Research Statement

Alex Delalande

April 2022

I study applied mathematics, and in particular optimal transport. During my PhD, I contributed to the theory of this field through the study of stability questions that naturally arise in applications. In what follows, I give a very brief overview of the quadratic optimal transportation problem. I then introduce the main objects I have studied and describe my different contributions to the field. I punctuate the description of these works with their potential impact and related future works.

Optimal Transport

Optimal transport [36] allows the comparison of probability measures in a geometrically faithful way. As such, it is now widely acknowledged as an important tool for numerical analysis, statistics and machine learning (see e.g. [33]). For $\Omega \subseteq \mathbb{R}^d$ and $\mathcal{P}_2(\Omega)$ the set of probability measures with finite second moment over Ω, the optimal transport problem between $\rho \in \mathcal{P}_2(\Omega)$ and $\mu \in \mathcal{P}_2(\Omega)$ with respect to the quadratic cost $c(x,y) = \|x-y\|^2$ corresponds to the following minimization problem, where the minimum is taken over the set $\Pi(\rho,\mu)$ of *transport plans* between ρ and μ, that is the set of probability measures over $\Omega \times \Omega$ with marginals ρ and μ:

$$\min_{\gamma \in \Pi(\rho,\mu)} \int_{\Omega\times\Omega} \|x-y\|^2 \,\mathrm{d}\gamma(x,y). \tag{1}$$

Wasserstein distances and spaces. The square root of the value of problem (1) is called the 2-*Wasserstein distance* between ρ and μ and is denoted $\mathrm{W}_2(\rho,\mu)$. This quantity is a proper distance that is known to metrize the weak convergence of probability measures. It is thus relevant to introduce the 2-*Wasserstein space* $(\mathcal{P}_2(\Omega), \mathrm{W}_2)$, that is a metric space of probability measures over Ω endowed with the 2-Wasserstein distance.

Optimal Transport Maps and their Stability

Optimal Transport Maps. A theorem of Brenier [7] asserts that if ρ is absolutely continuous with respect to the Lebesgue measure, the minimizer of the optimal transport problem (1) is unique and is induced by a map $T = \nabla\phi$, where ϕ is the unique convex function (up to additive constant) that verifies $\nabla\phi_{\#}\rho = \mu$ (the minimizer γ of (1) then reads $\gamma = (\mathrm{id}, T)_{\#}\rho$).

Linearized Optimal Transport. The Brenier map intervenes in the Riemannian interpretation of the 2-Wasserstein space $(\mathcal{P}_2(\Omega), \mathrm{W}_2)$ [30, 5]: in this interpretation, the tangent space to $(\mathcal{P}_2(\Omega), \mathrm{W}_2)$ at ρ a.c. is included in $\mathrm{L}^2(\rho, \mathbb{R}^d)$. The Brenier map T_μ from ρ to μ minus the identity, $T_\mu - \mathrm{id}$, can be regarded as the vector in the tangent space at ρ which supports the Wasserstein geodesic from ρ to μ. This fact has been leveraged in applications involving datasets of probability measures in order to *linearize* the Wasserstein geometry: the *Linearized Optimal Transport* (LOT) framework, introduced in [37], fixes an a.c. source reference measure $\rho \in \mathcal{P}_2(\Omega)$ and embeds $(\mathcal{P}_2(\Omega), \mathrm{W}_2)$ into $\mathrm{L}^2(\rho, \mathbb{R}^d)$ by representing any measure $\mu \in \mathcal{P}_2(\Omega)$ by the map $T_\mu - \mathrm{id}$. Working with this embedding is equivalent to replacing the Wasserstein distance $\mathrm{W}_2(\mu,\nu)$ between $\mu, \nu \in \mathcal{P}_2(\Omega)$ by the distance

$$\mathrm{W}_{2,\rho}(\mu,\nu) = \|T_\mu - T_\nu\|_{\mathrm{L}^2(\rho,\mathbb{R}^d)}.$$

This distance, for which geodesic curves are called *generalized geodesics* in the book of Ambrosio, Gigli, Savaré [5], has the advantage of being Hilbertian, which allows for the use of the whole Hilbertian toolbox of statistics and machine learning on measure-valued data, somehow consistently with the Wasserstein geometry. The use of $\mathrm{W}_{2,\rho}$ instead of W_2 may also reduce the computations by solving only as many optimal transport problems as there are elements in the measure-valued dataset at hand. A natural question that arises then when using $\mathrm{W}_{2,\rho}$ instead of W_2 is the question of how well $\mathrm{W}_{2,\rho}$ approximates W_2.

