Computational Optimal Transport

with Applications to Data Sciences

Other titles in Foundations and Trends® in Machine Learning

Non-convex Optimization for Machine Learningy
Prateek Jain and Purushottam Ka
ISBN: 978-1-68083-368-3

Kernel Mean Embedding of Distributions: A Review and Beyond
Krikamol Muandet, Kenji Fukumizu, Bharath Sriperumbudur and
Bernhard Scholkopf
ISBN: 978-1-68083-288-4

*Tensor Networks for Dimensionality Reduction and Large-scale
Optimization: Part 1 Low-Rank Tensor Decompositions*
Andrzej Cichocki, Anh-Huy Phan, Qibin Zhao, Namgil Lee,
Ivan Oseledets, Masashi Sugiyama and Danilo P. Mandic
ISBN: 978-1-68083-222-8

*Tensor Networks for Dimensionality Reduction and Large-scale
Optimization: Part 2 Applications and Future Perspectives*
Andrzej Cichocki, Anh-Huy Phan, Qibin Zhao, Namgil Lee,
Ivan Oseledets, Masashi Sugiyama and Danilo P. Mandic
ISBN: 978-1-68083-276-1

Patterns of Scalable Bayesian Inference
Elaine Angelino, Matthew James Johnson and Ryan P. Adams
ISBN: 978-1-68083-218-1

Generalized Low Rank Models
Madeleine Udell, Corinne Horn, Reza Zadeh and Stephen Boyd
ISBN: 978-1-68083-140-5

Computational Optimal Transport

with Applications to Data Sciences

Gabriel Peyré
CNRS and DMA, ENS

Marco Cuturi
Google and CREST/ENSAE

Foundations and Trends® in Machine Learning

Foundations and Trends® in Machine Learning
Volume 11, Issue 5-6, 2018
Editorial Board

Editorial Scope

Topics

Foundations and Trends® in Machine Learning publishes survey and tutorial articles in the following topics:

- Adaptive control and signal processing
- Applications and case studies
- Behavioral, cognitive and neural learning
- Bayesian learning
- Classification and prediction
- Clustering
- Data mining
- Dimensionality reduction
- Evaluation
- Game theoretic learning
- Graphical models
- Independent component analysis

- Inductive logic programming
- Kernel methods
- Markov chain Monte Carlo
- Model choice
- Nonparametric methods
- Online learning
- Optimization
- Reinforcement learning
- Relational learning
- Robustness
- Spectral methods
- Statistical learning theory
- Variational inference
- Visualization

Information for Librarians

Contents

Computational Optimal Transport

Gabriel Peyré[1] and Marco Cuturi[2]

[1] *CNRS and DMA, ENS; gabriel.peyre@ens.fr*
[2] *Google and CREST/ENSAE; cuturi@google.com*

ABSTRACT

Optimal transport (OT) theory can be informally described using the words of the French mathematician Gaspard Monge (1746–1818): A worker with a shovel in hand has to move a large pile of sand lying on a construction site. The goal of the worker is to erect with all that sand a target pile with a prescribed shape (for example, that of a giant sand castle). Naturally, the worker wishes to minimize her total effort, quantified for instance as the total distance or time spent carrying shovelfuls of sand. Mathematicians interested in OT cast that problem as that of comparing two probability distributions—two different piles of sand of the same volume. They consider all of the many possible ways to morph, *transport* or reshape the first pile into the second, and associate a "global" cost to every such transport, using the "local" consideration of how much it costs to move a grain of sand from one place to another. Mathematicians are interested in the properties of that least costly transport, as well as in its efficient computation. That smallest cost not only defines a distance between distributions, but it also entails a rich geometric structure on the space of probability distributions. That structure is canonical in the sense that it borrows key geometric properties of the underlying "ground" space on which these distributions are defined. For instance, when the underlying space is Euclidean, key concepts such

---

as interpolation, barycenters, convexity or gradients of functions extend naturally to the space of distributions endowed with an OT geometry.

OT has been (re)discovered in many settings and under different forms, giving it a rich history. While Monge's seminal work was motivated by an engineering problem, Tolstoi in the 1920s and Hitchcock, Kantorovich and Koopmans in the 1940s established its significance to logistics and economics. Dantzig solved it numerically in 1949 within the framework of linear programming, giving OT a firm footing in optimization. OT was later revisited by analysts in the 1990s, notably Brenier, while also gaining fame in computer vision under the name of earth mover's distances. Recent years have witnessed yet another revolution in the spread of OT, thanks to the emergence of approximate solvers that can scale to large problem dimensions. As a consequence, OT is being increasingly used to unlock various problems in imaging sciences (such as color or texture processing), graphics (for shape manipulation) or machine learning (for regression, classification and generative modeling).

This paper reviews OT with a bias toward numerical methods, and covers the theoretical properties of OT that can guide the design of new algorithms. We focus in particular on the recent wave of efficient algorithms that have helped OT find relevance in data sciences. We give a prominent place to the many generalizations of OT that have been proposed in but a few years, and connect them with related approaches originating from statistical inference, kernel methods and information theory. All of the figures can be reproduced using code made available in a companion website[1]. This website hosts the book project Computational Optimal Transport. You will also find slides and computational resources.

---

[1] https://optimaltransport.github.io/

1

Introduction

The shortest path principle guides most decisions in life and sciences:
When a commodity, a person or a single bit of information is available
at a given point and needs to be sent at a target point, one should favor
using the least possible effort. This is typically reached by moving an
item along a straight line when in the plane or along geodesic curves in
more involved metric spaces. The theory of optimal transport generalizes
that intuition in the case where, instead of moving only one item at a
time, one is concerned with the problem of moving simultaneously several
items (or a continuous distribution thereof) from one configuration onto
another. As schoolteachers might attest, planning the transportation of
a group of individuals, with the constraint that they reach a given target
configuration upon arrival, is substantially more involved than carrying
it out for a single individual. Indeed, thinking in terms of groups or
distributions requires a more advanced mathematical formalism which
was first hinted at in the seminal work of Monge (1781). Yet, no matter
how complicated that formalism might look at first sight, that problem
has deep and concrete connections with our daily life. Transportation,
be it of people, commodities or information, very rarely involves moving
only one item. All major economic problems, in logistics, production

planning or network routing, involve moving distributions, and that thread appears in all of the seminal references on optimal transport. Indeed Tolstoı (1930), Hitchcock (1941) and Kantorovich (1942) were all guided by practical concerns. It was only a few years later, mostly after the 1980s, that mathematicians discovered, thanks to the works of Brenier (1991) and others, that this theory provided a fertile ground for research, with deep connections to convexity, partial differential equations and statistics. At the turn of the millenium, researchers in computer, imaging and more generally data sciences understood that optimal transport theory provided very powerful tools to study distributions in a different and more abstract context, that of comparing distributions readily available to them under the form of bags-of-features or descriptors.

Several reference books have been written on optimal transport, including the two recent monographs by Villani (2003; 2009), those by Rachev and Rüschendorf (1998; 1998) and more recently that by San-tambrogio (2015). As exemplified by these books, the more formal and abstract concepts in that theory deserve in and by themselves several hundred pages. Now that optimal transport has gradually established itself as an applied tool (for instance, in economics, as put forward recently by Galichon (2016)), we have tried to balance that rich litera-ture with a computational viewpoint, centered on applications to data science, notably imaging sciences and machine learning. We follow in that sense the motivation of the recent review by Kolouri *et al.* (2017) but try to cover more ground. Ultimately, our goal is to present an overview of the main theoretical insights that support the practical effectiveness of OT and spend more time explaining how to turn these insights into fast computational schemes. The main body of Chapters 2, 3, 4, 9, and 10 is devoted solely to the study of the geometry induced by optimal transport in the space of probability vectors or discrete histograms. Targeting more advanced readers, we also give in the same chapters, in light gray boxes, a more general mathematical exposition of optimal transport tailored for discrete measures. Discrete measures are defined by their probability weights, but also by the location at which these weights are defined. These locations are usually taken in a continuous metric space, giving a second important degree of freedom

to model random phenomena. Lastly, the third and most technical layer of exposition is indicated in dark gray boxes and deals with arbitrary measures that need not be discrete, and which can have in particular a density w.r.t. a base measure. This is traditionally the default setting for most classic textbooks on OT theory, but one that plays a less important role in general for practical applications. Chapters 5 to 8 deal with the interplay between continuous and discrete measures and are thus targeting a more mathematically inclined audience.

The field of computational optimal transport is at the time of this writing still an extremely active one. There are therefore a wide variety of topics that we have not touched upon in this survey. Let us cite in no particular order the subjects of distributionally robust optimization (Shafieezadeh Abadeh *et al.*, 2015; Esfahani and Kuhn, 2018; Lee and Raginsky, 2018; GAO *et al.*, 2018), in which parameter estimation is carried out by minimizing the worst posssible empirical risk of any data measure taken within a certain Wasserstein distance of the input data; convergence of the Langevin Monte Carlo sampling algorithm in the Wasserstein geometry (Dalalyan and Karagulyan, 2017; Dalalyan, 2017; Bernton, 2018); other numerical methods to solve OT with a squared Euclidian cost in low-dimensional settings using the Monge-Ampère equation (Froese and Oberman, 2011; Benamou *et al.*, 2014; Sulman *et al.*, 2011) which are only briefly mentioned in Remark 2.25.

Notation

- $[\![n]\!]$: set of integers $\{1, \ldots, n\}$.

- $\mathbb{1}_{n,m}$: matrix of $\mathbb{R}^{n \times m}$ with all entries identically set to 1. $\mathbb{1}_n$: vector of ones.

- $\mathbb{I}_n$: identity matrix of size $n \times n$.

- For $u \in \mathbb{R}^n$, $\operatorname{diag}(u)$ is the $n \times n$ matrix with diagonal u and zero otherwise.

- Σ_n: probability simplex with n bins, namely the set of probability vectors in $\mathbb{R}^n_+$.

- $(\mathbf{a}, \mathbf{b})$: histograms in the simplices $\Sigma_n \times \Sigma_m$.

- (α, β): measures, defined on spaces $(\mathcal{X}, \mathcal{Y})$.

- $\frac{d\alpha}{d\beta}$: relative density of a measure α with respect to β.

- $\rho_\alpha = \frac{d\alpha}{dx}$: density of a measure α with respect to Lebesgue measure.

- $(\alpha = \sum_i \mathbf{a}_i \delta_{x_i}, \beta = \sum_j \mathbf{b}_j \delta_{y_j})$: discrete measures supported on $x_1, \ldots, x_n \in \mathcal{X}$ and $y_1, \ldots, y_m \in \mathcal{Y}$.

- $c(x, y)$: ground cost, with associated pairwise cost matrix $\mathbf{C}_{i,j} = (c(x_i, y_j))_{i,j}$ evaluated on the support of α, β.

- π: coupling measure between α and β, namely such that for any $A \subset \mathcal{X}, \pi(A \times \mathcal{Y}) = \alpha(A)$, and for any subset $B \subset \mathcal{Y}, \pi(\mathcal{X} \times B) = \beta(B)$. For discrete measures $\pi = \sum_{i,j} \mathbf{P}_{i,j} \delta_{(x_i, y_j)}$.

- $\mathcal{U}(\alpha, \beta)$: set of coupling measures, for discrete measures $\mathbf{U}(\mathbf{a}, \mathbf{b})$.

- $\mathcal{R}(c)$: set of admissible dual potentials; for discrete measures $\mathbf{R}(\mathbf{C})$.

- $T : \mathcal{X} \to \mathcal{Y}$: Monge map, typically such that $T_\sharp \alpha = \beta$.

- $(\alpha_t)_{t=0}^1$: dynamic measures, with $\alpha_{t=0} = \alpha_0$ and $\alpha_{t=1} = \alpha_1$.

