

EMBEDDING OPTIMAL TRANSPORTS IN STATISTICAL MANIFOLDS¹

Soumik Pal

Department of Mathematics, University of Washington, Seattle, WA 98195

e-mail: soumikpal@gmail.com

(Received 22 July 2016; accepted 30 August 2016)

We consider Monge-Kantorovich optimal transport problems on $\mathbb{R}^d$, $d \geq 1$, with a convex cost function given by the cumulant generating function of a probability measure. Examples include the Wasserstein-2 transport whose cost function is the square of the Euclidean distance and corresponds to the cumulant generating function of the multivariate standard normal distribution. The optimal coupling is usually described via an extended notion of convex/concave functions and their gradient maps. These extended notions are nonintuitive and do not satisfy useful inequalities such as Jensen's inequality. Under mild regularity conditions, we show that all such extended gradient maps can be recovered as the usual supergradients of a nonnegative concave function on the space of probability distributions. This embedding provides a universal geometry for all such optimal transports and an unexpected connection with information geometry of exponential families of distributions.

Key words : Optimal transport; exponentially concave functions; information geometry; exponential families.

1. INTRODUCTION

Fix $d \in \mathbb{N}$, where $\mathbb{N}$ is the set of natural numbers. Fix a Borel probability measure μ_0 on $\mathbb{R}^d$. This will be called the base measure. For any μ_0 integrable function f , we will denote the μ_0 expectation of f by $\mu_0(f(x))$ or $\mu_0(f)$. Let Λ_0 denote the cumulant generating function of μ_0 . That is, for any $\theta \in \mathbb{R}^d$, we have

$$\Lambda_0(\theta) := \log \mu_0 \left(e^{\langle \theta, x \rangle} \right).$$

We will throughout assume that its domain is the entire space, i.e., $\text{dom}(\Lambda_0) = \mathbb{R}^d$.

¹Dedicated to Prof. B. V. Rao. This research is partially supported by NSF grant DMS-1308340 and DMS-1612483.

It is well-known that Λ_0 is a convex function, which is strict whenever the support of μ_0 is not a singleton. Moreover, $\Lambda_0(0) = 0$. For $\theta, \psi \in \mathbb{R}^d$, we define a *cost function* on $\mathbb{R}^d \times \mathbb{R}^d$ given by

$$c(\theta, \psi) := \Lambda_0(\theta - \psi) = \log \mu_0 \left(e^{\langle \theta - \psi, x \rangle} \right).$$

Suppose P and Q are two probability measures on $\mathbb{R}^d$. Let $\Pi(P, Q)$ be the set of couplings of (P, Q) , i.e., the set of joint probability distributions on $\mathbb{R}^d \times \mathbb{R}^d$ whose first marginal is P and the second is Q . Consider the Monge-Kantorovich optimal transport problem of transporting P to Q with cost c . That is, we find the minimizer in the optimization problem

$$\min_{R \in \Pi(P, Q)} R[\Lambda_0(\theta - \psi)], \quad (\theta, \psi) \sim R. \quad (1)$$

The optimal coupling R , if exists, is said to solve the Monge problem if ψ is a deterministic function of θ , R almost surely. Let us consider two known examples.

Example 1 : Let $\mu_0 \sim N(0, I)$ be the standard Gaussian distribution on $\mathbb{R}^d$. In this case, $\Lambda_0(\theta - \psi) = \|\theta - \psi\|^2 / 2$ and the optimal coupling between P and Q is the well-known Wasserstein-2 or W_2 coupling.

Example 2 : Let $e_i, i \in [d]$, be the standard basis in $\mathbb{R}^d$. Additionally, let e_0 denote the zero vector in $\mathbb{R}^d$. Let μ_0 denote the probability measure that puts mass $1/(d+1)$ at each e_i . That is, if δ refers to the unit Dirac delta mass, then

$$\mu_0 = \frac{1}{d+1} \sum_{i=0}^d \delta_{e_i}.$$

Thus $\Lambda_0(\theta) = \log \left[1 + \sum_{i=1}^d e^{\theta_i} \right] - \log(d+1)$. The corresponding optimal transport problem has appeared recently in [5-7] in connection with portfolio theory and information geometry.

