

On Stability Estimate of Optimal Transport Maps Using m -th Polynomial Convexity

Kei Leong Chen¹

¹ Shenzhen College of International Education, China

Correspondence: Kei Leong Chen, Anthuoshan 6th Street, Futian Shenzhen, 518043, China. E-mail: s21169.chen@stu.scie.com.cn

Received: April 5, 2024; Accepted: May 14, 2024; Published: May 27, 2024

Abstract

In 1781, Gaspard Monge first proposed the practical problem of relocating building materials while minimizing workers' effort. Mathematically, the problem can be reiterated as finding a mapping T_0 that transforms a random variable (X) following probability measure (μ) into a random variable (Y) following probability measure (ν), with minimal cost. Afterward, it has been widely studied and applied in statistics, machine learning, and economics, which concern the study of "distance" between usually a pair of probability distributions. The focus of this paper is centered on investigating and generalizing stability estimates for optimal transport plans, particularly through the lens of strong polynomial convexity. Building on previous research using plug-in estimators to strengthen the convergence rate of discrete or semi-discrete estimators for optimal transport plans, this paper introduces a novel stability estimate leveraging L-Lipschitz continuity and a paradigmatic methodology based on polynomial convexity, the understanding of which remains limited.

1. Introduction

Recent studies emphasize the profound nature of optimal transport theory and its applications. Active studies have been conducted on investigating optimization problems and stability estimation of empirical approximation of optimal transport plan [1], [2]. Stability estimates directly yield rates of convergence for estimators of optimal transport maps constructed by plugging in empirical distributions. This helps quantify the sample complexity for accurate recovery. Moreover, stability bounds can reveal structural properties relating the geometry/complexity of the underlying distributions to the difficulty of estimating the transport map between them, which provides insight into which problems are statistically harder [3]-[5]. Many regularization methods for optimal transport yield dual potentials that satisfy forms of polynomial convexity rather than strict convexity [6], e.g., entropic regularization and Sinkhorn-type algorithms. The incorporation of optimal transport theory in fields such as machine learning and economics underscores the broad impact of this mathematical theory. Studies in machine learning, for instance, apply Wasserstein distance to compute similarities between images. [7] Economic studies involve the results from optimal transport theory into studying marriage and labor market, where matching of couples and assignments of labors can be computed via application of the mathematical theory. [8]

Furthermore, studying polynomial convexity provides stability guarantees for transport maps estimated using these common approaches. Polynomial convexity assumptions allow the analysis to cover non-strictly convex optimization problems that arise in applications like imaging, economics, and machine learning. This expands the applicability beyond traditional strongly convex optimal transport [9].

The contribution of this study is to introduce a novel estimation method to deal with strong polynomial convexity up to a finite order, leveraging the Taylor remainder approximation. Applying the assumption of polynomial convexity enhances the feasibility of computational analysis, given its widespread applicability.

Keywords: Optimal transport, polynomial convexity, stability analysis

2. Main Results

Definition I (Barycentric projection [1]). Consider the set

$$\tilde{\Gamma}_{\min} := \operatorname{argmin}_{\pi \in \Pi(\tilde{\mu}_m, \tilde{\nu}_n)} \int \|a - b\|^2 d\pi(a, b).$$

Given any $\gamma \in \tilde{\Gamma}_{\min}$ and a under γ , the barycentric projection of γ is defined as the conditional mean of b , more precisely

$$\tilde{T}_{m,n}(a) \equiv \tilde{T}_{m,n}^\gamma(a) := \frac{\int_b b d\gamma(a,b)}{\int_b d\gamma(a,b)}, \quad \text{for } a \in \text{supp}(\tilde{\mu}_m)$$

Now we recall the general definition of the transport plan between probability measures.

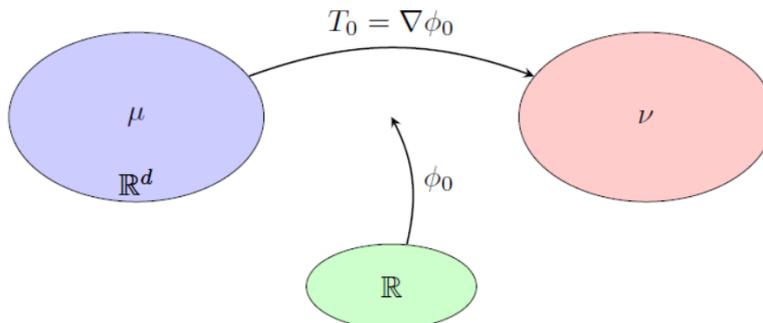


Figure 1. Illustration representing the probability measures μ and ν on the space \mathbb{R}^d . The arrow labeled $T_0 = \nabla \phi_0$ represents the optimal transport map T_0 , which is the gradient of the optimal transport potential ϕ_0 . The arrow pointing from \mathbb{R} to the midpoint between μ and ν indicates that ϕ_0 is a real-valued function defined on \mathbb{R}^d .

