Optimal Transport

Numerical methods with entropic regularization

Katya Filimoshina

Contents

Optimal Transport problem

Notation and definitions

Definition 1 (Probability simplex) *A standard probability simplex is a set of the form*

$$S_n(1) := \{s \in \mathbb{R}_n^+ : \langle s, 1_\mathrm{n} \rangle := \sum_{i=1}^{n} s_i = 1\}. \tag{1}$$

Definition 2 (Coupling matrix (Transportation plan)) *A coupling matrix is a matrix $\pi \in \mathbb{R}_+^{n \times n}$ with elements π_{ij} prescribing the amount of mass moved from the source point i to the target point j.*

Suppose $p \in S_n(1)$ and $q \in S_n(1)$ are two discrete probability measures.

Definition 3 (Transportation polytope) *Transportation polytope is a set (a convex hull of finite set of matrices) of all coupling matrices of size $n \times n$:*

$$\begin{aligned} U(p, q) \; &:= \; \{\pi \in \mathbb{R}_+^{n \times n} : \quad \pi 1_n = p, \quad \pi^T 1_n = q\} & (2) \\ &= \; \{\pi \in \mathbb{R}_+^{n \times n} : \quad \sum_{j=1}^{n} \pi_{ij} = p_i, \quad \forall i = 1, \ldots, n; \quad \sum_{i=1}^{n} \pi_{ij} = q_j, \quad \forall j = 1, \ldots, n\}. & (3) \end{aligned}$$

Definition 4 (Transportation cost matrix) *A transportation cost matrix $C \in \mathbb{R}_+^{n \times n}$ is a symmetric matrix with elements C_{ij} describing cost of transportation of a unit of mass from the source point i to the target point j.*

Optimal Transport (Monge–Kantorovich) problem

Primal problem

Monge–Kantorovich problem is a problem of finding a transportation plan π that minimizes the total cost of transportation of the distribution p to the distribution q:

$$\mathrm{L}_C(p,q) := \min_{\pi \in U(p,q)} \langle C, \pi \rangle := \min_{\pi \in U(p,q)} \sum_{i,j} C_{ij}\pi_{ij}, \tag{4}$$

where $\langle C, \pi \rangle$ is a Frobenius inner product of C and π. The problem (4) is a linear programming problem and can be solved by the corresponding methods.

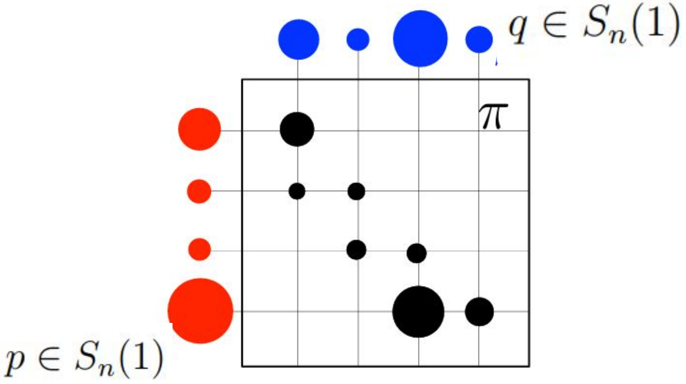
Optimal Transport (Monge–Kantorovich) problem

Primal problem

Monge–Kantorovich problem is a problem of finding a transportation plan π that minimizes the total cost of transportation of the distribution p to the distribution q:

$$\mathrm{L}_C(p,q) := \min_{\pi \in U(p,q)} \langle C, \pi \rangle := \min_{\pi \in U(p,q)} \sum_{i,j} C_{ij}\pi_{ij}, \tag{4}$$

where $\langle C, \pi \rangle$ is a Frobenius inner product of C and π. The problem (4) is a linear programming problem and can be solved by the corresponding methods.



Optimal Transport (Monge–Kantorovich) problem

Dual problem

Monge–Kantorovich problem is a problem of finding a transportation plan π that minimizes the total cost of transportation of the distribution p to the distribution q:

$$\mathrm{L}_C(p,q) := \min_{\pi \in U(p,q)} \langle C, \pi \rangle := \min_{\pi \in U(p,q)} \sum_{i,j} C_{ij}\pi_{ij}, \tag{4}$$

The problem (4) can be naturally paired with the following dual problem, which is a constrained concave maximization problem.

Proposition 1 *The Monge–Kantorovich problem (4) admits the dual*

$$\mathrm{L}_C(p,q) = \max_{(f,g)\in \mathrm{R}(C)} \langle f, p \rangle + \langle g, q \rangle, \tag{5}$$

where $\mathrm{R}(C)$ *is the set of admissible dual variables:*

$$\begin{aligned} \mathrm{R}(C) \; &:= \; \{(f,g) \in \mathbb{R}^n \times \mathbb{R}^n : \quad f \oplus g := f \mathbb{1}_n + \mathbb{1}_n g^T \leq C\} && (6) \\ &= \; \{(f,g) \in \mathbb{R}^n \times \mathbb{R}^n : \quad f_i + g_j \leq C_{ij}, \quad \forall i,j = 1,\dots,n\}. && (7) \end{aligned}$$

The dual variables (f,g) are usually called *Kantorovich potentials*.

