Strong Duality and Double Convexification

Presenter: Kaiwen Hou Scriber: Qiuran (Rita) Lyu Observer: Wenxin Zhang

Oct 13, 2024

1 Strong duality of KP formulation

We have discussed the KP formulation of finding the optimal transport plan, and solving the optimization question:

$$\inf_{\pi} \int_{X\times Y} c(x,y)\mathrm{d}\pi, \text{ s.t. } \pi \in \Pi(\mu,\nu), \tag{1}$$

where $\Pi(\mu,\nu)$ is the coupling over $X \times Y$, such that the mass conversation holds.

Before introducing the duality, we start with the discretization formulation of the above problem,

$$\begin{aligned}
&\min_{\pi} \sum_{i,j} \pi_{ij}c_{ij}, \\
&\text{s.t. } \mu = \sum_{i=1}^{n} \mu_i \delta_{x_i}, \ \nu = \sum_{j=1}^{m} \nu_j \delta_{y_j} \\
&\qquad \sum_{i=1}^{n} \pi_{ij} = \nu_j, \ \forall j \\
&\qquad \sum_{j=1}^{m} \pi_{ij} = \mu_i, \ \forall i \\
&\qquad \pi_{ij} \geq 0, \ \forall i,j
\end{aligned}$$

This is a linear programming with linear constraints. By using the dual form of LP, we could have the **dual** problem:

$$\begin{aligned}
&\max_{u,v} \sum_{i=1}^{n} \mu_i u_i + \sum_{j=1}^{m} \nu_j v_j \\
&\text{s.t. } u_i + v_j \leq c_{ij}, \forall i,j.
\end{aligned}$$

Thus, we jump back to the original case and the optimization problem will be formulated as:

$$\sup_{u,v} \left\{ \int_X u(x)\mathrm{d}\mu(x) + \int_Y v(y)\mathrm{d}\nu(y) \quad \text{s.t. } u(x)+v(y) \leq c(x,y), \forall x,y \right\} \tag{2}$$

$$\iff \sup_{u,v} \left\{ \int_{X\times Y} u(x)\mathrm{d}\pi(x,y) + \int_{X\times Y} v(y)\mathrm{d}\pi(x,y) \quad \text{s.t. } u(x)+v(y) \leq c(x,y), \forall x,y \right\}. \tag{3}$$

This equivalence comes from the mass conservation condition, such that

$$\pi[x \times Y] = \mu(x), \quad \pi[X \times y] = \nu(y). \tag{4}$$

From Eq (2), we could see the weak duality, because

$$\sup_{u,v} \left\{ \int_{X \times Y} u(x) \mathrm{d}\pi(x,y) + \int_{X \times Y} v(y) \mathrm{d}\pi(x,y) \quad \text{s.t. } u(x) + v(y) \leq c(x,y), \forall x, y \right\} \leq \inf_{\pi} \int_{X \times Y} c(x,y) \mathrm{d}\pi(x,y).$$

We then prove the strong duality by first relaxing the constrained optimization problem to the unconstrained optimization problem by imposing the indicator function and the infinite penalty parameter.

$$\inf_{\pi \in \Pi} \int c(x,y) \mathrm{d}\pi = \inf_{\pi} \left\{ \int c(x,y) \mathrm{d}\pi(x,y) + \lim_{\lambda \to \infty} \lambda(1 - \mathbb{I}(\pi \in \Pi)) \right\}. \tag{5}$$

For the second term, if the chosen π is in the constrained space Π, its value is 0, while its value is ∞ when π does not belong to Π. We then show

$$\text{Eq (5)} = \inf_{\pi} \left\{ \sup_{u,v} \left(\int_X u(x) \mathrm{d}\mu(x) + \int_Y v(y) \mathrm{d}\nu(y) - \int_{X \times Y} (u(x) + v(y)) \mathrm{d}\pi(x,y) \right) \right\} \tag{6}$$

This is because in Eq (4), we have shown that if $\pi \in \Pi$, then

$$\int_X u(x) \mathrm{d}\mu(x) + \int_Y v(y) \mathrm{d}\nu(y) - \int_{X \times Y} (u(x) + v(y)) \mathrm{d}\pi(x,y) = 0.$$

If $\pi \notin \Pi$, once we find u and v such that

$$\int_X u(x) \mathrm{d}\mu(x) + \int_Y v(y) \mathrm{d}\nu(y) - \int_{X \times Y} (u(x) + v(y)) \mathrm{d}\pi(x,y) > 0,$$

we could always give a positive constant M and let M go to infinity to make the term become infinity, and the same holds if

$$\int_X u(x) \mathrm{d}\mu(x) + \int_Y v(y) \mathrm{d}\nu(y) - \int_{X \times Y} (u(x) + v(y)) \mathrm{d}\pi(x,y) < 0,$$

where we could choose $-M, M > 0$, and let M go to infinity. As c has nothing to do with $u(x)$ and $v(y)$,

$$\text{Eq (6)} = \inf_{\pi} \sup_{u,v} \left\{ \int u(x) \mathrm{d}\mu(x) + \int v(y) \mathrm{d}\nu(y) + \int (c(x,y) - (u(x) + v(y)) \mathrm{d}\pi(x,y) \right\}. \tag{7}$$