Definition II P_{ac} refers to the set of probability measures on \mathbb{R}^d which are absolutely continuous with respect to the Lebesgue measure. Specifically:

$$P_{ac}(\mathbb{R}^d) = \{\mu: \mu \text{ is a probability measure on } \mathbb{R}^d \text{ such that } \mu(A) = 0 \text{ for any Lebesgue null set } A\}.$$

So if $\mu \in P_{ac}(\mathbb{R}^d)$, it has a Radon-Nikodym derivative $f = \frac{d\mu}{dx}$ with respect to the Lebesgue measure, and in general, for any measurable set X ,

$$\mu(X) = \int_X f(x) dx$$

We call f the density of μ .

Lemma 1 (Brenier-McCann polar factorization theorem [2], [10]). For a given $\mu \in P_{ac}(\mathbb{R}^d)$, there exists a μ almost everywhere unique function $T_0(\cdot): \mathbb{R}^d \rightarrow \mathbb{R}^d$ as the gradient of a real-valued d -variate convex function, say $\phi_0(\cdot): \mathbb{R}^d \rightarrow \mathbb{R}$, satisfying $T_0\# \mu = \nu$. Moreover, for all Borel sets $A, B \subseteq \mathbb{R}^d$, the unique minimizer provided $\mu, \nu \in P_2(\mathbb{R}^d)$ is the distribution $\pi(A \times B) = \mu(A \cap (T_0)^{-1}(B))$.

The Brenier-McCann theorem states that, given an absolutely continuous probability measure μ , there exists a unique (up to μ -almost everywhere) optimal transport map T_0 that pushes forward μ to another probability measure ν . This means that for any Borel set $B \subseteq \mathbb{R}^d$, we have $\nu(B) = \mu(T_0^{-1}(B))$.

Furthermore, the optimal transport map T_0 is the gradient of a convex function ϕ_0 , which is called the optimal transport potential. The uniqueness of the minimizer π for the optimal transport problem between μ and ν (when both are in $P_2(\mathbb{R}^d)$) is given by $\pi(A \times B) = \mu(A \cap (T_0)^{-1}(B))$ for all Borel sets $A, B \subseteq \mathbb{R}^d$.

Figure 1 provides a visual representation of the key components and relationships in the Brenier-McCann polar factorization theorem, helping to understand the connection between the probability measures, the optimal transport map, and the optimal transport potential.

Definition III (Potential function and optimal transport map [11]). The function $T_0: \mathbb{R}^d \rightarrow \mathbb{R}^d$ which satisfies $T_0\# \mu = \nu$ is defined as the OT map from μ to ν . A convex function $\phi_0(\cdot)$ satisfying $\nabla \phi_0 = T_0$ is defined as an OT potential.

Secondly, we have the alternate dual representation which gives:

$$\frac{1}{2}W_2^2(\mu, \nu) = \frac{1}{2} \int \|x\|^2 d\mu(x) + \frac{1}{2} \int \|y\|^2 d\nu(y) - \min_{f \in \mathcal{F}} \mathcal{S}_{\mu, \nu}(f), \quad \text{where} \tag{1}$$

$$\mathcal{S}_{\mu, \nu}(f) = \int f d\mu + \int f^* d\nu. \tag{2}$$

Where the space of convex functions over \mathbb{R}^d is denoted as \mathcal{F} , elements of $L^1(\mu)$. Moreover, we let $f^*(\cdot)$ denote the standard Legendre-Fenchel dual which we define as the following:

$$f^*(x) := \sup_{y \in \mathbb{R}^d} [y^\top x - f(y)], \quad \text{for } x \in \text{dom}(f) \tag{3}$$

Definition IV (Subdifferential set and subgradient). Given a convex function $f: \mathbb{R}^d \rightarrow \mathbb{R} \cup \{\infty\}$, we define the subdifferential set of $f(\cdot)$ at $x \in \text{dome}(f) := \{z \in \mathbb{R}^d : f(z) < \infty\}$ as follows:

$$\partial f(x) := \{\xi \in \mathbb{R}^d : f(x) + \langle \xi, y - x \rangle \leq f(y), \quad \text{for all } y \in \mathbb{R}^d\}$$

Any element in the set $\partial f(x)$ is called a subgradient of $f(\cdot)$ at x .