Optimal Transport (Monge–Kantorovich) problem

Relation between primal and dual problem

Primal problem:

$$\mathrm{L}_C(p,q) := \min_{\pi \in U(p,q)} \langle C, \pi \rangle := \min_{\pi \in U(p,q)} \sum_{i,j} C_{ij}\pi_{ij}, \tag{4}$$

Dual problem:

$$\mathrm{L}_C(p,q) = \max_{(f,g)\in \mathrm{R}(C)} \langle f, p \rangle + \langle g, q \rangle, \tag{5}$$

Remark 1. Before we prove Proposition 1, let us note that the right-hand side of the equation (5) is the lower-bound of $\mathrm{L}_C(p,q)$ (4):

$$\min_{\pi \in U(p,q)} \sum_{i,j} C_{ij}\pi_{ij} \geq \max_{(f,g)\in \mathrm{R}(C)} \langle f, p \rangle + \langle g, q \rangle. \tag{8}$$

This is because for any transport plan π (including the optimal one) and for any $(f,g) \in \mathrm{R}(C)$,

$$\sum_{i,j} \pi_{ij} C_{ij} \geq \sum_{i,j} \pi_{ij}(f_i + g_j) = (\sum_i f_i \sum_j \pi_{ij}) + (\sum_j g_j \sum_i \pi_{ij}) = \langle f, p \rangle + \langle g, q \rangle. \tag{9}$$

Optimal Transport (Monge–Kantorovich) problem

Relation between primal and dual problem

Primal problem:

$$\mathrm{L}_C(p, q) := \min_{\pi \in U(p,q)} \langle C, \pi \rangle := \min_{\pi \in U(p,q)} \sum_{i,j} C_{ij} \pi_{ij}, \tag{4}$$

Proposition 1 *The Monge–Kantorovich problem (4) admits the dual*

$$\mathrm{L}_C(p, q) = \max_{(f,g) \in \mathrm{R}(C)} \langle f, p \rangle + \langle g, q \rangle, \tag{5}$$

where $\mathrm{R}(C)$ *is the set of admissible dual variables:*

$$\mathrm{R}(C) \ := \ \{(f, g) \in \mathbb{R}^n \times \mathbb{R}^n : \quad f \oplus g := f 1_n + 1_n g^T \leq C\} \tag{6}$$

Optimal Transport (Monge–Kantorovich) problem

Relation between primal and dual problem

Primal problem:
$$\mathrm{L}_C(p,q) := \min_{\pi \in U(p,q)} \langle C, \pi \rangle := \min_{\pi \in U(p,q)} \sum_{i,j} C_{ij}\pi_{ij}, \tag{4}$$

Proposition 1 *The Monge–Kantorovich problem (4) admits the dual*

$$\mathrm{L}_C(p,q) = \max_{(f,g)\in \mathrm{R}(C)} \langle f, p \rangle + \langle g, q \rangle, \tag{5}$$

Proof. [of Proposition 1] Let us write down the Lagrangian to the problem (4):

$$\min_{\pi \in \mathbb{R}_+^{n\times n}} \max_{(f,g)\in \mathbb{R}^n \times \mathbb{R}^n} \langle C, \pi \rangle + \langle p - \pi 1_n, f \rangle + \langle q - \pi^T 1_n, g \rangle, \tag{10}$$

which is equivalent to

$$\min_{\pi \in \mathbb{R}_+^{n\times n}} \max_{(f,g)\in \mathbb{R}^n \times \mathbb{R}^n} \langle C, \pi \rangle + \langle p, f \rangle - \langle f, \pi 1_n \rangle + \langle q, g \rangle - \langle g, \pi^T 1_n \rangle \quad \Leftrightarrow \tag{11}$$

$$\min_{\pi \in \mathbb{R}_+^{n\times n}} \max_{(f,g)\in \mathbb{R}^n \times \mathbb{R}^n} \langle C, \pi \rangle + \langle p, f \rangle - \langle f 1_n^T, \pi \rangle + \langle q, g \rangle - \langle 1_n g^T, \pi \rangle \quad \Leftrightarrow \tag{12}$$

$$\min_{\pi \in \mathbb{R}_+^{n\times n}} \max_{(f,g)\in \mathbb{R}^n \times \mathbb{R}^n} \langle p, f \rangle + \langle q, g \rangle + \langle C - f 1_n^T - 1_n g^T, \pi \rangle. \tag{13}$$

Optimal Transport (Monge–Kantorovich) problem

Relation between primal and dual problem

Primal problem:

$$\mathrm{L}_C(p,q) := \min_{\pi\in U(p,q)} \langle C,\pi\rangle := \min_{\pi\in U(p,q)} \sum_{i,j} C_{ij}\pi_{ij}, \tag{4}$$

Proposition 1 *The Monge–Kantorovich problem (4) admits the dual*

$$\mathrm{L}_C(p,q) = \max_{(f,g)\in\mathrm{R}(C)} \langle f,p\rangle + \langle g,q\rangle, \tag{5}$$

Proof. Now we can exchange min and max, which is always possible in linear programs in finite dimension, and get:

$$\max_{(f,g)\in\mathbb{R}^n\times\mathbb{R}^n} \langle p,f\rangle + \langle q,g\rangle + \min_{\pi\in\mathbb{R}_+^{n\times n}} \langle C - f1_n^T - 1_n g^T, \pi\rangle. \tag{14}$$

Now let us note that

$$\min_{\pi\in\mathbb{R}_+^{n\times n}} \langle Q,\pi\rangle = \begin{cases} 0, & \text{if } Q\in\mathbb{R}_+^{n\times n}, \\ -\infty, & \text{otherwise.} \end{cases} \tag{15}$$

Therefore, from the constraint, we get $C - f1_n^T - 1_n g^T = C - f\oplus g \geq 0$, and the statement is proved. ■

Entropic regularization

Definition of Entropy of a coupling matrix

Entropic regularization is an important technique used in the context of optimal transport to introduce a smoothing term that enhances computational efficiency and stability. Below, we consider this concept.

Definition 5 (Discrete entropy of a coupling matrix) *Discrete entropy of a coupling matrix is defined as*

$$\mathrm{H}(\pi) = -\sum_{i,j} \pi_{ij}(log(\pi_{ij}) - 1). \tag{16}$$

The definition for vectors $a \in \mathbb{R}^n$ is similar, with the convention that $H(a) = +\infty$ if one of the entries a_j is ≤ 0.

