On Unbalanced Optimal Transport: an Analysis of Sinkhorn Algorithm

Khiem Pham†, Khang Le†, Nhat Ho ‡, Tung Pham †, Hung Bui†

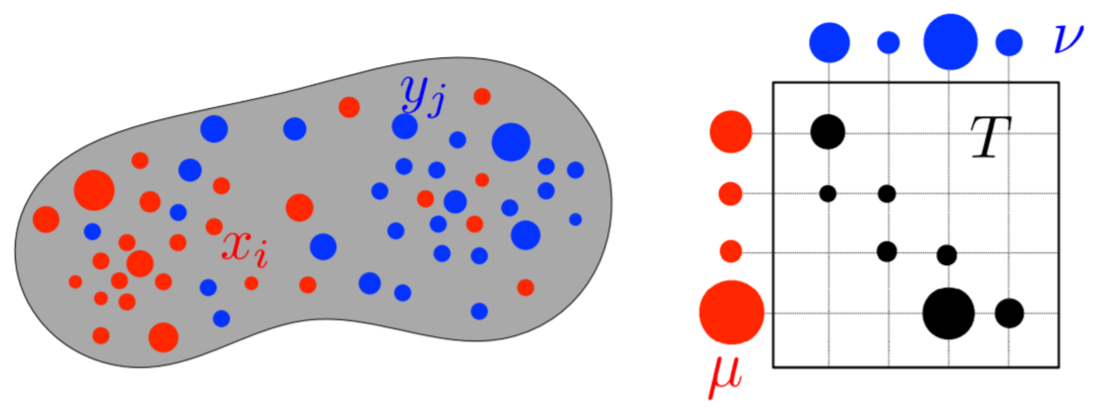
†VinAI Research ‡University of California, Berkeley

Introduction

We provide an complexity analysis for the Sinkhorn algorithm for solving the Unbalanced Optimal Transport problem, which is a generalized version of the Optimal Transport problem where marginal constraints are relaxed.

Concretely, the complexity is of order $\widetilde{\mathcal{O}}(n^2/\varepsilon^2)$.

Unbalanced Optimal Transport



Given two *probability* vectors $\mathbf{a} \in \mathbb{R}^n, \mathbf{b} \in \mathbb{R}^m$ *(i.e.* $\alpha = \beta = 1$, *where* $\alpha = \sum_i \mathbf{a}_i, \beta = \sum_j \mathbf{b}_j$*)*, and a cost matrix $C \in \mathbb{R}_+^{n\times m}$ the classical Optimal Transport (OT) problem aims to find a *transportation plan* $X \in \mathbb{R}_+^{n\times m}$ that

$$\begin{aligned} &\textit{minimizes} \quad \langle C, X\rangle, \\ &\textit{subject to} \quad X\mathbf{1}_m = \mathbf{a}, X^T\mathbf{1}_n = \mathbf{b}. \end{aligned}$$

If we relax $\mathbf{a}, \mathbf{b}$ to be arbitrary non-negative vectors, we have the Unbalanced Optimal Transport (UOT), that finds

$$\min_{X\in\mathbb{R}_+^{n\times m}} f(X) := \langle C, X\rangle + \tau \mathrm{KL}(X\mathbf{1}_n||\mathbf{a}) + \tau \mathrm{KL}(X^\top \mathbf{1}_n||\mathbf{b}),$$

where $\tau > 0$ tells how much we want the solutions to be close to two marginals.

Remark 1. *When* $\mathbf{a}^T\mathbf{1}_n = \mathbf{b}^T\mathbf{1}_m$ *and* $\tau \to \infty$*, UOT becomes OT.*

Remark 2. *Advantages of UOT over OT:*

- *enable transportation between measures of arbitrary masses,*
- *reduce effect of outliers via soft constraints.*

Dual Entropic-Regularized UOT

By introducing an entropy term $H(X) = \sum_{i,j=1}^n X_{ij}(\log(X_{ij}) - 1)$ to the UOT objective, we have the entropic UOT problem, i.e.

$$\min_{X\in\mathbb{R}^{n\times m}} g(X) := \langle C, X\rangle - \eta H(X) + \tau \mathrm{KL}(X\mathbf{1}_n||\mathbf{a}) + \tau \mathrm{KL}(X^\top \mathbf{1}_n||\mathbf{b}),$$

which removes the positive constrain on the variable. Besides, the optimal solution of this problem can be derived from the optimal solution of its dual problem, that reads

$$\min_{u,v\in\mathbb{R}^n} h(u,v) := \eta \sum_{i,j} e^{\frac{u_i+v_j-C_{ij}}{\eta}} + \tau \left\langle e^{-u/\tau}, \mathbf{a}\right\rangle + \tau \left\langle e^{-v/\tau}, \mathbf{b}\right\rangle,$$

which can be solved using the Sinkhorn algorithm.