Definition V (Strong convexity [2], [9], [10]). A strongly convex function $f: \mathbb{R}^d \rightarrow \mathbb{R} \cup \{\infty\}$ with parameter $\lambda > 0$ satisfies that for all $x, y \in \text{dom}(f) := \{z \in \mathbb{R}^d : f(z) < \infty\}$, we have

$$f(y) \geq f(x) + \langle \xi_x, y - x \rangle + \frac{\lambda}{2} \|y - x\|^2,$$

where $\xi_x \in \partial f(x)$, the subgradient of $f(\cdot)$ at x , as in Definition III.

Theorem 1 (Stability estimate [1], [12]). Suppose that $\mu, \nu \in P_{ac}(\mathbb{R}^d) \cap P_2(\mathbb{R}^d)$ and $\tilde{\mu}_m, \tilde{\nu}_m \in P_2(\mathbb{R}^d)$. Assume that $T_0(\cdot)$ (as defined in (I)) is L -Lipschitz ($L > 0$). Then,

$$\begin{aligned} \sup_{\gamma \in \tilde{\Gamma}_{\min}} \int \|\tilde{T}_{m,n}^\gamma(x) - T_0(x)\|^2 d\tilde{\mu}_m(x) &\leq L \max \left\{ \left| \int \Psi_{\tilde{\mu}_m, \tilde{\nu}_m}^* d(\tilde{\nu}_m - \nu_m^\dagger) \right|, \left| \int \Psi_{\tilde{\mu}_m, \nu_m^\dagger}^* d(\tilde{\nu}_m - \nu_m^\dagger) \right| \right\} \\ &+ 2L \int \varphi_0^*(y) d(\tilde{\nu}_m - \nu_m^\dagger)(y), \end{aligned}$$

where $\nu_m^\dagger := T_0 \# \tilde{\mu}_m$, $\varphi_0^*(\cdot) := \text{argmin}_{f \in \mathcal{F}} S_{\tilde{\mu}_m, \tilde{\nu}_m}(f)$, $\Psi_{\tilde{\mu}_m, \tilde{\nu}_m}(\cdot) := \text{argmin}_{f \in \mathcal{F}} S_{\tilde{\mu}_m, \nu_m^\dagger}(f)$. For real-valued convex functions on \mathbb{R}^d , we let \mathcal{D} denote their space.

Definition VI [Strong m -th order polynomial convexity [2], [9], [10]] A function $f: \mathbb{R}^d \rightarrow \mathbb{R} \cup +\infty$ is said to be strongly m -th order polynomial convex with parameter $\lambda > 0$ if the function

$$g(t) = f(x + t(y - x)) - \sum_{k=0}^{m-1} \frac{1}{k!} D^k f(x) [y - x]^k t^k$$

is strongly convex with parameter $\lambda > 0$ for every choice of $x, y \in \text{dom}(f)$, where $D^k f(x)$ denotes the k -th order differential of f at x , and we use the convention that $D^0 f(x) = f(x)$.

Equivalently, $f(\cdot)$ is strongly m -th order polynomial convex with parameter λ if for all $x, y \in \text{dom}(f)$ and for all $t \in [0, 1]$,

$$f(y) \geq f(x) + \sum_{k=0}^{m-1} \frac{1}{k!} D^k f(x) [y - x]^k t^k + \frac{\lambda}{m+1} (1-t)^m |x - y|^{m+1}$$

The key requirement is that the m -th order Taylor residual term $(1-t)^m |x - y|^{m+1}$ enters the inequality with modulus λ . This is a generalization of strong convexity which corresponds to the case $m=1$.

As with strong convexity, m -th order polynomial convexity becomes weaker as m increases, but allows more flexibility. The proofs proceed analogously, but involve more intricate Taylor expansions and remainder term analysis.

