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 problem 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 feasible point of this reduced problem whose induced transport map satisfies a c-convex first-order certificate makes the upper bound tight for both the Monge and Kantorovich formulations, with the optimal map realized by a group element. For the quadratic cost on a finite-dimensional Hilbert space and affine-induced actions, the c-convex certificate reduces to an algebraic condition: the candidate map must have self-adjoint positive semidefinite linear part. We give a structural criterion, based on Cartan theory, that guarantees this condition. When the linear image of the acting group admits a global Cartan decomposition and its fixed-point subgroup is contained in the linear image of the stabilizer of the reference law, the compact component can be absorbed by the stabilizer, yielding a transport map with a self-adjoint positive definite linear part. This orbit-based viewpoint unifies known closed-form solutions, such as elliptical distributions, and yields new closed-form solutions for Wishart, inverse-Wishart, and matrix beta type II distributions under the squared Frobenius cost.

1. Introduction

Originating in the seminal work of Monge [21] and later given its analytic form by Kantorovich [15], 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 [16]) 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 [29]. Likewise, the optimal transport problem between a generic (possibly continuous) measure and a discrete measure is also known to be #P-hard [28]. 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 optimal transport between two continuous, non-atomic measures, where mass is spread over a continuum rather than finitely many atoms, would be equally formidable. Curiously, however, certain instances of the

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

optimal transport problem between continuous probability measures admit explicit closed-form solutions, rendering these problems uniquely tractable despite their inherently infinite-dimensional structure.

Early breakthroughs showed that infinite-dimensionality need not preclude analytic tractability. In one dimension, [6] 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, [7] and [23] derived an explicit solution for the optimal transport problem induced by the quadratic cost function in terms of the means and covariance matrices. [11] 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 [24, §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 an equivalence among distributions, allowing familiar families to be recognized as members of a single orbit. In this setting, the Monge problem admits an upper bound obtained by optimizing over the stabilizer subgroup of the reference law. In the examples we study, this reduced problem is finite-dimensional and sometimes trivial. Under standard regularity conditions, we prove that this orbit upper bound is tight for both the Monge and Kantorovich formulations whenever the candidate map admits a c-convex first-order certificate. In that case, the optimal transport is realized by a suitable group element.

For the quadratic cost and affine-induced actions, we give a structural criterion based on Cartan theory. When the linear image of the acting group admits a global Cartan decomposition and its fixed-point subgroup is contained in the linear image of the stabilizer of the reference law, the compact component can be absorbed by the stabilizer, yielding an orbit-induced transport map with a self-adjoint positive definite linear part. Thus, in these cases, the reduced stabilizer problem need not be solved as an independent optimization problem; the Cartan factorization itself produces a stabilizing element whose orbit-induced map satisfies the algebraic optimality certificate. This Cartan-based mechanism recovers the classical closed-form formulas for elliptical transport and yields new closed-form quadratic transport maps and costs for Wishart, inverse-Wishart, and matrix beta type II distributions on the positive definite cone under the squared Frobenius cost.

The results in this paper unfold in layers. Theorem 4.1 provides the general orbit-reduction statement: it identifies a structured class of admissible maps and gives an upper bound on the Monge value, together with an optimality certificate under which this upper bound becomes tight. Theorem 4.2 specializes this certificate to quadratic costs and actions induced by affine representations. In this setting, optimality reduces to the algebraic condition that the linear part of the candidate map be self-adjoint and positive semidefinite. Finally, Theorem 4.3 gives the main group-theoretic mechanism: when the fixed-point subgroup of the linear image is contained in the linear image of the stabilizer of the reference law, the Cartan decomposition allows the compact factor to be absorbed by a stabilizing element, thereby producing a candidate map that satisfies the quadratic certificate. The examples in Section 5 verify these hypotheses for the affine and congruence mechanisms.

Notation. We write $\bar{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$ for the upper extended real line. 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 nonempty open subset of a finite-dimensional Euclidean space. In the examples considered in this paper (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 $\mathbb{S}^d_{\succ}$. 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 $\mathrm{Sym}(d) = \{\mathbf{H} \in \mathbb{R}^{d\times d} \mid \mathbf{H}^\top = \mathbf{H}\}$. 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 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 $T_\# \mu(\mathcal{A}) = \mu(T^{-1}(\mathcal{A}))$, $\mathcal{A} \in \mathcal{B}(\mathcal{X})$. We write $\mathcal{U}([0,1])$ for the uniform distribution on $[0,1]$. The multivariate gamma function is denoted by $\mathsf{\Gamma}_d(a) = \pi^{d(d-1)/4} \prod_{j=1}^d \Gamma(a - \frac{j-1}{2})$, $a > \frac{d-1}{2}$. The multivariate beta function is denoted by $\mathsf{B}_d(a,b) = \frac{\mathsf{\Gamma}_d(a)\mathsf{\Gamma}_d(b)}{\mathsf{\Gamma}_d(a+b)}$, $a, b > \frac{d-1}{2}$. For $a, b \in \mathbb{Z}_{>0}$, $\delta_{ab} = 1$ if $a = b$ and $\delta_{ab} = 0$ otherwise. For $1 \leq a, b \leq d$, $\boldsymbol{E}_{ab}$ denotes the $d \times d$ matrix with a 1 in position (a,b) and zeros elsewhere. The Hadamard (element-wise) product is denoted by $\odot$. The symbol $\oplus$ denotes a direct sum, and $\rtimes$ denotes a semidirect product of groups.

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 [21] formulated the problem of transporting one distribution of mass into another at minimal cost. Formally, given a Borel-measurable cost function $c : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ and two measures $\mu_0, \mu_1 \in \mathcal{P}(\mathcal{X})$, the *Monge problem* induced by cost function c is defined as

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

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 (MP) is an infinite-dimensional optimization over a generally nonconvex feasible set.

In 1942, Kantorovich [15] introduced a convex relaxation of (MP) by replacing maps with probability couplings. Given $\mu_0, \mu_1 \in \mathcal{P}(\mathcal{X})$, 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). Under suitable regularity assumptions, optimal Kantorovich plans concentrate on graphs of transport maps, and these maps solve the corresponding Monge problem. Existence of Monge solutions for the quadratic cost on Euclidean space was first developed in [3, 26]; for non-quadratic costs, existence was investigated in [25, 27, 20]. One of the most general results linking Monge solutions to gradients of c-convex functions is due to [30]. For ease of reference, we now adapt [30, 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 $c : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ 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 $\mathcal{X} \ni \boldsymbol{y} \mapsto \nabla_{\boldsymbol{x}} c(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^n$ is injective.*

Assumption 1 contains the twist condition (see [30, §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 $\boldsymbol{x}$.

Definition 1 (c-convexity and c-subdifferential)**.** *Suppose $c : \mathcal{X} \times \mathcal{X} \to \bar{\mathbb{R}}$. A function $\psi : \mathcal{X} \to \bar{\mathbb{R}}$ is said to be* c-convex *if it is not identically $+\infty$ and there exists a function $\phi : \mathcal{X} \to \mathbb{R} \cup \{\pm\infty\}$ such that $\psi(\boldsymbol{x}) = \sup_{\boldsymbol{y} \in \mathcal{X}} \{\phi(\boldsymbol{y}) - c(\boldsymbol{x}, \boldsymbol{y})\}$, $\forall \boldsymbol{x} \in \mathcal{X}$. For a c-convex function $\psi : \mathcal{X} \to \bar{\mathbb{R}}$, its* c-subdifferential *at $\boldsymbol{x} \in \mathcal{X}$ is $\partial^c \psi(\boldsymbol{x}) = \{\boldsymbol{y} \in \mathcal{X} : \psi(\boldsymbol{z}) \geq \psi(\boldsymbol{x}) + c(\boldsymbol{x}, \boldsymbol{y}) - c(\boldsymbol{z}, \boldsymbol{y}) \ \forall \boldsymbol{z} \in \mathcal{X}\}$. We say that ψ is c-subdifferentiable at $\boldsymbol{x}$ if $\partial^c \psi(\boldsymbol{x}) \neq \varnothing$, and we call* $\mathrm{dom}(\partial^c \psi) = \{\boldsymbol{x} \in \mathcal{X} : \partial^c \psi(\boldsymbol{x}) \neq \varnothing\}$ *the* domain of c-subdifferentiability.

The notion of c-convexity is a cost-dependent analogue of ordinary convexity. In classical convex analysis, convex functions arise as pointwise suprema of affine functions. Here, the affine family is replaced by the cost-generated family $\boldsymbol{x} \mapsto \phi(\boldsymbol{y}) - c(\boldsymbol{x}, \boldsymbol{y})$, $\boldsymbol{y} \in \mathcal{X}$, which leads to the definition above. In the same spirit, the notion of the c-subdifferential is the corresponding cost-dependent analogue of the ordinary subdifferential. In classical convex analysis, a vector $\boldsymbol{p} \in \mathbb{R}^d$ belongs to the subdifferential of a convex function ψ at $\boldsymbol{x}$ if the affine function $\boldsymbol{z} \mapsto \psi(\boldsymbol{x}) + \langle \boldsymbol{p}, \boldsymbol{z} - \boldsymbol{x} \rangle$ agrees with ψ at $\boldsymbol{x}$ and lies below ψ for all $\boldsymbol{z} \in \mathcal{X}$. Here, affine supporting functions are replaced by the cost-adjusted family $\boldsymbol{z} \mapsto \psi(\boldsymbol{x}) + c(\boldsymbol{x}, \boldsymbol{y}) - c(\boldsymbol{z}, \boldsymbol{y})$, $\boldsymbol{y} \in \mathcal{X}$, which leads to the definition above.

Assumption 2. *The cost function $c : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ and the measure $\mu_0 \in \mathcal{P}(\mathcal{X})$ are such that, for every c-convex function $\psi : \mathcal{X} \to \bar{\mathbb{R}}$, the function ψ is differentiable μ_0-almost everywhere on* $\mathrm{dom}(\partial^c \psi)$.

Assumption 2 allows the geometric condition $\boldsymbol{y} \in \partial^c \psi(\boldsymbol{x})$ to be converted into a first-order identity. Indeed, let ψ be c-convex, let $\boldsymbol{x} \in \mathrm{dom}(\partial^c \psi)$ be a point at which ψ is differentiable, and choose $\boldsymbol{y} \in \partial^c \psi(\boldsymbol{x})$. By the definition of the c-subdifferential, $\psi(\boldsymbol{z}) \geq \psi(\boldsymbol{x}) + c(\boldsymbol{x}, \boldsymbol{y}) - c(\boldsymbol{z}, \boldsymbol{y})$ for all $\boldsymbol{z} \in \mathcal{X}$. Equivalently, $\psi(\boldsymbol{z}) + c(\boldsymbol{z}, \boldsymbol{y}) \geq \psi(\boldsymbol{x}) + c(\boldsymbol{x}, \boldsymbol{y})$ for all $\boldsymbol{z} \in \mathcal{X}$. Hence the function $\boldsymbol{z} \mapsto \psi(\boldsymbol{z}) + c(\boldsymbol{z}, \boldsymbol{y})$ attains a minimum at $\boldsymbol{z} = \boldsymbol{x}$. Since $\mathcal{X}$ is open and $\boldsymbol{z} \mapsto c(\boldsymbol{z}, \boldsymbol{y})$ is $\mathcal{C}^1$, the first-order optimality condition gives $\nabla_{\boldsymbol{x}} \psi(\boldsymbol{x}) + \nabla_{\boldsymbol{x}} c(\boldsymbol{x}, \boldsymbol{y}) = 0$. Together with the twist condition in Assumption 1, this identity identifies $\boldsymbol{y}$ uniquely.

Assumption 3. *The cost $c : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ satisfies condition (H_∞) in the sense of [30, Chapter 10]. More explicitly, for a set $\mathcal{S} \subseteq \mathcal{X}$ and a point $\boldsymbol{x} \in \overline{\mathcal{S}}$, write*

$$\mathcal{T}(\mathcal{S}, \boldsymbol{x}) = \left\{ \lim_{k \to \infty} \frac{\boldsymbol{x}_k - \boldsymbol{x}}{t_k} : \boldsymbol{x}_k \in \mathcal{S}, \boldsymbol{x}_k \to \boldsymbol{x}, t_k > 0,\ t_k \to 0 \right\}$$

for the tangent cone to $\mathcal{S}$ at $\boldsymbol{x}$. Then:

- (i) *For every $\boldsymbol{x} \in \mathcal{X}$ and every measurable set $\mathcal{S} \subseteq \mathcal{X}$ whose tangent cone $\mathcal{T}(\mathcal{S}, \boldsymbol{x})$ is not contained in a half-space, there exist points $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_k \in \mathcal{S}$ and a ball $B \subseteq \mathcal{X}$ containing $\boldsymbol{x}$ such that, for all $\boldsymbol{y}$ outside a compact subset of $\mathcal{X}$, $\inf_{\boldsymbol{w} \in B} c(\boldsymbol{w}, \boldsymbol{y}) \geq \inf_{1 \leq j \leq k} c(\boldsymbol{z}_j, \boldsymbol{y})$.*
- (ii) *For every $\boldsymbol{x} \in \mathcal{X}$ and every neighborhood $\mathcal{U} \subseteq \mathcal{X}$ of $\boldsymbol{x}$, there exists a ball $B \subseteq \mathcal{X}$ containing $\boldsymbol{x}$ such that $\lim_{\boldsymbol{y} \to \infty} \sup_{\boldsymbol{w} \in B} \inf_{\boldsymbol{z} \in \mathcal{U}} \{c(\boldsymbol{z}, \boldsymbol{y}) - c(\boldsymbol{w}, \boldsymbol{y})\} = -\infty$.*

Here $\boldsymbol{y} \to \infty$ means that $\boldsymbol{y}$ eventually leaves every compact subset of $\mathcal{X}$.

Assumption 3 is a no-escape-at-infinity condition on the cost. It controls what happens when the second argument $\boldsymbol{y}$ leaves every compact subset of $\mathcal{X}$. The first part says that, near any point $\boldsymbol{x}$ and along any set $\mathcal{S}$ that has enough directions around $\boldsymbol{x}$, the cost of sending nearby points $\boldsymbol{w} \in B$ to a far-away target $\boldsymbol{y}$ is bounded from below by the cost of sending finitely many comparison points $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_k \in \mathcal{S}$ to the same target. Thus far-away targets cannot distinguish the point $\boldsymbol{x}$ from its surrounding directions in an uncontrolled way. The second part is a stronger separation condition. It says that, if $\boldsymbol{z}$ is allowed to range in any prescribed neighborhood $\mathcal{U}$ of $\boldsymbol{x}$, then for targets $\boldsymbol{y}$ going to infinity the cost difference $c(\boldsymbol{z}, \boldsymbol{y}) - c(\boldsymbol{w}, \boldsymbol{y})$ can be made uniformly very negative, with $\boldsymbol{w}$ ranging over a small ball B around $\boldsymbol{x}$. In other words, when the target moves far away, points near $\boldsymbol{x}$ cannot all behave as equally good first-order competitors. The assumption therefore rules out pathological behavior caused by target points escaping to infinity.

For ease of reference, we now recall the relevant consequences of [30, Theorem 10.28] in the notation of this paper.

Theorem 2.1. *Let $\mu_0, \mu_1 \in \mathcal{P}(\mathcal{X})$. Suppose that $c : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ satisfies Assumption 1, that the pair (c, μ_0) satisfies Assumption 2, and that $\mathbb{K}_c(\mu_0, \mu_1) < \infty$. Then:*

- (i) (KTP) *admits a unique (in law) optimal transportation plan $\Gamma^\star \in \Pi(\mu_0, \mu_1)$.*
- (ii) *There exists a unique optimal Monge map $T^\star : \mathcal{X} \to \mathcal{X}$, up to μ_0-almost everywhere equality solving* (MP)*.*
- (iii) *There exists a c-convex function $\varphi : \mathcal{X} \to \bar{\mathbb{R}}$ such that*

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

- (iv) *$\Gamma^\star$ is concentrated on the graph $T^\star$, that is, $\Gamma^\star = (\mathrm{id}_{\mathcal{X}}, T^\star)_{\#}\mu_0$.*
- (v) *If, in addition, c satisfies Assumption 3, then* ($\heartsuit$) *characterizes the optimal coupling in the following sense: if $\Gamma \in \Pi(\mu_0, \mu_1)$ is concentrated on pairs $(\boldsymbol{x}, \boldsymbol{y})$ satisfying $\nabla_{\boldsymbol{x}} \varphi(\boldsymbol{x}) + \nabla_{\boldsymbol{x}} c(\boldsymbol{x}, \boldsymbol{y}) = 0$ for some c-convex function φ, then Γ is the unique optimal Kantorovich plan. In particular, if a feasible map $T : \mathcal{X} \to \mathcal{X}$ satisfies* ($\heartsuit$) *for some c-convex function φ, then T is the unique optimal Monge map up to μ_0-almost everywhere equality.*

Under the Assumptions 1, 2 and 3, Theorem 2.1(v) delineates conditions under which the Monge map is characterized via the gradient condition (♡).

2.2. **Lie group theory.** Originating with [18], 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 [2, 8, 14].

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

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 [14, Chapter 1].

Definition 3 (Left action and orbit)**.** *Let G be a Lie group and let $\mathcal{M}$ be a set. A* left action *of G on $\mathcal{M}$ is a map $\phi : G \times \mathcal{M} \to \mathcal{M}$ such that:*

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

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

Orbits induce an equivalence relation on $\mathcal{M}$: two points $\rho_1, \rho_2 \in \mathcal{M}$ are equivalent whenever one can be reached from the other by applying an element of G, that is, whenever there exists $g \in G$ such that $\rho_2 = \phi(g, \rho_1)$. Thus, the equivalence classes are the orbits of the action.

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 $\mathrm{Diff}(\mathcal{X}) = \{\Phi : \mathcal{X} \to \mathcal{X} \mid \Phi \text{ is a } \mathcal{C}^\infty \text{ bijection and } \Phi^{-1} \text{ is } \mathcal{C}^\infty\}$. The group operation on $\mathrm{Diff}(\mathcal{X})$ is composition, the identity element is $\mathrm{id}_\mathcal{X}$, and inversion is the usual map inverse.

In the remainder of the paper, let G be a Lie group acting smoothly on $\mathcal{X}$. Thus we are given a smooth map $G \times \mathcal{X} \to \mathcal{X}$, $(g, \boldsymbol{x}) \mapsto \alpha(g)(\boldsymbol{x})$, such that $\alpha(e) = \mathrm{id}_\mathcal{X}$ and $\alpha(gh) = \alpha(g) \circ \alpha(h)$ for all $g, h \in G$. Equivalently, the action determines a group homomorphism $\alpha : G \to \mathrm{Diff}(\mathcal{X})$, $g \mapsto \alpha(g)$, whose associated evaluation map $(g, \boldsymbol{x}) \mapsto \alpha(g)(\boldsymbol{x})$ is smooth. We refer to the image $\alpha(G) \subseteq \mathrm{Diff}(\mathcal{X})$ as the associated acting transformation group. For example, let $G = \mathrm{GL}(d)$ and $\mathcal{X} = \mathbb{R}^d$, and define $\alpha : G \to \mathrm{Diff}(\mathbb{R}^d)$ by $\alpha(\boldsymbol{A})(x) = \boldsymbol{A}x$. Then α is a group homomorphism, and $\alpha(G)$ is the group of linear diffeomorphisms of $\mathbb{R}^d$.

The goal of this paper is to identify when and explain why the optimal transport problem admits closed-form solutions. 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.

The action α induces a left action of G on $\mathcal{P}(\mathcal{X})$ by pushforward: $\phi : G \times \mathcal{P}(\mathcal{X}) \to \mathcal{P}(\mathcal{X})$, $\phi(g, \mu) = \alpha(g)_{\#}\mu$. Indeed, since $\alpha(e) = \mathrm{id}_{\mathcal{X}}$ and $\alpha(gh) = \alpha(g) \circ \alpha(h)$, it follows that $\alpha(e)_{\#}\mu = \mu$, and $\alpha(gh)_{\#}\mu = \alpha(g)_{\#}(\alpha(h)_{\#}\mu)$ for all $g, h \in G$ and $\mu \in \mathcal{P}(\mathcal{X})$.

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

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

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

Lemma 1 (Orbit measures remain absolutely continuous)**.** *If $\rho \ll \mathcal{L}^n$, then for every $g \in G$ we have $\alpha(g)_{\#}\rho \ll \mathcal{L}^n$ and, for $\mathcal{L}^n$-almost every $\boldsymbol{y} \in \mathcal{X}$,*

$$\frac{\mathrm{d}(\alpha(g)_{\#}\rho)}{\mathrm{d}\mathcal{L}^n}(\boldsymbol{y}) = \frac{\mathrm{d}\rho}{\mathrm{d}\mathcal{L}^n}\big(\alpha(g)^{-1}(\boldsymbol{y})\big)\,\big|\det\big(D(\alpha(g)^{-1})(\boldsymbol{y})\big)\big|,$$

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

Proof. Since $\alpha(g) \in \mathrm{Diff}(\mathcal{X})$, the map $\alpha(g)$ is a $\mathcal{C}^1$-diffeomorphism of $\mathcal{X}$. The conclusion therefore follows from the change-of-variables formula. □

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

Thus $\mathrm{Stab}_G(\rho)$ consists precisely of those elements of G whose induced push-forward action leaves the reference measure ρ invariant. $\mathrm{Stab}_G(\rho)$ is nonempty because $\alpha(e) = \mathrm{id}_{\mathcal{X}}$, and hence $\alpha(e)_{\#}\rho = \rho$.

