Computational Optimal Transport

Basic Properties and Sinkhorn's Algorithm

WU Haoyu
Supervisor: CAI Jianfeng

HKUST Math Department

SCIE2500 Presentation, 12 May 2022

Table of Contents

1. Definition of Optimal Transport
   - Discrete Case
   - Continuous Case
2. Entropic Regularization
3. Sinkhorn's Algorithm
   - Lagrangian Method
   - Sinkhorn's Algorithm
   - Hilbert Projective Metric
   - Convergence of Sinkhorn's algorithm
4. Complexity and Application to Solve Original Problem

Contents

1. Definition of Optimal Transport
   - Discrete Case
   - Continuous Case
2. Entropic Regularization
3. Sinkhorn's Algorithm
   - Lagrangian Method
   - Sinkhorn's Algorithm
   - Hilbert Projective Metric
   - Convergence of Sinkhorn's algorithm
4. Complexity and Application to Solve Original Problem

Definition of Optimal Transport

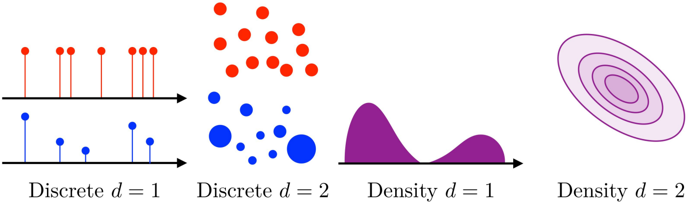

Figure: Mass Distrebution

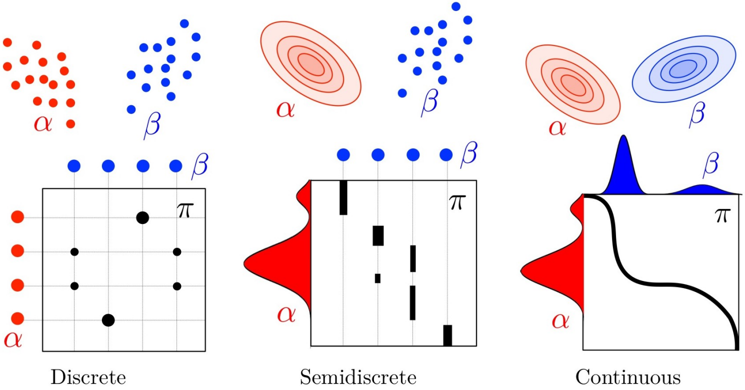

Figure: Coupling Method

Discrete Case

Using vector to denote the mass distrebution

$$\mathrm{a} \in \mathbb{R}^m, \sum_{k=1}^{m} a_k = 1, a_k \geq 0, \quad \mathrm{b} \in \mathbb{R}^n, \sum_{k=1}^{n} b_k = 1, b_k \geq 0$$

Then the coupling matrix set:

$$\mathrm{U}(\mathrm{a}, \mathrm{b}) = \{\mathrm{P} \in \mathbb{R}_+^{n \times m} : \mathrm{P}\mathbb{1}_m = a, \mathrm{P}^T \mathbb{1}_n = \mathrm{b}\}$$

with given Cost Matrix $\mathrm{C} \in \mathbb{R}_+^{n \times m}$, where C_{ij} represents the cost of 1 mass transports from j^{th} location (of a) to i^{th} location (of b).
Then optimal transport problem is:

$$L_{\mathrm{C}}(\mathrm{a}, \mathrm{b}) \overset{\text{def.}}{=} \min_{\mathrm{P} \in \mathrm{U}(\mathrm{a},\mathrm{b})} \langle \mathrm{C}, \mathrm{P} \rangle \overset{\text{def.}}{=} \sum_{i,j} C_{i,j} P_{i,j}$$

Continuous Case

Let $(X, \mathcal{M}_\alpha, \alpha), (X, \mathcal{M}_\beta, \beta)$ be measure space satisfying

$$\int_X d\alpha = \int_X d\beta = 1$$

For any $A \in \mathcal{M}_\alpha$, $\alpha(A)$ represents the mass located in the set A (at the beginning), similar for $\beta(B)$. Then any coupling method could be represented by a product measure

$$\mu : \mathcal{M}_\alpha \times \mathcal{M}_\beta \to \mathbb{R}^+ \cup \{0\}$$

where input (A, B), output the mass transport from A to B.

