EXPONENTIALLY CONCAVE FUNCTIONS AND A NEW INFORMATION GEOMETRY[1]

By Soumik Pal and Ting-Kam Leonard Wong

University of Washington and University of Southern California

A function is exponentially concave if its exponential is concave. We consider exponentially concave functions on the unit simplex. In a previous paper, we showed that gradient maps of exponentially concave functions provide solutions to a Monge–Kantorovich optimal transport problem and give a better gradient approximation than those of ordinary concave functions. The approximation error, called L-divergence, is different from the usual Bregman divergence. Using tools of information geometry and optimal transport, we show that L-divergence induces a new information geometry on the simplex consisting of a Riemannian metric and a pair of dually coupled affine connections which defines two kinds of geodesics. We show that the induced geometry is dually projectively flat but not flat. Nevertheless, we prove an analogue of the celebrated generalized Pythagorean theorem from classical information geometry. On the other hand, we consider displacement interpolation under a Lagrangian integral action that is consistent with the optimal transport problem and show that the action minimizing curves are dual geodesics. The Pythagorean theorem is also shown to have an interesting application of determining the optimal trading frequency in stochastic portfolio theory.

1. Introduction.

DEFINITION 1.1 (Exponential concavity). Let $\Omega \subset \mathbb{R}^d$ be convex. We say that a function $\varphi : \Omega \to \mathbb{R} \cup \{-\infty\}$ is exponentially concave if $\Phi = e^{\varphi}$ is concave. (By convention, we set $e^{-\infty} = 0$.)

Throughout this paper, we let Ω be the open unit simplex

$$(1.1) \qquad \Delta_n = \left\{ p = (p_1, \ldots, p_n) \in \mathbb{R}^n : p_i > 0, \sum_{i=1}^{n} p_i = 1 \right\},$$

regarded as the collection of strictly positive probability distributions on a set with n elements. This is due to the applications we have in mind, although many generalizations are possible. An interesting property of exponentially concave functions

is that their gradient maps give a better first-order approximation those of than ordinary concave functions. In [45], we introduced the concept of L-divergence. Let $\varphi : \Delta_n \to \mathbb{R}$ be a differentiable exponentially concave function. For $p, q \in \Delta_n$, concavity of e^φ implies that

$$(1.2) \qquad \varphi(p) + \log\big(1 + \nabla\varphi(p) \cdot (q - p)\big) \geq \varphi(q),$$

where $\nabla\varphi$ is the Euclidean gradient. Clearly this approximation is sharper than the linear approximation of φ itself. The L-divergence of φ is the error in this approximation:

$$(1.3) \qquad T(q|p) := \log\big(1 + \nabla\varphi(p) \cdot (q - p)\big) - \big(\varphi(q) - \varphi(p)\big) \geq 0.$$

The extra concavity of exponentially concave functions have found several recent applications in analysis, probability and optimization. For example, in [22], the equivalence of entropic curvature-dimension conditions and Bochner's inequality on metric measure spaces is established using the notion of (K, N) convexity. When $K \geq 0$ and $N < \infty$, the negative of a (K, N) convex function is exponentially concave. Better gradient approximation has also led to better algorithms in optimization and machine learning such as those in [29, 30, 35], although the authors tend to replace the logarithmic term in (1.2) by a quadratic approximation.

One of our primary applications in mind is related to stochastic portfolio theory. In [23, 24], the author considers the gradient map of an exponentially concave function as a map from Δ_n to its closure $\overline{\Delta}_n$. The following restatement can be found in [45], Proposition 6. Let φ be a differentiable exponentially concave function on Δ_n. For $p \in \Delta_n$, define $\boldsymbol{\pi}(p) \in \mathbb{R}^n$ by

$$(1.4) \qquad \boldsymbol{\pi}_i(p) = p_i\big(1 + \nabla\varphi(p) \cdot \big(e^{(i)} - p\big)\big), \qquad i = 1, \ldots, n,$$

where $e^{(i)} = (0, \ldots, 1, \ldots, 0)$ is the ith standard basis of $\mathbb{R}^n$. Then, it can be shown that $\boldsymbol{\pi} : \Delta_n \to \overline{\Delta}_n$. In keeping with standard definitions in the subject we will call this map the portfolio map. In this vein, also see articles [6, 25, 26, 31, 42, 49, 53, 54].

The L-divergence of φ should be distinguished from the Bregman divergence of φ defined by

$$(1.5) \qquad D(q|p) := \nabla\varphi(p) \cdot (q - p) - \big(\varphi(q) - \varphi(p)\big).$$

Bregman divergence was introduced in [12] and is widely applied in statistics and optimization. To see the difference, consider two fundamental examples. For $q, p \in \Delta_n$, the Kullback–Leibler divergence (also known as relative entropy) is given by

$$H(q|p) = \sum_{i=1}^{n} q_i \log \frac{q_i}{p_i}.$$

It can be shown that the relative entropy is the Bregman divergence of the Shannon entropy $H(p) = -\sum_{i=1}^{n} p_i \log p_i$. On the other hand, fix $\pi \in \Delta_n$ and consider the cross entropy $\varphi(p) = \sum_{i=1}^{n} \pi_i \log p_i$. This is an exponentially concave function of p whose associated portfolio map (1.4) is constant: $\boldsymbol{\pi}(p) \equiv \pi$. The corresponding L-divergence is given by

$$(1.6) \qquad T(q|p) = \log\left(\sum_{i=1}^{n} \pi_i \frac{q_i}{p_i}\right) - \sum_{i=1}^{n} \pi_i \log\left(\frac{q_i}{p_i}\right).$$

This quantity is sometimes referred to as the free energy in statistical physics. In finance it is called the diversification return in [9, 14, 21, 52], the excess growth rate in [23, 27, 44], the rebalancing premium in [10], and the volatility return in [28].

In [45], we introduced a Monge–Kantorovich optimal transport problem on $\mathbb{R}^{n-1}$ which can be solved using exponentially concave functions on the unit simplex. The cost function is defined for $(\theta, \phi) \in \mathbb{R}^{n-1} \times \mathbb{R}^{n-1}$ by

$$(1.7) \qquad c(\theta, \phi) := \psi(\theta - \phi), \qquad \text{where } \psi(x) := \log\left(1 + \sum_{i=1}^{n-1} e^{x_i}\right)$$

is strictly convex on $\mathbb{R}^{n-1}$. We will recall the details of this transport problem in Section 2.1 and its relationship to exponentially concave functions. It suffices to say for now that, given a pair of Borel probability measures P and Q on $\mathbb{R}^{n-1}$ the optimal coupling of the two with respect to the above cost can be expressed in terms of the portfolio map of an exponentially concave function on the simplex. A related cost function appears in [41] in the completely different context of finding polytopes with given geometric data. It also appears to be related to the study of moment measures as introduced in [15] (see page 3836 in particular).

1.1. *Our contributions.*

In this paper, we show that information geometry provides an elegant geometric structure underlying exponential concavity, L-divergence and the optimal transport problem. Here is a motivating question which is the starting point of this work. Suppose φ is an exponentially concave function on Δ_n with its associated L-divergence $T(\cdot|\cdot)$. Can we geometrically characterize triplets $(p, q, r) \in (\Delta_n)^3$ such that $T(q|p) + T(r|q) \leq T(r|p)$? The answer to this question determines the optimal frequency of rebalancing the portfolio generated by φ (see Section 5.4). Also see Section 3.3 for a transport interpretation of this inequality.

Using tools of information geometry, we show that exponentially concave functions on Δ_n and their L-divergences induce a new geometric structure on the simplex Δ_n regarded as a smooth manifold of probability distributions. Let φ be an exponentially concave function on Δ_n. We only require that φ is smooth and the Euclidean Hessian of e^φ is strictly positive definite everywhere (see Assumption 2.5).
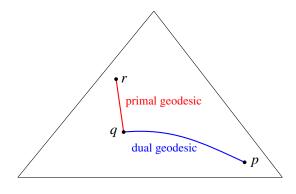
FIG. 1. *Generalized Pythagorean theorem for L-divergence.*

The induced geometric structure consists of a Riemannian metric g and a dual pair of affine connections ∇ and ∇^*. These connections define via parallel transports two kinds of geodesic curves on Δ_n called primal and dual geodesics. Interestingly, the duality in this geometry goes hand in hand with the duality in the related Monge–Kantorovich optimal transport problem, and this work is the first which exploits this connection. We summarize our main results as follows. First, we give the answer of the motivating question.

THEOREM 1.2 (Generalized Pythagorean theorem). *Given* $(p, q, r) \in (\Delta_n)^3$, *consider the dual geodesic joining q and p and the primal geodesic joining q and r. Consider the Riemannian angle between the geodesics at q.* (*See Proposition* 4.4 *which expresses the Riemannian metric g as a normalized Euclidean Hessian of* $\Phi := e^\varphi$.) *Then the difference*

$$(1.8) \qquad T(q|p) + T(r|q) - T(r|p)$$

is positive, zero or negative depending on whether the angle is less than, equal to, or greater than 90 *degrees* (*see Figure* 1).

We also prove other remarkable properties of the geodesics: (i) There exist explicit coordinate systems under which the primal and dual geodesics are time changes of Euclidean straight lines (Theorem 5.1). In other words, the new geometry is dually projectively flat. In particular, the primal geodesics are Euclidean straight lines up to time reparameterization. Moreover, the primal and dual connections have constant sectional curvature -1 with respect to the Riemannian metric, and thus satisfy an Einstein condition (Corollary 4.10). The primal and dual geodesics can also be constructed as time changes of Riemannian gradient flows for the functions $T(r|\cdot)$ and $T(\cdot|p)$ (Theorem 5.5). This is remarkable because while the geodesic equations depend only on the local properties of $T(\xi|\xi')$ near $\xi = \xi'$, the gradient flows are global as they involve the derivatives of $T(r|\cdot)$

and $T(\cdot|p)$. Indeed, this relation is known only for limited families of divergence including Bregman divergence and α-divergence [5].

As shown in [2], Chapter 1, the generalized Pythagorean theorem holds for any Bregman divergence which induces a dually flat geometry. We will prove that the resulting geometries from L-divergences are not flat for $n \geq 3$ (Theorem 4.9). While some extensions of the generalized Pythagorean theorem hold in certain nonflat spaces (see, e.g., [2], Theorem 4.5), they involve some extra terms. To the best of our knowledge Theorem 1.2 is the first exact Pythagorean theorem that holds in a geometry which is not dually flat. The difference (1.8) can also be given an optimal transport interpretation (Section 3.3).

(ii) We extend the static transport problem (1.7) to a time-dependent transport problem with a corresponding convex Lagrangian action. In Theorem 6.2, we show that the action minimizing curves are the (reparametrized) dual geodesics which, in addition, satisfy the intermediate time optimality condition. This allows for a consistent displacement interpolation formulation between probability measures on the unit simplex. Previously, such studies focused almost exclusively on the Wasserstein spaces corresponding to the cost functions $C(x, y) = d(x, y)^\alpha$ (here d is a metric on a Polish space with suitable properties and $\alpha \geq 1$). Displacement interpolation and the related concept of displacement convexity were introduced in [38] and in the thesis [37]. These ideas have grown to be immensely important in classical Wasserstein transport with fundamental implications in geometry, physics, probability and PDE. See [51], Chapter 7, for a thorough discussion. Our Lagrangian, although convex, is not superlinear, and, therefore, is not covered by the standard theory. However, we expect it to lead to many equally remarkable properties.

These results suggest plenty of problems for further research. Generalizing Theorem 1.2 to more than three points is of interest in stochastic portfolio theory. Displacement interpolation has become an extremely important topic in optimal transport theory. Extensions to Riemanninan manifolds, done in [16], have led to new functional inequalities. In another vein, [34] defines Ricci curvatures on metric measure spaces in terms of displacement interpolation and displacement convexity. We expect that the displacement interpolation in this paper will lead to a new Otto calculus ([51], Chapter 15) and related PDEs (such as Hamilton–Jacobi equations). It appears that Bregman divergence and L-divergence are only two of an entire family of divergences with special properties and corresponding optimal transport problems. For example, see [43] which extends the optimal transport problem (1.7) via the cumulant generating function of a general probability distribution. We also believe that this new information geometry will be useful in dynamic optimization problems where the objective function is multiplicative in time. Finally, it is naturally of interest to study exponential concavity on general convex domains.

1.2. *Related literature.* We have mentioned the L-divergence and the Bregman divergence. In general, a divergence on a set (usually a manifold of probability distributions) is a nonnegative function $D(q|p)$ such that $D(q|p) = 0$ if and only if $p = q$. Divergences are not metrics in general since they may be asymmetric and may not satisfy the triangle inequality. Apart from Bregman divergence, many families of divergences (such as α-divergence and f-divergence) have been applied in information theory, statistics and other areas; see the survey [7] for a catalog of these divergences. Among these divergences, Bregman divergence plays a special role because it induces a dually flat geometry on the underlying space. First studied in the context of exponential families in statistical inference by [40], it gave rise to information geometry—the geometric study of manifolds of probability distributions. Furthermore, Bregman divergence enjoys properties such as the generalized Pythagorean theorem and projection theorem which led to numerous applications. See [2, 3, 13, 32, 39] for introductions to this beautiful theory. The related concept of dual affine connection is also useful in affine differential geometry (see [17, 33, 47]). In [36], dually projectively flat manifolds are characterized in terms of the Bartlett tensors and conformal flatness. Here we identify a new and important class of examples and show that they have concrete applications.

This work is motivated by our study in mathematical finance. Recently optimal transport has been applied to financial problems such as robust asset pricing; see, for example, [1, 8, 18]. This line of work has a somewhat different flavor than ours although they share the same goal: development of model-free mathematical finance. Portfolios generated by exponentially concave functions generate profit due to fluctuations of a sequence in Δ_n representing the stock market. This idea is sometimes called volatility harvesting and leads naturally to the transport problem (1.7), as shown in [45]. In this philosophy, our work can be interpreted as developing a notion of model-free volatility.

1.3. *Outline of the paper.* In the next section, we recall the optimal transport problem formulated in [45] using the exponential coordinate system. Its relation with functionally generated portfolio is also reviewed. In Section 3, we relate exponential concavity with c-concavity and give a transport-motivated definition of L-divergence. Here duality plays a crucial role. After reviewing the basic concepts of information geometry, we derive in Section 4 the geometric structure induced by an exponentially concave function. The properties of this new geometry are then studied in Section 5. In particular, we characterize the primal and dual geodesics and prove the generalized Pythagorean theorem which has an interesting application in mathematical finance. Finally, in Section 6 we apply the geometric structure to construct a displacement interpolation for the associated optimal transport problem. Some technical and computational details are gathered in the Appendix.

2. Optimal transport and portfolio maps. In this section, we recall the optimal transport problem in [45] using the exponential coordinate system. We also

review the definition of functionally generated portfolio and explain how it relates to the transport problem.

2.1. *Exponential coordinate system.* The exponential coordinate system defines a global coordinate system on Δ_n regarded as an $(n-1)$-dimensional smooth manifold [2], Section 2.2.

DEFINITION 2.1 (Exponential coordinate system). The exponential coordinate $\theta = (\theta_1, \ldots, \theta_{n-1}) \in \mathbb{R}^{n-1}$ of $p \in \Delta_n$ is given by

$$\tag{2.1} \theta_i = \log \frac{p_i}{p_n}, \qquad i = 1, \ldots, n-1.$$

We denote this map by $\boldsymbol{\theta} : \Delta_n \to \mathbb{R}^{n-1}$. By convention, we set $\theta_n \equiv 0$. The inverse transformation $\mathbf{p} := \boldsymbol{\theta}^{-1}$ is given by

$$\tag{2.2} p_i = \mathbf{p}_i(\theta) = e^{\theta_i - \psi(\theta)}, \qquad 1 \le i \le n,$$

where $\psi(\theta) = \log(1 + \sum_{i=1}^{n-1} e^{\theta_i}) = \log(\sum_{i=1}^{n} e^{\theta_i})$ as defined in (1.7).

The exponential coordinate system is the first of several coordinate systems we will introduce on the simplex. By changing coordinate systems, any function on Δ_n can be expressed as a function on $\mathbb{R}^{n-1}$ and vice versa. Explicitly, a function φ on Δ_n can be expressed in exponential coordinates by $\theta \mapsto \varphi(\mathbf{p}(\theta))$. To simplify the notations, we simply write $\varphi(p)$ or $\varphi(\theta)$ depending on the coordinate system used. For example, if $\varphi(p) = \sum_{i=1}^{n} \pi_i \log p_i$ is the cross entropy where $\pi \in \Delta_n$, then $\varphi(\theta) = \sum_{i=1}^{n-1} \pi_i \theta_i - \psi(\theta)$.

2.2. *The transport problem.* We refer the reader to the books [4, 51] for introductions to optimal transport and its interplay with analysis, probability and geometry. Let $\mathcal{X} = \mathcal{Y} = \mathbb{R}^{n-1}$ be equipped with the standard Euclidean metric and topology. Let P and Q be Borel probability measures on $\mathcal{X}$ and $\mathcal{Y}$, respectively. By a coupling of P and Q we mean a Borel probability measure R on $\mathcal{X} \times \mathcal{Y}$ whose marginals are P and Q, respectively. Let $\Pi(P, Q)$ be the set of all couplings of P and Q. This set is always nonempty as it contains the product measure $P \otimes Q$.