Note that the function $-H$ is 1-strongly convex, because its Hessian is

$$\partial^2(-\mathrm{H}(\pi)) = -\partial^2(\mathrm{H}(\pi)) = -(-diag(\frac{1}{\pi_{ij}})) = diag(\frac{1}{\pi_{ij}})$$

and $\pi_{ij} \leq 1$, so $\partial^2(-\mathrm{H}(\pi)) \geq I_{n^2}$.

Entropic regularization

OT problem with regularization term

The idea of **entropic regularization** of the optimal transport problem (4) is to use $-\mathrm{H}$ as a regularizing function to obtain approximate solutions of (4). This results in an approximation of the original problem that is easier to solve:

$$\mathrm{L}_C^{\varepsilon}(p, q) := \min_{\pi \in U(p,q)} \langle \pi, C \rangle - \varepsilon \mathrm{H}(\pi). \tag{17}$$

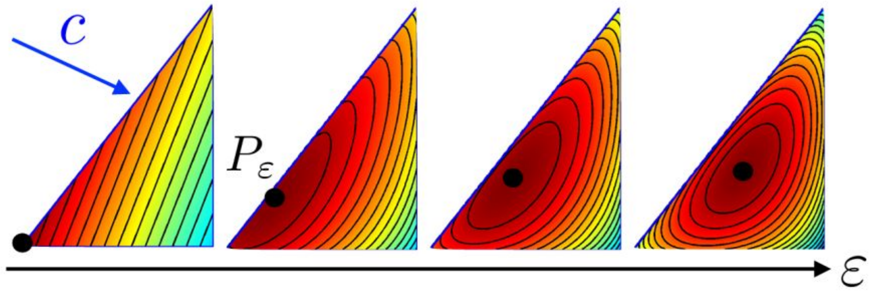
The objective is ε-strongly convex function, and problem (17) has a unique optimal solution.

Entropic regularization

OT problem with regularization term: impact of eps

Example 1. Impact of ε on solution [7] Figure 1 illustrates the effect of the entropy to regularize a linear program over the simplex $S_1(3)$ (which can thus be visualized as a triangle in two dimensions). The entropy pushes the original LP solution away from the boundary of the triangle. The optimal $P_\varepsilon := \pi_\varepsilon$ progressively moves toward an "entropic center" of the triangle.

Figure 1: Example: impact of ε on the optimization of a linear function on the simplex, solving $P_\varepsilon := \pi_\varepsilon = \arg\min_\pi \langle \pi, C \rangle - \varepsilon \mathrm{H}(\pi)$ for a varying ε.



Entropic regularization

OT problem with regularization term: impact of eps

Primal problem:
$$\mathrm{L}_C(p,q) := \min_{\pi \in U(p,q)} \langle C, \pi \rangle := \min_{\pi \in U(p,q)} \sum_{i,j} C_{ij}\pi_{ij}, \tag{4}$$

With reg.:
$$\mathrm{L}_C^\varepsilon(p,q) := \min_{\pi \in U(p,q)} \langle \pi, C \rangle - \varepsilon \mathrm{H}(\pi). \tag{17}$$

Proposition 2 (Convergence with ε [7]) *The unique solution π_ε of (17) converges to the optimal solution with maximal entropy within the set of all optimal solutions of the initial problem (4):*

$$\pi_\varepsilon \overset{\varepsilon \to 0}{\to} \arg\min_{\pi} \{-\mathrm{H}(\pi) : \quad \pi \in (p,q), \quad \langle \pi, C \rangle = \mathrm{L}_C(p,q)\} \tag{18}$$

so that in particular

$$\mathrm{L}_C^\varepsilon(p,q) \overset{\varepsilon \to 0}{\to} \mathrm{L}_C(p,q). \tag{19}$$

One also has

$$\pi_\varepsilon \overset{\varepsilon \to \infty}{\to} ab^T. \tag{20}$$

Numerical algorithms for OT

with entropic regularization

1. Sinkhorn's algorithm
2. Sinkhorn–Knopp algorithm
3. Greenkhorn algorithm
4. Adaptive Primal-Dual Accelerated Gradient Descent (APDAGD)
5. Primal-Dual Accelerated Alternating Minimization Algorithm (AAM)
6. ...

Sinkhorn's algorithm

OT problem with regularization

$$\mathrm{L}_C^{\varepsilon}(p,q) := \min_{\pi \in U(p,q)} \langle \pi, C \rangle - \varepsilon \mathrm{H}(\pi). \tag{17}$$

Definition 6 (Gibbs kernel associated to cost matrix) *A Gibbs kernel associated to a cost matrix $C \in \mathbb{R}_+^{n \times n}$ is a matrix $K \in \mathbb{R}_+^{n \times n}$ defined as*

$$K_{ij} = e^{\frac{-C_{ij}}{\varepsilon}}$$

Proposition 3 *The solution to the OT problem with entropic regularization (17) is unique and has the form*

$$\pi_{ij} = u_i K_{ij} v_j, \tag{21}$$

for two (unknown) scaling variables $(u, v) \in \mathbb{R}_+^n \times \mathbb{R}_+^n$.

Sinkhorn's algorithm

OT problem with regularization

$$\mathrm{L}_C^{\varepsilon}(p,q) := \min_{\pi \in U(p,q)} \langle \pi, C \rangle - \varepsilon \mathrm{H}(\pi). \tag{17}$$

Proposition 3 *The solution to the OT problem with entropic regularization (17) is unique and has the form*

$$\pi_{ij} = u_i K_{ij} v_j, \tag{21}$$

for two (unknown) scaling variables $(u,v) \in \mathbb{R}_+^n \times \mathbb{R}_+^n$.

