OPTIMAL TRANSPORT PROBLEMS

JIAFENG CHEN AND FRANCISCO RIVERA

1. Revisions

By and large, our project goal has remained the same. We have, however, focused our scope to focus primarily on the Kantorovich optimal transport problem. This is contrast to the more general but intractable continuous Wassserstein problem.

2. Theoretical foundations

The following is a concise review of selected topics in Chapters 2 and 3 in [**?**].

Let $\mathbf{a} \in \mathbb{R}^m_{\geq 0}$ and $\mathbf{b} \in \mathbb{R}^n_{\geq 0}$ be two vectors of nonnegative real numbers such that $\mathbf{a}^T \mathbb{1} = \mathbf{b}^T \mathbb{1} = 1$, where a_i and b_j represents the amount of "mass" at different positions. We are interested in a matrix $\mathbf{P} \in \mathbb{R}^{m\times n}_{\geq 0}$ where P_{ij} prescribes the amount of mass flowing from a_i to b_j. Naturally, we must have $\sum_j P_{ij} = a_i$ and $\sum_i P_{ij} = b_j$, so that mass is preserved. Let $\mathbf{C} \in \mathbb{R}^{m\times n}$ be a *cost matrix*, where C_{ij} represents the cost of transporting one unit of a_i to b_j. The *Kantorovich optimal transport problem* is the following linear program:

$$\begin{aligned} &\min_{\mathbf{P}\in\mathbb{R}^{m\times n}_{\geq 0}} \sum_{i,j} P_{ij} C_{ij} \\ \text{subjected to } &\mathbf{P}\mathbb{1} = \mathbf{a} \\ &\mathbf{P}^T \mathbb{1} = \mathbf{b} \end{aligned}$$

Remark 1 (A probabilistic view). Let X, Y be discrete random variables with support $\mathbf{x} \in \mathcal{X}^m$ and $\mathbf{y} \in \mathcal{Y}^n$, such that their marginal distributions are prescribed by $P(X = x_i) = a_i$ and $P(Y = y_j) = b_j$. Let $P_{ij} = P(X = x_i, Y = y_j)$, making $\mathbf{P} = (P_{ij})_{ij}$ a joint distribution. Let $C : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be a cost function. Then the Kantorovich problem is equivalent to the following:

$$\min_{\mathbf{P}} \mathbb{E}_{(X,Y)\sim\mathbf{P}} \left[C(X, Y)\right],$$

where the expectation is taken over valid joint distributions of (X, Y) respecting the marginal distributions. Note that when $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$ and $C(x, y) = \|x - y\|^p$,

Date: March 27, 2018.

then the optimal value of the above problem for X, Y is the p-Wasserstein distance between the marginal distributions μ_X, μ_Y (up to a power of $1/p$), for discrete, d-dimensional distributions with finite support.

It can be shown that the dual problem of the above LP takes the following form:

$$\max_{(\mathbf{f},\mathbf{g})\in\mathbb{R}^m\times\mathbb{R}^n} \sum_i a_i f_i + \sum_j b_j g_j$$
$$\text{subjected to } f_i + g_j \leq C_{ij} \text{ for all } i, j.$$

The dual problem has a neat economic intuition. Suppose Kevin needs to transport $\mathbf{a}$ to $\mathbf{b}$ but does not understand how. Francisco, his profit-maximizing colleague, offers him a deal, where Kevin will pay f_i for Francisco to pick up a unit of mass at a_i and pay g_j for Francisco to dropoff a unit of mass at b_j. At minimum, Kevin knows that if any $f_i + g_j > C_{ij}$, then Francisco is ripping him off, as transporting one unit from a_i to b_j costs exactly C_{ij}. If, on the other hand, $f_i + g_j \leq C_{ij}$ is satisfied, then given any $\mathbf{P}$, we have

$$\sum_{i,j} C_{ij}P_{ij} \geq \sum_{i,j} f_i P_{ij} + \sum_{i,j} g_j P_{ij} = \sum_i f_i a_i + \sum_j g_j b_j, \tag{Weak duality}$$

which means that Kevin cannot lose by taking Francisco's deal. Strong duality, which indeed holds, here imply that Francisco's optimal profit is zero, and the trade is fair.