Given P and Q we consider the Monge–Kantorovich optimal transport problem with cost c defined by (1.7):

$$\tag{2.3} \inf_{R \in \Pi(P, Q)} \mathbb{E}\big[c(\theta, \phi)\big].$$

Here the expectation is taken under the probability measure R under which the random element (θ, ϕ) has distribution R. If an optimal coupling takes the form $(\theta, F(\theta))$ for some measurable map $F : \mathcal{X} \to \mathcal{Y}$, we say that F is a Monge transport map.

In general, we may consider the optimal transport problem (2.3) with c replaced by a general cost function denoted by $C(\cdot, \cdot)$ and $\mathcal{X}$, $\mathcal{Y}$ are general Polish spaces. The classical example is where $\mathcal{X} = \mathcal{Y}$ and C is a power of the underlying metric: $C(x, y) = d(x, y)^\alpha$ (especially $\alpha = 1$ and 2). For these costs rich and delicate theories have been developed on Euclidean spaces, Riemannian manifolds and geodesic metric measure spaces. However, we consider the cost function c defined by (1.7).

REMARK 2.2. The cost function

$$\widetilde{c}(\theta, \phi) := \log\left(\frac{1}{n} + \frac{1}{n}\sum_{i=1}^{n-1} e^{\theta_i - \phi_i}\right) - \sum_{i=1}^{n-1}\frac{1}{n}(\theta_i - \phi_i)$$

differs from $c(\cdot, \cdot)$ in a linear term which plays no role in optimal transport. Thus we may consider $\widetilde{c}$ instead. The advantage of $\widetilde{c}$ is that it is nonnegative and, by Jensen's inequality, equals zero if and only if $\theta = \phi$. To be consistent with the notations in [45], we will use the cost function c in this paper.

DEFINITION 2.3 (c-cyclical monotonicity). A nonempty subset $A \subset \mathcal{X} \times \mathcal{Y}$ is c-cyclical monotone if and only if it satisfies the following property. For any finite collection $\{(\theta_j, \phi_j)\}_{j=1}^m$ in A and any permutation σ of the set $\{1, 2, \ldots, m\}$, we have

$$(2.4) \qquad \sum_{j=1}^n c(\theta_j, \phi_j) \le \sum_{j=1}^m c(\theta_j, \phi_{\sigma(j)}).$$

It is well known that c-cyclical monotonicity is, under mild technical conditions, a necessary and sufficient solution criteria of the general optimal transport problem (see [4], Chapter 1). In particular, a coupling R of (P, Q) is optimal if and only if the support of R is c-cyclical monotone.

2.3. *Functionally generated portfolio.* At this point, it is convenient to introduce the concept of functionally generated portfolio. Although it is possible to present the theory without reference to finance-motivated concepts, we stress that the portfolio map gives an additional structure to the transport problem not found in other cases. Also, the main examples of the theory as well as the key quantities (such as the induced Riemannian metric) are best expressed in terms of portfolios. Mathematically, the portfolio can be regarded as a normalized gradient of φ. In Section 5.4, we apply our information geometry to functionally generated portfolios.

Functionally generated portfolio was introduced in [24] and the following refined definition is taken from [45].

DEFINITION 2.4 (Functionally generated portfolio).   By a portfolio map, we mean a function $\boldsymbol{\pi} : \Delta_n \to \overline{\Delta}_n$. Let $\varphi : \Delta_n \to \mathbb{R}$ be exponentially concave. We say that a portfolio map $\boldsymbol{\pi} : \Delta_n \to \overline{\Delta}_n$ is generated by φ if for any $p, q \in \Delta_n$ we have

$$(2.5) \qquad \sum_{i=1}^{n} \boldsymbol{\pi}_i(p) \frac{q_i}{p_i} \geq e^{\varphi(q) - \varphi(p)}.$$

We call φ the log generating function of $\boldsymbol{\pi}$ and $\Phi := e^{\varphi}$ (which is positive and concave) the generating function. It is known that φ is unique (for a given $\boldsymbol{\pi}$) up to an additive constant. If φ is differentiable, then $\boldsymbol{\pi}$ is necessarily given by (1.4).

Throughout this paper, we impose the following regularity conditions on the exponentially concave function φ.

ASSUMPTION 2.5 (Regularity conditions).   (i) The function φ is smooth (i.e., infinitely differentiable) on Δ_n.

(ii) The (Euclidean) Hessian of $\Phi = e^{\varphi}$ is strictly negative definite everywhere on Δ_n. In particular, Φ is strictly concave. Moreover, it can be shown that the function $\boldsymbol{\pi}$ defined by (1.4) maps Δ_n into Δ_n.

Let us discuss these conditions briefly. Differentiability is needed to define differential geometric structures on Δ_n in terms of the derivatives of the L-divergence. Our theory requires the L-divergence $T(\cdot|\cdot)$ to be three times continuously differentiable, and for convenience we simply assume that φ is smooth. Strict concavity guarantees that the L-divergence is nondegenerate, that is, $T(q|p) = 0$ only if $q = p$, and strict positive definiteness of the Hessian implies that the induced Riemannian metric is nondegenerate.

Henceforth, we let $\varphi : \Delta_n \to \mathbb{R}$ be an exponentially concave function satisfying Assumption 2.5 and let $\boldsymbol{\pi} : \Delta_n \to \Delta_n$ given by (1.4) be the portfolio map generated by φ. The cost function c always refers to the one defined in (1.7), and a general cost function is denoted by C. Using (1.4), it can be shown that the L-divergence (1.3) of φ can be expressed in the form

$$(2.6) \qquad T(q|p) = \log\left(\sum_{i=1}^{n} \boldsymbol{\pi}_i(p) \frac{q_i}{p_i}\right) - (\varphi(q) - \varphi(p)).$$

Now we give several examples of functionally generated portfolios and their log generating functions. Many more examples can be found in [23], Chapter 3. In particular, the constant-weighted portfolios play a special role and will be taken as the basic example of the theory.

EXAMPLE 2.6 (Examples of functionally generated portfolios).   (i) (Market portfolio) The identity map $\boldsymbol{\pi}(p) \equiv p$ is generated by the constant function $\varphi(p) \equiv 0$. (Here Assumption 2.5 does not hold.)

(ii) (Constant-weighted portfolio) The constant map $\boldsymbol{\pi}(p) \equiv \pi \in \Delta_n$ is generated by the cross-entropy $\varphi(p) = \sum_{i=1}^{n} \pi_i \log p_i$. The special case $\pi = (\frac{1}{n}, \ldots, \frac{1}{n})$ is called the equal-weighted portfolio.

(iii) (Diversity-weighted portfolio) Let $\lambda \in (0, 1)$ be a fixed parameter. Consider the portfolio map $\boldsymbol{\pi} : \Delta_n \to \Delta_n$ defined by

$$\boldsymbol{\pi}_i(p) = \frac{p_i^\lambda}{\sum_{j=1}^{n} p_j^\lambda}, \qquad i = 1, \ldots, n.$$

It can be shown that the generating function is $\varphi(p) = \frac{1}{\lambda} \log(\sum_{j=1}^{n} p_j^\lambda)$.

(iv) (Convex combinations) It is known that the set of functionally generated portfolios is convex. Indeed, let $\boldsymbol{\pi}^{(1)}$ be generated by $\varphi^{(1)}$ and $\boldsymbol{\pi}^{(2)}$ be generated by $\varphi^{(2)}$. Then for $\lambda \in (0, 1)$, the portfolio map $\boldsymbol{\pi} = (1-\lambda)\boldsymbol{\pi}^{(1)} + \lambda\boldsymbol{\pi}^{(2)}$ is generated by $\varphi = (1 - \lambda)\varphi^{(1)} + \lambda\varphi^{(2)}$. Its generating function $\Phi = e^\varphi$ is then the geometric mean $(\Phi^{(1)})^{1-\lambda}(\Phi^{(2)})^\lambda$. This fact was used in [53, 54] to formulate and study nonparametric estimation of functionally generated portfolio.

The following result is taken from [45].

PROPOSITION 2.7. *For any portfolio map $\boldsymbol{\pi} : \Delta_n \to \Delta_n$ the following statements are equivalent.*

(i) *There exists an exponentially concave function φ on Δ_n which generates $\boldsymbol{\pi}$ in the sense of* (2.5).

(ii) *The portfolio map is multiplicatively cyclical monotone (MCM) in the following sense: for any sequence $\{\mu(t)\}_{t=0}^{m+1}$ in Δ_n satisfying $\mu(m + 1) = \mu(0)$, we have*

$$\prod_{t=0}^{m} \left(\sum_{i=1}^{n} \boldsymbol{\pi}_i(\mu(t)) \frac{\mu_i(t + 1)}{\mu_i(t)} \right) \geq 1.$$

(iii) *Define a map $\theta \mapsto \phi$ by*

$$\text{(2.7)} \qquad \phi_i = \theta_i - \log \frac{\boldsymbol{\pi}_i(\theta)}{\boldsymbol{\pi}_n(\theta)}, \qquad i = 1, \ldots, n - 1.$$

Here $\boldsymbol{\pi}$ is regarded as a function of the exponential coordinate. In words, we define ϕ in such a way that the exponential coordinate of $\boldsymbol{\pi}(\theta)$ is $\theta - \phi$. Then the graph of this map is c-cyclical monotone.

Using this result, we showed in [45] how the optimal transport problem (2.3) can be solved in terms of functionally generated portfolios. Here is a simple but interesting explicit example which is a direct generalization of the one-dimensional case treated in [45], Section 4.

EXAMPLE 2.8 (Product of Gaussian distributions).   In the transport problem (2.3), let P be a product of one-dimensional Gaussian distributions:

$$P = \bigotimes_{i=1}^{n-1} N(a_i, \sigma_i^2),$$

where $a_i \in \mathbb{R}$ and $\sigma_i > 0$. Also let

$$Q = \bigotimes_{i=1}^{n-1} N(b_i, (1-\lambda)\sigma_i^2),$$

where $b_i \in \mathbb{R}$ and $0 < \lambda < 1$. Then the optimal transport map for the measures P and Q is given by the map (2.7), where the portfolio map $\boldsymbol{\pi}$ is the following variant of the diversity-weighted portfolio discussed in Example 2.6(iii):

$$(2.8) \qquad \boldsymbol{\pi}_i(p) = \frac{w_i\, p_i^\lambda}{\sum_{j=1}^n w_j\, p_j^\lambda}, \qquad \varphi(p) = \frac{1}{\lambda} \log\left(\sum_{j=1}^n w_j\, p_j^\lambda \right),$$

where the coefficients w_i are chosen such that $(1-\lambda)a_i - \log \frac{w_i}{w_n} = b_i$ for all i.

3. Optimal transport and duality.

3.1. *c-Concavity and duality.*   Now we make use of the notion of *c*-concavity in optimal transport theory. The definitions we use are standard and can be found in [4], Chapter 1. Again, c refers to our cost function (1.7). Also recall that $\mathcal{X} = \mathbb{R}^{n-1}$ and $\mathcal{Y} = \mathbb{R}^{n-1}$ are the underlying spaces of the variables θ and ϕ, respectively. For $f : \mathcal{X} \to \mathbb{R} \cup \{\pm\infty\}$ we define its *c*-transform by

$$f^*(\phi) := \inf_{\theta \in \mathcal{X}} (c(\theta, \phi) - f(\theta)), \qquad \phi \in \mathcal{Y}.$$

Similarly, the *c*-transform of a function $g : \mathcal{Y} \to \mathbb{R} \cup \{\pm\infty\}$ is defined by

$$g^*(\theta) := \inf_{\phi \in \mathcal{Y}} (c(\theta, \phi) - g(\phi)), \qquad \theta \in \mathcal{X}.$$

We say that $f : \mathcal{X} \to \mathbb{R} \cup \{-\infty\}$ is *c*-concave if there exists $g : \mathcal{Y} \to \mathbb{R} \cup \{-\infty\}$ such that $f = g^*$ (similar for *c*-concave functions on $\mathcal{Y}$). A function h (on $\mathcal{X}$ or $\mathcal{Y}$) is *c*-concave if and only if $h^{**} = h$.

If $f : \mathcal{X} \to \mathbb{R} \cup \{-\infty\}$ is *c*-concave, its *c*-superdifferential is defined by

$$(3.1) \qquad \partial^c f := \{(\theta, \phi) \in \mathcal{X} \times \mathcal{Y} : f(\theta) + f^*(\phi) = c(\theta, \phi)\}.$$

For $\theta \in \mathcal{X}$ we define $\partial^c f(\theta) := \{\phi \in \mathcal{Y} : (\theta, \phi) \in \partial^c f\}$. If this set is a singleton $\{\phi\}$, we call ϕ the *c*-supergradient of f at θ and write $\phi = \nabla^c f(\theta)$. Similar definitions hold for a *c*-concave function g on $\mathcal{Y}$.

Let $f : \mathcal{X} \to \mathbb{R} \cup \{-\infty\}$ be *c*-concave. By definition of f^*, we have

$$(3.2) \qquad\qquad\qquad f(\theta) + f^*(\phi) \le c(\theta, \phi)$$

for every pair $(\theta, \phi) \in \mathcal{X} \times \mathcal{Y}$, and equality holds in (3.2) if and only if $(\theta, \phi) \in \partial^c f$. This is a generalized version of Fenchel's identity (see [46], Section 12) and will be used frequently in this paper.

Our first lemma relates exponential concavity on Δ_n with c-concavity on $\mathcal{X}$ and on $\mathcal{Y}$. Note that the cost function is asymmetric, and c-concavity on $\mathcal{X}$ is equivalent to c-concavity on $\mathcal{Y}$ after a change of variable.

LEMMA 3.1 (Exponential concavity and c-concavity). *For $\varphi : \Delta_n \to \mathbb{R} \cup \{-\infty\}$ the following statements are equivalent.*

  (i) *φ is exponentially concave on Δ_n.*
  (ii) *The function $f : \mathcal{X} \to \mathbb{R} \cup \{-\infty\}$ defined by*
$$f(\theta) = \varphi(\mathbf{p}(\theta)) + \psi(\theta)$$
is c-concave on $\mathcal{X}$.
  (iii) *The function $g : \mathcal{Y} \to \mathbb{R} \cup \{-\infty\}$ defined by*
$$g(\phi) = \varphi(\mathbf{p}(-\phi)) + \psi(-\phi),$$
where $-\phi$ is the exponential coordinate, is c-concave on $\mathcal{Y}$.

PROOF. We prove the implication (i) $\Rightarrow$ (ii) and the others can be proved similarly. Suppose (i) holds and consider the nonnegative concave function $\Phi = e^\varphi$ on Δ_n. By [46], Theorem 10.3, we can extend Φ continuously up to $\overline{\Delta}_n$, the closure of Δ_n in $\mathbb{R}^n$. We further extend Φ to the affine hull H of Δ_n in $\mathbb{R}^n$ by setting $\Phi(p) = -\infty$ for $p \notin \overline{\Delta}_n$. The extended function Φ is then a closed concave function on H. By convex duality (see [46], Theorem 12.1), there exists a family $\mathcal{C}$ of affine functions on H such that

$$(3.3) \qquad \Phi(p) = \inf_{\ell \in \mathcal{C}} \ell(p), \qquad p \in \Delta_n.$$

Since Φ is nonnegative on Δ_n, each $\ell \in \mathcal{C}$ is nonnegative on Δ_n. Replacing ℓ by the sequence $\ell_k = \ell + \frac{1}{k}$, $k \geq 1$, we may assume without loss of generality that each $\ell \in \mathcal{C}$ is strictly positive on Δ_n. We parameterize each $\ell \in \mathcal{C}$ in the form $\ell(p) = \sum_{i=1}^n a_i p_i$ where $a_1, \ldots, a_n$ are positive constants. (Note that an extra constant term is not required since $p_1 + \cdots + p_n = 1$.) Writing $\phi_i := -\log \frac{a_i}{a_n}$, $i = 1, \ldots, n-1$ and switching to exponential coordinates, we have

$$\log \ell(p) = \log\left(\sum_{i=1}^n a_i p_i\right)$$

$$= \log\left(1 + \sum_{i=1}^{n-1} \frac{a_i}{a_n} \frac{p_i}{p_n}\right) + \log p_n + \log a_n$$

$$= \log\left(1 + \sum_{i=1}^{n-1} e^{\theta_i - \phi_i}\right) - \psi(\theta) + \log a_n$$

$$= c(\theta - \phi) - \psi(\theta) + \log a_n.$$

It follows from (3.3) that

$$(3.4) \qquad f(\theta) = \varphi(\theta) + \psi(\theta) = \inf_{\ell \in \mathcal{C}} \big(c(\theta - \phi) + \log a_n \big).$$

Define $h : \mathcal{Y} \to \mathbb{R} \cup \{-\infty\}$ by setting

$$h(\phi) = \inf \left\{ -\log a_n : \exists \ell(p) = \sum_{i=1}^{n} a_i \, p_i \in \mathcal{C} \text{ s.t. } \phi_i = -\log \frac{a_i}{a_n} \, \forall i \right\},$$

where the infimum of the empty set is $-\infty$. From (3.4), we have

$$f(\theta) = \inf_{\phi \in \mathcal{Y}} \big(c(\theta - \phi) - h(\phi) \big) = h^*(\theta)$$

which shows that f is c-concave on $\mathcal{X}$. $\square$

The following is the c-concave analogue of the classical Legendre transformation [46]. Its proof is standard but lengthy and will be given in the Appendix.

THEOREM 3.2 (c-Legendre transformation). *Let φ be an exponentially concave function φ satisfying Assumption 2.5, and let $\boldsymbol{\pi}$, defined by (1.4), be the portfolio map generated by φ. Given φ, consider the c-concave function*

$$(3.5) \qquad f(\theta) := \varphi(\theta) + \psi(\theta)$$

defined on $\mathcal{X} = \mathbb{R}^{n-1}$ via the exponential coordinate system.

