最適輸送問題の定式化

IPS 専攻 最適輸送本 輪読会 #2

髙橋 優輝

2023.2.23

システム数理学講座 B4

目次

2.1 線形計画による定式化

2.1 線形計画による定式化

2.1.1 点群の比較

点群の比較-問題設定-

まず，空間 $\mathcal{X}$ 上で定義された重み付き点群

$$\alpha = \sum_{i=1}^{n} a_i \delta_{x_i}, \quad \beta = \sum_{j=1}^{m} b_j \delta_{y_j} \tag{1}$$

の比較を考える．

- $\alpha \in \mathcal{P}(\mathcal{X})$
  $\rightarrow$ 点 x_i を確率 a_i でとる確率分布で，$\boldsymbol{a} \in \Sigma_n$.
- $\beta \in \mathcal{P}(\mathcal{X})$ も同様 ($\boldsymbol{b} \in \Sigma_m$)
- $C(x, y) \colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$
  $\rightarrow$ 点 $x \in \mathcal{X}$ から点 $y \in \mathcal{X}$ に 1 単位の質量を輸送するのにかかるコスト；**コスト関数**

点群の比較-定式化-

問題 2.1(点群の場合の最適輸送問題)

$$
\begin{aligned}
\underset{\boldsymbol{P}\in\mathbb{R}^{n\times m}}{\text{minimize}} \quad & \sum_{i=1}^{n}\sum_{j=1}^{m} C(x_i, y_j) P_{ij} \\
\text{subject to} \quad & P_{ij} \geq 0 \qquad (\forall i \in [n], \forall j \in [m]) \\
& \sum_{j=1}^{m} P_{ij} = a_i \qquad (\forall i \in [n]) \\
& \sum_{i=1}^{n} P_{ij} = b_j \qquad (\forall j \in [m])
\end{aligned} \tag{2}
$$

- 最適値

  最適輸送コストといい，OT(α, β, C) と表す．
- 最適解

  最適輸送行列あるいは単に**最適輸送**という．

点群の比較-定式化-

問題 2.1(点群の場合の最適輸送問題)

$$
\begin{aligned}
\underset{\boldsymbol{P} \in \mathbb{R}^{n \times m}}{\text{minimize}} \quad & \sum_{i=1}^{n} \sum_{j=1}^{m} C(x_i, y_j) P_{ij} \\
\text{subject to} \quad & P_{ij} \geq 0 \qquad (\forall i \in [n], \forall j \in [m]) \\
& \sum_{j=1}^{m} P_{ij} = a_i \qquad (\forall i \in [n]) \\
& \sum_{i=1}^{n} P_{ij} = b_j \qquad (\forall j \in [m])
\end{aligned} \tag{2}
$$

制約式の意味

- 1 つめ：輸送量は非負
- 2 つめ：点 x_i から輸送される総量は a_i
- 3 つめ：点 y_j に輸送される総量は b_j
- 2 つめ，3 つめ：**質量保存制約**

点群の比較-定式化-

問題 2.1(点群の場合の最適輸送問題)

$$\begin{aligned} \underset{\boldsymbol{P}\in\mathbb{R}^{n\times m}}{\text{minimize}} \quad & \sum_{i=1}^{n}\sum_{j=1}^{m} C(x_i, y_j)P_{ij} \\ \text{subject to} \quad & P_{ij} \geq 0 \qquad (\forall i \in [n], \forall j \in [m]) \\ & \sum_{j=1}^{m} P_{ij} = a_i \qquad (\forall i \in [n]) \\ & \sum_{i=1}^{n} P_{ij} = b_j \qquad (\forall j \in [m]) \end{aligned} \tag{2}$$

質量保存制約は，

$$\begin{aligned} \boldsymbol{P}\mathbf{1}_m &= \boldsymbol{a} \\ \boldsymbol{P}^\top\mathbf{1}_n &= \boldsymbol{b} \end{aligned} \tag{3}$$

と書き表すこともできる．

点群の比較-定式化-

問題 2.1(点群の場合の最適輸送問題)

$$
\begin{aligned}
\underset{\boldsymbol{P} \in \mathbb{R}^{n \times m}}{\text{minimize}} \quad & \sum_{i=1}^{n} \sum_{j=1}^{m} C(x_i, y_j) P_{ij} \\
\text{subject to} \quad & P_{ij} \geq 0 \qquad (\forall i \in [n], \forall j \in [m]) \\
& \sum_{j=1}^{m} P_{ij} = a_i \qquad (\forall i \in [n]) \\
& \sum_{i=1}^{n} P_{ij} = b_j \qquad (\forall j \in [m])
\end{aligned} \tag{2}
$$

この問題は線形計画問題である.

