§1 Monge and Kantorovich Problem[*]

Jiayao Zhang[†]

October 24, 2019

1 Some notions from probability

We will assume as given a **Polish space** (complete, separable metric space) (X, d) on which we may associate topology τ and Borel σ-algebra $\mathcal{F}$ to define the set of bounded continuous functions $\mathcal{C}_\mathrm{b}$ and the set of probability measures $\mathcal{P}(X) \coloneqq \mathcal{M}_+^1(X)$. The support $\operatorname{supp}(\mu)$ of a probability measure $\mu \in \mathcal{M}_+^1$ is the smallest closed set whose complement is a μ-null set.

Given two Polish spaces X, Y, $\mu \in \mathcal{P}(X)$ and a map $T : X \to Y$. The **push forward** of μ by T, $\nu = T_\# \mu \in \mathcal{P}(Y)$ is defined by

$$\begin{aligned} \nu(B) = \nu(T^{-1}(B)) &= \mu(x : T(x) \in B), && \forall B \in \mathcal{B}(Y) \\ \Leftrightarrow \quad \int f \, \mathrm{d}v &= \int (f \circ T) \, \mathrm{d}\mu, && \forall f \in \mathcal{C}_\mathrm{b}(Y). \end{aligned} \tag{1}$$

For any $\varphi \in \mathcal{C}_\mathrm{b}(Y)$, the **pull back** of φ by T, $T^\# \varphi$ is defined as $T \circ \varphi : X \to \mathbb{R}$. If we view integration between φ and μ as a bilinear map $\langle \cdot, \cdot \rangle : \mathcal{C}_\mathrm{b} \times \mathcal{P} \to \mathbb{R}$, $T_\#$ and $T^\#$ are adjoint in the sense that

$$\int_Y \varphi \, \mathrm{d}T_\# \mu = \langle \varphi, T_\# \mu \rangle = \left\langle T^\# \varphi, \mu \right\rangle = \int_X T^\# \varphi \, \mathrm{d}\mu. \tag{2}$$

If $X = Y = \mathbb{R}^d$, $\mu, \nu \in \mathcal{P}(X)$ are absolutely continuous wrt the Lebesgue measure (i.e., they admit densities ρ_1 and ρ_2), and T homeomorphic, by the change of variable formula, for any $\varphi \in \mathcal{C}_\mathrm{b}(Y)$,

$$\begin{aligned} \int_{\mathbb{R}^d} \varphi \, \mathrm{d}T_\# \mu = \int_{\mathbb{R}^d} \varphi(y) \rho_2(y) \, \mathrm{d}y &= \int_{\mathbb{R}^d} \varphi(T(x)) \rho_2(T(x)) |\det T'| \, \mathrm{d}x \\ &= \int_{\mathbb{R}^d} T^\# \varphi \rho_2(T(x)) |\det T'| \, \mathrm{d}x \equiv \int_{\mathbb{R}^d} T^\# \varphi \rho_1(x) \, \mathrm{d}x. \end{aligned} \tag{3}$$

This implies

$$\rho_1(x) = \rho_2(T(x)) \cdot |\det T'(x)|, \text{ a.e.,} \tag{4}$$

that is,

$$\nu = T_\# \mu \Rightarrow \frac{\mathrm{d}\nu}{\mathrm{d}x} = |\det T'| \frac{\mathrm{d}\mu}{\mathrm{d}x}, \tag{5}$$

which means push forward acts linearly on densities, and indeed the Radon-Nykodym derivative $\frac{\mathrm{d}\nu}{\mathrm{d}\mu} = |\det T'|$.

[*] This note is based on [PC+19] and [AG13], and was presented at Penn optimal transport study group.

[†] University of Pennsylvania, jiayaozhang@acm.org.

2 Monge's and Kantorovich's formulation

Definition 1. ***Monge's optimal transport problem.***

Let $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$ be a measurable map, $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$. Monge's optimal transport problem aims at

$$\begin{aligned} \min_{T} \quad & \mathbb{E}_{Z\sim\mu}[c(Z, T(Z)] = \int_X c(x, T(x))\,\mathrm{d}\mu, \\ \text{subject to} \quad & T_{\#}\mu = \nu, \end{aligned} \tag{6}$$

*that is, among all **transport map** T.*

When $|X|, |Y| < \infty$, μ, ν discrete, Monge's problem is equivalent to the minimum weight bipartite matching problem and can be solved by (integer) linear programming. When the solution exists, the minimizers include an integral solution, which can be argued the same way as we do for maximum matching, maximum set cover and other problems of similar flavour.