(i) *The c-supergradient of f is given by (2.7), that is,*

$$(3.6) \qquad \nabla^c f(\theta) = \left(\theta_i - \log \frac{\boldsymbol{\pi}_i(\theta)}{\boldsymbol{\pi}_n(\theta)} \right)_{1 \le i \le n-1}, \qquad \theta \in \mathcal{X}.$$

Moreover, the map $\nabla^c f : \mathcal{X} \to \mathcal{Y}$ is injective.

(ii) *Let $\mathcal{Y}' \subset \mathcal{Y}$ be the range of $\nabla^c f$. Then the c-supergradient of f^* is given on $\mathcal{Y}'$ by*

$$\nabla^c f^*(\phi) = (\nabla^c f)^{-1}(\phi), \qquad \phi \in \mathcal{Y}'.$$

In fact, the map $\nabla^c f$ is a diffeomorphism from $\mathcal{X}$ to $\mathcal{Y}'$ whose inverse is $\nabla^c f^*$. Also, the function f^* is smooth on the open set $\mathcal{Y}'$.

Although $\mathcal{Y}'$ is in general a strict subset of $\mathcal{Y}$, by Theorem 3.2 the dual variable $\phi = \nabla^c f(\theta)$ defines a global coordinate system of the manifold Δ_n. In Theorem 5.1, we will use another coordinate system on Δ_n called the dual Euclidean coordinate system. Thus, we have four coordinate systems on Δ_n: Euclidean, primal, dual and dual Euclidean (see Definition 3.3). In the following, we will frequently switch between coordinate systems to facilitate computations. To avoid confusions, let us state once for all the conventions used. We let φ and $f = \varphi + \psi$ be given.
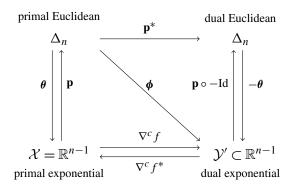
FIG. 2. *Coordinate systems on Δ_n.*

DEFINITION 3.3 (Coordinate systems). For the unit simplex Δ_n [defined by (1.1)], we call the identity map

$$p = (p_1, \ldots, p_n), \qquad p_i > 0, \qquad \sum_{i=1}^{n} p_i = 1$$

the (primal) Euclidean coordinate system with range Δ_n. We let

$$\theta = \boldsymbol{\theta}(p) = \left(\log \frac{p_1}{p_n}, \ldots, \log \frac{p_{n-1}}{p_n} \right)$$

be the primal (exponential) coordinate system with range $\mathcal{X}$ and

$$\phi = \boldsymbol{\phi}(p) := \nabla^c f(\theta)$$

be the dual (exponential) coordinate system with range $\mathcal{Y}'$. The dual Euclidean coordinate system is defined by the composition

$$p^* = \mathbf{p}^*(p) := \mathbf{p}\big(-\boldsymbol{\phi}(p)\big).$$

See Figure 2 for an illustration. From now on p, p^*, θ and ϕ always represent the same point of Δ_n. In particular, unless otherwise specified θ and ϕ are dual to each other in the sense that $\phi = \nabla^c f(\theta)$. By convention we let $\theta_n = \phi_n = 0$ for any $p \in \Delta_n$.

NOTATION 3.4 (Switching coordinate systems). We identify the spaces Δ_n, $\mathcal{X}$ and $\mathcal{Y}'$ using the coordinate systems in Definition 3.3. If h is a function on any one of these spaces, we write $h(p) = h(\theta) = h(\phi) = h(p^*)$ depending on the coordinate system used.

We also record a useful fact. A formula analogous to the first statement is derived in [49].

LEMMA 3.5.   *For $1 \le i \le n - 1$, we have*

$$\frac{\partial}{\partial \theta_i} f(\theta) = \boldsymbol{\pi}_i(\theta), \qquad \theta \in \mathcal{X},$$

$$\frac{\partial}{\partial \phi_i} f^*(\phi) = -\boldsymbol{\pi}_i(\phi), \qquad \phi \in \mathcal{Y}'.$$

PROOF.   The first statement is derived in the proof of Theorem 3.2. The second statement can be proved by differentiating $f^*(\phi) = c(\theta, \phi) - f(\theta)$ (Fenchel's identity).   □

3.2. *c-Divergence.*   By duality, we show that a pair of natural divergences on Δ_n can be defined for the c-concave functions f and f^*. Moreover, they coincide with L-divergence. Clearly we can consider other cost functions C other than c. When C is the squared Euclidean distance, the analogue of Definition 3.6 below gives the classical Bregman divergence. This covers both L-divergence and Bregman divergence under the same framework. To the best of our knowledge these definitions, which depend crucially on the interplay between transport and divergence, are new. We will use the triple representation (p, θ, ϕ) for each point in Δ_n.

DEFINITION 3.6 (*c*-divergence).   Consider the c-concave function f defined by (3.5) and its c-transform f^*.

 (i) The c-divergence of f is defined by

$$(3.7) \qquad D(p|p') = c(\theta, \phi') - c(\theta', \phi') - (f(\theta) - f(\theta')), \qquad p, p' \in \Delta_n.$$

 (ii) The c-divergence of f^* is defined by

$$(3.8) \qquad D^*(p|p') = c(\theta', \phi) - c(\theta', \phi') - (f^*(\phi) - f^*(\phi')), \qquad p, p' \in \Delta_n.$$

From Fenchel's identity (3.2) we see that D and D^* are nonnegative and nondegnereate, that is, they vanish only on the diagonal of $\Delta_n \times \Delta_n$. The following is a generalization of the self-dual expression of Bregman divergence (see [2], Theorem 1.1).

PROPOSITION 3.7 (Self-dual expressions).   *We have*

$$(3.9) \qquad\qquad D(p|p') = c(\theta, \phi') - f(\theta) - f^*(\phi'),$$

$$(3.10) \qquad\qquad D^*(p|p') = c(\theta', \phi) - f^*(\phi) - f(\theta').$$

In particular, for $p, p' \in \Delta_n$ we have $D(p|p') = D^*(p'|p)$.

PROOF. To prove (3.9), we use the Fenchel identity $f(\theta') + f^*(\phi') = c(\theta', \phi')$. Starting from (3.7), we have

$$D(p|p') = c(\theta, \phi') - c(\theta', \phi') - (f(\theta) - f(\theta'))$$
$$= c(\theta, \phi') - f(\theta) - f^*(\phi').$$

The proof of (3.10) is similar. $\square$

Now we show that L-divergence is a c-divergence where $c(\theta, \phi) = \psi(\theta - \phi)$.

THEOREM 3.8 (L-divergence as c-divergence). *The c-divergence of f is the L-divergence of φ. Namely, for $p, p' \in \Delta_n$ we have*

$$D(p|p') = T(p|p').$$

PROOF. Using the primal-dual relation (3.6), we have

$$\psi(\theta - \phi') = \log\left(\sum_{i=1}^{n} e^{\theta_i - \theta_i' + \log \frac{\pi_i(\theta')}{\pi_n(\theta')}}\right) = \log\left(\boldsymbol{\pi}(p') \cdot \frac{p}{p'}\right) - \log\left(\boldsymbol{\pi}_n(p') \frac{p_n}{p_n'}\right).$$

Next, by Fenchel's identity [see (3.2)], we have

$$f^*(\phi') = \psi(\theta' - \phi') - f(\theta') = \psi(\theta' - \phi') - \varphi(\theta') - \psi(\theta').$$

Using these identities and (2.6), we compute

$$D(p|p') = \psi(\theta - \phi') - f(\theta) - f^*(\phi')$$
$$= \log\left(\boldsymbol{\pi}(p') \cdot \frac{p}{p'}\right) - \log\left(\boldsymbol{\pi}_n(p') \frac{p_n}{p_n'}\right)$$
$$- (\varphi(\theta) + \psi(\theta)) - (\psi(\theta' - \phi') - \varphi(\theta') - \psi(\theta'))$$
$$= \log\left(\boldsymbol{\pi}(p') \cdot \frac{p}{p'}\right) - (\varphi(\theta) - \varphi(\theta'))$$
$$= T(p|p'). \qquad \square$$

For computations it is convenient to express $T(p|p')$ solely in terms of either the primal or dual coordinates. We omit the details of the computations.

LEMMA 3.9 (Coordinate representations). *For $p, p' \in \Delta_n$ we have*

$$T(p|p') = \log\left(\sum_{\ell=1}^{n} \boldsymbol{\pi}_\ell(\theta') e^{\theta_\ell - \theta_\ell'}\right) - (f(\theta) - f(\theta')),$$

$$T(p|p') = \log\left(\sum_{\ell=1}^{n} \boldsymbol{\pi}_\ell(\phi) e^{\phi_\ell - \phi_\ell'}\right) - (f^*(\phi') - f^*(\phi)).$$

3.3. *Transport interpretation of the generalized Pythagorean theorem.* Using Proposition 3.8, we give an interesting transport interpretation of the expression (1.8) in the generalized Pythagorean theorem (Theorem 1.2). Let $p, q, r \in \Delta_n$ be given. Let $(\theta^{(j)}, \phi^{(j)})_{1 \leq j \leq 3}$ be the primal and dual coordinates of p, q and r respectively. By Proposition 2.7, the coupling $(\theta, \phi = \nabla f^c(\theta))$ is c-cyclical monotone. Hence, coupling $\theta^{(j)}$ with $\phi^{(j)}$ is optimal.

Consider two (suboptimal) perturbations of the optimal coupling:

(i) (Cyclical perturbation) Couple $\theta^{(1)}$ with $\phi^{(3)}$, $\theta^{(2)}$ with $\phi^{(1)}$, and $\theta^{(3)}$ with $\phi^{(2)}$. The associated cost is

$$c(\theta^{(1)}, \phi^{(3)}) + c(\theta^{(2)}, \phi^{(1)}) + c(\theta^{(3)}, \phi^{(2)}).$$

(ii) (Transposition) Couple $\theta^{(1)}$ with $\phi^{(3)}$, $\theta^{(3)}$ with $\phi^{(1)}$, and keep the coupling $(\theta^{(2)}, \phi^{(2)})$. The associated cost is

$$c(\theta^{(1)}, \phi^{(3)}) + c(\theta^{(3)}, \phi^{(1)}) + c(\theta^{(2)}, \phi^{(2)}).$$

Now we ask which perturbation has lower cost. The difference (i)–(ii) is

$$c(\theta^{(2)}, \phi^{(1)}) + c(\theta^{(3)}, \phi^{(2)}) - c(\theta^{(3)}, \phi^{(1)}) - c(\theta^{(2)}, \phi^{(2)}).$$

By Proposition 3.8, this is nothing but the difference $T(q|p) + T(r|q) - T(r|p)$. Thus the generalized Pythagorean theorem gives an information geometric characterization of the relative costs of the two perturbations.

3.4. *Examples.* We consider the portfolios in Example 2.6.

EXAMPLE 3.10 (Constant-weighted portfolio). Let $\pi \in \Delta_n$ be a constant-weighted portfolio. Then φ is the cross entropy and we have

$$f(\theta) = \varphi(\theta) + \psi(\theta) = \sum_{i=1}^{n-1} \pi_i \theta_i,$$

which is an affine function on $\mathcal{X}$. Its c-transform is also affine. Indeed, we have

$$f^*(\phi) = \sum_{i=1}^{n-1} \pi_i(-\phi_i) + H(\pi),$$

where $H(\pi) := -\sum_{i=1}^{n} \pi_i \log \pi_i$ is the Shannon entropy of π. For this reason, we say that the constant-weighted portfolios are self-dual. The transport map in this case is given by a translation: $\phi = \theta - (\log \frac{\pi_1}{\pi_n}, \ldots, \log \frac{\pi_{n-1}}{\pi_n})$. Its L-divergence (1.6) is given in primal coordinates (see Lemma 3.9) by

$$T(p|p') = \log \left(\sum_{\ell=1}^{n} \pi_\ell e^{\theta_\ell - \theta_\ell'} \right) - \sum_{\ell=1}^{n} \pi_\ell(\theta_\ell - \theta_\ell'),$$

which is translation invariant. This property is equivalent to the following numéraire invariance property [44], Lemma 3.2: for any $w_1, \ldots, w_n > 0$, we have $T(q|p) = T(\widetilde{q}|\widetilde{p})$ under the mapping

$$p \mapsto \widetilde{p} = \left(\frac{w_i p_i}{w_1 p_1 + \cdots + w_n p_n} \right)_{1 \leq i \leq n}.$$

In fact, it is not difficult to show that this property characterizes the constant-weighted portfolios among L-divergences of exponentially concave functions. Also see [44], Proposition 4.6, for a chain rule analogous to that of relative entropy.

EXAMPLE 3.11 (Diversity-weighted portfolio). We have $f(\theta) = \frac{1}{\lambda} \psi(\lambda \theta)$. Since

$$\log \frac{\pi_i(\theta)}{\pi_n(\theta)} = \lambda \theta_i,$$

the map $\theta \mapsto \phi$ is a scaling: $\phi = (1 - \lambda)\theta$. For the generalized diversity-weighted portfolio in Example 2.8, the transport map $\theta \mapsto \phi$ is the composition of a scaling and a translation.

4. Geometric structure induced by L-divergence. In this section, we derive the geometric structure induced by a given L-divergence $T(\cdot|\cdot)$. As always we impose the regularity conditions in Assumption 2.5. Using the primal and dual coordinate systems (Definition 3.3), we compute explicitly the Riemannian metric g, the primal connection ∇ (not to be confused with the Euclidean gradient) and the dual connection ∇^*. We call (g, ∇, ∇^*) the induced geometric structure. An important fact in information geometry is that the Levi-Civita connection $\nabla^{(0)}$ is not necessarily the right one to use. Nevertheless, by duality we always have $\nabla^{(0)} = \frac{1}{2}(\nabla + \nabla^*)$.

4.1. *Preliminaries.* For differential geometric concepts such as Riemannian metric and affine connection we refer the reader to [2], Chapters 5, whose notations are consistent with ours. For computational convenience, we define the geometric structure in terms of coordinate representations. The geometric structure is determined by the L-divergence $T(\cdot|\cdot)$ and is independent of the choice of coordinates; for intrinsic formulations we refer the reader to [13], Chapter 11. The following definition (which makes sense for a general divergence on a manifold) is taken from [2], Section 6.2.

DEFINITION 4.1 (Induced geometric structure). Given a coordinate system $\xi = (\xi_1, \ldots, \xi_{n-1})$ of Δ_n, the coefficients of the geometric structure (g, ∇, ∇^*) are given as follows.

(i) The Riemannian metric is given by

$$(4.1) \qquad g_{ij}(\xi) = -\frac{\partial}{\partial \xi_i} \frac{\partial}{\partial \xi_j'} T(\xi|\xi')\Big|_{\xi=\xi'}, \qquad i,j = 1, \ldots, n-1.$$

By Assumption 2.5, the matrix $(g_{ij}(\xi))$ is strictly positive definite. The Riemannian inner product and length are denoted by $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$ respectively.

(ii) The primal connection is given by

$$(4.2) \qquad \Gamma_{ijk}(\xi) = -\frac{\partial}{\partial \xi_i} \frac{\partial}{\partial \xi_j} \frac{\partial}{\partial \xi_k'} T(\xi|\xi')\Big|_{\xi=\xi'}, \qquad i,j,k = 1, \ldots, n-1.$$

(iii) The dual connection is given by

$$(4.3) \qquad \Gamma_{ijk}^*(\xi) = -\frac{\partial}{\partial \xi_k} \frac{\partial}{\partial \xi_i'} \frac{\partial}{\partial \xi_j'} T(\xi|\xi')\Big|_{\xi=\xi'}, \qquad i,j,k = 1, \ldots, n-1.$$

For a general divergence the above definitions were first introduced in [19, 20]. If we define the dual divergence by $T^*(p|p') := T(p'|p)$, the dual connection of T is the primal connection of T^*. The primal and dual connections are dual to each other with respect to the Riemannian metric g (see [2], Theorem 6.2). While any divergence induces a geometric structure, it may not enjoy nice properties. For the geometric structure induced by a Bregman divergence, it can be shown that the Riemann–Christoffel curvatures of the primal and dual connections both vanish. Thus, we say that the induced geometry is dually flat [2], Chapter 1. We will show that L-divergence gives rise to a different geometry with many interesting properties.

4.2. *Notations.* We begin by clarifying the notations. Following our convention (see Notation 3.4), we write $T(p|p') = T(\theta|\theta') = T(\phi|\phi')$ depending on the coordinate system used. The primal and dual coordinate representations have been computed in Lemma 3.9.

The Riemannian metric will be computed using both the primal and dual coordinate systems. To be explicit about the coordinate system, we use $g_{ij}(\theta)$ to denote its coefficients in primal coordinates, and $g_{ij}^*(\phi)$ for its coefficients in dual coordinates:

$$g_{ij}(\theta) := -\frac{\partial^2}{\partial \theta_i \partial \theta_j'} T(\theta|\theta')\Big|_{\theta=\theta'}, \qquad g_{ij}^*(\phi) := -\frac{\partial^2}{\partial \phi_i \partial \phi_j'} T(\phi|\phi')\Big|_{\phi=\phi'}.$$

The inverses of the matrices $(g_{ij}(\theta))$ and $(g_{ij}^*(\phi))$ are denoted by $(g^{ij}(\theta))$ and $(g^{*ij}(\phi))$ respectively.