Stability of Optimal Transport Maps. Getting a quantitative comparison between $\mathrm{W}_{2,\rho}$ and W_2 actually corresponds to finding quantitative stability estimates for the mapping $\mu \mapsto T_\mu$, i.e. stability estimates for optimal transport maps under variations of the target measures. This question has been the object of my master's thesis and of the first year of my PhD, and it has led to a publication at AISTATS2020 [26] and a submission to the Duke Mathematical Journal [15] (under review). In these works, together with my supervisors Quentin Mérigot and Frédéric Chazal, we showed the bi-Hölder behavior of the mapping $\mu \mapsto T_\mu$ for some fixed reference measure ρ, and μ belonging to some subset $\mathcal{S}$ of $\mathcal{P}_2(\mathbb{R}^d)$. Namely, for ρ a.c. with compact convex support and density bounded away from zero and infinity, we showed inequalities of the type

$$\forall \mu, \nu \in \mathcal{S}, \quad \mathrm{W}_2(\mu, \nu) \leq \|T_\mu - T_\nu\|_{\mathrm{L}^2(\rho, \mathbb{R}^d)} \leq C_{\rho,\mathcal{S}} \mathrm{W}_2(\mu, \nu)^{\alpha_\mathcal{S}}, \tag{2}$$

with $C_{\rho,\mathcal{S}}, \alpha_\mathcal{S}$ positive constants depending on the choices of ρ and $\mathcal{S}$. While the left-hand side of this inequality is trivial, its right-hand side is non-trivial to get in general. Indeed, negative results exist and in particular, (2) cannot hold on the whole set $\mathcal{S} = \mathcal{P}_2(\mathbb{R}^d)$ whenever $d \geq 3$ [6, Theorem 7]. In [26], we showed (2) for $\mathcal{S} = \mathcal{P}(\mathcal{Y})$ for some compact subset $\mathcal{Y} \subset \mathbb{R}^d$ and in [15], we extended (2) to the case where $\mathcal{S}$ is made of probability measures over $\mathbb{R}^d$ with bounded p-th moment for some $p > d$. Both of these bounds mainly relied on the derivation of stability estimates for the Brenier potentials (i.e. the convex functions ϕ_μ such that $T_\mu = \nabla \phi_\mu$). These potentials solve the dual problem to (1) and their stability results from the estimation of the local strong-convexity of the dual functional to (1). In turn, such local strong-convexity estimates were derived from the stability analysis of finite volumes discretization of elliptic PDEs in [26] and from the Brascamp-Lieb concentration inequality in [15].

Beyond the justification of the LOT approach in machine learning, the estimates of [26, 15] may find applications in fields where a measure μ is approached by a discrete measure μ_N (for instance in numerical analysis, helping to justify multi-scale approaches, or in statistics, helping to get statistical guarantees and rates of convergence). The strong-convexity estimates derived in [26, 15] may also find other applications themselves, either in the numerical resolution of optimal transport (in order to improve the design and convergence analysis of optimization algorithms) or in order to derive other stability estimates, such as for instance the stability of barycenters in $(\mathcal{P}_2(\Omega), \mathbb{R}^d)$ or the stability of the entropic regularization detailed in the next sections.