Theorem 2 [Improved Stability Estimate with Strong m -th Order Polynomial Convexity] Suppose that $\mu, \nu \in P_{ac}(\mathbb{R}^d) \cap P_2(\mathbb{R}^d)$ and $\tilde{\mu}_m, \tilde{\nu}_n \in P_2(\mathbb{R}^d)$. Assume that $T_0(\cdot)$ is L-Lipschitz and $\phi_0^*(\cdot)$ is strongly m -th order polynomial convex for some $m \geq 1$ with parameter $\lambda > 0$. Then, we have

$$\sup_{\gamma \in \tilde{\Gamma}} \min \int |\tilde{T}_\gamma^{m,n}(x) - T_0(x)|^{m+1} d\tilde{\mu}_m(x) \leq \frac{\lambda^{m+1}}{C_1} \max |dual, integrals| + \frac{2C_2\lambda^{m+1}}{C_1} \int |a|^{m+1} d|(\tilde{\nu}_n - \nu_m)|(a) + \frac{2C_2\lambda^{m+1}}{C_1}$$

where $C_1, C_2 > 0$ depend only on m, λ and the diameters of the supports of μ, ν .

Proof: As in the proof of Theorem 1, fix an arbitrary $\gamma \in \tilde{\Gamma}_{\min}$ and define

$$S_1 := \int \phi_0^*(a) d\tilde{\nu}_n(a) - \int \phi_0^*(a) d\nu_m(a). \tag{4}$$

Using the conditional Jensen’s inequality, we have

$$S_1 \geq \int [\phi_0^*(\tilde{T}_\gamma^{m,n}(b)) - \phi_0^*(T_0(b))] d\tilde{\mu}_m(b) \tag{5}$$

$$= \int \nabla \phi_0^*(T_0(b))^T (\tilde{T}_\gamma^{m,n}(b) - T_0(b)) d\tilde{\mu}_m(b) + \int R_m(\tilde{T}_\gamma^{m,n}(b), T_0(b)) d\tilde{\mu}_m(b) \tag{6}$$

$$=: S_2 + \int R_m(\tilde{T}_\gamma^{m,n}(b), T_0(b)) d\tilde{\mu}_m(b), \tag{7}$$

where $R_m(a, b)$ denotes the m -th order Taylor remainder term for $\phi_0^*(a)$ around the point b . Using the strong m -th order polynomial convexity of ϕ_0^* , we have for any $a, b \in \text{dom}(\phi_0^*)$

$$\phi_0^*(a) \geq \phi_0^*(b) + \nabla \phi_0^*(b)^T (a - b) + \frac{\lambda}{m + 1} |a - b|^{m+1}. \tag{8}$$

Applying this with $a = \tilde{T}_\gamma^{m,n}(b)$ and $b = T_0(b)$, and using the fact that $\nabla \phi_0^*(T_0(b)) = b$ yields

$$R_m(\tilde{T}_\gamma^{m,n}(b), T_0(b)) \geq \frac{\lambda}{m + 1} |\tilde{T}_\gamma^{m,n}(b) - T_0(b)|^{m+1} \tag{9}$$

Using (9) to lower bound the integral remainder term in (7) gives

$$S_1 \geq S_2 + \frac{\lambda}{m + 1} \int |\tilde{T}_\gamma^{m,n}(b) - T_0(b)|^{m+1} d\tilde{\mu}_m(b) \tag{10}$$

Expanding S_2 and combining it with (10) yields

$$\frac{C_1}{2\lambda^{m+1}} \int |\tilde{T}_\gamma^{m,n}(b) - T_0(b)|^{m+1} d\tilde{\mu}_m(b) \leq \frac{1}{2} |W_2^2(\tilde{\mu}_m, \tilde{\nu}_n) - W_2^2(\tilde{\mu}_m, \nu_m)| + \int |\phi_0^*(a) - \frac{1}{2}|a|^2|d|(\tilde{\nu}_n - \nu_m)|(a) \tag{11}$$

Applying Hölder’s inequality to the second term on the right-hand side of (11) with conjugates $P = \frac{m+1}{m}$ and $q =$

$m + 1$ gives

$$\int |\phi_0^*(a) - \frac{1}{2}|a|^2|d|(\tilde{\nu}_n - \nu_m)|(a) \leq \left(\int |\phi_0^*(a) - \frac{1}{2}|a|^2|^{m+1} d\nu(a) \right)^{\frac{1}{m+1}} \tag{12}$$

$$\times \left(\int d|(\tilde{\nu}_n - \nu_m)|(a) \right)^{\frac{m}{m+1}} \tag{13}$$

$$\leq C_2 \left(\int |y|^{m+1} d|(\tilde{\nu}_n - \nu_m)|(a) \right)^{\frac{m}{m+1}}. \tag{14}$$