- v: speed for Benamou–Brenier formulations; $J = \alpha v$: momentum.

- (f, g): dual potentials, for discrete measures $(\mathbf{f}, \mathbf{g})$ are dual variables.

- $(\mathbf{u}, \mathbf{v}) \stackrel{\text{def.}}{=} (e^{\mathbf{f}/\varepsilon}, e^{\mathbf{g}/\varepsilon})$: Sinkhorn scalings.

- $\mathbf{K} \stackrel{\text{def.}}{=} e^{-\mathbf{C}/\varepsilon}$: Gibbs kernel for Sinkhorn.

- s: flow for $\mathcal{W}_1$-like problem (optimization under divergence constraints).

- $\mathrm{L}_{\mathbf{C}}(\mathbf{a}, \mathbf{b})$ and $\mathcal{L}_c(\alpha, \beta)$: value of the optimization problem associated to the OT with cost $\mathbf{C}$ (histograms) and c (arbitrary measures).

- $W_p(\mathbf{a}, \mathbf{b})$ and $\mathcal{W}_p(\alpha, \beta)$: p-Wasserstein distance associated to ground distance matrix $\mathbf{D}$ (histograms) and distance d (arbitrary measures).

- $\lambda \in \Sigma_S$: weight vector used to compute the barycenters of S measures.

- $\langle \cdot, \cdot \rangle$: for the usual Euclidean dot-product between vectors; for two matrices of the same size A and B, $\langle A, B \rangle \stackrel{\text{def.}}{=} \text{tr}(A^\top B)$ is the Frobenius dot-product.

- $f \oplus g(x, y) \stackrel{\text{def.}}{=} f(x) + g(y)$, for two functions $f : \mathcal{X} \to \mathbb{R}, g : \mathcal{Y} \to \mathbb{R}$, defines $f \oplus g : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$.

- $\mathbf{f} \oplus \mathbf{g} \stackrel{\text{def.}}{=} \mathbf{f}\mathbb{1}_m^\top + \mathbb{1}_n \mathbf{g}^\top \in \mathbb{R}^{n \times m}$ for two vectors $\mathbf{f} \in \mathbb{R}^n$, $\mathbf{g} \in \mathbb{R}^m$.

- $\alpha \otimes \beta$ is the product measure on $\mathcal{X} \times \mathcal{Y}$, i.e. $\int_{\mathcal{X} \times \mathcal{Y}} g(x, y) \mathrm{d}(\alpha \otimes \beta)(x, y) \stackrel{\text{def.}}{=} \int_{\mathcal{X} \times \mathcal{Y}} g(x, y) \mathrm{d}\alpha(x) \mathrm{d}\beta(y)$.

- $\mathbf{a} \otimes \mathbf{b} \stackrel{\text{def.}}{=} \mathbf{a}\mathbf{b}^\top \in \mathbb{R}^{n \times m}$.

- $\mathbf{u} \odot \mathbf{v} = (\mathbf{u}_i \mathbf{v}_i) \in \mathbb{R}^n$ for $(\mathbf{u}, \mathbf{v}) \in (\mathbb{R}^n)^2$.

References

Abadi, M., A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, *et al.* 2016. "Tensorflow: large-scale machine learning on heterogeneous distributed systems". *arXiv preprint arXiv:1603.04467*.

Abraham, I., R. Abraham, M. Bergounioux, and G. Carlier. 2017. "Tomographic reconstruction from a few views: a multi-marginal optimal transport approach". *Applied Mathematics & Optimization*. 75(1): 55–73.

Adams, R. P. and R. S. Zemel. 2011. "Ranking via Sinkhorn propagation". *arXiv preprint arXiv:1106.1925*.

Agueh, M. and M. Bowles. 2013. "One-dimensional numerical algorithms for gradient flows in the p-Wasserstein spaces". *Acta Applicandae Mathematicae*. 125(1): 121–134.

Agueh, M. and G. Carlier. 2011. "Barycenters in the Wasserstein space". *SIAM Journal on Mathematical Analysis*. 43(2): 904–924.

Agueh, M. and G. Carlier. 2017. "Vers un théorème de la limite centrale dans l'espace de Wasserstein?" *Comptes Rendus Mathematique*. 355(7): 812–818.

Ali, S. M. and S. D. Silvey. 1966. "A general class of coefficients of divergence of one distribution from another". *Journal of the Royal Statistical Society. Series B (Methodological)*. 28(1): 131–142.

Allen-Zhu, Z., Y. Li, R. Oliveira, and A. Wigderson. 2017. "Much faster algorithms for matrix scaling". *arXiv preprint arXiv:1704.02315*.

Al-Rfou, R., G. Alain, A. Almahairi, C. Angermüller, D. Bahdanau, and N. B. et al. 2016. "Theano: A Python framework for fast computation of mathematical expressions". *CoRR*. abs/1605.02688.

Altschuler, J., J. Weed, and P. Rigollet. 2017. "Near-linear time approximation algorithms for optimal transport via Sinkhorn iteration". *arXiv preprint arXiv:1705.09634*.

Álvarez-Esteban, P. C., E. del Barrio, J. Cuesta-Albertos, and C. Matrán. 2016. "A fixed-point approach to barycenters in Wasserstein space". *Journal of Mathematical Analysis and Applications*. 441(2): 744–762.

Alvarez-Melis, D., S. Jegelka, and T. S. Jaakkola. 2018. "Towards Optimal Transport with Global Invariances". *arXiv preprint 1806.09277*.

Amari, S.-i., R. Karakida, and M. Oizumi. 2018. "Information geometry connecting Wasserstein distance and Kullback-Leibler divergence via the entropy-relaxed transportation problem". *Information Geometry*. (1): 13–37.

Ambrosio, L., N. Gigli, and G. Savaré. 2006. *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. Springer.

Ambrosio, L. and A. Figalli. 2009. "Geodesics in the space of measure-preserving maps and plans". *Archive for Rational Mechanics and Analysis*. 194(2): 421–462.

Anderes, E., S. Borgwardt, and J. Miller. 2016. "Discrete Wasserstein barycenters: optimal transport for discrete data". *Mathematical Methods of Operations Research*. 84(2): 389–409.

Andoni, A., P. Indyk, and R. Krauthgamer. 2008. "Earth mover distance over high-dimensional spaces". In: *Proceedings of the nineteenth annual ACM-SIAM Symposium on Discrete Algorithms*. Society for Industrial and Applied Mathematics. 343–352.

Andoni, A., A. Naor, and O. Neiman. 2018. "Snowflake universality of Wasserstein spaces". *Annales scientifiques de l'École normale supérieure*. 51: 657–700.

Arjovsky, M., S. Chintala, and L. Bottou. 2017. "Wasserstein generative adversarial networks". *Proceedings of the 34th International Conference on Machine Learning*. Proceedings of Machine Learning Research 70: 214–223. Ed. by D. Precup and Y. W. Teh.

Aurenhammer, F. 1987. "Power diagrams: properties, algorithms and applications". *SIAM Journal on Computing.* 16(1): 78–96.

Aurenhammer, F., F. Hoffmann, and B. Aronov. 1998. "Minkowski-type theorems and least-squares clustering". *Algorithmica.* 20(1): 61–76.

Averbuch, A., R. Coifman, D. Donoho, M. Israeli, and J. Walden. 2001. "Fast slant stack: a notion of Radon transform for data in a Cartesian grid which is rapidly computible, algebraically exact, geometrically faithful and invertible". *Tech. Rep., Stanford University.*

Bach, F. 2010. "Self-concordant analysis for logistic regression". *Electronic Journal of Statistics.* 4: 384–414.

Bach, F. R. 2014. "Adaptivity of averaged stochastic gradient descent to local strong convexity for logistic regression." *Journal of Machine Learning Research.* 15(1): 595–627.

Bacharach, M. 1965. "Estimating nonnegative matrices from marginal data". *International Economic Review.* 6(3): 294–310.

Barrio, E. del, J. Cuesta-Albertos, C. Matrán, and A. Mayo-Íscar. 2016. "Robust clustering tools based on optimal transportation". *arXiv preprint arXiv:1607.01179.*

Barvinok, A. 2002. *A Course in Convexity. Graduate Studies in Mathematics.* American Mathematical Society. ISBN: 9780821829684.

Bassetti, F., A. Bodini, and E. Regazzini. 2006. "On minimum Kantorovich distance estimators". *Statistics & Probability Letters.* 76(12): 1298–1302.

Bauschke, H. H. and P. L. Combettes. 2011. *Convex analysis and monotone operator theory in Hilbert spaces.* New York: Springer-Verlag.

Bauschke, H. H. and A. S. Lewis. 2000. "Dykstra's algorithm with Bregman projections: a convergence proof". *Optimization.* 48(4): 409–427.

Beck, A. and M. Teboulle. 2003. "Mirror descent and nonlinear projected subgradient methods for convex optimization". *Operations Research Letters.* 31(3): 167–175.

Beckmann, M. 1952. "A continuous model of transportation". *Econometrica.* 20: 643–660.

Beiglböck, M., P. Henry-Labordère, and F. Penkner. 2013. "Model-independent bounds for option prices: a mass transport approach". *Finance and Stochastics*. 17(3): 477–501.

Beirlant, J., E. J. Dudewicz, L. Gyorfi, and E. C. Van der Meulen. 1997. "Nonparametric entropy estimation: an overview". *International Journal of Mathematical and Statistical Sciences*. 6(1): 17–39.

Benamou, J.-D. 2003. "Numerical resolution of an "unbalanced" mass transport problem". *ESAIM: Mathematical Modelling and Numerical Analysis*. 37(05): 851–868.

Benamou, J.-D. and Y. Brenier. 2000. "A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem". *Numerische Mathematik*. 84(3): 375–393.

Benamou, J.-D. and G. Carlier. 2015. "Augmented Lagrangian methods for transport optimization, mean field games and degenerate elliptic equations". *Journal of Optimization Theory and Applications*. 167(1): 1–26.

Benamou, J.-D., G. Carlier, M. Cuturi, L. Nenna, and G. Peyré. 2015. "Iterative Bregman projections for regularized transportation problems". *SIAM Journal on Scientific Computing*. 37(2): A1111–A1138.

Benamou, J.-D., G. Carlier, Q. Mérigot, and E. Oudet. 2016a. "Discretization of functionals involving the Monge–Ampère operator". *Numerische Mathematik*. 134(3): 611–636.

Benamou, J.-D., F. Collino, and J.-M. Mirebeau. 2016b. "Monotone and consistent discretization of the Monge-Ampere operator". *Mathematics of Computation*. 85(302): 2743–2775.

Benamou, J.-D., B. D. Froese, and A. M. Oberman. 2014. "Numerical solution of the optimal transportation problem using the Monge–Ampere equation". *Journal of Computational Physics*. 260: 107–126.

Berg, C., J. P. R. Christensen, and P. Ressel. 1984. *Harmonic Analysis on Semigroups. Graduate Texts in Mathematics*. No. 100. Springer Verlag.

Berlinet, A. and C. Thomas-Agnan. 2003. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Kluwer Academic Publishers.

Bernton, E. 2018. "Langevin Monte Carlo and JKO splitting". In: *Proceedings of the 31st Conference On Learning Theory*. Ed. by S. Bubeck, V. Perchet, and P. Rigollet. Vol. 75. *Proceedings of Machine Learning Research*. PMLR. 1777–1798.

Bernton, E., P. E. Jacob, M. Gerber, and C. P. Robert. 2017. "Inference in generative models using the Wasserstein distance". *arXiv preprint arXiv:1701.05146*.

Bertsekas, D. P. 1981. "A new algorithm for the assignment problem". *Mathematical Programming*. 21(1): 152–171.

Bertsekas, D. P. 1992. "Auction algorithms for network flow problems: a tutorial introduction". *Computational Optimization and Applications*. 1(1): 7–66.