The modern theory of optimal transport is a vast area with important applications in analysis, geometry, and probability. We refer the reader to the book [1] for an introduction.

We now describe the solution to the optimization problem (1) as done in [4]. The optimal coupling can be described by an extension of the usual notion of concavity that we describe below. See, for example, [1, Definition 1.8 and eqn. 1.3, Chapter 1].

Definition 1 — A function $\psi : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{-\infty\}$ is said to be Λ_0 concave if

$$\psi(x) = \inf_{y \in \mathbb{R}^d} [\Lambda_0(x - y) - \psi(y)]$$

for some $\rho : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{-\infty\}$. The Λ_0 superdifferential of a function $\psi : \mathbb{R}^d \mapsto \mathbb{R} \cup \{-\infty\}$ at a point $\theta \in \mathbb{R}^d$ is given by the set of points $(\theta, y) \in \mathbb{R}^d \times \mathbb{R}^d$ such that

$$\psi(v) \leq \psi(\theta) + \Lambda_0(v - y) - \Lambda_0(\theta - y), \quad \text{for all } v \in \mathbb{R}^d.$$

The set of Λ_0 superdifferential pairs (θ, y) of ψ will be denoted by $\partial^{\Lambda_0}\psi$. In particular, $y \in \partial^{\Lambda_0}\psi(\theta)$ will mean $(\theta, y) \in \partial^{\Lambda_0}\psi$.

Define the Λ_0 -transform (or, conjugate) $\psi^0 : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{-\infty\}$ by

$$\psi^0(y) = \inf_{x \in \mathbb{R}^d} [\Lambda_0(x - y) - \psi(x)]. \quad (2)$$

Then, ψ^0 is dual Λ_0 concave, and $\psi(\theta) + \psi^0(y) = \Lambda_0(\theta - y)$ if and only if $(\theta, y) \in \partial^{\Lambda_0}\psi$. See [1, Section 3.1] for more details.

The following theorem is taken from [1, Theorem 1.13, Chapter 1] where it is referred to as the *Fundamental theorem of optimal transport*.

Theorem 1 — *For a joint distribution $R \in \Pi(P, Q)$ to be optimal for the minimization problem (1) it is sufficient that there exists a Λ_0 concave function ψ such that $\max(\psi, 0)$ is P integrable and the support of R is a subset of $\partial^{\Lambda_0}\psi$. The above is also necessary if Λ_0 is bounded below.*

In our context, Λ_0 is bounded below by zero if the mean of μ_0 is zero.

Although these extended notions of concavity and superdifferentiability gives us a theoretical picture of the optimal coupling, they can be nonintuitive and do not satisfy useful inequalities satisfied by the superdifferentials of ordinary concave functions. However, in this article we show that they can all be embedded as superdifferentials of an actual concave function on the space of probability distributions that are absolutely continuous with respect to μ_0 .

To understand the idea, one first needs to note that $\mathbb{R}^d$ itself corresponds to a space of probability distributions; as parameters of the natural exponential family of probability distributions with base measure μ_0 . For example, $\theta \mapsto N(\theta, I)$ embeds $\mathbb{R}^d$ as the mean of a normal distribution. This corresponds to Example 1. Hence, if we consider the convex set of all probability distributions that are absolutely continuous with respect to $N(0, I)$, we have an embedding of $\mathbb{R}^d$ into that set. What we show is that there exists a concave function on this convex set of probability measures whose superdifferentials (taken in the usual way) are also probability measures that are members of the same exponential family. The parameters of these superdifferentials recover the optimal coupling for problem (1).

This idea is closely related to our work [6] where we describe this result only for Example 2. It leads to a new information geometry based on the concept of L -divergence that extends the classical information geometry of Kullback-Leibler divergence (that corresponds to Example 2). See [2] for an introduction to information geometry. I expect a similar consequence for this general construction.