Plugging this back into (11), we obtain

$$\frac{C_1}{2\lambda^{m+1}} \int |\tilde{T}_\gamma^{m,n}(b) - T_0(b)|^{m+1} d\tilde{\mu}_m(b) \leq \frac{1}{2} |W_2^2(\tilde{\mu}_m, \tilde{\nu}_n) - W_2^2(\tilde{\mu}_m, \nu_m)| + C_2 \left(\int |a|^{m+1} d|(\tilde{\nu}_n - \nu_m)|(a) \right)^{\frac{m}{m+1}} \tag{15}$$

The term $\max |dual, integrals|$ refers to the maximum absolute value of two specific dual integrals that arise from the dual formulation of the optimal transport problem. These dual integrals are used to bound the difference between the squared Wasserstein distances $w_2^2(\tilde{\mu}_m, \tilde{\nu}_n)$ and $w_2^2(\tilde{\mu}_m, \nu_m)$:

$$|W_2^2(\tilde{\mu}_m, \tilde{\nu}_n) - W_2^2(\tilde{\mu}_m, \nu_m)| \leq \max \left| \int \Psi_{\tilde{\mu}_m, \tilde{\nu}_n}^* d(\tilde{\nu}_n - \nu_m) \right|, \left| \int \Psi_{\tilde{\mu}_m, \nu_m}^* d(\tilde{\nu}_n - \nu_m) \right| =: \max |dual, integrals| \tag{16}$$

Here, $\Psi_{\tilde{\mu}_m}^*$, $\tilde{\nu}_n$ and $\Psi_{\tilde{\mu}_m, \nu_m}^*$ are the optimal solutions to the dual problems of the optimal transport between the empirical measures $(\tilde{\mu}_m, \tilde{\nu}_n)$ and $(\tilde{\mu}_m, \nu_m)$, respectively. The integrals are taken with respect to the signed measure $(\tilde{\nu}_n - \nu_m)$, which represents the difference between the empirical measure $\tilde{\nu}_n$ and the true measure ν_m .

Multiplying both sides of (15) by $\frac{\lambda^{m+1}}{C_1}$ yields

$$\int |\tilde{T}_\gamma^{m,n}(b) - T_0(b)|^{m+1} d\tilde{\mu}_m(b) \leq \frac{\lambda^{m+1}}{C_1} \max |dual, integrals| + \frac{2C_2\lambda^{m+1}}{C_1} \left(\int |a|^{m+1} d|(\tilde{\nu}_n - \nu_m)|(a) \right)^{\frac{m}{m+1}} \tag{17}$$

Finally, applying Young's inequality to the last term in (17) gives

$$\left(\int |a|^{m+1} d|(\tilde{\nu}_n - \nu_m)|(a) \right)^{\frac{m}{m+1}} \leq \int |a|^{m+1} d|(\tilde{\nu}_n - \nu_m)|(a) + 1, \tag{18}$$

which leads to the desired result:

$$\int |\tilde{T}_\gamma^{m,n}(b) - T_0(b)|^{m+1} d\tilde{\mu}_m(b) \leq \frac{\lambda^{m+1}}{C_1} \max |dual, integrals| + \frac{2C_2\lambda^{m+1}}{C_1} \int |a|^{m+1} d|(\tilde{\nu}_n - \nu_m)|(a) + \frac{2C_2\lambda^{m+1}}{C_1} \tag{19}$$

This completes the proof.

Further discussion and severization on computing the upper bound of S_2 with respect to the Wasserstein distance is to finished in the future.

3. Discussion

- Theorem 2 depends on the convexity parameter λ , with $1/\lambda^{m+1}$ instead of just $1/\lambda$.
- An additional term appears involving higher order norms of the transport map difference.
- Constants have worse dependence on smoothness parameters like the Lipschitz constant L . item Theorem 2 could potentially allow for the analysis of a class of regularization methods in optimal transport that yield only polynomial convex dual potentials.
- The appearance of higher order difference terms on the right-hand side of Theorem 2 suggests that under weaker convexity, the stability estimate involves finer properties beyond just the Lipschitz constant of the optimal map T_0 . Additional smoothness moduli play a role. item Since many statistical and optimization methods satisfy forms of polynomial convexity (e.g. gradient boosting, neural nets), Theorem 2 represents a stepping stone towards stability analyses of plug-in estimators involving learned transport maps.