Noticed

$$\mu(A, B) \leq \min\{\alpha(A), \beta(B)\}, \quad \mu(A, X) = \alpha(A), \mu(X, B) = \beta(B)$$

And the cost of μ is

$$\int_X^2 C(x, y) d\mu$$

where $C : X^2 \to \mathbb{R}^+ \cup \{0\}$ is the cost function, and the optimal problem is

$$\mathrm{L_C}(\alpha, \beta) = \min_\mu \int_X^2 C(x, y) d\mu$$

where μ is a product meausure on X^2 satisfying

$$\mu(A, X) = \alpha(A), \mu(X, B) = \beta(B)$$

Then we just focus on discrete case.

Contents

1 Definition of Optimal Transport
- Discrete Case
- Continuous Case

2 Entropic Regularization

3 Sinkhorn's Algorithm
- Lagrangian Method
- Sinkhorn's Algorithm
- Hilbert Projective Metric
- Convergence of Sinkhorn's algorithm

4 Complexity and Application to Solve Original Problem

Entropic Regularizatione

Definition(Entropic)

For a coupling matrix P, the discrete entropy is defined as:

$$\mathrm{H}(\mathrm{P}) \stackrel{\text{def.}}{=} \sum_{i,j} \mathrm{P}_{i,j}(1 - \log(\mathrm{P}_{i,j}))$$

Regularization

it's always greater or equal 0 since $\mathrm{P}_{i,j} \in [0,1]$. The function $-H$ is 1-strongly convex since by computing the Hessian,

$$\partial^2 - H(P) = \mathrm{diag}(1/\mathrm{P}_{i,j})$$

and $\mathrm{P}_{i,j} \leq 1$.

Regularization

The idea of the entropic regularization is to make $-H$ the regularizing function ($L_{\mathrm{C}}^{\varepsilon}(\mathrm{a},\mathrm{b})$) to approch solutions or approximation of $L_{\mathrm{C}}(\mathrm{a},\mathrm{b})$, the original problem:

$$L_{\mathrm{C}}^{\varepsilon}(\mathrm{a},\mathrm{b}) \stackrel{\text{def.}}{=} \min_{\mathrm{P}\in\mathrm{U(a,b)}} \langle \mathrm{P},\mathrm{C}\rangle - \varepsilon \mathrm{H}(\mathrm{P})$$

Due to it is ε-strongly convex function, then $L_{\mathrm{C}}^{\varepsilon}(\mathrm{a},\mathrm{b})$ has a unique optimal solution.

For every different $\varepsilon > 0$, the solution P_{ε} of $L_{\mathrm{C}}^{\varepsilon}(\mathrm{a},\mathrm{b})$ is unique due to convexity. Then we will claim P_{ε} converges to the original optimal solution with maximal entropic, exactly:

$$\mathrm{P}_{\varepsilon} \longrightarrow \arg\min_{\mathrm{P}\in\mathrm{U(a,b)}} \{-\mathrm{H}(\mathrm{P}) : \langle \mathrm{P},\mathrm{C}\rangle = L_{\mathrm{C}}(\mathrm{a},\mathrm{b})\}, \text{ as } \varepsilon \rightarrow 0$$

In particular,

$$L_{\mathrm{C}}^{\varepsilon}(\mathrm{a},\mathrm{b}) \longrightarrow L_{\mathrm{C}}(\mathrm{a},\mathrm{b}), \text{ as } \varepsilon \rightarrow 0$$

Proof of Convergence with ε

Proof.

