Hessian Curvature and Optimal Transport

Gabriel Khan(✉) and Jun Zhang

University of Michigan, Ann Arbor, USA
{gabekhan,junz}@umich.edu

Abstract. We consider the problem of optimal transport where the cost function is given by a $\mathcal{D}_\Psi^{(\alpha)}$-divergence for some convex function Ψ [21], where $\alpha = \pm 1$ gives the Bregman divergence. For costs of this form, we introduce a new complex geometric interpretation of the optimal transport problem by considering an induced Sasaki metric on the tangent bundle of the domain of Ψ. In this framework, the Ma-Trudinger-Wang (MTW) tensor [12] is proportional to the orthogonal bisectional curvature. This geometric framework for optimal transport is complementary to the pseudo-Riemannian approach of Kim and McCann [10].

1 Introduction

Optimal transport is a classical field of mathematics which dates back to the work of Monge in 1781 [13]. In its original formulation, it considered finding the most efficient way to move piles of rubble from one configuration to another. In the modern framework, we consider X and Y as Borel subsets of two metric spaces, equipped with probability measures μ and ν, respectively, and a lower-semicontinuous cost function $c : X \times Y \to \mathbb{R}$. The optimal transport problem is to find the non-negative measure γ on $X \times Y$ which minimizes

$$W_c(\mu, \nu) = \min_{\gamma \in \Gamma(\mu,\nu)} \int_{X \times Y} c(x, y) d\gamma(x, y).$$

Here, $\Gamma(\mu, \nu)$ is the set of joint probabilities with the same marginal distributions as $\mu \otimes \nu$ and γ is known as the *optimal coupling.*

In this paper, we study the regularity theory of this problem, specializing to the case where the cost function is given by a $\mathcal{D}_\Psi^{(\alpha)}$-divergence.

Definition 1 **($\mathcal{D}_\Psi^{(\alpha)}$-divergence).** *Let $\Psi : M \to \mathbb{R}$ be a convex function on a convex domain M in Euclidean space. For two points $x, y \in M$ and $\alpha \in (-1, 1)$, a $\mathcal{D}_\Psi^{(\alpha)}$-divergence is a function of the form*

$$\mathcal{D}_\Psi^{(\alpha)}(x, y) = \frac{4}{1-\alpha^2}\left[\frac{1-\alpha}{2}\Psi(x) + \frac{1+\alpha}{2}\Psi(y) - \Psi\left(\frac{1-\alpha}{2}x + \frac{1+\alpha}{2}y\right)\right].$$

© Springer Nature Switzerland AG 2019
F. Nielsen and F. Barbaresco (Eds.): GSI 2019, LNCS 11712, pp. 423–430, 2019.
https://doi.org/10.1007/978-3-030-26980-7_44

These divergences were introduced by the second author [21] and form a one-parameter family of statistical divergences. As α converges to either ± 1, the $\mathcal{D}_{\Psi}^{(\alpha)}$-divergence converges to the Bregman divergence [1]. Bregman divergences play an important role in information geometry as they provide a generalization of distance functions which satisfy the generalized Pythagorean theorem. Due to this interpolation property, $\mathcal{D}_{\Psi}^{(\alpha)}$-divergences form a natural class worthy of investigation.

Our main results provide a new geometric interpretation for the necessary conditions to ensure that the associated optimal transport is smooth. More concretely, we want to understand whether the rubble is moved in a continuous way so that nearby piles remain close after transport. In this paper, we give a summary of our results, omitting proofs and a more complete exposition. We refer the interested reader to [9] for a complete description of our results, accompanied with proofs and extensive exposition.

2 Preliminaries on Optimal Transport

We briefly discuss some background on the regularity theory of optimal transport. For a more complete overview of the subject, we recommend the book by Villani [19] and the survey paper of De Philippis and Figalli [3]. In the following, we use c to refer to the cost function and $c_{I,J}$ to denote $\partial_{x^I}\partial_{y^J} c$ for multi-indices I and J. Furthermore, $c^{i,j}$ is the matrix inverse of the mixed derivative $c_{i,j}$.

In order to state the background results, it is necessary to first define the c-exponential map, which plays a crucial role throughout.

