p^2\,\mathrm{Tr}(\boldsymbol{UV}).\end{aligned}$$

This observation completes our proof. □

Theorem 5.2. *Suppose $g_i = f_{\boldsymbol{\Sigma}_i^{1/2}} \in \mathcal{G}_{\mathrm{cong}}$ for $i = 0, 1$. Then, orbit transport problem in (LGOP) induced by the cost function $c(\boldsymbol{X}, \boldsymbol{Y}) = \|\boldsymbol{X} - \boldsymbol{Y}\|_{\mathrm{F}}^2$ admits the following closed form expression*

$$\begin{aligned}\mathbb{J}_c(g_0, g_1) = p\left(\mathrm{Tr}(\boldsymbol{\Sigma}_0)^2 + \mathrm{Tr}(\boldsymbol{\Sigma}_1)^2 - 2\,\mathrm{Tr}(\boldsymbol{\Lambda})^2 - 2\,\mathrm{Tr}(\boldsymbol{\Lambda}^2)\right) - \\ 2p^2\,\mathrm{Tr}(\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}_1) + p(p+1)\left(\mathrm{Tr}(\boldsymbol{\Sigma}_0^2) + \mathrm{Tr}(\boldsymbol{\Sigma}_1^2)\right).\end{aligned}$$

Moreover, $h^\star = f_{\boldsymbol{VU}^\top}$ solves (LGOP), where $\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_1^{1/2}$ has the singular value decomposition $\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_1^{1/2} = \boldsymbol{U\Lambda V}^\top$.

Proof of Theorem 5.2. By Lemma 8, problem in (LGOP) is equivalent to

$$\min_{\boldsymbol{Q}\in\mathcal{O}(d)} \Psi(\boldsymbol{Q}), \tag{6}$$

where

$$\Psi(\boldsymbol{Q}) = \mathbb{E}_{\boldsymbol{X}\sim\rho}\left[\|\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{X}\boldsymbol{\Sigma}_0^{1/2} - \boldsymbol{\Sigma}_1^{1/2}\boldsymbol{QXQ}^\top\boldsymbol{\Sigma}_1^{1/2}\|_{\mathrm{F}}^2\right].$$

For brevity of expression, momentarily we define $Y(\boldsymbol{X}) = \boldsymbol{\Sigma}_0^{1/2}\boldsymbol{X}\boldsymbol{\Sigma}_0^{1/2}$ and $Z(\boldsymbol{X}) = \boldsymbol{\Sigma}_1^{1/2}\boldsymbol{QXQ}^\top\boldsymbol{\Sigma}_1^{1/2}$. Note that $Y(\boldsymbol{X}), \boldsymbol{Z}(\boldsymbol{X}) \in \mathbb{S}_{++}^d$ and we can rewrite $\Psi(\boldsymbol{Q})$ as

$$\Psi(\boldsymbol{Q}) = \mathbb{E}_{\boldsymbol{X}\sim\rho}[\mathrm{Tr}(Y(\boldsymbol{X})^2) + \mathrm{Tr}(Z(\boldsymbol{X})^2) - 2\,\mathrm{Tr}(Y(\boldsymbol{X})Z(\boldsymbol{X}))]. \tag{7}$$

By plugging in the expression of $\boldsymbol{Y}$, first term in (7) reads

$$\mathbb{E}_{\boldsymbol{X}\sim\rho}[\mathrm{Tr}(Y(\boldsymbol{X})^2)] = \mathbb{E}_{\boldsymbol{X}\sim\rho}[\mathrm{Tr}(\boldsymbol{\Sigma}_0\boldsymbol{X}\boldsymbol{\Sigma}_0\boldsymbol{X})] = p\,\mathrm{Tr}(\boldsymbol{\Sigma}_0)^2 + p(p+1)\,\mathrm{Tr}(\boldsymbol{\Sigma}_0^2), \tag{8}$$

where the second equality follows by Lemma 10.

Similarly, by plugging in the definition of $\boldsymbol{Z}$, the second term in (7), we have

$$\mathbb{E}_{\boldsymbol{X}\sim\rho}[\mathrm{Tr}(Z(\boldsymbol{X})^2)] = p\,\mathrm{Tr}(\boldsymbol{\Sigma}_1)^2 + p(p+1)\,\mathrm{Tr}(\boldsymbol{\Sigma}_1^2). \tag{9}$$