Consider a sequnce $(\varepsilon_l)_l$ s.t. $\varepsilon_l \downarrow 0$. Denote P_l the solution of $L_{\mathrm{C}}^{\varepsilon_l}(\mathrm{a}, \mathrm{b})$. Noticed $\mathrm{U}(\mathrm{a}, \mathrm{b})$ is bounded, by Bazano-Weierstrass Theorem, there is a subsequnce such that $\mathrm{P}_k \to \mathrm{P}^*$, and $\mathrm{P}^* \in \mathrm{U}(\mathrm{a}, \mathrm{b})$ because $\mathrm{U}(\mathrm{a}, \mathrm{b})$ is closed. Consider any P as the solution of $L_{\mathrm{C}}(\mathrm{a}, \mathrm{b})$, due to optimality:

$$\langle \mathrm{C}, \mathrm{P} \rangle \leq \langle \mathrm{C}, \mathrm{P}_k \rangle, \quad \langle \mathrm{C}, \mathrm{P}_k \rangle + \varepsilon_k \mathrm{H}(\mathrm{P}) \leq \langle \mathrm{C}, \mathrm{P}_k \rangle + \varepsilon_k \mathrm{H}(\mathrm{P})$$

$$\Longrightarrow 0 \leq \langle \mathrm{C}, \mathrm{P}_k \rangle - \langle \mathrm{C}, \mathrm{P} \rangle \leq \varepsilon_k (\mathrm{H}(\mathrm{P}_k) - \mathrm{H}(\mathrm{P}))$$

Noticed H is continuous, $\langle \mathrm{C}, \mathrm{P}^* \rangle = \langle \mathrm{C}, \mathrm{P} \rangle$ as $k \to \infty$. and $\mathrm{H}(\mathrm{P}) \leq \mathrm{H}(\mathrm{P}^*)$, which means that P^* is a solution in the set of all optimal solutions of $L_{\mathrm{C}}(\mathrm{a}, \mathrm{b})$ with maximal entropy. By strictly convexity, the solution is unique and just P^*.

$\square$

Contents

1. Definition of Optimal Transport
   - Discrete Case
   - Continuous Case
2. Entropic Regularization
3. Sinkhorn's Algorithm
   - Lagrangian Method
   - Sinkhorn's Algorithm
   - Hilbert Projective Metric
   - Convergence of Sinkhorn's algorithm
4. Complexity and Application to Solve Original Problem

Lagrangian Method

We use Lagrangian to show that the unique solution of $L_{\mathrm{C}}^{\varepsilon}(\mathrm{a},\mathrm{b})$ has specific form.

Proposition

There exist scaling variable $\mathrm{u} \in \mathbb{R}_{+}^{n}, \mathrm{v} \in \mathbb{R}_{+}^{m}$ such that the solution P of $L_{\mathrm{C}}^{\varepsilon}(\mathrm{a},\mathrm{b})$ satisfying:

$$\mathrm{P}_{i,j} = \mathrm{u}_i \mathrm{K}_{i,j} \mathrm{v}_j, \quad 1 \leq i \leq n, 1 \leq j \leq m, i,j \in \mathbb{N}$$

where $\mathrm{K} \in \mathbb{R}^{n \times m}$ with $\mathrm{K}_{i,j} = e^{-\frac{\mathrm{C}_{i,j}}{\varepsilon}}$.

Proof of Proposition

Proof.

Consider the Lagrangian of $L_C^\varepsilon(a, b)$ with variable $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$ for all marginal constraints:

$$L(P, x, y) = \langle C, P \rangle - \varepsilon H(P) - \langle x, P\mathbb{1}_m - a \rangle - \langle y, P^T\mathbb{1}_n - b \rangle$$

By the first order condition,

$$\frac{\partial L(P, x, y)}{\partial P_{i,j}} = C_{i,j} + \varepsilon \log(P_{i,j}) - x_i - y_j = 0$$

$$\Longrightarrow P_{i,j} = e^{\frac{x_i}{\varepsilon}} e^{\frac{-C_{i,j}}{\varepsilon}} e^{\frac{y_j}{\varepsilon}}, \quad (u)_i = e^{\frac{x_i}{\varepsilon}}, (v)_j = e^{\frac{y_j}{\varepsilon}}$$