Definition 2 (c-exponential map). *For any $x \in X, y \in Y, p \in \mathbb{R}^n$, the c-exponential map satisfies the following identity.*

$$c\text{-}\exp_x(p) = y \iff p = -c_x(x, y).$$

Our primary interest is in the case when the optimal coupling γ is supported on the graph of a function. In this case, the optimal transport is said to be *deterministic* and the associated function is known as the *optimal map*. The following theorem, originally proved by Brenier [2] and extended by Gangbo and McCann [8], provides sufficient conditions for optimal transport to be deterministic and shows that the associated optimal maps are induced by solutions to Monge-Ampere type equations.

Theorem 1. *Let X and Y be two open subsets of $\mathbb{R}^n$ and consider a cost function $c : X \times Y \to \mathbb{R}$. Suppose that $d\mu$ is a smooth probability density supported on X and that $d\nu$ is a smooth probability density supported on Y. Suppose that the following conditions hold:*

1. *The cost function c is of class C^4 with $\|c\|_{C^4(X\times Y)} < \infty$*
2. *For any $x \in X$, the map $Y \ni y \to c_x(x, y) \in \mathbb{R}^n$ is injective.*
3. *For any $y \in Y$, the map $X \ni x \to c_y(x, y) \in \mathbb{R}^n$ is injective.*

4. $\det(c_{x,y})(x,y) \neq 0$ *for all* $(x,y) \in X \times Y$.

Then there exists a c-convex function $u : X \to \mathbb{R}$ *such that the map* $T_u : X \to Y$ *defined by* $T_u(x) := c\text{-}\exp_x(\nabla u(x))$ *is the unique optimal transport map sending* μ *onto* ν*. Furthermore,* T_u *is injective* $d\mu$*-a.e.,*

$$|\det(\nabla T_u(x))| = \frac{d\mu(x)}{d\nu(T_u(x))} \qquad d\mu - a.e., \tag{1}$$

and its inverse is given by the optimal transport map sending ν *onto* μ*.*

The regularity problem for optimal transport studies the smoothness of the potential u. For this question, most of the initial work was done for the squared-distance cost $c(x,y) = \frac{1}{2}\|x-y\|^2$ in Euclidean space, known as the 2-Wasserstein distance. For more general cost functions, the breakthrough work was done by Ma, Trudinger and Wang [12], who gave three conditions that ensure smoothness for the solutions of Monge-Ampere equations. In this paper, we use a modified version of their result, originally proved by Trudinger and Wang [18].

Theorem 2. *Suppose that* $c : X \times Y \to \mathbb{R}$ *satisfies the hypothesis of the previous theorem, and that the smooth densities* $d\mu$ *and* $d\nu$ *are bounded away from zero and infinity on their respective supports* X *and* Y*. Suppose further that the following holds:*

1. X *and* Y *are smooth.*
2. The domain X *is strictly c-convex relative to the domain* Y*.*
3. The domain Y *is strictly* c^**-convex relative to the domain* X*.*
4. The following condition (known as MTW(0)) holds:

For all vectors $\xi, \eta \in \mathbb{R}^n$ *with* $\xi \perp \eta$*, the following inequality holds.*

$$\mathfrak{S}(\xi,\eta) := \sum_{i,j,k,l,p,q,r,s} (c_{ij,p}c^{p,q}c_{q,rs} - c_{ij,rs})c^{r,k}c^{s,l}\xi^i\xi^j\eta^k\eta^l \geq 0 \tag{2}$$

Then $u \in C^\infty(\overline{X})$ *and* $T : \overline{X} \to \overline{Y}$ *is a smooth diffeomorphism, where* $T(x) = c\text{-}\exp_x(\nabla u(x))$*.*

We will discuss the assumptions of Theorem 2 in a bit more detail. The first condition is self-explanatory, while the second and third define the proper notions of convexity, which are necessary to establish regularity of optimal transport [11].

Definition 3 (c-segment). *A c-segment in* X *with respect to a point* y *is a solution set* $\{x\}$ *to* $c_y(x,y) = z$ *for* z *on a line segment in* $\mathbb{R}^n$*. A* c^**-segment in* Y *with respect to a point* x *is a solution set* $\{y\}$ *to* $c_x(x,y) = z$ *for* z *on a line segment in* R^n*.*

Definition 4 (c-convexity). *A set* E *is c-convex relative to a set* E^* *if for any two points* $x_0, x_1 \in E$ *and any* $y \in E^*$*, the c-segment relative to* y *connecting* x_0 *and* x_1 *lies in* E*. Similarly we say* E^* *is* c^**-convex relative to* E *if for any two points* $y_0, y_1 \in E^*$ *and any* $x \in E$*, the* c^**-segment relative to* x *connecting* y_0 *and* y_1 *lies in* E^**.*

Finally, we discuss the inequality (2), which is known as the $MTW(0)$ condition, and is a weakened version of the $MTW(\kappa)$ condition.