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 reference probability measure $\rho \in \mathcal{P}(\mathcal{X})$. We study the Monge transportation problem (MP) induced by a Borel-measurable cost function $c : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, between two measures in the orbit of ρ, that is, $\mu_0, \mu_1 \in G_{\#}(\rho)$. Equivalently, there exists $g_0, g_1 \in G$, such that $\mu_0 = \alpha(g_0)_{\#}\rho$ and $\mu_1 = \alpha(g_1)_{\#}\rho$. We fix one such pair (g_0, g_1) throughout this section. We assume $\rho \ll \mathcal{L}^n$, then by Lemma 1, 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 G-orbit. First, although $\mathcal{T}(\mu_0, \mu_1)$ may be empty for arbitrary distributions, when measures reside on a common orbit, the map $T = \alpha(g_1) \circ \alpha(g_0)^{-1} \in \alpha(G)$ 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 2. *If a measurable map $T : \mathcal{X} \to \mathcal{X}$ satisfies $T_{\#}\mu_0 = \mu_1$, then the composite $H = \alpha(g_1)^{-1} \circ T \circ \alpha(g_0)$ is such that $H_{\#}\rho = \rho$ and $T = \alpha(g_1) \circ H \circ \alpha(g_0)^{-1}$.*

Proof. Because $\alpha(g_0)$ and $\alpha(g_1)$ are diffeomorphisms of $\mathcal{X}$, their inverses are Borel measurable. Hence the map $H = \alpha(g_1)^{-1} \circ T \circ \alpha(g_0)$ is measurable. For any $\mathcal{A} \in \mathcal{B}(\mathcal{X})$, we have

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

Thus $H_{\#}\rho = \rho$, which proves the first assertion. Finally, by the definition of H, $\alpha(g_1) \circ H \circ \alpha(g_0)^{-1} = \alpha(g_1) \circ \big(\alpha(g_1)^{-1} \circ T \circ \alpha(g_0)\big) \circ \alpha(g_0)^{-1} = T$. This proves the second assertion. □

Lemma 2 shows that any admissible map $T \in \mathcal{T}(\mu_0, \mu_1)$ can be pulled back to a ρ-preserving transformation H on the reference space, and then recovered by composing with the orbit representatives $\alpha(g_0)$ and $\alpha(g_1)$. Thus, the Monge problem can be viewed as a search over ρ-preserving maps on the reference space. If, in addition, H belongs to the acting transformation group $\alpha(G)$, say $H = \alpha(h)$ for some $h \in G$, then $h \in \mathrm{Stab}_G(\rho)$. To systematically exploit this structure, we introduce the *Lie group orbit transport problem* between $\alpha(g_0)_{\#}\rho$ and $\alpha(g_1)_{\#}\rho$ induced by the cost function c as follows:

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

Lemma 3. *Let $s_0, s_1 \in \mathrm{Stab}_G(\rho)$. Then $\mathbb{J}_c(g_0 s_0, g_1 s_1) = \mathbb{J}_c(g_0, g_1)$.*

Proof. By definition, $\mathbb{J}_c(g_0 s_0, g_1 s_1) = \inf_{h \in \mathrm{Stab}_G(\rho)} \int_{\mathcal{X}} c(\alpha(g_0 s_0)(\boldsymbol{x}), \alpha(g_1 s_1 h)(\boldsymbol{x})) \mathrm{d}\rho(\boldsymbol{x})$.

As $s_0 \in \mathrm{Stab}_G(\rho)$, the change of variables $\boldsymbol{z} = \alpha(s_0)(\boldsymbol{x})$ preserves ρ. Therefore $\mathbb{J}_c(g_0 s_0,\ g_1 s_1)$ equals

$$\inf_{h \in \mathrm{Stab}_G(\rho)} \int_{\mathcal{X}} c\big(\alpha(g_0)(\boldsymbol{z}), \alpha(g_1 s_1 h s_0^{-1})(\boldsymbol{z})\big) \mathrm{d}\rho(\boldsymbol{z}).$$

Note that $\mathrm{Stab}_G(\rho)$ is a subgroup of G. Indeed, $e \in \mathrm{Stab}_G(\rho)$. If $s, t \in \mathrm{Stab}_G(\rho)$, then $\alpha(st)_{\#}\rho = \alpha(s)_{\#}\alpha(t)_{\#}\rho = \rho$, and if $s \in \mathrm{Stab}_G(\rho)$, then $\alpha(s^{-1})_{\#}\rho = \alpha(s^{-1})_{\#}\alpha(s)_{\#}\rho = \rho$. Consequently, the map $h \mapsto s_1 h s_0^{-1}$ is a bijection of $\mathrm{Stab}_G(\rho)$ onto itself. Hence

$$\mathbb{J}_c(g_0 s_0, g_1 s_1) = \inf_{\widetilde{h} \in \mathrm{Stab}_G(\rho)} \int_{\mathcal{X}} c\big(\alpha(g_0)(\boldsymbol{z}), \alpha(g_1 \widetilde{h})(\boldsymbol{z})\big) \mathrm{d}\rho(\boldsymbol{z}) = \mathbb{J}_c(g_0, g_1).$$

□

By Lemma 3, one may equivalently regard $\mathbb{J}_c$ as a function of the quotient representatives $g_0 \,\mathrm{Stab}_G(\rho)$ and $g_1 \,\mathrm{Stab}_G(\rho)$, or equivalently as a function of the orbit measures μ_0, μ_1.

Moreover, because every $h \in \mathrm{Stab}_G(\rho)$ produces an admissible transport map $T_h = \alpha(g_1 h g_0^{-1})$, (LGOP) provides a structured upper bound on (MP) indexed by the stabilizer of the reference law. When $\mathrm{Stab}_G(\rho)$ is finite-dimensional, this reduces the search to a finite-dimensional problem; when the stabilizer is trivial, it reduces to evaluating a single candidate map.

The key issue is when this upper bound is tight. Under the hypotheses of Theorem 2.1, optimality of a candidate map is certified by the existence of a c-convex potential satisfying (♡). In general, finding such a potential is an analytic problem and, for smooth costs, leads to the partial differential equations of optimal transport [31, § 12], for which explicit solutions are rarely available.

The point of the orbit framework is that, for maps of the special form $T_h = \alpha(g_1 h g_0^{-1})$, this analytic certificate can sometimes be verified algebraically. When this happens, T_h is optimal for the Monge problem, the orbit upper bound is tight, and the same element h automatically solves (LGOP). The following theorem formalizes this certificate.

Theorem 4.1. *The following statements hold.*

(i) $\mathbb{M}_c(\mu_0, \mu_1) \leq \mathbb{J}_c(g_0, g_1)$.

(ii) *Suppose that* $\mathbb{K}_c(\mu_0, \mu_1) < \infty$*, that* c *satisfies Assumptions 1 and 3, and that the pair* (c, μ_0) *satisfies Assumption 2. Suppose further that there exist* $h \in \mathrm{Stab}_G(\rho)$ *and a* c*-convex function* φ *such that the map* $T_h = \alpha(g_1 h g_0^{-1})$ *satisfies* (♡) μ_0*-almost surely. Then:*

  (a) T_h *solves* (MP)*, and it is the unique optimal Monge map up to* μ_0*-almost everywhere equality,*

  (b) h *solves* (LGOP)*,*

  (c) $(\mathrm{id}_{\mathcal{X}}, T_h)_{\#}\mu_0$ *solves* (KTP) *and it is the unique optimal Kantorovich plan,*

  (d) $\mathbb{K}_c(\mu_0, \mu_1) = \mathbb{M}_c(\mu_0, \mu_1) = \mathbb{J}_c(g_0, g_1)$.

Proof. 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 $T_h = \alpha(g_1 h g_0^{-1})$. Since $h \in \mathrm{Stab}_G(\rho)$, we have

$$(T_h)_{\#}\mu_0 = \alpha(g_1 h g_0^{-1})_{\#}\alpha(g_0)_{\#}\rho = \alpha(g_1 h)_{\#}\rho = \alpha(g_1)_{\#}\alpha(h)_{\#}\rho = \alpha(g_1)_{\#}\rho = \mu_1.$$

Thus $T_h \in \mathcal{T}(\mu_0, \mu_1)$. We next verify that the objective value attained by h in (LGOP) coincides with the value attained by T_h in (MP):

$$\begin{aligned}\int_{\mathcal{X}} c\big(\alpha(g_0)(\boldsymbol{x}), \alpha(g_1 h)(\boldsymbol{x})\big)\mathrm{d}\rho(\boldsymbol{x}) &= \int_{\mathcal{X}} c\big(\alpha(g_0)(\boldsymbol{x}), (\alpha(g_1) \circ \alpha(h))(\boldsymbol{x})\big)\mathrm{d}\rho(\boldsymbol{x}) \\ &= \int_{\mathcal{X}} c\big(\alpha(g_0)(\boldsymbol{x}), (T_h \circ \alpha(g_0))(\boldsymbol{x})\big)\mathrm{d}\rho(\boldsymbol{x}) \\ &= \int_{\mathcal{X}} c\big(\boldsymbol{x}, T_h(\boldsymbol{x})\big)\mathrm{d}\mu_0(\boldsymbol{x}).\end{aligned}$$

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

In the remainder of the proof, assume that c and μ_0 satisfies Assumptions 1-3, and that there exist $h \in \mathrm{Stab}_G(\rho)$ and a c-convex function φ such that T_h satisfies (♡) μ_0-almost surely. Set $\Gamma_h = (\mathrm{id}_{\mathcal{X}}, T_h)_{\#}\mu_0$. Since $T_h \in \mathcal{T}(\mu_0, \mu_1)$, we have $\Gamma_h \in \Pi(\mu_0, \mu_1)$. Moreover, the assumed identity in (♡) is equivalent to

$$\nabla_{\boldsymbol{x}}\varphi(\boldsymbol{x}) + \nabla_{\boldsymbol{x}} c(\boldsymbol{x}, \boldsymbol{y}) = 0 \qquad \Gamma_h\text{-almost surely.}$$

By Theorem 2.1(v), the coupling Γ_h is the unique optimal Kantorovich plan. This proves assertion (ii)(c). Since Γ_h is induced by the deterministic map T_h, the map T_h solves the Monge problem. The uniqueness of T_h up to μ_0-almost everywhere equality follows from Theorem 2.1. This proves (ii)(a). Hence $\mathbb{M}_c(\mu_0, \mu_1) = \int_{\mathcal{X}} c(\boldsymbol{x}, T_h(\boldsymbol{x}))\mathrm{d}\mu_0(\boldsymbol{x})$. As $T_h = \alpha(g_1 h g_0^{-1})$, we have

$$\mathbb{M}_c(\mu_0, \mu_1) = \int_{\mathcal{X}} c\big(\alpha(g_0)(\boldsymbol{z}), \alpha(g_1 h)(\boldsymbol{z})\big)\mathrm{d}\rho(\boldsymbol{z}).$$

Since $h \in \mathrm{Stab}_G(\rho)$, this h is feasible for (LGOP). Therefore

$$\mathbb{J}_c(g_0, g_1) \leq \int_{\mathcal{X}} c\big(\alpha(g_0)(\boldsymbol{z}), \alpha(g_1 h)(\boldsymbol{z})\big)\mathrm{d}\rho(\boldsymbol{z}) = \mathbb{M}_c(\mu_0, \mu_1).$$

Together with assertion (i), the inequality above implies $\mathbb{M}_c(\mu_0,\mu_1) = \mathbb{J}_c(g_0,g_1)$, and that the feasible element h attains the infimum in (LGOP), proving (ii)(b).

Since $(\mathrm{id}_{\mathcal{X}}, T_h)_{\#}\mu_0$ has the same cost as the optimal Monge map, we obtain $\mathbb{K}_c(\mu_0,\mu_1) = \mathbb{M}_c(\mu_0,\mu_1) = \mathbb{J}_c(g_0,g_1)$. This proves (ii)(d). □

4.1. **Quadratic cost.** We now specialize to the quadratic cost. In this setting, the general c-convex certificate becomes a convex-gradient certificate. Moreover, when the group action is induced by an affine representation on the ambient Hilbert space, the orbit-induced transport maps are affine. The purpose of this subsection is to combine these two observations. First, we record the convex-analytic identities that underlie optimality for the quadratic cost. Then, under the affine-representation assumption, we show that self-adjointness and positive semidefiniteness of the linear part of an orbit-induced affine map produce an explicit quadratic c-convex potential. This gives a directly verifiable certificate that the map is optimal, and therefore that the Monge, Kantorovich, and orbit transport values coincide.

Let $\mathcal{H}$ be a finite-dimensional real Hilbert space with inner product $\langle\cdot,\cdot\rangle_{\mathcal{H}}$ and induced norm $\|\cdot\|_{\mathcal{H}}$, and let $\mathcal{X} \subset \mathcal{H}$ be a nonempty open convex set. Unless stated otherwise, the cost function is the quadratic cost $c(\boldsymbol{x},\boldsymbol{y}) = \|\boldsymbol{x}-\boldsymbol{y}\|_{\mathcal{H}}^2$, and the reference measure $\rho \in \mathcal{P}(\mathcal{X})$ satisfies $\rho \ll \mathcal{L}_{\mathcal{H}}$ and $\int_{\mathcal{X}} \|\boldsymbol{x}\|_{\mathcal{H}}^2 \mathrm{d}\rho(\boldsymbol{x}) < \infty$, where $\mathcal{L}_{\mathcal{H}}$ is the Lebesgue measure on $\mathcal{H}$ induced by the Hilbert structure. For a finite-dimensional real Hilbert space $\mathcal{H}$, we write $\mathrm{GL}(\mathcal{H}) = \{A : \mathcal{H} \to \mathcal{H} \mid A \text{ is linear and invertible}\}$ for its general linear group. The following lemma records the basic convex-analytic identities for the quadratic cost that will be used repeatedly in the remainder of the paper.

Lemma 4 (Quadratic c-convexity and c-subdifferential)**.** *For a function $\psi : \mathcal{X} \to \bar{\mathbb{R}}$, define $\bar{\psi}(\boldsymbol{x}) = \psi(\boldsymbol{x}) + \|\boldsymbol{x}\|_{\mathcal{H}}^2$. Then the following hold:*

- (i) *If ψ is c-convex, then $\bar{\psi}$ is convex.*
- (ii) *For $\boldsymbol{x},\boldsymbol{y} \in \mathcal{X}$, $\boldsymbol{y} \in \partial^c\psi(\boldsymbol{x})$ if and only if $2\boldsymbol{y} \in \partial\bar{\psi}(\boldsymbol{x})$, where $\partial\bar{\psi}(\boldsymbol{x})$ denotes the convex subdifferential of $\bar{\psi}$ at $\boldsymbol{x}$.*
- (iii) *If $\mu_0 \ll \mathcal{L}_{\mathcal{H}}$, then Assumption 2 holds.*
- (iv) *c satisfies Assumption 3 on $\mathcal{H}$.*

Proof. Recall that $\partial\bar{\psi}(\boldsymbol{x}) = \{\boldsymbol{p} \in \mathcal{H} : \bar{\psi}(\boldsymbol{z}) \geq \bar{\psi}(\boldsymbol{x}) + \langle \boldsymbol{p}, \boldsymbol{z}-\boldsymbol{x}\rangle_{\mathcal{H}}\ \forall \boldsymbol{z} \in \mathcal{X}\}$. For the quadratic cost, $c(\boldsymbol{x},\boldsymbol{y}) = \|\boldsymbol{x}-\boldsymbol{y}\|_{\mathcal{H}}^2 = \|\boldsymbol{x}\|_{\mathcal{H}}^2 - 2\langle \boldsymbol{x},\boldsymbol{y}\rangle_{\mathcal{H}} + \|\boldsymbol{y}\|_{\mathcal{H}}^2$.

(i) Suppose that ψ is c-convex. Then there exists a function $\phi : \mathcal{X} \to \mathbb{R} \cup \{\pm\infty\}$ such that $\psi(\boldsymbol{x}) = \sup_{\boldsymbol{y}\in\mathcal{X}}\{\phi(\boldsymbol{y}) - c(\boldsymbol{x},\boldsymbol{y})\}$ for all $\boldsymbol{x} \in \mathcal{X}$. Then, we have $\bar{\psi}(\boldsymbol{x}) = \sup_{\boldsymbol{y}\in\mathcal{X}}\{\phi(\boldsymbol{y}) - \|\boldsymbol{y}\|_{\mathcal{H}}^2 + 2\langle \boldsymbol{x},\boldsymbol{y}\rangle_{\mathcal{H}}\}$. For each fixed $\boldsymbol{y} \in \mathcal{X}$, the map $\boldsymbol{x} \mapsto \phi(\boldsymbol{y}) - \|\boldsymbol{y}\|_{\mathcal{H}}^2 + 2\langle \boldsymbol{x},\boldsymbol{y}\rangle_{\mathcal{H}}$ is affine on $\mathcal{H}$. Hence $\bar{\psi}$ is the pointwise supremum of affine functions, and therefore is convex on $\mathcal{X}$.

(ii) Fix $\boldsymbol{x},\boldsymbol{y} \in \mathcal{X}$. By definition, $\boldsymbol{y} \in \partial^c\psi(\boldsymbol{x})$ if and only if $\psi(\boldsymbol{z}) + \|\boldsymbol{z}\|_{\mathcal{H}}^2 \geq \psi(\boldsymbol{x}) + \|\boldsymbol{x}\|_{\mathcal{H}}^2 + 2\langle \boldsymbol{y}, \boldsymbol{z}-\boldsymbol{x}\rangle_{\mathcal{H}}$ for all $\boldsymbol{z} \in \mathcal{X}$, that is, $\bar{\psi}(\boldsymbol{z}) \geq \bar{\psi}(\boldsymbol{x}) + \langle 2\boldsymbol{y}, \boldsymbol{z}-\boldsymbol{x}\rangle_{\mathcal{H}}$ for all $\boldsymbol{z} \in \mathcal{X}$, which corresponds to the condition $2\boldsymbol{y} \in \partial\bar{\psi}(\boldsymbol{x})$.

(iii) Let ψ be c-convex. By part (i), $\bar{\psi}$ is convex. Let $\mathrm{dom}_{\mathrm{eff}}(\bar{\psi}) = \{\boldsymbol{x} \in \mathcal{X} : \bar{\psi}(\boldsymbol{x}) < \infty\}$ be the effective domain of $\bar{\psi}$. Since $\bar{\psi}$ is convex, $\mathrm{dom}_{\mathrm{eff}}(\bar{\psi})$ is convex. By [1, Theorem 2.1.12], $\bar{\psi}$ is locally Lipschitz on the interior of its effective domain. Therefore, by [10, Theorem 3.2], $\bar{\psi}$ is differentiable $\mathcal{L}_{\mathcal{H}}$-almost everywhere on $\mathrm{int}(\mathrm{dom}_{\mathrm{eff}}(\bar{\psi}))$. Since $\psi = \bar{\psi} - \|\cdot\|_{\mathcal{H}}^2$, and $\boldsymbol{x} \mapsto \|\boldsymbol{x}\|_{\mathcal{H}}^2$ is smooth, it follows that ψ is also differentiable $\mathcal{L}_{\mathcal{H}}$-almost everywhere on $\mathrm{int}(\mathrm{dom}_{\mathrm{eff}}(\bar{\psi}))$. Now let $\boldsymbol{x} \in \mathrm{dom}(\partial^c\psi)$. Then $\partial^c\psi(\boldsymbol{x}) \neq \varnothing$. By part (ii), $\partial\bar{\psi}(\boldsymbol{x}) \neq \varnothing$, and therefore $\bar{\psi}(\boldsymbol{x}) < \infty$. Thus

$\mathrm{dom}(\partial^c\psi) \subseteq \mathrm{dom}_{\mathrm{eff}}(\bar{\psi})$. Since $\mathrm{dom}_{\mathrm{eff}}(\bar{\psi})$ is convex, its boundary has $\mathcal{L}_{\mathcal{H}}$-measure zero. Hence ψ is $\mathcal{L}_{\mathcal{H}}$-almost everywhere differentiable on $\mathrm{dom}(\partial^c\psi)$. If $\mu_0 \ll \mathcal{L}_{\mathcal{H}}$, the same holds μ_0-almost surely. Hence, Assumption 2 holds.