2. PRELIMINARIES

2.1 Natural exponential families

We now define the exponential family of models with base measure μ_0 . See [2] for more details and historical references. Let $M_1 := M_1(\mathbb{R}^d)$ be the space of Borel probability measures on $\mathbb{R}^d$.

Definition 2 — For every $\theta \in \mathbb{R}^d$, define μ_θ by the exponential change of measure

$$\frac{d\mu_\theta}{d\mu_0}(x) = \exp(\langle \theta, x \rangle - \Lambda_0(\theta)).$$

Then $\mu_\theta \in M_1(\mathbb{R}^d)$ and the collection $\{\mu_\theta, \theta \in \mathbb{R}^d\}$ is called the natural exponential family of models with base measure μ_0 .

For example, when the base measure is the multidimensional standard normal, for any θ , the probability measure μ_θ is the multidimensional Gaussian law with mean θ and identity covariance. Now, consider the exponential family generated by μ_0 in Example 2. For any $\theta \in \mathbb{R}^d$, notice that μ_θ is still supported on the discrete set $\{e_i, i = 0, 1, \dots, d\}$. The mass it puts on e_i is proportional to $\exp(\langle \theta, e_i \rangle)$. Thus,

$$\mu_\theta(e_i) = \frac{e^{\theta_i}}{1 + \sum_{i=1}^d e^{\theta_i}}, \text{ for } i = 1, 2, \dots, d, \text{ and } \mu_\theta(e_0) = \frac{1}{1 + \sum_{i=1}^d e^{\theta_i}}.$$

As θ varies in $\mathbb{R}^d$ the natural exponential family corresponds to the open unit simplex of dimension d (i.e., with $(d + 1)$ coordinates).

Lemma 2 — For any $\theta, \gamma \in \mathbb{R}^d$, we have $\log \mu_\theta(e^{\langle \gamma, x \rangle}) = \Lambda_0(\gamma + \theta) - \Lambda_0(\theta)$.

PROOF : By a change of measure, $\log \mu_\theta(e^{\langle \gamma, x \rangle}) = \log \mu_0(e^{\langle \theta + \gamma, x \rangle}) - \Lambda_0(\theta)$. $\square$

2.1 Topological preliminaries

To perform convex analysis on the infinite dimensional space of probability measures we require a proper locally convex topological vector space (LCTVS). Let $M(\mathbb{R}^d)$ be the space of all finite signed measures on $\mathbb{R}^d$ which have finite exponential moments of all orders. For $\nu \in M(\mathbb{R}^d)$, we will follow the usual notation of denoting by ν^+ , ν^- , and $|\nu|$, the positive part, the negative part, and the absolute

value (or, variation) of the signed measure ν . For $i \in [d] := \{1, 2, \dots, d\}$ and $j \in \mathbb{N} \cup \{0\}$, define the set of functions from $M_0(\mathbb{R}^d)$:

$$p_j^{(i)}(\nu) = |\nu| \left(e^{j|x_i|} \right), \quad \nu \in \mathcal{V}.$$

Consider the vector space $\mathcal{V}$ of elements ν in $M(\mathbb{R}^d)$ such that $p_j^{(i)}(\nu) < \infty$ for every $i \in [d]$ and $j \in \mathbb{N}$. We now define a locally convex topology on $\mathcal{V}$.

Lemma 3 — The family $(p_j^{(i)}, i \in [d], j \geq 0)$ is a collection of seminorms that is total. Hence, it induces a locally convex Hausdorff metrizable topology on $\mathcal{V}$. In this topology, we have $\lim_{n \rightarrow \infty} \nu_n = \nu$ if and only if

$$\lim_{n \rightarrow \infty} |\nu_n| \left(e^{\langle \theta, x \rangle} \right) = |\nu| \left(e^{\langle \theta, x \rangle} \right), \quad \forall \theta \in \mathbb{R}^d. \quad (3)$$

Hence, when each ν_n is a probability measure, convergence in this topology is equivalent to the convergence of all exponential moments.