Bertsekas, D. P. 1998. *Network Optimization: Continuous and Discrete Models*. Athena Scientific.

Bertsekas, D. P. and J. Eckstein. 1988. "Dual coordinate step methods for linear network flow problems". *Mathematical Programming*. 42(1): 203–243.

Bertsimas, D. and J. N. Tsitsiklis. 1997. *Introduction to Linear Optimization*. Athena Scientific.

Bhatia, R., T. Jain, and Y. Lim. 2018. "On the Bures-Wasserstein distance between positive definite matrices". *Expositiones Mathematicae, to appear*.

Bigot, J., E. Cazelles, and N. Papadakis. 2017a. "Central limit theorems for Sinkhorn divergence between probability distributions on finite spaces and statistical applications". *arXiv preprint arXiv:1711.08947*.

Bigot, J., R. Gouet, T. Klein, and A. López. 2017b. "Geodesic PCA in the Wasserstein space by convex PCA". *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*. 53(1): 1–26.

Bigot, J. and T. Klein. 2012a. "Characterization of barycenters in the Wasserstein space by averaging optimal transport maps". *arXiv preprint arXiv:1212.2562*.

Bigot, J. and T. Klein. 2012b. "Consistent estimation of a population barycenter in the Wasserstein space". *arXiv Preprint arXiv:1212.-2562*.

Birkhoff, G. 1946. "Tres observaciones sobre el algebra lineal". *Universidad Nacional de Tucumán Revista Series A*. 5: 147–151.

Birkhoff, G. 1957. "Extensions of Jentzsch's theorem". *Transactions of the American Mathematical Society.* 85(1): 219–227.

Blanchet, A., V. Calvez, and J. A. Carrillo. 2008. "Convergence of the mass-transport steepest descent scheme for the subcritical Patlak-Keller-Segel model". *SIAM Journal on Numerical Analysis.* 46(2): 691–721.

Blanchet, A. and G. Carlier. 2015. "Optimal transport and Cournot-Nash equilibria". *Mathematics of Operations Research.* 41(1): 125–145.

Boissard, E. 2011. "Simple bounds for the convergence of empirical and occupation measures in 1-Wasserstein distance". *Electronic Journal of Probability.* 16: 2296–2333.

Boissard, E., T. Le Gouic, and J.-M. Loubes. 2015. "Distribution's template estimate with Wasserstein metrics". *Bernoulli.* 21(2): 740–759.

Bolley, F., A. Guillin, and C. Villani. 2007. "Quantitative concentration inequalities for empirical measures on non-compact spaces". *Probability Theory and Related Fields.* 137(3): 541–593.

Bonneel, N., G. Peyré, and M. Cuturi. 2016. "Wasserstein barycentric coordinates: histogram regression using optimal transport". *ACM Transactions on Graphics.* 35(4): 71:1–71:10.

Bonneel, N., J. Rabin, G. Peyré, and H. Pfister. 2015. "Sliced and Radon Wasserstein barycenters of measures". *Journal of Mathematical Imaging and Vision.* 51(1): 22–45.

Bonneel, N., M. Van De Panne, S. Paris, and W. Heidrich. 2011. "Displacement interpolation using Lagrangian mass transport". *ACM Transactions on Graphics.* 30(6): 158.

Borchardt, C. and C. Jocobi. 1865. "De investigando ordine systematis aequationum differentialium vulgarium cujuscunque." *Journal für die reine und angewandte Mathematik.* 64: 297–320.

Borg, I. and P. J. Groenen. 2005. *Modern Multidimensional Scaling: Theory and Applications.* Springer Science & Business Media.

Botsch, M., L. Kobbelt, M. Pauly, P. Alliez, and B. Lévy. 2010. *Polygon mesh processing.* Taylor & Francis.

Bousquet, O., S. Gelly, I. Tolstikhin, C.-J. Simon-Gabriel, and B. Schoelkopf. 2017. "From optimal transport to generative modeling: the VEGAN cookbook". *arXiv preprint arXiv:1705.07642.*

Boyd, S., N. Parikh, E. Chu, B. Peleato, and J. Eckstein. 2011. "Distributed optimization and statistical learning via the alternating direction method of multipliers". *Foundations and Trends in Machine Learning.* 3(1): 1–122.

Bregman, L. M. 1967. "The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming". *USSR Computational Mathematics and Mathematical Physics.* 7(3): 200–217.

Brenier, Y. 1987. "Décomposition polaire et réarrangement monotone des champs de vecteurs". *C. R. Acad. Sci. Paris Sér. I Math.* 305(19): 805–808.

Brenier, Y. 1990. "The least action principle and the related concept of generalized flows for incompressible perfect fluids". *Journal of the AMS.* 2: 225–255.

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

Brenier, Y. 1993. "The dual least action problem for an ideal, incompressible fluid". *Archive for Rational Mechanics and Analysis.* 122(4): 323–351.

Brenier, Y. 1999. "Minimal geodesics on groups of volume-preserving maps and generalized solutions of the Euler equations". *Communications on Pure and Applied Mathematics.* 52(4): 411–452.

Brenier, Y. 2008. "Generalized solutions and hydrostatic approximation of the Euler equations". *Physica D. Nonlinear Phenomena.* 237(14-17): 1982–1988.

Bronstein, A. M., M. M. Bronstein, and R. Kimmel. 2006. "Generalized multidimensional scaling: a framework for isometry-invariant partial surface matching". *Proceedings of the National Academy of Sciences.* 103(5): 1168–1172.

Bronstein, A. M., M. M. Bronstein, R. Kimmel, M. Mahmoudi, and G. Sapiro. 2010. "A Gromov-Hausdorff framework with diffusion geometry for topologically-robust non-rigid shape matching". *International Journal on Computer Vision.* 89(2-3): 266–286.

Brualdi, R. A. 2006. *Combinatorial Matrix Classes.* Vol. 108. Cambridge University Press.

Burago, D., Y. Burago, and S. Ivanov. 2001. *A Course in Metric Geometry.* Vol. 33. American Mathematical Society Providence, RI.

Bures, D. 1969. "An extension of Kakutani's theorem on infinite product measures to the tensor product of semifinite w^*-algebras". *Transactions of the American Mathematical Society.* 135: 199–212.

Burger, M., J. A. Carrillo de la Plata, and M.-T. Wolfram. 2010. "A mixed finite element method for nonlinear diffusion equations". *Kinetic and Related Models.* 3(1): 59–83.

Burger, M., M. Franek, and C.-B. Schönlieb. 2012. "Regularised regression and density estimation based on optimal transport". *Applied Mathematics Research Express.* 2: 209–253.

Buttazzo, G., L. De Pascale, and P. Gori-Giorgi. 2012. "Optimal-transport formulation of electronic density-functional theory". *Physical Review A.* 85(6): 062502.

Caffarelli, L. 2003. "The Monge-Ampere equation and optimal transportation, an elementary review". *Lecture Notes in Mathematics, Springer-Verlag*: 1–10.

Caffarelli, L. A., S. A. Kochengin, and V. I. Oliker. 1999. "Problem of reflector design with given far-field scattering data". In: *Monge Ampère equation: applications to geometry and optimization.* Vol. 226. 13.

Caffarelli, L. A. and R. J. McCann. 2010. "Free boundaries in optimal transport and Monge-Ampère obstacle problems". *Annals of Mathematics.* 171(2): 673–730.

Caffarelli, L., M. Feldman, and R. McCann. 2002. "Constructing optimal maps for Monge's transport problem as a limit of strictly convex costs". *Journal of the American Mathematical Society.* 15(1): 1–26.

Canas, G. and L. Rosasco. 2012. "Learning probability measures with respect to optimal transport metrics". In: *Advances in Neural Information Processing Systems 25*. Ed. by F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger. 2492–2500.

Carlen, E. A. and J. Maas. 2014. "An analog of the 2-Wasserstein metric in non-commutative probability under which the fermionic Fokker–Planck equation is gradient flow for the entropy". *Communications in Mathematical Physics*. 331(3): 887–926.

Carlier, G., V. Chernozhukov, and A. Galichon. 2016. "Vector quantile regression beyond correct specification". *arXiv preprint arXiv:1610.-06833*.

Carlier, G., V. Duval, G. Peyré, and B. Schmitzer. 2017. "Convergence of entropic schemes for optimal transport and gradient flows". *SIAM Journal on Mathematical Analysis*. 49(2): 1385–1418.

Carlier, G. and I. Ekeland. 2010. "Matching for teams". *Economic Theory*. 42(2): 397–418.

Carlier, G., A. Galichon, and F. Santambrogio. 2010. "From Knothe's transport to Brenier's map and a continuation method for optimal transport". *SIAM Journal on Mathematical Analysis*. 41(6): 2554–2576.

Carlier, G., C. Jimenez, and F. Santambrogio. 2008. "Optimal transportation with traffic congestion and Wardrop equilibria". *SIAM Journal on Control and Optimization*. 47(3): 1330–1350.

Carlier, G., A. Oberman, and E. Oudet. 2015. "Numerical methods for matching for teams and Wasserstein barycenters". *ESAIM: Mathematical Modelling and Numerical Analysis*. 49(6): 1621–1642.

Carlier, G. and C. Poon. 2017. "On the total variation Wasserstein gradient flow and the TV-JKO scheme". *arXiv preprint arXiv:1703.-00243*.

Carrillo, J. A., A. Chertock, and Y. Huang. 2015. "A finite-volume method for nonlinear nonlocal equations with a gradient flow structure". *Communications in Computational Physics*. 17(01): 233–258.

Carrillo, J. A. and J. S. Moll. 2009. "Numerical simulation of diffusive and aggregation phenomena in nonlinear continuity equations by evolving diffeomorphisms". *SIAM Journal on Scientific Computing*. 31(6): 4305–4329.

Censor, Y. and S. Reich. 1998. "The Dykstra algorithm with Bregman projections". *Communications in Applied Analysis*. 2: 407–419.

Censor, Y. and S. A. Zenios. 1992. "Proximal minimization algorithm with *D*-functions". *Journal of Optimization Theory and Applications*. 73(3): 451–464.

Champion, T., L. De Pascale, and P. Juutinen. 2008. "The ∞-Wasserstein distance: local solutions and existence of optimal transport maps". *SIAM Journal on Mathematical Analysis*. 40(1): 1–20.

Chan, T. M. 1996. "Optimal output-sensitive convex hull algorithms in two and three dimensions". *Discrete & Computational Geometry*. 16(4): 361–368.

Chen, Y., W. Gangbo, T. T. Georgiou, and A. Tannenbaum. 2017. "On the matrix Monge-Kantorovich problem". *arXiv preprint arXiv:1701.02826*.

Chen, Y., T. T. Georgiou, and M. Pavon. 2016a. "On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint". *Journal of Optimization Theory and Applications*. 169(2): 671–691.

Chen, Y., T. T. Georgiou, and A. Tannenbaum. 2016b. "Matrix optimal mass transport: a quantum mechanical approach". *arXiv preprint arXiv:1610.03041*.

Chizat, L., G. Peyré, B. Schmitzer, and F.-X. Vialard. 2018a. "An interpolating distance between optimal transport and Fisher–Rao metrics". *Foundations of Computational Mathematics*. 18(1): 1–44.

Chizat, L., G. Peyré, B. Schmitzer, and F.-X. Vialard. 2018b. "Scaling algorithms for unbalanced transport problems". *Mathematics of Computation*. 87: 2563–2609.

Chizat, L., G. Peyré, B. Schmitzer, and F.-X. Vialard. 2018c. "Unbalanced optimal transport: geometry and Kantorovich formulation". *Journal of Functional Analysis*. 274(11): 3090–3123.