The primal connection ∇ will be computed using the primal coordinate system:

$$\Gamma_{ijk}(\theta) := -\frac{\partial^3}{\partial \theta_i \partial \theta_j \partial \theta_k'} T(\theta|\theta')\Big|_{\theta=\theta'}, \qquad \Gamma_{ij}^k(\theta) := \sum_{m=1}^{n-1} \Gamma_{ijm}(\theta) g^{mk}(\theta).$$

The dual connection ∇^* will be computed using the dual coordinate system:

$$\Gamma_{ijk}^*(\phi) := -\frac{\partial^3}{\partial \phi_k \partial \phi_i' \partial \phi_j'} T(\phi|\phi')\Big|_{\phi=\phi'}, \qquad \Gamma_{ij}^{*k}(\phi) := \sum_{m=1}^{n-1} \Gamma_{ijm}^*(\phi) g^{*mk}(\phi).$$

The following notations are useful. For $1 \leq i \leq n$, we define

(4.4)
$$\Pi_i(\theta, \theta') := \frac{\boldsymbol{\pi}_i(\theta') e^{\theta_i - \theta_i'}}{\sum_{\ell=1}^n \boldsymbol{\pi}_\ell(\theta') e^{\theta_\ell - \theta_\ell'}},$$

$$\Pi_i^*(\phi, \phi') := \frac{\boldsymbol{\pi}_i(\phi) e^{\phi_i - \phi_i'}}{\sum_{\ell=1}^n \boldsymbol{\pi}_\ell(\phi) e^{\phi_\ell - \phi_\ell'}}.$$

As always we adopt the convention $\theta_n = \theta_n' = \phi_n = \phi_n' = 0$. Note that $\Pi_i(\theta, \theta')$ involves the portfolio at θ' (the second variable) while $\Pi_i^*(\phi, \phi')$ involves the portfolio at ϕ (the first variable). The partial derivatives of Π_i and Π_i^* are given in the next lemma and can be verified by direct differentiation. We let δ_{ij} be the Kronecker delta and $\delta_{ijk} = \delta_{ij}\delta_{jk}$.

LEMMA 4.2 (Derivatives of Π_i and Π_i^*).   (i) *For $1 \leq i \leq n$ and $1 \leq j \leq n-1$, we have*

$$\frac{\partial \Pi_i(\theta, \theta')}{\partial \theta_j} = \Pi_i(\theta, \theta')(\delta_{ij} - \Pi_j(\theta, \theta')),$$

$$\frac{\partial \Pi_i(\theta, \theta')}{\partial \theta_j'} = -\Pi_i(\theta, \theta')(\delta_{ij} - \Pi_j(\theta, \theta'))$$

$$+ \Pi_i(\theta, \theta')\left(\frac{1}{\boldsymbol{\pi}_i(\theta')} \frac{\partial \boldsymbol{\pi}_i}{\partial \theta_j'}(\theta') - \sum_{\ell=1}^n \Pi_\ell(\theta, \theta') \frac{1}{\boldsymbol{\pi}_\ell(\theta')} \frac{\partial \boldsymbol{\pi}_\ell}{\partial \theta_j'}(\theta')\right).$$

(ii) *For $1 \leq i \leq n$ and $1 \leq j \leq n-1$, we have*

$$\frac{\partial \Pi_i^*(\phi, \phi')}{\partial \phi_j} = \Pi_i^*(\phi, \phi')(\delta_{ij} - \Pi_j^*(\phi, \phi'))$$

$$+ \Pi_i^*(\phi, \phi')\left(\frac{1}{\boldsymbol{\pi}_i(\phi)} \frac{\partial \boldsymbol{\pi}_i}{\partial \phi_j}(\phi) - \sum_{\ell=1}^n \Pi_\ell^*(\phi, \phi') \frac{1}{\boldsymbol{\pi}_\ell(\phi)} \frac{\partial \boldsymbol{\pi}_\ell}{\partial \phi_j}(\phi)\right),$$

$$\frac{\partial \Pi_i(\phi, \phi')}{\partial \phi_j'} = -\Pi_i^*(\phi, \phi')(\delta_{ij} - \Pi_j^*(\phi, \phi')).$$

We also note the following easy fact which is used several times below. Since $\sum_{i=1}^n \Pi_i(\theta, \theta') \equiv \sum_{i=1}^n \Pi_i^*(\phi, \phi') \equiv 1$, the partial derivatives of $\sum_i \Pi_i$ and $\sum_i \Pi_i^*$ are all zero.

LEMMA 4.3 [Derivatives of $\boldsymbol{\pi}(\cdot)$].   *For $1 \le i \le n$ and $1 \le j \le n-1$, we have*

$$\frac{\partial \boldsymbol{\pi}_i}{\partial \theta_j}(\theta) = \boldsymbol{\pi}_i(\theta)\big(\delta_{ij} - \boldsymbol{\pi}_j(\theta)\big) - \boldsymbol{\pi}_i(\theta)\bigg(\frac{\partial \phi_i}{\partial \theta_j}(\theta) - \sum_{\ell=1}^{n-1} \boldsymbol{\pi}_\ell(\theta)\frac{\partial \phi_\ell}{\partial \theta_j}(\theta)\bigg),$$

(4.5)

$$\frac{\partial \boldsymbol{\pi}_i}{\partial \phi_j}(\phi) = -\boldsymbol{\pi}_i(\phi)\big(\delta_{ij} - \boldsymbol{\pi}_j(\phi)\big) + \boldsymbol{\pi}_i(\phi)\bigg(\frac{\partial \theta_i}{\partial \phi_j}(\phi) - \sum_{\ell=1}^{n-1} \boldsymbol{\pi}_\ell(\phi)\frac{\partial \theta_\ell}{\partial \phi_j}(\phi)\bigg).$$

PROOF.   We prove the second formula and the proof of the first is similar. Using (2.7), we write

$$\boldsymbol{\pi}_i(\phi) = \frac{e^{\theta_i - \phi_i}}{\sum_{\ell=1}^{n} e^{\theta_\ell - \phi_\ell}}$$

and regard θ as a function of ϕ (recall that $\theta_n = \phi_n = 0$). Then

$$\frac{\partial \boldsymbol{\pi}_i}{\partial \phi_j}(\phi) = \frac{e^{\theta_i - \phi_i}\big(\frac{\partial \theta_i}{\partial \phi_j}(\phi) - \delta_{ij}\big)}{\sum_{\ell=1}^{n} e^{\theta_\ell - \phi_\ell}} - \frac{e^{\theta_i - \phi_i}}{(\sum_{\ell=1}^{n} e^{\theta_\ell - \phi_\ell})^2}\sum_{\ell=1}^{n} e^{\theta_\ell - \phi_\ell}\bigg(\frac{\partial \theta_\ell}{\partial \phi_j}(\phi) - \delta_{\ell j}\bigg)$$

$$= -\boldsymbol{\pi}_i(\phi)\big(\delta_{ij} - \boldsymbol{\pi}_j(\phi)\big) + \boldsymbol{\pi}_i(\phi)\bigg(\frac{\partial \theta_i}{\partial \phi_j}(\phi) - \sum_{\ell=1}^{n-1} \boldsymbol{\pi}_\ell(\phi)\frac{\partial \theta_\ell}{\partial \phi_j}(\phi)\bigg).$$

Note that the nth term of the sum is omitted because $\theta_n = 0$.   □

Thanks to these formulas, computations in the primal and dual coordinates are very similar except for a change of sign. In the following, we will often give details for one coordinate system and leave the other one to the reader.

Last but not least, let $\frac{\partial \phi}{\partial \theta}(\theta) = (\frac{\partial \phi_i}{\partial \theta_j}(\theta))$ be the Jacobian of the change of coordinate map $\theta \mapsto \phi$. Similarly, we let $\frac{\partial \theta}{\partial \phi}(\phi) = (\frac{\partial \theta_i}{\partial \phi_j}(\phi))$ be the Jacobian of the inverse map $\phi \mapsto \theta$. The two Jacobians are inverses of each other, that is,

(4.6)
$$\frac{\partial \phi}{\partial \theta}(\theta)\frac{\partial \theta}{\partial \phi}(\phi) = I,$$

where I is the identity matrix. We denote the transpose of a matrix A by $A^\top$.

4.3. *Riemannian metric.*   For intuition, we first compute the Riemannian inner product using Euclidean coordinates. We let $T_p \Delta_n$ be the tangent space at p.

PROPOSITION 4.4.   *Let $u, v \in T_p \Delta_n$ be represented in Euclidean coordinates, that is, $u = (u_1, \dots, u_n) \in \mathbb{R}^n$ and $u_1 + \cdots + u_n = 0$, and similarly for v. Then*

$$\langle u, v \rangle = u^\top\big(-\operatorname{Hess}\varphi(p) - \nabla\varphi(p)\nabla\varphi(p)^\top\big)v$$

(4.7)

$$= u^\top\bigg(\frac{-1}{\Phi(p)}\operatorname{Hess}\Phi(p)\bigg)v,$$

where $\Phi = e^\varphi$.

PROOF. By [13], Proposition 11.3.1, we have $\|v\|^2 = \frac{d^2}{dt^2}T(p+tv|p)|_{t=0}$, where

$$T(p+tv|p) = \log(1 + t\nabla\varphi(p)\cdot v) - (\varphi(p+tv) - \varphi(p)).$$

Differentiating two times and setting $t = 0$ gives the first equality in (4.7) when $u = v$, and polarizing gives the general case. The second equality follows from the chain rule. □

THEOREM 4.5 (Riemannian metric). (i) *Under the primal coordinate system, the Riemannian metric is given by*

$$(4.8) \qquad g_{ij}(\theta) = \boldsymbol{\pi}_i(\theta)\big(\delta_{ij} - \boldsymbol{\pi}_j(\theta)\big) - \frac{\partial\boldsymbol{\pi}_i}{\partial\theta_j}(\theta).$$

Its inverse is given by

$$(4.9) \qquad g^{ij}(\theta) = \frac{1}{\boldsymbol{\pi}_j(\theta)}\frac{\partial\theta_i}{\partial\phi_j}(\phi) + \frac{1}{\boldsymbol{\pi}_n(\theta)}\sum_{\ell=1}^{n-1}\frac{\partial\theta_i}{\partial\phi_\ell}(\phi).$$

(ii) *Under the dual coordinate system, the Riemannian metric is given by*

$$(4.10) \qquad g^*_{ij}(\phi) = \boldsymbol{\pi}_i(\phi)\big(\delta_{ij} - \boldsymbol{\pi}_j(\phi)\big) + \frac{\partial\boldsymbol{\pi}_i}{\partial\phi_j}(\phi).$$

Its inverse is given by

$$(4.11) \qquad g^{*ij}(\phi) = \frac{1}{\boldsymbol{\pi}_j(\phi)}\frac{\partial\phi_i}{\partial\theta_j}(\theta) + \frac{1}{\boldsymbol{\pi}_n(\phi)}\sum_{\ell=1}^{n-1}\frac{\partial\phi_i}{\partial\theta_\ell}(\theta).$$

PROOF. (i) By Lemma 3.9 and Lemma 4.2, we compute

$$(4.12) \quad \frac{\partial}{\partial\theta_i}T(\theta|\theta') = \Pi_i(\theta,\theta') - \boldsymbol{\pi}_i(\theta),$$

$$(4.13) \quad \frac{\partial}{\partial\theta'_i}T(\theta|\theta') = -\Pi_i(\theta,\theta') + \boldsymbol{\pi}_i(\theta') + \sum_{\ell=1}^{n}\Pi_\ell(\theta,\theta')\frac{1}{\boldsymbol{\pi}_\ell(\theta')}\frac{\partial\boldsymbol{\pi}_\ell}{\partial\theta'_i}(\theta').$$

Differentiating (4.12) with respect to θ'_j, we have

$$\frac{\partial^2}{\partial\theta_i\,\partial\theta'_j}T(\theta|\theta')$$

$$(4.14) \qquad = -\Pi_i(\theta,\theta')\big(\delta_{ij} - \Pi_j(\theta,\theta')\big)$$

$$+ \Pi_i(\theta,\theta')\left(\frac{1}{\boldsymbol{\pi}_i(\theta')}\frac{\partial\boldsymbol{\pi}_i}{\partial\theta'_j}(\theta') - \sum_{\ell=1}^{n}\Pi_\ell(\theta,\theta')\frac{1}{\boldsymbol{\pi}_\ell(\theta')}\frac{\partial\boldsymbol{\pi}_\ell}{\partial\theta'_j}(\theta')\right).$$

Setting $\theta = \theta'$, we get $g_{ij}(\theta) = \boldsymbol{\pi}_i(\theta)(\delta_{ij} - \boldsymbol{\pi}_j(\theta)) - \frac{\partial \boldsymbol{\pi}_i}{\partial \theta_j}(\theta)$.

By Lemma 4.3, we have the alternative expression

$$(4.15) \qquad g_{ij}(\theta) = \boldsymbol{\pi}_i(\theta)\left(\frac{\partial \phi_i}{\partial \theta_j}(\theta) - \sum_{\ell=1}^{n-1} \boldsymbol{\pi}_\ell(\theta)\frac{\partial \phi_\ell}{\partial \theta_j}(\theta)\right).$$

Expressing (4.15) in matrix form, we have

$$(4.16) \qquad (g_{ij}(\theta)) = \mathrm{diag}(\boldsymbol{\pi}(\theta))(I - \mathbf{1}\boldsymbol{\pi}^\top(\theta))\frac{\partial \phi}{\partial \theta}(\theta),$$

where here $\boldsymbol{\pi}(\theta) = (\boldsymbol{\pi}_1(\theta), \ldots, \boldsymbol{\pi}_{n-1}(\theta))^\top$, $\mathbf{1} = \mathbf{1}_{n-1} = (1, \ldots, 1)^\top$ and $I = I_{n-1}$ is the identity matrix.

To invert (4.16) we use the fact that

$$(I - \mathbf{1}\boldsymbol{\pi}^\top(\theta))^{-1} = I + \frac{\mathbf{1}\boldsymbol{\pi}^\top(\theta)}{\boldsymbol{\pi}_n(\theta)}.$$

This can be verified directly or seen as a special case of the Sherman–Morrison formula. Using (4.6), we have

$$(g^{ij}(\theta)) = \frac{\partial \theta}{\partial \phi}(\phi)\left(I + \frac{\mathbf{1}\boldsymbol{\pi}^\top(\theta)}{\boldsymbol{\pi}_n(\theta)}\right)\mathrm{diag}\left(\frac{1}{\boldsymbol{\pi}(\theta)}\right).$$

Now (4.9) follows by expanding the matrix product.

(ii) The proofs of (4.10) and (4.11) follow the same lines. For later use, we record the following formulas:

$$(4.17) \qquad \frac{\partial}{\partial \phi_i}T(\phi|\phi') = \Pi_i^*(\phi, \phi') - \boldsymbol{\pi}_i(\phi) + \sum_{\ell=1}^{n} \frac{1}{\boldsymbol{\pi}_\ell(\phi)}\frac{\partial \boldsymbol{\pi}_\ell}{\partial \phi_i}(\phi)\Pi_\ell^*(\phi, \phi'),$$

$$(4.18) \qquad \frac{\partial}{\partial \phi_i'}T(\phi|\phi') = -\Pi_i^*(\phi, \phi') + \boldsymbol{\pi}_i(\phi'),$$

$$\frac{\partial^2}{\partial \phi_i\, \partial \phi_j'}T(\phi|\phi') = -\Pi_j^*(\phi, \phi')(\delta_{ij} - \Pi_i^*(\phi, \phi'))$$

$$(4.19) \qquad\qquad\qquad - \Pi_j^*(\phi, \phi')\left(\frac{1}{\boldsymbol{\pi}_j(\phi)}\frac{\partial \boldsymbol{\pi}_j}{\partial \phi_i}(\phi)\right.$$

$$\qquad\qquad\qquad\qquad \left. - \sum_{\ell=1}^{n}\frac{1}{\boldsymbol{\pi}_\ell(\phi)}\frac{\partial \boldsymbol{\pi}_\ell}{\partial \phi_i}(\phi)\Pi_\ell^*(\phi, \phi')\right). \qquad \square$$

REMARK 4.6.   By Lemma 3.5, we have

$$(4.20) \qquad\qquad \frac{\partial \boldsymbol{\pi}_i}{\partial \theta_j}(\theta) = \frac{\partial^2}{\partial \theta_i\, \partial \theta_j}f(\theta) = \frac{\partial \boldsymbol{\pi}_j}{\partial \theta_i}(\theta).$$

Thus the right-hand side of (4.8) is symmetric in i and j despite its appearance. Similarly, we have $\frac{\partial \pi_i}{\partial \phi_j} = \frac{\partial \pi_j}{\partial \phi_i}$.