Proof. Let us write down the Lagrangian of (17) with two dual variables $f, g \in \mathbb{R}^n$. Similarly to the proof of Proposition 1, we have

$$\min_{\pi \in \mathbb{R}_+^{n\times n}} \max_{(f,g)\in \mathbb{R}^n \times \mathbb{R}^n} \langle C, \pi \rangle - \varepsilon \mathrm{H}(\pi) - \langle f, \pi 1_n - p \rangle - \langle g, \pi^T 1_n - q \rangle. \tag{22}$$

Then the first order condition, by taking the derivative by π_{ij} of the expression, gives us

$$C_{ij} + \varepsilon \log(\pi_{ij}) - f_i - g_j = 0. \tag{23}$$

Therefore,

$$\pi_{ij} = e^{\frac{f_i}{\varepsilon} \frac{-C_{ij}}{\varepsilon} \frac{g_j}{\varepsilon}},$$

which can be rewritten in the form provided above with non-negative vectors u and v. ■

Sinkhorn's algorithm

OT problem with regularization

Proposition 3 *The solution to the OT problem with entropic regularization (17) is unique and has the form*

$$\pi_{ij} = u_i K_{ij} v_j, \tag{21}$$

for two (unknown) scaling variables $(u, v) \in \mathbb{R}_+^n \times \mathbb{R}_+^n$.

Remark 2. Matrix and vector forms of optimal solution Note that the factorization (21) of an optimal solution can be rewritten in matrix form:

$$\pi = \mathrm{diag}(u) K \mathrm{diag}(v). \tag{24}$$

By the definition of π, the following equalities should be satisfied:

$$\mathrm{diag}(u) K \mathrm{diag}(v) 1_n = p, \qquad \mathrm{diag}(v) K^T \mathrm{diag}(u) 1_n = q. \tag{25}$$

Note that $\mathrm{diag}(v) 1_n = v$ and $\mathrm{diag}(u) 1_n = u$, and we get

$$u \odot (Kv) = p, \qquad v \odot (K^T u) = q, \tag{26}$$

where $\odot$ stands for Hadamard (element-wise) product. That problem is known in the numerical analysis community as the matrix scaling problem (see [6]).

Sinkhorn's algorithm

OT problem with regularization

Proposition 3 *The solution to the OT problem with entropic regularization (17) is unique and has the form*

$$\pi_{ij} = u_i K_{ij} v_j, \tag{21}$$

for two (unknown) scaling variables $(u, v) \in \mathbb{R}_+^n \times \mathbb{R}_+^n$.

Remark 2. Matrix and vector forms of optimal solution Note that the factorization (21) of an optimal solution can be rewritten in matrix form:

$$\pi = \mathrm{diag}(u) K \mathrm{diag}(v). \tag{24}$$

By the definition of π, the following equalities should be satisfied:

$$\mathrm{diag}(u) K \mathrm{diag}(v) 1_n = p, \qquad \mathrm{diag}(v) K^T \mathrm{diag}(u) 1_n = q. \tag{25}$$

Note that $\mathrm{diag}(v) 1_n = v$ and $\mathrm{diag}(u) 1_n = u$, and we get

$$u \odot (Kv) = p, \qquad v \odot (K^T u) = q, \tag{26}$$

where $\odot$ stands for Hadamard (element-wise) product. That problem is known in the numerical analysis community as the matrix scaling problem (see [6]).

Sinkhorn's algorithm

OT problem with regularization

Proposition 3 *The solution to the OT problem with entropic regularization (17) is unique and has the form*

$$\pi_{ij} = u_i K_{ij} v_j, \tag{21}$$

for two (unknown) scaling variables $(u, v) \in \mathbb{R}^n_+ \times \mathbb{R}^n_+$.

$$u \odot (Kv) = p, \qquad v \odot (K^T u) = q, \tag{26}$$

Remark 3. Sinkhorn's algorithm updates An intuitive way to handle the equations (26) is to solve them iteratively, by modifying first u so that it satisfies the left-hand side of (26) and then v to satisfy its right-hand side. These two updates define Sinkhorn's algorithm:

$$u^{l+1} := \frac{p}{Kv^{(l)}}, \qquad v^{l+1} := \frac{q}{K^T u^{(l+1)}}, \tag{27}$$

where $v^{(0)}$ can be initialized as 1_n and the division is performed element-wise. Note that a different initialization will likely lead to a different solution for u, v, since u, v are only defined up to a multiplicative constant, but all result in the same optimal coupling $\mathrm{diag}(u)K\mathrm{diag}(v)$.

Sinkhorn's algorithm

OT problem with regularization

Algorithm 1 Sinkhorn's Algorithm

Input: Cost matrix $C \in \mathbb{R}^{n \times n}$, marginals $p \in \mathbb{R}^n$, $q \in \mathbb{R}^n$, reg. parameter $\varepsilon > 0$, tolerance $\tau > 0$

Initialize $K \leftarrow \exp\left(-\frac{C}{\varepsilon}\right)$ (element-wise exponent)

Initialize $v \leftarrow \mathbf{1}_n$ (vector of ones)

while not converged **do**

  $u \leftarrow p/(Kv)$ element-wise

  $v \leftarrow q/(K^T u)$ element-wise

  if $\|u \odot (Kv) - p\|_1 + \|v \odot (K^T \cdot u) - q\|_1 < \tau$ **then**

    break

  end if

end while

Output: Approximate matrix $\pi \in \mathbb{R}^{n \times n}$

Sinkhorn–Knopp algorithm

OT problem with regularization

Sinkhorn-Knopp Algorithm is based on the observation that the coupling matrix π can be efficiently adjusted through scaling operations to enforce the marginal constraints while minimizing the entropic regularized cost. The Sinkhorn-Knopp algorithm minimizes the regularized optimal transport objective (17) by alternating between two steps: row and column scaling.

