Optimal Transport

Artur Bulanbaev

December 2022

Problem statement

Consider the OT problem between two discrete probability measures $p \in S_n(1), q \in S_n(1)$, where $S_n(1) = \{s \in \mathbb{R}^n_+ : \langle s, \mathbf{1} \rangle = 1\}$ is the standard probability simplex. Transportation plan $\pi \in \mathbb{R}^{n \times n}_+$ with the elements π_{ij} prescribes the amount of mass moved from the source point i to the target point j. Admissible transportation plans form the transportation polytope $U(p, q)$ of all coupling matrices with marginals equal to the source p and target q. Formally,

$$U(p, q) = \{\pi \in \mathbb{R}^{n \times n}_+ : \pi \mathbf{1}_n = p, \pi^T \mathbf{1}_n = q\}.$$

Problem statement

Cost matrix $C \in \mathbb{R}_+^{n \times n}$ with the elements C_{ij} giving the cost of transportation of a unit of mass from the source point i to the target point j.
The Monge–Kantorovich problem of finding a transportation plan π that minimizes the total cost of transportation of the distribution π to the distribution q reads:

$$W(p, q) = \min_{\pi \in U(p,q)} \langle C, \pi \rangle,$$

where $\langle A, B \rangle = \sum_{i,j}^{n,n} A_{ij} B_{ij}$.

Dual problem

The Monge–Kantorovich problem is a constrained convex minimization problem, and as such, it can be naturally paired with a so-called dual problem, which is a constrained concave maximization problem. The responding dual problem is written as

$$W(p, q) = \max_{y,z \in R(C)} \langle y, p \rangle + \langle z, q \rangle,$$

where $R(C) = \{(y, z) \in \mathbb{R}^n \times \mathbb{R}^n : y \bigoplus z \leq C\}$.

Entropic regularization

The main idea of entropic regularization is to diversify transport plan and to make solution more robust. Let us consider the following entropy-regularized OT problem:

$$W_\gamma(p,q) = \min_{\pi \in U(p,q)} \{g(\pi) = \langle C, \pi \rangle + \gamma H(\pi)\},$$

where $H(\pi) = \langle \pi, \log \pi \rangle$ is the negative entropy and $\gamma > 0$ is regularization parameter.
As negative entropy is strongly convex, the objective W_γ is also strongly convex, thus the problem has unique solution.

Altogether

Combining duality and regularization, we may rewrite the problem as

$$\max_{y,z\in\mathbb{R}^n} \phi(y,z) = \max_{y,z\in\mathbb{R}^n} \langle y, p\rangle + \langle z, q\rangle + \frac{\gamma}{e}\sum_{i,j=1}^{n} e^{-\dfrac{y_i + z_j + C_{ij}}{\gamma}}$$

Making variable change we come to standard formulation

$$\min_{u,v\in\mathbb{R}^n} \{f(u,v) = \gamma(\mathbf{1}_n^T B(u,v)\mathbf{1}_n - \langle u, p\rangle - \langle v, q\rangle)\},$$

where $u = -\dfrac{y}{\gamma} - \dfrac{1}{2}, v = -\dfrac{z}{\gamma} - \dfrac{1}{2}, B(u,v)_{ij} = \exp(u_i + v_j - \dfrac{C_{ij}}{\gamma})$.

Algorithms

The state-of-the-art algorithm for solving the regularized OT problem is the Sinkhorn's algorithm:

Algorithm 1 Sinkhorn's algorithm

Input: $C, p, q, \gamma > 0$

for $t \geqslant 1$ **do**

  if $t \bmod 2 = 0$ **then**

    $u^{t+1} = u^t + \ln p - \ln\left(\pi\left(u^t, v^t\right)\mathbf{1}_n\right),\ v^{t+1} = v^t$

  else

    $u^{t+1} = u^t,\ v^{t+1} = v^t + \ln q - \ln\left(\pi\left(u^t, v^t\right)^T \mathbf{1}_n\right)$

  end if

  $t = t + 1$

end for

Output: $\check{\pi} = \pi(u^t, v^t)$

Complexity: $O(n^2 \|C\|_\infty^2 / \gamma\epsilon)$

Algorithms

The following modification uses projections to recalculate transport plan.