□

Sinkhorn's Algorithm

Noticed for the corrosponding (u, v), $\mathrm{P} = \mathrm{diag}(\mathrm{u})\mathrm{K}\mathrm{diag}(\mathrm{v})$. Therefore the variable (u, v) must satisfy the restriction of $\mathrm{U}(\mathrm{a}, \mathrm{b})$:

$$\mathrm{u} \odot (\mathrm{Kv}) = \mathrm{a}, \mathrm{v} \odot (\mathrm{K}^T \mathrm{u}) = \mathrm{b}$$

where $\odot$ is the vector product under entrywise $((\mathrm{a} \odot \mathrm{b})_i = a_i b_i)$.
The naturally way to find (u, v) is solve them iteratively, and the Sinkhorn's algorithm is defined by the idea.

Definition(Sinkhorn's Algorithm)

$$\mathrm{u}^{(k+1)} \stackrel{\text{def.}}{=} \frac{\mathrm{a}}{\mathrm{Kv}^{(k)}}, \quad \mathrm{v}^{(k+1)} \stackrel{\text{def.}}{=} \frac{\mathrm{b}}{\mathrm{Ku}^{(k+1)}}$$

where the division is also entrywise and initialized with an arbitrary positive vector $\mathrm{v}^{(0)} = \mathbb{1}_m$.

Remark

Noticed that if Sinkhorn's algorithm converges for any initialization(will be proced latet), different initialization may obtain different limit up to a multiplicative constant since if (u, v) satisfies the $\mathrm{U}(\mathrm{a}, \mathrm{b})$ restriction, then $(\lambda \mathrm{u}, \lambda^{-1} \mathrm{v})$ also.

Then we prove the global convergence of Sinkhorn's algorithm. It's first proved by [Franklin:1989], using the property of Hilbert projective metric introduced from [Birkhoff:1957].

Hilbert Projective Metric

Definition

The Hilbert metric is defined on $\mathbb{R}_+^n$:

$$\forall (\mathrm{u}, \mathrm{u}') \in (\mathbb{R}_+^n)^2, \quad d_H(\mathrm{u}, \mathrm{u}') \stackrel{\text{def.}}{=} \log \max_{i,j} \frac{\mathrm{u}_i \mathrm{u}'_j}{\mathrm{u}_j \mathrm{u}'_i}$$

It's actually a distance on the projective cone $\mathbb{R}_+^n / \sim$ with $\mathrm{u} \sim \mathrm{v}$ iff $\exists r > 0, \mathrm{v} = r\mathrm{u}$.

Important Property

For $\mathrm{K} \in \mathbb{R}_+^{n \times m}$, $(\mathrm{u}, \mathrm{u}') \in (\mathbb{R}_+^m)^2$:

$$d_H(\mathrm{Ku}, \mathrm{Ku}') \leq \lambda(\mathrm{K}) d_H(\mathrm{u}, \mathrm{u}')$$

where $\lambda(\mathrm{K}) = \frac{\sqrt{\mu(\mathrm{K})}-1}{\sqrt{\mu(\mathrm{K})}+1} < 1, \mu(\mathrm{K}) = \max_{i,j,k,l} \frac{\mathrm{K}_{i,j}\mathrm{K}_{k,l}}{\mathrm{K}_{i,l}\mathrm{K}_{j,k}}$.

Important Property

For $\mathrm{K} \in \mathbb{R}_+^{n\times m}$, $(\mathrm{u},\mathrm{u}') \in (\mathbb{R}_+^m)^2$:

$$d_H(\mathrm{Ku},\mathrm{Ku}') \leq \lambda(\mathrm{K}) d_H(\mathrm{u},\mathrm{u}'), \text{where } \begin{cases} \lambda(\mathrm{K}) = \frac{\sqrt{\mu(\mathrm{K})}-1}{\sqrt{\mu(\mathrm{K})}+1} < 1 \\ \mu(\mathrm{K}) = \max_{i,j,k,l} \frac{\mathrm{K}_{i,j}\mathrm{K}_{k,l}}{\mathrm{K}_{i,l}\mathrm{K}_{j,k}} \end{cases}$$

Remark

The inequality is also used to prove the Perron-Frobenius theorem, known as a matrix cases of contraction mapping theorem.

