OPTIMAL TRANSPORTATION

ATHARV SAMPATH

1. The Monge Transportation Problem

The Monge Transportation Problem is the main problem regarding Optimal Transport. It can be stated as follows: How do you best move given piles of sand to fill up given holes of the same total volume?

The mathematical formulation is as shown below.

Pile of Sand: a positive Radon measure μ^+ on a convex subset $X \subset \mathbb{R}^m$.

Hole: another positive Radon measure μ^- on X.

Same Volume: $0 < \mu^+(X) = \mu^-(X) < +\infty$. Usually, we normalize the mass to 1. So, both μ^+ and μ^- are probability measures.

Move: a Borel, one-to-one map $\psi : X \to X$

Fill: $\psi_\# \mu^+ = \mu^-$ (i.e. $\mu^-(A) = \psi_\# \mu^+(A) = \mu^+(\psi^{-1}(A))$.).

Best: minimum total "work" or transport cost.

Work or cost of $\psi : I(\psi) = \displaystyle\int_X |x - \psi(x)| d\mu^+(x)$

Therefore, we can describe Monge's problem as minimizing the cost given two probability measures on X :

$$I[\psi] := \int_X |x - \psi(x)| d\mu^+(x)$$

among all "transport maps" in $\mathcal{A} = \{\psi : X \to X \text{ Borel, one-to-one, } \psi_\# (\mu^+) = \mu^-\}$. In essence, $I[\psi]$ finds the sum of the distances (cost) of the path of a distinct element between one Radon measure to the other. In the more general case, considering other cost functions, we must minimize

$$I[\psi] := \int_X c(x, \psi(x)) d\mu^+(x)$$

for some given cost function $c : X \times X \to [0, +\infty)$ Some important cases are the linear cost, $c(x, y) = |x - y|$, the quadratic cost, $c(x, y) = |x - y|^2$, $c(x, y) = |x - y|^p$, for $0 < p < \infty$ and $c(x, y) = h(|x - y|)$ for some convex function h.

Date: November 2020.

2. Discrete Cases

Both

$$\mu^+ = \sum_i a_i \delta_{x_i} \text{ and } \mu^- = \sum_j b_j \delta_{y_j}$$

are atomic measures of equal total mass. We assume $\{x_i, y_j\}$ are distinct. For example,

$$\mu^+ = \frac{1}{3}\delta_{x_1} + \frac{1}{4}\delta_{x_2} + \frac{5}{12}\delta_{x_3}$$
$$\mu^- = \frac{1}{3}\delta_{y_1} + \frac{1}{4}\delta_{y_2} + \frac{5}{12}\delta_{y_3}$$

We find that there is only one admissible transport map here. Each x_i is mapped to y_i. It is easy to see that this is the only trivial case whenever $a_i \neq b_j$. So, for the nontrivial case, we may assume $a_i = b_j = \frac{1}{n}$. That is,

$$\mu^+ = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i} \text{ and } \mu^- = \frac{1}{n}\sum_{j=1}^{n} \delta_{y_j}$$

Each transport map is just a permutation σ of $\{1, \cdots, n\}$. In this case, Monge's problem becomes minimizing

$$\frac{1}{n}\sum_i c\left(x_i, y_{\sigma(i)}\right)$$

among all permutations $\sigma \in \mathcal{S}_n$. It corresponds to finding an optimal matching between the source points x_i and the target points y_j.

3. Continuous Cases

Assume both μ^+ and μ^- are absolutely continuous with respect to Lebesgue measure:

$$\mu^+ = f(x)dx \text{ and } \mu^- = g(x)dx$$

for some integrable density functions f and g on X. In this case, an one-to-one smooth map $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ satisfies the constraint $\varphi_{\#}\mu^+ = \mu^-$ if and only if

$$f(x) = g(\varphi(x))|\det(D\varphi(x))|$$

If f and g have finite second moments, there exists an optimal transport map $\varphi(x) = \nabla\psi(x)$ for some convex function $\psi(x)$ with respect to the quadratic cost $c(x, y) = |x - y|^2$. Thus, $\psi(x)$ solves a particular form of Monge-Ampe're equation

$$\det\left(D^2\psi(x)\right) = \frac{f(x)}{g(\nabla\psi(x))}$$

4. Difficulty Solving Monge's Problem

One difficulty is the highly nonlinear structure of

$$I[\psi] = \int_X c(x, \psi(x))d\mu^+(x)$$

Let $c(x, y) = |x - y|$. For example, let $X = [-1, 1], \mu^+ = \delta_0, \mu^- = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$. Then, there exists no transport map. So, no splitting of mass may cause non-existence of transport maps.