Chow, S.-N., W. Huang, Y. Li, and H. Zhou. 2012. "Fokker-Planck equations for a free energy functional or Markov process on a graph". *Archive for Rational Mechanics and Analysis*. 203(3): 969–1008.

Chow, S.-N., W. Li, and H. Zhou. 2017a. "A discrete Schrodinger equation via optimal transport on graphs". *arXiv preprint arXiv:1705.07583*.

Chow, S.-N., W. Li, and H. Zhou. 2017b. "Entropy dissipation of Fokker-Planck equations on graphs". *arXiv preprint arXiv:1701.04841.*

Chowdhury, S. and F. Mémoli. 2016. "Constructing geodesics on the space of compact metric spaces". *arXiv preprint arXiv:1603.02385.*

Chui, H. and A. Rangarajan. 2000. "A new algorithm for non-rigid point matching". In: *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on.* Vol. 2. IEEE. 44–51.

Ciszár, I. 1967. "Information-type measures of difference of probability distributions and indirect observations". *Studia Scientiarum Mathematicarum Hungarica.* 2: 299–318.

Cohen, M. B., A. Madry, D. Tsipras, and A. Vladu. 2017. "Matrix scaling and balancing via box constrained Newton's method and interior point methods". *arXiv preprint arXiv:1704.02310.*

Cohen, S. and L. Guibas. 1999. "The earth mover's distance under transformation sets". In: *Proceedings of the Seventh IEEE International Conference on Computer vision.* Vol. 2. IEEE. 1076–1083.

Combettes, P. L. and J.-C. Pesquet. 2007. "A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery". *IEEE Journal of Selected Topics in Signal Processing.* 1(4): 564–574.

Cominetti, R. and J. San Martín. 1994. "Asymptotic analysis of the exponential penalty trajectory in linear programming". *Mathematical Programming.* 67(1-3): 169–187.

Condat, L. 2015. "Fast projection onto the simplex and the ℓ_1 ball". *Math. Programming, Ser. A*: 1–11.

Costa, S. I., S. A. Santos, and J. E. Strapasson. 2015. "Fisher information distance: a geometrical reading". *Discrete Applied Mathematics.* 197: 59–69.

Cotar, C., G. Friesecke, and C. Klüppelberg. 2013. "Density functional theory and optimal transportation with Coulomb cost". *Communications on Pure and Applied Mathematics.* 66(4): 548–599.

Courty, N., R. Flamary, A. Habrard, and A. Rakotomamonjy. 2017a. "Joint distribution optimal transportation for domain adaptation". In: *Advances in Neural Information Processing Systems 30.* Ed. by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett. Curran Associates, Inc. 3730–3739.

Courty, N., R. Flamary, D. Tuia, and T. Corpetti. 2016. "Optimal transport for data fusion in remote sensing". In: *2016 IEEE International Geoscience and Remote Sensing Symposium*. IEEE. 3571–3574.

Courty, N., R. Flamary, D. Tuia, and A. Rakotomamonjy. 2017b. "Optimal transport for domain adaptation". *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 39(9): 1853–1865.

Crane, K., C. Weischedel, and M. Wardetzky. 2013. "Geodesics in heat: a new approach to computing distance based on heat flow". *ACM Transaction on Graphics*. 32(5): 152:1–152:11.

Cuesta, J. A. and C. Matran. 1989. "Notes on the Wasserstein Metric in Hilbert Spaces". *The Annals of Probability*. 17(3): 1264–1276.

Cuturi, M. 2012. "Positivity and transportation". *arXiv preprint 1209.2655*.

Cuturi, M. 2013. "Sinkhorn distances: lightspeed computation of optimal transport". In: *Advances in Neural Information Processing Systems 26*. 2292–2300.

Cuturi, M. and D. Avis. 2014. "Ground metric learning". *Journal of Machine Learning Research*. 15: 533–564.

Cuturi, M. and A. Doucet. 2014. "Fast computation of Wasserstein barycenters". In: *Proceedings of ICML*. Vol. 32. 685–693.

Cuturi, M. and K. Fukumizu. 2007. "Kernels on structured objects through nested histograms". In: *Advances in Neural Information Processing Systems 19*. Ed. by P. B. Schölkopf, J. C. Platt, and T. Hoffman. MIT Press. 329–336.

Cuturi, M. and G. Peyré. 2016. "A smoothed dual approach for variational Wasserstein problems". *SIAM Journal on Imaging Sciences*. 9(1): 320–343.

Dalalyan, A. 2017. "Further and stronger analogy between sampling and optimization: Langevin Monte Carlo and gradient descent". In: *Proceedings of the 2017 Conference on Learning Theory*. Vol. 65. *Proceedings of Machine Learning Research*. PMLR. 678–689.

Dalalyan, A. S. and A. G. Karagulyan. 2017. "User-friendly guarantees for the Langevin Monte Carlo with inaccurate gradient". *arXiv preprint arXiv:1710.00095*.

Dantzig, G. B. 1949. "Programming of interdependent activities: II mathematical model". *Econometrica*. 17(3/4): 200–211.

Dantzig, G. B. 1951. "Application of the simplex method to a trans-
    portation problem". *Activity Analysis of Production and Allocation.*
    13: 359–373.

Dantzig, G. B. 1983. "Reminiscences Aabout the origins of linear pro-
    gramming". In: *Mathematical Programming The State of the Art:
    Bonn 1982.* Ed. by A. Bachem, B. Korte, and M. Grötschel. Springer.
    78–86.

Dantzig, G. B. 1991. "Linear Programming". In: *History of mathemat-
    ical programming: a collection of personal reminiscences.* Ed. by
    J. K. Lenstra, A. H. G. R. Kan, and A. Schrijver. Elsevier Science
    Publishers. 257–282.

Dattorro, J. 2017. *Convex Optimization & Euclidean Distance Geometry.*
    Meboo Publishing.

De Goes, F., K. Breeden, V. Ostromoukhov, and M. Desbrun. 2012.
    "Blue noise through optimal transport". *ACM Transactions on
    Graphics.* 31(6): 171.

Delon, J. 2004. "Midway image equalization". *Journal of Mathematical
    Imaging and Vision.* 21(2): 119–134.

Delon, J., J. Salomon, and A. Sobolevski. 2010. "Fast transport opti-
    mization for Monge costs on the circle". *SIAM Journal on Applied
    Mathematics.* 70(7): 2239–2258.

Delon, J., J. Salomon, and A. Sobolevski. 2012. "Local matching indi-
    cators for transport problems with concave costs". *SIAM Journal
    on Discrete Mathematics.* 26(2): 801–827.

Deming, E. and F. F. Stephan. 1940. "On a least squares adjustment of
    a sampled frequency table when the expected marginal totals are
    known". *Annals of Mathematical Statistics.* 11(4): 427–444.

Dereich, S., M. Scheutzow, and R. Schottstedt. 2013. "Constructive
    quantization: Approximation by empirical measures". In: *Annales de
    l'Institut Henri Poincaré, Probabilités et Statistiques.* Vol. 49. No. 4.
    1183–1203.

Deriche, R. 1993. "Recursively implementating the Gaussian and its
    derivatives". *PhD thesis.* INRIA.

Dessein, A., N. Papadakis, and C.-A. Deledalle. 2017. "Parameter estimation in finite mixture models by regularized optimal transport: a unified framework for hard and soft clustering". *arXiv preprint arXiv:1711.04366.*

Dessein, A., N. Papadakis, and J.-L. Rouas. 2018. "Regularized Optimal Transport and the Rot Mover's Distance". *Journal of Machine Learning Research.* 19(15): 1–53.

Di Marino, S. and L. Chizat. 2017. "A tumor growth model of Hele-Shaw type as a gradient flow". *Arxiv.*

Do Ba, K., H. L. Nguyen, H. N. Nguyen, and R. Rubinfeld. 2011. "Sublinear time algorithms for earth mover's distance". *Theory of Computing Systems.* 48(2): 428–442.

Dolbeault, J., B. Nazaret, and G. Savaré. 2009. "A new class of transport distances between measures". *Calculus of Variations and Partial Differential Equations.* 34(2): 193–231.

Dolinsky, Y. and H. M. Soner. 2014. "Martingale optimal transport and robust hedging in continuous time". *Probability Theory and Related Fields.* 160(1-2): 391–427.

Dudley, R. M. 1969. "The speed of mean Glivenko-Cantelli convergence". *Annals of Mathematical Statistics.* 40(1): 40–50.

Dupuy, A., A. Galichon, and Y. Sun. 2016. "Estimating matching affinity matrix under low-rank constraints". *Arxiv:1612.09585.*

Dvurechenskii, P., D. Dvinskikh, A. Gasnikov, C. Uribe, and A. Nedich. 2018. "Decentralize and Randomize: Faster Algorithm for Wasserstein Barycenters". In: *Advances in Neural Information Processing Systems 31.* Ed. by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett. Curran Associates, Inc. 10783–10793.

Dvurechensky, P., A. Gasnikov, and A. Kroshnin. 2018. "Computational Optimal Transport: Complexity by Accelerated Gradient Descent Is Better Than by Sinkhorn's Algorithm". In: *Proceedings of the 35th International Conference on Machine Learning.* Ed. by J. Dy and A. Krause. Vol. 80. *Proceedings of Machine Learning Research.* PMLR. 1367–1376.

Dykstra, R. L. 1983. "An algorithm for restricted least squares regression". *Journal American Statistical Association.* 78(384): 839–842.

Dykstra, R. L. 1985. "An iterative procedure for obtaining *I*-projections onto the intersection of convex sets". *Annals of Probability.* 13(3): 975–984.

Eckstein, J. and D. P. Bertsekas. 1992. "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators". *Mathematical Programming.* 55(3): 293–318.

Edwards, D. A. 1975. "The structure of superspace". In: *Studies in topology.* Elsevier. 121–133.

El Moselhy, T. A. and Y. M. Marzouk. 2012. "Bayesian inference with optimal maps". *Journal of Computational Physics.* 231(23): 7815–7850.

Endres, D. M. and J. E. Schindelin. 2003. "A new metric for probability distributions". *IEEE Transactions on Information theory.* 49(7): 1858–1860.

Erbar, M. 2010. "The heat equation on manifolds as a gradient flow in the Wasserstein space". *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques.* 46(1): 1–23.

Erbar, M. and J. Maas. 2014. "Gradient flow structures for discrete porous medium equations". *Discrete and Continuous Dynamical Systems.* 34(4): 1355–1374.

Erlander, S. 1980. *Optimal Spatial Interaction and the Gravity Model.* Vol. 173. Springer-Verlag.

Erlander, S. and N. F. Stewart. 1990. *The Gravity Model in Transportation Analysis: Theory and Extensions.*

Esfahani, P. M. and D. Kuhn. 2018. "Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations". *Mathematical Programming.* 171(1-2): 115–166.

Essid, M. and J. Solomon. 2017. "Quadratically-regularized optimal transport on graphs". *arXiv preprint arXiv:1704.08200.*

Evans, L. C. and W. Gangbo. 1999. *Differential Equations Methods for the Monge-Kantorovich Mass Transfer Problem.* Vol. 653. American Mathematical Society.

Feldman, M. and R. McCann. 2002. "Monge's transport problem on a Riemannian manifold". *Transaction AMS.* 354(4): 1667–1697.

Feydy, J., B. Charlier, F.-X. Vialard, and G. Peyré. 2017. "Optimal transport for diffeomorphic registration". In: *Proceedings of MIC-CAI'17*. Springer. 291–299.

Feydy, J., T. Séjourné, F.-X. Vialard, S.-I. Amari, A. Trouvé, and G. Peyré. 2018. "Interpolating between optimal transport and MMD using Sinkhorn divergences". *arXiv preprint arXiv:1810.08278*.