Remark

The proof is basically first shown on $\mathbb{R}_+^2$ then generalize to common case.

Convergence of Sinkhorn's algorithm

We first has the unique solution P^* of $L_{\mathrm{C}}^{\varepsilon}(\mathrm{a},\mathrm{b})$, by [Franklin;1989], there exist $(\mathrm{u}^*,\mathrm{v}^*) \in \mathbb{R}^m \times \mathbb{R}^n$ such that:

$$\mathrm{P}^* = \mathrm{diag}(\mathrm{u}^*)\mathrm{K}\mathrm{diag}(\mathrm{v}^*), \quad \mathrm{u}^* \odot (\mathrm{K}\mathrm{v}^*) = \mathrm{a}, \mathrm{v}^* \odot (\mathrm{K}^T\mathrm{u}^*) = \mathrm{b}$$

Theorem

We have $(\mathrm{u}^{(k)},\mathrm{v}^{(k)}) \to (\mathrm{u}^*,\mathrm{v}^*)$ under Sinkhorn's algorithm, and:

$$d_H(\mathrm{u}^{(k)},\mathrm{u}^*) = O(\lambda(\mathrm{K})^{2k}), \quad d_H(\mathrm{v}^{(k)},\mathrm{v}^*) = O(\lambda(\mathrm{K})^{2k}) \tag{1}$$

$$d_H(\mathrm{u}^{(k)},\mathrm{u}^*) \le \frac{d_H(\mathrm{P}^{(k)}\mathbb{1}_m,\mathrm{a})}{1-\lambda(\mathrm{K})^2}, \quad d_H(\mathrm{v}^{(k)},\mathrm{v}^*) \le \frac{d_H(\mathrm{P}^{(k),T}\mathbb{1}_n,\mathrm{b})}{1-\lambda(\mathrm{K})^2} \tag{2}$$

$$||\log(\mathrm{P}^{(k)}) - \log(\mathrm{P}^*)||_\infty \le d_H(\mathrm{u}^{(k)},\mathrm{u}^*) + d_H(\mathrm{v}^{(k)},\mathrm{v}^*) \tag{3}$$

where $\mathrm{P}^{(k)} \stackrel{\text{def.}}{=} \mathrm{diag}(\mathrm{u}^k)\mathrm{K}\mathrm{diag}(\mathrm{v}^k)$.

Convergence of Sinkhorn's Algorithm

Proof of (1)

Noticed by definition of Hilbert Metric,

$$\forall (\mathrm{u}, \mathrm{u}') \in (\mathbb{R}_+^m)^2, \quad d_H(\mathrm{u}, \mathrm{u}') = d_H(\mathrm{u}/\mathrm{u}', \mathbb{1}_m) = d_H(\frac{\mathbb{1}_m}{\mathrm{u}}, \frac{\mathbb{1}_m}{\mathrm{u}'})$$

That means,

$$\begin{aligned} d_H(\mathrm{u}^{(k+1)}, \mathrm{u}^*) &= d_H\left(\frac{\mathrm{a}}{\mathrm{Kv}^{(k)}}, \frac{\mathrm{a}}{\mathrm{Kv}^*}\right) = d_H\left(\frac{\mathbb{1}_m}{\mathrm{Kv}^{(k)}}, \frac{\mathbb{1}_m}{\mathrm{Kv}^*}\right) \\ &= d_H(\mathrm{Kv}^{(k)}, \mathrm{Kv}^*) \leq \lambda(\mathrm{K}) d_H(\mathrm{v}^{(k)}, \mathrm{v}^*) \end{aligned}$$

the last step used inequality proved above, and mutatis mutandis,

$$d_H(\mathrm{v}^{(k)}, \mathrm{v}^*) \leq \lambda(\mathrm{K}) d_H(\mathrm{u}^{(k)}, \mathrm{u}^*) \Longrightarrow d_H(\mathrm{u}^{(k+1)}, \mathrm{u}^*) \leq \lambda^2(\mathrm{K}) d_H(\mathrm{u}^{(k)}, \mathrm{u}^*)$$

which proves (1).