PROOF : It is obvious that every $p_j^{(i)}$ is a seminorm. In fact, it is a norm, and thus, the family is total. Therefore, there is a corresponding locally convex Hausdorff topology which is the smallest topology under which each $p_j^{(i)}$ is continuous. The topology is metrizable since the family of seminorms is countable.

We now show (3). To see the *only if* part, consider some $\theta \in \mathbb{R}^d$ and let $j := \lceil \|\theta\|_1 \rceil$. Then

$$0 \leq e^{\langle \theta, x \rangle} \leq e^{j \max_i |x_i|} \leq \sum_{i=1}^d e^{j|x_i|}.$$

Thus the function $\nu \mapsto |\nu| \left(e^{\langle \theta, x \rangle} \right)$ is continuous in the locally convex topology constructed above which gives us the *only if* part.

For the *if* part, fix $i \in [d]$ and $j \geq 0$, and define $\theta^+ = je_i$ and $\theta^- = -je_i$. Now,

$$e^{j|x|_i} \leq e^{\langle \theta^+, x \rangle} + e^{\langle \theta^-, x \rangle},$$

whereby convergence of the right side gives convergence of the left. $\square$

We shall call this topology $\mathcal{E}$. Consider the dual space $\mathcal{V}^*$ of all linear $\mathcal{E}$ continuous functions on $\mathcal{V}$. Equip $\mathcal{V}^*$ with the weak* topology $\mathcal{E}^*$. This gives us a pair of locally convex topological vector spaces $(\mathcal{V}, \mathcal{V}^*)$ such that if $\nu \in \mathcal{V}$ and $h \in \mathcal{V}^*$ then the bilinear function $\langle h, \nu \rangle := \nu(h)$ is $\mathcal{E}$ continuous on $\mathcal{V}$ and $\mathcal{E}^*$ continuous on $\mathcal{V}^*$.

It follows from (3) that the function $x \mapsto e^{\langle \theta, x \rangle}$, for any $\theta \in \mathbb{R}^d$, can be thought of as an element in $\mathcal{V}^*$. We will use this identification without further remark.

3. EXPONENTIALLY CONCAVE FUNCTIONS AND Λ_0 -CONCAVE FUNCTIONS

Let Ω be a convex subset of M_1 . We generalize the definition of exponentially concave functions from [5, 6] where the reader can find more references and applications to various other fields.

Definition 3 — A function $\varphi : \Omega \rightarrow \mathbb{R} \cup \{-\infty\}$ will be called exponentially concave if $\exp(\varphi)$ is a nonnegative concave function on the convex set Ω . In particular, φ is itself concave.

Consider the base measure μ_0 and let Ω_0 denote the convex set of probability measures on $\mathbb{R}^d$ that are absolutely continuous with respect to μ_0 . Let φ denote an exponentially concave function on Ω_0 . We will throughout assume that φ is proper and is continuous in the interior of its effective domain, which is non-empty. By [3, Proposition 5.2], at any μ in the interior of the effective domain, the set of superdifferentials $\partial\varphi(\mu)$ is non-empty. Consider one such point $\mu \in \Omega_0$ and let $\mu^* \in \mathcal{V}^*$ be an element in the superdifferential $\partial\varphi(\mu)$.

Note that, since Ω_0 is a set of probability measures, the set of superdifferentials is closed under addition by a constant. Thus if we replace μ^* by $\widetilde{\mu^*} := \mu^* + 1 - \langle \mu^*, \mu \rangle$, then the latter is an element in $\partial\varphi(\mu)$ and satisfies $\langle \widetilde{\mu^*}, \mu \rangle = 1$. To keep our notations simple, we will assume that $\langle \mu^*, \mu \rangle = 1$. We will throughout use this normalization.

Our next lemma shows that supergradients of exponential concave functions can be expressed in terms of a probability measure. This generalizes [7, Proposition 5] where it is shown for the unit simplex (Example 2).