By [Villani, 2021] Section 1.1.6, we could switch the inf and sup from minimax principles. Then

$$\text{Eq (7)} = \sup_{u,v} \inf_{\pi} \left\{ \int u(x) \mathrm{d}\mu(x) + \int v(y) \mathrm{d}\nu(y) + \int (c(x,y) - (u(x) + v(y)) \mathrm{d}\pi(x,y) \right\} \tag{8}$$

$$= \sup_{u,v} \left\{ \int_X u(x) \mathrm{d}\mu(x) + \int_Y v(y) \mathrm{d}\nu(y) + \inf_{\pi} \int_{X \times Y} (c(x,y) - (u(x) + v(y))) \mathrm{d}\pi(x,y) \right\} \tag{9}$$

For the third term, we could see that if $u(x) + v(y) \leq c(x,y)$, then the inf could go to 0, and if there exists $u(x^*) + v(y^*) > c(x^*, y^*)$, we could select $\pi(x,y)$ with positive constant M such that $\pi = M\delta((x^*, y^*))$, and when M goes to infinity, this term will go to $-\infty$.

Using the same trick as in Eq (5), we could see the behavior of the second term is the same as the indicator penalty for the condition $u(x) + v(y) \leq c(x,y)$ in the constrained optimization. Thus, the strong duality holds, where

$$\text{Eq (8)} = \sup_{u,v} \left\{ \int_X u(x) \mathrm{d}\mu(x) + \int_Y v(y) \mathrm{d}\nu(y) \middle| u(x) + v(y) \leq c(x,y) \right\}. \tag{10}$$

Eq (10) transforms the original optimization problem to its dual form and shows the strong duality of such transformation. We successfully jump from the high dimension over the coupling function to the lower dimension over marginal functions.

2 Dual problem refinement

We define two functions

- $\varphi(x) = \frac{1}{2}||x||_2^2 - u(x)$,
- $\psi(y) = \frac{1}{2}||y||_2^2 - v(y)$.

Remark: we use $\frac{1}{2}||x||_2^2$ and $\frac{1}{2}||y||_2^2$ here to construct the Legendre-Fenchel transform (see [Villani, 2021] p.23 or [Legendre, 1787]).
Eq (10) could be further written as

$$\sup_{\substack{\varphi,\psi \\ (\varphi,\psi)\in\Phi(\varphi,\psi)}} \int_X \left(\frac{1}{2}||x||_2^2 - \varphi(x)\right) \mathrm{d}\mu(x) + \int_Y \left(\frac{1}{2}||y||_2^2 - \psi(y)\right) \mathrm{d}\nu(y). \tag{11}$$

Choose the cost function $c(x,y) = \frac{1}{2}||x-y||_2^2$, we know the new feasible region becomes

$$(\varphi(x), \psi(y)) \in \Phi(\varphi,\psi), \text{ s.t. } \varphi(x) + \psi(y) \geq x^\top y.$$

2.1 Legendre-Fenchel transform (convex conjugate)

Before moving forward, we introduce the Legendre-Fenchel transform, for the function $f(x)$, its corresponding Legendre-Fenchel transform is:

$$f^*(y) = \sup_x x^\top y - f(x).$$

The Legendre transform has the following properties:

- For any function f, $f^*(y)$ is convex. This could be proved by definition:

$$\begin{aligned} f^*(\lambda y + (1-\lambda)x) &= \sup_z z^\top(\lambda y + (1-\lambda)x) - f(z) \\ &\leq \sup_z z^\top \lambda y - \lambda f(z) + \sup_z z^\top (1-\lambda)x - (1-\lambda)f(z) \\ &= \lambda f^*(y) + (1-\lambda)f^*(x). \end{aligned}$$

- $f(x) + f^*(y) \geq x^\top y, \forall x, y$. The definition could also prove this:

$$f^*(y) = \sup_x x^\top y - f(x) \geq x^\top y - f(x), \forall x, y.$$

- $f(x) \geq f^{**}(x)$, which could be proved through:

$$\begin{aligned} f^{**}(x) &= \sup_y x^\top y - f^*(x) \\ &= \sup_y \left\{ x^\top y - \sup_z \{z^\top x - f(x)\} \right\} \\ &\leq \sup_y \left\{ x^\top y - \{y^\top x - f(x)\} \right\} = f(x). \end{aligned}$$