Algorithm 2 KL projection form Sinkhorn

Input: Cost matrix C, probability measures $p, q, \gamma > 0$, starting transport plans $\pi^0 := \exp\left(-\frac{C}{\gamma}\right)$

for t=0,1,... **do**
    if $t \bmod 2 = 0$ **then**
        $\pi^{t+1} := \underset{\pi \in C_1}{\operatorname{argmin}} KL\left(\pi | \pi^t\right)$
    else
        $\pi^{t+1} := \underset{\pi \in C_2}{\operatorname{argmin}} KL\left(\pi | \pi^t\right)$
    end if
    $t := t + 1$
end for

Output: π^t

Algorithms

where $KL(\pi|\pi') := \sum_{i,j=1}^{n} \left(\pi_{ij} \ln\left(\frac{\pi_{ij}}{\pi'_{ij}}\right) - \pi_{ij} + \pi'_{ij}\right) = \langle \pi, \ln \pi - \ln \pi' \rangle + \langle \pi' - \pi, \mathbf{1}_n \mathbf{1}_n^T \rangle$, $K = \exp(-C/\gamma)$ and the affine convex sets C_1 and C_2 with

$$C_1 = \{\pi : \pi \mathbf{1}_n = p\}, \quad C_2 = \left\{\pi : \pi^T \mathbf{1}_n = q\right\}. \tag{17}$$

Algorithms

Algorithm 4 Approximate OT by Sinkhorn

Input: Accuracy ε.

1: Set $\varepsilon' = \frac{\varepsilon}{8\|C\|_\infty}$.

2: Set $(\check{p}, \check{q}) = \left(1 - \frac{\varepsilon'}{8}\right)\left((p, q) + \frac{\varepsilon'}{8n}(\mathbf{1}_n, \mathbf{1}_n)\right)$

3: Calculate $\check{\pi}$ s.t. $\|\check{\pi}\mathbf{1}_n - \check{p}\|_1 + \|\check{\pi}^T\mathbf{1}_n - \check{q}\|_1 \leqslant \varepsilon'/2$ by Algorithm 1 with marginals $\check{p}, \check{q}$ and $\gamma = \frac{\varepsilon}{4\ln n}$

4: Find $\hat{\pi}$ as the projection of $\check{\pi}$ on $U(p, q)$ by Algorithm 3.

Output: $\hat{\pi}$

Algorithms

Theorem 3. *For a given $\varepsilon > 0$, Algorithm 4 returns $\hat{\pi} \in U(p,q)$ s.t.*

$$\langle C, \hat{\pi} \rangle - \langle C, \pi^* \rangle \leqslant \varepsilon,$$

and requires

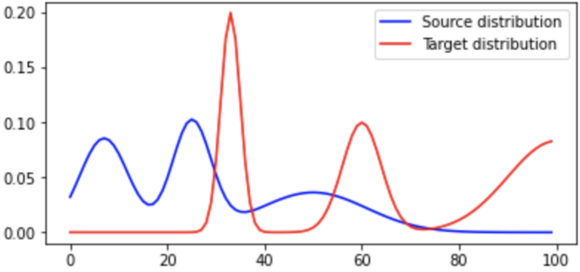
$$O\left(\left(\frac{\|C\|_\infty}{\varepsilon} \right)^2 M_n \ln n \right)$$

arithmetic operations, where M_n is a time complexity of one iteration of Algorithm 1.

As each iteration of Algorithm 1 requires a matrix-vector multiplication, the general bound is $M_n = O(n^2)$. However, for some specific forms of matrix C it is possible to achieve better complexity, e.g. $M_n = O(n \log n)$ via FFT [52].

Experiments

Few experiments were made on synthetic gaussian mixture data.



Experiments