Finally, the third term in (7) reads

$$\begin{aligned}\mathbb{E}_{\boldsymbol{X}\sim\rho}[\mathrm{Tr}(Y(\boldsymbol{X})Z(\boldsymbol{X}))] &= \mathbb{E}_{\boldsymbol{X}\sim\rho}[\mathrm{Tr}(\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{X}\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{QXQ}^\top\boldsymbol{\Sigma}_1^{1/2})] \\ &= \mathbb{E}_{\boldsymbol{X}\sim\rho}[\mathrm{Tr}(\boldsymbol{XMQXQ}^\top\boldsymbol{M}^\top)] \\ &= p\,\mathrm{Tr}(\boldsymbol{MQ})^2 + p\,\mathrm{Tr}(\boldsymbol{MQMQ}) + p^2\,\mathrm{Tr}(\boldsymbol{MM}^\top) \qquad (10) \\ &\le p\,\mathrm{Tr}(\boldsymbol{\Lambda})^2 + p\,\mathrm{Tr}(\boldsymbol{MQMQ}) + p^2\,\mathrm{Tr}(\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}_1) \\ &\le p\,\mathrm{Tr}(\boldsymbol{\Lambda})^2 + p\,\mathrm{Tr}(\boldsymbol{\Lambda}^2) + p^2\,\mathrm{Tr}(\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}_1). \qquad (11)\end{aligned}$$

where $\boldsymbol{M} = \boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_1^{1/2}$, and its singular value decomposition is $\boldsymbol{M} = \boldsymbol{U\Lambda V}^\top$. The third equality follows by Lemma 10. The first inequality follows by von Neumann's trace inequality and the second inequality follows by Cauchy-Schwarz. Next, we will set $\boldsymbol{Q}$ to $\boldsymbol{Q}^\star = \boldsymbol{VU}^\top$ and evaluate (10):

$$p\,\mathrm{Tr}(\boldsymbol{U\Lambda V}^\top\boldsymbol{VU}^\top) + p\,\mathrm{Tr}(\boldsymbol{U\Lambda V}^\top\boldsymbol{VU}^\top\boldsymbol{U\Lambda V}^\top\boldsymbol{VU}^\top) + p^2\,\mathrm{Tr}((\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_1^{1/2})^2),$$

which is equivalent to the upper bound on (11) because $\boldsymbol{U}^\top \boldsymbol{U} = \boldsymbol{V}^\top \boldsymbol{V} = \boldsymbol{I}_d$. As $\boldsymbol{Q}^\star \in \mathcal{O}(d)$, $\boldsymbol{Q}^\star$ solves (6). Finally, combination of (8), (9), (11) results in

$$\begin{aligned}&\min_{h \in \mathrm{Stab}_{\mathcal{G}_{\mathrm{cong}}}(\rho)} J_c(g_0, g_1, h)\\ &\quad = p\left(\mathrm{Tr}(\boldsymbol{\Sigma}_0)^2 + \mathrm{Tr}(\boldsymbol{\Sigma}_1)^2 - 2\,\mathrm{Tr}(\boldsymbol{\Lambda})^2 - 2\,\mathrm{Tr}(\boldsymbol{\Lambda}^2)\right) - 2p^2\,\mathrm{Tr}(\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}_1) +\\ &\qquad p(p+1)\left(\mathrm{Tr}(\boldsymbol{\Sigma}_0^2) + \mathrm{Tr}(\boldsymbol{\Sigma}_1^2)\right).\end{aligned}$$

This observation concludes our proof. □

Lemma 11. *Suppose $g_i = f_{\boldsymbol{\Sigma}_i^{1/2}} \in \mathcal{G}_{\mathrm{cong}}$ for $i = 0, 1$ and $c(\boldsymbol{X}, \boldsymbol{Y}) = \|\boldsymbol{X} - \boldsymbol{Y}\|_{\mathrm{F}}^2$. Then $T^\star = g_1 \circ h^\star \circ g_0^{-1}$, where $h^\star$ is as defined in Theorem 5.2, admits the following form*

$$T^\star(\boldsymbol{X}) = \boldsymbol{T}\boldsymbol{X}\boldsymbol{T}^\top, \quad \text{where} \quad \boldsymbol{T} = \boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{1/2})^{1/2}\boldsymbol{\Sigma}_0^{-1/2}. \tag{12}$$

Additionally, $T^\star$ satisfies ($\star$) for some c-convex function φ.

Proof of Lemma 11. We write the singular value decomposition of $\boldsymbol{M} = \boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_1^{1/2} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^\top$, where $\boldsymbol{\Lambda}$ is a d-dimensional diagonal matrix. By Theorem 5.2, $h^\star = f_{(\boldsymbol{V}\boldsymbol{U}^\top)}$ and as $T^\star = g_1 \circ h^\star \circ g_0^{-1}$, we have $T^\star(\boldsymbol{X}) = \boldsymbol{T}\boldsymbol{X}\boldsymbol{T}^\top$, where $\boldsymbol{T} = \boldsymbol{\Sigma}_1^{1/2}\boldsymbol{V}\boldsymbol{U}^\top\boldsymbol{\Sigma}_0^{-1/2}$. Similarly to the proof Lemma 6, one can show that $\boldsymbol{T} = \boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{1/2})^{1/2}\boldsymbol{\Sigma}_0^{-1/2}$ and $\boldsymbol{T}$ is a symmetric matrix satisfying $\boldsymbol{T} \succ 0$. Next, define