Lemma 4 — There exists an element $\pi = \pi(\mu) \in \Omega_0$ that induces μ^* in the following way:

$$\langle \mu^*, \nu \rangle = \pi \left(\frac{d\nu}{d\mu} \right), \quad \text{for all } \nu \ll \mu. \quad (4)$$

We will denote μ^* by $d\pi/d\mu$ in view of the above lemma. Consistent with the terminology developed in [7, 6] we will call the map $\pi : \Omega_0 \rightarrow \Omega_0$ to be a portfolio map generated by φ .

PROOF OF LEMMA 4 : Let $\Phi = \exp(\varphi)$. Then Φ is a nonnegative concave function on Ω_0 and $\Phi(\mu)\mu^* \in \partial\Phi(\mu)$. Consider $\nu \in \Omega_0$. Since $\Phi \geq 0$ and concave, for $0 < t < 1$, we get

$$-\Phi(\mu) \leq \Phi(\mu + t(\nu - \mu)) - \Phi(\mu) \leq \langle \Phi(\mu)\mu^*, t(\nu - \mu) \rangle = t\Phi(\mu) \langle \mu^*, \nu - \mu \rangle. \quad (5)$$

Since $\Phi(\mu) > 0$, we divide both sides by it and take $t \uparrow 1$ to get

$$\langle \mu^*, \nu \rangle \geq \langle \mu^*, \mu \rangle - 1 = 0, \quad (6)$$

due to our chosen normalization.

Consider the space of continuous functions supported on $B_K := \{x : \|x\| \leq K\}$. Let g be one such function such that $g \geq 0$ and $\mu(g) = 1$. Then define $\mu_g \in \Omega_0$ by the change of measure $d\mu_g/d\mu = g$. Consider the map $g \mapsto \Gamma(g) := \langle \mu^*, \mu_g \rangle$. It can be extended to all continuous g supported on B_K by using linearity. Since all our functions are supported on B_K , uniform convergence of functions imply convergence in the $\mathcal{E}$ topology for the corresponding sequence of probability measures. Thus Γ is a continuous map in the uniform topology that takes nonnegative functions to nonnegative values by (6). By the Riesz representation theorem, there exists a nonnegative measure π_K such that $\Gamma(g) = \pi_K(g)$. Consistency over K gives us a π whose restriction to B_K is π_K . That π is a probability is due to our normalization. This shows (6) for $\nu = \mu_g$ for all bounded continuous g . Generalization to non-continuous g with bounded support follows by uniform continuous approximation. If g has unbounded support, consider a collection of increasing Borel maps g_i , $i \in \mathbb{N}$, such that $\lim_{i \rightarrow \infty} g_i = g$. By monotone convergence theorem, $\lim_{i \rightarrow \infty} \mu_{g_i} = \mu_g$ in the $\mathcal{E}$ topology. This proves (4) for all $\nu \ll \mu$. $\square$

We now consider a special class of exponentially concave functions. Let $h : \mathbb{R}^d \rightarrow [-\infty, \infty)$ be a measurable function. Define φ on Ω_0 by

$$\varphi(\mu) = \inf_{\theta \in \mathbb{R}^d} \left[\log \mu \left(e^{\langle -\theta, x \rangle} \right) - h(\theta) \right], \quad \mu \in \Omega_0. \quad (7)$$

Clearly, φ is exponentially concave. Our next result connects exponentially concave functions with Λ_0 concave functions. Recall the natural exponential family with base measure μ_0 . Notice that the natural exponential family with base measure μ_0 is a subset of Ω_0 .

