Large-scale Optimal Transport

Weijie Chen[*]
School of Physics
Peking University
1500011335
1500011335@pku.edu.cn

Dinghuai Zhang
School of Mathematics
Peking University
1600013525
1600013525@pku.edu.cn

Abstract

1 Introduction to Optimal Transport

2 Problem Statement

The standard formulation of optimal transport are derived from couplings. [**Villani2009**] That is, let $(\mathcal{X}, \mu)$ and $(\mathcal{Y}, \nu)$ be two probability spaces, and a probability distribution π on $\mathcal{X} \times \mathcal{Y}$ is called *coupling* if $proj_{\mathcal{X}}(\pi) = \mu$ and $proj_{\mathcal{Y}}(\pi) = \nu$. An optimal transport between $(\mathcal{X}, \mu)$ and $(\mathcal{Y}, \nu)$, or an optimal coupling, is a coupling minimize

$$\int_{\mathcal{X}\times\mathcal{Y}} c(x, y) d\pi(x, y) \tag{1}$$

Optimal transport problems can be categorized according to the discreteness of μ and ν. In this report, we only consider discrete optimal tranport problems, where the two distributions are distributions of finite weighted points.

A discrete optimal transport problem can be formulated into a linear program as

$$\begin{aligned}
&\min_{\pi} \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} \pi_{ij} \\
&s.t. \sum_{j=1}^{n} \pi_{ij} = \mu_i, \forall i \\
&\quad\;\; \sum_{i=1}^{n} \pi_{ij} = \nu_j, \forall j \\
&\quad\;\; \pi_{ij} \geq 0,
\end{aligned} \tag{2}$$

where c stands for the cost and s for the transportation plan, while μ and ν are restrictions. Note that we always suppose $c \geq 0$, $\mu \geq 0$, $\nu \geq 0$ and $\sum_{i=1}^{m} \mu_i = \sum_{j=1}^{n} \nu_j = 1$ implicitly. From realistic background, c is always valued the squared Euclidean distanced or some other norms. Note that there are mn variables in this formulation, and this leads to intensive computation.

[*]Pre-admission 2019 PKU AAIS

3 Algorithms

3.1 ADMM for Primal Problem

3.2 ADMM for Dual Problem

3.3 Sinkhorn Method with Entropy Regularization

The discrete entropy of a coupling matrix is defined as

$$\mathbf{H}(\mathbf{P}) \stackrel{\text{def}}{=} -\sum_{i,j} \mathbf{P}_{i,j} \left(\log\left(\mathbf{P}_{i,j}\right) - 1\right) \tag{3}$$

The function $\mathbf{H}$ is strongly concave.

The idea of the entropic regularization of optimal transport is to use $-\mathbf{H}$ as a regularizing function to obtain approximate solutions to the original transport problem:

$$\mathrm{L}_{\mathrm{C}}^{\varepsilon}(\mathbf{a}, \mathbf{b}) \stackrel{\text{def}}{=} \min_{\mathbf{P} \in \mathbf{U}(\mathbf{a}, \mathbf{b})} \langle \mathbf{P}, \mathbf{C} \rangle - \varepsilon \mathbf{H}(\mathbf{P}) \tag{4}$$

One can show that the solution to **??** has the form of

$$\mathbf{P}_{i,j} = \mathbf{u}_i \mathbf{K}_{i,j} \mathbf{v}_j \tag{5}$$

where $\mathbf{K}_{i,j} = e^{-\mathbf{C}_{i,j}/\epsilon}$ by calculating the KKT condition: Introducing two dual variables $\mathbf{f} \in \mathbb{R}^n, \mathbf{g} \in \mathbb{R}^n$ and calculate the lagrangian:

$$\mathcal{L}(\mathbf{P}, \mathbf{f}, \mathbf{g}) = \langle \mathbf{P}, \mathbf{C} \rangle - \varepsilon \mathbf{H}(\mathbf{P}) - \langle \mathbf{f}, \mathbf{P} \mathbf{1}_n - \mathbf{a} \rangle - \left\langle \mathbf{g}, \mathbf{P}^{\mathrm{T}} \mathbf{1}_n - \mathbf{b} \right\rangle \tag{6}$$

take first order gradient and we get

$$\frac{\partial \mathcal{L}(\mathbf{P}, \mathbf{f}, \mathbf{g})}{\partial \mathbf{P}_{i,j}} = \mathbf{C}_{i,j} + \varepsilon \log\left(\mathbf{P}_{i,j}\right) - \mathbf{f}_i - \mathbf{g}_j = 0 \tag{7}$$

$$\Rightarrow \mathbf{P}_{i,j} = e^{\mathbf{f}_i/\varepsilon} e^{-\mathbf{C}_{i,j}/\varepsilon} e^{\mathbf{g}_j/\varepsilon} \tag{8}$$

Based on the constrain that:

$$\operatorname{diag}(\mathbf{u}) \mathbf{K} \operatorname{diag}(\mathbf{v}) \mathbf{1}_m = \mathbf{a} \tag{9}$$

$$\operatorname{diag}(\mathbf{v}) \mathbf{K}^{\top} \operatorname{diag}(\mathbf{u}) \mathbf{1}_n = \mathbf{b} \tag{10}$$

or :

$$\mathbf{u} \odot (\mathbf{K} \mathbf{v}) = \mathbf{a} \quad \text{and} \quad \mathbf{v} \odot \left(\mathbf{K}^{\mathrm{T}} \mathbf{u}\right) = \mathbf{b} \tag{11}$$

(where $\odot$ means entry-wise multiplication of vectors) we can develop our algorithm as iteratively updating $\mathbf{u}$ and $\mathbf{v}$:

$$\mathbf{u}^{(\ell+1)} = \frac{\mathbf{a}}{\mathbf{K} \mathbf{v}^{(\ell)}} \text{ and } \mathbf{v}^{(\ell+1)} = \frac{\mathbf{b}}{\mathbf{K}^{\mathrm{T}} \mathbf{u}^{(\ell+1)}} \tag{12}$$

with $\mathbf{v}^{(0)} = \mathbf{1}_m$ and $\mathbf{K}_{i,j} = e^{-\mathbf{C}_{i,j}/\epsilon}$.

Acknowledgments

Use unnumbered third level headings for the acknowledgments. All acknowledgments go at the end of the paper. Do not include acknowledgments in the anonymized submission, only in the final paper.

References

References follow the acknowledgments. Use unnumbered first-level heading for the references. Any choice of citation style is acceptable as long as you are consistent. It is permissible to reduce the font size to `small` (9 point) when listing the references. **Remember that you can go over 8 pages as long as the subsequent ones contain *only* cited references.**

[1] Alexander, J.A. & Mozer, M.C. (1995) Template-based algorithms for connectionist rule extraction. In G. Tesauro, D.S. Touretzky and T.K. Leen (eds.), *Advances in Neural Information Processing Systems 7*, pp. 609–616. Cambridge, MA: MIT Press.

[2] Bower, J.M. & Beeman, D. (1995) *The Book of GENESIS: Exploring Realistic Neural Models with the GEneral NEural SImulation System.* New York: TELOS/Springer–Verlag.

[3] Hasselmo, M.E., Schnell, E. & Barkai, E. (1995) Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3. *Journal of Neuroscience* **15**(7):5249-5262.