また，実行可能領域が非空かつ，問題が有界なので，常に最適解が存在する.

点群の比較-輸送多面体-

制約条件を満たす行列全体の集合を

$$\mathcal{U}(\boldsymbol{a}, \boldsymbol{b}) \overset{\text{def}}{=} \{\boldsymbol{P} \in \mathbb{R}^{n \times m} \mid P_{ij} \geq 0, \boldsymbol{P}\mathbf{1}_m = \boldsymbol{a}, \boldsymbol{P}^\top \mathbf{1}_n = \boldsymbol{b}\} \tag{3}$$

と表し，$\mathcal{U}(\boldsymbol{a}, \boldsymbol{b})$ を輸送多面体と呼び，この元を輸送行列と呼ぶ．

ここで，$\tilde{\boldsymbol{P}} \overset{\text{def}}{=} \boldsymbol{a}\boldsymbol{b}^\top \in \mathbb{R}^{n \times m}$ とおくと，$\boldsymbol{a} \in \Sigma_n, \boldsymbol{b} \in \Sigma_m$ に注意すると，$\tilde{P}_{ij} \geq 0$ かつ，

$$\begin{aligned} \tilde{\boldsymbol{P}}\mathbf{1}_m &= \boldsymbol{a}\boldsymbol{b}^\top \mathbf{1}_m = \boldsymbol{a} \\ \tilde{\boldsymbol{P}}^\top \mathbf{1}_n &= (\boldsymbol{a}\boldsymbol{b}^\top)^\top \mathbf{1}_n = \boldsymbol{b}\boldsymbol{a}^\top \mathbf{1}_n = \boldsymbol{b} \end{aligned} \tag{4}$$

が成り立つので，$\tilde{\boldsymbol{P}} \in \mathcal{U}(\boldsymbol{a}, \boldsymbol{b})$ となる． $\longrightarrow$ $\mathcal{U}(\boldsymbol{a}, \boldsymbol{b}) \neq \emptyset$

点群の比較-すっきりとした表現-

各成分が $C_{ij} = C(x_i, y_j)$ で与えられる行列 $\boldsymbol{C} \in \mathbb{R}^{n \times m}$ を**コスト行列**と呼ぶ．

また，内積 $\langle \bullet, \bullet \rangle : \mathbb{R}^{n \times m} \times \mathbb{R}^{n \times m} \to \mathbb{R}$ を以下のように定義する．

$$\langle \boldsymbol{A}, \boldsymbol{B} \rangle \stackrel{\mathrm{def}}{=} \sum_{i=1}^{n} \sum_{j=1}^{m} A_{ij} B_{ij} \tag{5}$$

これらの記法を用いると，最適輸送問題は以下のように表せる．

$$\mathrm{OT}(\alpha, \beta, C) = \min_{\boldsymbol{P} \in \mathcal{U}(\boldsymbol{a}, \boldsymbol{b})} \langle \boldsymbol{C}, \boldsymbol{P} \rangle \tag{6}$$

点群の比較-同時分布を求める問題との等価性-

輸送行列 $\boldsymbol{P} \in \mathcal{U}(\boldsymbol{a}, \boldsymbol{b})$ を同時分布の確率を並べた表だとみなすと，α, β は $\boldsymbol{P}$ の周辺分布となる．

また，C_{ij} の期待値 (期待コスト) は $\langle \boldsymbol{C}, \boldsymbol{P} \rangle$ となる．

よって，最適輸送問題は二つの周辺分布を入力として受け取り，期待コストの最も低い同時分布を求める問題と見ることができる．

2.1 線形計画による定式化

2.1.2 ヒストグラムの比較

ヒストグラムの比較-問題設定-

n 個のカテゴリーからなるヒストグラム $\boldsymbol{a}, \boldsymbol{b} \in \Sigma_n$ の比較を考える．

ここで，各要素 C_{ij} がカテゴリー i からカテゴリ j へのコストを表す行列を $\boldsymbol{C} \in \mathbb{R}^{n \times n}$ とすると，最適輸送問題は次のように定式化される．