Wasserstein Barycenters and their Stability

Wasserstein Barycenters. The consideration of Fréchet means in the 2-Wasserstein space $(\mathcal{P}_2(\Omega), \mathrm{W}_2)$ allows to define notions of *averages* of probability measures that are relevant to the optimal transport geometry. Indeed for any $\mathbb{P} \in \mathcal{P}(\mathcal{P}_2(\Omega))$ one may define a *Wasserstein barycenter* [2] of $\mathbb{P}$ as a minimizer $\mu_\mathbb{P} \in \mathcal{P}_2(\Omega)$ of

$$\min_{\mu \in \mathcal{P}_2(\Omega)} \frac{1}{2} \int_{\mathcal{P}_2(\Omega)} \mathrm{W}_2^2(\rho, \mu) \mathrm{d}\mathbb{P}(\rho).$$

Such a minimizer exists, and it is unique whenever $\mathbb{P}(\mathcal{P}_2^{a.c.}(\Omega)) > 0$ where $\mathcal{P}_2^{a.c.}(\Omega)$ denotes the subset of a.c. measures of $\mathcal{P}_2(\Omega)$. This notion of average has proved useful in statistics, image processing and machine learning (see e.g. [22, 32, 35, 39]). In such applications, one may not always have access to the measure $\mathbb{P}$ of interest but only to an empirical version $\mathbb{P}_N$ resulting from N-samples of $\mathbb{P}$: in this context, one may wonder how well a barycenter $\mu_{\mathbb{P}_N}$ of $\mathbb{P}_N$ approaches a barycenter $\mu_\mathbb{P}$ of $\mathbb{P}$. This leads to the question of the stability of Wasserstein barycenters with respect to their marginals.

Stability of Wasserstein Barycenters. In a recent work – still in writing – with Quentin Mérigot and Guillaume Carlier, we have studied this question. More precisely, for $\Omega \subset \mathbb{R}^d$ compact and $\mathbb{P}, \mathbb{Q} \in \mathcal{P}(\mathcal{P}(\Omega))$, we have bounded the distance $\mathrm{W}_2(\mu_\mathbb{P}, \mu_\mathbb{Q})$ between two barycenters $\mu_\mathbb{P}, \mu_\mathbb{Q}$ of $\mathbb{P}, \mathbb{Q}$ by a power of the following 1-Wasserstein distance between $\mathbb{P}$ and $\mathbb{Q}$:

$$\mathcal{W}_1(\mathbb{P}, \mathbb{Q}) = \min_{\pi \in \Pi(\mathbb{P}, \mathbb{Q})} \int_{\mathcal{P}(\Omega) \times \mathcal{P}(\Omega)} \mathrm{W}_2(\rho, \tilde{\rho}) \mathrm{d}\pi(\rho, \tilde{\rho}).$$

In order to derive such bound, we had to ensure at least uniqueness of one of the barycenters (otherwise the derivation of stability results is hopeless). We actually noticed that additional regularity assumptions were necessary, and precisely showed that if for some $m_\mathbb{P}, \alpha_\mathbb{P} > 0$,

$$\mathbb{P}\left(\{\rho \in \mathcal{P}^{a.c.}(\Omega), \mathrm{supp}(\rho) \text{ is convex}, m_\mathbb{P} \leq \rho \leq \frac{1}{m_\mathbb{P}} \text{ on } \mathrm{supp}(\rho)\}\right) = \alpha_\mathbb{P},$$

then one has the bound

$$\mathrm{W}_2(\mu_{\mathbb{P}}, \mu_{\mathbb{Q}}) \leq C_{\Omega, m_{\mathbb{P}}} \left(\frac{\mathcal{W}_1(\mathbb{P}, \mathbb{Q})}{\alpha_{\mathbb{P}}} \right)^{1/6},$$

for some finite constant $C_{\Omega, m_{\mathbb{P}}} > 0$ depending on Ω and $m_{\mathbb{P}}$. This bound was derived from two stability results: one involving the stability of optimal transport maps, for which the results of [26, 15] presented in the last section almost readily gave estimates; and one involving the stability of the push-forward operation under the gradient of a convex Lipschitz map that we proved in this recent work. Only one previous work in the literature [21] studied this question and derived bounds in the statistical context where $\mathbb{Q} = \mathbb{P}_N$ and for a very regular $\mathbb{P}$. Our stability estimate for Wasserstein barycenters is therefore the first of its kind to hold in that generality, and may again find computational and statistical applications.