Figalli, A. 2010. "The optimal partial transport problem". *Archive for Rational Mechanics and Analysis*. 195(2): 533–560.

Flamary, R., C. Févotte, N. Courty, and V. Emiya. 2016. "Optimal spectral transportation with application to music transcription". In: *Advances in Neural Information Processing Systems*. 703–711.

Ford, L. R. and D. R. Fulkerson. 1962. *Flows in Networks*. Princeton University Press.

Forrester, P. J. and M. Kieburg. 2016. "Relating the Bures measure to the Cauchy two-matrix model". *Communications in Mathematical Physics*. 342(1): 151–187.

Fournier, N. and A. Guillin. 2015. "On the rate of convergence in Wasserstein distance of the empirical measure". *Probability Theory and Related Fields*. 162(3-4): 707–738.

Franklin, J. and J. Lorenz. 1989. "On the scaling of multidimensional matrices". *Linear Algebra and its Applications*. 114: 717–735.

Frisch, U., S. Matarrese, R. Mohayaee, and A. Sobolevski. 2002. "A reconstruction of the initial conditions of the universe by optimal mass transportation". *Nature*. 417(6886): 260–262.

Froese, B. D. and A. M. Oberman. 2011. "Convergent finite difference solvers for viscosity solutions of the elliptic Monge–Ampère equation in dimensions two and higher". *SIAM Journal on Numerical Analysis*. 49(4): 1692–1714.

Frogner, C., C. Zhang, H. Mobahi, M. Araya, and T. A. Poggio. 2015. "Learning with a Wasserstein loss". In: *Advances in Neural Information Processing Systems*. 2053–2061.

Gabay, D. and B. Mercier. 1976. "A dual algorithm for the solution of nonlinear variational problems via finite element approximation". *Computers & Mathematics with Applications*. 2(1): 17–40.

Galichon, A. 2016. *Optimal Transport Methods in Economics*. Princeton University Press.

Galichon, A., P. Henry-Labordère, and N. Touzi. 2014. "A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options". *Annals of Applied Probability*. 24(1): 312–336.

Galichon, A. and B. Salanié. 2009. "Matching with trade-offs: revealed preferences over competing characteristics". *Tech. rep.* Preprint SSRN-1487307.

Gallouët, T. O. and Q. Mérigot. 2017. "A Lagrangian scheme à la Brenier for the incompressible Euler equations". *Foundations of Computational Mathematics*. 18: 1–31.

Gallouët, T. O. and L. Monsaingeon. 2017. "A JKO splitting scheme for Kantorovich–Fisher–Rao gradient flows". *SIAM Journal on Mathematical Analysis*. 49(2): 1100–1130.

Gangbo, W. and R. J. McCann. 1996. "The geometry of optimal transportation". *Acta Mathematica*. 177(2): 113–161.

Gangbo, W. and A. Swiech. 1998. "Optimal maps for the multidimensional Monge-Kantorovich problem". *Communications on Pure and Applied Mathematics*. 51(1): 23–45.

GAO, R., L. Xie, Y. Xie, and H. Xu. 2018. "Robust Hypothesis Testing Using Wasserstein Uncertainty Sets". In: *Advances in Neural Information Processing Systems 31*. Ed. by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett. Curran Associates, Inc. 7913–7923.

Gelbrich, M. 1990. "On a formula for the L^2 Wasserstein metric between measures on Euclidean and Hilbert spaces". *Mathematische Nachrichten*. 147(1): 185–203.

Genevay, A., L. Chizat, F. Bach, M. Cuturi, and G. Peyré. 2018a. "Sample complexity of Sinkhorn divergences". *arXiv preprint arXiv:1810.02733*.

Genevay, A., M. Cuturi, G. Peyré, and F. Bach. 2016. "Stochastic optimization for large-scale optimal transport". In: *Advances in Neural Information Processing Systems*. 3440–3448.

Genevay, A., G. Peyré, and M. Cuturi. 2017. "GAN and VAE from an optimal transport point of view". (arXiv preprint arXiv:1706.01807).

Genevay, A., G. Peyré, and M. Cuturi. 2018b. "Learning generative models with Sinkhorn divergences". In: *Proceedings of AISTATS*. 1608–1617.

Gentil, I., C. Léonard, and L. Ripani. 2015. "About the analogy between optimal transport and minimal entropy". *arXiv preprint arXiv:1510.08230*.

George, A. and J. W. Liu. 1989. "The evolution of the minimum degree ordering algorithm". *SIAM Review*. 31(1): 1–19.

Georgiou, T. T. and M. Pavon. 2015. "Positive contraction mappings for classical and quantum Schrödinger systems". *Journal of Mathematical Physics*. 56(3): 033301.

Getreuer, P. 2013. "A survey of Gaussian convolution algorithms". *Image Processing On Line*. 2013: 286–310.

Gianazza, U., G. Savaré, and G. Toscani. 2009. "The Wasserstein gradient flow of the Fisher information and the quantum drift-diffusion equation". *Archive for Rational Mechanics and Analysis*. 194(1): 133–220.

Gibbs, A. L. and F. E. Su. 2002. "On choosing and bounding probability metrics". *International Statistical Review*. 70(3): 419–435.

Glaunes, J., A. Trouvé, and L. Younes. 2004. "Diffeomorphic matching of distributions: a new approach for unlabelled point-sets and submanifolds matching". In: *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Vol. 2.

Glorot, X., A. Bordes, and Y. Bengio. 2011. "Domain adaptation for large-scale sentiment classification: A deep learning approach". In: *Proceedings of the 28th International Conference on Machine Learning*. 513–520.

Glowinski, R. and A. Marroco. 1975. "Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires". *ESAIM: Mathematical Modelling and Numerical Analysis*. 9(R2): 41–76.

Goes, F. de, C. Wallez, J. Huang, D. Pavlov, and M. Desbrun. 2015. "Power particles: an incompressible fluid solver based on power diagrams". *ACM Transaction Graphics*. 34(4): 50:1–50:11.

Gold, S. and A. Rangarajan. 1996. "A graduated assignment algorithm for graph matching". *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 18(4): 377–388.

Gold, S., A. Rangarajan, C.-P. Lu, S. Pappu, and E. Mjolsness. 1998. "New algorithms for 2D and 3D point matching: pose estimation and correspondence". *Pattern Recognition.* 31(8): 1019–1031.

Gómez, E., M. A. Gómez-Villegas, and J. M. Marín. 2003. "A survey on continuous elliptical vector distributions". *Rev. Mat. Complut.* 16: 345–361.

Gori-Giorgi, P., M. Seidl, and G. Vignale. 2009. "Density-functional theory for strongly interacting electrons". *Physical Review Letters.* 103(16): 166402.

Gramfort, A., G. Peyré, and M. Cuturi. 2015. "Fast Optimal Transport Averaging of Neuroimaging Data". In: *Information Processing in Medical Imaging - 24th International Conference, IPMI 2015.* 261–272.

Grauman, K. and T. Darrell. 2005. "The pyramid match kernel: discriminative classification with sets of image features". In: *Tenth IEEE International Conference on Computer Vision.* Vol. 2. IEEE. 1458–1465.

Grave, E., A. Joulin, and Q. Berthet. 2018. "Unsupervised Alignment of Embeddings with Wasserstein Procrustes". *arXiv preprint arXiv:1805.11222.*

Gretton, A., K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola. 2012. "A kernel two-sample test". *Journal of Machine Learning Research.* 13(Mar): 723–773.

Gretton, A., K. M. Borgwardt, M. Rasch, B. Schölkopf, and A. J. Smola. 2007. "A kernel method for the two-sample-problem". In: *Advances in Neural Information Processing Systems.* 513–520.

Griewank, A. 1992. "Achieving logarithmic growth of temporal and spatial complexity in reverse automatic differentiation". *Optimization Methods and Software.* 1(1): 35–54.

Griewank, A. and A. Walther. 2008. *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation.* SIAM.

Gromov, M. 2001. *Metric Structures for Riemannian and Non-Riemannian Spaces. Progress in Mathematics.* Birkhäuser.

Guo, G. and J. Obloj. 2017. "Computational methods for martingale optimal transport problems". *arXiv preprint arXiv:1710.07911*.

Gurvits, L. 2004. "Classical complexity and quantum entanglement". *Journal of Computer and System Sciences*. 69(3): 448–484.

Gutiérrez, C. E. 2016. *The Monge-Ampere Equation*. Springer.

Gutierrez, J., J. Rabin, B. Galerne, and T. Hurtut. 2017. "Optimal patch assignment for statistically constrained texture synthesis". In: *International Conference on Scale Space and Variational Methods in Computer Vision*. Springer. 172–183.

Hadjidimos, A. 2000. "Successive overrelaxation (SOR) and related methods". *Journal of Computational and Applied Mathematics*. 123(1): 177–199.

Haker, S., L. Zhu, A. Tannenbaum, and S. Angenent. 2004. "Optimal mass transport for registration and warping". *International Journal of Computer Vision*. 60(3): 225–240.

Hanin, L. G. 1992. "Kantorovich-Rubinstein norm and its application in the theory of Lipschitz spaces". *Proceedings of the American Mathematical Society*. 115(2): 345–352.

Hashimoto, T., D. Gifford, and T. Jaakkola. 2016. "Learning population-level diffusions with generative RNNs". In: *International Conference on Machine Learning*. 2417–2426.

Hilbert, D. 1895. "Über die gerade Linie als kürzeste Verbindung zweier Punkte". *Mathematische Annalen*. 46(1): 91–96.

Hitchcock, F. L. 1941. "The distribution of a product from several sources to numerous localities". *Studies in Applied Mathematics*. 20(1-4): 224–230.

Ho, N., X. Nguyen, M. Yurochkin, H. H. Bui, V. Huynh, and D. Phung. 2017. "Multilevel clustering via Wasserstein means". *arXiv preprint arXiv:1706.03883*.

Hofmann, T., B. Schölkopf, and A. J. Smola. 2008. "Kernel methods in machine learning". *Annals of Statistics*. 36(3): 1171–1220.

Hosmer Jr, D. W., S. Lemeshow, and R. X. Sturdivant. 2013. *Applied Logistic Regression*. Vol. 398. John Wiley & Sons.

Huang, G., C. Guo, M. J. Kusner, Y. Sun, F. Sha, and K. Q. Weinberger. 2016. "Supervised word mover's distance". In: *Advances in Neural Information Processing Systems*. 4862–4870.

Idel, M. 2016. "A review of matrix scaling and Sinkhorn's normal form for matrices and positive maps". *arXiv preprint arXiv:1609.06349.*

Indyk, P. and E. Price. 2011. "K-median clustering, model-based compressive sensing, and sparse recovery for earth mover distance". In: *Proceedings of the forty-third annual ACM Symposium on Theory of Computing.* ACM. 627–636.

Indyk, P. and N. Thaper. 2003. "Fast image retrieval via embeddings". In: *3rd International Workshop on Statistical and Computational Theories of Vision.*

Jiang, X., L. Ning, and T. T. Georgiou. 2012. "Distances and Riemannian metrics for multivariate spectral densities". *IEEE Transactions on Automatic Control.* 57(7): 1723–1735.

Johnson, W. B. and J. Lindenstrauss. 1984. "Extensions of Lipschitz mappings into a Hilbert space". In: *Conference in Modern Analysis and Probability (New Haven, Conn., 1982).* Vol. 26. *Contemporary Mathematics.* American Mathematical Society. 189–206.

Jordan, R., D. Kinderlehrer, and F. Otto. 1998. "The variational formulation of the Fokker-Planck equation". *SIAM Journal on Mathematical Analysis.* 29(1): 1–17.

Kantorovich, L. 1942. "On the transfer of masses (in Russian)". *Doklady Akademii Nauk.* 37(2): 227–229.