4.4. *Primal and dual connections.*

THEOREM 4.7 (Primal and dual connections). (i) *Under the primal coordinate system, the coefficients of the primal connection ∇ is given by*

$$\Gamma_{ijk}(\theta) = \delta_{ij} g_{ik}(\theta) - \pi_i(\theta) g_{jk}(\theta) - \pi_j(\theta) g_{ik}(\theta), \tag{4.21}$$

$$\Gamma_{ij}^k(\theta) = \delta_{ijk} - \delta_{ik} \pi_j(\theta) - \delta_{jk} \pi_i(\theta). \tag{4.22}$$

(ii) *Under the dual coordinate system, the coefficients of the dual connection ∇^* is given by*

$$\Gamma_{ijk}^*(\phi) = -\delta_{ij} g_{ik}^*(\phi) + \pi_i(\phi) g_{jk}^*(\phi) + \pi_j(\phi) g_{ik}^*(\phi), \tag{4.23}$$

$$\Gamma_{ij}^{*k}(\phi) = -\delta_{ijk} + \delta_{ik} \pi_j(\phi) + \delta_{jk} \pi_i(\phi). \tag{4.24}$$

PROOF. We prove (ii) and leave (i) to the reader. By (4.19), we have

$$\frac{\partial^2}{\partial \phi_k \, \partial \phi_i'} T(\phi|\phi') = -\Pi_i^*(\phi, \phi')(\delta_{ik} - \Pi_k^*(\phi, \phi'))$$

$$- \Pi_i^*(\phi, \phi') \left(\frac{1}{\pi_i(\phi)} \frac{\partial \pi_i}{\partial \phi_k}(\phi) - \sum_{\ell=1}^n \frac{1}{\pi_\ell(\phi)} \frac{\partial \pi_\ell}{\partial \phi_k}(\phi) \Pi_\ell(\phi, \phi') \right).$$

For notational convenience we momentarily suppress ϕ and ϕ' in the computation (later we will do so without comment). Differentiating one more time, we have

$$\frac{\partial^3}{\partial \phi_k \, \partial \phi_i' \, \partial \phi_j'} T(\phi|\phi')$$

$$= -\delta_{ik} \frac{\partial \Pi_i^*}{\partial \phi_j'} + \Pi_i^* \frac{\partial \Pi_k^*}{\partial \phi_j'} + \Pi_k^* \frac{\partial \Pi_i^*}{\partial \phi_j'}$$

$$- \frac{\partial \Pi_i^*}{\partial \phi_j'} \left(\frac{1}{\pi_i} \frac{\partial \pi_i}{\partial \phi_k} - \sum_{\ell=1}^n \frac{1}{\pi_\ell} \frac{\partial \pi_\ell}{\partial \phi_k} \Pi_\ell^* \right) + \Pi_i \sum_{\ell=1}^n \frac{1}{\pi_\ell} \frac{\partial \pi_\ell}{\partial \phi_k} \frac{\partial \Pi_\ell^*}{\partial \phi_j'}$$

$$= \delta_{ik} \Pi_i^* (\delta_{ij} - \Pi_j^*) - \Pi_i^* \Pi_k^* (\delta_{jk} - \Pi_j^*) - \Pi_k^* \Pi_i^* (\delta_{ij} - \Pi_j^*)$$

$$+ \Pi_i^* (\delta_{ij} - \Pi_j^*) \left(\frac{1}{\pi_i} \frac{\partial \pi_i}{\partial \phi_k} - \sum_{\ell=1}^n \frac{1}{\pi_\ell} \frac{\partial \pi_\ell}{\partial \phi_k} \Pi_\ell^* \right)$$

$$- \Pi_i^* \sum_{\ell=1}^n \frac{1}{\pi_\ell} \frac{\partial \pi_\ell}{\partial \phi_k} \Pi_\ell^* (\delta_{\ell j} - \Pi_j^*).$$

Evaluating at $\phi = \phi'$ and simplifying, we get

$$
\begin{aligned}
\Gamma^*_{ijk}(\phi) = {}& -\delta_{ijk}\boldsymbol{\pi}_i - 2\boldsymbol{\pi}_i\boldsymbol{\pi}_j\boldsymbol{\pi}_k + \delta_{ij}\boldsymbol{\pi}_i\boldsymbol{\pi}_k + \delta_{jk}\boldsymbol{\pi}_j\boldsymbol{\pi}_i + \delta_{ki}\boldsymbol{\pi}_k\boldsymbol{\pi}_j \\
& - \delta_{ij}\frac{\partial \boldsymbol{\pi}_i}{\partial \phi_k} + \boldsymbol{\pi}_j\frac{\partial \boldsymbol{\pi}_i}{\partial \phi_k} + \boldsymbol{\pi}_i\frac{\partial \boldsymbol{\pi}_j}{\partial \phi_k}.
\end{aligned}
$$
(4.25)

By (4.10), we have $\frac{\partial \boldsymbol{\pi}_i}{\partial \phi_j} = g^*_{ij} - \boldsymbol{\pi}_i(\delta_{ij} - \boldsymbol{\pi}_j)$. Plugging this into (4.25) and simplifying, we have $\Gamma^*_{ijk}(\phi) = -\delta_{ij}g^*_{ik}(\phi) + \boldsymbol{\pi}_i(\phi)g^*_{jk}(\phi) + \boldsymbol{\pi}_j(\phi)g^*_{ik}(\phi)$. Finally, we compute

$$
\begin{aligned}
\Gamma^{*k}_{ij}(\phi) &= \sum_{m=0}^{n-1}\left(-\delta_{ij}g^*_{im}(\phi) + \boldsymbol{\pi}_i(\phi)g^*_{jm}(\phi) + \boldsymbol{\pi}_j(\phi)g^*_{im}(\phi)\right)g^{*mk}(\phi) \\
&= -\delta_{ijk} + \delta_{ik}\boldsymbol{\pi}_j(\phi) + \delta_{jk}\boldsymbol{\pi}_i(\phi). \qquad\qquad \square
\end{aligned}
$$

REMARK 4.8.   It is interesting to note that although the connections are defined in terms of the third order derivatives of $T(\cdot|\cdot)$, the coefficients $\Gamma^k_{ij}(\theta)$ and $\Gamma^{*k}_{ij}(\phi)$ are given in terms of the portfolio $\boldsymbol{\pi}$ which is a normalized gradient of φ.

4.5. *Curvatures.*   It is well known (see [2], Chapter 1) that the induced geometric structure of any Bregman divergence is dually flat. This is not the case for the geometry of L-divergence whenever $n \geq 3$ (when $n = 2$ the simplex Δ_2 is one-dimensional). To verify this, we compute the Riemann–Christoffel curvature tensors of the primal and dual connections. In this (and only this) subsection, we adopt the Einstein summation notation (see [2], p. 20) to avoid writing a lot of summation signs.

The Riemann–Christoffel (RC) curvature tensor of a connection ∇ is defined for smooth vector fields X, Y and Z by

$$
R(X,Y)Z = \nabla_X(\nabla_Y Z) - \nabla_Y(\nabla_X Z) - \nabla_{[X,Y]}Z,
$$

where $[X, Y]$ is the Lie bracket. Given a coordinate system ξ, its coefficients are given by R^ℓ_{ijk} which satisfy $R(\frac{\partial}{\partial \xi_i}, \frac{\partial}{\partial \xi_j})\frac{\partial}{\partial \xi_k} = \sum_\ell R^\ell_{ijk}\frac{\partial}{\partial \xi_\ell}$. By [2], (5.66), we have

$$
R^\ell_{ijk} = \frac{\partial}{\partial \theta_i}\Gamma^\ell_{jk} - \frac{\partial}{\partial \theta_j}\Gamma^\ell_{ik} + \Gamma^\ell_{im}\Gamma^m_{jk} - \Gamma^\ell_{jm}\Gamma^m_{ik}.
$$

THEOREM 4.9 (Primal and dual Riemann–Christoffel curvatures).   *Let R and R^* be the RC curvature tensors of the primal and dual connections respectively.*

   (i) *In primal coordinates, the coefficients of R are given by*

$$
R^\ell_{ijk}(\theta) = \delta_{\ell j}g_{ik}(\theta) - \delta_{\ell i}g_{jk}(\theta).
$$
(4.26)

   (ii) *In dual coordinates, the coefficients of R^* are given by*

$$
R^{*\ell}_{ijk}(\phi) = \delta_{\ell j}g^*_{ik}(\phi) - \delta_{\ell i}g^*_{jk}(\phi).
$$
(4.27)

In particular, for $n \geq 3$ both R and R^* are nonzero everywhere on Δ_n.

PROOF. We prove the statements for R. Using (4.21) and suppressing the argument, we have

$$\frac{\partial}{\partial \theta_i} \Gamma_{jk}^{\ell} = -\delta_{\ell j} \frac{\partial \boldsymbol{\pi}_k}{\partial \theta_i} - \delta_{\ell k} \frac{\partial \boldsymbol{\pi}_j}{\partial \theta_i}, \qquad \frac{\partial}{\partial \theta_j} \Gamma_{ik}^{\ell} = -\delta_{\ell i} \frac{\partial \boldsymbol{\pi}_k}{\partial \theta_j} - \delta_{\ell k} \frac{\partial \boldsymbol{\pi}_i}{\partial \theta_j}.$$

From (4.20), it follows that

$$(4.28) \qquad \frac{\partial}{\partial \theta_i} \Gamma_{jk}^{\ell} - \frac{\partial}{\partial \theta_j} \Gamma_{ik}^{\ell} = -\delta_{\ell j} \frac{\partial \boldsymbol{\pi}_k}{\partial \theta_i} + \delta_{\ell i} \frac{\partial \boldsymbol{\pi}_k}{\partial \theta_j}.$$

Next, we compute (with some work)

$$(4.29) \qquad \Gamma_{im}^{\ell} \Gamma_{jk}^{m} - \Gamma_{jm}^{\ell} \Gamma_{ik}^{m} = -\delta_{\ell i} \delta_{jk} \boldsymbol{\pi}_j + \delta_{\ell j} \delta_{ik} \boldsymbol{\pi}_i + \delta_{\ell i} \boldsymbol{\pi}_j \boldsymbol{\pi}_k - \delta_{\ell j} \boldsymbol{\pi}_i \boldsymbol{\pi}_k.$$

Combining (4.28) and (4.29), we have

$$R_{ijk}^{\ell}(\theta) = -\delta_{\ell j} \frac{\partial \boldsymbol{\pi}_k}{\partial \theta_i} + \delta_{\ell i} \frac{\partial \boldsymbol{\pi}_k}{\partial \theta_j} - \delta_{\ell i} \delta_{jk} \boldsymbol{\pi}_j + \delta_{\ell j} \delta_{ik} \boldsymbol{\pi}_i + \delta_{\ell i} \boldsymbol{\pi}_j \boldsymbol{\pi}_k - \delta_{\ell j} \boldsymbol{\pi}_i \boldsymbol{\pi}_k$$

$$= \delta_{\ell j} \left(\delta_{ik} \boldsymbol{\pi}_i - \boldsymbol{\pi}_i \boldsymbol{\pi}_k - \frac{\partial \boldsymbol{\pi}_k}{\partial \theta_i} \right) - \delta_{\ell i} \left(\delta_{jk} \boldsymbol{\pi}_j - \boldsymbol{\pi}_j \boldsymbol{\pi}_k - \frac{\partial \boldsymbol{\pi}_k}{\partial \theta_j} \right)$$

$$= \delta_{\ell j} g_{ik} - \delta_{\ell i} g_{jk}.$$

To see that R does not vanish for $n \geq 3$, suppose on the contrary that $R(\theta) = 0$. Then $R_{ijk}^{\ell}(\theta) = \delta_{\ell j} g_{ik}(\theta) - \delta_{\ell i} g_{jk}(\theta) = 0$ for all values of i, j, k, ℓ. Fix i and k. Letting $\ell = j$, we have $g_{ik}(\theta) = \delta_{ij} g_{jk}(\theta)$. Next, let $j \neq i$ (here we need dim $\Delta_n = n - 1 \geq 2$). Then we get $g_{ik}(\theta) = 0$. Since i and k are arbitrary, we have $g(\theta) = 0$ which is a contradiction. $\square$

We end this section by showing that the primal and dual connections have constant sectional curvature -1. See [13], Definition 7.10.5, for the definition of Ricci curvature.

COROLLARY 4.10 (Primal and dual sectional curvatures). *The primal and dual connections have constant sectional curvature -1 with respect to g. In particular, the primal and dual Ricci curvatures satisfy the Einstein condition*

$$\mathrm{Ric} = \mathrm{Ric}^* = -(n-2)g.$$

PROOF. The primal connection ∇ has constant sectional curvature k with respect to g if and only if

$$R(X, Y)Z \equiv k(\langle Y, Z \rangle X - \langle X, Z \rangle Y)$$

for all smooth vector fields X, Y and Z (see [13], (9.7.41)). For the primal Riemann–Christoffel curvature tensor, we have

$$R\left(\frac{\partial}{\partial \theta_i}, \frac{\partial}{\partial \theta_j} \right) \frac{\partial}{\partial \theta_k} = R_{ijk}^{\ell} \frac{\partial}{\partial \theta_{\ell}} = -\left(\left\langle \frac{\partial}{\partial \theta_j}, \frac{\partial}{\partial \theta_k} \right\rangle \frac{\partial}{\partial \theta_i} - \left\langle \frac{\partial}{\partial \theta_i}, \frac{\partial}{\partial \theta_k} \right\rangle \frac{\partial}{\partial \theta_j} \right),$$

which implies that the sectional curvature is $k = -1$. The claim for Ricci curvature follows immediately by taking trace (see, for example, [48], (4.31)). The proof for the dual curvatures is the same. $\quad\square$

5. Geodesics and generalized Pythagorean theorem. Armed with the primal and dual connections we can formulate the primal and dual geodesic equations. Their solutions are the primal and dual geodesics which will be studied in this section. The highlight of this section is the generalized Pythagorean theorem (Theorem 1.2). Along the way we will prove some remarkable properties of the geometric structure (g, ∇, ∇^*).

5.1. *Primal and dual geodesics.* Note that in Figure 1 the primal geodesic is drawn as a straight line in Δ_n. We now prove that this is indeed the case. The same is true for the dual geodesic in dual Euclidean coordinates.

Let $\gamma : [0, 1] \to \Delta_n$ be a smooth curve. We denote time derivatives by $\dot{\gamma}(t)$. Let $\theta(t)$ and $\phi(t)$ be the primal and dual coordinate representations of γ. We say that γ is a primal geodesic if its satisfies

$$\ddot{\theta}_k(t) + \sum_{i,j=1}^{n-1} \Gamma_{ij}^k(\theta(t))\dot{\theta}_i(t)\dot{\theta}_j(t) = 0, \qquad k = 1, \ldots, n-1.$$

It is a dual geodesic if its satisfies

$$\ddot{\phi}_k(t) + \sum_{i,j=1}^{n-1} \Gamma_{ij}^{*k}(\phi(t))\dot{\phi}_i(t)\dot{\phi}_j(t) = 0, \qquad k = 1, \ldots, n-1.$$

By Theorem 4.7, the primal geodesic equation in primal coordinates is

$$(5.1) \qquad \ddot{\theta}_k(t) + 2\dot{\theta}_k(t)\sum_{\ell=1}^{n-1} \pi_\ell(\theta(t))\dot{\theta}_\ell(t) = 0, \quad k = 1, \ldots, n-1.$$

The dual geodesic equation in dual coordinates is

$$(5.2) \qquad \ddot{\phi}_k(t) - 2\dot{\phi}_k(t)\sum_{\ell=1}^{n-1} \pi_\ell(\phi(t))\dot{\phi}_\ell(t) = 0, \qquad k = 1, \ldots, n-1.$$

THEOREM 5.1 (Primal and dual geodesics). (i) *Let $\gamma : [0, 1] \to \Delta_n$ be a primal geodesic. Then the trace of γ in Δ_n is the Euclidean straight line in Δ_n joining $\gamma(0)$ and $\gamma(1)$.*

(ii) *Let $\gamma^* : [0, 1] \to \Delta_n$ be a dual geodesic. For each t, let $p^*(t)$ be the dual Euclidean coordinate of $\gamma(t)$. Then the trace of p^* in Δ_n is the Euclidean straight line in Δ_n joining $p^*(0)$ and $p^*(1)$.*

We will prove (i) and leave (ii) to the reader. The proof makes use of the following lemmas.

LEMMA 5.2. *Let $q, r \in \Delta_n$ and let θ^q and θ^r their primal coordinates respectively. Consider the differential equation*

$$(5.3) \qquad h''(t) - 2(h'(t))^2 \sum_{\ell=1}^{n-1} \pi_\ell(\theta(t)) \frac{\theta_\ell^r - \theta_\ell^q}{(1 - h(t))e^{\theta_\ell^q} + h(t)e^{\theta_\ell^r}} = 0,$$

where $\theta(t)$ is defined by

$$(5.4) \qquad \theta_k(t) = \log((1 - h(t))e^{\theta_k^q} + h(t)e^{\theta_k^r}), \qquad k = 1, \ldots, n - 1.$$

Then there exists a unique solution $h : [0, 1] \to \mathbb{R}$ satisfying $h(0) = 0$, $h'(t) > 0$ for $t \in [0, 1]$, and $h(1) = 1$.