Theorem 5 — Consider φ from (7). Define the function $\psi : \mathbb{R}^d \rightarrow \mathbb{R}$ by

$$\psi(\alpha) = \varphi(\mu_\alpha) + \Lambda_0(\alpha), \quad \alpha \in \mathbb{R}^d. \quad (8)$$

Then $\psi(\cdot)$ is Λ_0 concave on $\mathbb{R}^d$. Conversely, suppose that a Λ_0 concave function ψ is given. Let ψ^0 denote its Λ_0 conjugate. Define an exponentially concave function on Ω_0 by

$$\varphi(\mu) := \inf_{\theta \in \mathbb{R}^d} \left[\log \mu \left(e^{-\langle \theta, x \rangle} \right) - \psi^0(\theta) \right], \quad \mu \in \Omega_0. \quad (9)$$

Then, for all $\alpha \in \mathbb{R}^d$, we have

$$\varphi(\mu_\alpha) = \psi(\alpha) - \Lambda_0(\alpha). \quad (10)$$

PROOF : The first part follows by direct evaluation. For any $\alpha \in \mathbb{R}^d$, we have

$$\begin{aligned} \varphi(\mu_\alpha) &:= \inf_{\theta \in \mathbb{R}^d} \left[\log \mu_\alpha \left(e^{-\langle \theta, x \rangle} \right) - h(\theta) \right] \\ &= \inf_{\theta \in \mathbb{R}^d} [\Lambda_0(\alpha - \theta) - h(\theta)] - \Lambda_0(\alpha). \end{aligned} \quad (11)$$

Thus $\psi(\alpha) = \varphi(\mu_\alpha) + \Lambda_0(\alpha)$ satisfies the representation in Definition 1. The converse is a consequence of (11) replacing h by ψ^0 . $\square$

We now show that the Λ_0 gradient of ψ is related to the portfolio map of φ . Let (V, V^*) be any pair of LCTVS and its topological dual. For each θ in some collection Θ , let $f_\theta : V \rightarrow \overline{\mathbb{R}}$ be a proper concave function. Consider the proper concave function $F : V \rightarrow \overline{\mathbb{R}}$ given by $F(v) := \inf_{\theta \in \Theta} f_\theta(v)$. Let $\partial f(u)$ refer to the set of superdifferentials of a concave function f at a point u .

Lemma 6 — For $v_0 \in \text{dom}(F)$, $\partial f_\theta(v_0) \subseteq \partial F(v_0)$, for any θ such that $F(v_0) = f_\theta(v_0)$.

PROOF : The proof is trivial. Pick any $v_0^* \in \partial f_\theta(v_0)$ and any other $u \in V$. Then, by the superdifferentiability of f_θ we get

$$F(u) \leq f_\theta(u) \leq f_\theta(v_0) + \langle v_0^*, u - v_0 \rangle = F(v_0) + \langle v_0^*, u - v_0 \rangle.$$

Varying $u \in V$ proves the claim. $\square$

Theorem 7 — Consider the exponentially concave function (9). For $\alpha \in \mathbb{R}^d$, let θ be an element in $\mathbb{R}^d$ such that $(\alpha, \theta) \in \partial^{\Lambda_0} \psi$. Then, there exists an element $\mu_\alpha^* \in \mathcal{V}^*$ such that μ_α^* is a supergradient of φ at μ_α and for any $\nu \in \mathcal{V}$, we have $\langle \mu_\alpha^*, \nu \rangle = \nu(h_\alpha(x))$, where

$$h_\alpha(x) := \exp(-\langle \theta, x \rangle - \Lambda_0(\alpha - \theta) + \Lambda_0(\alpha)).$$

In particular, the probability measure $\pi_\alpha := \pi(\mu_\alpha)$ in Lemma 4 is given by $\mu_{\alpha-\theta}$.

PROOF : We apply Lemma 6 to the choice: $f_\theta(\mu) = \log \mu(e^{-\langle \theta, x \rangle}) - \psi^0(\theta)$, $\theta \in \mathbb{R}^d$. By (10) and the assumption that $(\alpha, \theta) \in \partial^{\Lambda_0} \psi$, we have

$$\begin{aligned} \varphi(\mu_\alpha) &= \psi(\alpha) - \Lambda_0(\alpha) = \Lambda_0(\alpha - \theta) - \Lambda_0(\alpha) - \psi^0(\theta) \\ &= \log \mu_\alpha(e^{-\langle \theta, x \rangle}) - \psi^0(\theta) = f_\theta(\mu_\alpha). \end{aligned}$$

The first equality in the second line in the above display is due to Lemma 2. From (9) we get $\varphi(\mu) = \inf_\theta f_\theta(\mu)$. However from the above display, we also get $F(\mu_\alpha) = f_\theta(\mu_\alpha)$ for each $(\alpha, \theta) \in \partial^{\Lambda_0} \psi$.