Convergence of Sinkhorn's Algorithm

Proof of (2)

Consider Triangular inequality,

$$
\begin{aligned}
d_H(\mathrm{u}^{(k)}, \mathrm{u}^*) &\le d_H(\mathrm{u}^{(k+1)}, \mathrm{u}^{(k)}) + d_H(\mathrm{u}^{(k+1)}, \mathrm{u}^*) \\
&\le d_H\left(\frac{\mathrm{a}}{\mathrm{K}\mathrm{v}^{(k)}}, \mathrm{u}^{(l)}\right) + \lambda^2(\mathrm{K}) d_H(\mathrm{u}^{(k)}, \mathrm{u}^*) \\
&= d_H\left(\mathrm{a}, \mathrm{u}^{(k)} \odot \mathrm{K}\mathrm{v}^{(k)}\right) + \lambda^2(\mathrm{K}) d_H(\mathrm{u}^{(k)}, \mathrm{u}^*) \\
&= d_H\left(\mathrm{a}, \mathrm{P}^{(k)} \mathbb{1}_m\right) + \lambda^2(\mathrm{K}) d_H(\mathrm{u}^{(k)}, \mathrm{u}^*)
\end{aligned}
$$

$$
\Longrightarrow d_H(\mathrm{u}^{(k)}, \mathrm{u}^*) \le \frac{d_H(\mathrm{P}^{(k)} \mathbb{1}_m, \mathrm{a})}{1 - \lambda(\mathrm{K})^2}
$$

This proves the first part of (2), the latter part is similar.

Convergence of Sinkhorn's Algorithm

Proff of (3)

Denote $M_k = \exp\left(d_H(\mathrm{u}^{(k)}, \mathrm{u}^*) + d_H(\mathrm{v}^{(k)}, \mathrm{v}^*)\right) > 1$. Noticed $\mathrm{P}^{(k)} = \mathrm{diag}(\mathrm{u}^k)\mathrm{K}\mathrm{diag}(\mathrm{v}^k)$, $\mathrm{P}^{(k)} = \mathrm{diag}(\mathrm{u}^*)\mathrm{K}\mathrm{diag}(\mathrm{v}^*)$:

$$\mathrm{P}^* = \mathrm{diag}(\mathrm{u}^*/\mathrm{u}^{(k)})\mathrm{P}^{(k)}\mathrm{diag}(\mathrm{v}^*/\mathrm{v}^{(k)})$$

$$d_H(\mathrm{u}^*, \mathrm{u}^{(k)}) = d_H(\mathrm{u}^*/\mathrm{u}^{(k)}, \mathbb{1}_m), d_H(\mathrm{u}^*, \mathrm{u}^{(k)}) = d_H(\mathrm{v}^*/\mathrm{v}^{(k)}, \mathbb{1}_n)$$

Let $(\mathrm{u}^*/\mathrm{u}^{(k)})_i$ be normalized by dividing the smallest entry among $(\mathrm{u}^*/\mathrm{u}^{(k)})_i$ to obtain $\left(\mathrm{u}^*/\mathrm{u}^{(k)}\right)'$.

$$1 \leq \left(\mathrm{u}^*/\mathrm{u}^{(k)}\right)'_i \leq M_k, \forall i$$

And times the value divided to $\left(\mathrm{v}^*/\mathrm{v}^{(k)}\right)$ to obtain $\left(\mathrm{v}^*/\mathrm{v}^{(k)}\right)'$.