$$\varphi(\boldsymbol{X}) = \mathrm{Tr}(\boldsymbol{T}\boldsymbol{X}\boldsymbol{T}^\top\boldsymbol{X}) - \|\boldsymbol{X}\|_F^2$$

and as $\boldsymbol{T}$ is symmetric positive definite, we have $\nabla_{\boldsymbol{X}}\varphi(\boldsymbol{X}) = 2\boldsymbol{T}\boldsymbol{X}\boldsymbol{T}^\top - 2\boldsymbol{X}$. Next, we evaluate ($\star$):

$$2\boldsymbol{T}\boldsymbol{X}\boldsymbol{T}^\top - 2\boldsymbol{X} + 2(\boldsymbol{X} - T^\star(\boldsymbol{X})) = 0.$$

As $\boldsymbol{T} \succ 0$, $\nabla^2_{\boldsymbol{X}}\varphi(\boldsymbol{X})[\boldsymbol{H}] = 2\boldsymbol{T}\boldsymbol{H}\boldsymbol{T} - 2\boldsymbol{H}$ for $\boldsymbol{H} \in \mathrm{Sym}(d)$. Note that

$$\langle \boldsymbol{H}, (\nabla^2_{\boldsymbol{X}}\varphi(\boldsymbol{X})(\boldsymbol{H}) + 2\boldsymbol{I}_{\mathrm{Sym}(d)})\rangle = \langle \boldsymbol{H}, 2\boldsymbol{T}\boldsymbol{H}\boldsymbol{T}\rangle = 2\|\boldsymbol{T}^{1/2}\boldsymbol{H}\boldsymbol{T}^{1/2}\|_{\mathrm{F}}^2 \geq 0,$$

where $\boldsymbol{I}_{\mathrm{Sym}(d)}$ is the identity operator on Sym(d). Consequently, by (Villani, 2008, Example 13.6) φ is c-convex. Hence, the optimality criteria in ($\star$) is satisfied by $T^\star$, and this observation completes our proof. □

Theorem 5.3. *Suppose that $\mu_i = \mathcal{W}_d(\boldsymbol{\Sigma}_i, p)$ for some $\boldsymbol{\Sigma}_i \in \mathbb{S}^d_{++}$, $i = 0, 1$, and $c(\boldsymbol{X}, \boldsymbol{Y}) = \|\boldsymbol{X} - \boldsymbol{Y}\|_{\mathrm{F}}^2$. Then, we have*

(i) *$\mathbb{M}_c(\mu_0, \mu_1) = \mathbb{J}_c(g_0, g_1)$ and $T^\star$ as expressed in (12) solves (MP),*
(ii) *$\mathbb{K}_c(\mu_0, \mu_1) = \mathbb{J}_c(g_0, g_1)$ and $(\mathrm{id}_{\mathcal{X}}, T^\star)_{\#}\mu_0$ solves (KTP),*

where $g_i = f_{\boldsymbol{\Sigma}_i^{1/2}} \in \mathcal{G}_{\mathrm{cong}}$ for $i = 0, 1$.

Proof of Theorem 5.3. By Lemma 8, we have $(f_{\boldsymbol{\Sigma}_i^{1/2}})_{\#}\rho = \mu_i$ for $i = 0, 1$; hence $\mu_0, \mu_1 \in \mathcal{G}_{\mathrm{cong}\#}\rho$. Define $T^\star = g_1 \circ h^\star \circ g_0^{-1}$, where $h^\star$ is the map in Theorem 5.2. Then, by Lemma 11, $T^\star$ satisfies the requirement in Theorem 4.1 (iii). Hence, the assertions (i) and (ii) follow directly from Theorem 4.1 (iii)-(a) and (iii)-(b), respectively. □

Wishart laws are not elliptical and they arise from congruence, not affine, symmetries. To our knowledge, a closed-form solution for the optimal transport problem induced by the squared-Frobenius norm between two Wishart distributions sharing the same degrees-of-freedom parameter has not previously appeared in the literature.

5.3. **Product of Exponential distributions:** Probability density function of the exponential distribution with rate parameter $\beta > 0$ denoted by $\mathrm{Exp}(\beta)$ is

$$x \mapsto \beta \exp(-\beta x) \mathbb{1}_{x>0}.$$

(i) We set the reference distribution as the product of d exponential distributions each with parameter $\beta_i = 1$, that is, $\rho = \bigotimes_{i=1}^d \mathrm{Exp}(1)$ with probability density function

$$r(\boldsymbol{x}) = \prod_{i=1}^d \exp(-x_i) \mathbb{1}_{x_i>0}.$$

(ii) The acting group is the diagonal scaling group:

$$\mathcal{G}_{\text{scale}} = \left\{ f_{\boldsymbol{\beta}} : \boldsymbol{z} \mapsto \boldsymbol{\beta} \odot \boldsymbol{z} : \boldsymbol{\beta} \in (0, \infty)^d \right\}.$$