Entropic Optimal Transport and its Convergence to Optimal Transport

Entropic Optimal Transport. It is recognized that the original formulation (1) of the transport problem suffers in applications from poor computationability and statistical behavior with respect to the ambient dimension, and that some form of regularization can be helpful. In this state of mind, the entropic regularization of the optimal transport problem, which dates back to Schrödinger [34] and was revisited in the recent years by Cuturi [13], has proven to be a relevant choice of regularization. This regularization consists in adding to (1) the (strongly-convex) relative entropy (or Kullback-Leibler divergence) between the candidate solution γ and the trivial transport plan $\rho \otimes \mu$, weighting this term by a parameter $\varepsilon > 0$:

$$\min_{\gamma \in \Pi(\rho, \mu)} \int_{\Omega \times \Omega} \|x - y\|^2 \, \mathrm{d}\gamma(x, y) + \varepsilon \mathrm{KL}(\gamma | \rho \otimes \mu). \tag{3}$$

Setting $\varepsilon > 0$ in (3) has many computational and statistical advantages (see e.g. [13, 3, 19, 25, 20]). Thus, introducing for a solution γ^ε to (3) the quantity

$$\mathrm{W}_{2,\varepsilon}(\rho, \mu) = \left(\int_{\mathcal{X} \times \mathcal{Y}} \|x - y\|^2 \, \mathrm{d}\gamma^\varepsilon(x, y) \right)^{1/2},$$

one may hope that $\mathrm{W}_{2,\varepsilon}$ approximates W_2 well when ε is not too big. This fact has been the object of a long line of works, going to very recent developments. The convergence of $\mathrm{W}_{2,\varepsilon}$ to W_2 as ε goes to zero is established in general settings [27, 24]. It has been quantified in more specific settings, from an asymptotic linear rate of convergence when both ρ and μ are a.c. [1, 31, 11], to a non-asymptotic exponential rate of convergence when both ρ and μ are finitely supported [10, 38]. However, very little was known – until recently [4] – on the intermediate setting of semi-discrete optimal transport, where ρ is absolutely continuous and μ is finitely supported. This setting is of particular importance in some applications in statistics [9], numerical analysis [29, 12, 18, 8, 28] or image processing [16, 17, 23].

Convergence bounds for Semi-discrete Entropic Optimal Transport. In a paper accepted at AISTATS2022 [14], I produced a non-asymptotic analysis of the trajectory in ε of the solutions of the dual problem to (3) in the semi-discrete setting. Denoting ψ^ε such dual solution to problem (3) (sometimes called a Schrödinger or Sinkhorn potential), I showed that $\varepsilon \mapsto \psi^\varepsilon$ satisfies a specific ODE. Bounding the terms of this ODE then allowed to derive regularity estimates for the mapping $\varepsilon \mapsto \psi^\varepsilon$. The terms of this ODE could be bounded using ideas similiar to Laplace's method and a strong convexity estimate reminiscent of the ones introduced above [26, 15], obtained this time from the Prékopa-Leindler inequality. Eventually, assuming the source ρ to be a.c. with compact convex support and α-Hölder continuous density bounded away from zero and infinity, I showed that for $0 < \varepsilon' \leq \varepsilon$, one has for any $\alpha' \in (0, \alpha)$

$$\left\| \psi^\varepsilon - \psi^{\varepsilon'} \right\|_\infty \leq C_{\rho,\mu} \varepsilon^{\alpha'} (\varepsilon - \varepsilon'), \tag{4}$$

where $C_{\rho,\mu} > 0$ is a constant depending on ρ and μ that can be upper bounded explicitly. This result may have consequences in the numerical resolution of optimal transport, where practitioners often use (unjustifiably but functional) ε-scaling techniques which consist in solving (3) with a starting large value of ε and then gradually decrease the value of ε over the course of the iterations of the optimization algorithm. Letting ε' go to 0 in (4) then allows us to give super-linear (resp. exponential a.e.) convergence rates of ψ^ε to ψ^0 (resp. of γ^ε to γ^0). Finally, such bounds also allow the derivation the following tight convergence result: for any $\alpha' \in (0, \alpha)$,

$$\left| \mathrm{W}^2_{2,\varepsilon}(\rho, \mu) - \mathrm{W}^2_2(\rho, \mu) - \varepsilon^2 \ell_{\rho,\mu} \right| \leq C_{\rho,\mu} \varepsilon^{2+\alpha'}, \tag{5}$$

where $\ell_{\rho,\mu}$ is an explicit term depending on ρ and μ derived in [4] and $C_{\rho,\mu} > 0$ is a constant depending on ρ and μ that can be explicitly upper bounded. Interestingly, the tightness of (5) implies that the Taylor expansion produced in [4] does not admit a third order term whenever $\alpha < 1$.