5. Optimality Test for $c(x,y) = |x-y|$

Suppose φ is a transport map from μ^+ to μ^-. How to know if φ is optimal? Let $u : X \to \mathbb{R}$ be a Lipschitz function with $\mathrm{Lip}(u) \leq 1$. (We denote $u \in \mathrm{Lip}_1(X)$) $(u(x) - u(y) \leq |x-y|.)$ Then

$$\int_X u(x) d\left(\mu^- - \mu^+\right) = \int_X [u(\varphi(x)) - u(x)] d\mu^+ \leq \int_X |\varphi(x) - x| d\mu^+$$

Thus, if the equality is achieved for some $u \in \mathrm{Lip}_1$, then φ must be an optimal transport map. One way to see this is the book shifting problem, which is the Monge Problem in $\mathbb{R}$.

$$X = \mathbb{R}, \mu^+ = \chi_{[0,n]} dx, \mu^- = \chi_{[1,n+1]} dx$$

Both $\varphi(x) = x + 1$ and $\phi(x) = \begin{cases} x + n \text{ on } [0,1) \\ x \quad \text{ on } [1,n] \end{cases}$ are optimal transport maps. In fact, if we take $u(x) = x$, then $u \in \mathrm{Lip}_1$ and

$$\int_X u(x) d\left(\mu^- - \mu^+\right) = \int_1^{n+1} x dx - \int_0^n x dx = n$$

On the other hand,

$$\int_X |\varphi(x) - x| d\mu^+ = \int_X |\psi(x) - x| d\mu^+ = n$$

The book shifting problem intuitively asks the question: if we have a row of books (μ^+) that spans from 0 to n, on a bookshelf (the real line), what is the easiest way to move the row such that it spans from 1 to $n+1$? So, the above solutions show that we actually have two minimum cost solutions, we can move the books from $[0,1]$ to $[n, n+1]$, which has cost n. Or, we can move each book by one, which also has cost n. Thus, this example shows that minimizers are not necessarily unique.

We in fact have another issue, that the limit of the minimizing sequence of maps ψ_i may actually fail to be a map itself.

6. Kantarovich Relaxation

The key idea that Kantarovich formulated was the idea of associating each map with a measure. This takes away the problem with the limit, as the limits of maps may fail to be maps, but the limit of measures will always be measures. So, we can view maps as positive measures in the product space. We can associate with each transport map ψ the measure

$$\gamma_\psi = (Id \times \psi)_\# \mu^+$$

in the product space $X \times X$. Let π_i be the projection map from $X \times X$ to its i^{th} coordinates. Then $\pi_{1\#}\left(\gamma_\psi\right) = \mu^+$ and $\pi_{2\#}\left(\gamma_\psi\right) = \psi_\# \mu^+ = \mu^-$. Moreover,

$$\int_{X \times X} c(x,y) d\gamma_\psi = \int_X c(x, \psi(x)) d\mu^+$$

Thus, our goal is now to minimize

$$J(\gamma) := \int_{X \times X} c(x,y) d\gamma(x,y)$$

in the class of transport plans

$$\Pi\left(\mu^+, \mu^-\right) = \left\{\gamma \in P(X \times X) \mid \pi_{1\#}\gamma = \mu^+, \pi_{2\#}\gamma = \mu^-\right\}$$

Existence: from a simple compactness argument of probability measures (for simplicity, assume X is compact here.)

7. The Discrete Case

Let

$$\mu^+ = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i} \text{ and } \mu^- = \frac{1}{n}\sum_{j=1}^{n} \delta_{y_j}.$$

Then, any transport plan from μ^+ to μ^- can be represented as a bistochastic $n \times n$ matrix $\pi = (\pi_{ij})$. Here by bistochasticity we mean that all the π_{ij} are nonnegative and that

$$\forall j, \sum_i \pi_{ij} = 1; \text{ and } \forall i, \sum_j \pi_{ij} = 1.$$

So, in this case, the Kantorovich problem reduces to minimize

$$\frac{1}{n}\sum_{i,j} \pi_{ij} c\left(x_i, y_j\right)$$

among all bistochastic $n \times n$ matrices π. This is a linear minimizing problem on a bounded convex set. By Choquet's theorem and Birkhoff's theorem, its solutions are given by permutation matrices π, (i.e. $\pi_{ij} = \delta_{j,\sigma(i)}$ for some permutation σ.) Thus, in this case, optimal transport plans in Kantorovich's problem coincide with solutions of Monge's problem

$$\inf\left\{\frac{1}{n}\sum_i c\left(x_i, y_{\sigma(i)}\right); \sigma \in \mathcal{S}_n\right\}.$$