Composition is component-wise multiplication

$$(f_{\boldsymbol{\beta}} \circ f_{\tilde{\boldsymbol{\beta}}})(\boldsymbol{z}) = f_{\boldsymbol{\beta}}(\tilde{\boldsymbol{\beta}} \odot \boldsymbol{z}) = \tilde{\boldsymbol{\beta}} \odot \boldsymbol{\beta} \odot \boldsymbol{z} = f_{\tilde{\boldsymbol{\beta}} \odot \boldsymbol{\beta}}(\boldsymbol{z})$$

and inverse is $f_{\boldsymbol{\beta}}^{-1}(\boldsymbol{x}) = \boldsymbol{\beta}^{-1} \odot \boldsymbol{x}$, where $\boldsymbol{\beta}^{-1} = (\beta_1^{-1}, \ldots, \beta_d^{-1})$.

Lemma 12. *For any $\boldsymbol{\beta} \in (0, +\infty)^d$, $f_{\boldsymbol{\beta}^{-1}\#}\rho = \otimes_{i=1}^d \mathrm{Exp}(\beta_i)$ and $\{g_\# \rho : g \in \mathcal{G}_{\text{scale}}\} = \{\bigotimes_{i=1}^d \mathrm{Exp}(\beta_i) : \boldsymbol{\beta} \in (0, \infty)^d\}$.*

Proof of Lemma 12. Now, let $\boldsymbol{\beta} \in (0, +\infty)^d$, then for every Borel $\mathcal{A} \subset \mathbb{R}_+^d$, we have

$$\begin{aligned} (f_{\boldsymbol{\beta}\#}\rho)(\mathcal{A}) &= \int_{\mathcal{A}} r\left(f_{\boldsymbol{\beta}}^{-1}(\boldsymbol{x})\right) \left|\det(Df_{\boldsymbol{\beta}}(\boldsymbol{x}))^{-1}\right| \mathrm{d}\boldsymbol{x} \\ &= \int_{\mathcal{A}} \prod_{i=1}^d \beta_i^{-1} \exp(-\beta_i^{-1} x_i) \mathrm{d}\boldsymbol{x}, \end{aligned}$$

where the first equality follows by the definition of the push-forward operator. The second equality follows by the multivariate change-of-variables formula. Hence, we may conclude that $f_{\boldsymbol{\beta}\#}\rho = \otimes_{i=1}^d \mathrm{Exp}(\beta_i^{-1})$, implying that

$$\{g_\# \rho : g \in \mathcal{G}_{\text{scale}}\} \subseteq \left\{ \bigotimes_{i=1}^d \mathrm{Exp}(\beta_i) : \boldsymbol{\beta} \in (0, \infty)^d \right\}.$$

The reverse inclusion follows directly by noticing for any $\otimes_{i=1}^d \mathrm{Exp}(\beta_i)$ with $\boldsymbol{\beta} \in (0, \infty)^d$, $f_{\boldsymbol{\beta}^{-1}\#}\rho = \otimes_{i=1}^d \mathrm{Exp}(\beta_i)$. This observation completes our proof. □

Lemma 13. *The stabilizer $\rho = \otimes_{i=1}^d \mathrm{Exp}(1)$ under $\mathcal{G}_{\text{scale}}$ is $\mathrm{Stab}_{\mathcal{G}_{\text{scale}}}(\rho) = \{f_{\boldsymbol{1}_d}\}$.*

Proof Lemma 13. Now, suppose that $f_{\boldsymbol{\beta}} \in \mathcal{G}_{\text{scale}}$ satisfies $f_{\boldsymbol{\beta}\#}\rho = \rho$. For each coordinate i its one-dimensional marginal densities must coincide:

$$\beta_i^{-1} e^{-x_i/\beta_i} = e^{-x_i} \quad \text{for all } x_i > 0.$$

Taking log and comparing the coefficients of x_i gives $\beta_i = 1$ for every i. □

Lemma 14. *Suppose $g_i = f_{\boldsymbol{\beta}_i^{-1}} \in \mathcal{G}_{\text{scale}}$ for $i = 0, 1$. Then, orbit transport problem in* (LGOP) *induced by the cost function $c(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|^2$ admits the following* *closed form expression* $\mathbb{J}_c(g_0, g_1) = 2\sum_{i=1}^{d}(\beta_{0,i}^{-1} - \beta_{1,i}^{-1})^2$. *Additionally,* $h^\star = f_{\mathbf{1}_d}$ *solves* (LGOP).