Monge's formulation may be ill-posed as it is possible that -

- No map satisfies the mass conservation constraint. E.g., when $|X| < |Y|$.
- From an optimization's perspective, given a sequence of $T^{(k)}$ such that $T^{(k)}_{\#}\mu = \nu$, $T^{(k)} \to T$ weakly for some T does not imply $T_{\#}\mu = \nu$. E.g., $\mu = x\mathbb{1}_{\{0\leq x\leq 1\}}$, $\nu = (\delta_{-1} + \delta_1)/2$, and $T^{(k)} = f(kx)$, where $f(x) = \mathbb{1}_{\{0\leq x<1/2\}} - \mathbb{1}_{\{1/2\leq x\leq 1\}}$.

Furthermore, even the optimal transport map (the Monge map) exists, it may be asymmetric. Kantorovich's relaxation circumvent these potential pathologies by allowing the *split of mass.*

As an example, if μ, ν admits densities, this is equivalent to say that we move the mass at x in X probabilistically to Y according to $p(y|x)$. The cost at x would be $\int_Y c(x, y)p(y|x)\,\mathrm{d}y$ and the total cost would be integrating again wrt $p(x)$, $\int_X p(x) \int_Y c(x, y)p(y|x)\,\mathrm{d}y\,\mathrm{d}x$. Let $\gamma \in \mathcal{P}(X \times Y)$ has density $p(x, y)$, this is equivalent to minimizing

$$\mathbb{E}_{(x,y)\sim\gamma} c(x, y) = \int_{X\times Y} c(x, y)p(x, y)\,\mathrm{d}x\,\mathrm{d}y. \tag{7}$$

The mass conservation constraint would be γ has $p(x)$ as marginal on X and $p(y)$ on Y. In the more general case where μ, ν do not necessarily admit densities, we have the following.

Definition 2. ***Kantorovich's relaxation.***

Let $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$ be a measurable map, $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$. Kantorovich's optimal transport problem aims at

$$\begin{aligned} \min_{\gamma} \quad & \mathbb{E}_{(x,y)\sim\gamma}[c(x, y)] = \int_{X\times Y} c(x, y)\,\mathrm{d}\gamma(x, y), \\ \text{subject to} \quad & \gamma \in \Pi(\mu, \nu), \end{aligned} \tag{8}$$

where

$$\Pi(\mu, \nu) \coloneqq \left\{\gamma \in \mathcal{P}(X \times Y) : \pi^X_{\#}\gamma = \mu, \pi^Y_{\#}\gamma = \nu\right\} \tag{9}$$

*is the set of all **couplings** between μ and ν, $\pi^{(\cdot)}$ is the natural projection.*

In other words, the coupling contains all probability measures on $X \times Y$ whose marginal on X (resp. Y) is μ (resp. ν).

The advantages of Kantorovich's relaxation include

- The constraint set Π is non empty (it contains $\mu \times \nu$).
- Π is convex and compact (wrt the narrow topology, for our purposes this means closed under weak convergence).
- The objective is linear.
- Contains all feasible points to the Monge's formulation (given T such that $T_{\#}\mu = \nu$, $(\mathrm{id} \times T)_{\#}\mu = \mu \times \nu \in \Pi$).

Furthermore, the minimizer to Kantorovich's problem *always* exists, as shown next.

Theorem 3. *If $c(\cdot,\cdot)$ is lower semi-continuous (l.s.b., $\forall x_0$, $\liminf_{x \to x_0} f(x) \geq f(x_0)$) and bounded from below, then the solution to Equation (8) always exists.*

Proof. This is a result from the **direct method** in the calculus of variations, which asserts that the minimizer of a bounded below, l.s.c. function on a compact set always exists[1]. To see this, let f be a function defined on a compact set Ω that is bounded from below and l.s.c. Then there exists a sequenece $x_k \in \Omega$ such that $\lim f(x_k) = \inf_{x \in \Omega} f(x) =: f_0$. Compactness implies x_k is bounded, hence there is a convergent subsequence $x_{k'} \to x_0 \in \Omega$. By lower semi-continuity, $f_0 \leq f(x_0) \leq \liminf f(x_k) = f_0$.