- $f(x) = f^{**}(x)$ if and only if $f(x)$ itself is convex and lower semi-continuous by the Fenchel–Moreau theorem, see Fenchel–Moreau theorem or [Borwein and Lewis, 2006]. To prove this, we only need to prove the right-hand side $f(x) \leq f^{**}(x)$. As $f(x)$ itself is convex and lower semi-continuous, $f(x)$ could be represented as:

$$f(x) = \sup_{\alpha\in\mathcal{A}} L^\alpha(x),$$

where L^α are hyper-planes indexed by the index set $\mathcal{A}$, and because hyper-planes have linear forms, we know

$$(L^\alpha)^{**} = L^\alpha.$$

Then, $\forall \alpha$, we know

$$\begin{aligned}
&f(x) \geq L^\alpha(x) \\
&f^*(x) \leq (L^\alpha)^* \\
&f^{**} \geq (L^\alpha)^{**} = L^\alpha \\
&\Rightarrow f^{**} \geq \sup_\alpha L^\alpha(x) = f(x).
\end{aligned}$$

2.2 Reduce to conjugate Pairs

Let us reconsider the dual optimization problem:

$$\begin{aligned}
&\sup_{\substack{\varphi,\psi \\ (\varphi,\psi)\in\Phi(\varphi,\psi)}} \int_X \left(\frac{1}{2}||x||_2^2 - \varphi(x)\right) \mathrm{d}\mu(x) + \int_Y \left(\frac{1}{2}||y||_2^2 - \psi(y)\right) \mathrm{d}\nu(y) \\
&(\varphi(x), \psi(y)) \in \Phi(\varphi, \psi), \text{ s.t. } \varphi(x) + \psi(y) \geq x^\top y.
\end{aligned}$$

This problem is equivalent to

$$\begin{aligned}
&\inf_{\substack{\varphi,\psi \\ (\varphi,\psi)\in\Phi(\varphi,\psi)}} \int_X \varphi(x) \mathrm{d}\mu(x) + \int_Y \psi(y) \mathrm{d}\nu(y) \\
&(\varphi(x), \psi(y)) \in \Phi(\varphi, \psi), \text{ s.t. } \varphi(x) + \psi(y) \geq x^\top y,
\end{aligned}$$

where $\varphi(x) = \frac{1}{2}||x||_2^2 - u(x)$ and $\psi(y) = \frac{1}{2}||y||_2^2 - v(y)$.

First we need to check the feasible set is non empty. We could easily use $\varphi(x) = \frac{1}{2}||x||_2^2$ and $\psi(y) = \frac{1}{2}||y||_2^2$. By the basis inequality, we know the feasible set is non-empty.

Second, we would like to reduce the pair $(\varphi(x), \psi(y))$ to the pair of one function with its convex conjugate. For pair (φ, ψ), we define

$$\inf_{\substack{\varphi,\psi \\ (\varphi,\psi)\in\Phi(\varphi,\psi)}} \int_X \varphi(x) \mathrm{d}\mu(x) + \int_Y \psi(y) \mathrm{d}\nu(y) = J[\varphi, \psi].$$

We could prove if $(\varphi, \psi) \in \Phi$, then $(\varphi^*, \varphi^{**}) \in \Phi$, and $J[\varphi^*, \varphi^{**}] \leq J[\varphi, \psi]$. Thus, we could limit our searching space to all convex conjugate pairs because if we have solution pair (φ', ψ'), we could always find the convex conjugate pair $(\varphi'^*, \varphi'^{**})$ to achieve smaller values.

To prove this argument, we first show $\varphi^* + \varphi^{**} \geq x^\top y, \forall x, y$. This is easily satisfied by the second property of Legendre-Fenchel transform. We also know $\varphi(x) \geq \varphi^{**}(x)$. Thus, we only need to show

$$\psi(y) \geq \varphi^*(y).$$

This could be proved as follows:

$$\begin{aligned}
&(\varphi, \psi) \in \Phi \Rightarrow \varphi(x) + \psi(y) \geq x^\top y, \forall x, y \\
&\psi(y) \geq x^\top y - \varphi(x) \\
&\Rightarrow \psi(y) \geq \sup_x \{x^\top y - \varphi(x)\} = \varphi(y)^*.
\end{aligned}$$

Our problem is reduced to convex conjugate pairs, that is,

$$\inf_{\varphi^*,\varphi^{**}} \int_X \varphi^{**}(x)\mathrm{d}\mu(x) + \int_Y \varphi^*(y)\mathrm{d}\nu(y). \tag{12}$$

Now, we successfully translate the original constrained optimization problem Eq (1) to unconstrained Eq (12), an optimization problem with only one function. This trick is known as "Double Convexification".

References

[Borwein and Lewis, 2006] Borwein, J. and Lewis, A. (2006). *Convex Analysis*. Springer.

[Legendre, 1787] Legendre, A. M. (1787). *Mémoire sur l'intégration de quelques équations aux différences partielles*. Imprimerie royale.

[Villani, 2021] Villani, C. (2021). *Topics in optimal transportation*, volume 58. American Mathematical Soc.