(iv) $\mathcal{H}$ is a flat Riemannian manifold, its Riemannian distance is $d(\boldsymbol{x},\boldsymbol{y}) = \|\boldsymbol{x}-\boldsymbol{y}\|_{\mathcal{H}}$, and its sectional curvature is identically zero. Therefore, [30, Example 10.36] applied with $M = \mathcal{X} = \mathcal{Y} = \mathcal{H}$ and $c(\boldsymbol{x},\boldsymbol{y}) = d(\boldsymbol{x},\boldsymbol{y})^2$, implies that the quadratic cost $c(x,y) = \|x-y\|_{\mathcal{H}}^2$ satisfies Assumption 3. ☐

The preceding lemma shows that, for the quadratic cost, the shifted potential $\bar{\psi} = \psi + \|\cdot\|_{\mathcal{H}}^2$ is convex, and the c-subdifferential relation becomes an ordinary subgradient relation for this shifted potential. Consequently, when the potential is differentiable, the first-order condition in Theorem 2.1 identifies the optimal map as the gradient of a convex function, in agreement with Brenier's convex-gradient characterization of quadratic optimal transport [4, Theorem 1.1-1.2].

This observation makes the quadratic case especially amenable to the orbit framework. If the group action is affine on the ambient Hilbert space, then every orbit-induced candidate map has an affine form. For such maps, being the gradient of a convex quadratic potential reduces to an algebraic condition: the linear part must be self-adjoint and positive semidefinite. We therefore introduce affine actions in a form that separates the translation part from the linear part.

Definition 6. *Let G be a Lie group acting on $\mathcal{X}$ via a homomorphism $\alpha : G \to \mathrm{Diff}(\mathcal{X})$. The action α is induced by an affine representation on $\mathcal{H}$ if there exist maps $b : G \to \mathcal{H}$, $\pi : G \to \mathrm{GL}(\mathcal{H})$, such that $\alpha(g)(\boldsymbol{x}) = b(g) + \pi(g)\boldsymbol{x} \in \mathcal{X}$ for all $g \in G$, $\forall \boldsymbol{x} \in \mathcal{X}$, and for all $g_1, g_2 \in G$: $\pi(g_1g_2) = \pi(g_1)\pi(g_2)$, and $b(g_1g_2) = b(g_1) + \pi(g_1)b(g_2)$.*

Assumption 4. *The action $\alpha : G \to \mathrm{Diff}(\mathcal{X})$ is induced by an affine representation on $\mathcal{H}$ in the sense of Definition 6, with associated maps $b : G \to \mathcal{H}$ and $\pi : G \to \mathrm{GL}(\mathcal{H})$.*

Under Assumption 4, the Lie group orbit transport problem (LGOP) becomes

$$\mathbb{J}_c(g_0,g_1) = \inf_{h \in \mathrm{Stab}_G(\rho)} \int_{\mathcal{X}} \|b(g_0) - b(g_1h) + (\pi(g_0) - \pi(g_1h))\boldsymbol{z}\|_{\mathcal{H}}^2 \,\mathrm{d}\rho(\boldsymbol{z}).$$

Since the integrand is the square of an affine function of $\boldsymbol{z}$ and ρ has finite second moment, $\mathbb{J}_c(g_0,g_1) < \infty$. Moreover, for $i = 0,1$, the measure $\mu_i = \alpha(g_i)_{\#}\rho$ is absolutely continuous with respect to $\mathcal{L}_{\mathcal{H}}$ by Lemma 1. Finally, for every $h \in \mathrm{Stab}_G(\rho)$, the associated transport map $T_h = \alpha(g_1hg_0^{-1})$ admits the following explicit form $T_h(\boldsymbol{x}) = b(g_1hg_0^{-1}) + \pi(g_1hg_0^{-1})\boldsymbol{x}$ and belongs to $\mathcal{T}(\mu_0,\mu_1)$ by Lemma 2. Therefore, $\mathbb{K}_c(\mu_0,\mu_1) \le \mathbb{M}_c(\mu_0,\mu_1) \le \mathbb{J}_c(g_0,g_1) < \infty$.

The orbit problem thus yields a concrete family of admissible affine transport maps, one for each $h \in \mathrm{Stab}_G(\rho)$. Since an affine map is the gradient of a convex quadratic if and only if its linear part is self-adjoint and positive semidefinite, Theorem 2.1 reduces the optimality question to a purely algebraic condition on $\pi(g_1hg_0^{-1})$. The following result shows that, whenever this condition holds for some $h \in \mathrm{Stab}_G(\rho)$, the orbit upper bound $\mathbb{J}_c$ is tight and coincides with both $\mathbb{M}_c$ and $\mathbb{K}_c$.

Theorem 4.2. *Suppose Assumption 4 holds. If there exists $h \in \mathrm{Stab}_G(\rho)$ such that the operator $\pi(g_1hg_0^{-1})$ is self-adjoint and positive semidefinite on $\mathcal{H}$, then*

(i) *$T_h(\boldsymbol{x}) = b(g_1hg_0^{-1}) + \pi(g_1hg_0^{-1})\boldsymbol{x}$ solves (MP), and it is the unique optimal Monge map up to μ_0-almost everywhere equality,*
(ii) *$(\mathrm{id}_{\mathcal{X}}, T_h)_{\#}\mu_0 \in \Pi(\mu_0,\mu_1)$ solves (KTP), and it is the unique optimal Kantorovich plan,*
(iii) *h solves (LGOP),*
(iv) *$\mathbb{K}_c(\mu_0,\mu_1) = \mathbb{M}_c(\mu_0,\mu_1) = \mathbb{J}_c(g_0,g_1)$.*

Proof. For convenience, define $m_h = b(g_1 h g_0^{-1})$ and $A_h = \pi(g_1 h g_0^{-1})$. Then, $T_h(\boldsymbol{x}) = m_h + A_h \boldsymbol{x}$, $\boldsymbol{x} \in \mathcal{X}$. To verify that T_h satisfies (♡) for some c-convex function, define $\varphi_h(\boldsymbol{x}) = -\|\boldsymbol{x}\|_{\mathcal{H}}^2 + 2\langle m_h, \boldsymbol{x}\rangle_{\mathcal{H}} + \langle \boldsymbol{x}, A_h \boldsymbol{x}\rangle_{\mathcal{H}}$, $\boldsymbol{x} \in \mathcal{X}$. Set $u_h(\boldsymbol{x}) \triangleq \varphi_h(\boldsymbol{x}) + \|\boldsymbol{x}\|_{\mathcal{H}}^2 = 2\langle m_h, \boldsymbol{x}\rangle_{\mathcal{H}} + \langle \boldsymbol{x}, A_h \boldsymbol{x}\rangle_{\mathcal{H}}$. Since A_h is self-adjoint and positive semidefinite, u_h is convex. Moreover, $\nabla_{\boldsymbol{x}} u_h(\boldsymbol{z}) = 2m_h + 2A_h \boldsymbol{z} = 2T_h(\boldsymbol{z})$. Thus the supporting hyperplane inequality for u_h gives

$$u_h(\boldsymbol{x}) \geq u_h(\boldsymbol{z}) + 2\langle T_h(\boldsymbol{z}), \boldsymbol{x} - \boldsymbol{z}\rangle_{\mathcal{H}} \qquad \forall \boldsymbol{x}, \boldsymbol{z} \in \mathcal{X}.$$

Equivalently, because $c(\boldsymbol{z}, T_h(\boldsymbol{z})) - c(\boldsymbol{x}, T_h(\boldsymbol{z})) = 2\langle \boldsymbol{x} - \boldsymbol{z}, T_h(\boldsymbol{z})\rangle_{\mathcal{H}} + \|\boldsymbol{z}\|_{\mathcal{H}}^2 - \|\boldsymbol{x}\|_{\mathcal{H}}^2$, we have

$$\varphi_h(\boldsymbol{x}) \geq \varphi_h(\boldsymbol{z}) + c(\boldsymbol{z}, T_h(\boldsymbol{z})) - c(\boldsymbol{x}, T_h(\boldsymbol{z})) \qquad \forall \boldsymbol{x}, \boldsymbol{z} \in \mathcal{X}.$$

Now define $\eta_h(\boldsymbol{y}) \triangleq \sup\{\varphi_h(\boldsymbol{z}) + c(\boldsymbol{z}, T_h(\boldsymbol{z})) \mid \boldsymbol{z} \in \mathcal{Z}_{\boldsymbol{y}}\}$, where $\mathcal{Z}_{\boldsymbol{y}} = \{\boldsymbol{z} \in \mathcal{X} \mid T_h(\boldsymbol{z}) = \boldsymbol{y}\}$ with the convention that the supremum over the empty set is $-\infty$. The preceding inequality implies $\varphi_h(\boldsymbol{x}) \geq \eta_h(\boldsymbol{y}) - c(\boldsymbol{x}, \boldsymbol{y})$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$. Taking supremum over $\boldsymbol{y}$ results in

$$\varphi_h(\boldsymbol{x}) \geq \sup_{\boldsymbol{y} \in \mathcal{X}} \eta_h(\boldsymbol{y}) - c(\boldsymbol{x}, \boldsymbol{y}) \quad \forall \boldsymbol{x} \in \mathcal{X}.$$

Conversely, fix $\boldsymbol{x} \in \mathcal{X}$ and take $\boldsymbol{y} = T_h(\boldsymbol{x})$. Then $\boldsymbol{x} \in \mathcal{Z}_{\boldsymbol{y}}$, so $\eta_h(T_h(\boldsymbol{x})) \geq \varphi_h(\boldsymbol{x}) + c(\boldsymbol{x}, T_h(\boldsymbol{x}))$. Therefore

$$\sup_{\boldsymbol{y} \in \mathcal{X}} \eta_h(\boldsymbol{y}) - c(\boldsymbol{x}, \boldsymbol{y}) \geq \eta_h(T_h(\boldsymbol{x})) - c(\boldsymbol{x}, T_h(\boldsymbol{x})) \geq \varphi_h(\boldsymbol{x}).$$

Combining the two inequalities gives $\varphi_h(\boldsymbol{x}) = \sup_{\boldsymbol{y} \in \mathcal{X}} \eta_h(\boldsymbol{y}) - c(\boldsymbol{x}, \boldsymbol{y})$, so φ_h is c-convex.

Since φ_h is a quadratic polynomial on $\mathcal{H}$, it is of class $\mathcal{C}^1$ on $\mathcal{X}$. Furthermore, $\nabla_{\boldsymbol{x}} \langle \boldsymbol{x}, A_h \boldsymbol{x}\rangle_{\mathcal{H}} = (A_h + A_h^*)\boldsymbol{x}$, and because A_h is self-adjoint, we have $\nabla_{\boldsymbol{x}} \langle \boldsymbol{x}, A_h \boldsymbol{x}\rangle_{\mathcal{H}} = 2A_h \boldsymbol{x}$. Hence $\nabla_{\boldsymbol{x}} \varphi_h(\boldsymbol{x}) = -2\boldsymbol{x} + 2m_h + 2A_h \boldsymbol{x}$. On the other hand, $T_h(\boldsymbol{x}) = m_h + A_h \boldsymbol{x}$, so

$$\nabla_{\boldsymbol{x}} c(\boldsymbol{x}, T_h(\boldsymbol{x})) = 2(\boldsymbol{x} - T_h(\boldsymbol{x})) = 2\boldsymbol{x} - 2m_h - 2A_h \boldsymbol{x} = -\nabla_{\boldsymbol{x}} \varphi_h(\boldsymbol{x}).$$

Thus T_h satisfies (♡) everywhere on $\mathcal{X}$ with the c-convex potential φ_h.

Observe first that the preceding c-convexity argument holds verbatim on the ambient Hilbert space $\mathcal{H}$, since u_h is a convex quadratic on $\mathcal{H}$. Let $\widetilde{c}(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|_{\mathcal{H}}^2$, and view μ_i as probability measures $\widetilde{\mu}_i$ on $\mathcal{H}$ by setting $\widetilde{\mu}_i(B) = \mu_i(B \cap \mathcal{X})$ for every $B \in \mathcal{B}(\mathcal{H})$. Then $\widetilde{\mu}_0 \ll \mathcal{L}_{\mathcal{H}}$, the cost $\widetilde{c}$ satisfies Assumption 1, and $\nabla_{\boldsymbol{x}} \widetilde{c}(\boldsymbol{x}, \boldsymbol{y}) = 2(\boldsymbol{x} - \boldsymbol{y})$, so injectivity in $\boldsymbol{y}$ is immediate. By Lemma 4(iii)-(iv), Assumptions 2 and 3 hold on $\mathcal{H}$. Moreover, $\mathbb{K}_{\widetilde{c}}(\widetilde{\mu}_0, \widetilde{\mu}_1) < \infty$, because the affine candidate has finite quadratic cost. Hence Theorem 2.1 applied on $\mathcal{H}$ shows that T_h is the unique optimal Monge map and that $(\mathrm{id}_{\mathcal{H}}, T_h)_{\#}\widetilde{\mu}_0$ is the unique optimal Kantorovich plan. Since the extended measures are concentrated on $\mathcal{X}$ and $\{T_h(\boldsymbol{x}) : \boldsymbol{x} \in \mathcal{X}\} \subseteq \mathcal{X}$, these ambient optimality statements restrict to the original problem on $\mathcal{X}$. This proves (i) and (ii).

Next, by (i), we have $\mathbb{M}_c(\mu_0, \mu_1) = \int_{\mathcal{X}} c(\boldsymbol{x}, T_h(\boldsymbol{x}))\mathrm{d}\mu_0(\boldsymbol{x})$. Using $\mu_0 = \alpha(g_0)_{\#}\rho$ and the identity $T_h \circ \alpha(g_0) = \alpha(g_1) \circ \alpha(h) \circ \alpha(g_0)^{-1} \circ \alpha(g_0) = \alpha(g_1) \circ \alpha(h)$, we obtain

$$\mathbb{M}_c(\mu_0, \mu_1) = \int_{\mathcal{X}} c\big(\alpha(g_0)(\boldsymbol{z}), (T_h \circ \alpha(g_0))(\boldsymbol{z})\big)\mathrm{d}\rho(\boldsymbol{z}) = \int_{\mathcal{X}} c\big(\alpha(g_0)(\boldsymbol{z}), (\alpha(g_1) \circ \alpha(h))(\boldsymbol{z})\big)\mathrm{d}\rho(\boldsymbol{z}).$$

Since $h \in \mathrm{Stab}_G(\rho)$, h is feasible for (LGOP). Hence, by definition of $\mathbb{J}_c(g_0, g_1)$,

$$\mathbb{J}_c(g_0, g_1) \leq \int_{\mathcal{X}} c\big(\alpha(g_0)(\boldsymbol{z}), (\alpha(g_1) \circ \alpha(h))(\boldsymbol{z})\big)\mathrm{d}\rho(\boldsymbol{z}) = \mathbb{M}_c(\mu_0, \mu_1).$$

On the other hand, Theorem 4.1(i) yields $\mathbb{M}_c(\mu_0, \mu_1) \leq \mathbb{J}_c(g_0, g_1)$. Therefore, $\mathbb{M}_c(\mu_0, \mu_1) = \mathbb{J}_c(g_0, g_1)$, and consequently h attains the infimum in (LGOP), proving (iii).

Finally, since T_h solves (MP) and $(\mathrm{id}_{\mathcal{X}}, T_h)_{\#}\mu_0$ solves (KTP), both problems attain the same cost. This observation proves assertion (iv) and completes the proof. □

Indeed, Theorem 4.2 reduces optimality to an algebraic question: can one choose a stabilizing element $h \in \mathrm{Stab}_G(\rho)$ such that the linear part $\pi(g_1 h g_0^{-1})$ is self-adjoint and positive semidefinite on $\mathcal{H}$? We now give a structural criterion, based on Cartan theory, that guarantees the existence of such a stabilizing element.

We begin with the necessary Lie-algebraic background. Let L be a Lie group with identity element e. We write $\mathrm{Lie}(L)$ for its Lie algebra, that is, the tangent space $T_e L$ endowed with its canonical Lie bracket. If $\theta : L \to L$ is a Lie group involution, then $\theta^2 = \mathrm{id}_L$. Differentiating this identity at the identity element gives $(\mathrm{d}\theta_e)^2 = \mathrm{id}_{\mathrm{Lie}(L)}$. Thus $\mathrm{d}\theta_e$ is a linear involution on $\mathrm{Lie}(L)$. Since the polynomial $t^2 - 1 = (t-1)(t+1)$ has distinct roots, $\mathrm{Lie}(L)$ splits as the direct sum of the $+1$ and -1 eigenspaces of $\mathrm{d}\theta_e$: $\mathrm{Lie}(L) = \mathfrak{k} \oplus \mathfrak{p}$, where

$$\mathfrak{k} = \{X \in \mathrm{Lie}(L) : \mathrm{d}\theta_e(X) = X\}, \qquad \mathfrak{p} = \{X \in \mathrm{Lie}(L) : \mathrm{d}\theta_e(X) = -X\}.$$

This Lie-algebra decomposition follows only from the fact that θ is an involution.

The Cartan property used below is stronger than this infinitesimal splitting. It is a global factorization statement: every element of the group can be written uniquely and smoothly as a product of a fixed-point factor and the exponential of an element from the -1-eigenspace. In this factorization, the fixed-point subgroup plays the role of a rotation-like part, while the exponential factor plays the role of a symmetric or stretching part. A standard sufficient condition for this global factorization is that L is a real reductive linear Lie group and that θ is a Cartan involution; see, for example, [17, § VII.2].

In our application, this Cartan structure need not be imposed on the full acting group G. The quadratic optimality certificate presented in Theorem 4.2 depends only on the linear part of the candidate map. We therefore impose the Cartan condition on the linear image $L \triangleq \pi(G) \subseteq \mathrm{GL}(\mathcal{H})$.

Definition 7 (Cartan decomposition)**.** *Let $L \subseteq \mathrm{GL}(\mathcal{H})$ be a Lie subgroup, and let $\theta : L \to L$ be a Lie group involution. Write*

$$\mathfrak{l} = \mathrm{Lie}(L), \qquad \mathfrak{k} = \{X \in \mathfrak{l} : \mathrm{d}\theta_e(X) = X\}, \qquad \mathfrak{p} = \{X \in \mathfrak{l} : \mathrm{d}\theta_e(X) = -X\},$$

and let $L^\theta = \{\ell \in L : \theta(\ell) = \ell\}$. We say that (L, θ) admits a global Cartan decomposition *if the map $L^\theta \times \mathfrak{p} \to L$, $(k, S) \mapsto k \exp(S)$, is a diffeomorphism.*

Lemma 5 (Cartan absorption)**.** *Let (L, θ) admit a global Cartan decomposition in the sense of Definition 7. Then, for every $\ell_0, \ell_1 \in L$, there exists $k \in L^\theta$ such that $\ell_1 k \ell_0^{-1} \in \exp(\mathfrak{p})$.*

Proof. Define $\eta : L \to L$ such that $\eta(\ell) = \theta(\ell)^{-1}\ell$. We will first show that $\{\eta(\ell) : \ell \in L\} = \exp(\mathfrak{p})$. Let $\ell \in L$. By the Cartan decomposition, there exist $k \in L^\theta$ and $X \in \mathfrak{p}$ such that $\ell = k\exp(X)$. Since $\theta(k) = k$ and $\mathrm{d}\theta_e(X) = -X$, we have

$$\theta(\exp(X)) = \exp(\mathrm{d}\theta_e(X)) = \exp(-X). \tag{1}$$

Therefore

$$\begin{aligned}\eta(\ell) &= \theta(k\exp(X))^{-1}k\exp(X) = (\theta(k)\theta(\exp(X)))^{-1}k\exp(X)\\ &= (k\exp(-X))^{-1}k\exp(X) = \exp(2X) \in \exp(\mathfrak{p}),\end{aligned}$$

which implies $\{\eta(\ell) : \ell \in L\} \subseteq \exp(\mathfrak{p})$. Conversely, if $p = \exp(Y) \in \exp(\mathfrak{p})$ with $Y \in \mathfrak{p}$, then

$$\eta(\exp(Y/2)) = \theta(\exp(Y/2))^{-1}\exp(Y/2) = \exp(-Y/2)^{-1}\exp(Y/2) = \exp(Y) = p,$$

where the second equality follows by (1) Hence $\exp(\mathfrak{p}) \subseteq \{\eta(\ell) : \ell \in L\}$, and thus $\{\eta(\ell) : \ell \in L\} = \exp(\mathfrak{p})$.