Given optimal solutions to the primal and dual, $\mathbf{P}$, $(\mathbf{f}, \mathbf{g})$, respectively, *complementary slackness* implies that

$$P_{ij}(C_{ij} - f_i - g_j) = 0 \tag{Complementary slackness}$$

holds, which means that $\mathbf{P}_{ij} \neq 0 \implies C_{ij} = f_i - g_j$ and vice versa.

2.1. **Dual-ascent algorithms and Hungarian algorithm.** Let $S \subset \{1, \ldots, m\}$ and $S' \subset \{1, \ldots, n\}$. Let $\mathbb{1}_S$ be a vector of zeros with ones at location prescribed by S. Given $\mathbf{f}, \mathbf{g}$ a feasible pair for the dual, call (i, j) a balanced pair if the condition $f_i + g_j = C_{ij}$ binds, and inactive otherwise. Then

Proposition 1. Let $(\tilde{\mathbf{f}}, \tilde{\mathbf{g}}) = (\mathbf{f}, \mathbf{g}) + \epsilon(\mathbb{1}_S, -\mathbb{1}_{S'})$. Then $(\tilde{\mathbf{f}}, \tilde{\mathbf{g}})$ is dual feasible for a sufficiently small ϵ if for all i,

$$(i, j) \text{ is balanced } \implies j \in S'.$$

Proof. For all $i \in S$, consider

$$\epsilon_i = \min_{j:(i,j) \text{ inactive}} C_{ij} - f_i - g_j$$

Let $\epsilon = \min_i \epsilon_i$. If (i,j) is inactive, then

$$\tilde{f}_i + \tilde{g}_j \le f_i + \epsilon + g_j \le C_{ij},$$

which continues to be inactive. If (i,j) is balanced, then $\tilde{f}_i + \tilde{g}_j = f_i + g_j = C_{ij}$. In either case, $\tilde{\mathbf{f}}, \tilde{\mathbf{g}}$ continues to be feasible. □

Proposition 2. If a dual feasible solution $(\mathbf{f}, \mathbf{g})$ is not optimal, then there exists S, S' such that $(\tilde{\mathbf{f}}, \tilde{\mathbf{g}}) = (\mathbf{f}, \mathbf{g}) + \epsilon(\mathbb{1}_S, -\mathbb{1}_{S'})$ is feasible for small enough ϵ with a strictly better objective.

Proof. Let $\mathcal{B}$ be the set of balanced edges (i,j), and consider the bipartite graph on $\{1, \dots, m\} \sqcup \{1, \dots, n\}$ with edges in $\mathcal{B}$. Let there be a source node s with capacitated edges to $\{1, \dots, m\}$ with capacity a_i, and a sink node t analogously with capacities b_j. Consider the maximum flow $\mathbf{F}$ on the network via Ford-Fulkerson. If the throughput of the flow is one, then let $P_{ij} = F_{ij}$, which generates a primal feasible solution $\mathbf{P}$. Strong duality would imply that $\mathbf{P}, (\mathbf{f}, \mathbf{g})$ are both optimal.

Thus we may assume that the throughput is less than one. We label the nodes reached from s where $\mathbf{F}$ does not saturate capacity, as well as the nodes that contribute flow to any labeled nodes. Store the labeled nodes in Q. Since the graph is bipartite, Q can be split into S, S'. Note that if (i,j) is balanced and $i \in S$, then $j \in S'$. By Proposition 1, $(\tilde{\mathbf{f}}, \tilde{\mathbf{g}})$ so defined is feasible. We wish to show that

$$\mathbb{1}_S^T \mathbf{a} - \mathbb{1}_{S'}^T \mathbf{b} > 0.$$

We readily conclude that this is indeed the case by accounting for the flow on the network. □

Proposition 2 provides a template for the celebrated Hungarian algorithm [**?**]. The Hungarian algorithm solves the Kantorovich problem in the case when both $\mathbf{a}, \mathbf{b} \propto \mathbb{1}$. In a probability context for the Wasserstein distance between empirical measures, if the measures μ, ν are both d-dimensional with n observations, the Hungarian algorithm is $O(n^3)$ for $d > 1$ and $O(n \log n)$ for[1] $d = 1$ [**?**].

[1]In the $d = 1$ case, the optimal transport problem reduces to sorting.

3. Completed Steps

4. Next Steps

5. Application value

[?] is an example of recent work applying the Wasserstein distance to statistics and machine learning.