Proof of Lemma 14. By Lemma 13, the stabilizer group is a singleton, and thus $h^\star = f_{\mathbf{1}_d}$ solves (LGOP). Next, we evaluate $\mathbb{J}_c$:

$$\begin{aligned}\mathbb{J}_c(g_0, g_1) &= \int_{\mathcal{X}} \|g_0(\boldsymbol{x}) - g_1(\boldsymbol{x})\|^2 \mathrm{d}\rho(\boldsymbol{x}) \\ &= \int_{\mathcal{X}} \sum_{i=1}^{d} (\beta_{0,i}^{-1} - \beta_{1,i}^{-1})^2 x_i^2 \mathrm{d}\rho(\boldsymbol{x}) = 2\sum_{i=1}^{d}(\beta_{0,i}^{-1} - \beta_{1,i}^{-1})^2,\end{aligned}$$

where the last equality follows because $\mathbb{E}_{x\sim\mathrm{Exp}(1)}[x^2] = 2$. □

Theorem 5.4. *Suppose* $\mu_i = \bigotimes_{k=1}^{d} \mathrm{Exp}(\beta_{i,k})$ *for some* $\boldsymbol{\beta} \in (0, +\infty)^d$, $i = 0, 1$ *and* $c(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|^2$. *Then, we have*

(i) $\mathbb{M}_c(\mu_0, \mu_1) = \mathbb{J}_c(g_0, g_1)$ *and* $T^\star(\boldsymbol{x}) = (\beta_{0,1}/\beta_{1,1}x_1, \ldots, \beta_{0,d}/\beta_{1,d}x_d)$ *solves* (MP),

(ii) $\mathbb{K}_c(\mu_0, \mu_1) = \mathbb{J}_c(g_0, g_1)$ *and* $(\mathrm{id}_{\mathcal{X}}, T^\star)_{\#}\mu_0$ *solves* (KTP),

where $g_i = f_{\boldsymbol{\beta}_i^{-1}} \in \mathcal{G}_{\mathrm{scale}}$.

Proof of Theorem 5.4. By Lemma 12, $g_i = f_{\boldsymbol{\beta}_i^{-1}} \in \mathcal{G}_{\mathrm{scale}}$ such that $\mu_i = g_{i\#}\rho$ for $i = 0, 1$. Next, we define $T^\star = g_1 \circ h^\star \circ g_0^{-1}$, where $h^\star = f_{\mathbf{1}_d}$. Then, a simple calculation shows that $T^\star$ is in the form stated in assertion (i). We define $\varphi(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{B}\boldsymbol{x} - \|\boldsymbol{x}\|^2$, where $\boldsymbol{B} = \mathrm{diag}(\beta_{0,1}/\beta_{1,1}, \ldots, \beta_{0,d}/\beta_{1,d})$. Then, we compute its gradient $\nabla_{\boldsymbol{x}}\varphi(\boldsymbol{x}) = 2\boldsymbol{B}\boldsymbol{x} - 2\boldsymbol{x}$ and evaluate ($\star$):

$$2\boldsymbol{B}\boldsymbol{x} - 2\boldsymbol{x} + 2(\boldsymbol{x} - T^\star(\boldsymbol{x})) = 0.$$

Because $\beta_{0,k}/\beta_{1,k} > 0$ for all $k = 1, \ldots, d$, $\nabla_{\boldsymbol{x}}^2\varphi(\boldsymbol{x}) = 2\boldsymbol{B} - 2\boldsymbol{I}_d \succeq -2\boldsymbol{I}_d$. Consequently, by (Villani, 2008, Example 13.6), φ is c-convex. Hence, $T^\star$ satisfies the requirement in Theorem 4.1 (iii). Subsequently, the assertions (i) and (ii) follow directly from Theorem 4.1 (iii)-(a) and (iii)-(b), respectively. □

Many other product-form families with coordinate-wise scale symmetries admit the same analysis as in Section 5.3. Consequently, the optimal map is again diagonal (pure rescaling for scale families; coordinate-wise affine for location-scale families), and the closed-form cost reduces to a one-dimensional expectation, derived similarly to Theorem 5.4. Representative examples include products of Weibull, Rayleigh, Gamma, inverse-Gamma, Pareto, Lognormal, and generalized-Gamma. Note that because the cost function is separable, and the probability measures in the Lie group orbits are products, the Kantorovich problem decouples across coordinates; one optimal plan is the product of the one-dimensional optimizers, and the Kantorovich value is the sum of the corresponding one-dimensional costs. The one-dimensional case is treated in the following subsection.