Next, for each $a \in L$, define the map $\tau_a : \exp(\mathfrak{p}) \to L$ as $\tau_a(p) = \theta(a)^{-1}pa$. In what follows, we will show that $\{\tau_a(p) : p \in \exp(\mathfrak{p})\} \subseteq \exp(\mathfrak{p})$ for every $a \in L$. Indeed, if $p \in \exp(\mathfrak{p})$, then by $\{\eta(\ell) : \ell \in L\} = \exp(\mathfrak{p})$, there exists $x \in L$ such that $p = \eta(x) = \theta(x)^{-1}x$. Therefore

$$\tau_a(p) = \theta(a)^{-1}\theta(x)^{-1}xa = \theta(xa)^{-1}(xa) = \eta(xa) \in \exp(\mathfrak{p}).$$

Thus $\exp(\mathfrak{p})$ is invariant under τ_a. Now fix $\ell_0, \ell_1 \in L$. By the Cartan decomposition, we may write $\theta(\ell_0)^{-1}\ell_1 = kp$ for some $k \in L^\theta,\ p \in \exp(\mathfrak{p})$. Set $h = k^{-1} \in L^\theta$. Since $k \in L^\theta$, we have $\theta(k) = k$ by definition of L^θ. As group automorphisms preserve inverses $\theta(k^{-1}) = \theta(k)^{-1} = k^{-1}$. Hence k^{-1} indeed belongs to L^θ and we have $\theta(k^{-1})^{-1} = k$. Therefore,

$$kpk^{-1} = \theta(k^{-1})^{-1}pk^{-1} = \tau_{k^{-1}}(p) \in \exp(\mathfrak{p}).$$

Now, define $q = kpk^{-1} \in \exp(\mathfrak{p})$. Then

$$\ell_1 h\ell_0^{-1} = \ell_1 k^{-1}\ell_0^{-1} = \theta(\ell_0)kpk^{-1}\ell_0^{-1} = \theta(\ell_0)q\ell_0^{-1} = \tau_{\ell_0^{-1}}(q) \in \exp(\mathfrak{p}),$$

where the second equality follows because $\theta(\ell_0)^{-1}\ell_1 = kp$, and thus $\ell_1 = \theta(\ell_0)kp$, and the inclusion follows because $\exp(\mathfrak{p})$ is invariant under τ_a. This concludes our proof. □

In words, Lemma 5 shows that, after right-composing by a suitable element of L^θ, the relative element between ℓ_0 and ℓ_1 can always be moved into the symmetric factor $\exp(\mathfrak{p})$. Thus the compact part of the Cartan decomposition can be absorbed into a stabilizing group element, leaving only the noncompact factor.

Theorem 4.3 (Cartan criterion)**.** *Suppose Assumption 4 holds, and set* $L = \pi(G) \subseteq \mathrm{GL}(\mathcal{H})$. *Assume that* (L, θ) *admits a global Cartan decomposition in the sense of Definition 7, with* $\mathrm{Lie}(L) = \mathfrak{k} \oplus \mathfrak{p}$. *Assume further that:*

(i) *every* $S \in \mathfrak{p}$ *is self-adjoint on* $\mathcal{H}$*,*
(ii) $L^\theta \subseteq \pi(\mathrm{Stab}_G(\rho))$.

Then there exists $h \in \mathrm{Stab}_G(\rho)$*, with* $\pi(h) \in L^\theta$*, such that:*

(a) $\pi(g_1 h g_0^{-1})$ *is self-adjoint and positive definite on* $\mathcal{H}$*,*
(b) $T_h = \alpha(g_1 h g_0^{-1})$ *solves* (MP)*, and is the unique optimal Monge map up to* μ_0*-almost everywhere equality,*
(c) $(\mathrm{id}_{\mathcal{X}}, T_h)_{\#}\mu_0$ *solves* (KTP)*, and is the unique optimal Kantorovich plan,*
(d) h *solves* (LGOP)*,*
(e) $\mathbb{K}_c(\mu_0, \mu_1) = \mathbb{M}_c(\mu_0, \mu_1) = \mathbb{J}_c(g_0, g_1)$.

Proof. Set $\ell_i = \pi(g_i)$ for $i = 0, 1$. By Lemma 5, there exists $k \in L^\theta$ such that $\ell_1 k\ell_0^{-1} \in \exp(\mathfrak{p})$. By assumption (ii), choose $h \in \mathrm{Stab}_G(\rho)$ such that $\pi(h) = k$. Then $\pi(g_1 h g_0^{-1}) = \ell_1 k \ell_0^{-1} = \exp(S)$ for some $S \in \mathfrak{p}$. By assumption (i), the operator S is self-adjoint on $\mathcal{H}$. Hence $\exp(S)$ is self-adjoint and positive definite on $\mathcal{H}$. This proves (a). The remaining assertions follow from Theorem 4.2. □

5. Examples

In the examples below, we separate the group-theoretic mechanism from the choice of reference law. For each mechanism, we first specify the state space, the acting group, and the affine representation inducing the action. We then identify the linear image $L = \pi(G)$ and verify the Cartan structure and self-adjointness hypotheses of Theorem 4.3. Next, we identify reference laws ρ for which the compact invariance condition $L^\theta \subseteq \pi(\mathrm{Stab}_G(\rho))$ holds. Once these structural hypotheses are verified, we apply Theorem 4.3 to obtain the optimal Monge map, the optimal Kantorovich plan, and the closed-form transport value.

5.1. **The affine mechanism on $\mathbb{R}^d$.** Let the state space be $\mathcal{X} = \mathbb{R}^d$, and let G be the affine group $G_{\text{aff}} = \mathbb{R}^d \rtimes \text{GL}(d)$ with multiplication $(\boldsymbol{m}, \boldsymbol{A})(\boldsymbol{n}, \boldsymbol{B}) = (\boldsymbol{m} + \boldsymbol{A}\boldsymbol{n}, \boldsymbol{A}\boldsymbol{B})$ acting on $\mathbb{R}^d$ by $\alpha : G_{\text{aff}} \times \mathbb{R}^d \to \mathbb{R}^d,\ \alpha((\boldsymbol{m}, \boldsymbol{A}))(\boldsymbol{x}) = \boldsymbol{m} + \boldsymbol{A}\boldsymbol{x}$. The action α is induced by an affine representation on $\mathcal{H} = \mathbb{R}^d$, equipped with the standard inner product $\langle \boldsymbol{x}, \boldsymbol{y} \rangle_{\mathcal{H}} = \boldsymbol{x}^\top \boldsymbol{y}$ in the sense of Definition 6, with associated maps $b : G_{\text{aff}} \to \mathbb{R}^d$ and $\pi : G_{\text{aff}} \to \text{GL}(\mathbb{R}^d)$ given by $b((\boldsymbol{m}, \boldsymbol{A})) = \boldsymbol{m}$ and $\pi((\boldsymbol{m}, \boldsymbol{A})) = \boldsymbol{A}$. Thus $\alpha(g)(\boldsymbol{x}) = b(g) + \pi(g)(\boldsymbol{x})$ for every $g \in G_{\text{aff}}$ and $\boldsymbol{x} \in \mathbb{R}^d$ and Assumption 4 holds.
Cartan structure. The relevant Cartan structure is carried by the linear part: $L = \pi(G_{\text{aff}}) = \text{GL}(d)$, equipped with the Cartan involution $\theta(\boldsymbol{A}) = \boldsymbol{A}^{-\top}$, and thus $L^\theta = \mathcal{O}(d)$. The differential of θ at the identity is $\text{d}\theta_e(\boldsymbol{X}) = -\boldsymbol{X}^\top$, so $\mathfrak{p} = \{\boldsymbol{X} \in \mathfrak{gl}(d) : \text{d}\theta_e(\boldsymbol{X}) = -\boldsymbol{X}\} = \{\boldsymbol{X} \in \mathbb{R}^{d\times d} : \boldsymbol{X}^\top = \boldsymbol{X}\} = \text{Sym}(d)$. Since $L = \text{GL}(d)$ and $\theta(\boldsymbol{A}) = \boldsymbol{A}^{-\top}$ is the standard Cartan involution, the global Cartan decomposition is the polar decomposition $\text{GL}(d) = \mathcal{O}(d)\exp(\text{Sym}(d))$. Thus the map $\mathcal{O}(d) \times \text{Sym}(d) \to \text{GL}(d)$, $(\boldsymbol{Q}, \boldsymbol{S}) \mapsto \boldsymbol{Q}\exp(\boldsymbol{S})$, is a diffeomorphism, and (L, θ) admits a global Cartan decomposition in the sense of Definition 7. Finally, note that every $\boldsymbol{S} \in \mathfrak{p}$ satisfies $\langle \boldsymbol{S}\boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^\top \boldsymbol{S}\boldsymbol{y} = \langle \boldsymbol{x}, \boldsymbol{S}\boldsymbol{y} \rangle$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$. Therefore, the linear part of the affine mechanism satisfies hypothesis (i) of Theorem 4.3. Thus the structural part of Theorem 4.3 is already in place for the affine mechanism.
Reference distributions. We now identify a class of reference laws ρ for which the compact invariance condition (ii) of Theorem 4.3 also holds, so that the theorem applies on the corresponding affine orbit.

Definition 8 (Elliptical family)**.** *A random vector $\boldsymbol{x} \in \mathbb{R}^d$ has an elliptical distribution with characteristic generator $\varsigma : \mathbb{R}_{\geq} \to \mathbb{R}$ if its characteristic function evaluated at $\boldsymbol{t} \in \mathbb{R}^d$ 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 $\boldsymbol{S} \in \mathbb{S}^d_{\succ}$; and we write $\boldsymbol{x} \sim \mathcal{E}_\varsigma(\boldsymbol{m}, \boldsymbol{S})$. When $\varsigma'(0)$ exists and is finite, then the covariance matrix of $\boldsymbol{x} \sim \mathcal{E}_\varsigma(\boldsymbol{m}, \boldsymbol{S})$, satisfies $\text{Cov}(\boldsymbol{x}) = (-2\varsigma'(0))\boldsymbol{S}$. Reparameterizing the generator ς as $\tilde{\varsigma}(s) = \varsigma(s/(-2\varsigma'(0)))$ yields $\text{Cov}(\boldsymbol{x}) = \boldsymbol{S}$ with $\boldsymbol{x} \sim \mathcal{E}_{\tilde{\varsigma}}(\boldsymbol{m}, \boldsymbol{S})$.*

The elliptical class is broad and includes heavy-tailed examples such as Cauchy and multivariate stable laws. In this subsection, however, we restrict to absolutely continuous elliptical laws with finite second moment, so that the quadratic cost is finite. This finite-moment subclass includes the multivariate Gaussian, finite-variance Student-t, and multivariate Laplace families. We work throughout with the reparametrized generator $\tilde{\varsigma}$ and set $\rho = \mathcal{E}_{\tilde{\varsigma}}(\boldsymbol{0}_d, \boldsymbol{I}_d)$.

Lemma 6. (i) $\{\alpha(g)_\# \rho : g \in G_{\text{aff}}\} = \{\mathcal{E}_{\tilde{\varsigma}}(\boldsymbol{m}, \boldsymbol{S}) : \boldsymbol{m} \in \mathbb{R}^d, \boldsymbol{S} \in \mathbb{S}^d_{\succ}\}$, (ii) $\text{Stab}_{G_{\text{aff}}}(\rho) = \{(\boldsymbol{0}_d, \boldsymbol{Q}) : \boldsymbol{Q} \in \mathcal{O}(d)\}$.

Proof. (i) Follows by [22, § 1.5]. Indeed, if $\boldsymbol{X} \sim \rho$ and $\boldsymbol{Y} = \boldsymbol{m} + \boldsymbol{A}\boldsymbol{X}$, then characteristic function of $\boldsymbol{Y}$ evaluated at $\boldsymbol{t} \in \mathbb{R}^d$ is $\exp(i\boldsymbol{t}^\top \boldsymbol{m})\tilde{\varsigma}(\boldsymbol{t}^\top \boldsymbol{A}\boldsymbol{A}^\top \boldsymbol{t})$, so $\boldsymbol{Y} \sim \mathcal{E}_{\tilde{\varsigma}}(\boldsymbol{m}, \boldsymbol{A}\boldsymbol{A}^\top)$. Conversely, every $\boldsymbol{S} \in \mathbb{S}^d_{\succ}$ can be written as $\boldsymbol{S} = \boldsymbol{A}\boldsymbol{A}^\top$ with $\boldsymbol{A} = \boldsymbol{S}^{1/2}$.

(ii) Let $(\boldsymbol{u}, \boldsymbol{A}) \in \text{Stab}_{G_{\text{aff}}}(\rho)$ and let $\boldsymbol{x} \sim \rho$. Then $\boldsymbol{u} + \boldsymbol{A}\boldsymbol{x} \sim \rho$. Since ρ has mean $\boldsymbol{0}_d$, taking expectations gives $\boldsymbol{u} + \boldsymbol{A}\mathbb{E}_{\boldsymbol{X}\sim\rho}[\boldsymbol{X}] = \boldsymbol{0}_d$, and hence $\boldsymbol{u} = \boldsymbol{0}_d$. Since ρ has covariance $\boldsymbol{I}_d$, taking covariances gives $\boldsymbol{A}\text{Cov}(\boldsymbol{x})\boldsymbol{A}^\top = \text{Cov}(\boldsymbol{x})$, that is, $\boldsymbol{A}\boldsymbol{A}^\top = \boldsymbol{I}_d$. Thus $\boldsymbol{A} \in \mathcal{O}(d)$, so $\text{Stab}_{G_{\text{aff}}}(\rho) \subseteq \{(\boldsymbol{0}_d, \boldsymbol{Q}) : \boldsymbol{Q} \in \mathcal{O}(d)\}$. Conversely, let $\boldsymbol{Q} \in \mathcal{O}(d)$. The characteristic function of ρ is $\phi_\rho(\boldsymbol{t}) = \tilde{\varsigma}(\|\boldsymbol{t}\|^2)$. The characteristic function of $\boldsymbol{Q}\boldsymbol{x}$ is $\boldsymbol{t} \mapsto \phi_\rho(\boldsymbol{Q}^\top \boldsymbol{t}) = \tilde{\varsigma}(\|\boldsymbol{Q}^\top \boldsymbol{t}\|^2) = \tilde{\varsigma}(\|\boldsymbol{t}\|^2) = \phi_\rho(\boldsymbol{t})$. Hence $\boldsymbol{Q}\boldsymbol{x} \sim \boldsymbol{x}$, and therefore $(\boldsymbol{0}_d, \boldsymbol{Q}) \in \text{Stab}_{G_{\text{aff}}}(\rho)$. ☐

We now verify the hypotheses of Theorem 4.3. By Lemma 6(ii) $\pi(\text{Stab}_{G_{\text{aff}}}(\rho)) = \mathcal{O}(d) = L^\theta$. Thus hypothesis (ii) of Theorem 4.3 also holds. Consequently, all hypotheses of Theorem 4.3 are satisfied for the affine mechanism.

Corollary 1 (Optimal transport between elliptical distributions)**.** *Let $\mu_i = \mathcal{E}_{\tilde{\varsigma}}(\boldsymbol{m}_i, \boldsymbol{\Sigma}_i)$ with $\boldsymbol{\Sigma}_i \in \mathbb{S}^d_{\succ}$, $i = 0, 1$ and $c(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|^2_{\mathcal{H}}$. Then:*

(a) *the unique optimal Monge map from μ_0 to μ_1 is $T^\star(\boldsymbol{x}) = \boldsymbol{m}_1 + \boldsymbol{A}^\star(\boldsymbol{x} - \boldsymbol{m}_0)$, where $\boldsymbol{A}^\star = \boldsymbol{\Sigma}_0^{-1/2}\big(\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{1/2}\big)^{1/2}\boldsymbol{\Sigma}_0^{-1/2}$,*
(b) *$(\mathrm{id}_{\mathbb{R}^d}, T^\star)_\# \mu_0$ is the unique optimal Kantorovich plan,*
(c) *$\mathbb{K}_c(\mu_0, \mu_1) = \|\boldsymbol{m}_0 - \boldsymbol{m}_1\|^2 + \mathrm{Tr}(\boldsymbol{\Sigma}_0) + \mathrm{Tr}(\boldsymbol{\Sigma}_1) - 2\,\mathrm{Tr}\big((\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{1/2})^{1/2}\big)$,*
(d) *$\mathbb{M}_c(\mu_0, \mu_1) = \mathbb{K}_c(\mu_0, \mu_1)$.*

Proof. For $i = 0, 1$, write $g_i = (\boldsymbol{m}_i, \boldsymbol{\Sigma}_i^{1/2}) \in G_{\mathrm{aff}}$ so that $\mu_i = \alpha(g_i)_\# \rho$, where $\rho = \mathcal{E}_{\tilde{\varsigma}}(\boldsymbol{0}_d, \boldsymbol{I}_d)$. By Theorem 4.3, there exists $h \in \mathrm{Stab}_{G_{\mathrm{aff}}}(\rho)$ with $\pi(h) \in \mathcal{O}(d)$ such that $\pi(g_1 h g_0^{-1})$ is self-adjoint and positive definite. By Lemma 6(ii), $h = (\boldsymbol{0}, \boldsymbol{Q})$ for some $\boldsymbol{Q} \in \mathcal{O}(d)$. A direct computation gives $T_h = \alpha(g_1) \circ \alpha(h) \circ \alpha(g_0)^{-1}$, and hence $T_h(\boldsymbol{x}) = \boldsymbol{m}_1 + \boldsymbol{A}_h(\boldsymbol{x} - \boldsymbol{m}_0)$, where $\boldsymbol{A}_h = \pi(g_1)\pi(h)\pi(g_0)^{-1} = \boldsymbol{\Sigma}_1^{1/2}\boldsymbol{Q}\boldsymbol{\Sigma}_0^{-1/2}$. Therefore, by Theorem 4.3(b), T_h is the unique optimal Monge map from μ_0 to μ_1, up to μ_0-almost everywhere equality.
Identification of $\boldsymbol{A}_h$. Set $\boldsymbol{R} = \boldsymbol{\Sigma}_0^{1/2}\boldsymbol{A}_h\boldsymbol{\Sigma}_0^{1/2} = \boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{Q}$. Because $\pi(g_1 h g_0^{-1})$ is self-adjoint and positive definite on $\mathbb{R}^d$, we have $\boldsymbol{A}_h = \boldsymbol{A}_h^\top \succ 0$. This implies $\boldsymbol{R} = \boldsymbol{R}^\top \succ 0$. Moreover, since $\boldsymbol{Q}$ is orthogonal,

$$\boldsymbol{R}^2 = \boldsymbol{R}\boldsymbol{R}^\top = \boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{Q}\boldsymbol{Q}^\top\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{\Sigma}_0^{1/2} = \boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{1/2},$$

so $\boldsymbol{R} = (\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{1/2})^{1/2}$ and $\boldsymbol{A}_h = \boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{R}\boldsymbol{\Sigma}_0^{-1/2} = \boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{1/2})^{1/2}\boldsymbol{\Sigma}_0^{-1/2}$. Thus $\boldsymbol{A}_h = \boldsymbol{A}^\star$, and therefore $T_h = T^\star$.

Transport value. By Theorem 4.3(d), $h = (\boldsymbol{0}, \boldsymbol{Q})$ solves (LGOP). Hence, we have

$$\begin{aligned}
\mathbb{J}_c(g_0, g_1) &= \int_{\mathbb{R}^d} \big\|(\boldsymbol{m}_0 - \boldsymbol{m}_1) + (\boldsymbol{\Sigma}_0^{1/2} - \boldsymbol{\Sigma}_1^{1/2}\boldsymbol{Q})\boldsymbol{z}\big\|^2 \mathrm{d}\rho(\boldsymbol{z}) \\
&= \|\boldsymbol{m}_0 - \boldsymbol{m}_1\|^2 + 2(\boldsymbol{m}_0 - \boldsymbol{m}_1)^\top(\boldsymbol{\Sigma}_0^{1/2} - \boldsymbol{\Sigma}_1^{1/2}\boldsymbol{Q}) \int_{\mathbb{R}^d} \boldsymbol{z}\,\mathrm{d}\rho(\boldsymbol{z}) \\
&\quad + \mathrm{Tr}\big((\boldsymbol{\Sigma}_0^{1/2} - \boldsymbol{\Sigma}_1^{1/2}\boldsymbol{Q})(\boldsymbol{\Sigma}_0^{1/2} - \boldsymbol{\Sigma}_1^{1/2}\boldsymbol{Q})^\top\big) \\
&= \|\boldsymbol{m}_0 - \boldsymbol{m}_1\|^2 + \mathrm{Tr}(\boldsymbol{\Sigma}_0) + \mathrm{Tr}(\boldsymbol{\Sigma}_1) - 2\,\mathrm{Tr}(\boldsymbol{R}),
\end{aligned}$$

where the last equality follows because $\int_{\mathbb{R}^d} \boldsymbol{z}\,\mathrm{d}\rho(\boldsymbol{z}) = 0$, and the trace expands using $\boldsymbol{Q}\boldsymbol{Q}^\top = \boldsymbol{I}_d$. Plugging in the value of $\boldsymbol{R}$ to the expression above results in the displayed equation in the theorem statement. Equality of $\mathbb{K}_c$, $\mathbb{M}_c$ and $\mathbb{J}_c$ follows by Theorem 4.3(e). □

When $\tilde{\varsigma}(s) = \exp(-s/2)$, the reference measure ρ is the standard Gaussian, Corollary 1 then recovers the classical result of [7, 23]; the extension to general elliptical families is due to [11]. The orbit perspective makes transparent that both results share the same algebraic source: the Cartan involution $\theta(\boldsymbol{A}) = \boldsymbol{A}^{-\top}$ on $\mathrm{GL}(d)$.