Sinkhorn Algorithm

Input: $k = 0$ and $u^0 = v^0 = 0$, accuracy ε
while $k < \tilde{c}\frac{\log(n)}{\varepsilon}$ **do**
  $a^k = B(u^k, v^k)\mathbf{1}_n, \quad b^k = B(u^k, v^k)^\top \mathbf{1}_n.$
  if k is even **then**

$$u^{k+1} = \left[\frac{u^k}{\eta} + \log(a) - \log\left(a^k\right)\right]\frac{\eta\tau}{\eta+\tau}, \quad v^{k+1} = v^k$$

  else

$$v^{k+1} = \left[\frac{v^k}{\eta} + \log(b) - \log\left(b^k\right)\right]\frac{\eta\tau}{\eta+\tau}, \quad u^{k+1} = u^k.$$

  end if
  $k = k + 1.$
end while
Output: $B(u^k, v^k)$.

ε-approximation

For any $\varepsilon > 0$, we call X an ε-approximation transportation plan if the following holds

$$\begin{aligned} &\langle C, X\rangle + \tau \mathrm{KL}(X\mathbf{1}_n||\mathbf{a}) + \tau \mathrm{KL}(X^\top\mathbf{1}_n||\mathbf{b}) \\ &\le \left\langle C, \widehat{X}\right\rangle + \tau \mathrm{KL}(\widehat{X}\mathbf{1}_n||\mathbf{a}) + \tau \mathrm{KL}(\widehat{X}^\top\mathbf{1}_n||\mathbf{b}) + \varepsilon, \end{aligned}$$

where $\widehat{X}$ is an optimal transportation plan for the UOT problem.

Our contribution

We show that the complexity of the Sinkhorn algorithm is

$$\mathcal{O}\left(\frac{\tau(\alpha+\beta)n^2}{\varepsilon}\log(n)\left[\log(\|C\|_\infty) + \log(\log(n)) + \log\left(\frac{1}{\varepsilon}\right)\right]\right).$$

Detailed Analysis

Denote u^k, v^k the partial solution at the k-th iteration and u^*, v^* the optimal solution of the dual problem. Our main theorems are:

Theorem 1. *The update* (u^{k+1}, v^{k+1}) *from Sinkhorn Algorithm satisfies the following bound*

$$\max\left\{||u^{k+1} - u^*||_\infty, ||v^{k+1} - v^*||_\infty\right\} \le \left(\frac{\tau}{\tau+\eta}\right)^k \times \tau \times R$$

where $R := \|\log(\mathbf{a})\|_\infty + \|\log(\mathbf{b})\|_\infty + \max\left\{\log(n), \frac{1}{\eta}\|C\|_\infty - \log(n)\right\}$.

→ This shows that the Sinkhorn algorithm converges at a geometric rate.

Theorem 2. *Let*

$$S = \frac{1}{2}(\alpha+\beta) + \frac{1}{2} + \frac{1}{4\log(n)}, \quad S = \widetilde{\mathcal{O}}(\alpha+\beta)$$

$$T = \left(\frac{\alpha+\beta}{2}\right)\left[\log\left(\frac{\alpha+\beta}{2}\right) + 2\log(n) - 1\right] + \log(n) + \frac{5}{2}, \quad T = \widetilde{\mathcal{O}}((\alpha+\beta)\log(n))$$

$$U = \max\left\{S+T, 2\varepsilon, \frac{4\varepsilon\log(n)}{\tau}, \frac{4\varepsilon(\alpha+\beta)\log(n)}{\tau}\right\}. \quad U = \widetilde{\mathcal{O}}((\alpha+\beta)\log(n))$$

For $\eta = \frac{\varepsilon}{U}$ *and* $k \ge 1 + \left(\frac{\tau U}{\varepsilon} + 1\right)\left[\log(8\eta R) + \log(\tau(\tau+1)) + 3\log\left(\frac{U}{\varepsilon}\right)\right]$*, the solution* X^k *is an* ε*-approximation of* $\widehat{X}$.

→ With an additional $O(n^2)$ per iteration, we get the claimed complexity.

Experiment