PROOF. First, we note that if $h(t)$ satisfies (5.3), then for any $c > 0$ the map $t \mapsto h(ct)$ also does (with domain scaled by $1/c$). Also, if $h'(t_0) = 0$ then $h(t) = h(t_0)$ for all $t \geq t_0$. Let $h_0(t)$ be a maximal solution to (5.3) defined on an interval $[0, t_{\max})$ with $h_0(0) = 0$ and $h'_0(0) > 0$. Since the equation can be solved in a neighborhood of $t = 0$, we have $t_{\max} > 0$. By the previous remark, h_0 is strictly increasing on $[0, t_{\max})$. If $h_0(t)$ hits 1 at some $t = t_0 < t_{\max}$, the function $h(t) = h_0(t/t_0)$ is a solution with the desired properties. In fact, we claim that

$$(5.5) \qquad \lim_{t \uparrow t_{\max}} h_0(t) = \sup_{t < t_{\max}} h_0(t) = M,$$

where M is defined by

$$M := \inf\left\{s > 0 : \min_{1 \leq \ell \leq n-1} (1 - s)e^{\theta_\ell^q} + se^{\theta_\ell^r} = 0\right\}$$

$$= \min_{1 \leq \ell \leq n-1} \left(\frac{e^{\theta_\ell^q}}{e^{\theta_\ell^q} - e^{\theta_\ell^r}} 1_{\{\theta_\ell^q > \theta_\ell^r\}} + \infty \cdot 1_{\{\theta_\ell^q \leq \theta_\ell^r\}}\right) > 1.$$

Thus as $h(t)$ approaches M from below, at least one of the fractions in (5.3) blows up to $+\infty$. It follows that $M' := \sup_{t < t_{\max}} h(t) \leq M$. Suppose on the contrary that $M' < M$. Let $h_1(t)$, $t \in (-\epsilon, \epsilon)$ be a solution to (5.3) satisfying $h_1(0) = M'$ and $h'_1(0) > 0$. Note that h_1 exists because the fractions in (5.3) are finite and continuous near $M' < M$. Then the range of h_1 contains an open interval containing M'. Thus, there exists $t_0 < t_{\max}$, $c > 0$ and $t_1 < 0$ such that $ct_1 > -\epsilon$, $h_0(t_0) = h_1(ct_1)$ and $h'_0(t_0) = \frac{d}{dt}h_1(ct)|_{t=t_1}$. This allows us to extend the range of h_0 beyond M' which contradicts the maximality of M'. By the uniqueness theorem for ODE [note that (5.3) has smooth coefficients], the solution h is unique. $\square$

LEMMA 5.3. *Let h be the solution in Lemma 5.2, and consider the curve $\gamma : [0, 1] \to \Delta_n$ given in exponential coordinates by (5.4). Then γ is a primal geodesic from q to r. Moreover, the trace of γ in Δ_n is the Euclidean straight line in Δ_n joining $q = \gamma(0)$ and $r = \gamma(1)$.*

PROOF. That γ is a primal geodesic from q to r can be verified directly by differentiating (5.4) and plugging into the primal geodesic equation (5.1). We omit the computational details.

To see that the trace of γ is a Euclidean straight line in Δ_n, consider its Euclidean representation $p(t) = (p_1(t), \ldots, p_n(t))$. By (5.4) we have

$$e^{\psi(\theta(t))} = \frac{1}{p_n(t)} = (1 - h(t))e^{\psi(\theta^q)} + h(t)e^{\psi(\theta^r)}.$$

Solving for $h(t)$ gives

$$(5.6) \qquad h(t) = \frac{e^{\psi(\theta(t))} - e^{\psi(\theta^q)}}{e^{\psi(\theta^r)} - e^{\psi(\theta^q)}}.$$

Expressing (5.4) in Euclidean coordinates and using (5.6), we get after some algebra that

$$p_k(t) = e^{\theta_k^q} p_n(t) + \left(1 - e^{\psi(\theta^q)} p_n(t)\right) \frac{e^{\theta_k^r} - e^{\theta_k^q}}{e^{\psi(\theta^r)} - e^{\psi(\theta^q)}}, \qquad k = 1, \ldots, n-1.$$

Hence, there exists a_k, b_k such that

$$(5.7) \qquad p_k(t) = a_k + b_k p_n(t), \qquad k = 1, \ldots, n-1.$$

Together with the identity $p_1(t) + \cdots + p_n(t) \equiv 1$, (5.7) shows that γ is a time change of the Euclidean straight line from q to r. $\quad\square$

PROOF OF THEOREM 5.1. (i) We have shown in Lemma 5.3 that for any pair of points (q, r) in Δ_n, there exists a primal geodesic from q to r which is a time-changed Euclidean straight line. It remains to observe that the geodesic is unique. Indeed, let γ be any primal geodesic. Then it solves the primal geodesic equation (5.1). Consider the initial velocity $\gamma'(0)$. Let $q = \gamma(0)$. By varying $r \in \Delta_n$ as well as the initial speed, there exists a primal geodesic $\widetilde{\gamma}$ in the form (5.4) from q to r such that $\dot{\gamma}(0) = \dot{\widetilde{\gamma}}(0)$. [This is because $T_q \Delta_n = \bigcup_{r \in \Delta_n, \kappa > 0} \kappa(r - q)$.] By the uniqueness theorem of ODE, we have $\gamma(\cdot) = \widetilde{\gamma}(\cdot)$.

Using the dual coordinate system ϕ and dual Euclidean coordinate system p^*, (ii) can be proved in a similar way by considering the curve defined by

$$(5.8) \qquad \phi_k(t) = \log\left(\frac{1}{(1 - h(t))e^{-\phi_k^q} + h(t)e^{-\phi_k^p}}\right), \qquad k = 1, \ldots, n-1. \qquad \square$$

A connection is projectively flat if there is a coordinate system under which the geodesics are straight lines up to time reparameterizations. We say that the geometric structure (g, ∇, ∇^*) is dually projectively flat if both ∇ and ∇^* are projectively flat. In view of Theorem 5.1, we have the following corollary.

COROLLARY 5.4. *The manifold Δ_n equipped with the geometric structure (g, ∇, ∇^*) is dually projectively flat, but is not flat for $n \geq 3$.*

5.2. *Gradient flows and inverse exponential maps.* Motivated by the recent paper [5] we relate the primal and dual geodesics with gradient flows under the L-divergence. Fix $p, q, r \in \Delta_n$. Consider the following gradient flows starting at q:

$$(5.9) \qquad \begin{cases} \dot{\gamma}(t) = -\mathrm{grad}T(r|\cdot)(\gamma(t)) \\ \gamma(0) = q \end{cases} \qquad \text{(primal flow)}$$

and

$$(5.10) \qquad \begin{cases} \dot{\gamma}^*(t) = -\mathrm{grad}T(\cdot|p)(\gamma^*(t)) \\ \gamma^*(0) = q \end{cases} \qquad \text{(dual flow)}.$$

Here grad denotes the Riemannian gradient with respect to the metric g. We call (5.9) the primal flow and (5.10) the dual flow.

It can be verified easily that

$$\frac{d}{dt}T(r|\gamma(t)) = -\|\dot{\gamma}(t)\|^2 \quad \text{and} \quad \frac{d}{dt}T(\gamma^*(t)|p) = -\|\dot{\gamma}^*(t)\|^2.$$

Since $T(q|p) = 0$ if and only if $p = q$, by standard ODE theory it can be shown that the solutions $\gamma(t)$ and $\gamma^*(t)$ are defined for $t \in [0, \infty)$ and

$$\lim_{t \to \infty} \gamma(t) = r, \qquad \lim_{t \to \infty} \gamma^*(t) = p.$$

In other words, both gradient flows converge to the unique minimizers.

THEOREM 5.5 (Gradient flows). (i) *The primal flow $\gamma(t)$ is a time change of the primal geodesic from q to r.* (ii) *The the dual flow $\gamma^*(t)$ is a time change of the dual geodesic from q to p.*

Recall the concept of exponential map. For $q \in \Delta_n$ and $v \in T_q \Delta_n$, consider the primal geodesic γ starting at q with initial velocity v. If γ is defined up to time 1, we define $\exp_q(v) = \gamma(1)$. The dual exponential map $\exp^*$ is defined analogously. As a corollary of Theorems 5.1 and 5.5, we have the following characterization of the primal and dual inverse exponential maps.

COROLLARY 5.6 (Inverse exponential maps). *Let* exp *and* exp* *be the exponential maps with respect to the primal and dual connections, respectively. For $p, q \in \Delta_n$ we have*

(i) $\exp_q^{-1}(p) \propto -\mathrm{grad}T(p|\cdot)(q)$.
(ii) $(\exp_q^*)^{-1}(p) \propto -\mathrm{grad}T(\cdot|p)(q)$.

To prove Theorem 5.5, we begin by computing the Riemannian gradients of $T(r|\cdot)$ and $T(\cdot|p)$. The computation is somewhat tricky and will be given in the Appendix. Recall the notations in (4.4).

LEMMA 5.7 (Riemannian gradients).   *Let $p, q, r \in \Delta_n$.*

(i) *Under the primal coordinate system, we have*

$$\mathrm{grad}T(r|\cdot)(q) = \sum_{i=1}^{n-1}\left(-\frac{\boldsymbol{\pi}_i(\theta^r, \theta^q)}{\boldsymbol{\pi}_i(\theta^q)} + \frac{\boldsymbol{\pi}_n(\theta^r, \theta^q)}{\boldsymbol{\pi}_n(\theta^q)}\right)\frac{\partial}{\partial\theta_i^q}$$

(5.11)

$$= \frac{1}{\sum_{\ell=1}^n \boldsymbol{\pi}_\ell(\theta^q)e^{\theta_\ell^r - \theta_\ell^q}}\sum_{i=1}^{n-1}(-e^{\theta_i^r - \theta_i^q} + 1)\frac{\partial}{\partial\theta_i^q}.$$

(ii) *Under the dual coordinate system, we have*

$$\mathrm{grad}T(\cdot|p)(q) = \sum_{i=1}^{n-1}\left(\frac{\boldsymbol{\pi}_i^*(\phi^q, \phi^p)}{\boldsymbol{\pi}_i(\phi^q)} - \frac{\boldsymbol{\pi}_n^*(\phi^q, \phi^p)}{\boldsymbol{\pi}_n(\phi^q)}\right)\frac{\partial}{\partial\phi_i^q}$$

(5.12)

$$= \frac{1}{\sum_{\ell=1}^n \boldsymbol{\pi}_\ell(\phi^q)e^{\phi_\ell^q - \phi_\ell^p}}\sum_{i=1}^{n-1}(e^{\phi_i^q - \phi_i^p} - 1)\frac{\partial}{\partial\phi_i^q}.$$

PROOF OF THEOREM 5.5.   We prove (i) and leave (ii) to the reader. Let $\theta(t)$ be the primal representation of the primal flow starting at q. By Lemma 5.7, at any time t we have

$$\dot{\theta}_k(t) \propto e^{\theta_k^r - \theta_k(t)} - 1 = \frac{1}{e^{\theta_k(t)}}\big(e^{\theta_k^r} - e^{\theta_k(t)}\big),$$

where the constant of proportionality depends on $\theta(t)$ but is independent of k. It follows that

(5.13) $$\frac{d}{dt}e^{\theta_k(t)} \propto e^{\theta_k^r} - e^{\theta_k(t)}, \qquad k = 1, \ldots, n-1.$$

Comparing (5.13) and (5.4) we see that the primal flow is a time change of the primal geodesic.   □

5.3. *Generalized Pythagorean theorem.*   Having characterized the primal and dual geodesics, we are ready to prove the generalized Pythagorean theorem. Our proof makes use of the Riemannian gradients given in Lemma 5.7. The reason is that these gradients appear to have the correct scaling which is easier to handle, as can be seen in the proof [see (5.16)].

PROOF OF THEOREM 1.2.   Given $p, q, r \in \Delta_n$, consider the primal geodesic from q to r and the dual geodesic from q to p. Let

(5.14) $$u = -\mathrm{grad}T(\cdot|p)(q) \quad \text{and} \quad v = -\mathrm{grad}T(r|\cdot)(q).$$

By Theorem 5.5, u and v are proportional to the initial velocities of the two geodesics. Thus, it suffices to prove that the sign of (1.8) is the same as that of $\langle u, v \rangle$. This claim will be established by the following two lemmas.

LEMMA 5.8.   *The sign of $T(q|p) + T(r|q) - T(r|p)$ is the same as that of*

$$(5.15) \qquad 1 - \sum_{k=1}^{n} \frac{\Pi_k(q, p)\Pi_k(r, q)}{\boldsymbol{\pi}_k(q)}.$$

PROOF.   By Lemma 3.9, the sign of $T(q|p) + T(r|q) - T(r|p)$ is the same as that of

$$\left(\sum_{i=1}^{n} \boldsymbol{\pi}_i(\theta^p)e^{\theta_i^q - \theta_i^p}\right)\left(\sum_{j=1}^{n} \boldsymbol{\pi}_j(\theta^q)e^{\theta_j^r - \theta_j^q}\right) - \sum_{i=1}^{n} \boldsymbol{\pi}_i(\theta^p)e^{\theta_i^r - \theta_i^p}.$$

Rearranging, we have

$$-\sum_{i,j=1}^{n} \boldsymbol{\pi}_i(\theta^p)(\delta_{ij} - \boldsymbol{\pi}_j(\theta^q))e^{\theta_j^r - \theta_j^q}e^{\theta_i^q - \theta_i^p}.$$

Since scaling does not change sign, we may consider instead the quantity

$$-\sum_{i,j=1}^{n} \boldsymbol{\pi}_i(\theta^p)(\delta_{ij} - \boldsymbol{\pi}_j(\theta^q))\frac{\Pi_j(\theta^r, \theta^q)}{\boldsymbol{\pi}_j(\theta^q)}\frac{\Pi_i(\theta^q, \theta^p)}{\boldsymbol{\pi}_i(\theta^p)}.$$

We get (5.15) by expanding. $\square$

LEMMA 5.9.   *Consider the tangent vectors u and v defined by (5.14). Then*

$$(5.16) \qquad \langle u, v \rangle = 1 - \sum_{k=1}^{n} \frac{\Pi_k(\theta^q, \theta^p)\Pi_k(\theta^r, \theta^q)}{\boldsymbol{\pi}_k(\theta^q)}.$$

PROOF.   For this computation, we use the primal coordinate system. We have

$$u = -\text{grad}T(\cdot|p)(q) = -\sum_{i,k=1}^{n-1} g^{ik}(q)\frac{\partial}{\partial\theta_k^q}T(\cdot|p)(\theta^q)\frac{\partial}{\partial\theta_i^q},$$

$$v = -\text{grad}T(r|\cdot)(q) = -\sum_{j,\ell=1}^{n-1} g^{j\ell}(q)\frac{\partial}{\partial\theta_\ell^q}T(r|\cdot)(\theta^q)\frac{\partial}{\partial\theta_j^q}.$$

Using the definition of the Riemannian inner product, we compute

$$\langle u, v \rangle = \sum_{i,j=1}^{n-1} g_{ij}(q) \sum_{k,\ell=1}^{n-1} g^{ik}(q)g^{j\ell}(q)\frac{\partial}{\partial\theta_k^q}T(\cdot|p)(\theta^q)\frac{\partial}{\partial\theta_\ell^q}T(r|\cdot)(\theta^q)$$

$$= \sum_{k,\ell=1}^{n-1} g^{k\ell}(\theta^q)\frac{\partial}{\partial\theta_k^q}T(\cdot|p)(\theta^q)\frac{\partial}{\partial\theta_\ell^q}T(r|\cdot)(\theta^q).$$

By (4.9), (4.18) and (A.6), we have

$$g^{k\ell}(q) = \frac{1}{\boldsymbol{\pi}_k(q)} \frac{\partial \theta_\ell}{\partial \phi_k}(\phi^q) + \frac{1}{\boldsymbol{\pi}_n(\theta^q)} \sum_{\alpha=1}^{n-1} \frac{\partial \theta_\ell}{\partial \phi_\alpha}(\phi^q),$$

$$\frac{\partial}{\partial \theta_k^q} T(\cdot|p)(\theta^q) = \Pi_k(\theta^q, \theta^p) - \boldsymbol{\pi}_k(\theta^q),$$

$$\frac{\partial}{\partial \theta_\ell^q} T(r|\cdot)(\theta^q) = \sum_{\beta=1}^{n-1} \frac{\partial \phi_\beta}{\partial \theta_\ell^q}(\theta^q)(\boldsymbol{\pi}_\beta(\theta^q) - \Pi_\beta(\theta^r, \theta^q)).$$

Claim. We have

$$\langle u, v \rangle = \sum_{k,\ell=1}^{n-1} (\Pi_k(\theta^q, \theta^p) - \boldsymbol{\pi}_k(\theta^q))(\boldsymbol{\pi}_\ell(\theta^q)$$

(5.17)

$$- \Pi_\ell(\theta^r, \theta^q))\left(\frac{\delta_{k\ell}}{\boldsymbol{\pi}_k(\theta^q)} + \frac{1}{\boldsymbol{\pi}_n(\theta^q)}\right).$$

To see this, write

$$\langle u, v \rangle = \sum_{k,\beta=1}^{n-1} (\Pi_k(\theta^q, \theta^p) - \boldsymbol{\pi}_k(\theta^q))(\boldsymbol{\pi}_\beta(\theta^q) - \Pi_\beta(\theta^r, \theta^q))$$

$$\times \sum_{\ell=1}^{n-1}\left(\frac{1}{\boldsymbol{\pi}_k(q)} \frac{\partial \phi_\beta}{\partial \theta_\ell^q}(\theta^q)\frac{\partial \theta_\ell}{\partial \theta_k}(\phi^q) + \frac{1}{\boldsymbol{\pi}_n(q)} \sum_{\alpha=1}^{n-1} \frac{\partial \phi_\beta}{\partial \theta_\ell^q}(\theta^q)\frac{\partial \theta_\ell}{\partial \phi_\alpha}(\phi^q)\right).$$

The last expression can be simplified using the identities

$$\frac{\partial \phi_\beta}{\partial \theta_\ell^q}(\theta^q)\frac{\partial \theta_\ell}{\partial \theta_k}(\phi^q) = \delta_{\beta k}, \qquad \frac{\partial \phi_\beta}{\partial \theta_\ell^q}(\theta^q)\frac{\partial \theta_\ell}{\partial \phi_\alpha}(\phi^q) = \delta_{\alpha\beta},$$

and this gives the claim.