5.2. **The congruence mechanism on $\mathbb{S}^d_{\succ}$.** Let the state space be $\mathcal{X} = \mathbb{S}^d_{\succ}$, the cone of $d \times d$ symmetric positive definite matrices, and let $G = \mathrm{GL}(d)$ act on $\mathcal{X}$ by congruence: $\alpha(\boldsymbol{A})(\boldsymbol{X}) = \boldsymbol{A}\boldsymbol{X}\boldsymbol{A}^\top$, $\boldsymbol{A} \in \mathrm{GL}(d)$. The action α is induced by a linear representation on the Hilbert space $\mathcal{H} = \mathrm{Sym}(d)$ equipped with the Frobenius inner product $\langle \boldsymbol{X}, \boldsymbol{Y} \rangle_{\mathcal{H}} = \mathrm{Tr}(\boldsymbol{X}\boldsymbol{Y})$, with $b \equiv 0$ and $\pi = \alpha$. Thus Assumption 4 holds.

Cartan structure. The Cartan structure is carried by the image $L = \pi(\mathrm{GL}(d)) = \{\pi(\boldsymbol{A}) : \boldsymbol{A} \in \mathrm{GL}(d)\}$. The Cartan involution $\boldsymbol{A} \mapsto \boldsymbol{A}^{-\top}$ on $\mathrm{GL}(d)$ induces a Cartan involution on L via $\theta(\pi(\boldsymbol{A})) = \pi(\boldsymbol{A}^{-\top})$. The map is well-defined because the only ambiguity in representing an element of $L = \pi(\mathrm{GL}(d))$ is sign: if $\pi(\boldsymbol{A}) = \pi(\boldsymbol{B})$, then $\boldsymbol{B} = \pm\boldsymbol{A}$ by [12, Proposition 4 (ii)]. Hence $\pi(\boldsymbol{B}^{-\top}) = \pi(\boldsymbol{A}^{-\top})$. Then, the fixed-point subgroup of L is $L^\theta = \{\pi(\boldsymbol{Q}) : \boldsymbol{Q} \in \mathcal{O}(d)\}$. Indeed, if $\boldsymbol{Q} \in \mathcal{O}(d)$, then $\boldsymbol{Q}^{-\top} = \boldsymbol{Q}$, so $\theta(\pi(\boldsymbol{Q})) = \pi(\boldsymbol{Q}^{-\top}) = \pi(\boldsymbol{Q})$. Conversely, if $\pi(\boldsymbol{A}) \in L^\theta$, then $\pi(\boldsymbol{A}) = \theta(\pi(\boldsymbol{A})) = \pi(\boldsymbol{A}^{-\top})$. Evaluating both operators, $\pi(\boldsymbol{A})$ and $\pi(\boldsymbol{A}^{-\top})$, at $\boldsymbol{I}_d$ gives $\boldsymbol{A}\boldsymbol{A}^\top = \boldsymbol{A}^{-\top}\boldsymbol{A}^{-1} = (\boldsymbol{A}\boldsymbol{A}^\top)^{-1}$. Hence $(\boldsymbol{A}\boldsymbol{A}^\top)^2 = \boldsymbol{I}_d$. Since $\boldsymbol{A}\boldsymbol{A}^\top$ is symmetric positive definite, all its eigenvalues are positive and satisfy $\lambda^2 = 1$, hence $\lambda = 1$. Therefore $\boldsymbol{A}\boldsymbol{A}^\top = \boldsymbol{I}_d$, so $\boldsymbol{A} \in \mathcal{O}(d)$. The Lie algebra of L consists of operators

$$\mathcal{D}_{\boldsymbol{H}} : \mathrm{Sym}(d) \to \mathrm{Sym}(d), \qquad \mathcal{D}_{\boldsymbol{H}}(\boldsymbol{X}) = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{X}\boldsymbol{H}^\top, \qquad \boldsymbol{H} \in \mathfrak{gl}(d),$$

where $\mathfrak{gl}(d) = \mathbb{R}^{d\times d}$. Indeed, if $A(t) = \boldsymbol{I}_d + t\boldsymbol{H} + o(t)$, then $\pi(A(t))(\boldsymbol{X}) = A(t)\boldsymbol{X}A(t)^\top = \boldsymbol{X} + t(\boldsymbol{H}\boldsymbol{X} + \boldsymbol{X}\boldsymbol{H}^\top) + o(t)$. Thus $\frac{\mathrm{d}}{\mathrm{d}t}|_{t=0}\pi(A(t)) = \mathcal{D}_{\boldsymbol{H}}$ and $\mathrm{Lie}(L) = \{\mathcal{D}_{\boldsymbol{H}} : \boldsymbol{H} \in \mathfrak{gl}(d)\}$. Since the Cartan involution on L is given by $\theta(\pi(\boldsymbol{A})) = \pi(\boldsymbol{A}^{-\top})$, its differential satisfies $\mathrm{d}\theta_e(\mathcal{D}_{\boldsymbol{H}}) = \mathcal{D}_{-\boldsymbol{H}^\top}$. Now write

$$\boldsymbol{H} = \boldsymbol{S} + \boldsymbol{K}, \qquad \boldsymbol{S} = \frac{\boldsymbol{H} + \boldsymbol{H}^\top}{2} \in \mathrm{Sym}(d), \qquad \boldsymbol{K} = \frac{\boldsymbol{H} - \boldsymbol{H}^\top}{2}, \quad \boldsymbol{K}^\top = -\boldsymbol{K}.$$

Then $\mathcal{D}_{\boldsymbol{H}} = \mathcal{D}_{\boldsymbol{S}} + \mathcal{D}_{\boldsymbol{K}}$, and $\mathrm{d}\theta_e(\mathcal{D}_{\boldsymbol{S}}) = \mathcal{D}_{-\boldsymbol{S}} = -\mathcal{D}_{\boldsymbol{S}}$, $\mathrm{d}\theta_e(\mathcal{D}_{\boldsymbol{K}}) = \mathcal{D}_{\boldsymbol{K}}$. Hence the (-1)-eigenspace is $\mathfrak{p} = \{\mathcal{D}_{\boldsymbol{S}} : \boldsymbol{S} \in \mathrm{Sym}(d)\}$.

The polar decomposition gives a diffeomorphism $\mathcal{O}(d) \times \mathrm{Sym}(d) \to \mathrm{GL}(d)$, $(\boldsymbol{Q}, \boldsymbol{S}) \mapsto \boldsymbol{Q}\exp(\boldsymbol{S})$. For $\boldsymbol{S} \in \mathrm{Sym}(d)$, compatibility of Lie group homomorphisms with exponential maps gives $\exp(\mathcal{D}_{\boldsymbol{S}}) = \pi(\exp(\boldsymbol{S}))$. For the congruence representation, the only ambiguity is the sign: $\pi(\boldsymbol{A}) = \pi(\boldsymbol{B})$ if and only if $\boldsymbol{B} = \pm\boldsymbol{A}$. If $\boldsymbol{A} = \boldsymbol{Q}\exp(\boldsymbol{S})$ is the polar decomposition of $\boldsymbol{A}$, then the polar decomposition of $-\boldsymbol{A}$ is $-\boldsymbol{A} = (-\boldsymbol{Q})\exp(\boldsymbol{S})$. Thus the quantities $\pi(\boldsymbol{Q})$ and $\mathcal{D}_{\boldsymbol{S}}$ do not depend on the choice of representative $\boldsymbol{A}$ of $\pi(\boldsymbol{A})$. Therefore the polar decomposition induces a well-defined and smooth bijection $L^\theta \times \mathfrak{p} \to L$, $(\pi(\boldsymbol{Q}), \mathcal{D}_{\boldsymbol{S}}) \mapsto \pi(\boldsymbol{Q})\exp(\mathcal{D}_{\boldsymbol{S}})$. Its inverse is also smooth because it is induced by the smooth inverse of the polar decomposition on $\mathrm{GL}(d)$. Hence this map is a diffeomorphism, and (L, θ) admits a global Cartan decomposition in the sense of Definition 7.

Observe that every element of $\mathfrak{p}$ is self-adjoint on $\mathcal{H}$. Indeed, for $\boldsymbol{S}, \boldsymbol{X}, \boldsymbol{Y} \in \mathrm{Sym}(d)$, we have: $\langle \mathcal{D}_{\boldsymbol{S}}\boldsymbol{X}, \boldsymbol{Y}\rangle_{\mathcal{H}} = \mathrm{Tr}((\boldsymbol{S}\boldsymbol{X} + \boldsymbol{X}\boldsymbol{S})\boldsymbol{Y}) = \mathrm{Tr}(\boldsymbol{X}(\boldsymbol{Y}\boldsymbol{S} + \boldsymbol{S}\boldsymbol{Y})) = \langle \boldsymbol{X}, \mathcal{D}_{\boldsymbol{S}}\boldsymbol{Y}\rangle_{\mathcal{H}}$. Hence, hypothesis (i) of Theorem 4.3 holds.

Let $\mathcal{L}_{\mathrm{Sym}(d)}$ denote the Euclidean volume measure on the finite-dimensional vector space $\mathrm{Sym}(d)$ induced by the Frobenius inner product; equivalently, it is the $d(d+1)/2$-dimensional Hausdorff measure induced by the Frobenius norm. All densities on $\mathbb{S}^d_\succ$ in this subsection are taken with respect to the restriction $\mathcal{L}_{\mathrm{Sym}(d)}|_{\mathbb{S}^d_\succ}$.

Reference distributions. It remains to identify reference laws ρ on $\mathbb{S}^d_\succ$ for which the compact invariance condition (ii) $L^\theta \subseteq \pi(\mathrm{Stab}_{\mathrm{GL}(d)}(\rho))$ holds. Since $L^\theta = \pi(\mathcal{O}(d))$, this condition is guaranteed by $\mathcal{O}(d) \subseteq \mathrm{Stab}_{\mathrm{GL}(d)}(\rho)$. Because the congruence action of $\boldsymbol{Q} \in \mathcal{O}(d)$ preserves the eigenvalues of $\boldsymbol{X}$, any distribution whose density depends on $\boldsymbol{X}$ only through its eigenvalues is $\mathcal{O}(d)$-invariant. We call such a distribution *spectrally invariant*. Before stating the main result, we establish the second-moment structure that governs the optimal transport value for every spectrally invariant reference.

Lemma 7 (Second-moment tensor of a spectral law)**.** *Let $d \geq 2$ and let $\rho \in \mathcal{P}(\mathbb{S}^d_\succ)$ be spectrally invariant, i.e., $\alpha(\boldsymbol{Q})_\#\rho = \rho$ for every $\boldsymbol{Q} \in \mathcal{O}(d)$. Assume moreover that $\int_{\mathbb{S}^d_\succ} \|\boldsymbol{X}\|_{\mathrm{F}}^2 \mathrm{d}\rho(\boldsymbol{X}) < \infty$.*

Then there exist constants $a_\rho, b_\rho \in \mathbb{R}$ depending on ρ such that

$$\mathbb{E}_{\boldsymbol{X}\sim\rho}[X_{ij}X_{kl}] = a_\rho \delta_{ij}\delta_{kl} + b_\rho\big(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\big) \tag{2}$$

for all $i, j, k, l \in \{1, \dots, d\}$. Consequently, for every $\boldsymbol{U}, \boldsymbol{V} \in \mathbb{R}^{d\times d}$,

$$\mathbb{E}_{\boldsymbol{X}\sim\rho}[\mathrm{Tr}(\boldsymbol{UXVX})] = a_\rho \,\mathrm{Tr}(\boldsymbol{UV}) + b_\rho\big(\mathrm{Tr}(\boldsymbol{UV}^\top) + \mathrm{Tr}(\boldsymbol{U})\,\mathrm{Tr}(\boldsymbol{V})\big). \tag{3}$$

The constants a_ρ and b_ρ are determined by the moments $\mathbb{E}_{\boldsymbol{X}\sim\rho}[(\mathrm{Tr}(\boldsymbol{X}))^2]$ and $\mathbb{E}_{\boldsymbol{X}\sim\rho}[\mathrm{Tr}(\boldsymbol{X}^2)]$ via

$$b_\rho = \frac{\mathbb{E}_{\boldsymbol{X}\sim\rho}[\mathrm{Tr}(\boldsymbol{X}^2)] - \frac{1}{d}\mathbb{E}_{\boldsymbol{X}\sim\rho}[(\mathrm{Tr}(\boldsymbol{X}))^2]}{d^2+d-2}, \qquad a_\rho = \frac{\mathbb{E}_{\boldsymbol{X}\sim\rho}[(\mathrm{Tr}(\boldsymbol{X}))^2] - 2b_\rho d}{d^2}. \tag{4}$$

For $d = 1$, the pair (a_ρ, b_ρ) is not uniquely identified; only the combination $a_\rho + 2b_\rho = \mathbb{E}_{X\sim\rho}[X^2]$ is determined and $\mathbb{E}_{X\sim\rho}[\mathrm{Tr}(UXVX)] = (a_\rho + 2b_\rho)UV$.

Proof. Let $\boldsymbol{X} \sim \rho$. Since ρ is invariant under orthogonal congruence, we have $\boldsymbol{QXQ}^\top \sim \boldsymbol{X}$ for every $\boldsymbol{Q} \in \mathcal{O}(d)$. Hence, for every $i, j, k, l \in \{1, \dots, d\}$,

$$T_{ijkl} = \mathbb{E}_{\boldsymbol{X}\sim\rho}[X_{ij}X_{kl}] = \mathbb{E}_{\boldsymbol{X}\sim\rho}[(\boldsymbol{QXQ}^\top)_{ij}(\boldsymbol{QXQ}^\top)_{kl}].$$

Using $(\boldsymbol{QXQ}^\top)_{ij} = \sum_{i',j'=1}^d Q_{ii'}X_{i'j'}Q_{jj'}$, we obtain $T_{ijkl} = \sum_{i',j',k',l'=1}^d Q_{ii'}Q_{jj'}Q_{kk'}Q_{ll'}\,T_{i'j'k'l'}$.

In other words, the tensor $T_{ijkl} = \mathbb{E}_{\boldsymbol{X}\sim\rho}[X_{ij}X_{kl}]$ is invariant under the simultaneous congruence $T \mapsto \boldsymbol{Q}^{\otimes 4}T$ for every $\boldsymbol{Q} \in \mathcal{O}(d)$. By the First Fundamental Theorem for the orthogonal group, every $O(d)$-invariant rank-four tensor is a linear combination of the three pairings $\delta_{ij}\delta_{kl},\ \delta_{ik}\delta_{jl},\ \delta_{il}\delta_{jk}$. Hence, for some scalars A, B, C, $T_{ijkl} = A\delta_{ij}\delta_{kl} + B\delta_{ik}\delta_{jl} + C\delta_{il}\delta_{jk}$. Since $\boldsymbol{X}$ is symmetric, we have $T_{ijkl} = T_{jikl} = T_{ijlk}$, which forces $B = C$. Writing $a_\rho = A$ and $b_\rho = B$ yields $T_{ijkl} = a_\rho\delta_{ij}\delta_{kl} + b_\rho(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk})$. For the identity (3), note that $\mathrm{Tr}(\boldsymbol{UXVX}) = \sum_{i,j,k,l} U_{ij}V_{kl}X_{jk}X_{li}$. Substituting (2) for $\mathbb{E}_{\boldsymbol{X}\sim\rho}[X_{jk}X_{li}]$: the term $a_\rho\delta_{jk}\delta_{li}$ contributes $a_\rho\sum_{i,j} U_{ij}V_{ji} = a_\rho\,\mathrm{Tr}(\boldsymbol{UV})$; the term $b_\rho\delta_{jl}\delta_{ki}$ contributes $b_\rho\sum_{j,k} U_{kj}V_{kj} = b_\rho\,\mathrm{Tr}(\boldsymbol{UV}^\top)$; and the term $b_\rho\delta_{ji}\delta_{kl}$ contributes $b_\rho = \sum_{i=1}^d U_{ii}\sum_{k=1}^d V_{kk} = b_\rho\,\mathrm{Tr}(\boldsymbol{U})\,\mathrm{Tr}(\boldsymbol{V})$.

Finally, evaluating (2) with $i = j, k = l$ gives $\mathbb{E}_{\boldsymbol{X}\sim\rho}[(\mathrm{Tr}(\boldsymbol{X}))^2] = a_\rho d^2 + 2b_\rho d$, and with $i = k, j = l$ gives $\mathbb{E}_{\boldsymbol{X}\sim\rho}[\mathrm{Tr}(\boldsymbol{X}^2)] = a_\rho d + b_\rho d(d+1)$. Solving this 2×2 system yields (4). □

Lemma 8. *Let $\boldsymbol{A} \in \mathrm{GL}(d)$ and define $\pi(\boldsymbol{A}) : \mathrm{Sym}(d) \to \mathrm{Sym}(d)$, $\pi(\boldsymbol{A})\boldsymbol{X} = \boldsymbol{AXA}^\top$. If $\pi(\boldsymbol{A})$ is self-adjoint and positive definite on* Sym(d) *with respect to the Frobenius inner product, then $\boldsymbol{A} \in \mathrm{Sym}(d)$ and $\boldsymbol{A} \succ 0$ or $\boldsymbol{A} \prec 0$.*

Proof. For $\boldsymbol{X}, \boldsymbol{Y} \in \mathrm{Sym}(d)$, $\langle\pi(\boldsymbol{A})\boldsymbol{X}, \boldsymbol{Y}\rangle = \mathrm{Tr}(\boldsymbol{AXA}^\top\boldsymbol{Y}) = \mathrm{Tr}(\boldsymbol{XA}^\top\boldsymbol{YA}) = \langle\boldsymbol{X}, \pi(\boldsymbol{A}^\top)\boldsymbol{Y}\rangle$. Thus $\pi(\boldsymbol{A})^* = \pi(\boldsymbol{A}^\top)$. Since $\pi(\boldsymbol{A})$ is self-adjoint, $\pi(\boldsymbol{A}) = \pi(\boldsymbol{A}^\top)$. By [12, Proposition 4 (ii)] if $\boldsymbol{AXA}^\top = \boldsymbol{BXB}^\top\ \forall\boldsymbol{X} \in \mathrm{Sym}(d)$, then $\boldsymbol{B} = \pm\boldsymbol{A}$. Applying this with $\boldsymbol{B} = \boldsymbol{A}^\top$, we obtain $\boldsymbol{A}^\top = \pm\boldsymbol{A}$. Note that the skew-symmetric case $\boldsymbol{A}^\top = -\boldsymbol{A}$ is impossible. Indeed, for any nonzero $\boldsymbol{x} \in \mathbb{R}^d$, set $\boldsymbol{X} = \boldsymbol{xx}^\top$. Then $\boldsymbol{X} \in \mathrm{Sym}(d)$ and $\langle\pi(\boldsymbol{A})\boldsymbol{X}, \boldsymbol{X}\rangle = \mathrm{Tr}(\boldsymbol{AXA}^\top\boldsymbol{X}) = (\boldsymbol{x}^\top\boldsymbol{Ax})^2 = 0$, contradicting positive definiteness of $\pi(\boldsymbol{A})$. Thus $\boldsymbol{A} = \boldsymbol{A}^\top$.

Now diagonalize $\boldsymbol{A} = \boldsymbol{U}\mathrm{diag}(\lambda_1, \dots, \lambda_d)\boldsymbol{U}^\top$. If $d = 1$, then $\lambda_1 \neq 0$, so $\boldsymbol{A}$ is either positive or negative definite. If $d \geq 2$, then for $i \neq j$ set $\boldsymbol{X}_{ij} = \boldsymbol{U}(\boldsymbol{e}_i\boldsymbol{e}_j^\top + \boldsymbol{e}_j\boldsymbol{e}_i^\top)\boldsymbol{U}^\top$. Positive definiteness gives $0 < \langle\pi(\boldsymbol{A})\boldsymbol{X}_{ij}, \boldsymbol{X}_{ij}\rangle = 2\lambda_i\lambda_j$. Therefore $\lambda_i\lambda_j > 0$ for all $i \neq j$, so all eigenvalues of $\boldsymbol{A}$ have the same sign. Hence either $\boldsymbol{A} \succ 0$ or $\boldsymbol{A} \prec 0$. □

Theorem 5.1 (Congruence template on $\mathbb{S}^d_\succ$)**.** *Let $\rho \in \mathcal{P}(\mathbb{S}^d_\succ)$ be absolutely continuous, spectrally invariant, and have finite second moment with respect to the Frobenius norm. Let a_ρ and b_ρ be as in (4). For $\boldsymbol{\Sigma}_i \in \mathbb{S}^d_\succ$, $i = 0, 1$, define $\mu_i = \alpha(\boldsymbol{\Sigma}_i^{1/2})_\#\rho$ and set $\boldsymbol{C} = (\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{1/2})^{1/2}$. Then for the quadratic cost $c(\boldsymbol{X}, \boldsymbol{Y}) = \|\boldsymbol{X} - \boldsymbol{Y}\|_{\mathrm{F}}^2$:*

(a) *the unique optimal Monge map from* μ_0 *to* μ_1 *is*

$$T^\star(\boldsymbol{X}) = \boldsymbol{A}^\star \boldsymbol{X} \boldsymbol{A}^\star, \qquad \boldsymbol{A}^\star = \boldsymbol{\Sigma}_0^{-1/2}\big(\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{1/2}\big)^{1/2}\boldsymbol{\Sigma}_0^{-1/2}, \tag{5}$$

where $\boldsymbol{A}^\star$ *is symmetric positive definite;*

(b) $(\mathrm{id}, T^\star)_\# \mu_0$ *is the unique optimal Kantorovich plan,*

(c) *the optimal transport cost is*

$$\mathbb{K}_c(\mu_0,\mu_1) = (a_\rho + b_\rho)\|\boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}_1\|_{\mathrm{F}}^2 + b_\rho\big((\mathrm{Tr}(\boldsymbol{\Sigma}_0))^2 + (\mathrm{Tr}(\boldsymbol{\Sigma}_1))^2 - 2(\mathrm{Tr}\,\boldsymbol{C})^2\big), \tag{6}$$

where a_ρ *and* b_ρ *are as defined in* (4). *For* $d = 1$, a_ρ *and* b_ρ *are not separately identified, but the two quadratic terms in* (6) *coincide. Thus the value depends only on the identifiable combination* $a_\rho + 2b_\rho = \mathbb{E}_{X\sim\rho}[X^2]$.