問題 2.2(ヒストグラムの場合の最適輸送問題)

$$
\begin{aligned}
\underset{\boldsymbol{P} \in \mathbb{R}^{n \times n}}{\text{minimize}} \quad & \sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij} P_{ij} \\
\text{subject to} \quad & P_{ij} \geq 0 \qquad (\forall i \in [n], \forall j \in [n]) \\
& \sum_{j=1}^{n} P_{ij} = a_i \qquad (\forall i \in [n]) \\
& \sum_{i=1}^{n} P_{ij} = b_j \qquad (\forall j \in [n])
\end{aligned}
\tag{7}
$$

点群の場合とヒストグラムの場合の最適輸送問題の比較

問題 2.1(点群の場合)

$$
\begin{aligned}
\underset{\boldsymbol{P} \in \mathbb{R}^{n \times m}}{\text{minimize}} \quad & \sum_{i=1}^{n} \sum_{j=1}^{m} C(x_i, y_j) P_{ij} \\
\text{subject to} \quad & P_{ij} \geq 0 \qquad (\forall i \in [n], \forall j \in [m]) \\
& \sum_{j=1}^{m} P_{ij} = a_i \qquad (\forall i \in [n]) \\
& \sum_{i=1}^{n} P_{ij} = b_j \qquad (\forall j \in [m])
\end{aligned} \tag{8}
$$

問題 2.2(ヒストグラムの場合)

$$
\begin{aligned}
\underset{\boldsymbol{P} \in \mathbb{R}^{n \times n}}{\text{minimize}} \quad & \sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij} P_{ij} \\
\text{subject to} \quad & P_{ij} \geq 0 \qquad (\forall i \in [n], \forall j \in [n]) \\
& \sum_{j=1}^{n} P_{ij} = a_i \qquad (\forall i \in [n]) \\
& \sum_{i=1}^{n} P_{ij} = b_j \qquad (\forall j \in [n])
\end{aligned} \tag{9}
$$

ほとんど同じ．

2.1 線形計画による定式化

2.1.3 連続分布を含む一般の確率分布の場合

Why 確率測度？

- 測度
  - → 集合を受け取り非負の実数を返す関数
  - → 入力集合にどれだけの「量」が含まれているか (直観的理解)
- 確率測度
  - → 総量が 1 となる測度の特別な場合
  - → 入力集合が表現する現象が実現する確率 (直観的理解)
- OT の文脈で確率測度を用いるメリット
  - → 離散分布と連続分布を区別せず扱える
    - ▷ 現象の背景に仮定する連続分布と，そこからサンプルされた点群の両方を区別なく扱う際に有用
    - ▷ 定理の記述の際に，確率測度で表現すれば場合分けしなくてよい

一般の確率分布の場合-点群の場合の拡張-

点群の場合

輸送行列 $\boldsymbol{P}$ は，α, β を周辺分布として持つ，同時分布の確率の表だと解釈でき，最適輸送問題は期待コストの最も低い同時分布を求める問題と見ることができた． $\longrightarrow$ 一般の確率分布の場合に拡張

点群の場合

輸送行列 $\boldsymbol{P}$ は，α, β を周辺分布として持つ，同時分布の確率の表だと解釈でき，最適輸送問題は期待コストの最も低い同時分布を求める問題と見ることができた． $\longrightarrow$ 一般の確率分布の場合に拡張

問題 2.3(一般の場合の最適輸送問題)

$$
\begin{aligned}
\underset{\pi \in \mathcal{P}(\mathcal{X}\times\mathcal{X})^{*1}}{\text{minimize}} \quad & \int_{\mathcal{X}\times\mathcal{X}} C(x,y)\,d\pi(x,y) \\
\text{subject to} \quad & \pi(\mathcal{A}\times\mathcal{B}) \geq 0 \qquad (\forall \mathcal{A},\mathcal{B} \in \mathcal{F}(\mathcal{X})^{*2}) \\
& \pi(\mathcal{A}\times\mathcal{X}) = \alpha(\mathcal{A}) \qquad (\forall \mathcal{A} \in \mathcal{F}(\mathcal{X})) \\
& \pi(\mathcal{X}\times\mathcal{B}) = \beta(\mathcal{B}) \qquad (\forall \mathcal{B} \in \mathcal{F}(\mathcal{X}))
\end{aligned} \tag{10}
$$

[*1] $\mathcal{X}\times\mathcal{X}$ 上の確率分布全体の集合

[*2] $\mathcal{F}(\mathcal{X})$ は $\mathcal{X}$ 上の σ-加法族 (直観的には確率を測ることのできるすべての事象)