Beyond the justification of the ε-scaling heuristic used in the numerical resolution of optimal transport, the ODE derived in [14] may be used itself to help improve existing algorithms solving optimal transport, using ideas reminiscent of central-path following approaches used in interior-points methods.

References

[1] Stefan Adams, Nicolas Dirr, Mark A. Peletier, and Johannes Zimmer. From a Large-Deviations Principle to the Wasserstein Gradient Flow: A New Micro-Macro Passage. *Communications in Mathematical Physics*, 307(3):791, Sep 2011.

[2] Martial Agueh and Guillaume Carlier. Barycenters in the Wasserstein space. *SIAM Journal on Mathematical Analysis*, 43(2):904–924, 2011.

[3] Jason Altschuler, Jonathan Weed, and Philippe Rigollet. Near-linear time approximation algorithms for optimal transport via sinkhorn iteration. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 1964–1974. Curran Associates, Inc., 2017.

[4] Jason M. Altschuler, Jonathan Niles-Weed, and Austin J. Stromme. Asymptotics for semidiscrete entropic optimal transport. *SIAM Journal on Mathematical Analysis*, 54(2):1718–1741, 2022.

[5] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2008.

[6] Alexandr Andoni, Assaf Naor, and Ofer Neiman. Snowflake universality of Wasserstein spaces. *Annales Scientifiques de l'Ecole Normale Superieure*, 51(3):657–700, 2018.

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

[8] Luis A Caffarelli, Sergey A Kochengin, and Vladimir I Oliker. Problem of reflector design with given far-field scattering data. *Monge Ampère equation: applications to geometry and optimization*, 226:13, 1999.

[9] Victor Chernozhukov, Alfred Galichon, Marc Hallin, and Marc Henry. Monge–Kantorovich depth, quantiles, ranks and signs. *Annals of Statistics*, 45(1):223–256, 2017.

[10] R. Cominetti and J. San Martín. Asymptotic analysis of the exponential penalty trajectory in linear programming. *Mathematical Programming*, 67(1):169–187, Oct 1994.

[11] Giovanni Conforti and Luca Tamanini. A formula for the time derivative of the entropic cost and applications. *Journal of Functional Analysis*, 280(11):108964, 2021.

[12] MJP Cullen, J Norbury, and RJ Purser. Generalised lagrangian solutions for atmospheric and oceanic flows. *SIAM Journal on Applied Mathematics*, 51(1):20–31, 1991.

[13] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'13, pages 2292–2300, 2013.

[14] Alex Delalande. Nearly tight convergence bounds for semi-discrete entropic optimal transport. *Twenty-Fifth International Conference on Artificial Intelligence and Statistics (to appear)*, 2022.

[15] Alex Delalande and Quentin Merigot. Quantitative Stability of Optimal Transport Maps under Variations of the Target Measure. working paper or preprint, March 2021.

[16] Bruno Galerne, Arthur Leclaire, and Julien Rabin. Semi-Discrete Optimal Transport in Patch Space for Enriching Gaussian Textures. In *Geometric Science of Information*, volume 10589 of *Lecture Notes in Computer Science*, Paris, France, November 2017.

[17] Bruno Galerne, Arthur Leclaire, and Julien Rabin. A Texture Synthesis Model Based on Semi-discrete Optimal Transport in Patch Space. *SIAM Journal on Imaging Sciences*, 2018.

[18] Wilfrid Gangbo and Robert J McCann. The geometry of optimal transportation. *Acta Mathematica*, 177(2):113–161, 1996.