The proof of the compactness of Π replies on the Prokhov's theorem, and the l.s.c. of the objective from the l.s.c. assumption on $c(\cdot,\cdot)$.

To illustrate the idea, first consider a sequence of $\gamma_n \in \Pi(\mu,\nu)$ that converges weakly to γ. We must now show that $\gamma \in \Pi(\mu,\nu)$. Indeed for any $\varphi \in \mathcal{C}_{\mathrm{b}}(X)$,

$$\int_X \varphi \, \mathrm{d}\pi^X_{\#}\gamma = \int_{X \times Y} \varphi \, \mathrm{d}\gamma = \lim_{n \to \infty} \int_{X \times Y} \varphi \, \mathrm{d}\gamma_n = \lim_{n \to \infty} \int_X \varphi \, \mathrm{d}\pi^X_{\#}\gamma_n = \int_X \varphi \, \mathrm{d}\mu, \tag{10}$$

which implies $\pi^X_{\#}\gamma = \mu$. Similarly $\pi^Y_{\#}\gamma = \nu$, hence $\gamma \in \Pi(\mu,\nu)$.

Now we will show $\gamma \mapsto \int_{X \times Y} c \, \mathrm{d}\gamma$ is l.s.c. Since c is l.s.c., we can approximate it by an increasing sequence of bounded continuous functions c_m such that $c(x,y) = \sup_m c_m(x,y)$, $\forall (x,y) \in X \times Y$. Note that for any m, $c_m \in \mathcal{C}_{\mathrm{b}}(X \times Y)$, hence

$$\int c \, \mathrm{d}\gamma_n \geq \int c_m \, \mathrm{d}\gamma_n, \quad \forall m \quad \Rightarrow \quad \liminf_n \int c \, \mathrm{d}\gamma_n \geq \int c_m \, \mathrm{d}\gamma, \quad \forall m. \tag{11}$$

Taking supremum over m we have

$$\liminf_n \int c \, \mathrm{d}\gamma_n \geq \sup_m \int c_m \, \mathrm{d}\gamma = \int c \, \mathrm{d}\gamma, \tag{12}$$

by the monotone convergence theorem. □

[1] The assumption on the compactness can be replaced by requiring the function to be *coercive*, that is, every sequence with bounded function value is bounded.

Theorem 4. *If $c(\cdot,\cdot)$ is continuous and μ non-atomic (i.e., does not contain any Dirac mass), then*

$$\inf\{\textit{Monge}\} = \min\{\textit{Kantorovich}\}. \tag{13}$$

Theorem 5. ***Brenier's Theorem***

Let $X = Y = \mathbb{R}^d$, $\mu \in \mathcal{P}(\mathbb{R}^d)$ with finite second moment, then TFAE:

- *μ is regular;*
- *For every $\nu \in \mathcal{P}(\mathbb{R}^d)$ with finite second moment, there exists only one* ***transport plan*** *from μ to ν and the plan is induced by a map T.*

Either one would imply the Monge map can be recovered by taking gradient of some convex function.

A ***c-c hypersurface*** *(convex-minus-convex) in $\mathbb{R}^d$ is the set of the form*

$$\{(\boldsymbol{x}, y) : \boldsymbol{x} \in \mathbb{R}^{d-1}, y \in \mathbb{R}, y = f(\boldsymbol{x}) - g(\boldsymbol{x})\}, \tag{14}$$

for some convex function $f, g : \mathbb{R}^{d-1} \to \mathbb{R}$. A measure μ is ***regular*** *if $\mu(E) = 0$ for every c-c hypersurface E.*

In the textbook, Theorem 2.1 asserts that if c is the L_2 distance and μ admits a density, then the Monge map is given by $T = \nabla\Phi$ where the convex function $\Phi(x) = \|x\|^2/2 - \varphi(x)$ and φ is the **dual potential**.

3 The dual formulation

The Kantorovich's formulation involves minimizing a linear functional with affine constraints. In the case of discrete measurable spaces, this is a linear program and one can use the strong duality of the LP to consider the dual problem. As we will see, a similar result holds in the general case.