(d) $\mathbb{M}_c(\mu_0,\mu_1) = \mathbb{K}_c(\mu_0,\mu_1) = \mathbb{J}_c(\boldsymbol{\Sigma}_0^{1/2}, \boldsymbol{\Sigma}_1^{1/2})$.

Proof. Set $g_i = \boldsymbol{\Sigma}_i^{1/2}$, $i = 0, 1$, so that $\mu_i = \alpha(g_i)_\#\rho$. Both hypotheses of Theorem 4.3 have been verified above: $\mathcal{D}_{\boldsymbol{S}}$ is self-adjoint on $\mathrm{Sym}(d)$ for every $\boldsymbol{S} \in \mathrm{Sym}(d)$, and $\mathcal{O}(d) \subseteq \mathrm{Stab}_{\mathrm{GL}(d)}(\rho)$ by the spectral invariance of ρ.

By Theorem 4.3, there exists $h \in \mathrm{Stab}_{\mathrm{GL}(d)}(\rho)$, with $\pi(h) \in L^\theta$, such that $\pi(g_1 h g_0^{-1})$ is self-adjoint and positive definite on $\mathcal{H} = \mathrm{Sym}(d)$. Since $L^\theta = \pi(\mathcal{O}(d))$, we may choose $\boldsymbol{Q}^\star \in \mathcal{O}(d)$ such that $\pi(h) = \pi(\boldsymbol{Q}^\star)$. Then, we have $\pi(g_1 h g_0^{-1}) = \pi(\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{Q}^\star\boldsymbol{\Sigma}_0^{-1/2})$. Set $\boldsymbol{A}_h = \boldsymbol{\Sigma}_1^{1/2}\boldsymbol{Q}^\star\boldsymbol{\Sigma}_0^{-1/2}$. Then $\pi(\boldsymbol{A}_h)$ is self-adjoint and positive definite on $\mathrm{Sym}(d)$, and the induced congruence map is the unique optimal Monge map from μ_0 to μ_1 by Theorem 4.3(b).

Identification of $\boldsymbol{A}_h$*.* By Lemma 8, the operator $\pi(\boldsymbol{A}_h)$ admits a symmetric positive definite representative. Since $\pi(\boldsymbol{A}_h) = \pi(-\boldsymbol{A}_h)$, replacing $\boldsymbol{A}_h$ by $-\boldsymbol{A}_h$ if necessary, equivalently replacing $\boldsymbol{Q}^\star$ by $-\boldsymbol{Q}^\star$, does not change the transport map. We may therefore assume $\boldsymbol{A}_h = \boldsymbol{A}_h^\top \succ 0$. Set $\boldsymbol{R} = \boldsymbol{\Sigma}_0^{1/2}\boldsymbol{A}_h\boldsymbol{\Sigma}_0^{1/2} = \boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{Q}^\star$. Then $\boldsymbol{R} = \boldsymbol{R}^\top \succ 0$. Since $\boldsymbol{Q}^\star(\boldsymbol{Q}^\star)^\top = \boldsymbol{I}_d$, we have $\boldsymbol{R}^2 = \boldsymbol{R}\boldsymbol{R}^\top = \boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{Q}^\star(\boldsymbol{Q}^\star)^\top\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{\Sigma}_0^{1/2} = \boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{1/2}$. Therefore $\boldsymbol{R}$ is the unique positive definite square root of $\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{1/2}$. Hence $\boldsymbol{R} = \boldsymbol{C} = \big(\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{1/2}\big)^{1/2}$. Consequently, $\boldsymbol{A}_h = \boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{C}\boldsymbol{\Sigma}_0^{-1/2} = \boldsymbol{A}^\star$. Since the possible sign change does not affect the congruence operator, we conclude in all cases that $T_h(\boldsymbol{X}) = \boldsymbol{A}^\star\boldsymbol{X}\boldsymbol{A}^\star = T^\star(\boldsymbol{X})$.

Transport value. By Theorem 4.3, the optimal value equals

$$\mathbb{K}_c(\mu_0,\mu_1) = \mathbb{E}_{\boldsymbol{X}\sim\rho}\big[\|\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{X}\boldsymbol{\Sigma}_0^{1/2} - \boldsymbol{\Sigma}_1^{1/2}\boldsymbol{Q}^\star\boldsymbol{X}\boldsymbol{Q}^{\star\top}\boldsymbol{\Sigma}_1^{1/2}\|_{\mathrm{F}}^2\big]. \tag{7}$$

Expanding the squared Frobenius norm yields:

$$\begin{aligned}\mathbb{K}_c(\mu_0,\mu_1) = \mathbb{E}_{\boldsymbol{X}\sim\rho}&\left[\mathrm{Tr}\left((\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{X}\boldsymbol{\Sigma}_0^{1/2})^2\right)\right] + \mathbb{E}_{\boldsymbol{X}\sim\rho}\left[\mathrm{Tr}\left((\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{Q}^\star\boldsymbol{X}\boldsymbol{Q}^{\star\top}\boldsymbol{\Sigma}_1^{1/2})^2\right)\right] \\ &- 2\mathbb{E}_{\boldsymbol{X}\sim\rho}\left[\mathrm{Tr}\left(\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{X}\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{Q}^\star\boldsymbol{X}\boldsymbol{Q}^{\star\top}\boldsymbol{\Sigma}_1^{1/2}\right)\right].\end{aligned}$$

Since $\boldsymbol{Q}^\star\boldsymbol{X}\boldsymbol{Q}^{\star\top} \sim \boldsymbol{X}$, the second term equals $\mathbb{E}_{\boldsymbol{X}\sim\rho}[\mathrm{Tr}(\boldsymbol{\Sigma}_1\boldsymbol{X}\boldsymbol{\Sigma}_1\boldsymbol{X})]$. By cyclicity of the trace, the cross term can be written as $\mathbb{E}_\rho[\mathrm{Tr}(\boldsymbol{M}\boldsymbol{Q}^\star\boldsymbol{X}\boldsymbol{Q}^{\star\top}\boldsymbol{M}^\top\boldsymbol{X})]$, where $\boldsymbol{M} = \boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_1^{1/2}$. Hence,

$$\begin{aligned}\mathbb{K}_c(\mu_0,\mu_1) = \mathbb{E}_{\boldsymbol{X}\sim\rho}&[\mathrm{Tr}(\boldsymbol{\Sigma}_0\boldsymbol{X}\boldsymbol{\Sigma}_0\boldsymbol{X})] + \mathbb{E}_{\boldsymbol{X}\sim\rho}[\mathrm{Tr}(\boldsymbol{\Sigma}_1\boldsymbol{X}\boldsymbol{\Sigma}_1\boldsymbol{X})] \\ &- 2\mathbb{E}_{\boldsymbol{X}\sim\rho}[\mathrm{Tr}(\boldsymbol{M}\boldsymbol{Q}^\star\boldsymbol{X}\boldsymbol{Q}^{\star\top}\boldsymbol{M}^\top\boldsymbol{X})].\end{aligned} \tag{8}$$

Each of the three terms above is of the form $\mathbb{E}_{\boldsymbol{X}\sim\rho}[\mathrm{Tr}(\boldsymbol{U}\boldsymbol{X}\boldsymbol{V}\boldsymbol{X})]$ for some $\boldsymbol{U}, \boldsymbol{V} \in \mathbb{R}^{d\times d}$, applying Lemma 7 to each term yields:

- $\boldsymbol{U} = \boldsymbol{V} = \boldsymbol{\Sigma}_i$ (symmetric): $\mathbb{E}_{\boldsymbol{X}\sim\rho}[\mathrm{Tr}(\boldsymbol{\Sigma}_i\boldsymbol{X}\boldsymbol{\Sigma}_i\boldsymbol{X})] = (a_\rho + b_\rho)\,\mathrm{Tr}(\boldsymbol{\Sigma}_i^2) + b_\rho(\mathrm{Tr}\,\boldsymbol{\Sigma}_i)^2$,

- $\boldsymbol{U} = \boldsymbol{M}\boldsymbol{Q}^\star$, $\boldsymbol{V} = \boldsymbol{Q}^{\star\top}\boldsymbol{M}^\top$: Then $\mathrm{Tr}(\boldsymbol{U}\boldsymbol{V}) = \mathrm{Tr}(\boldsymbol{M}\boldsymbol{M}^\top) = \mathrm{Tr}(\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}_1)$. Moreover, since $\boldsymbol{M}\boldsymbol{Q}^\star = \boldsymbol{C}$, $\mathrm{Tr}(\boldsymbol{U}\boldsymbol{V}^\top) = \mathrm{Tr}((\boldsymbol{M}\boldsymbol{Q}^\star)^2) = \mathrm{Tr}(\boldsymbol{C}^2) = \mathrm{Tr}(\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}_1)$, and $\mathrm{Tr}(\boldsymbol{U})\,\mathrm{Tr}(\boldsymbol{V}) = (\mathrm{Tr}\,\boldsymbol{C})^2$. Therefore $\mathbb{E}_{\boldsymbol{X}\sim\rho}[\mathrm{Tr}(\boldsymbol{M}\boldsymbol{Q}^\star\boldsymbol{X}\boldsymbol{Q}^{\star\top}\boldsymbol{M}^\top\boldsymbol{X})] = (a_\rho + b_\rho)\,\mathrm{Tr}(\boldsymbol{\Sigma}_0\ \boldsymbol{\Sigma}_1) + b_\rho(\mathrm{Tr}\,\boldsymbol{C})^2$.

Substituting into (8): $\mathbb{K}_c(\mu_0,\mu_1) = (a_\rho + b_\rho)(\mathrm{Tr}(\boldsymbol{\Sigma}_0^2) + \mathrm{Tr}(\boldsymbol{\Sigma}_1^2) - 2\,\mathrm{Tr}(\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}_1)) + b_\rho((\mathrm{Tr}(\boldsymbol{\Sigma}_0))^2 + (\mathrm{Tr}(\boldsymbol{\Sigma}_1))^2 - 2(\mathrm{Tr}(\boldsymbol{C}))^2)$. Since $\mathrm{Tr}(\boldsymbol{\Sigma}_0^2) + \mathrm{Tr}(\boldsymbol{\Sigma}_1^2) - 2\,\mathrm{Tr}(\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}_1) = \|\boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}_1\|_{\mathrm{F}}^2$, the closed-form solution coincides with (6). The equality $\mathbb{M}_c = \mathbb{K}_c$ follows from Theorem 4.3(e). □

Note that the optimal map (5) is independent of the choice of spectrally invariant reference ρ; only the transport value (6) depends on ρ, through the pair (a_ρ, b_ρ).

We next present three families of spectrally invariant reference measures, each yielding a closed-form optimal transport value through (6). Table 1 summarizes these families and the corresponding coefficients. We derive the Wishart family as a representative case and relegate the analogous derivations for the other families to Section C.

TABLE 1. Spectrally invariant families on $\mathbb{S}^d_\succ$ under the congruence action, their densities, reference laws, and second-moment constants entering the transport value (6).

Family	**Density** ($\propto$)	ρ	a_ρ	b_ρ
Wishart ($\boldsymbol{\Sigma} \in \mathbb{S}^d_\succ$, $p > d-1$)				
$\mathcal{W}_d(\boldsymbol{\Sigma}, p)$	$\det(\boldsymbol{X})^{(p-d-1)/2}\exp(-\frac{1}{2}\,\mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{X}))$	$\mathcal{W}_d(\boldsymbol{I}_d, p)$	p^2	p
Inverse-Wishart ($\boldsymbol{\Psi} \in \mathbb{S}^d_\succ$, $p > d{+}3$, $m = p{-}d$)				
$\mathcal{IW}_d(\boldsymbol{\Psi}, p)$	$\det(\boldsymbol{X})^{-(p+d+1)/2}\exp(-\frac{1}{2}\,\mathrm{Tr}(\boldsymbol{\Psi}\boldsymbol{X}^{-1}))$	$\mathcal{IW}_d(\boldsymbol{I}_d, p)$	$\frac{m-2}{m(m-1)(m-3)}$	$\frac{1}{m(m-1)(m-3)}$
Matrix beta type II ($\boldsymbol{\Sigma} \in \mathbb{S}^d_\succ$, $q_1 > (d{-}1)/2$, $2q_2 > d{+}3$, $q = 2q_2{-}d$)				
$\mathcal{MB}^{\mathrm{II}}_d(q_1, q_2, \boldsymbol{\Sigma})$	$\det(\boldsymbol{X})^{q_1-(d+1)/2}\det(\boldsymbol{\Sigma}{+}\boldsymbol{X})^{-(q_1+q_2)}$	$\mathcal{MB}^{\mathrm{II}}_d(q_1, q_2, \boldsymbol{I}_d)$	$\frac{4q_1(q_1(q-2)+1)}{q(q-1)(q-3)}$	$\frac{2q_1(2q_1+q-1)}{q(q-1)(q-3)}$

Wishart distributions. The Wishart distribution $\mathcal{W}_d(\boldsymbol{\Sigma}, p)$ with scale matrix $\boldsymbol{\Sigma} \in \mathbb{S}^d_\succ$ and degrees of freedom $p > d-1$ is a probability law on $\mathbb{S}^d_\succ$ with density

$$\boldsymbol{X} \mapsto \frac{\det(\boldsymbol{X})^{\frac{(p-d-1)}{2}}\exp\left(-\frac{1}{2}\,\mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{X})\right)}{2^{\frac{pd}{2}}\det(\boldsymbol{\Sigma})^{\frac{p}{2}}\Gamma_d(\frac{p}{2})} \tag{9}$$

with respect to $\mathcal{L}_{\mathrm{Sym}(d)}|_{\mathbb{S}^d_\succ}$. When $d = 1$, this reduces to a scaled χ^2 distribution with p degrees of freedom and scale parameter $\Sigma > 0$.

Lemma 9 (Wishart orbit, stabilizer, and second moments)**.** *Let* $\rho = \mathcal{W}_d(\boldsymbol{I}_d, p)$ *with* $p > d-1$*. Then:*

(i) $\alpha(\boldsymbol{A})_\#\rho = \mathcal{W}_d(\boldsymbol{A}\boldsymbol{A}^\top, p)$ *for every* $\boldsymbol{A} \in \mathrm{GL}(d)$*,*
(ii) $\mathrm{Stab}_{\mathrm{GL}(d)}(\rho) = \mathcal{O}(d)$*,*
(iii) *for* $d \geq 2$*, the constants in Lemma 7 are* $a_\rho = p^2$*,* $b_\rho = p$*. For* $d = 1$*, the identifiable combination is* $a_\rho + 2b_\rho = p(p+2)$*.*

Proof. (i) This is the standard transformation property of the Wishart distribution; see, for example, [9, Proposition 8.1] or [22, Theorem 3.2.5]. (ii) We next identify the stabilizer. If $\boldsymbol{Q} \in \mathcal{O}(d)$, then by (i), we have $\alpha(\boldsymbol{Q})_{\#}\rho = \mathcal{W}_d(\boldsymbol{I}_d, p) = \rho$. Thus $\mathcal{O}(d) \subseteq \mathrm{Stab}_{\mathrm{GL}(d)}(\rho)$. Conversely, suppose $\boldsymbol{A} \in \mathrm{Stab}_{\mathrm{GL}(d)}(\rho)$. Then $\mathcal{W}_d(\boldsymbol{A}\boldsymbol{A}^\top, p) = \alpha(\boldsymbol{A})_{\#}\rho = \rho = \mathcal{W}_d(\boldsymbol{I}_d, p)$. Then, the equality of the two densities implies $\det(\boldsymbol{A}\boldsymbol{A}^\top)^{-p/2}\exp(-\frac{1}{2}\operatorname{Tr}(((\boldsymbol{A}\boldsymbol{A}^\top)^{-1} - \boldsymbol{I}_d)\boldsymbol{Y})) = 1$ for almost every $\boldsymbol{Y} \in \mathbb{S}^d_{\succ}$. Since both sides are continuous in $\boldsymbol{Y}$, the identity holds for all $\boldsymbol{Y} \in \mathbb{S}^d_{\succ}$. The exponent is a linear function of $\boldsymbol{Y}$. Since the displayed identity holds for every $\boldsymbol{Y} \in \mathbb{S}^d_{\succ}$, this linear function is constant on the open cone $\mathbb{S}^d_{\succ}$. A linear function that is constant on a nonempty open set must vanish identically; so $(\boldsymbol{A}\boldsymbol{A}^\top)^{-1} - \boldsymbol{I}_d = 0$. Thus $\boldsymbol{A}\boldsymbol{A}^\top = \boldsymbol{I}_d$, and hence $\boldsymbol{A} \in \mathcal{O}(d)$. This proves $\mathrm{Stab}_{\mathrm{GL}(d)}(\rho) = \mathcal{O}(d)$.

(iii) It remains to compute the constants a_ρ and b_ρ. We compute them from the two scalar moments $M_1 = \mathbb{E}_{\boldsymbol{X}\sim\rho}[(\operatorname{Tr}(\boldsymbol{X}))^2]$ and $M_2 = \mathbb{E}_{\boldsymbol{X}\sim\rho}[\operatorname{Tr}(\boldsymbol{X}^2)]$.

Let $Y_{\boldsymbol{H}} = \operatorname{Tr}(\boldsymbol{H}\boldsymbol{X})$, $\boldsymbol{H} \in \mathrm{Sym}(d)$, where $\boldsymbol{X} \sim \rho = \mathcal{W}_d(\boldsymbol{I}_d, p)$. For t in a neighborhood of 0, the Laplace transform of $Y_{\boldsymbol{H}}$ is $L_{\boldsymbol{H}}(t) = \mathbb{E}_{\boldsymbol{X}\sim\rho}[\exp(-tY_{\boldsymbol{H}})] = \mathbb{E}_{\boldsymbol{X}\sim\rho}[\exp(-t\operatorname{Tr}(\boldsymbol{H}\boldsymbol{X}))] = \det(\boldsymbol{I}_d + 2t\boldsymbol{H})^{-p/2}$, where the last equality follows from the characteristic function formula for the Wishart distribution in [22, Theorem 3.2.3] by evaluating it at the imaginary argument $it\boldsymbol{H}$. Since $L_{\boldsymbol{H}}(t) = \mathbb{E}_{\boldsymbol{X}\sim\rho}[\exp(-tY_{\boldsymbol{H}})]$, differentiation at $t = 0$ gives $L'_{\boldsymbol{H}}(0) = -\mathbb{E}_{\boldsymbol{X}\sim\rho}[Y_{\boldsymbol{H}}]$ and $L''_{\boldsymbol{H}}(0) = \mathbb{E}_{\boldsymbol{X}\sim\rho}[Y^2_{\boldsymbol{H}}]$. Equivalently, $L'_{\boldsymbol{H}}(0) = -\mathbb{E}_{\boldsymbol{X}\sim\rho}[\operatorname{Tr}(\boldsymbol{H}\boldsymbol{X})]$ and $L''_{\boldsymbol{H}}(0) = \mathbb{E}_{\boldsymbol{X}\sim\rho}[(\operatorname{Tr}(\boldsymbol{H}\boldsymbol{X}))^2]$.

To compute these quantities, we differentiate $\log(L_{\boldsymbol{H}}(t)) = -\frac{p}{2}\operatorname{Tr}(\log(\boldsymbol{I}_d + 2t\boldsymbol{H}))$. Hence $(\log(L_{\boldsymbol{H}}))'(t) = -p\operatorname{Tr}\big((\boldsymbol{I}_d + 2t\boldsymbol{H})^{-1}\boldsymbol{H}\big)$, and

$$(\log(L_{\boldsymbol{H}}))''(t) = 2p\operatorname{Tr}\big((\boldsymbol{I}_d + 2t\boldsymbol{H})^{-1}\boldsymbol{H}(\boldsymbol{I}_d + 2t\boldsymbol{H})^{-1}\boldsymbol{H}\big).$$

Evaluating the first and second derivatives of $\log(L_{\boldsymbol{H}})$ at $t = 0$ yields $(\log(L_{\boldsymbol{H}}))'(0) = -p\operatorname{Tr}(\boldsymbol{H})$ and $(\log(L_{\boldsymbol{H}}))''(0) = 2p\operatorname{Tr}(\boldsymbol{H}^2)$.