Finally, expanding and simplifying (5.17), we obtain the desired identity (5.16).
□

With these two lemmas the proof is complete. □

5.4. *Application to mathematical finance.* In this subsection, we explain how the new information geometry can be applied to finance. For financial background and further details, we refer the reader to [45]. Consider sequential investment in a stock market with n stocks. At time $t = 0, 1, 2, \ldots$, let $\mu(t) \in \Delta_n$ be the distribution of capital in the market. Explicitly, let $X_i(t) > 0$ be the market capitalization of stock i. Then

(5.18) $$\mu_i(t) = \frac{X_i(t)}{X_1(t) + \cdots + X_n(t)}$$

is the market weight of stock i.

Let $\boldsymbol{\pi} : \Delta_n \to \Delta_n$ be a portfolio map. It defines a self-financing investment strategy such that if the current state of the market is $\mu(t) = p$, we distribute our capital among the stocks according to the vector $\boldsymbol{\pi}(p)$. For example, if $\boldsymbol{\pi}(p) = (0.4, 0.5, 0.1)$, we invest 40% in stock 1, 50% in stock 2, and the rest in stock 3. After each period, we perform the necessary trading such that the proportions of capitals invested are rebalanced to $\boldsymbol{\pi}(\mu(t + 1))$. If one chooses the identity map $\boldsymbol{\pi}(p) \equiv p$, the resulting portfolio is called the market portfolio which we denote by μ. The market portfolio plays the role of benchmark.

Now let $\boldsymbol{\pi}$ be the portfolio map generated by an exponentially concave function φ (Definition 2.4), and consider the value of the resulting portfolio beginning with $1 at time 0. Also consider the value of the market portfolio starting with $1. Consider the ratio

$$V(t) := \frac{\text{value of the portfolio } \boldsymbol{\pi} \text{ at time } t}{\text{value of the portfolio } \mu \text{ at time } t}$$

called the relative value of the portfolio $\boldsymbol{\pi}$. Under suitable conditions (see [44, 45]) it can be shown that $V(0) = 1$ and

$$\frac{V(t + 1)}{V(t)} = \sum_{i=1}^{n} \pi_i\big(\mu(t)\big) \frac{\mu_i(t + 1)}{\mu_i(t)}.$$

PROPOSITION 5.10 (Fernholz's decomposition). [24, 45] *For all time t, we have*

$$\log V(t) = \big(\varphi(\mu(t)) - \varphi(\mu(0))\big) + \sum_{s=0}^{t-1} T\big(\mu(s + 1)|\mu(s)\big).$$

In this sense, the cumulative L-divergence measures the volatility harvested by the portfolio $\boldsymbol{\pi}$. See [54] for optimization of this quantity.

The above discussion assumes that the portfolio rebalances every period (say every week). In practice, due to transaction costs and other considerations, we may want to rebalance at other frequencies. Let $0 = t_0 < t_1 < t_2$ be three time points and consider two ways of implementing the portfolio $\boldsymbol{\pi}$: (i) rebalance at times t_0 and t_1 (ii) rebalance at time t_0 only. By Fernholz's decomposition, the relative values of the two implementations at time t_2 are

$$\log V^{(1)}(t_2) = \big(\varphi(\mu(t_2)) - \varphi(\mu(t_0))\big) + T\big(\mu(t_1)|\mu(t_0)\big) + T\big(\mu(t_2)|\mu(t_1)\big),$$

$$\log V^{(2)}(t_2) = \big(\varphi(\mu(t_2)) - \varphi(\mu(t_0))\big) + T\big(\mu(t_2)|\mu(t_0)\big).$$

Letting $\mu(t_0) = p$, $\mu(t_1) = q$ and $\mu(t_2) = r$, the difference between the two values is

$$\log V^{(1)}(t_2) - \log V^{(2)}(t_2) = T(q|p) + T(r|q) - T(r|p).$$
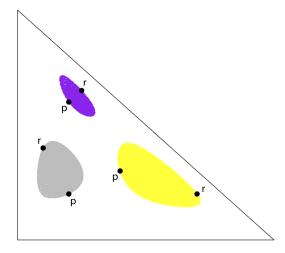
FIG. 3. *Plots of the region $\{q \in \Delta_n : T(q|p) + T(r|q) \leq T(r|p)\}$ for the equal-weighted portfolio $\pi \equiv (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, for several values of p and r. Each point q on the boundary gives a 'right-angled geodesic triangle' in the sense of Theorem* 1.2.

By the generalized Pythagorean theorem, the sign of the difference is determined by the angle between the dual geodesic from q to p and the primal geodesic from q to r. In Figure 3, we illustrate the result for the equal-weighted portfolio of 3 stocks. In the figure, rebalancing at time t_1 creates extra profit if and only if q lies outside the region. This shows convincingly that rebalancing more frequently is not always better, even in the absence of transaction costs. More importantly, our framework provides a geometric and model-independent way of saying that the sequence (p, q, r) is 'more volatile' than the subsequence (p, r).

6. Displacement interpolation. In this final section, we consider displacement interpolation for the optimal transport problem formulated in Section 2.2. We refer the reader to [50], Chapter 5, and [51], Chapter 7, for introductions to displacement interpolation.

6.1. *Time dependent transport problem.* Let $P^{(0)}$ and $P^{(1)}$ be Borel probability measures on $\mathbb{R}^{n-1}$. Consider the transport problem with cost c given by (1.7). Suppose the transport problem is solved in terms of the exponentially concave function φ on Δ_n. Letting $f = \varphi + \psi$, the (Monge) optimal transport map is given by the c-supergradient of f. In particular, $P^{(1)}$ is the pushforward of $P^{(0)}$ under $F := \nabla^c f$:

$$P^{(1)} = F_\# P^{(0)}.$$

The idea of displacement interpolation is to introduce an additional time structure. We want to define an 'action' $\mathcal{A}(\cdot)$ on curves such that the cost function is

given by

$$c(\theta, \phi) = \min_{\gamma} \mathcal{A}(\gamma),$$

where the minimum is taken over smooth curves $\gamma : [0, 1] \to \mathbb{R}^{n-1}$ satisfying $\gamma(0) = \theta$ and $\gamma(1) = \phi$. For each pair (θ, ϕ), a minimizing curve γ gives a time-dependent map transporting θ to ϕ along a continuous path. Let $F^{(t)} : \mathbb{R}^{n-1} \to \mathbb{R}^{n-1}$ be defined by $F^{(t)}(\theta) = \gamma(t)$, where γ is the minimizing curve for the pair $(\theta, F(\theta))$. We want to define $\mathcal{A}$ in such a way that $F^{(t)}$ is an optimal transport map for the probability measures $P^{(0)}$ and $P^{(t)}$ where

$$P^{(t)} = (F^{(t)})_{\#} P_0.$$

For the classical Euclidean case with cost $|x - y|^2$ the action is $\mathcal{A}(\gamma) = \int_0^1 |\dot{\gamma}(t)|^2 \, dt$ and the optimal transport map has the form $F(x) = x - \nabla h(x)$ where h is an ordinary concave function. The displacement interpolations are linear interpolations:

$$F^{(t)} = (1 - t)\mathrm{Id} + t F.$$

(See [50], Theorem 5.5, Theorem 5.6.) In particular, the individual trajectories (minimizing curves) are Euclidean straight lines which can be regarded as the geodesics of a flat geometry. In this section, we formulate and prove an analogous statement for our transport problem.

6.2. *Lagrangian action and portfolio interpolation.* We begin by defining an appropriate action. Let $\gamma : [0, 1] \to \mathbb{R}^{n-1}$ be a smooth curve with $\gamma(0) = \theta$. For each t, define $q(t) \in \Delta_n$ such that its exponential coordinate is $\theta - \gamma(t)$, that is,

$$(6.1) \qquad \frac{q_i(t)}{q_n(t)} = e^{\theta_i - \gamma_i(t)}, \qquad 1 \le i \le n - 1.$$

Equivalently, we have $q_i(t) = e^{\theta_i - \gamma_i(t) - \psi(\theta - \gamma(t))}$, for $1 \le i \le n - 1$. Intuitively, we think of $q(t)$ as the portfolio at time t (in the sense of interpolation). Note that $q(0) = (\frac{1}{n}, \ldots, \frac{1}{n})$.

We define the Lagrangian action by

$$(6.2) \qquad \mathcal{A}(\gamma) = \int_0^1 -\log\left(\frac{1}{n} + \dot{q}_n(t)\right) dt.$$

We take $-\log(\cdot) = \infty$ if the argument is not in $(0, \infty)$. An alternative representation of the action is

$$(6.3) \qquad \mathcal{A}(\gamma) = \int_0^1 -\log\left(\frac{1}{n} + \frac{d}{dt} e^{-\psi(\gamma(0) - \gamma(t))}\right) dt.$$

LEMMA 6.1.   *For any $\theta, \phi \in \mathbb{R}^{n-1}$ we have*

$$(6.4) \qquad c(\theta, \phi) = \psi(\theta - \phi) = \min\{\mathcal{A}(\gamma) : \gamma(0) = \theta, \gamma(1) = \phi\}.$$

The action is minimized by the curve

$$(6.5) \qquad \gamma_i(t) = \theta_i - \log \frac{(1-t)\frac{1}{n} + t q_i(1)}{(1-t)\frac{1}{n} + t q_n(1)}, \qquad 1 \le i \le n - 1.$$

In particular, for this minimizing curve we have

$$(6.6) \qquad q(t) = (1-t)\left(\frac{1}{n}, \ldots, \frac{1}{n}\right) + t q(1).$$

PROOF.   Fix a smooth curve $\gamma : [0, 1] \to \mathbb{R}^{n-1}$ from θ to ϕ. Since $-\log$ is convex, by Jensen's inequality we have

$$
\begin{aligned}
\int_0^1 -\log\left(\frac{1}{n} + \dot{q}_n(t)\right) dt &\ge -\log\left(\frac{1}{n} + \int_0^1 \dot{q}_n(t)\, dt\right) \\
(6.7) \qquad\qquad &= -\log\left(\frac{1}{n} + q_n(1) - q_n(0)\right) = -\log q_n(1) \\
&= \psi(\theta - \phi).
\end{aligned}
$$

For the curve defined by (6.5), $\dot{q}(t) = q(1) - \frac{1}{n}$ is constant and so equality holds in (6.7). Finally, (6.6) follows by a direct calculation.   □

6.3. *Displacement interpolation.*   We work under the following setting. Let $P^{(0)}$ and $P^{(1)}$ be Borel probability measures on $\mathbb{R}^{n-1}$. Let $\varphi : \Delta_n \to \mathbb{R}$ be an exponentially concave function, satisfying Assumption 2.5, such that $F^{(1)} := \nabla^c f$ is an optimal transport map (here f is the c-concave function $\varphi + \psi$). Let $\pi^{(1)} : \Delta_n \to \Delta_n$ be the portfolio map generated by $\varphi^{(1)} := \varphi$.

Consider the flow $(t, \theta) \mapsto \phi^{(t)}(\theta)$ defined by the minimizing curves (6.5), that is,

$$(6.8) \qquad \phi_i^{(t)}(\theta) = \theta_i - \log \frac{\pi_i^{(t)}(\theta)}{\pi_n^{(t)}(\theta)}, \qquad 1 \le i \le n - 1, t \in [0, 1],$$

where each $\pi^{(t)} : \Delta_n \to \Delta_n$ is the portfolio map defined by

$$(6.9) \qquad \pi^{(t)} := (1-t)\left(\frac{1}{n}, \ldots, \frac{1}{n}\right) + t\pi^{(1)}, \qquad t \in [0, 1].$$

The following is the main result of this section. It is interesting to note that the displacement interpolation can be interpreted naturally as the linear interpolation between the equal-weighted portfolio and the terminal portfolio.

THEOREM 6.2 (Displacement interpolation). *Consider the setting of Section* 6.3.

(i) *For each $t \in [0, 1]$, the portfolio map $\boldsymbol{\pi}^{(t)}$ is generated by the exponentially concave function $\varphi^{(t)}$ on Δ_n defined by*

$$(6.10) \qquad \varphi^{(t)}(p) = (1 - t) \sum_{i=1}^{n} \frac{1}{n} \log p_i + t\varphi(p), \qquad p \in \Delta_n.$$

(ii) *For each $t \in [0, 1]$, let $f^{(t)} = \varphi^{(t)} + \psi$ and let $F^{(t)} = \nabla^c f^{(t)}$. If θ is distributed according to $P^{(0)}$, then $\theta^{(t)}$ is distributed according to $P^{(t)}$ where*

$$P^{(t)} = \left(F^{(t)}\right)_{\#} P^{(0)}.$$

Moreover, $F^{(t)}$ is an optimal transport map for the transport problem for $(P^{(0)}, P^{(t)})$.

(iii) *Endow Δ_n with the geometric structure induced by the L-divergence of φ. We further assume that the c-gradient $F^{(1)} = \nabla^c f : \mathbb{R}^{n-1} \to \mathbb{R}^{n-1}$ is surjective. For each $\theta \in \mathbb{R}^{n-1}$ fixed, consider the curve $t \mapsto \phi^{(t)}(\theta)$ in dual coordinates. Then the trace of the curve is the dual geodesic joining θ and $\phi^{(1)}(\theta)$.*

PROOF. (i) Follows directly from Example 2.6(iv).

(ii) It is clear that $(\theta, F^{(t)}(\theta))$ is a coupling of $(P^{(0)}, P^{(t)})$. By Proposition 2.7, the graph of the map $F^{(t)}$ is c-cyclical monotone. This proves that $F^{(t)}$ is an optimal transport map.

(iii) We write (6.8) in the form

$$
\begin{aligned}
e^{-\phi_i^{(t)}} &= e^{-\theta_i} \frac{(1-t)\frac{1}{n} + t\boldsymbol{\pi}_i(\theta)}{(1-t)\frac{1}{n} + t\boldsymbol{\pi}_n(\theta)} \\
&= \frac{(1-t)\frac{1}{n}}{(1-t)\frac{1}{n} + t\boldsymbol{\pi}_n(\theta)} e^{-\theta_i} + \frac{t\boldsymbol{\pi}_n(\theta)}{(1-t)\frac{1}{n} + t\boldsymbol{\pi}_n(\theta)} e^{-\theta_i + \log \frac{\boldsymbol{\pi}_i(\theta)}{\boldsymbol{\pi}_n(\theta)}} \\
&=: \left(1 - h(t)\right) e^{-\theta_i} + h(t) e^{-\phi_i^{(1)}(\theta)}.
\end{aligned}
$$

By (5.8), we see that $t \mapsto \phi^{(t)}$ is a time change of a dual geodesic. The surjectivity assumption guarantees that the curve lies within $\mathcal{Y}'$, the range of the dual coordinate system. $\square$

6.4. *Another interpolation.* From the financial perspective, there is another natural interpolation, namely the linear interpolation between the market portfolio μ [Example 2.6(i)] and the portfolio $\boldsymbol{\pi}$:

$$(6.11) \qquad\qquad \boldsymbol{\pi}^{(t)} = (1 - t)\boldsymbol{\pi} + t\mu.$$

The corresponding log generating function is $\varphi^{(t)} = (1 - t)\varphi$. From the transport perspective, the market portfolio corresponds to the trivial transport map $F(\theta) \equiv 0$

[recall in Proposition 2.7(iii) that the portfolio has exponential coordinate given by $\theta - F(\theta)$]. By the argument of Theorem 6.2, we have the following result.

PROPOSITION 6.3. *Consider the geometric structure induced by φ and assume that the range of the dual coordinate system is $\mathbb{R}^{n-1}$. Consider the flow $(t, \theta) \mapsto \phi^{(t)}(\theta)$ in (6.8) where $\boldsymbol{\pi}^{(t)}$ is given by the interpolation (6.11). Then for each θ, in dual coordinates, the trace of the curve $t \mapsto \phi^{(t)}(\theta)$ is a time change of the dual geodesic from $\phi^{(0)}(\theta)$ to 0.*

APPENDIX: TECHNICAL PROOFS

PROOF OF THEOREM 3.2.   In this proof, we treat θ and ϕ as independent variables.

We prove (i) and (ii) together. We begin by observing that

$$(\text{A.1}) \qquad \frac{\partial}{\partial \theta_i} f(\theta) = \boldsymbol{\pi}_i(\theta), \qquad 1 \le i \le n - 1.$$

To see this, write $p_i = e^{\theta_i - \psi(\theta)}$. Switching coordinates and using the chain rule, we have

$$\frac{\partial}{\partial \theta_i} f(\theta) = \sum_{j=1}^{n} \frac{\partial \varphi}{\partial p_j}(p)(-p_i p_j) + \frac{\partial \varphi}{\partial p_i}(p) p_i + p_i$$

$$= p_i \left(1 + \frac{\partial \varphi}{\partial p_i}(p) - \sum_{j=1}^{n} p_j \frac{\partial \varphi}{\partial p_j}(p) \right).$$

From (1.4) and a bit of computation, we see that this equals $\boldsymbol{\pi}_i(\theta)$.

Consider the c-transform of f given by

$$(\text{A.2}) \qquad f^*(\phi) = \inf_{\theta \in \mathcal{X}} (\psi(\theta - \phi) - f(\theta)).$$

Differentiating $\psi(\theta - \phi) - f(\theta)$ and using (A.1), we see that $\theta \in \mathcal{X}$ attains the infimum in (A.2) if and only if

$$\frac{e^{\theta_i - \phi_i}}{\sum_{j=1}^{n} e^{\theta_j - \phi_j}} = \boldsymbol{\pi}_i(\theta), \qquad i = 1, \ldots, n - 1.$$

Rearranging, we have

$$(\text{A.3}) \qquad \phi_i = \theta_i - \log \frac{\boldsymbol{\pi}_i(\theta)}{\boldsymbol{\pi}_n(\theta)}, \qquad i = 1, \ldots, n - 1.$$

This proves that equality holds in

$$(\text{A.4}) \qquad f(\theta) + f^*(\phi) \le \psi(\theta - \phi)$$

if and only if θ and ϕ satisfies the relation (A.3). In particular, for $\theta \in \mathcal{X}$ the c-supergradient $\nabla^c f(\theta)$ is given by (A.3).