Algorithm 2 Sinkhorn-Knopp Algorithm

Input: Cost matrix C, source distribution p, target distribution q, regularization parameter ε

Initialize $\pi^{(1)} \leftarrow \exp\left(-\frac{C}{\varepsilon}\right)$

for each iteration $k = 1, \ldots, N$ **do**

  $\pi^{(k+\frac{1}{2})} \leftarrow \text{diag}\left(\frac{p}{\pi^{(k)} 1_n}\right) \pi^{(k)}$ ▷ Row scaling (element-wise division)

  $\pi^{(k+1)} \leftarrow \pi^{(k+\frac{1}{2})} \left(\frac{q}{(\pi^{(k+\frac{1}{2})})^T 1_n}\right)$ ▷ Column scaling (element-wise division)

end for

Output: Approximate matrix $\pi \in \mathbb{R}^{n \times n}$

Greenkhorn algorithm

OT problem with regularization

In practice, Greenkhorn algorithm usually gives better results than Sinkhorn's algorithm [4].

In each iteration of the Sinkhorn's (and Sinkhorn–Knopp) algorithm, all the elements of $u^{(k)}$ or $v^{(k)}$ are updated simultaneously such that the row sum of $\pi^{(k)}$ equals p or the column sum of $\pi^{(k)}$ equals q. In the Greenkhorn algorithm, in contrast, only a single element of $u^{(k)}$ or $v^{(k)}$ is updated at a time, such that only one element of the row sum or column sum of $\pi^{(k)}$ is equal to the target value. To determine which element of $u^{(k)}$ or $v^{(k)}$ is updated, the following scalar version of the KL-divergence is used to quantify the mismatch between the elements of p or q and the corresponding elements of $\pi 1_n$ or $\pi^T 1_n$:

Definition 7 (Scalar-value KL-divergence) *For two scalars $x, y \in \mathbb{R}$, KL-divergence is defined as*

$$\rho(x, y) := y - x + x \log \frac{x}{y}. \tag{28}$$

Note that $\rho(x, y)$ is indeed the Bregman distance between x and y associated with the function $\phi(t) = t \log t$.

Greenkhorn algorithm

OT problem with regularization

Algorithm 3 Greenkhorn Algorithm

Input: Cost matrix $C \in \mathbb{R}^{n \times n}$, marginals $p \in \mathbb{R}^n$, $q \in \mathbb{R}^n$, reg. parameter $\varepsilon > 0$, tolerance $\tau > 0$

Initialize $K \leftarrow \exp\left(-\frac{C}{\varepsilon}\right)$ ▷ (element-wise exponent)

Initialize $u^{(0)} \leftarrow p$, $v^{(0)} \leftarrow q$, $\pi^{(0)} \leftarrow \mathrm{diag}(u_0) K \mathrm{diag}(v_0)$

while not converged **do**

  $I \leftarrow \arg\max_i \rho(p_i, (\pi^{(k)} 1_n)_i)$ ▷ Choose row of $\pi^{(k)}$ "farest" from p

  $J \leftarrow \arg\max_j \rho(q_j, ((\pi^{(k)})^T 1_n)_j)$ ▷ Choose column of $\pi^{(k)}$ "farest" from q

  $u^{(k+1)} \leftarrow u^{(k)}$, $v^{(k+1)} \leftarrow v^{(k)}$,

  if $\rho(p_I, (\pi^{(k)} 1_n)_I) > \rho(q_J, ((\pi^{(k)})^T 1_n)_J)$ **then** ▷ If current rows are "worse" than columns

    $(u^{(k+1)})_I \leftarrow \frac{p_I}{(K v^{(k)})_I}$

  else $(v^{(k+1)})_J \leftarrow \frac{q_J}{(K^T u^{(k)})_J}$

  end if

  $\pi^{(k+1)} \leftarrow \mathrm{diag}(u^{(k+1)}) K \mathrm{diag}(v^{(k+1)})$

  if $\|u^{(k+1)} \odot (K v^{(k+1)}) - p\|_1 + \|v^{(k+1)} \odot (K^T \cdot u^{(k+1)}) - q\|_1 < \tau$ **then**

    break

  end if

  $k \leftarrow k + 1$

end while

Output: Matrix $\pi^{(k)} \in \mathbb{R}^{n \times n}$

Complexities of different algorithms

for OT problem with regularization

Table 1: Complexities of various algorithms.

Algorithm	**Complexity**
Sinkhorn	$O(n^2\|C^3\|_\infty \log n/\delta^3)$ [5]
Sinkhorn–Knopp	$O(n^2\|C^3\|_\infty \log n/\delta^3)$ [5]
Greenkhorn	$O(n^2\|C^3\|_\infty \log n/\delta^3)$ [5]
Primal-Dual AGD	$O(\frac{n^{\frac{5}{2}}\sqrt{\log n}\|C\|_\infty}{\delta})$[1]
Primal-Dual AAM	$O(\frac{n^{\frac{5}{2}}\sqrt{\log n}\|C\|_\infty}{\delta})$ [2]

Complexities of different algorithms

for OT problem with regularization

Table 1: Complexities of various algorithms.