Definition 6. ***Dual formulation of Kantorovich's problem.***

The problem given in Equation (8) is ***equivalent*** *to*

$$\begin{aligned} \text{max} \quad & \int \varphi \, \mathrm{d}\mu + \int \psi \, \mathrm{d}\nu \\ \text{subject to} \quad & \varphi(x) + \psi(y) \le c(x,y), \quad \forall (x,y) \in X \times Y, \\ & \varphi \in L^1(\mu), \quad \psi \in L^1(\nu). \end{aligned} \tag{15}$$

Furthermore, we have

$$\inf_{\gamma \in \Pi(\mu,\nu)} \int c(x,y) \, \mathrm{d}\gamma(x,y) = \sup_{\varphi,\psi} \int \varphi \, \mathrm{d}\mu + \int \psi \, \mathrm{d}\nu. \tag{16}$$

Proof. This argument is based on the **min-max principle**, where we construct a penalty term and enlarge the feasible set.

Define for $\gamma \in \mathcal{M}_+(X \times Y)$, $\chi(\gamma)$ to be 0 if $\gamma \in \Pi(\mu,\nu)$ and $+\infty$ otherwise, then

$$\inf_{\gamma \in \Pi(\mu,\nu)} c \, \mathrm{d}\gamma = \inf_{\gamma \in \mathcal{M}_+} \int c \, \mathrm{d}\gamma + \chi(\gamma). \tag{17}$$

Now observe that the following is a valid candidate for χ:

$$\chi(\gamma) = \sup_{(\varphi,\psi)\in\mathcal{C}_\mathrm{b}(X)\times\mathcal{C}_\mathrm{b}(Y)} \left\{ \int_X \varphi \,\mathrm{d}\mu + \int_Y \psi \,\mathrm{d}\nu - \int_{X\times Y} (\varphi(x) + \psi(y)) \,\mathrm{d}\gamma(x,y) \right\}. \tag{18}$$

Hence

$$\inf_{\gamma\in\Pi(\mu,\nu)} \int c \,\mathrm{d}\gamma = \inf_{\gamma\in\mathcal{M}_+} \sup_{\varphi,\psi} F(\gamma,\varphi,\psi;\mu,\nu), \tag{19}$$

where we write

$$F(\gamma,\varphi,\psi;\mu,\nu) = \int_{X\times Y} c(x,y) \,\mathrm{d}\gamma(x,y) + \int_X \varphi \,\mathrm{d}\mu + \int_Y \psi \,\mathrm{d}\nu - \int_{X\times Y} (\varphi(x) + \psi(y)) \,\mathrm{d}\gamma(x,y). \tag{20}$$

Observe that F is linear in γ (hence convex) and linear in φ, ψ (hence concave). Thus we can swap the order of inf and sup to write

$$\begin{aligned} \inf_{\gamma\in\Pi(\mu,\nu)} \int c \,\mathrm{d}\gamma &= \sup_{\varphi,\psi} \inf_{\gamma\in\mathcal{M}_+} F \\ &= \sup_{\varphi,\psi} \left\{ \int \varphi \,\mathrm{d}\mu + \int \psi \,\mathrm{d}\nu + \inf_{\gamma\in\mathcal{M}_+} \int \left(c(x,y) - \varphi(x) - \psi(y)\right) \,\mathrm{d}\gamma(x,y) \right\}. \end{aligned} \tag{21}$$

Consider the quantity

$$\inf_{\gamma\in\mathcal{M}_+} \int \left(c(x,y) - \varphi(x) - \psi(y)\right) \,\mathrm{d}\gamma(x,y), \tag{22}$$

if $\varphi(x) + \psi(y) \leq c(x,y)$ for all x, y, setting γ to be the null measure would drive the infimum to 0; on the other hand if $\varphi(x) + \psi(t) > c(x,y)$ for some (x_0, y_0), setting $\gamma(n) = n \cdot \delta_{(x_0,y_0)}$ makes the infimum goes to $-\infty$. Hence

$$\inf_{\gamma\in\Pi(\mu,\nu)} \int c \,\mathrm{d}\gamma = \sup_{\varphi,\psi} \int \varphi \,\mathrm{d}\mu + \int \psi \,\mathrm{d}\nu, \tag{23}$$

for L_1 functions φ, ψ such that $\varphi(x) + \psi(y) \leq c(x,y)$ *everywhere.* □

The next theorem gives condition for strong duality.