Kantorovich, L. and G. Rubinstein. 1958. "On a space of totally additive functions". *Vestn Leningrad Universitet.* 13: 52–59.

Karcher, H. 2014. "Riemannian center of mass and so called Karcher mean". *arXiv preprint arXiv:1407.2087.*

Karlsson, J. and A. Ringh. 2016. "Generalized Sinkhorn iterations for regularizing inverse problems using optimal mass transport". *arXiv preprint arXiv:1612.02273.*

Kim, S., R. Ma, D. Mesa, and T. P. Coleman. 2013. "Efficient Bayesian inference methods via convex optimization and optimal transport". In: *IEEE International Symposium on Information Theory.* IEEE. 2259–2263.

Kinderlehrer, D. and N. J. Walkington. 1999. "Approximation of parabolic equations using the Wasserstein metric". *ESAIM: Mathematical Modelling and Numerical Analysis.* 33(04): 837–852.

Kitagawa, J., Q. Mérigot, and B. Thibert. 2016. "A Newton algorithm for semi-discrete optimal transport". *arXiv preprint arXiv:1603.05579.*

Knight, P. A. 2008. "The Sinkhorn–Knopp algorithm: convergence and applications". *SIAM Journal on Matrix Analysis and Applications.* 30(1): 261–275.

Knight, P. A. and D. Ruiz. 2013. "A fast algorithm for matrix balancing". *IMA Journal of Numerical Analysis.* 33(3): 1029–1047.

Knight, P. A., D. Ruiz, and B. Uccar. 2014. "A symmetry preserving algorithm for matrix scaling". *SIAM Journal on Matrix Analysis and Applications.* 35(3): 931–955.

Knott, M. and C. S. Smith. 1984. "On the optimal mapping of distributions". *Journal of Optimization Theory and Applications.* 43(1): 39–49.

Knott, M. and C. S. Smith. 1994. "On a generalization of cyclic monotonicity and distances among random vectors". *Linear Algebra and Its Applications.* 199: 363–371.

Kolouri, S., S. R. Park, M. Thorpe, D. Slepcev, and G. K. Rohde. 2017. "Optimal mass transport: signal processing and machine-learning applications". *IEEE Signal Processing Magazine.* 34(4): 43–59.

Kolouri, S., Y. Zou, and G. K. Rohde. 2016. "Sliced Wasserstein kernels for probability distributions". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 5258–5267.

Kondratyev, S., L. Monsaingeon, and D. Vofnikov. 2016. "A new optimal transport distance on the space of finite Radon measures". *Advances in Differential Equations.* 21(11/12): 1117–1164.

Koopmans, T. C. 1949. "Optimum utilization of the transportation system". *Econometrica: Journal of the Econometric Society*: 136–146.

Korman, J. and R. McCann. 2015. "Optimal transportation with capacity constraints". *Transactions of the American Mathematical Society.* 367(3): 1501–1521.

Korte, B. and J. Vygen. 2012. *Combinatorial Optimization.* Springer.

Kosowsky, J. and A. L. Yuille. 1994. "The invisible hand algorithm: Solving the assignment problem with statistical physics". *Neural networks.* 7(3): 477–490.

Kruithof, J. 1937. "Telefoonverkeersrekening". *De Ingenieur*. 52: E15–E25.

Kuhn, H. W. 1955. "The Hungarian method for the assignment problem". *Naval Research Logistics Quarterly*. 2: 83–97.

Kulis, B. 2012. "Metric learning: a survey". *Foundations and Trends in Machine Learning*. 5(4): 287–364.

Kusner, M., Y. Sun, N. Kolkin, and K. Weinberger. 2015. "From word embeddings to document distances". In: *International Conference on Machine Learning*. 957–966.

Lacombe, T., M. Cuturi, and S. Oudot. 2018. "Large scale computation of means and clusters for persistence diagrams using optimal transport". *Advances in Neural Information Processing Systems 31*: 9792–9802. Ed. by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett.

Lai, R. and H. Zhao. 2014. "Multi-scale non-rigid point cloud registration using robust sliced-Wasserstein distance via Laplace-Beltrami eigenmap". *arXiv preprint arXiv:1406.3758*.

Lavenant, H. 2017. "Harmonic mappings valued in the Wasserstein space". *Preprint cvgmt 3649*.

Lazebnik, S., C. Schmid, and J. Ponce. 2006. "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories". In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Vol. 2. IEEE. 2169–2178.

Le Gouic, T. and J.-M. Loubes. 2016. "Existence and consistency of Wasserstein barycenters". *Probability Theory and Related Fields*. 168: 901–917.

Lee, D. D. and H. S. Seung. 1999. "Learning the parts of objects by non-negative matrix factorization". *Nature*. 401(6755): 788–791.

Lee, J. and M. Raginsky. 2018. "Minimax Statistical Learning with Wasserstein distances". In: *Advances in Neural Information Processing Systems 31*. Ed. by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett. Curran Associates, Inc. 2692–2701.

Leeb, W. and R. Coifman. 2016. "Hölder–Lipschitz norms and their duals on spaces with semigroups, with applications to earth mover's distance". *Journal of Fourier Analysis and Applications.* 22(4): 910–953.

Lellmann, J., D. A. Lorenz, C. Schönlieb, and T. Valkonen. 2014. "Imaging with Kantorovich–Rubinstein discrepancy". *SIAM Journal on Imaging Sciences.* 7(4): 2833–2859.

Lemmens, B. and R. Nussbaum. 2012. *Nonlinear Perron-Frobenius Theory.* Vol. 189. Cambridge University Press.

Léonard, C. 2012. "From the Schrödinger problem to the Monge–Kantorovich problem". *Journal of Functional Analysis.* 262(4): 1879–1920.

Léonard, C. 2014. "A survey of the Schrödinger problem and some of its connections with optimal transport". *Discrete Continuous Dynamical Systems Series A.* 34(4): 1533–1574.

Lévy, B. 2015. "A numerical algorithm for L^2 semi-discrete optimal transport in 3D". *ESAIM: Mathematical Modelling and Numerical Analysis.* 49(6): 1693–1715.

Lévy, B. and E. L. Schwindt. 2018. "Notions of optimal transport theory and how to implement them on a computer". *Computers & Graphics.* 72: 135–148.

Li, P., Q. Wang, and L. Zhang. 2013. "A novel earth mover's distance methodology for image matching with Gaussian mixture models". In: *Proceedings of the IEEE International Conference on Computer Vision.* 1689–1696.

Li, W., E. K. Ryu, S. Osher, W. Yin, and W. Gangbo. 2018a. "A parallel method for Earth Mover's distance". *Journal of Scientific Computing.* 75(1): 182–197.

Li, Y., W. Li, and G. Cao. 2018b. "Image segmentation via L^1 Monge-Kantorovich problem". *CAM report 17-73.*

Liero, M., A. Mielke, and G. Savaré. 2016. "Optimal transport in competition with reaction: the Hellinger–Kantorovich distance and geodesic curves". *SIAM Journal on Mathematical Analysis.* 48(4): 2869–2911.

Liero, M., A. Mielke, and G. Savaré. 2018. "Optimal entropy-transport problems and a new Hellinger–Kantorovich distance between positive measures". *Inventiones Mathematicae.* 211(3): 969–1117.

Ling, H. and K. Okada. 2006. "Diffusion distance for histogram comparison". In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition.* Vol. 1. IEEE. 246–253.

Ling, H. and K. Okada. 2007. "An efficient earth mover's distance algorithm for robust histogram comparison". *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 29(5): 840–853.

Linial, N., A. Samorodnitsky, and A. Wigderson. 1998. "A deterministic strongly polynomial algorithm for matrix scaling and approximate permanents". In: *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing.* ACM. 644–652.

Lions, P.-L. and B. Mercier. 1979. "Splitting algorithms for the sum of two nonlinear operators". *SIAM Journal on Numerical Analysis.* 16: 964–979.

Loftsgaarden, D. O. and C. P. Quesenberry. 1965. "A nonparametric estimate of a multivariate density function". *Annals of Mathematical Statistics.* 36(3): 1049–1051.

Loiola, E. M., N. M. M. de Abreu, P. O. Boaventura-Netto, P. Hahn, and T. Querido. 2007. "A survey for the quadratic assignment problem". *European Journal Operational Research.* 176(2): 657–690.

Lyzinski, V., D. E. Fishkind, M. Fiori, J. T. Vogelstein, C. E. Priebe, and G. Sapiro. 2016. "Graph matching: relax at your own risk". *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 38(1): 60–73.

Maas, J. 2011. "Gradient flows of the entropy for finite Markov chains". *Journal of Functional Analysis.* 261(8): 2250–2292.

Maas, J., M. Rumpf, C. Schönlieb, and S. Simon. 2015. "A generalized model for optimal transport of images including dissipation and density modulation". *ESAIM: Mathematical Modelling and Numerical Analysis.* 49(6): 1745–1769.

Maas, J., M. Rumpf, and S. Simon. 2016. "Generalized optimal transport with singular sources". *arXiv preprint arXiv:1607.01186.*

Makihara, Y. and Y. Yagi. 2010. "Earth mover's morphing: Topology-free shape morphing using cluster-based EMD flows". In: *Asian Conference on Computer Vision*. Springer. 202–215.

Malagò, L., L. Montrucchio, and G. Pistone. 2018. "Wasserstein-Riemannian geometry of positive-definite matrices". *arXiv preprint arXiv:1801.09269*.

Mallasto, A. and A. Feragen. 2017. "Learning from uncertain curves: The 2-Wasserstein metric for Gaussian processes". In: *Advances in Neural Information Processing Systems 30*. Ed. by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett. Curran Associates, Inc. 5660–5670.

Mallat, S. 2008. *A Wavelet Tour of Signal Processing: the Sparse Way*. Academic press.

Mathon, B., F. Cayre, P. Bas, and B. Macq. 2014. "Optimal transport for secure spread-spectrum watermarking of still images". *IEEE Transactions on Image Processing*. 23(4): 1694–1705.

Matthes, D. and H. Osberger. 2014. "Convergence of a variational Lagrangian scheme for a nonlinear drift diffusion equation". *ESAIM: Mathematical Modelling and Numerical Analysis*. 48(3): 697–726.

Matthes, D. and H. Osberger. 2017. "A convergent Lagrangian discretization for a nonlinear fourth-order equation". *Foundations of Computational Mathematics*. 17(1): 73–126.

Maury, B. and A. Preux. 2017. "Pressureless Euler equations with maximal density constraint: a time-splitting scheme". *Topological Optimization and Optimal Transport: In the Applied Sciences*. 17: 333.

Maury, B., A. Roudneff-Chupin, and F. Santambrogio. 2010. "A macroscopic crowd motion model of gradient flow type". *Mathematical Models and Methods in Applied Sciences*. 20(10): 1787–1821.

McCann, R. J. 1997. "A convexity principle for interacting gases". *Advances in Mathematics*. 128(1): 153–179.

Mémoli, F. 2007. "On the use of Gromov–Hausdorff distances for shape comparison". In: *Symposium on Point Based Graphics*. 81–90.

Mémoli, F. 2011. "Gromov–Wasserstein distances and the metric approach to object matching". *Foundations of Computational Mathematics*. 11(4): 417–487.

Mémoli, F. and G. Sapiro. 2005. "A theoretical and computational framework for isometry invariant recognition of point cloud data". *Foundations of Computational Mathematics*. 5(3): 313–347.

Mérigot, Q. 2011. "A multiscale approach to optimal transport". *Computer Graphics Forum*. 30(5): 1583–1592.

Mérigot, Q., J. Meyron, and B. Thibert. 2017. "Light in power: a general and parameter-free algorithm for caustic design". *arXiv preprint arXiv:1708.04820*.