I implemented:

Algorithm	Complexity
Sinkhorn	$O(n^2\|C^3\|_\infty \log n/\delta^3)$ [5]
Sinkhorn–Knopp	$O(n^2\|C^3\|_\infty \log n/\delta^3)$ [5]
Greenkhorn	$O(n^2\|C^3\|_\infty \log n/\delta^3)$ [5]
Primal-Dual AGD	$O(\frac{n^{\frac{5}{2}}\sqrt{\log n}\|C\|_\infty}{\delta})$ [1]
Primal-Dual AAM	$O(\frac{n^{\frac{5}{2}}\sqrt{\log n}\|C\|_\infty}{\delta})$ [2]

Experiment 1

Cost matrix 3x3

```
Cost matrix:
 [[0 1 1]
 [1 0 1]
 [1 1 0]]
p:
 [0.4 0.3 0.3]
q:
 [0.5 0.2 0.3]
```

Experiment 1

Cost matrix 3x3

Regularization param: $\varepsilon = 0.1$

Sinkhorn

```
Transport Matrix (pi):
 [[4.00000073e-01 1.66823892e-09 2.12797232e-07]
 [9.88420955e-02 1.99999892e-01 1.15822429e-03]
 [1.15783176e-03 1.06362480e-07 2.98841563e-01]]
Final transport cost: 0.10115847239941189
Converged in 786 iterations
```

Sinkhorn–Knopp

```
Transport Matrix (pi):
 [[4.00000066e-01 1.66823888e-09 2.12796228e-07]
 [9.88420963e-02 1.99999892e-01 1.15821885e-03]
 [1.15783722e-03 1.06362981e-07 2.98841568e-01]]
Final transport cost: 0.10115847323685895
Converged in 788 iterations
```

Greenkhorn

```
Transport Matrix (pi):
 [[4.00000159e-01 1.66824492e-09 2.12751914e-07]
 [9.88417612e-02 1.99999892e-01 1.15797346e-03]
 [1.15807977e-03 1.06385623e-07 2.98841855e-01]]
Final transport cost: 0.10115813526360722
Converged in 146 iterations
```

AGD

```
Transport Matrix (pi):
 [[4.00000000e-01 1.25978650e-17 2.88216950e-17]
 [9.99657538e-02 2.00000000e-01 3.42461672e-05]
 [3.42461672e-05 2.14187923e-17 2.99965754e-01]]
Final transport cost: 0.1000342461672061 6
Number of iterations: 56
```

Experiment 2

Regularization param: $\varepsilon = 0.0001$

Cost matrix 3x3

Sinkhorn

Sinkhorn–Knopp

Greenkhorn

```
Max_iter is achieved
Transport Matrix (pi):
 [[0.5 0.  0. ]
 [0.  0.2 0. ]
 [0.  0.  0.3]]
Final transport cost: 0.0
Converged in 1000 iterations

Correspondence of pi to p: False
Correspondence of pi to q: True
```

AGD

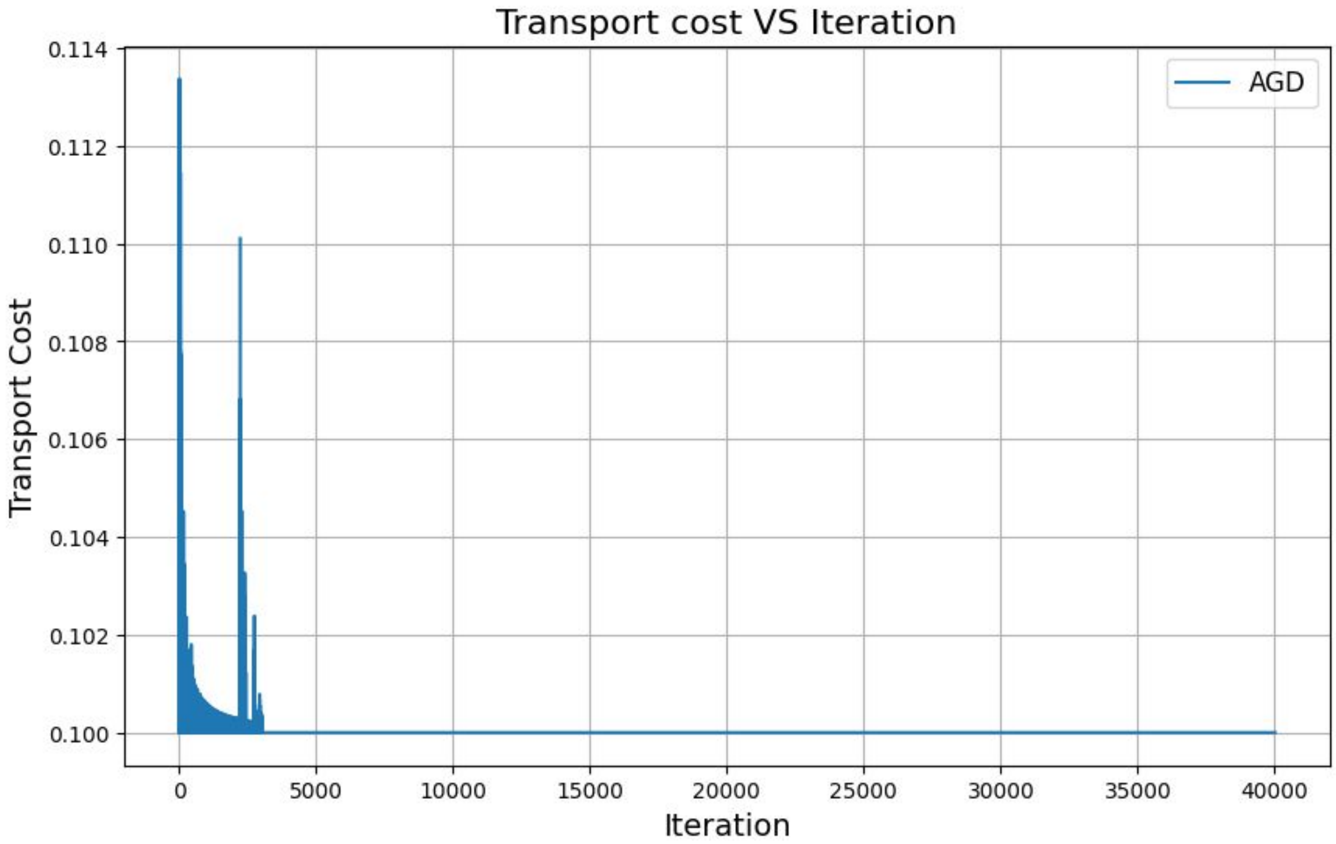
```
Transport Matrix (pi):
 [[4.00000000e-01 0.00000000e+00 0.00000000e+00]
 [9.99999998e-02 2.00000000e-01 1.68300773e-10]
 [1.68300773e-10 0.00000000e+00 3.00000000e-01]]
Final transport cost: 0.10000000016830074
Number of iterations: 19998

Correspondence of pi to p: True
Correspondence of pi to q: True
```