Theorem 7. ***Duality.***

If c is continuous and bounded from below and for some $f \in L_1(\mu)$, $g \in L_1(\nu)$ we have for all $(x,y) \in X \times Y$,

$$c(x,y) \leq f(x) + g(y), \tag{24}$$

then the minimum of the primal problem is equal to the supremum of the dual problem. Furthermore, the supremum of the dual is attained.

4 Wasserstein W_p distance

Given (X, d) Polish, the **Wasserstein W_p distance**, defined on $\mathcal{P}(X)$ is given by

$$W_p^p(\mu,\nu) = \inf_{\gamma\in\Pi(\mu,\nu)} \int d^p(x,y)\,\mathrm{d}\gamma(x,y) = \sup_{\substack{\varphi,\psi\in\mathcal{C}_\mathrm{b}(X)\\ \varphi(x)+\psi(y)\le d^p(x,y)}} \int \varphi\,\mathrm{d}\mu + \int \psi\,\mathrm{d}\nu, \tag{25}$$

that is, the optimal to the Kantorovich's problem with d^p being the cost. We will next prove W_p is indeed a distance, which relies on the following lemmas.

Theorem 8. ***Disintegration.*** *Let X, Y be Polish, $\mu \in \mathcal{P}(X)$, $T : X \to Y$, and $\nu = T_\#\mu$. Then there exists a (ν-a.e.) unique family of measures $\{\mu_y \in \mathcal{P}(X)\}_{y\in Y}$ such that*

- *The map $\Phi(y) = \mu_y$ is Borel measurable. That is, for every $B \in \mathcal{B}(Y)$, the function $\varphi(y) = \Phi(B) = \mu_y(B) : Y \to \mathbb{R}$ is Borel measurable.*
- *μ_y concentrates on the **fiber** $T^{-1}(y)$, that is, for ν-a.e. y,*
$$\mu_y(X/T^{-1}(y)) = 0. \tag{26}$$
- *For any $f : X \to [0, \infty]$,*
$$\int_X f(x)\,\mathrm{d}\mu = \int_Y \int_{T^{-1}(y)} f(x)\,\mathrm{d}\mu_y(x)\,\mathrm{d}\nu. \tag{27}$$
In other words, we can "decompose" the measure μ on X as a product of the push forward measure and some unique measure on Y, in the sense that
$$\mathrm{d}\mu = \mathrm{d}\mu_y(x)\,\mathrm{d}\nu. \tag{28}$$

Theorem 9. ***Gluing-together lemma.*** *Let X, Y, Z be Polish spaces and $\gamma^{XY} \in \mathcal{P}(X \times Y)$, $\gamma^{YZ} \in \mathcal{P}(Y \times Z)$ be such that*

$$\pi^Y_\#\gamma^{XY} = \pi^Y_\#\gamma^{YZ}. \tag{29}$$

Then there exists $\gamma \in \mathcal{P}(X \times Y \times Z)$ such that

$$\pi^{X\times Y}_\#\gamma = \gamma^{XY}, \quad \pi^{Y\times Z}_\#\gamma = \gamma^{YZ}. \tag{30}$$

Proof. Let $\mu = \pi^Y_\#\gamma^{XY} = \pi^Y_\#\gamma^{YZ} \in \mathcal{P}(Y)$, by the third conclusion of the disintegration theorem, we have the following decomposition

$$\mathrm{d}\gamma^{XY}(x,y) = \mathrm{d}\gamma^{XY}_y(x)\,\mathrm{d}\mu(y), \quad \mathrm{d}\gamma^{YZ}(y,z) = \mathrm{d}\gamma^{YZ}_y(z)\,\mathrm{d}\mu(y). \tag{31}$$

We can define γ to be such that

$$\mathrm{d}\gamma(x,y,z) = \mathrm{d}\mu(y)\,\mathrm{d}(\gamma^{XY}_y \times \gamma^{YZ}_y)(x,z), \tag{32}$$

which completes the proof. □

Theorem 10. ***Wasserstein distance.*** *$W_p(\cdot,\cdot)$ defines a distance on $\mathcal{P}(X)$ for $p \ge 1$.*

Proof. Write $W = W_p$, consider:

- Clearly[2] $W(\mu,\mu) = 0$. Suppose $W(\mu,\nu) = 0$, let γ be an optimal transport plan, then

[2] Left as an exercise.