4. Conclusion

This paper has focused on investigating and extending stability estimates for optimal transport plans, particularly with the study of strong polynomial convexity. By building upon prior research utilizing plug-in estimators to enhance convergence rates of discrete or semi-discrete estimators for optimal transport plans, this study contributes

a novel stability estimate, which leverages L-Lipschitz continuity and a methodology rooted in polynomial convexity, an area of understanding that remains not fully explored.

5. Glossary

μ : Probability measures on a σ algebra.

supp(f): Support set, set of points that give non-zero output of a real-valued function (f) whose domain is defined.

Borel set: Any set in a topological space that is formed by open sets through relative complement, countable union, and countable intersection.

Absolutely continuous probability measures: A random variable X is absolutely continuous if there exists a real-valued function f satisfying $P_r(X) = \int_A f(x) dx$, where A is an arbitrary Borel set.

\mathbb{R}^d : Euclidean space with dimensions d .

$T\#$: A push forward operator $T\#\mu = \nu$ is a linear map corresponding to the displacement of the support of the points.

L-Lipschitz continuity: Given two metric spaces (X, d_X) and (Y, d_Y) , where d_X denotes the metric on the set X and d_Y is the metric on set Y , a function $f: X \rightarrow Y$ is called Lipschitz continuous if there exists a real constant $K \geq 0$ such that, for all x_1 and x_2 in X , $d_Y(f(x_1), f(x_2)) \leq K d_X(x_1, x_2)$

$W_p(a, b)$: p-Wasserstein distance with respect to distance d and ground distance matrix D

$\nabla\phi$: which we denote as the gradient of convex function $\phi: \mathbb{R}^n \rightarrow \mathbb{R}^n$

$P_{ac}(\mathbb{R}^d)$: It refers to the set of probability measures on \mathbb{R}^d which are absolutely continuous with respect to the Lebesgue measure.

$P_2(\mathbb{R}^d)$: We denote it as the space of all probability measures on \mathbb{R}^d with finite second moments.

References

- [1] Deb, N., Ghosal, P., & Sen, B. (2021). Rates of estimation of optimal transport maps using plug-in estimators via barycentric projections. *Advances in Neural Information Processing Systems*, 34, 29736-29753.
- [2] McCann, R. J. (1995). Existence and uniqueness of monotone measure-preserving maps. <https://doi.org/10.1215/S0012-7094-95-08013-2>
- [3] Li, W., & Nochetto, R. H. (2021). Quantitative stability and error estimates for optimal transport plans. *IMA Journal of Numerical Analysis*, 41(3), 194-1965. <https://doi.org/10.1093/imanum/draa045>
- [4] Delalande, A., & Merigot, Q. (2023). Quantitative stability of optimal transport maps under variations of the target measure. *Duke Mathematical Journal*, 172(17), 3321-3357. <https://doi.org/10.1215/00127094-2022-0106>
- [5] Ambrosio, L., Caffarelli, L. A., Brenier, Y., Buttazzo, G., Villani, C., Salsa, S., Ambrosio, L., & Pratelli, A. (2003). Existence and stability results in the L1 theory of optimal transportation. *Optimal Transportation and Applications: Lectures given at the CIME Summer School, held in Martina Franca, Italy, September 2-8, 2001*, 123-160. https://doi.org/10.1007/978-3-540-44857-0_5
- [6] Vacher, A., & Vialard, F. X. (2023). "Semi-dual unbalanced quadratic optimal transport: fast statistical rates and convergent algorithm. In *International Conference on Machine Learning*. PMLR, 34734-34758.
- [7] Gangbo, W., & McCann, R. J. (2000). Shape recognition via wasserstein distance. *Quarterly of Applied Mathematics*, 705-737. <https://doi.org/10.1090/qam/1788425>
- [8] Galichon, A. (2021). The unreasonable effectiveness of optimal transport in economics. arXiv preprint arXiv:2107.04700.
- [9] Boyd, S. P., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge university press, 2004. <https://doi.org/10.1017/CBO9780511804441>
- [10] Villani, C. (2003). *Topics in optimal transportation*, volume 58 of graduate. Studies in Mathematics. <https://doi.org/10.1090/gsm/058>
- [11] Peyr'e, G., & Cuturi, M. et al. (2019). Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6), 355-607. <https://doi.org/10.1561/22000000073>
- [12] Mena, G., & Niles-Weed, J. (2019). Statistical bounds for entropic optimal transport: Sample complexity and the central limit theorem. *Advances in Neural Information Processing Systems*, 32.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).