Experiment 2

Regularization param: $\varepsilon = 0.0001$

Cost matrix 3x3



Experiment 3

Regularization param: $\varepsilon = 0.5$

Cost matrix 3x3

Sinkhorn

```
Transport Matrix (pi):
 [[0.35993932 0.01348355 0.02657745]
 [0.08354499 0.17087298 0.045582  ]
 [0.0565157  0.01564347 0.22784055]]
Final transport cost: 0.24134715672460128
Converged in 16 iterations
```

Sinkhorn–Knopp

```
Transport Matrix (pi):
 [[0.35993919 0.01348354 0.02657741]
 [0.08354504 0.17087298 0.04558197]
 [0.05651578 0.01564348 0.22784061]]
Final transport cost: 0.24134721923914415
Converged in 17 iterations
```

Greenkhorn

```
Transport Matrix (pi):
 [[0.35993907 0.01348349 0.02657744]
 [0.0835452  0.17087268 0.04558213]
 [0.05651575 0.01564342 0.22784083]]
Final transport cost: 0.24134741608259402
Converged in 58 iterations
```

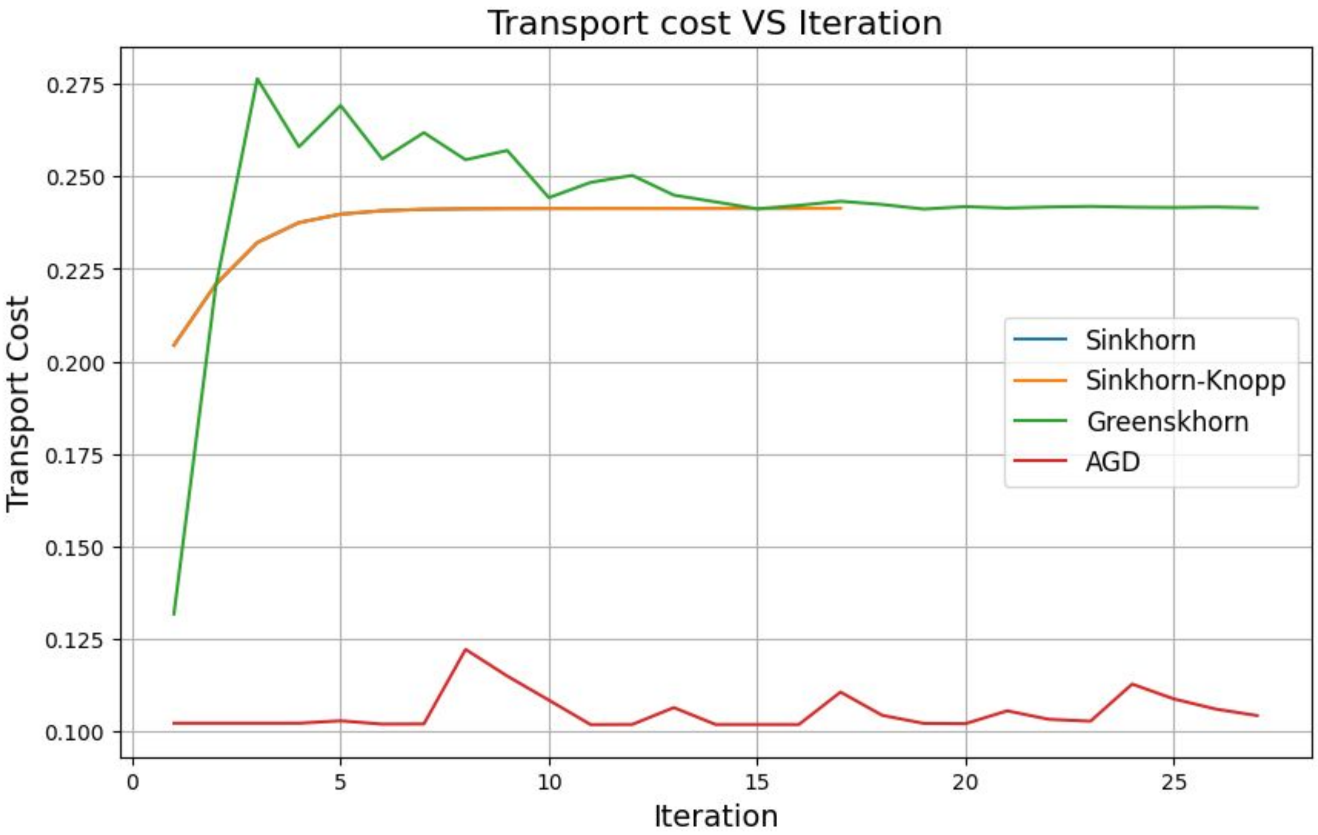
AGD

```
Transport Matrix (pi):
 [[3.99795590e-01 5.75413775e-05 1.46868451e-04]
 [9.64077774e-02 1.99841909e-01 3.75031374e-03]
 [3.79663241e-03 1.00549783e-04 2.96102818e-01]]
Final transport cost: 0.10425968317895995
Number of iterations: 11
```