Now we use the identity: $(\log(L_{\boldsymbol{H}}))''(0) = L''_{\boldsymbol{H}}(0)/L_{\boldsymbol{H}}(0) - (L'_{\boldsymbol{H}}(0)/L_{\boldsymbol{H}}(0))^2$. Since $L_{\boldsymbol{H}}(0) = 1$, we have $L''_{\boldsymbol{H}}(0) = (\log L_{\boldsymbol{H}})''(0) + ((\log L_{\boldsymbol{H}})'(0))^2$. Therefore

$$\mathbb{E}_{\boldsymbol{X}\sim\rho}[(\operatorname{Tr}(\boldsymbol{H}\boldsymbol{X}))^2] = 2p\operatorname{Tr}(\boldsymbol{H}^2) + p^2(\operatorname{Tr}(\boldsymbol{H}))^2. \tag{10}$$

Taking $\boldsymbol{H} = \boldsymbol{I}_d$ gives $M_1 = \mathbb{E}_{\boldsymbol{X}\sim\rho}[(\operatorname{Tr}(\boldsymbol{X}))^2] = 2p\operatorname{Tr}(\boldsymbol{I}_d^2) + p^2(\operatorname{Tr}(\boldsymbol{I}_d))^2 = 2pd + p^2d^2$.

To compute M_2, let

$$\mathcal{ON} = \{\boldsymbol{E}_{aa} : 1 \le a \le d\} \cup \left\{\frac{\boldsymbol{E}_{ab} + \boldsymbol{E}_{ba}}{\sqrt{2}} : 1 \le a < b \le d\right\}$$

be the standard Frobenius-orthonormal basis of $\mathrm{Sym}(d)$. Since $\operatorname{Tr}(\boldsymbol{X}^2) = \|\boldsymbol{X}\|_{\mathrm{F}}^2 = \sum_{\boldsymbol{H}\in\mathcal{ON}} (\operatorname{Tr}(\boldsymbol{H}\boldsymbol{X}))^2$, where the sum is over this orthonormal basis, and thus we have

$$\begin{aligned}
M_2 &= \mathbb{E}_{\boldsymbol{X}\sim\rho}[\operatorname{Tr}(\boldsymbol{X}^2)] \\
&= \mathbb{E}_{\boldsymbol{X}\sim\rho}\left[\sum_{a=1}^d (\operatorname{Tr}(\boldsymbol{E}_{aa}\boldsymbol{X}))^2 + \sum_{1\le a<b\le d}\left(\operatorname{Tr}\left(\frac{\boldsymbol{E}_{ab}+\boldsymbol{E}_{ba}}{\sqrt{2}}\boldsymbol{X}\right)\right)^2\right] \\
&= \sum_{a=1}^d \big(p^2(\operatorname{Tr}(\boldsymbol{E}_{aa}))^2 + 2p\operatorname{Tr}(\boldsymbol{E}_{aa}^2)\big) \\
&\quad + \sum_{1\le a<b\le d}\left(p^2\left(\operatorname{Tr}\left(\frac{\boldsymbol{E}_{ab}+\boldsymbol{E}_{ba}}{\sqrt{2}}\right)\right)^2 + 2p\operatorname{Tr}\left(\left(\frac{\boldsymbol{E}_{ab}+\boldsymbol{E}_{ba}}{\sqrt{2}}\right)^2\right)\right) \\
&= d(p^2 + 2p) + \frac{d(d-1)}{2}2p = p^2d + pd(d+1).
\end{aligned}$$

where the third equality follows by applying (10) to each element of the Frobenius orthonormal basis above.

For $d \geq 2$, Lemma 7 gives $b_\rho = \frac{M_2 - \frac{1}{d}M_1}{d^2+d-2}$. Substituting the values of M_1 and M_2,

$$b_\rho = \frac{p^2 d + pd(d+1) - \frac{1}{d}(p^2 d^2 + 2pd)}{d^2+d-2} = \frac{p(d^2+d-2)}{d^2+d-2} = p.$$

Then $a_\rho = \frac{M_1 - 2b_\rho d}{d^2} = \frac{p^2d^2+2pd-2pd}{d^2} = p^2$. For $d = 1$, we can only identify the combination $a_\rho + 2b_\rho = \mathbb{E}_\rho[X^2] = p^2 + 2p = p(p+2)$, as claimed. □

Corollary 2 (Optimal transport between Wishart distributions)**.** *Let* $\mu_i = \mathcal{W}_d(\boldsymbol{\Sigma}_i, p)$ *with* $\boldsymbol{\Sigma}_i \in \mathbb{S}^d_\succ$, $p > d-1$, $i = 0, 1$, *and let* $c(\boldsymbol{X}, \boldsymbol{Y}) = \|\boldsymbol{X} - \boldsymbol{Y}\|_{\mathrm{F}}^2$. *Set* $\boldsymbol{C} = (\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{1/2})^{1/2}$. *Then:*

(a) *the unique optimal Monge map from* μ_0 *to* μ_1 *is* $T^\star$, *where* $T^\star$ *is as defined in* (5),
(b) $(\mathrm{id}_{\mathbb{S}_\succ}, T^\star)_\# \mu_0$ *is the unique optimal Kantorovich plan,*
(c) $\mathbb{K}_c(\mu_0, \mu_1) = p(p+1)\|\boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}_1\|_{\mathrm{F}}^2 + p\left((\mathrm{Tr}(\boldsymbol{\Sigma}_0))^2 + (\mathrm{Tr}(\boldsymbol{\Sigma}_1))^2 - 2(\mathrm{Tr}(\boldsymbol{C}))^2\right)$,
(d) $\mathbb{M}_c(\mu_0, \mu_1) = \mathbb{K}_c(\mu_0, \mu_1)$.

Proof. By Lemma 9, $\rho = \mathcal{W}_d(\boldsymbol{I}_d, p)$ satisfies the hypotheses of Theorem 5.1 for $p > d-1$. Substituting the corresponding values of a_ρ and b_ρ from Lemma 9 into (6) yields the stated value of $\mathbb{K}_c$. The remaining claims follow directly from Theorem 5.1. □

References

[1] Jonathan M Borwein and Jon D Vanderwerff. *Convex functions: Constructions, characterizations and counterexamples*, volume 109. Cambridge University Press Cambridge, 2010.

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

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

[4] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on Pure and Applied Mathematics*, 44(4):375–417, 1991.

[5] 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.

[6] 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.

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

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

[9] Morris L. Eaton. *Multivariate Statistics: A Vector Space Approach*. John Wiley & Sons, 1983.

[10] Lawrence C Evans. *Measure theory and fine properties of functions*. Chapman and Hall/CRC, 2015.

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

[12] M. Seetharama Gowda, Roman Sznajder, and Jiyuan Tao. The automorphism group of a completely positive cone and its lie algebra. *Linear Algebra and its Applications*, 438:3862–3871, 2013.

[13] Arjun K Gupta and Daya K Nagar. *Matrix variate distributions*. Chapman and Hall/CRC, 2018.

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

[15] Leonid Kantorovich. On the transfer of masses. In *Doklady Akademii Nauk*, volume 37, page 227, 1942.

[16] 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.

[17] Anthony W. Knapp. *Lie Groups Beyond an Introduction*, volume 140 of *Progress in Mathematics*. Birkhäuser, 1996.

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

[19] Kanti V. Mardia, J. T. Kent, and J. M. Bibby. *Multivariate Analysis*. Academic Press, 1979.

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

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

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

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

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

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

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

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

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

[29] 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, 2023.

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

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

Appendix

This document contains supplementary material for the paper *Optimal Transport on Lie Group Orbits*.

Appendix A. Beyond Lie group orbits: one-dimensional distributions

The preceding examples fall within the finite-dimensional Lie group framework developed in the main paper. We now record the classical one-dimensional case as a complementary example. This example still has an orbit and stabilizer structure, but the acting transformation group is infinite-dimensional and is not a finite-dimensional Lie group. The results in this section should therefore be read as an orbit-based reformulation of the classical monotone rearrangement formula [5], rather than as an application of the finite-dimensional theory of the main paper.

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

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

Fix a reference probability measure ρ with a smooth, strictly positive density $r \in \mathcal{C}^\infty(\mathbb{R})$; a concrete example is the logistic density $r(x) = \frac{1}{4}\operatorname{sech}^2\big(x/2\big)$, 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 $\mathcal{C}^\infty$, strictly increasing bijection. By the inverse-function theorem, F_ρ is in fact a $\mathcal{C}^\infty$ diffeomorphism with smooth inverse $F_\rho^{-1} : (0,1) \to \mathbb{R}$.

Let $\mathcal{G}^+$ be the group of orientation-preserving $\mathcal{C}^1$ diffeomorphisms of $\mathbb{R}$: $\mathcal{G}^+ = \{g \in \mathcal{C}^1(\mathbb{R};\mathbb{R}) : g'(x) > 0\ \forall x \in \mathbb{R},\ g \text{ is bijective},\ g^{-1} \in \mathcal{C}^1(\mathbb{R};\mathbb{R})\}$, equipped with composition.

We regard $\mathcal{G}^+$ only as a transformation group. In particular, this one-dimensional example is not an application of the finite-dimensional Lie-group framework developed in the main paper. Nevertheless, it has the same orbit-stabilizer structure at the level of transformations. Namely, if $\mu_i = (g_i)_\# \rho$, $g_i \in \mathcal{G}^+$, $i = 0, 1$, then every map of the form $g_1 \circ h \circ g_0^{-1}$, $h \in \operatorname{Stab}_{\mathcal{G}^+}(\rho)$, transports μ_0 to μ_1. Thus we introduce the one-dimensional analogue of the orbit-reduced problem (LGOP). For this subsection only, we define

$$\mathbb{J}^1_c(g_0, g_1) = \inf_{h \in \operatorname{Stab}_{\mathcal{G}^+}(\rho)} \int_{\mathbb{R}} c\big(g_0(x), (g_1 \circ h)(x)\big)\,\mathrm{d}\rho(x), \tag{11}$$

where $\operatorname{Stab}_{\mathcal{G}^+}(\rho) = \{h \in \mathcal{G}^+ : h_\# \rho = \rho\}$.

Lemma 10. *For any $\mu \in \mathcal{P}^+$, $(F_\mu^{-1} \circ F_\rho)_\# \rho = \mu$ and $\mathcal{G}^+_\# \rho = \mathcal{P}^+$.*

Proof. First, we will show that $\mathcal{G}^+{}_\# \rho \subseteq \mathcal{P}^+$. Take $g \in \mathcal{G}^+$ 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 $F_\rho : \mathbb{R} \to (0,1)$ and $g^{-1} : \mathbb{R} \to \mathbb{R}$ are both $\mathcal{C}^1$ and strictly increasing, their composition $F_\rho \circ g^{-1}$ is $\mathcal{C}^1$ and strictly increasing on $\mathbb{R}$. 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 $\mathcal{P}^+ \subseteq \mathcal{G}^+_\# \rho$. 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 belongs to $\mathcal{C}^1((0,1))$ and strictly increasing. Define $g = F_\mu^{-1} \circ F_\rho : \mathbb{R} \to \mathbb{R}$, which also belongs to $\mathcal{C}^1(\mathbb{R})$. Hence, $g \in \mathcal{G}^+$. For any $\mathcal{A} \in \mathcal{B}(\mathbb{R})$,

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

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 $\mathcal{P}^+ \subseteq \mathcal{G}^+_\#\rho$. This observation completes our proof. □

Lemma 11. *The stabilizer of* ρ *under* $\mathcal{G}^+$ *is* $\mathrm{Stab}_{\mathcal{G}^+}(\rho) = \{\mathrm{id}_\mathbb{R}\}$.

Proof. Fix $g \in \mathcal{G}^+$ with $g_\#\rho = \rho$. For every $x \in \mathbb{R}$ we have

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

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 12. *Suppose that* $g_i = F_{\mu_i}^{-1} \circ F_\rho$, $i = 0, 1$. *Then,* (11) *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*

$$\mathbb{J}_c^1(g_0, g_1) = \int_0^1 (F_{\mu_1}^{-1}(t) - F_{\mu_0}^{-1}(t))^2 \mathrm{d}t. \tag{12}$$

Proof. By Lemma 11, the stabilizer group is a singleton, and thus $h^\star = \mathrm{id}_\mathbb{R}$ solves (11). Next, we evaluate $\mathbb{J}_c^1$:

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

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. $\square$

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

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

where $g_i = F_{\mu_i}^{-1} \circ F_\rho \in \mathcal{G}^+$ *for* $i = 0,1$.

Proof. By Lemma 10, we have $(g_i)_\# \rho = \mu_i$ for $i = 0,1$; hence $\mu_0, \mu_1 \in \mathcal{G}^+_\# \rho$. Define $T^\star = g_1 \circ h^\star \circ g_0^{-1}$, where $h^\star$ is the map defined in Lemma 12. 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 [30, Example 13.6], φ is c-convex. Next, we evaluate (♡):

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

By Lemma 4, the quadratic cost satisfies all hypotheses of Theorem 2.1, and the finite-second-moment assumption gives $\mathcal{K}_c(\mu_0,\mu_1) < \infty$. Therefore, by Theorem 2.1(v) $T^\star$ solves (MP) and the optimal Kantorovich plan is concentrated on the graph of $T^\star$.

If $X \sim \mu_0$, then $U = F_{\mu_0}(X) \sim \mathcal{U}([0,1])$. Hence

$$\int_\mathbb{R} |x - T^\star(x)|^2 \,\mathrm{d}\mu_0(x) = \int_0^1 \left(F_{\mu_0}^{-1}(t) - F_{\mu_1}^{-1}(t)\right)^2 \,\mathrm{d}t.$$

The equality with $\mathbb{J}_c^1(g_0,g_1)$ follows from Lemma 12. $\square$

Proposition 1 coincides with [5, Corollary 2.7] when restricted to the quadratic cost.

Appendix B. Additional example for Section 5

B.1. **The diagonal scaling mechanism on $\mathbb{R}^d_{>0}$.** Let the state space be $\mathcal{X} = \mathbb{R}^d_{>0}$, and let $G_{\text{diag}} = (\mathbb{R}^d_{>0}, \odot)$ be the componentwise multiplicative group. It acts on $\mathcal{X}$ by coordinatewise scaling: $\alpha(\boldsymbol{a})(\boldsymbol{x}) = \boldsymbol{a} \odot \boldsymbol{x}$, $\boldsymbol{a} \in \mathbb{R}^d_{>0}, \boldsymbol{x} \in \mathbb{R}^d_{>0}$. Equivalently, writing $D(\boldsymbol{a}) = \mathrm{diag}(a_1,\dots,a_d)$, we have $\alpha(\boldsymbol{a})(\boldsymbol{x}) = D(\boldsymbol{a})\boldsymbol{x}$. Thus the action is induced by a linear representation on $\mathcal{H} = \mathbb{R}^d$ with the standard inner product, with $b \equiv 0$, $\pi(\boldsymbol{a}) = D(\boldsymbol{a})$. Hence Assumption 4 holds.
Cartan structure. The linear image is $L = \pi(G_{\text{diag}}) = \{D(\boldsymbol{a}) : \boldsymbol{a} \in \mathbb{R}^d_{>0}\}$. We equip L with the involution $\theta(D(\boldsymbol{a})) = D(\boldsymbol{a})^{-\top} = D(\boldsymbol{a}^{-1})$. Then $L^\theta = \{D(\boldsymbol{a}) \in L : D(\boldsymbol{a}) = D(\boldsymbol{a}^{-1})\} = \{\boldsymbol{I}_d\}$. The Lie algebra of L is the space of diagonal matrices, $\mathfrak{l} = \{\mathrm{diag}(s_1,\dots,s_d) : s_i \in \mathbb{R}\}$. Since $\mathrm{d}\theta_{\boldsymbol{I}_d}(\boldsymbol{S}) = -\boldsymbol{S}$ for all $\boldsymbol{S} \in \mathfrak{l}$, we have $\mathfrak{k} = \{0\}$ and $\mathfrak{p} = \mathfrak{l}$. The exponential map restricts to a diffeomorphism $\mathfrak{p} \to L$, $\mathrm{diag}(s_1,\dots,s_d) \mapsto \mathrm{diag}(e^{s_1},\dots,e^{s_d})$. Therefore $L^\theta \times \mathfrak{p} \to L$, $(\boldsymbol{I}_d, \boldsymbol{S}) \mapsto \exp(\boldsymbol{S})$, is a diffeomorphism, so (L,θ) admits a global Cartan decomposition in the sense of Definition 7. Moreover every $\boldsymbol{S} \in \mathfrak{p}$ is diagonal, hence self-adjoint on $\mathbb{R}^d$. Thus hypothesis (i) of Theorem 4.3 holds.

Reference distribution. Let $\rho = \bigotimes_{i=1}^{d} \mathrm{Exp}(1)$, with density

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

The compact invariance condition of Theorem 4.3 is automatic in this example, because $L^\theta = \{\boldsymbol{I}_d\} \subseteq \pi(\mathrm{Stab}_{G_{\mathrm{diag}}}(\rho))$. Indeed, the identity element always belongs to the stabilizer.

Lemma 13 (Product exponential orbit and stabilizer)**.** *Let* $\rho = \bigotimes_{i=1}^{d} \mathrm{Exp}(1)$. *Then:*

(i) *for every* $\boldsymbol{\beta} \in \mathbb{R}^d_{>0}$, $\alpha(\boldsymbol{\beta}^{-1})_{\#}\rho = \bigotimes_{i=1}^{d} \mathrm{Exp}(\beta_i)$, *hence* $G_{\mathrm{diag}\,\#}\rho = \{\bigotimes_{i=1}^{d} \mathrm{Exp}(\beta_i) : \boldsymbol{\beta} \in \mathbb{R}^d_{>0}\}$,

(ii) $\mathrm{Stab}_{G_{\mathrm{diag}}}(\rho) = \{\mathbf{1}_d\}$.

Proof. Let $\boldsymbol{Z} \sim \rho$ and set $\boldsymbol{X} = \boldsymbol{a} \odot \boldsymbol{Z}$. For each coordinate, $X_i = a_i Z_i$. Since $Z_i \sim \mathrm{Exp}(1)$, the density of X_i is $x_i \mapsto a_i^{-1} \exp(-x_i/a_i) \mathbb{1}_{\{x_i>0\}}$. Taking $\boldsymbol{a} = \boldsymbol{\beta}^{-1}$ gives $X_i \sim \mathrm{Exp}(\beta_i)$, and the independence of the coordinates is preserved. This proves (i).

For (ii), suppose $\boldsymbol{a} \in \mathrm{Stab}_{G_{\mathrm{diag}}}(\rho)$. Then $\boldsymbol{a} \odot \boldsymbol{Z} \sim \boldsymbol{Z}$. Comparing the one-dimensional marginal densities gives, for every coordinate i, $a_i^{-1} \exp(-x_i/a_i) = \exp(-x_i)$ for all $x_i > 0$. Taking logarithms and comparing the coefficient of x_i gives $a_i = 1$. Hence $\boldsymbol{a} = \mathbf{1}_d$. □

Corollary 3 (Optimal transport between products of exponential laws)**.** *Let* $\mu_j = \bigotimes_{i=1}^{d} \mathrm{Exp}(\beta_{j,i})$, $\boldsymbol{\beta}_j = (\beta_{j,1}, \dots, \beta_{j,d}) \in \mathbb{R}^d_{>0}$, $j = 0, 1$. *For the quadratic cost* $c(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|^2$, *the following hold:*

(a) *the unique optimal Monge map from* μ_0 *to* μ_1 *is* $T^\star(\boldsymbol{x}) = (\frac{\beta_{0,1}}{\beta_{1,1}} x_1, \dots, \frac{\beta_{0,d}}{\beta_{1,d}} x_d)$,

(b) $(\mathrm{id}_{\mathbb{R}^d_{>0}}, T^\star)_{\#}\mu_0$ *is the unique optimal Kantorovich plan;*

(c) *the optimal transport value is* $\mathbb{K}_c(\mu_0, \mu_1) = 2 \sum_{i=1}^{d} (\beta_{0,i}^{-1} - \beta_{1,i}^{-1})^2$,

(d) $\mathbb{M}_c(\mu_0, \mu_1) = \mathbb{K}_c(\mu_0, \mu_1)$.