[19] Aude Genevay, Lénaïc Chizat, Francis Bach, Marco Cuturi, and Gabriel Peyré. Sample complexity of sinkhorn divergences. In Kamalika Chaudhuri and Masashi Sugiyama, editors, *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pages 1574–1583. PMLR, 16–18 Apr 2019.

[20] Aude Genevay, Marco Cuturi, Gabriel Peyré, and Francis Bach. Stochastic optimization for large-scale optimal transport. In *Advances in neural information processing systems*, pages 3440–3448, 2016.

[21] Thibaut Le Gouic, Quentin Paris, Philippe Rigollet, and Austin Stromme. Fast convergence of empirical barycenters in alexandrov spaces and the wasserstein space. *J. Eur. Math. Soc. (to appear)*, 2019.

[22] Rabin Julien, Gabriel Peyré, Julie Delon, and Bernot Marc. Wasserstein Barycenter and its Application to Texture Mixing. In *SSVM'11*, pages 435–446, Israel, 2011. Springer.

[23] Arthur Leclaire and Julien Rabin. A Stochastic Multi-layer Algorithm for Semi-Discrete Optimal Transport with Applications to Texture Synthesis and Style Transfer. *Journal of Mathematical Imaging and Vision*, July 2020.

[24] Christian Léonard. From the schrödinger problem to the monge-kantorovich problem. *Journal of Functional Analysis*, 262, 02 2012.

[25] Gonzalo Mena and Jonathan Niles-Weed. Statistical bounds for entropic optimal transport: sample complexity and the central limit theorem. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[26] Quentin Mérigot, Alex Delalande, and Frederic Chazal. Quantitative stability of optimal transport maps and linearization of the 2-Wasserstein space. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108, pages 3186–3196, 26–28 Aug 2020.

[27] Toshio Mikami. Monge's problem with a quadratic cost by the zero-noise limit of h-path processes. *Probability Theory and Related Fields*, 129(2):245–260, Jun 2004.

[28] Jean-Marie Mirebeau. Discretization of the 3d monge-ampere operator, between wide stencils and power diagrams. *ESAIM: M2AN*, 49(5):1511–1523, 2015.

[29] Vladimir I Oliker and Laird D Prussner. On the numerical solution of the equation $\frac{\partial^2 z}{\partial x^2}\frac{\partial^2 z}{\partial y^2} - \left(\frac{\partial^2 z}{\partial x \partial y}\right) = f$ and its discretizations, i. *Numerische Mathematik*, 54(3):271–293, 1989.

[30] Felix Otto. The geometry of dissipative evolution equations: the porous medium equation. *Communications in Partial Differential Equations*, 26:101–174, 2001.

[31] Soumik Pal. On the difference between entropic cost and the optimal transport cost. *arXiv: Probability*, 2019.

[32] Victor Panaretos and Yoav Zemel. Statistical aspects of wasserstein distances. *Annual Review of Statistics and Its Application*, 6, 03 2019.

[33] Gabriel Peyré and Marco Cuturi. Computational optimal transport. *Foundations and Trends in Machine Learning*, 11(5-6):355–607, 2019.

[34] Erwin Schrödinger. Über die umkehrung der naturgesetze. *Angewandte Chemie*, 44(30):636–636, 1931.

[35] Sanvesh Srivastava, Cheng Li, and David B. Dunson. Scalable bayes via barycenter in wasserstein space. *Journal of Machine Learning Research*, 19(8):1–35, 2018.

[36] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.

[37] Wei Wang, Dejan Slepčev, Saurav Basu, John A. Ozolek, and Gustavo K. Rohde. A linear optimal transportation framework for quantifying and visualizing variations in sets of images. *Int. J. Comput. Vision*, 101(2):254–269, January 2013.

[38] Jonathan Weed. An explicit analysis of the entropic penalty in linear programming. In Sébastien Bubeck, Vianney Perchet, and Philippe Rigollet, editors, *Conference On Learning Theory, COLT 2018, Stockholm, Sweden, 6-9 July 2018*, volume 75 of *Proceedings of Machine Learning Research*, pages 1841–1855. PMLR, 2018.

[39] Yoav Zemel and Victor Panaretos. Fréchet means and procrustes analysis in wasserstein space. *Bernoulli*, 01 2017.