Next, we prove that the minimizer in (A.2), if exists, is unique. Consider instead maximization of the quantity

$$e^{f(\theta)-\psi(\theta-\phi)} = e^{\varphi(\theta)+\psi(\theta)-\psi(\theta-\phi)} = \Phi(\theta)\frac{e^{\psi(\theta)}}{e^{\psi(\theta-\phi)}}.$$

Expanding and switching to Euclidean coordinates, this equals

$$(A.5) \qquad \Phi(p)\frac{\sum_{i=1}^{n} e^{\theta_i}}{\sum_{i=1}^{n} e^{\theta_i-\phi_i}} = \Phi(p)\frac{1}{\sum_{i=1}^{n} a_i\,p_i},$$

where $a_i = e^{-\phi_i} > 0$. Being the quotient of a strictly concave function and an affine function, the right hand side of (A.5) is strictly quasi-concave, that is, its superlevel sets are strictly convex (see [11], Example 3.38). This shows that the minimizer θ in (A.2) is unique if it exists.

Let $\phi \in \mathcal{Y}'$. Then there exists unique $\theta \in \mathcal{X}$ such that equality holds in (A.4) and $\phi = \nabla^c f(\theta)$. In particular, the c-supergradient $\partial^c f^*(\phi)$ is θ and $\nabla^c f^*(\nabla^c f(\theta)) = \theta$. This completes the proof of (i) and (ii).

Next, we prove that $\nabla^c f^* = (\nabla^c f)^{-1}$ is smooth, so that $\nabla^c f : \mathcal{X} \to \mathcal{Y}'$ is a diffeomorphism. Since $\nabla^c f : \mathcal{X} \to \mathcal{Y}$ is smooth and injective, by the inverse function theorem it suffices to show that the Jacobian of $\nabla^c f$ is invertible on $\mathcal{X}$. Invertibility of the Jacobian follows from (4.16) and the fact that $(g_{ij}(\xi))$ is strictly positive definite.

Finally, by Fenchel's identity we can express f^* in the form

$$f^*(\phi) = c(\nabla^c f^*(\phi), \phi) - f(\nabla^c f^*(\phi)).$$

Since $\nabla^c f^*$ and the cost function c are smooth, we see that f^* is smooth as well. $\square$

PROOF OF LEMMA 5.7. (i) To prove the first formula in (5.11), we compute, using (4.9) and (4.12),

$$\big(\mathrm{grad}\,T\big(\cdot|\theta^p\big)(\theta^q)\big)_i$$

$$= \sum_{j=1}^{n-1} g^{ij}(\theta^q)\frac{\partial}{\partial\theta_j^q}T\big(\cdot|\theta^p\big)(\theta^q)$$

$$= \sum_{j=1}^{n-1}\left(\frac{1}{\pi_j(\theta^q)}\frac{\partial\theta_i}{\partial\phi_j}(\theta^q) + \frac{1}{\pi_n(\theta^q)}\sum_{k=1}^{n-1}\frac{\partial\theta_i}{\partial\phi_k}(\theta^q)\right)\big(\Pi_j(\theta^q,\theta^p) - \pi_j(\theta^q)\big)$$

$$= \sum_{j=1}^{n-1}\left(\frac{\Pi_j(\theta^q,\theta^p)}{\pi_j(\theta^q)} - \frac{\Pi_n(\theta^q,\theta^p)}{\pi_n(\theta^q)}\right)\frac{\partial\theta_i}{\partial\phi_j}(\theta^q).$$

For the second formula, we first prove the following.

Claim. We have

$$(A.6) \qquad \frac{\partial}{\partial \theta_j^q} T(\theta^r | \theta^q) = \sum_{\ell=1}^{n-1} \frac{\partial \phi_\ell}{\partial \theta_j^q}(\theta^q)(\boldsymbol{\pi}_\ell(\theta^q) - \Pi_\ell(\theta^r, \theta^q)).$$

To see this, we use (4.13), (4.5), and compute as follows:

$$\frac{\partial}{\partial \theta_j^q} T(\theta^r | \theta^q)$$

$$= -\Pi_j(\theta^r, \theta^q) + \boldsymbol{\pi}_j(\theta^q) + \sum_{\ell=1}^{n} \frac{1}{\boldsymbol{\pi}_\ell(\theta^q)} \frac{\partial \boldsymbol{\pi}_\ell}{\partial \theta_j^q}(\theta^q) \Pi_\ell(\theta^r, \theta^q)$$

$$= -\Pi_j(\theta^r, \theta^q) + \boldsymbol{\pi}_j(\theta^q)$$

$$\qquad + \sum_{\ell=1}^{n} \left(\delta_{\ell j} - \boldsymbol{\pi}_j(\theta^q) - \left(\frac{\partial \phi_\ell}{\partial \theta_j^q}(\theta^q) - \sum_{m=1}^{n-1} \boldsymbol{\pi}_m(\theta^q) \frac{\partial \phi_m}{\partial \theta_j^q}(\theta^q) \right) \Pi_\ell(\theta^r, \theta^q) \right)$$

$$= \sum_{\ell=1}^{n-1} \frac{\partial \phi_\ell}{\partial \theta_j^q}(\theta^q)(\boldsymbol{\pi}_\ell(\theta^q) - \Pi_\ell(\theta^r, \theta^q)).$$

Now we compute, using Theorem 4.5 and the symmetry of g^{ij},

$$\big(\mathrm{grad} T(r|\cdot)(q) \big)_i$$

$$= \sum_{j=1}^{n-1} g^{ij}(\theta^q) \frac{\partial}{\partial \theta_j^q} T(\theta^r|\cdot)(\theta^q)$$

$$= \sum_{j=1}^{n-1} \left(\frac{1}{\boldsymbol{\pi}_i(\theta^q)} \frac{\partial \theta_j}{\partial \phi_i}(\phi^q) + \frac{1}{\boldsymbol{\pi}_n(\theta^q)} \sum_{k=1}^{n-1} \frac{\partial \theta_j}{\partial \phi_k}(\phi^q) \right)$$

$$\qquad \times \sum_{\ell=1}^{n-1} \frac{\partial \phi_\ell}{\partial \theta_j^q}(\theta^q)(\boldsymbol{\pi}_\ell(\theta^q) - \Pi_\ell(\theta^r, \theta^q))$$

$$= \sum_{\ell=1}^{n-1} (\boldsymbol{\pi}_\ell(\theta^q) - \Pi_\ell(\theta^r, \theta^q))$$

$$\qquad \times \sum_{j=1}^{n-1} \left(\frac{1}{\boldsymbol{\pi}_i(\theta^q)} \frac{\partial \phi_\ell}{\partial \theta_j^q}(\phi^q) \frac{\partial \theta_j}{\partial \phi_i}(\phi^q) + \frac{1}{\boldsymbol{\pi}_n(\theta^q)} \sum_{k=1}^{n-1} \frac{\partial \phi_\ell}{\partial \theta_j^q}(\phi^q) \frac{\partial \theta_j}{\partial \phi_k}(\phi^q) \right)$$

$$= \sum_{\ell=1}^{n-1} \bigl(\boldsymbol{\pi}_\ell(\theta^q) - \Pi_\ell(\theta^r, \theta^q)\bigr)\biggl(\frac{1}{\boldsymbol{\pi}_i(\theta^q)}\delta_{\ell i} + \frac{1}{\boldsymbol{\pi}_n(\theta^q)}\biggr)$$

$$= -\frac{\Pi_i(\theta^r, \theta^q)}{\boldsymbol{\pi}_i(\theta^q)} + \frac{\Pi_n(\theta^r, \theta^q)}{\boldsymbol{\pi}_n(\theta^q)}.$$

In the second last equality we used (4.6). The proof of (ii) is similar. $\quad\square$

REFERENCES

[1] ACCIAIO, B., BEIGLBÖCK, M., PENKNER, F. and SCHACHERMAYER, W. (2016). A model-free version of the fundamental theorem of asset pricing and the super-replication theorem. *Math. Finance* **26** 233–251. MR3481303

[2] AMARI, S. (2016). *Information Geometry and Its Applications. Applied Mathematical Sciences* **194**. Springer, Berlin. MR3495836

[3] AMARI, S. and NAGAOKA, H. (2000). *Methods of Information Geometry. Translations of Mathematical Monographs* **191**. Amer. Math. Soc., Providence, RI. MR1800071

[4] AMBROSIO, L. and GIGLI, N. (2013). A user's guide to optimal transport. In *Modelling and Optimisation of Flows on Networks. Lecture Notes in Math.* **2062** 1–155. Springer, Heidelberg. MR3050280

[5] AY, N. and AMARI, S.-I. (2015). A novel approach to canonical divergences within information geometry. *Entropy* **17** 8111–8129.

[6] BANNER, A. D. and FERNHOLZ, D. (2008). Short-term relative arbitrage in volatility-stabilized markets. *Ann. Finance* **4** 445–454.

[7] BASSEVILLE, M. (2013). Divergence measures for statistical data processing—an annotated bibliography. *Signal Process.* **93** 621–633.

[8] BEIGLBÖCK, M., HENRY-LABORDÈRE, P. and PENKNER, F. (2013). Model-independent bounds for option prices—a mass transport approach. *Finance Stoch.* **17** 477–501. MR3066985

[9] BOOTH, D. G. and FAMA, E. F. (1992). Diversification returns and asset contributions. *Financ. Anal. J.* **48** 26–32.

[10] BOUCHEY, P., NEMTCHINOV, V., PAULSEN, A. and STEIN, D. M. (2012). Volatility harvesting: Why does diversifying and rebalancing create portfolio growth? *J. Wealth Manag.* **15** 26–35.

[11] BOYD, S. and VANDENBERGHE, L. (2004). *Convex Optimization*. Cambridge Univ. Press, Cambridge. MR2061575

[12] BRÈGMAN, L. M. (1967). A relaxation method of finding a common point of convex sets and its application to the solution of problems in convex programming. *Ž. VyČisl. Mat. i Mat. Fiz.* **7** 620–631. MR0215617

[13] CALIN, O. and UDRIȘTE, C. (2014). *Geometric Modeling in Probability and Statistics*. Springer, Cham. MR3308142

[14] CHAMBERS, D. R. and ZDANOWICZ, J. S. (2014). The limitations of diversification return. *J. Portf. Manag.* **40** 65–76.

[15] CORDERO-ERAUSQUIN, D. and KLARTAG, B. (2015). Moment measures. *J. Funct. Anal.* **268** 3834–3866. MR3341966

[16] CORDERO-ERAUSQUIN, D., MCCANN, R. J. and SCHMUCKENSCHLÄGER, M. (2001). A Riemannian interpolation inequality à la Borell, Brascamp and Lieb. *Invent. Math.* **146** 219–257. MR1865396

[17] DILLEN, F., NOMIZU, K. and VRANKEN, L. (1990). Conjugate connections and Radon's theorem in affine differential geometry. *Monatsh. Math.* **109** 221–235. MR1058409

[18] DOLINSKY, Y. and SONER, H. M. (2014). Martingale optimal transport and robust hedging in continuous time. *Probab. Theory Related Fields* **160** 391–427. MR3256817

[19] EGUCHI, S. (1983). Second order efficiency of minimum contrast estimators in a curved exponential family. *Ann. Statist.* **11** 793–803. MR0707930

[20] EGUCHI, S. (1992). Geometry of minimum contrast. *Hiroshima Math. J.* **22** 631–647. MR1194056

[21] ERB, C. B. and HARVEY, C. R. (2006). The strategic and tactical value of commodity futures. *Financ. Anal. J.* **62** 69–97.

[22] ERBAR, M., KUWADA, K. and STURM, K.-T. (2015). On the equivalence of the entropic curvature-dimension condition and Bochner's inequality on metric measure spaces. *Invent. Math.* **201** 993–1071. MR3385639

[23] FERNHOLZ, E. R. (2002). *Stochastic Portfolio Theory*. Springer, Berlin.

[24] FERNHOLZ, R. (1999). Portfolio generating functions. In *Quantitative Analysis in Financial Markets* (M. Avellaneda, ed.) World Scientific, Singapore.

[25] FERNHOLZ, R. (2001). Equity portfolios generated by functions of ranked market weights. *Finance Stoch.* **5** 469–486. MR1861997

[26] FERNHOLZ, R., KARATZAS, I. and KARDARAS, C. (2005). Diversity and relative arbitrage in equity markets. *Finance Stoch.* **9** 1–27. MR2210925

[27] FERNHOLZ, R. and SHAY, B. (1982). Stochastic portfolio theory and stock market equilibrium. *J. Finance* **37** 615–624.

[28] HALLERBACH, W. G. (2014). Disentangling rebalancing return. *J. Asset Manag.* **15** 301–316.

[29] HAZAN, E., AGARWAL, A. and KALE, S. (2007). Logarithmic regret algorithms for online convex optimization. *Mach. Learn.* **69** 169–192.

[30] JUDITSKY, A., RIGOLLET, P. and TSYBAKOV, A. B. (2008). Learning by mirror averaging. *Ann. Statist.* **36** 2183–2206. MR2458184

[31] KARATZAS, I. and RUF, J. (2016). Trading strategies generated by Lyapunov functions. Available at arXiv:1603.08245.

[32] KASS, R. E. and VOS, P. W. (1997). *Geometrical Foundations of Asymptotic Inference*. Wiley, New York. MR1461540

[33] KUROSE, T. (1990). Dual connections and affine geometry. *Math. Z.* **203** 115–121. MR1030710

[34] LOTT, J. and VILLANI, C. (2009). Ricci curvature for metric-measure spaces via optimal transport. *Ann. of Math.* (2) **169** 903–991. MR2480619

[35] MAHDAVI, M., ZHANG, L. and JIN, R. (2015). Lower and upper bounds on the generalization of stochastic exponentially concave optimization. In *Proceedings of the* 28*th Conference on Learning Theory* 1305–1320.

[36] MATSUZOE, H. (1999). Geometry of contrast functions and conformal geometry. *Hiroshima Math. J.* **29** 175–191. MR1679582

[37] MCCANN, R. J. (1994). A convexity theory for interacting gases and equilibrium crystals. Ph.D. thesis, Princeton University.

[38] MCCANN, R. J. (1997). A convexity principle for interacting gases. *Adv. Math.* **128** 153–179. MR1451422

[39] MURRAY, M. K. and RICE, J. W. (1993). *Differential Geometry and Statistics*. *Monographs on Statistics and Applied Probability* **48**. Chapman & Hall, London. MR1293124

[40] NAGAOKA, H. and AMARI, S.-I. (1982). Differential geometry of smooth families of probability distributions. *Univ. Tokyo*, *Japan*, *METR* 82–7.

[41] OLIKER, V. (2007). Embedding $\mathbf{S}^n$ into $\mathbf{R}^{n+1}$ with given integral Gauss curvature and optimal mass transport on $\mathbf{S}^n$. *Adv. Math.* **213** 600–620. MR2332603

[42] PAL, S. (2016). Exponentially concave functions and high dimensional stochastic portfolio theory. Available at arXiv:1603.01865.

[43] PAL, S. (2017). Embedding optimal transports in statistical manifolds. *Indian J. Pure Appl. Math.* To appear.

[44] PAL, S. and WONG, T.-K. L. (2013). Energy, entropy, and arbitrage. Available at arXiv:1308.5376.

[45] PAL, S. and WONG, T.-K. L. (2016). The geometry of relative arbitrage. *Math. Financ. Econ.* **10** 263–293. MR3500452

[46] ROCKAFELLAR, R. T. (1997). *Convex Analysis*. Princeton Univ. Press, Princeton, NJ. MR1451876

[47] SHIMA, H. (2007). *The Geometry of Hessian Structures*. World Scientific Co. Pte., Hackensack, NJ. MR2293045

[48] STERNBERG, S. (2012). *Curvature in Mathematics and Physics*. Dover Publications, Mineola, NY. MR2985540

[49] STRONG, W. (2014). Generalizations of functionally generated portfolios with applications to statistical arbitrage. *SIAM J. Financial Math.* **5** 472–492. MR3246899

[50] VILLANI, C. (2003). *Topics in Optimal Transportation. Graduate Studies in Mathematics* **58**. Amer. Math. Soc., Providence, RI. MR1964483

[51] VILLANI, C. (2009). *Optimal Transport*: *Old and New*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **338**. Springer, Berlin. MR2459454

[52] WILLENBROCK, S. (2011). Diversification return, portfolio rebalancing, and the commodity return puzzle. *Financ. Anal. J.* **67** 42–49.

[53] WONG, T.-K. L. (2015). Optimization of relative arbitrage. *Ann. Finance* **11** 345–382. MR3426111

[54] WONG, T.-K. L. (2015). Universal portfolios in stochastic portfolio theory. Available at arXiv:1510.02808.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF WASHINGTON
C-547 PADELFORD HALL
SEATTLE, WASHINGTON 98195
USA
E-MAIL: soumikpal@gmail.com

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SOUTHERN CALIFORNIA
406H KAPRIELIAN HALL
LOS ANGELES, CALIFORNIA 90089
USA
E-MAIL: tkleonardwong@gmail.com