$W^p = \int d^p(x,y)\,\mathrm{d}\gamma(x,y) = 0$. Since d is a metric, $\gamma(x,y)$ concentrates on the diagonal of $X \times X$, i.e., $\gamma(\{(x,x) : x \in X\}) = 1$. Hence the projection to the two components π^1 and π^2 are equal γ-a.e., hence $\mu = \pi^1_\# \gamma = \pi^2_\# \gamma = \nu$.

- Clearly $W(\mu,\nu) = W(\nu,\mu)$.
- We now prove the triangle inequality. Let $\mu_1, \mu_2, \mu_3 \in \mathcal{P}(X)$, and $\gamma^{12}, \gamma^{23} \in \mathcal{P}(X \times X)$ be optimal transport plans from μ_1 to μ_2 and μ_2 to μ_3 resp. By the gluing-together lemma, there exists some $\gamma \in \mathcal{P}(X \times X \times X)$ such that

$$\pi^{12}_\# \gamma = \gamma^{12}, \quad \pi^{23}_\# \gamma = \gamma^{23}. \tag{33}$$

Then, writing $X^2 = X \times X$, $X^3 = X \times X \times X$, using Minkowski's inequality,

$$\begin{aligned} W(\mu_1,\mu_3) &\le \left(\int_{X^2} d^p(x_1,x_3)\,\mathrm{d}(\pi^{13}_\# \gamma)(x_1,x_3) \right)^{1/p} \\ &= \left(\int_{X^3} d^p(x_1,x_3)\,\mathrm{d}\gamma(x_1,x_2,x_3) \right)^{1/p} \\ &\le \left(\int_{X^3} d^p(x_1,x_2)\,\mathrm{d}\gamma(x_1,x_2,x_3) \right)^{1/p} + \left(\int_{X^3} d^p(x_2,x_3)\,\mathrm{d}\gamma(x_1,x_2,x_3) \right)^{1/p} \\ &= \left(\int_{X^2} d^p(x_1,x_2)\,\mathrm{d}\gamma^{12}(x_1,x_2) \right)^{1/p} + \left(\int_{X^2} d^p(x_2,x_3)\,\mathrm{d}\gamma^{23}(x_2,x_3) \right)^{1/p} \\ &= W(\mu_1,\mu_2) + W(\mu_2,\mu_3). \end{aligned} \tag{34}$$

□

5 Examples

Example 11. ***Kronecker cost.***

Let $X = Y = \Omega$ and $c(x,y) = \mathbb{1}_{\{x \neq y\}}$, i.e., 1 if $x \neq y$ and 0 otherwise. Then the Kantorovich's problem

$$\inf_{\gamma \in \Pi(\mu,\nu)} \inf \mathbb{1}_{\{x \neq y\}}\,\mathrm{d}\gamma = \inf_{\gamma \in \Pi(\mu,\nu)} \mathbb{E}_{x,y \sim \gamma} \mathbb{1}_{\{x \neq y\}} = \sup_{E \in \mathcal{B}(\Omega)} |\mu(E) - \nu(E)| \equiv \|\mu - \nu\|_{TV}. \tag{35}$$

Example 12. ***1-D case.*** *Let $X = Y = \mathbb{R}$ and F, G by the CDF of $\mu, \nu \in \mathcal{P}(\mathbb{R})$, i.e.,*

$$F(x) = \int_{-\infty}^{x} \mathrm{d}\mu, \quad G(x) = \int_{-\infty}^{x} \mathrm{d}\nu. \tag{36}$$

Then

$$W_p^p(\mu,\nu) = \int_{\mathbb{R}} |F^{-1}(x) - G^{-1}(x)|^p\,\mathrm{d}x. \tag{37}$$

References

[AG13] Luigi Ambrosio and Nicola Gigli. A user's guide to optimal transport. In *Modelling and optimisation of flows on networks*, pages 1–155. Springer, 2013.

[PC$^+$19] Gabriel Peyré, Marco Cuturi, et al. Computational optimal transport. *Foundations and Trends in Machine Learning*, 11(5-6):355–607, 2019.