8. Relations between Monge and Kantarovich

Since every transport map determines a transport plan of the same cost,

$$\min_{\gamma\in\Pi(\mu^+,\mu^-)} \int_{X\times X} c(x,y)d\gamma(x,y) \leq \inf_{\psi_{\#}\mu^+=\mu^-} \int_X |x-\psi(x)|d\mu^+(x)$$

In general, when will solutions to the Kantorovich and Monge problems coincide? Will an optimal transport plan come from an optimal transport map? Let $X = \mathbb{R}^n, c(x,y) = |x-y|^p, 0 < p < +\infty$, and μ^+, μ^- are compactly supported. Then, if $p > 1$, the strictly convexity of $c(x,y) = |x-y|^p$ guarantees that, if μ^+, μ^- are absolutely continuous with respect to Lebesgue measure, then there is a unique solution to the Kantorovich problem, which turns out to be also the solution to the Monge problem. The same result holds true when μ^+ is nonatomic (i.e. contains no atomics: $\mu^+(\{x\}) = 0$ for all $x \in X$.) In the case of a quadratic case, $p = 2$, these optimal maps are the (restrictions of) gradients of convex functions $\mathbb{R}^n$. Sometimes, optimal transport plans may have to split mass and solution to the Monge problem may fail to exist. If $p = 1, \mu^+, \mu^-$ are absolutely continuous with respect to Lebesgue measure, then there are solutions of the Monge problem which are also solutions of the Kantorovich problem. However,uniqueness may fail here. If $0 < p < 1$, there is in general no solution of the Monge problem, except if μ^+ and μ^- are concentrated on disjoint sets.

9. One Dimensional Case

Let $X=[a,b]$ be a closed interval in $\mathbb{R}$. Let $c(x,y)=|x-y|^p$ for some $p\geq 1$ If μ^+ is nonatomic (i.e. $\mu^+(\{x\})=0$, for all x), then there exists a unique (modulo countable sets) nondecreasing function $\psi:\operatorname{spt}(\mu^+)\to X$ such that $\psi_{\#}\mu^+=\mu^-$ This function ψ is given by

$$\psi(s):=\sup\left\{t\in[a,b]:\mu^-([a,t])\leq\mu^+([a,s])\right\}$$

The function ψ is an optimal transport map. When $p>1$, it is the unique optimal transport.

This is a sharp result in the one-dimensional case. μ^+ has atoms may cause non-existence of maps. Dropping monotonicity may cause non-uniqueness.

10. Finite moments (in the case X is unbounded)

For any $p>0$, let $P_p(X)$ be the set of all probability measures with finite moments of order p. i.e. those measures μ such that for some (and thus any) $x_0\in X$

$$\int_X|x-x_0|^p\,d\mu(x)<+\infty$$

Of course, if X is bounded, then $P_p(X)$ coincides with the set $P(X)$ of all probability measures on X. We now assume both μ^+ and μ^- are in $P_p(X)$. This condition ensures that the cost of any transport plan $\gamma\in\Pi\left(\mu^+,\mu^-\right)$ is always finite. Indeed,

$$\int_{X\times X}|x-y|^p d\gamma(x,y)\leq\int_{X\times X}2^p\left(|x|^p+|y|^p\right)d\gamma(x,y)<+\infty$$

since $\Pi\left(\mu^+,\mu^-\right)$ is compact with respect to weak $*$ convergence of probability measures, the Kantorovich minimization problem $\inf\left\{J(\gamma):\gamma\in\Pi\left(\mu^+,\mu^-\right\}\right.$ admits a minimizer.

11. Wasserstein Distances on $P(X)$

Definition. Given $p\in(0,+\infty)$ (usually $[1,+\infty)$, for any two probability measures $\mu^+,\mu^-\in P_p(X)$, define

$$W_p\left(\mu^+,\mu^-\right):=\left[\min_{\gamma\in\Pi(\mu^+,\mu^-)}\int_{X\times X}|x-y|^p d\gamma(x,y)\right]^{\min(1,1/p)}$$

distance between measures = minimal cost to suitable powers Proposition. W_p is a distance on $P_p(X)$ and metrizes the weak $*$ topology of $P_p(X)$.