点群の場合と一般の場合の最適輸送問題の比較

問題 2.1(点群の場合)

$$\begin{aligned} \underset{\boldsymbol{P}\in\mathbb{R}^{n\times m}}{\text{minimize}} \quad & \sum_{i,j} C(x_i, y_j)P_{ij} \\ \text{subject to} \quad & P_{ij} \geq 0 \quad (\forall i \in [n], \forall j \in [m]) \\ & \boldsymbol{P}\mathbf{1}_m = \boldsymbol{a} \\ & \boldsymbol{P}^\top \mathbf{1}_n = \boldsymbol{b} \end{aligned} \tag{11}$$

問題 2.3(一般の場合の最適輸送問題)

$$\begin{aligned} \underset{\pi\in\mathcal{P}(\mathcal{X}\times\mathcal{X})}{\text{minimize}} \quad & \int_{\mathcal{X}\times\mathcal{X}} C(x, y)d\pi(x, y) \\ \text{subject to} \quad & \pi(\mathcal{A}\times\mathcal{B}) \geq 0 \quad (\forall \mathcal{A}, \mathcal{B} \in \mathcal{F}(\mathcal{X})) \\ & \pi(\mathcal{A}\times\mathcal{X}) = \alpha(\mathcal{A}) \quad (\forall \mathcal{A} \in \mathcal{F}(\mathcal{X})) \\ & \pi(\mathcal{X}\times\mathcal{B}) = \beta(\mathcal{B}) \quad (\forall \mathcal{B} \in \mathcal{F}(\mathcal{X})) \end{aligned} \tag{12}$$

制約式は，カップリング π を周辺化したときに α, β となることを表している．

- 一般の場合の最適輸送問題は LP じゃない

  $\rightarrow$ α, β からのサンプルを用いて点群の比較に帰着させる

  or

  $\rightarrow$ この問題の双対問題の決定変数となる連続関数を NN でモデリングして解く (4 章)

2.1 線形計画による定式化

2.1.4 特殊例：ワッサースタイン距離

距離の公理

定義：距離の公理

集合 $\mathcal{X}$ 上の写像 $d\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ が任意の $x, y, z \in \mathcal{X}$ に対して，以下の 3 つの条件 (**距離の公理**) をすべて満たすとき，写像 d を $\mathcal{X}$ 上の距離関数という．

非退化性 $d(x, y) = 0 \iff x = y$

対称性 $d(x, y) = d(y, x)$

三角不等式 $d(x, y) + d(y, z) \geq d(x, z)$

Q.OT(α, β, C) は距離の公理を満たすか？

A. 反例あり．例えば，C を恒等的に 0 を返すコスト関数とすると，任意の α, β に対して，OT$(\alpha, \beta, C) = 0$ となるので，非退化性を満たさない．

$\longrightarrow$ OT(α, β, C) が距離の公理を満たすための十分条件は？

ワッサースタイン距離-定義-

コスト関数に制約を課すことで，距離の公理を満たすように $\alpha, \beta \in \mathcal{P}(\mathcal{X})$ 間の距離を定めることができる．

定義：ワッサースタイン距離

$\mathcal{X}$ 上の距離関数 $d\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ と実数 $p \geq 1$ について，コスト関数を $C(x, y) = d(x, y)^p$ と定義する．このとき，$\alpha, \beta \in \mathcal{P}(\mathcal{X})$ について，

$$W_p(\alpha, \beta) \stackrel{\text{def}}{=} \mathrm{OT}(\alpha, \beta, C)^{1/p} \tag{13}$$

を α と β の p-ワッサースタイン距離という．

$\mathcal{X} = \mathbb{R}^N$ のときには，$C(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|_2$ とした 1-ワッサースタイン距離や，$C(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|_2^2$ とした 2-ワッサースタイン距離がよく用いられる．

次に，ワッサースタイン距離が距離の公理を満たすことを示す．

ワッサースタイン距離-距離の公理を満たす-

定理：ワッサースタイン距離は距離の公理を満たす

p-ワッサースタイン距離は距離の公理を満たす．すなわち，

1. $W_p(\alpha, \beta) = 0$ のときかつそのときのみ $\alpha = \beta$
2. $W_p(\alpha) = W_p(\beta) \qquad \forall \alpha, \beta$
3. $W_p(\alpha, \beta) + W_p(\beta, \gamma) \geq W_p(\alpha, \gamma) \qquad \forall \alpha, \beta, \gamma$

- 点対 $x, y \in \mathcal{X}$ の距離を定めることで，分布 $\alpha, \beta \in \mathcal{P}(\mathcal{X})$ の距離を定めることができる．(問題の簡略化)
- この定理の証明は，点群についてのみ行なう．
- 一般の確率測度についての証明は，Villani[*3]

[*3] C.Villani. *Optimal Transport: Old and New.* Springer, 2009.