Proof. By Lemma 13, for $j = 0, 1$ we may write $\mu_j = \alpha(g_j)_{\#}\rho$, $g_j = \boldsymbol{\beta}_j^{-1} \in G_{\mathrm{diag}}$. The structural hypotheses of Theorem 4.3 were verified above. Since the stabilizer is trivial, the only possible stabilizing element is $h = \mathbf{1}_d$. Therefore $T_h = \alpha(g_1 h g_0^{-1}) = \alpha(\boldsymbol{\beta}_1^{-1} \odot \boldsymbol{\beta}_0)$, which gives $T_h(\boldsymbol{x}) = (\frac{\beta_{0,1}}{\beta_{1,1}} x_1, \dots, \frac{\beta_{0,d}}{\beta_{1,d}} x_d)$. The linear part is $\mathrm{diag}(\frac{\beta_{0,1}}{\beta_{1,1}}, \dots, \frac{\beta_{0,d}}{\beta_{1,d}})$, which is self-adjoint and positive definite. Hence Theorem 4.2 applies and proves (a), (b), and (d).

It remains only to compute the value. Since $h = \mathbf{1}_d$ is the unique stabilizer element,

$$\mathbb{J}_c(g_0, g_1) = \int_{\mathbb{R}^d_{>0}} \|g_0 \odot \boldsymbol{z} - g_1 \odot \boldsymbol{z}\|^2 \, \mathrm{d}\rho(\boldsymbol{z}) = \sum_{i=1}^{d} (\beta_{0,i}^{-1} - \beta_{1,i}^{-1})^2 \mathbb{E}_\rho[Z_i^2].$$

Since $Z_i \sim \mathrm{Exp}(1)$, $\mathbb{E}[Z_i^2] = 2$. Thus $\mathbb{J}_c(g_0, g_1) = 2 \sum_{i=1}^{d} (\beta_{0,i}^{-1} - \beta_{1,i}^{-1})^2$. The equality $\mathbb{K}_c(\mu_0, \mu_1) = \mathbb{J}_c(g_0, g_1)$ follows from Theorem 4.2, proving (c). □

Remark 1. *In this example the Cartan decomposition is degenerate: the fixed-point subgroup* L^θ *is trivial. Thus there is no nontrivial stabilizer optimization or rotational component to absorb. The example nevertheless fits the same structural template as the affine and congruence mechanisms: the orbit map has a self-adjoint positive definite linear part, and this algebraic fact certifies optimality.*

Many other product-form scale families admit the same analysis, provided the shape parameters are fixed and the relevant second moments are finite. In such cases, the orbit is generated by coordinate-wise positive scalings, the stabilizer is trivial, and the optimal map is again diagonal. The closed-form value reduces to a sum of one-dimensional second-moment calculations, exactly as in Corollary 3. Examples include products of Weibull, Rayleigh, Gamma, inverse-Gamma, Pareto, lognormal, and generalized-Gamma laws, under the corresponding parameter restrictions ensuring finite quadratic cost.

More generally, whenever both the cost and the measures factorize across coordinates, the Kantorovich problem decouples into one-dimensional problems: a product of one-dimensional optimal plans is optimal, and the Kantorovich value is the sum of the one-dimensional values. The one-dimensional case is treated in Section A.

Appendix C. Proofs and auxiliary results for Section 5.2

This appendix collects the orbit, stabilizer, and second-moment computations for the distributional families summarized in Table 1. Each subsection verifies the hypotheses of Theorem 5.1 and derives the constants a_ρ and b_ρ that enter the optimal transport cost formula (6).

Inverse-Wishart distributions. The inverse-Wishart distribution $\mathcal{IW}_d(\boldsymbol{\Psi}, p)$ with scale matrix $\boldsymbol{\Psi} \in \mathbb{S}^d_\succ$ and degrees of freedom $p > d-1$ is a probability law on $\mathbb{S}^d_\succ$ with density

$$\boldsymbol{X} \mapsto \frac{\det(\boldsymbol{\Psi})^{p/2}}{2^{pd/2}\Gamma_d(p/2)} \det(\boldsymbol{X})^{-(p+d+1)/2} \exp\left(-\frac{1}{2}\operatorname{Tr}(\boldsymbol{\Psi}\boldsymbol{X}^{-1})\right), \qquad \boldsymbol{X} \in \mathbb{S}^d_\succ. \tag{13}$$

Its mean exists for $p > d+1$, and the second moments of its matrix entries exist for $p > d+3$.

Lemma 14 (Inverse-Wishart orbit, stabilizer, and second moments)**.** *Let $\rho = \mathcal{IW}_d(\boldsymbol{I}_d, p)$ with $p > d+3$, and set $m = p - d$. Then:*

(i) *For every $\boldsymbol{A} \in \mathrm{GL}(d)$, $\alpha(\boldsymbol{A})_\# \rho = \mathcal{IW}_d(\boldsymbol{A}\boldsymbol{A}^\top, p)$.*
(ii) $\mathrm{Stab}_{\mathrm{GL}(d)}(\rho) = \mathcal{O}(d)$.
(iii) *For $d \geq 2$, the constants in Lemma 7 are $a_\rho = (m-2)/(m(m-1)(m-3))$ and $b_\rho = (m(m-1)(m-3))^{-1}$. If $d = 1$, only the combination $a_\rho + 2b_\rho$ is identifiable, and in this case $a_\rho + 2b_\rho = ((p-2)(p-4))^{-1}$.*

Proof. Parts (i) and (ii) are proved by the same arguments as in the proof of Lemma 9. We therefore omit the proofs of those assertions.
(iii) The condition $p > d+3$ is equivalent to $m = p-d > 3$, which is the condition under which the second moments of the inverse-Wishart entries are finite. Then, we have $\mathbb{E}_{\boldsymbol{X}\sim\rho}[\boldsymbol{X}] = (m-1)^{-1}\boldsymbol{I}_d$, so $\mathbb{E}_{\boldsymbol{X}\sim\rho}[X_{ab}] = \delta_{ab}/(m-1)$. Moreover, the entrywise covariance is

$$\operatorname{Cov}(X_{ab}, X_{cd}) = \frac{2\delta_{ab}\delta_{cd} + (m-1)(\delta_{ac}\delta_{bd} + \delta_{ad}\delta_{bc})}{m(m-1)^2(m-3)};$$

see, for example, [19, §3.8]. Using $\mathbb{E}_{\boldsymbol{X}\sim\rho}[X_{ab}X_{cd}] = \mathbb{E}_{\boldsymbol{X}\sim\rho}[X_{ab}]\mathbb{E}_{\boldsymbol{X}\sim\rho}[X_{cd}] + \operatorname{Cov}(X_{ab}, X_{cd})$, we get

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{X}\sim\rho}[X_{ab}X_{cd}] &= \frac{\delta_{ab}\delta_{cd}}{(m-1)^2} + \frac{2\delta_{ab}\delta_{cd} + (m-1)(\delta_{ac}\delta_{bd} + \delta_{ad}\delta_{bc})}{m(m-1)^2(m-3)} \\
&= \left(\frac{1}{(m-1)^2} + \frac{2}{m(m-1)^2(m-3)}\right)\delta_{ab}\delta_{cd} \\
&\quad + \frac{1}{m(m-1)(m-3)}(\delta_{ac}\delta_{bd} + \delta_{ad}\delta_{bc}).
\end{aligned}$$

A routine calculation shows that the coefficient of $\delta_{ab}\delta_{cd}$ simplifies as $\frac{m-2}{m(m-1)(m-3)}$. Therefore

$$\mathbb{E}[X_{ab}X_{cd}] = \frac{m-2}{m(m-1)(m-3)}\delta_{ab}\delta_{cd} + \frac{1}{m(m-1)(m-3)}(\delta_{ac}\delta_{bd} + \delta_{ad}\delta_{bc}).$$

Comparing this with the tensor form in Lemma 7 yields, for $d \geq 2$,

$$a_\rho = \frac{m-2}{m(m-1)(m-3)}, \qquad b_\rho = \frac{1}{m(m-1)(m-3)}.$$

For $d = 1$, the coefficients a_ρ and b_ρ are not separately identifiable. Only the combination $a_\rho + 2b_\rho$ is meaningful. Substituting $d = 1$, so that $m = p - 1$, gives

$$a_\rho + 2b_\rho = \frac{m-2+2}{m(m-1)(m-3)} = \frac{1}{(m-1)(m-3)} = \frac{1}{(p-2)(p-4)}.$$

This completes the proof. □

Corollary 4 (Optimal transport between inverse-Wishart distributions)**.** *For $i = 0, 1$, let $\mu_i = \mathcal{IW}_d(\boldsymbol{\Sigma}_i, p)$ with $\boldsymbol{\Sigma}_i \in \mathbb{S}^d_\succ$, $p > d+3$, and let $c(\boldsymbol{X}, \boldsymbol{Y}) = \|\boldsymbol{X} - \boldsymbol{Y}\|_{\mathrm{F}}^2$. Set $m = p - d$ and $\boldsymbol{C} = (\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{1/2})^{1/2}$. Then,*

(a) *the unique optimal Monge map from μ_0 to μ_1 is $T^\star$, where $T^\star$ is as defined in (5),*
(b) *$(\mathrm{id}_{\mathbb{S}^d_\succ}, T^\star)_\# \mu_0$ is the unique optimal Kantorovich plan,*
(c) *$\mathbb{K}_c(\mu_0, \mu_1) = (m(m-3))^{-1}\|\boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}_1\|_{\mathrm{F}}^2 + \frac{1}{m(m-1)(m-3)}\big((\mathrm{Tr}\,\boldsymbol{\Sigma}_0)^2 + (\mathrm{Tr}\,\boldsymbol{\Sigma}_1)^2 - 2(\mathrm{Tr}\,\boldsymbol{C})^2\big)$,*
(d) *$\mathbb{M}_c(\mu_0, \mu_1) = \mathbb{K}_c(\mu_0, \mu_1)$.*

Proof. By Lemma 14, $\rho = \mathcal{IW}_d(\boldsymbol{I}_d, p)$ satisfies the hypotheses of Theorem 5.1 for $p > d+3$. Substituting the corresponding values of a_ρ and b_ρ from Lemma 14 into (6) yields the stated value of $\mathbb{K}_c$. The remaining claims follow directly from Theorem 5.1. □

Matrix beta type II distributions. The matrix beta type II distribution $\mathcal{MB}^{\mathrm{II}}_d(q_1,\, q_2, \boldsymbol{\Sigma})$ with scale matrix $\boldsymbol{\Sigma} \in \mathbb{S}^d_\succ$ and shape parameters $q_1 > (d-1)/2$ and $q_2 > (d-1)/2$ is a probability law on $\mathbb{S}^d_\succ$ with density

$$\boldsymbol{X} \mapsto \frac{\det(\boldsymbol{\Sigma})^{q_2}}{\mathsf{B}_d(q_1, q_2)} \det(\boldsymbol{X})^{q_1 - (d+1)/2} \det(\boldsymbol{\Sigma} + \boldsymbol{X})^{-(q_1+q_2)}. \tag{14}$$

When $q_1 > (d-1)/2$ and $2q_2 > d+3$, the entry-wise second moments are finite; see [13, Theorem 5.3.20].

Lemma 15 (Matrix beta II orbit, stabilizer, and second moments)**.** *Let $\rho = \mathcal{MB}^{\mathrm{II}}_d(q_1, q_2, \boldsymbol{I}_d)$ with $q_1 > (d-1)/2$ and $2q_2 > d+3$, and set $q = 2q_2 - d$. Then:*

(i) *$\alpha(\boldsymbol{A})_\#\rho = \mathcal{MB}^{\mathrm{II}}_d(q_1, q_2, \boldsymbol{A}\boldsymbol{A}^\top)$ for every $\boldsymbol{A} \in \mathrm{GL}(d)$;*
(ii) *$\mathrm{Stab}_{\mathrm{GL}(d)}(\rho) = \mathcal{O}(d)$;*
(iii) *for $d \geq 2$, the constants in Lemma 7 are*

$$a_\rho = \frac{4q_1\big(q_1(q-2)+1\big)}{q(q-1)(q-3)} \text{ and } b_\rho = \frac{2q_1(2q_1+q-1)}{q(q-1)(q-3)}.$$

Proof. (i) follows by [13, Theorem 5.2.2].

(ii) If $\alpha(\boldsymbol{A})_\#\rho = \rho$, then by part (i), $\mathcal{MB}^{\mathrm{II}}_d(q_1, q_2, \boldsymbol{A}\boldsymbol{A}^\top) = \mathcal{MB}^{\mathrm{II}}_d(q_1, q_2, \boldsymbol{I}_d)$. Since the scale parameter is uniquely determined by the density in (14), it follows that $\boldsymbol{A}\boldsymbol{A}^\top = \boldsymbol{I}_d$.

(iii) By [13, Theorem 5.2.5], if $\boldsymbol{W}_1 \sim \mathcal{W}_d(\boldsymbol{I}_d, 2q_1)$ and $\boldsymbol{W}_2 \sim \mathcal{W}_d(\boldsymbol{I}_d, 2q_2)$ are independent, then $\boldsymbol{X} = \boldsymbol{W}_2^{-1/2}\boldsymbol{W}_1\boldsymbol{W}_2^{-1/2}$ satisfies $\boldsymbol{X} \sim \mathcal{MB}^{\mathrm{II}}_d(q_1, q_2, \boldsymbol{I}_d)$.

To compute the second moments of $\boldsymbol{X}$, we condition on $\boldsymbol{W}_2$. Given $\boldsymbol{W}_2$, the matrix $\boldsymbol{W}_2^{-1/2}$ is deterministic and $\boldsymbol{W}_1$ remains distributed as $\mathcal{W}_d(\boldsymbol{I}_d, 2q_1)$, because $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ are independent. Therefore, by the congruence transformation rule for the Wishart distribution,

$$\boldsymbol{X} \mid \boldsymbol{W}_2 = \boldsymbol{W}_2^{-1/2}\boldsymbol{W}_1\boldsymbol{W}_2^{-1/2} \sim \mathcal{W}_d(\boldsymbol{W}_2^{-1}, 2q_1).$$

Here the scale matrix $\boldsymbol{W}_2^{-1}$ is understood conditionally on $\boldsymbol{W}_2$.

Now we apply the Wishart second-moment formula to $\boldsymbol{X} \mid \boldsymbol{W}_2$. If $\boldsymbol{Y} \sim \mathcal{W}_d(\boldsymbol{\Sigma}, p)$, then by Lemma 9(iii)

$$\mathbb{E}_{\boldsymbol{Y}\sim\mathcal{W}_d(\boldsymbol{\Sigma},p)}[Y_{ab}Y_{cd}] = p^2\Sigma_{ab}\Sigma_{cd} + p\big(\Sigma_{ac}\Sigma_{bd} + \Sigma_{ad}\Sigma_{bc}\big).$$

Taking here $p = 2q_1$ and $\boldsymbol{\Sigma} = \boldsymbol{W}_2^{-1}$ yields

$$\begin{aligned} &\mathbb{E}_{\boldsymbol{X}\sim\mathcal{MB}_d^{\mathrm{II}}(q_1,q_2,\boldsymbol{I}_d)}[X_{ij}X_{kl} \mid \boldsymbol{W}_2] = \\ &(2q_1)^2(W_2^{-1})_{ij}(W_2^{-1})_{kl} + 2q_1\big((W_2^{-1})_{ik}(W_2^{-1})_{jl} + (W_2^{-1})_{il}(W_2^{-1})_{jk}\big). \end{aligned} \tag{15}$$

Next, since $\boldsymbol{W}_2^{-1} \sim \mathcal{IW}_d(\boldsymbol{I}_d, 2q_2)$, we may use the inverse-Wishart second-moment tensor from Lemma 14. Writing

$$\mathbb{E}_{\boldsymbol{W}_2^{-1}\sim\mathcal{IW}_d(\boldsymbol{I}_d,2q_2)}[(W_2^{-1})_{ij}(W_2^{-1})_{k\ell}] = \alpha_{\mathrm{IW}}\delta_{ij}\delta_{k\ell} + \beta_{\mathrm{IW}}\big(\delta_{ik}\delta_{j\ell} + \delta_{i\ell}\delta_{jk}\big),$$

with

$$\alpha_{\mathrm{IW}} = \frac{q-2}{q(q-1)(q-3)}, \qquad \beta_{\mathrm{IW}} = \frac{1}{q(q-1)(q-3)}, \qquad q = 2q_2 - d,$$

we take expectations in (15).

For the first term,

$$\mathbb{E}_{\boldsymbol{W}_2^{-1}\sim\mathcal{IW}_d(\boldsymbol{I}_d,2q_2)}[(W_2^{-1})_{ij}(W_2^{-1})_{kl}] = \alpha_{\mathrm{IW}}\delta_{ij}\delta_{kl} + \beta_{\mathrm{IW}}\big(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\big).$$

For the second term, apply the same formula twice:

$$\begin{aligned} \mathbb{E}_{\boldsymbol{W}_2^{-1}\sim\mathcal{IW}_d(\boldsymbol{I}_d,2q_2)}[(W_2^{-1})_{ik}(W_2^{-1})_{jl}] &= \alpha_{\mathrm{IW}}\delta_{ik}\delta_{jl} + \beta_{\mathrm{IW}}\big(\delta_{ij}\delta_{kl} + \delta_{il}\delta_{jk}\big), \\ \mathbb{E}_{\boldsymbol{W}_2^{-1}\sim\mathcal{IW}_d(\boldsymbol{I}_d,2q_2)}[(W_2^{-1})_{il}(W_2^{-1})_{jk}] &= \alpha_{\mathrm{IW}}\delta_{il}\delta_{jk} + \beta_{\mathrm{IW}}\big(\delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl}\big). \end{aligned}$$

Summing these two identities gives

$$\begin{aligned} &\mathbb{E}_{\boldsymbol{W}_2^{-1}\sim\mathcal{IW}_d(\boldsymbol{I}_d,2q_2)}[(W_2^{-1})_{ik}(W_2^{-1})_{jl}] + \mathbb{E}[(W_2^{-1})_{il}(W_2^{-1})_{jk}] \\ &\quad = 2\beta_{\mathrm{IW}}\delta_{ij}\delta_{kl} + (\alpha_{\mathrm{IW}} + \beta_{\mathrm{IW}})\big(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\big). \end{aligned}$$

Substituting back into (15), we obtain $\mathbb{E}[X_{ij}X_{kl}] = a_\rho\delta_{ij}\delta_{kl} + b_\rho(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk})$, where

$$a_\rho = (2q_1)^2\alpha_{\mathrm{IW}} + 2q_1 \cdot 2\beta_{\mathrm{IW}}, \quad b_\rho = (2q_1)^2\beta_{\mathrm{IW}} + 2q_1(\alpha_{\mathrm{IW}} + \beta_{\mathrm{IW}}).$$

Finally, inserting the values of α_{IW} and β_{IW} yields

$$\begin{aligned} a_\rho &= \frac{4q_1^2(q-2) + 4q_1}{q(q-1)(q-3)} = \frac{4q_1\big(q_1(q-2)+1\big)}{q(q-1)(q-3)}, \\ b_\rho &= \frac{4q_1^2 + 2q_1(q-1)}{q(q-1)(q-3)} = \frac{2q_1(2q_1+q-1)}{q(q-1)(q-3)}. \end{aligned}$$

□

Corollary 5 (Optimal transport between matrix beta II distributions). *For $i = 0, 1$ let $\mu_i = \mathcal{MB}_d^{\mathrm{II}}(q_1, q_2, \boldsymbol{\Sigma}_i)$ with $\boldsymbol{\Sigma}_i \in \mathbb{S}_\succ^d$, $q_1 > (d-1)/2$, $2q_2 > d+3$, $i = 0, 1$, and let $c(\boldsymbol{X}, \boldsymbol{Y}) = \|\boldsymbol{X} - \boldsymbol{Y}\|_{\mathrm{F}}^2$. Set $q = 2q_2 - d$ and $\boldsymbol{C} = (\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{1/2})^{1/2}$. Then,*

(a) *the unique optimal Monge map from μ_0 to μ_1 is $T^\star$, where $T^\star$ is as defined in (5),*
(b) $(\mathrm{id}_{\mathbb{S}_\succ^d}, T^\star)_\#\mu_0$ *is the unique optimal Kantorovich plan,*

(c)

$$\mathbb{K}_c(\mu_0, \mu_1) = \frac{2q_1(2q_1 q - 2q_1 + q + 1)}{q(q-1)(q-3)} \|\boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}_1\|_{\mathrm{F}}^2 + \frac{2q_1(2q_1 + q - 1)}{q(q-1)(q-3)} \left((\operatorname{Tr} \boldsymbol{\Sigma}_0)^2 + (\operatorname{Tr} \boldsymbol{\Sigma}_1)^2 - 2(\operatorname{Tr} \boldsymbol{C})^2 \right),$$

(d) $\mathbb{M}_c(\mu_0, \mu_1) = \mathbb{K}_c(\mu_0, \mu_1)$.

Proof. By Lemma 15, $\rho = \mathcal{MB}_d^{\mathrm{II}}(q_1, q_2, \boldsymbol{I}_d)$ satisfies the hypotheses of Theorem 5.1 for $q_1 > (d-1)/2$ and $2q_2 > d+3$. Substituting the corresponding values of a_ρ and b_ρ from Lemma 15 into (6) yields the stated value of $\mathbb{K}_c$. The remaining claims follow directly from Theorem 5.1. □