12. Kantorovich Duality

Kantorovich problem is a linear minimization problem with convex constraints, so it also admits a dual formulation:

$$\begin{aligned}&\inf_{\gamma\in\Pi(\mu^+,\mu^-)}\int_{X\times X}c(x,y)d\gamma(x,y)\\ =&\sup_{\Phi_c}\left[\int_X\varphi(x)d\mu^++\int_X\psi d\mu^-\right]\end{aligned}$$

Here,

$$\Phi_c=\left\{(\varphi,\psi)\in L^1\left(d\mu^+\right)\times L^1\left(d\mu^-\right)\text{ with }\varphi(x)+\psi(y)\leq c(x,y)\right\}.$$

Here is one way to think about it more intuitively, in the form of the Shipper's Problem, quoted from Villani's book: Suppose for instance that you are an industrial willing to transfer a huge amount of coal from your mines to your factories. You can hire trucks to do this transportation problem, but you have to pay them c(x; y) for each ton of coal which is transported from place x to place y. Both the amount of coal which you can extract from each mine, and the amount which each factory will receive, are fixed. As you are trying to solve the associated Monge-Kantorovich problem in order to minimize the price you have to pay, another mathematician comes to you and tells you "My friend, let me handle this for you: I will ship all your coal with my own trucks and you won't have to care of what goes where. I will just set a price $\varphi(x)$ for loading one ton of coal at place x, and a price $\psi(y)$ for unloading it at destination y. I will set the prices in such a way that your financial interest will be to let me handle all your transportation ! Indeed, you can check very easily that for all x and all y, the sum $\varphi(x) + \psi(y)$ will always be less that the cost $c(x; y)$ (in order to achieve this goal, I am even ready to give financial compensations for some places, in the form of negative prices !)". Of course you accept the deal. Now, what Kantorovich's duality tells you is that if this shipper is clever enough, then he can arrange the prices in such a way that you will pay him (almost) as much as you would have been ready to spend by the other method.

13. Informal Solution to Kantorovich Duality

The main idea of this proof uses the minimax principle. Let $M_+(X \times X)$ be the space of all nonnegative Borel measures on $X \times X$. Then we have

$$\inf_{\gamma \in \Pi(\mu^+,\mu^-)} \int_{X\times X} c(x,y)d\gamma(x,y)$$

$$= \inf_{\gamma \in M_+(X\times X)} \left[\int_{X\times X} c(x,y)d\gamma(x,y) + \left\{ \begin{array}{ll} 0, & \text{if } \gamma \in \Pi(\mu^+,\mu^-) \\ +\infty & \end{array} \right. \right]$$

$$= \inf_{\gamma \in M_+(X\times X)} \left\{ \int_{X\times X} c(x,y)d\gamma(x,y) + \sup_{(\varphi,\psi)} \left[\int \varphi d\mu + \int \psi d\mu - \int [\varphi(x)+\psi(y)]d\gamma(x,y) \right] \right\}$$

$$= \inf_{\gamma \in M_+(X\times X)} \sup_{(\varphi,\psi)} \left[\int_{X\times X} c(x,y)d\gamma(x,y) + \int \varphi d\mu^+ + \int \psi d\mu - \int [\varphi(x)+\psi(y)]d\gamma(x,y) \right]$$

Invoking the minimax principle, we have,

$$\sup_{(\varphi,\psi)} \inf_{\gamma \in M_+(X\times X)} \left\{ \int_{X\times X} c(x,y)d\gamma(x,y) + \int \varphi d\mu^+ + \int \psi d\mu - \int [\varphi(x)+\psi(y)]d\gamma(x,y) \right\}$$

$$= \sup_{(\varphi,\psi)} \left\{ \int \varphi d\mu^+ + \int \psi d\mu + \inf_{\gamma \in M_+(X\times X)} \int [c(x,y) - \varphi(x) - \psi(y)]d\gamma \right\}$$

$$= \sup_{(\varphi,\psi)} \left[\int \varphi d\mu^+ + \int \psi d\mu^- + \left\{ \begin{array}{ll} 0, & \text{if } (\varphi,\psi) \in \Phi_c \\ -\infty & \text{else} \end{array} \right. \right]$$

$$= \sup_{\Phi_c} \left[\int_X \varphi(x)d\mu^+ + \int_X \psi d\mu^- \right].$$

14. Sources

Villani C., Topics in Optimal Transportation; Graduate Studies in mathematics, Volume 58.

Ambrosio L., Lecture notes on optimal transport problems.

Thorpe M., Introduction to Optimal Transport.

Mikami T., A Simple Proof od Duality Theorem For Monge-Kantorovich Problem.