Definition 5 ($MTW(\kappa)$). *A cost function c satisfies the $MTW(\kappa)$ condition if for any orthogonal vector-covector pair η and ξ, $\mathfrak{S}(\eta,\xi) \geq \kappa|\eta|^2|\xi|^2$ for $\kappa > 0$.*

Ma, Trudinger and Wang's original work used $MTW(\kappa)$, and this stronger assumption is used in many applications. There is another strengthening of the $MTW(0)$ assumption that appears in the literature.

Definition 6 (Non-negative cross-curvature). *A cost function c has non-negative (resp. strictly positive) cross-curvature if, for any vector-covector pair η and ξ,*

$$\mathfrak{S}(\eta,\xi) \geq 0 \ \textit{(resp. } \kappa|\eta|^2|\xi|^2).$$

Non-negative cross-curvature is stronger than $MTW(0)$, as the non-negativity must hold for all pairs η and ξ, not simply orthogonal ones. This notion was introduced by Figalli, Kim, and McCann [5] to study a problem in microeconomics and in later work, they showed that stronger regularity for optimal maps can be proven with this assumption [6].

The geometric significance of these notions is a topic of active research. Although it is not immediately clear, $\mathfrak{S}$ is in fact tensorial (coordinate-invariant) and transforms quadratically in η and ξ [10]. On a Riemannian manifold, Loeper [11] gave some insight into the behavior of the MTW tensor by showing that for the 2-Wasserstein distance, the tensor is proportional to the sectional curvature on the diagonal. This was extended by Kim and McCann, who gave a pseudo-Riemannian framework for optimal transport [10], in which the MTW-tensor becomes the curvature of light-like planes.

3 Geometry of TM

In order to associate optimal transport with Kähler geometry, we consider the tangent bundle TM where M is the domain of Ψ. On any Riemannian manifold (M,g) endowed with an affine connection D, it is possible to induce TM with an almost Hermitian structure known as the Sasaki metric [4]. For brevity, we will not review the construction here, but complete details can be found in the paper by Satoh [16].

We are primarily interested in the case where TM is Kähler, which occurs when M is Hessian (also known as *affine-Kähler*). There are two equivalent definitions for such manifolds; with the former definition primarily used in differential geometry and the latter primarily used in information geometry. For details on how these definitions are equivalent, we refer readers to the book by Shima [17].

Definition 7 (Differential geometric). *A Riemannian manifold* (M, g) *is said to be Hessian if there is an atlas of local coordinates* $\{u^i\}_{i=1}^n$ *so that for each coordinate chart, there is a convex potential* Ψ *such that*

$$g_{ij} = \frac{\partial^2 \Psi}{\partial u^i u^j}.$$

Furthermore, the transition maps between these coordinate charts are affine.

Definition 8 (Information geometric). *A Riemannian manifold* (M, g) *is said to be Hessian if it admits dually flat connections. Namely, it admits two flat (torsion- and curvature-free) connections* D *and* D^* *satisfying*

$$\mathcal{X}(g(\mathcal{Y}, \mathcal{Z})) = g(D_{\mathcal{X}}\mathcal{Y}, \mathcal{Z}) + g(\mathcal{Y}, D^*_{\mathcal{X}}\mathcal{Z}) \tag{3}$$

for all vector fields $\mathcal{X}$, $\mathcal{Y}$, *and* $\mathcal{Z}$.

For our purposes, we are interested in the curvature of this metric, which can be derived using the work of Satoh [16].

Proposition 1. *Let* (M, g, D) *be a Hessian manifold with metric* g *and flat connection* D. *Suppose* $\{x^i\}$ *are the coordinates where the Christoffel symbols of* D *vanishes and* Ψ *is the Hessian potential of* g. *In the associated holomorphic coordinates* $\{z^i\}$ *on* TM, *the holomorphic bisectional curvature of the Sasaki metric satisfies the following identity:*

$$\tilde{R}_{\tilde{g}^D}\left(\partial_{z^i}, \overline{\partial}_{z^j}, \partial_{z^k}, \overline{\partial}_{z^l}\right) = -\frac{1}{2}\Psi_{ijkl} + \frac{1}{2}\Psi^{rs}\Psi_{iks}\Psi_{jlr}.$$

It is worth noting that the holomorphic bisectional curvature is negative to what Shima defined as the *Hessian curvature* [17, p. 38].