Experiment 3

Regularization param: $\varepsilon = 0.5$

Cost matrix 3x3



Experiments 4-6

Cost matrix 100x100

results are similar to 3x3 case

Experiment 7

Cost matrix 10000x10000

Sinkhorn

```
Transport Matrix (pi):
 [[1.05030954e-07 1.13011698e-11 2.68083992e-11 ... 1.10078039e-09
  2.82950427e-10 9.85805909e-10]
 [2.17214178e-10 1.31915397e-08 1.02598764e-08 ... 3.39939821e-08
  7.54519575e-09 1.01674458e-08]
 [1.76926978e-10 3.52290178e-09 1.00948370e-07 ... 5.34546086e-08
  1.52530808e-08 4.50454270e-09]
 ...
 [1.83815262e-10 2.95337199e-10 1.35251739e-09 ... 3.70415803e-08
  6.12198003e-11 2.13558473e-11]
 [6.32324668e-11 8.77274510e-11 5.16492671e-10 ... 8.19295929e-11
  1.21849347e-07 1.41227079e-11]
 [4.38800324e-09 2.35462970e-09 3.03810691e-09 ... 5.69261021e-10
  2.81296111e-10 1.42827218e-08]]
Final transport cost: 0.1928657204237296
Converged in 3 iterations
```

```
Correspondence of pi to p: True
Correspondence of pi to q: True
```

Experiment 8:

OT applied to color transportation in CV

Input:

Image 1

Image 2

Experiment 8:

OT applied to color transportation in CV

Output obtained with Sinkhorn:

SinkhornTransport

Conclusions

What have been done:

Within this project, I have:

- studied the concepts of OT and dual tasks with and without entropic regularization.
- considered the algorithms for OT with entropic regularization: Sinkhorn, Sinkhorn–Knopp, Greenkhorn, AGD, AAM.
- implemented in Python: Sinkhorn, Sinkhorn–Knopp, Greenkhorn, AGD algorithms.
- coded experiments for these algorithms in the case of $\mathbb{R}^3$, $\mathbb{R}^{100}$, and $\mathbb{R}^{10000}$ and $\varepsilon = 0.1, 0.0001, 0.5$.
- studied and coded application of OT Sinkhorn algorithm to color transfer task.

Conclusions

Why is it convenient to consider dual task in OT?

+ The dual formulation of OT has already made numerical algorithms (for more general tasks in linear programming).

+ As discussed above, the dual solution provides a tight lower bound for the primal problem.

Conclusions

Advantages and drawbacks of entropic regularization

The advantages of entropic regularization and ideas behind it:

+ As ε increases, the optimal coupling becomes less and less sparse (in the sense of having entries larger than a prescribed threshold), which in turn has the effect of both accelerating computational algorithms (as we see, for example, in Sikhorn's algorithm) and leading to faster convergence.

+ The idea to regularize the optimal transport problem by an entropic term can be traced back to modeling ideas in transportation theory [Wilson, 1969]: Actual traffic patterns in a network do not agree with those predicted by the solution of the optimal transport problem. Indeed, the former are more diffuse than the latter, which tend to rely on a few routes as a result of the sparsity of optimal couplings for (4).

+ Adding entropy prevents numerical instability and ill-conditioning, which can occur in the unregularized OT problem.

+ Entropic regularization allows the use of more efficient algorithms, like Sinkhorn's algorithm, which can be faster and more scalable than direct methods.

Conclusions

Advantages and drawbacks of entropic regularization

The drawbacks of entropic regularization in OT:

- The solution with entropy regularization is only an approximation of the true OT solution. As the regularization parameter ε decreases, the solution becomes closer to the true OT solution, but this requires more computation.

- The performance depends on the choice of the regularization parameter ε, which must be tuned. A poorly chosen parameter can lead to suboptimal performance.

References

[1] Pavel Dvurechensky, Alexander Gasnikov, and Alexey Kroshnin. Computational optimal transport: Complexity by accelerated gradient descent is better than by sinkhorn's algorithm. *arXiv preprint arXiv:1802.04367*, 2018.

[2] Sergey Guminov, Pavel Dvurechensky, Nazarii Tupitsa, and Alexander Gasnikov. On a combination of alternating minimization and nesterov's momentum. 2021.

[3] Alexey Kroshnin, Darina Dvinskikh, Nazarii Tupitsa, Pavel Dvurechensky, Alexander Gasnikov, and Cesar Uribe. On the complexity of approximating wasserstein barycenters. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97, pages 3530–3540, 2019. arXiv:1901.08686.

[4] Tianyi Lin, Nhat Ho, and Michael Jordan. On efficient optimal transport: An analysis of greedy and accelerated mirror descent algorithms. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97, pages 3982–3991. PMLR, 2019.

References

[5] Jianzhou Luo, Dingchuan Yang, and Ke Wei. Improved complexity analysis of the sinkhorn and greenkhorn algorithms for optimal transport. *arXiv preprint arXiv:2305.14939*, 2023.

[6] Jose Rafael Espinosa Mena. On the convergence of the sinkhorn-knopp algorithm with sparse cost matrices. *arXiv preprint arXiv:2405.20528v3 [math.OC]*, 2024.

[7] Arkadi Nemirovski and Uriel Rothblum. On complexity of matrix scaling. *Linear Algebra and its Applications*, 302-303:435–460, 1999.

[8] Gabriel Peyré and Marco Cuturi. Computational optimal transport. *arXiv preprint arXiv:1803.00567*, 2020.

[9] Nazarii Tupitsa, Pavel Dvurechensky, Darina Dvinskikh, and Alexander Gasnikov. Numerical methods for large-scale optimal transport. *arXiv preprint arXiv:2210.11368*, 2022.

Thank you!