Métivier, L., R. Brossier, Q. Merigot, E. Oudet, and J. Virieux. 2016. "An optimal transport approach for seismic tomography: Application to 3D full waveform inversion". *Inverse Problems*. 32(11): 115008.

Mielke, A. 2013. "Geodesic convexity of the relative entropy in reversible Markov chains". *Calculus of Variations and Partial Differential Equations*. 48(1-2): 1–31.

Mikolov, T., K. Chen, G. Corrado, and J. Dean. 2013. "Efficient estimation of word representations in vector space". *arXiv preprint arXiv:1301.3781*.

Mirebeau, J.-M. 2015. "Discretization of the 3D Monge-Ampere operator, between wide stencils and power diagrams". *ESAIM: Mathematical Modelling and Numerical Analysis*. 49(5): 1511–1523.

Monge, G. 1781. "Mémoire sur la théorie des déblais et des remblais". *Histoire de l'Académie Royale des Sciences*: 666–704.

Montavon, G., K.-R. Müller, and M. Cuturi. 2016. "Wasserstein training of restricted Boltzmann machines". In: *Advances in Neural Information Processing Systems 29*. Ed. by D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett. 3718–3726.

Moon, K. and A. Hero. 2014. "Multivariate f-divergence estimation with confidence". In: *Advances in Neural Information Processing Systems*. 2420–2428.

Museyko, O., M. Stiglmayr, K. Klamroth, and G. Leugering. 2009. "On the application of the Monge–Kantorovich problem to image registration". *SIAM Journal on Imaging Sciences*. 2(4): 1068–1097.

Muzellec, B. and M. Cuturi. 2018. "Generalizing Point Embeddings using the Wasserstein Space of Elliptical Distributions". In: *Advances in Neural Information Processing Systems 31*. Ed. by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett. Curran Associates, Inc. 10258–10269.

Muzellec, B., R. Nock, G. Patrini, and F. Nielsen. 2017. "Tsallis regularized optimal transport and ecological inference." In: *AAAI*. 2387–2393.

Myronenko, A. and X. Song. 2010. "Point set registration: coherent point drift". *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 32(12): 2262–2275.

Naor, A. and G. Schechtman. 2007. "Planar Earthmover Is Not in L_1". *SIAM Journal on Computing.* 37(3): 804–826.

Neidinger, R. D. 2010. "Introduction to automatic differentiation and Matlab object-oriented programming". *SIAM Review.* 52(3): 545–563.

Nemirovski, A. and U. Rothblum. 1999. "On complexity of matrix scaling". *Linear Algebra and its Applications.* 302: 435–460.

Nesterov, Y. and A. Nemirovskii. 1994. *Interior-point polynomial algorithms in convex programming.* Vol. 13. SIAM.

Ni, K., X. Bresson, T. Chan, and S. Esedoglu. 2009. "Local histogram based segmentation using the Wasserstein distance". *International Journal of Computer Vision.* 84(1): 97–111.

Ning, L. and T. T. Georgiou. 2014. "Metrics for matrix-valued measures via test functions". In: *53rd IEEE Conference on Decision and Control.* IEEE. 2642–2647.

Ning, L., T. T. Georgiou, and A. Tannenbaum. 2015. "On matrix-valued Monge–Kantorovich optimal mass transport". *IEEE Transactions on Automatic Control.* 60(2): 373–382.

Nocedal, J. and S. J. Wright. 1999. *Numerical Optimization.* Springer-Verlag.

Oberman, A. M. and Y. Ruan. 2015. "An efficient linear programming method for optimal transportation". *arXiv preprint arXiv:1509.03668.*

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

Oliva, A. and A. Torralba. 2001. "Modeling the shape of the scene: a holistic representation of the spatial envelope". *International Journal of Computer Vision.* 42(3): 145–175.

Oliver, D. S. 2014. "Minimization for conditional simulation: Relationship to optimal transport". *Journal of Computational Physics.* 265: 1–15.

Orlin, J. B. 1997. "A polynomial time primal network simplex algorithm for minimum cost flows". *Mathematical Programming.* 78(2): 109–129.

Österreicher, F. and I. Vajda. 2003. "A new class of metric divergences on probability spaces and its applicability in statistics". *Annals of the Institute of Statistical Mathematics.* 55(3): 639–653.

Otto, F. 2001. "The geometry of dissipative evolution equations: the porous medium equation". *Communications in Partial Differential Equations.* 26(1-2): 101–174.

Owen, A. B. 2001. *Empirical Likelihood.* Wiley Online Library.

Pan, S. J. and Q. Yang. 2010. "A survey on transfer learning". *IEEE Transactions on knowledge and data engineering.* 22(10): 1345–1359.

Panaretos, V. M. and Y. Zemel. 2016. "Amplitude and phase variation of point processes". *Annals of Statistics.* 44(2): 771–812.

Papadakis, N., G. Peyré, and E. Oudet. 2014. "Optimal transport with proximal splitting". *SIAM Journal on Imaging Sciences.* 7(1): 212–238.

Pass, B. 2012. "On the local structure of optimal measures in the multi-marginal optimal transportation problem". *Calculus of Variations and Partial Differential Equations.* 43(3-4): 529–536.

Pass, B. 2015. "Multi-marginal optimal transport: theory and applications". *ESAIM: Mathematical Modelling and Numerical Analysis.* 49(6): 1771–1790.

Pele, O. and B. Taskar. 2013. "The tangent earth mover's distance". In: *Geometric Science of Information.* Springer. 397–404.

Pele, O. and M. Werman. 2008. "A linear time histogram metric for improved sift matching". *Computer Vision–ECCV 2008*: 495–508.

Pele, O. and M. Werman. 2009. "Fast and robust earth mover's distances". In: *IEEE 12th International Conference on Computer Vision.* 460–467.

Perthame, B., F. Quirós, and J. L. Vázquez. 2014. "The Hele-Shaw asymptotics for mechanical models of tumor growth". *Archive for Rational Mechanics and Analysis.* 212(1): 93–127.

Peyré, G. 2015. "Entropic approximation of Wasserstein gradient flows". *SIAM Journal on Imaging Sciences.* 8(4): 2323–2351.

Peyré, G., L. Chizat, F.-X. Vialard, and J. Solomon. 2017. "Quantum entropic regularization of matrix-valued optimal transport". *to appear in European Journal of Applied Mathematics.*

Peyré, G., M. Cuturi, and J. Solomon. 2016. "Gromov-Wasserstein averaging of kernel and distance matrices". In: *International Conference on Machine Learning.* 2664–2672.

Peyré, G., J. Fadili, and J. Rabin. 2012. "Wasserstein active contours". In: *19th IEEE International Conference on Image Processing.* IEEE. 2541–2544.

Peyre, R. 2011. "Comparison between W_2 distance and H^{-1} norm, and localisation of Wasserstein distance". *arXiv preprint arXiv:1104.4631.*

Piccoli, B. and F. Rossi. 2014. "Generalized Wasserstein distance and its application to transport equations with source". *Archive for Rational Mechanics and Analysis.* 211(1): 335–358.

Pitié, F., A. C. Kokaram, and R. Dahyot. 2007. "Automated colour grading using colour distribution transfer". *Computer Vision and Image Understanding.* 107(1): 123–137.

Pytorch. 2017. "Pytorch library". *http://pytorch.org/.*

Rabin, J. and N. Papadakis. 2015. "Convex color image segmentation with optimal transport distances". In: *Proceedings of SSVM'15.* 256–269.

Rabin, J., G. Peyré, J. Delon, and M. Bernot. 2011. "Wasserstein barycenter and its application to texture mixing". In: *International Conference on Scale Space and Variational Methods in Computer Vision.* Springer. 435–446.

Rachev, S. T. and L. Rüschendorf. 1998a. *Mass Transportation Problems: Volume I: Theory.* Springer Science & Business Media.

Rachev, S. T. and L. Rüschendorf. 1998b. *Mass Transportation Problems: Volume II: Applications.* Springer Science & Business Media.

Rall, L. B. 1981. *Automatic Differentiation: Techniques and Applications.* Springer.

Ramdas, A., N. G. Trillos, and M. Cuturi. 2017. "On Wasserstein two-sample testing and related families of nonparametric tests". *Entropy.* 19(2): 47.

Rangarajan, A., A. L. Yuille, S. Gold, and E. Mjolsness. 1999. "Convergence properties of the softassign quadratic assignment algorithm". *Neural Computation.* 11(6): 1455–1474.

Reich, S. 2013. "A nonparametric ensemble transform method for Bayesian inference". *SIAM Journal on Scientific Computing.* 35(4): A2013–A2024.

Rockafellar, R. T. 1976. "Monotone operators and the proximal point algorithm". *SIAM Journal on Control and Optimization.* 14(5): 877–898.

Rolet, A., M. Cuturi, and G. Peyré. 2016. "Fast dictionary learning with a smoothed Wasserstein loss". In: *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics.* Vol. 51. *Proceedings of Machine Learning Research.* 630–638.

Rubner, Y., C. Tomasi, and L. J. Guibas. 2000. "The earth mover's distance as a metric for image retrieval". *International Journal of Computer Vision.* 40(2): 99–121.

Ruschendorf, L. 1995. "Convergence of the iterative proportional fitting Procedure". *Annals of Statistics.* 23(4): 1160–1174.

Rüschendorf, L. and S. T. Rachev. 1990. "A characterization of random variables with minimum L2-distance". *Journal of Multivariate Analysis.* 32(1): 48–54.

Rüschendorf, L. and W. Thomsen. 1998. "Closedness of sum spaces and the generalized Schrodinger problem". *Theory of Probability and its Applications.* 42(3): 483–494.

Rüschendorf, L. and L. Uckelmann. 2002. "On the n-coupling problem". *Journal of Multivariate Analysis.* 81(2): 242–258.

Ryu, E. K., Y. Chen, W. Li, and S. Osher. 2017a. "Vector and matrix optimal mass transport: theory, algorithm, and applications". *SIAM Journal on Scientific Compututing.* 40(5): A3675–A3698.

Ryu, E. K., W. Li, P. Yin, and S. Osher. 2017b. "Unbalanced and partial L^1 Monge-Kantorovich problem: a scalable parallel first-order method". *Journal of Scientific Computing.* 75(3): 1596–1613.

Salimans, T., H. Zhang, A. Radford, and D. Metaxas. 2018. "Improving GANs Using Optimal Transport". In: *International Conference on Learning Representations.*

Samelson, H. *et al.* 1957. "On the Perron-Frobenius theorem." *Michigan Mathematical Journal.* 4(1): 57–59.

Sandler, R. and M. Lindenbaum. 2011. "Nonnegative matrix factorization with earth mover's distance metric for image analysis". *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 33(8): 1590–1602.

Santambrogio, F. 2015. *Optimal transport for applied mathematicians.* Birkhauser.

Santambrogio, F. 2017. "{Euclidean, metric, and Wasserstein} gradient flows: an overview". *Bulletin of Mathematical Sciences.* 7(1): 87–154.

Santambrogio, F. 2018. "Crowd motion and population dynamics under density constraints". *GMT preprint 3728.*

Saumier, L.-P., B. Khouider, and M. Agueh. 2015. "Optimal transport for particle image velocimetry". *Communications in Mathematical Sciences.* 13(1): 269–296.

Schiebinger, G., J. Shu, M. Tabaka, B. Cleary, V. Subramanian, A. Solomon, S. Liu, S. Lin, P. Berube, L. Lee, *et al.* 2017. "Reconstruction of developmental landscapes by optimal-transport analysis of single-cell gene expression sheds light on cellular reprogramming." *bioRxiv:* 191056.

Schmitzer, B. 2016a. "A sparse multiscale algorithm for dense optimal transport". *Journal of Mathematical Imaging and Vision.* 56(2): 238–259.