4 Our Results

Our main result is to relate the MTW tensor to the bisectional curvature of the Sasaki metric. To do so, we note that if $c : M \times M \to \mathbb{R}$ is a $D_{\Psi}^{(\alpha)}$-divergence, then the MTW tensor takes the following form:

$$\mathfrak{S}_{(x,y)}(\xi, \eta) = \frac{1-\alpha^2}{4}\left(\Psi_{ijp}\Psi_{rsq}\Psi^{pq} - \Psi_{ijrs}\right)\Psi^{rk}\Psi^{sl}\xi^i\xi^j\eta^k\eta^l. \tag{4}$$

To relate this to a complex metric, we define M to be the domain of Ψ and use Ψ as a potential for a Riemannian metric on M. This immediately implies the main theorem of our current paper.

Theorem 3. *Let* X *and* Y *be open sets in* $\mathbb{R}^n$ *and* c *be a* $D_{\Psi}^{(\alpha)}$*-divergence. Then the MTW tensor is proportional to the orthogonal bisectional curvature of the Sasaki metric on* (TM, g^D, J^D), *after flatting the latter two indices (i.e. treating* η *as a covector with* $\eta(\xi) = 0$*). Furthermore, the cross curvature is proportional to the bisectional curvature of* (TM, g^D, J^D).

As an immediate consequence, this shows that the MTW tensor is non-negative if the orthogonal bisectional holomorphic curvature of TM is non-negative. By considering the dually flat structure of a Hessian manifold, we can also give an alternative characterization of relative c-convexity.

Theorem 4. *For a $D_\Psi^{(\alpha)}$-divergence, a set Y is c-convex relative to X if and only if, for all $x \in X$, the set $\frac{1+\alpha}{2}x + \frac{1-\alpha}{2}Y \subset M$ is geodesically convex with respect to the dual connection D^*.*

Our results also hold for cost functions of the form $c(x, y) = \Psi(x - y)$ with $\Psi : \mathbb{R}^n \to \mathbb{R}$ a strongly convex function. This includes many of the common examples in the literature, including the p-Wasserstein costs in Euclidean space. These costs were studied by Gangbo and McCann [7] and Ma, Trudinger and Wang [12], derived an expression proportional to Eq. (4) for their MTW tensor.

5 The Regularity of Pseudo-Arbitrages

For brevity, we will focus on a single application and refer readers to the full paper for others [9]. A recent series of papers by Pal and Wong (see, e.g., [14] and [20]) have studied the problem of finding *pseudo-arbitrages*, which are investment strategies which outperform the market portfolio under "mild and realistic assumptions". Their work combines information geometry, optimal transport and mathematical finance to reduce this problem to solving optimal transport problems where the cost function is given by a so-called log-divergence.

A central result in [14] shows that a portfolio map T outperforms the market portfolio almost surely in the long run iff it is a solution to the Monge problem for the cost function $c : \mathbb{R}^{n-1} \times \mathbb{R}^{n-1} \to \mathbb{R}$ given by

$$c(x, y) := \log\left(1 + \sum_{i=1}^{n-1} e^{x^i - y^i}\right) - \log(n) - \frac{1}{n}\sum_{i=1}^{n-1} x^i - y^i. \tag{5}$$

For this problem, we study the Kähler geometry of the Sasaki metric associated to the potential $\Psi(u) = \log\left(1 + \sum_{i=1}^{n-1} e^{u^i}\right)$. Doing so, we find that the Sasaki metric has vanishing orthogonal bisectional curvature and constant positive holomorphic sectional curvature [9]. Applying our main results and Theorem 2, we find the following regularity theorem.

Theorem 5. *Suppose μ and ν are probability measures supported respectively on subsets X and Y of the probability simplex. Suppose further that the following regularity assumptions hold:*

1. *X and Y are smooth, strictly convex and uniformly bounded from the boundary of the probability simplex. More precisely, there exists $\delta > 0$, so that for all $x \in X$, $1 \leq i \leq n$, $x^i \geq \delta$.*
2. *μ and ν are absolutely continuous with respect to the Lebesgue measure and $d\mu$ and $d\nu$ are smooth functions which are bounded away from zero and infinity on their supports.*

Let $\hat{c}(p,q)$ be the cost function given by

$$\hat{c}(p,q) = \log\left(\frac{1}{n}\sum_{i=1}^{n}\frac{q_i}{p_i}\right) - \frac{1}{n}\sum_{i=1}^{n}\log\frac{q_i}{p_i}.$$

Then the $\hat{c}$-optimal map T_u taking μ to ν is smooth.