5.4. **Absolutely continuous one-dimensional distributions with positive density.** Fix an absolutely continuous probability measure $\mu \ll \mathcal{L}^1$ on $\mathbb{R}$ and let F_μ and r_μ denote its cumulative distribution function, and probability density function, respectively. Throughout we consider the class

$$\mathcal{P}^+ = \left\{\mu \in \mathcal{P}(\mathbb{R}) : \mu \ll \mathcal{L}^1,\ F_\mu \in \mathcal{C}^1,\ F_\mu'(x) > 0\ \forall x \in \mathbb{R}\right\}$$

(i) Fix a reference probability measure ρ with a smooth, strictly positive density $r \in C^\infty(\mathbb{R})$; a concrete example is the logistic density

$$r(x) = \tfrac{1}{4}\,\mathrm{sech}^2\big(x/2\big),$$

and its cumulative distribution function $F_\rho(x) = \int_{-\infty}^x r(s)\mathrm{d}s$ satisfies $F_\rho'(x) = r(x) > 0$ for all x, so $F_\rho : \mathbb{R} \to (0,1)$ is a C^∞, strictly increasing bijection. By the inverse-function theorem, F_ρ is in fact a C^∞ diffeomorphism with smooth inverse $F_\rho^{-1} : (0,1) \to \mathbb{R}$.

(ii) Let $\mathrm{Diff}^+(\mathbb{R})$ be the orientation-preserving $\mathcal{C}^1$ diffeomorphisms:

$$\mathrm{Diff}^+(\mathbb{R}) = \left\{g \in \mathcal{C}^1(\mathbb{R};\mathbb{R}) : g'(x) > 0 \ \forall x \in \mathbb{R},\ g \text{ bijective}\right\},$$

equipped with composition.

Lemma 15. *For any* $\mu \in \mathcal{P}^+$, $(F_\mu^{-1} \circ F_\rho)_\# \rho = \mu$ *and* $\mathrm{Diff}^+(\mathbb{R})_\# \rho = \mathcal{P}^+$.

Proof of Lemma 15. First, we will show that $\mathrm{Diff}^+(\mathbb{R})_\#\rho \subseteq \mathcal{P}^+$. Take $g \in \mathrm{Diff}^+(\mathbb{R})$ and set $\mu = g_\#\rho$. Since g is $\mathcal{C}^1$, strictly increasing and surjective, its inverse is likewise $\mathcal{C}^1$ and strictly increasing. Additionally, by the inverse function theorem $(g^{-1})'(t) = 1/g'(g^{-1}(t)) > 0$. A direct computation shows that the cumulative distribution function of μ is $F_\mu(x) = \int_{-\infty}^x r(g^{-1}(t))(g^{-1})'(t)\mathrm{d}t$. By applying the change of variables in the form of $s = g^{-1}(t)$ (hence $\mathrm{d}t = g'(s)\mathrm{d}s$), we have $F_\mu(x) = \int_{-\infty}^{g^{-1}(x)} r(s)\mathrm{d}s = F_\rho(g^{-1}(x))$. Since F_ρ is a smooth, strictly increasing bijection $\mathbb{R} \to (0,1)$, its composition with the smooth, strictly increasing g^{-1} is continuous on $\mathbb{R}$ and strictly increasing on the set where its values lie between 0 and 1. Moreover,

$$F_\mu'(x) = r\big(g^{-1}(x)\big)(g^{-1})'(x) = \frac{r(g^{-1}(x))}{g'(g^{-1}(x))} > 0.$$

Thus F_μ satisfies the defining properties of $\mathcal{P}^+$, so $\mu \in \mathcal{P}^+$.

Next, we will show that $\mathcal{P}^+ \subseteq \mathrm{Diff}^+(\mathbb{R})_\#\rho$. Take an arbitrary $\mu \in \mathcal{P}^+$ and write F_μ for its cumulative distribution function and r_μ for its density. Because F_μ is strictly increasing, the inverse $F_\mu^{-1} : (0,1) \to \mathbb{R}$ is well defined and rests in $\mathcal{C}^1(\mathbb{R})$ on $(0,1)$ and strictly increasing. Define

$$g = F_\mu^{-1} \circ F_\rho : \mathbb{R} \longrightarrow \mathbb{R}.$$

Both factors are in $\mathcal{C}^1(\mathbb{R})$ and strictly increasing, hence $g \in \mathrm{Diff}^+(\mathbb{R})$. For any $\mathcal{A} \subseteq \mathcal{B}(\mathbb{R})$,

$$\begin{aligned} g_\#\rho(\mathcal{A}) &= \rho\left(g^{-1}(\mathcal{A})\right) \\ &= \rho\left(\left\{x \in \mathbb{R} : F_\mu^{-1}(F_\rho(x)) \in \mathcal{A}\right\}\right) \\ &= \rho\left(\{x \in \mathbb{R} : F_\rho(x) \in F_\mu(\mathcal{A})\}\right) \\ &= (F_\rho)_\#\rho(F_\mu(\mathcal{A})) = \lambda\left(F_\mu(\mathcal{A})\right) = \mu(\mathcal{A}), \end{aligned}$$

where λ is the Lebesgue measure on the unit interval $(0,1)$. The first equality follows by the definition of the push-forward operator, the second equality by construction of g, the third equality follows because F_μ^{-1} is the inverse of the strictly increasing function F_μ. The fourth equality follows by the definition of the push-forward operator, the fifth equality follows because by the probability-integral transform $(F_\rho)_\#\rho = \lambda$. The last equality follows because $(F_\mu)_\#\mu = \lambda$, which equivalently means that for every $\mathcal{B}' \in \mathcal{B}((0,1))$, we have $\lambda(\mathcal{B}') = \mu(F_\mu^{-1}(\mathcal{B}'))$. Taking $\mathcal{B}' =$