Convergence of Sinkhorn's Algorithm

Proof of (3) (Continuous)

By definition of $\mathrm{P}^*, \mathrm{P}^{(k)}$,

$$\mathrm{P}^{(k),T}\mathbb{1}_n = \mathrm{P}^{*,T}\mathbb{1}_m = \mathrm{b}$$

By the diagnoalization:

$$\mathrm{diag}\left(\mathrm{v}^*/\mathrm{v}^{(k)}\right)'\mathrm{P}^{(k),T} = \mathrm{P}^{*,T}\mathrm{diag}\left(\mathrm{u}^*/\mathrm{u}^{(k)}\right)'^{-1}$$

By the first equality,

$$\mathrm{diag}\left(\mathrm{v}^*/\mathrm{v}^{(k)}\right)'\mathrm{b} = \mathrm{diag}\left(\mathrm{v}^*/\mathrm{v}^{(k)}\right)\mathrm{P}^{(k),T}\mathbb{1}_m$$

By the second equality,

$$\mathrm{diag}\left(\mathrm{v}^*/\mathrm{v}^{(k)}\right)'\mathrm{P}^{(k),T}\mathbb{1}_m = \mathrm{P}^{*,T}\mathrm{diag}\left(\mathrm{u}^*/\mathrm{u}^{(k)}\right)^{-1}\mathbb{1}_m$$

Proof of (3)(Continuous)

Recall the range of $\left(\mathrm{u}^*/\mathrm{u}^{(k)}\right)$:

$$M_k^{-1} \leq \left(\mathrm{u}^*/\mathrm{u}^{(k)}\right)_i^{-1} \leq 1, \forall i$$

Consider "times" $\mathrm{P}^{*,T}$:

$$\begin{aligned} M_k^{-1}\mathrm{b}_j = M_k^{-1}\left(\mathrm{P}^{*,T}\mathbb{1}_m\right)_j &\leq \left(\mathrm{P}^{*,T}\mathrm{diag}\left(\mathrm{u}^*/\mathrm{u}^{(k)}\right)^{-1}\mathbb{1}_m\right)_j \\ &\leq \left(\mathrm{P}^{*,T}\mathbb{1}_m\right)_j = \mathrm{b}_j \end{aligned}$$

By the result from second equality and $\mathrm{P}^{(k),T}\mathbb{1}_m = \mathrm{b}$:

$$M_k^{-1}\mathrm{b}_j \leq \left[\left(\mathrm{v}^*/\mathrm{v}^{(k)}\right)'\mathrm{b}\right]_j \leq \mathrm{b}_j, \forall j \Longrightarrow 1 \leq \left(\mathrm{v}^*/\mathrm{v}^{(k)}\right)_j' \leq M_k$$

Convergence of Sinkhorn's Algorithm

Proof of (3)(Continuous)

$$M_k^{-1} \le \left(\mathrm{u}^*/\mathrm{u}^{(k)}\right)_i' \left(\mathrm{v}^*/\mathrm{v}^{(k)}\right)_j' = \left(\mathrm{u}^*/\mathrm{u}^{(k)}\right)_i \left(\mathrm{v}^*/\mathrm{v}^{(k)}\right)_j = \mathrm{P}_{ij}^*/\mathrm{P}_{ij}^{(k)} \le M_k$$

which can be represented to the form (3).

Contents

1. Definition of Optimal Transport
   - Discrete Case
   - Continuous Case
2. Entropic Regularization
3. Sinkhorn's Algorithm
   - Lagrangian Method
   - Sinkhorn's Algorithm
   - Hilbert Projective Metric
   - Convergence of Sinkhorn's algorithm
4. Complexity and Application to Solve Original Problem

Complexity and Application to Solve Original Problem

It's mainly the result from [Altschuler:2017].

Main Theorem

Sinkhorn's algorithm with a rounding step returns a point $\hat{P} \in \mathrm{U(a, b)}$ satifying

$$\langle \hat{P}, \mathrm{C} \rangle \leq \min_{\mathrm{P} \in \mathrm{U(a,b)}} \langle \mathrm{P}, \mathrm{C} \rangle + \varepsilon$$

in time $O(n^2 L^3 (\log n)\varepsilon^{-3})$, where L in the upper bound of entry of C.

"Conclusion"

As the development of computer science, people can use more stronger method to compute some tradictional question. By using the entropic as the regularization, the orginal optimal problem can be evaluated more easily due the convexity.