This also provides a regularity theorem for the associated displacement interpolation, which was asked in [15].

Corollary 1. *Suppose μ and ν are smooth probability measures satisfying the assumptions of Corollary 5 and that T_u is the $\hat{c}$-optimal map transporting μ to ν. Suppose further that $T(t)\mu$ is the displacement interpolation from μ to ν defined by $T(t) = t \cdot Id + (1-t)T_u$. Then $T(t)$ is smooth, both as a map for fixed t and also in terms of the t parameter.*

Acknowledgment. The first author would like to thank Fangyang Zheng and Bo Guan for their helpful discussions about this project. The project is supported by DARPA/ARO Grant W911NF-16-1-0383 ("Information Geometry: Geometrization of Science of Information", PI: Zhang).

References

1. Bregman, L.M.: The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. USSR Comput. Math. Math. Phys. **7**(3), 200–217 (1967)
2. Brenier, Y.: Décomposition polaire et réarrangement monotone des champs de vecteurs. C.R. Acad. Sci. Paris Sér. I Math. **305**, 805–808 (1987)
3. De Philippis, G., Figalli, A.: The Monge-Ampère equation and its link to optimal transportation. Bull. Am. Math. Soc. **51**(4), 527–580 (2014)
4. Dombrowski, P.: On the geometry of the tangent bundle. Journal für Mathematik. Bd **210**(1/2), 10 (1962)
5. Figalli, A., Kim, Y.H., McCann, R.J.: When is multidimensional screening a convex program? J. Econ. Theory **146**(2), 454–478 (2011)
6. Figalli, A., Kim, Y.H., McCann, R.J.: Hölder continuity and injectivity of optimal maps. Arch. Rat. Mech. Anal. **209**(3), 747–795 (2013)
7. Gangbo, W., McCann, R.J.: Optimal maps in Monge's mass transport problem. Comptes Rendus de l'Academie des Sciences-Serie I-Mathematique **321**(12), 1653 (1995)
8. Gangbo, W., McCann, R.J.: The geometry of optimal transportation. Acta Math. **177**(2), 113–161 (1996)
9. Khan, G., Zhang, J.: On the Kähler geometry of certain optimal transport problems (2018). arXiv preprint arXiv:1812.00032
10. Kim, Y.H., McCann, R.J.: Continuity, curvature, and the general covariance of optimal transportation (2007). arXiv preprint arXiv:0712.3077
11. Loeper, G.: On the regularity of solutions of optimal transportation problems. Acta Math. **202**(2), 241–283 (2009)
12. Ma, X.N., Trudinger, N.S., Wang, X.J.: Regularity of potential functions of the optimal transportation problem. Arch. Rat. Mech. Anal. **177**(2), 151–183 (2005)

13. Monge, G.: Mémoire sur la théorie des déblais et des remblais. Histoire de l'Académie Royale des Sciences de Paris (1781)
14. Pal, S., Wong, T.K.L.: Exponentially concave functions and a new information geometry. Ann. Probab. **46**(2), 1070–1113 (2018)
15. Pal, S., Wong, T.K.L.: Multiplicative Schrödinger problem and the Dirichlet transport (2018). arXiv preprint arXiv:1807.05649
16. Satoh, H.: Almost Hermitian structures on tangent bundles. In: Proceedings of the Eleventh International Workshop on Differential. Geometry, Kyungpook Nat. Univ., Taegu, vol. 11, pp. 105–118 (2007)
17. Shima, H.: Geometry of Hessian Structures. World Scientific Publishing Co., Singapore (2007)
18. Trudinger, N.S., Wang, X.J.: On the second boundary value problem for Monge-Ampere type equations and optimal transportation. Annali della Scuola Normale Superiore di Pisa-Classe di Scienze-Serie IV **8**(1), 143 (2009)
19. Villani, C.: Optimal Transport: Old and New, vol. 338. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-71050-9
20. Wong, T.K.L.: Logarithmic divergences from optimal transport and Rényi geometry. Inf. Geom. **1**(1), 39–78 (2018)
21. Zhang, J.: Divergence function, duality, and convex analysis. Neural Comput. **16**(1), 159–195 (2004)