$F_\mu(\mathcal{A})$ gives $\lambda(F_\mu(\mathcal{A})) = \mu(\mathcal{A})$. Hence, we may conclude that $g_{\#}\rho = \mu$, which implies that $\mathcal{P}^+ \subseteq \mathrm{Diff}^+(\mathbb{R})_{\#}\rho$. This observation completes our proof. ☐

Lemma 16. *The stabilizer of* ρ *under* $\mathrm{Diff}^+(\mathbb{R})$ *is* $\mathrm{Stab}_{\mathrm{Diff}^+(\mathbb{R})}(\rho) = \{\mathrm{id}_\mathbb{R}\}$.

Proof of Lemma 16. Fix $g \in \mathrm{Diff}^+(\mathbb{R})$ with $g_{\#}\rho = \rho$. For every $x \in \mathbb{R}$ we have

$$F_\rho(x) = \rho((-\infty, x]) = \rho\left(g^{-1}((-\infty, x])\right) = \rho\left((-\infty, g^{-1}(x)]\right) = F_\rho\left(g^{-1}(x)\right).$$

Thus $F_\rho(x) = F_\rho(g^{-1}(x))$ for all $x \in \mathbb{R}$. Since $r > 0$, the function F_ρ is strictly increasing on $\mathbb{R}$. Hence F_ρ is injective, and the identity above implies $g^{-1}(x) = x$ for every $x \in \mathbb{R}$; equivalently, $g(x) = x$. Therefore $g = \mathrm{id}_\mathbb{R}$, completing the proof. ☐

Lemma 17. *Suppose that* $g_i = F_{\mu_i}^{-1} \circ F_\rho$, $i = 0, 1$. *Then,* (LGOP) *induced by the cost function* $c(x, y) = (x-y)^2$ *is solved by* $h^\star = \mathrm{id}_\mathbb{R}$ *and admits the following closed form expression*

$$\mathbb{J}_c(g_0, g_1) = \int_0^1 (F_{\mu_1}^{-1}(t) - F_{\mu_0}^{-1}(t))^2 \mathrm{d}t. \tag{13}$$

Proof of Lemma 17. By Lemma 16, the stabilizer group is a singleton, and thus $h^\star = \mathrm{id}_\mathbb{R}$ solves (LGOP). Next, we evaluate $\mathbb{J}_c$:

$$\begin{aligned}\mathbb{J}_c(g_0, g_1) &= \int_\mathbb{R} (g_0(x) - g_1(x))^2 \mathrm{d}\rho(x) \\ &= \int_\mathbb{R} (F_{\mu_1}^{-1}(F_\rho(x)) - F_{\mu_0}^{-1}(F_\rho(x)))^2 r(x) \mathrm{d}x \\ &= \int_0^1 (F_{\mu_1}^{-1}(t) - F_{\mu_0}^{-1}(t))^2 \mathrm{d}t,\end{aligned}$$

where the last equality follows by the change-of-variables formula $t \leftarrow F_\rho(x)$, and thus $\mathrm{d}t = r(x)\mathrm{d}x$. This observation completes our proof. ☐

Proposition 1. *Suppose that* $\mu_0, \mu_1 \in \mathcal{P}^+ \cap \mathcal{P}_c(\mathbb{R})$, $c(x, y) = (x - y)^2$. *Then, we have*

(i) $\mathbb{M}_c(\mu_0, \mu_1) = \mathbb{J}_c(g_0, g_1)$, *and* $T^\star = F_{\mu_1}^{-1} \circ F_{\mu_0}$ *solves* (MP),
(ii) $\mathbb{K}_c(\mu_0, \mu_1) = \mathbb{J}_c(g_0, g_1)$, *and* $(\mathrm{id}_\mathcal{X}, T^\star)_{\#}\mu_0$ *solves* (KTP),

where $g_i = F_{\mu_i}^{-1} \circ F_\rho \in \mathrm{Diff}^+(\mathbb{R})$ *for* $i = 0, 1$.

Proof of Proposition 1. By Lemma 15, we have $(g_i)_{\#}\rho = \mu_i$ for $i = 0, 1$; hence $\mu_0, \mu_1 \in \mathrm{Diff}^+(\mathbb{R})_{\#}\rho$. Define $T^\star = g_1 \circ h^\star \circ g_0^{-1}$, where $h^\star$ is the map defined in Lemma 17. Consequently, we have $T^\star(x) = F_{\mu_1}^{-1}(F_{\mu_0}(x))$.