Schmitzer, B. 2016b. "Stabilized sparse scaling algorithms for entropy regularized transport problems". *arXiv preprint arXiv:1610.06519.*

Schmitzer, B. and C. Schnörr. 2013a. "Modelling convex shape priors and matching based on the Gromov-Wasserstein distance". *Journal of Mathematical Imaging and Vision*. 46(1): 143–159.

Schmitzer, B. and C. Schnörr. 2013b. "Object segmentation by shape matching with Wasserstein modes". In: *International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*. Springer. 123–136.

Schmitzer, B. and B. Wirth. 2017. "A framework for Wasserstein-1-type metrics". *arXiv preprint arXiv:1701.01945*.

Schoenberg, I. J. 1938. "Metric spaces and positive definite functions". *Transactions of the American Mathematical Society*. 38: 522–356.

Schölkopf, B. and A. J. Smola. 2002. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press.

Schrödinger, E. 1931. "Über die Umkehrung der Naturgesetze". *Sitzungsberichte Preuss. Akad. Wiss. Berlin. Phys. Math.* 144: 144–153.

Seguy, V. and M. Cuturi. 2015. "Principal geodesic analysis for probability measures under the optimal transport metric". In: *Advances in Neural Information Processing Systems 28*. 3294–3302.

Seguy, V., B. B. Damodaran, R. Flamary, N. Courty, A. Rolet, and M. Blondel. 2018. "Large-scale optimal transport and mapping estimation". In: *Proceedings of ICLR 2018*.

Shafieezadeh Abadeh, S., P. M. Mohajerin Esfahani, and D. Kuhn. 2015. "Distributionally Robust Logistic Regression". In: *Advances in Neural Information Processing Systems 28*. Ed. by C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett. Curran Associates, Inc. 1576–1584.

Shafieezadeh Abadeh, S., V. A. Nguyen, D. Kuhn, and P. M. Mohajerin Esfahani. 2018. "Wasserstein Distributionally Robust Kalman Filtering". In: *Advances in Neural Information Processing Systems 31*. Ed. by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett. Curran Associates, Inc. 8483–8492.

Shirdhonkar, S. and D. W. Jacobs. 2008. "Approximate earth mover's distance in linear time". In: *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE. 1–8.

Silverman, B. W. 1986. *Density Estimation for Statistics and Data Analysis.* Vol. 26. CRC press.

Sinkhorn, R. 1964. "A relationship between arbitrary positive matrices and doubly stochastic matrices". *Annals of Mathematical Statistics.* 35: 876–879.

Slomp, M., M. Mikamo, B. Raytchev, T. Tamaki, and K. Kaneda. 2011. "GPU-based softAssign for maximizing image utilization in photomosaics". *International Journal of Networking and Computing.* 1(2): 211–229.

Solomon, J., F. De Goes, G. Peyré, M. Cuturi, A. Butscher, A. Nguyen, T. Du, and L. Guibas. 2015. "Convolutional Wasserstein distances: efficient optimal transportation on geometric domains". *ACM Transactions on Graphics.* 34(4): 66:1–66:11.

Solomon, J., L. Guibas, and A. Butscher. 2013. "Dirichlet energy for analysis and synthesis of soft maps". In: *Computer Graphics Forum.* Vol. 32. No. 5. Wiley Online Library. 197–206.

Solomon, J., G. Peyré, V. G. Kim, and S. Sra. 2016a. "Entropic metric alignment for correspondence problems". *ACM Transactions on Graphics.* 35(4): 72:1–72:13.

Solomon, J., R. Rustamov, L. Guibas, and A. Butscher. 2014a. "Earth mover's distances on discrete surfaces". *Transaction on Graphics.* 33(4).

Solomon, J., R. Rustamov, L. Guibas, and A. Butscher. 2016b. "Continuous-flow graph transportation distances". *arXiv preprint arXiv:1603.06927.*

Solomon, J., R. Rustamov, G. Leonidas, and A. Butscher. 2014b. "Wasserstein propagation for semi-supervised learning". In: *Proceedings of the 31st International Conference on Machine Learning.* 306–314.

Sommerfeld, M. and A. Munk. 2018. "Inference for empirical Wasserstein distances on finite spaces". *Journal of the Royal Statistical Society: Series B (Statistical Methodology).* 80(1): 219–238.

Sriperumbudur, B. K., K. Fukumizu, A. Gretton, B. Schölkopf, and G. R. Lanckriet. 2009. "On integral probability metrics,φ-divergences and binary classification". *arXiv preprint arXiv:0901.2698.*

Sriperumbudur, B. K., K. Fukumizu, A. Gretton, B. Schölkopf, and G. R. Lanckriet. 2012. "On the empirical estimation of integral probability metrics". *Electronic Journal of Statistics*. 6: 1550–1599.

Srivastava, S., V. Cevher, Q. Dinh, and D. Dunson. 2015a. "WASP: scalable Bayes via barycenters of subset posteriors". In: *Artificial Intelligence and Statistics*. 912–920.

Srivastava, S., C. Li, and D. B. Dunson. 2015b. "Scalable Bayes via barycenter in Wasserstein space". *arXiv preprint arXiv:1508.05880*.

Staib, M., S. Claici, J. M. Solomon, and S. Jegelka. 2017a. "Parallel Streaming Wasserstein Barycenters". In: *Advances in Neural Information Processing Systems 30*. Ed. by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett. Curran Associates, Inc. 2647–2658.

Staib, M., S. Claici, J. Solomon, and S. Jegelka. 2017b. "Parallel streaming Wasserstein barycenters". *arXiv preprint arXiv:1705.07443*.

Stougie, L. 2002. "A polynomial bound on the diameter of the transportation polytope". *Tech. rep.* TU/e, Technische Universiteit Eindhoven, Department of Mathematics and Computing Science.

Sturm, K.-T. 2012. "The space of spaces: curvature bounds and gradient flows on the space of metric measure spaces". *Preprint* No. 1208.0434. arXiv.

Su, Z., Y. Wang, R. Shi, W. Zeng, J. Sun, F. Luo, and X. Gu. 2015. "Optimal mass transport for shape matching and comparison". *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 37(11): 2246–2259.

Sudakov, V. N. 1979. *Geometric Problems in the Theory of Infinite-dimensional Probability Distributions*. No. 141. American Mathematical Society.

Sugiyama, M., H. Nakahara, and K. Tsuda. 2017. "Tensor balancing on statistical manifold". *arXiv preprint arXiv:1702.08142*.

Sulman, M. M., J. Williams, and R. D. Russell. 2011. "An efficient approach for the numerical solution of the Monge–Ampère equation". *Applied Numerical Mathematics*. 61(3): 298–307.

Swoboda, P. and C. Schnörr. 2013. "Convex variational image restoration with histogram priors". *SIAM Journal on Imaging Sciences*. 6(3): 1719–1735.

Székely, G. J. and M. L. Rizzo. 2004. "Testing for equal distributions in high dimension". *InterStat.* 5(16.10).

Takatsu, A. 2011. "Wasserstein geometry of Gaussian measures". *Osaka Journal of Mathematics.* 48(4): 1005–1026.

Tan, X. and N. Touzi. 2013. "Optimal transportation under controlled stochastic dynamics". *Annals of Probability.* 41(5): 3201–3240.

Tarjan, R. E. 1997. "Dynamic trees as search trees via Euler tours, applied to the network simplex algorithm". *Mathematical Programming.* 78(2): 169–177.

Tartavel, G., G. Peyré, and Y. Gousseau. 2016. "Wasserstein loss for image synthesis and restoration". *SIAM Journal on Imaging Sciences.* 9(4): 1726–1755.

Thorpe, M., S. Park, S. Kolouri, G. K. Rohde, and D. Slepčev. 2017. "A transportation L^p distance for signal analysis". *Journal of Mathematical Imaging and Vision.* 59(2): 187–210.

Tolstoı, A. 1930. "Metody nakhozhdeniya naimen'shego summovogo kilome-trazha pri planirovanii perevozok v prostranstve (Russian; Methods of finding the minimal total kilometrage in cargo transportation planning in space)". *TransPress of the National Commissariat of Transportation*: 23–55.

Tolstoı, A. 1939. "Metody ustraneniya neratsional'nykh perevozok pri-planirovanii [Russian; Methods of removing irrational transportation in planning]". *Sotsialisticheskii Transport.* 9: 28–51.

Trouvé, A. and L. Younes. 2005. "Metamorphoses through Lie group action". *Foundations of Computational Mathematics.* 5(2): 173–198.

Trudinger, N. S. and X.-J. Wang. 2001. "On the Monge mass transfer problem". *Calculus of Variations and Partial Differential Equations.* 13(1): 19–31.

Vaillant, M. and J. Glaunès. 2005. "Surface matching via currents". In: *Information Processing in Medical Imaging.* Springer. 1–5.

Varadhan, S. R. S. 1967. "On the behavior of the fundamental solution of the heat equation with variable coefficients". *Communications on Pure and Applied Mathematics.* 20(2): 431–455.

Villani, C. 2003. *Topics in Optimal Transportation. Graduate Studies in Mathematics Series.* American Mathematical Society. ISBN: 9780821833124.

Villani, C. 2009. *Optimal Transport: Old and New*. Vol. 338. Springer Verlag.

Vogt, T. and J. Lellmann. 2018. "Measure-valued variational models with applications to diffusion-weighted imaging". *Journal of Mathematical Imaging and Vision*. 60(9): 1482–1502.

Wang, F. and L. J. Guibas. 2012. "Supervised earth mover's distance learning and its computer vision applications". *ECCV2012*: 442–455.

Wang, W., J. A. Ozolek, D. Slepcev, A. B. Lee, C. Chen, and G. K. Rohde. 2011. "An optimal transportation approach for nuclear structure-based pathology". *IEEE Transactions on Medical Imaging*. 30(3): 621–631.

Wang, W., D. Slepčev, S. Basu, J. A. Ozolek, and G. K. Rohde. 2013. "A linear optimal transportation framework for quantifying and visualizing variations in sets of images". *International Journal of Computer Vision*. 101(2): 254–269.

Weed, J. and F. Bach. 2017. "Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance". *arXiv preprint arXiv:1707.00087*.

Weinberger, K. Q. and L. K. Saul. 2009. "Distance metric learning for large margin nearest neighbor classification". *Journal of Machine Learning Research*. 10: 207–244.

Westdickenberg, M. and J. Wilkening. 2010. "Variational particle schemes for the porous medium equation and for the system of isentropic Euler equations". *ESAIM: Mathematical Modelling and Numerical Analysis*. 44(1): 133–166.

Wilson, A. G. 1969. "The use of entropy maximizing models, in the theory of trip distribution, mode split and route split". *Journal of Transport Economics and Policy*: 108–126.

Xia, G.-S., S. Ferradans, G. Peyré, and J.-F. Aujol. 2014. "Synthesizing and mixing stationary Gaussian texture models". *SIAM Journal on Imaging Sciences*. 7(1): 476–508.

Yule, G. U. 1912. "On the methods of measuring association between two attributes". *Journal of the Royal Statistical Society*. 75(6): 579–652.

Zemel, Y. and V. M. Panaretos. 2018. "Fréchet means and procrustes analysis in Wasserstein space". *to appear in Bernoulli*.

Zen, G., E. Ricci, and N. Sebe. 2014. "Simultaneous ground metric learning and matrix factorization with earth mover's distance". *Proceedings of ICPR'14*: 3690–3695.

Zhu, L., Y. Yang, S. Haker, and A. Tannenbaum. 2007. "An image morphing technique based on optimal mass preserving mapping". *IEEE Transactions on Image Processing.* 16(6): 1481–1495.

Zinsl, J. and D. Matthes. 2015. "Transport distances and geodesic convexity for systems of degenerate diffusion equations". *Calculus of Variations and Partial Differential Equations.* 54(4): 3397–3438.