By Lemma 6, it is sufficient to show the existence of the superdifferential with the claimed properties for f_θ . In fact, we find the Gâteaux derivative of f_θ at μ_α . To do this, fix $\nu \in \Omega_0$ and consider the family of probability measures $\mu(t) = (1-t)\mu_\alpha + t\nu$, for $0 \leq t \leq 1$. Clearly $\mu(0) = \mu_\alpha$. We are interested in the limit:

$$f'_\theta(\mu_\alpha) := \lim_{t \rightarrow 0+} \frac{f_\theta(\mu(t)) - f_\theta(\mu_\alpha)}{t} = \frac{d}{dt} \Big|_{t=0+} \log [\mu(t) (\exp(-\langle \theta, x \rangle))].$$

The last expression, via differentiation within the expectation, gives us

$$\frac{\nu(\exp(-\langle \theta, x \rangle))}{\mu_\alpha(\exp(-\langle \theta, x \rangle))} - 1 = \nu(h_\alpha(x)) - 1.$$

That proves that f_θ is Gâteaux differentiable at μ_α with the above derivative.

Notice that $\mu_\alpha(h_\alpha(x)) = 1$ by Lemma 2. Hence $\nu(h_\alpha(x)) - 1 = (\nu - \mu_\alpha)(h_\alpha(x))$. Since $t \mapsto f_\theta(\mu(t))$ is concave, we also get the following expression of a supergradient:

$$f_\theta(\nu) \leq f_\theta(\mu_\alpha) + (\nu - \mu_\alpha)(h_\alpha(x)).$$

By Lemma 6 this proves the claim regarding the existence of the supergradient of φ at μ_α . The formula for π_α follows immediately from the Radon-Nikodym derivative

$$\frac{d\pi_\alpha}{d\mu_\alpha}(x) = h_\alpha(x),$$

and an application of Lemma 4. □

Thus the parameter of the portfolio map at μ_α is $\alpha - \theta$, where $(\alpha, \theta) \in \partial^{\Lambda_0} \varphi$. Hence, as mentioned in the Introduction, if the support of the optimal transport includes the point (α, θ) , we see it embedded as the pair of elements $(\mu_\alpha, \mu_{\alpha-\theta})$ in the subdifferential of a true concave function on the space of probability measures.

ACKNOWLEDGEMENT

I am indebted to Prof. B. V. Rao for teaching me probability. He remains one of the finest teachers I have encountered in my life and interactions with him are some of my fondest memories from ISI. I wish him all the best on the occasion of his 70th birthday. I also thank an anonymous referee for useful comments.

REFERENCES

1. Luigi Ambrosio and Nicola Gigli, *A user's guide to optimal transport*, In: *Modelling and Optimisation of Flows on Networks*, 1-155. Springer (2013).
2. Shun-ichi Amari, *Information geometry and its applications*, Springer (2016).
3. Ivar Ekeland and Roger Teraam, *Convex analysis and variational problems*, volume 1 of *Studies in Mathematics and its Applications*, North-Holland American Elsevier, (1976).
4. Wilfrid Gangbo and Robert J. McCann, The geometry of optimal transportation, *Acta Math.*, **177** (1996), 113-161.

5. Soumik Pal, *Exponentially concave functions and high dimensional stochastic portfolio theory*, ArXiv e-prints 1603.01865, (2016).
6. S. Pal and T.-K. L. Wong, *Exponentially concave functions and a new information geometry*, Preprint available on arXiv 1605.05819, (2016).
7. Soumik Pal and Ting-Karn Leonard Wong, The geometry of relative arbitrage, *Mathematics and Financial Economics*, **10** (2016), 263-293.