In what follows, we define $\varphi(x) = 2\int_{x_0}^x T^\star(s)\mathrm{d}s - x^2$ for some $x_0 \in \mathbb{R}$, and we compute its first and second order derivatives

$$\varphi'(x) = 2T^\star(x) - 2x \text{ and } \varphi''(x) = 2(T^\star)'(x) - 2.$$

As $T^\star$ is increasing, $(T^\star)'(x) \geq 0$ for all $x \in \mathbb{R}$, implying $\varphi''(x) \geq -2$. Consequently, by (Villani, 2008, Example 13.6), φ is c-convex. Next, we evaluate ($\star$):

$$2T^\star(x) - 2x + 2(x - T^\star(x)) = 0.$$

Hence, $T^\star$ satisfies the requirement in Theorem 4.1 (iii). Accordingly, assertions (i) and (ii) follow directly from Theorem 4.1 (iii)-(a) and (iii)-(b), respectively. ☐

Section 5.4 coincides with (Cuesta-Albertos et al., 1993, Corollary 2.7) when restricted to the quadratic cost.

Acknowledgements

The author thanks Ya-Ping Hsieh, Wouter Jongeneel, and Teng Andrea Xu for helpful feedback.

References

Nicolas Boumal. *An introduction to optimization on smooth manifolds*. Cambridge University Press, 2023.

Yann Brenier. Décomposition polaire et réarrangement monotone des champs de vecteurs. *CR Acad. Sci. Paris Sér. I Math.*, 305:805–808, 1987.

Juan A Cuesta-Albertos, Ludger Rüschendorf, and Araceli Tuerodiaz. Optimal coupling of multivariate distributions and stochastic processes. *Journal of Multivariate Analysis*, 46(2):335–361, 1993.

Giorgio Dall'Aglio. Sugli estremi dei momenti delle funzioni di ripartizione doppia. *Annali della Scuola Normale Superiore di Pisa-Scienze Fisiche e Matematiche*, 10(1-2):35–74, 1956.

Richard Dowson and Henry Landau. The Fréchet distance between multivariate normal distributions. *Journal of Multivariate Analysis*, 12:450–455, 1982.

Johannes Jisse Duistermaat and Johan AC Kolk. *Lie groups*. Springer Science & Business Media, 2012.

Martin Gelbrich. On a formula for the L^2 wasserstein metric between measures on Euclidean and Hilbert spaces. *Mathematische Nachrichten*, 147:185–203, 1990.

Brian C Hall. Lie groups, Lie algebras, and representations. In *Quantum Theory for Mathematicians*, pages 333–366. Springer, 2013.

Leon Isserlis. On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*, 12(1/2): 134–139, 1918.

Leonid Kantorovich. On the transfer of masses (in russian). In *Doklady Akademii Nauk*, volume 37, page 227, 1942.

Narendra Karmarkar. A new polynomial-time algorithm for linear programming. In *Proceedings of the sixteenth annual ACM symposium on Theory of computing*, pages 302–311, 1984.

Sophus Lie. *Vorlesungen über Differentialgleichungen mit bekannten infinitesimalen Transformationen*. BG Teubner, 1891.

Robert J. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Mathematical Journal*, 80(2):309–323, 1995.

Gaspard Monge. Memoir on the theory of excavations and embankments. pages 666–704, 1781.

Robb J Muirhead. *Aspects of multivariate statistical theory*. John Wiley & Sons, 2009.

Ingram Olkin and Friedrich Pukelsheim. The distance between two random vectors with given dispersion matrices. *Linear Algebra and its Applications*, 48:257–263, 1982.

Svetlozar T Rachev and Ludger Rüschendorf. *Mass Transportation Problems: Volume I: Theory*. Springer, 1998.

Ludger Rüschendorf. Fréchet-bounds and their applications. In *Advances in Probability Distributions with Given Marginals: Beyond the copulas*, pages 151–187. Springer, 1991.

Ludger Rüschendorf and Svetlozar T Rachev. A characterization of random variables with minimum $l2$-distance. *Journal of Multivariate Analysis*, 32(1):48–54, 1990.

Cyril Smith and Martin Knott. On Hoeffding-Fréchet bounds and cyclic monotone relations. *Journal of multivariate analysis*, 40(2):328–334, 1992.

Bahar Taşkesen, Soroosh Shafieezadeh-Abadeh, and Daniel Kuhn. Semi-discrete optimal transport: Hardness, regularization and numerical solution. *Mathematical Programming*, 199(1):1033–1106, 2023a.

Bahar Taşkesen, Soroosh Shafieezadeh-Abadeh, Daniel Kuhn, and Karthik Natarajan. Discrete optimal transport with independent marginals is #P-hard. *SIAM Journal on Optimization*, 33(2):589–614, 2023b.

Cédric Villani. *Optimal transport: Old and new*, volume 338. Springer, 2008.

Cédric Villani. *Topics in Optimal Transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, 2003.