ワッサースタイン距離-証明の準備-

- $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$ とする．ただし，要素の重複なし．
- 距離関数 $d \colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ を任意にとり，$C(x, y) = d(x, y)^p$ とする．ただし，$p \geq 1$.
- 確率ベクトル $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \in \Sigma_n$ を任意にとり，$\alpha, \beta, \gamma \in \mathcal{P}(\mathcal{X})$ をそれぞれ以下のように定義する．

$$\alpha = \sum_{i=1}^{n} a_i \delta_{x_i} \tag{14}$$

$$\beta = \sum_{i=1}^{n} b_i \delta_{x_i} \tag{15}$$

$$\gamma = \sum_{i=1}^{n} c_i \delta_{x_i} \tag{16}$$

- (α, β) 間の最適輸送行列，(β, γ) 間の最適輸送行列をそれぞれ $\boldsymbol{P}^* \in \mathbb{R}_{\geq 0}^{n \times n}, \boldsymbol{Q}^* \in \mathbb{R}_{\geq 0}^{n \times n}$ とする．

ワッサースタイン距離-非退化性の証明-

1.$W_p(\alpha,\beta)=0 \iff \alpha=\beta$ を示す．
(i).$W_p(\alpha,\beta)=0 \implies \alpha=\beta$ を示す．

proof.

$W_p(\alpha,\beta)=0$ とすると，W_p の定義 [*4] から，以下の式が成り立つ．

$$\begin{aligned} W_p(\alpha,\beta)^p &= \mathsf{OT}(\alpha,\beta,C) \\ &= \sum_{i,j} C(x_i,x_j)P^*_{ij} \\ &= \sum_{i,j} d(x_i,x_j)^p P^*_{ij} = 0 \end{aligned} \tag{17}$$

ここで，任意の i,j について $d(x_i,x_j)\geq 0, P^*_{ij}\geq 0$ が成り立つので，式 (17) の最後の等式は，任意の i,j について $d(x_i,x_j)=0$ または $P^*_{ij}=0$ が成り立つことと同値である．

[*4] $W_p(\alpha,\beta) \stackrel{\text{def}}{=} \mathsf{OT}(\alpha,\beta,C)^{1/p}$

proof.(続き)

ところで，$i \neq j$ すなわち，$x_i \neq x_j$ のとき，$d(x_i, x_j) > 0$ となるので，$P^*_{ij} = 0$ となる．よって，$\boldsymbol{P}^*$ は対角行列となる．

このとき，$\boldsymbol{P}^* \in \mathcal{U}(\boldsymbol{a}, \boldsymbol{b})$[*5] に注意すると，

$$\boldsymbol{a} = \boldsymbol{P}^* \mathbf{1}_n = (\boldsymbol{P}^*)^\top \mathbf{1}_n = \boldsymbol{b} \tag{18}$$

が成り立ち，これは $\alpha = \beta$ を意味する．

[*5] $\mathcal{U}(\boldsymbol{a}, \boldsymbol{b}) \overset{\text{def}}{=} \{\boldsymbol{P} \in \mathbb{R}^{n \times n} \mid P_{ij} \geq 0, \boldsymbol{P}\mathbf{1}_n = \boldsymbol{a}, \boldsymbol{P}^\top \mathbf{1}_n = \boldsymbol{b}\}$

(ii).$\alpha = \beta \implies W_p(\alpha, \beta) = 0$ を示す.

proof.

$\alpha = \beta$ すなわち $\boldsymbol{a} = \boldsymbol{b}$ とする．このとき，
$\boldsymbol{P} = \mathrm{diag}(\boldsymbol{a}) = \mathrm{diag}(\boldsymbol{b}) \in \mathcal{U}(\boldsymbol{a}, \boldsymbol{b})$ であり，さらに，

$$\begin{aligned} \sum_{i,j} C(x_i, x_j) P_{ij} &= \sum_{i,j} d(x_i, x_j)^p P_{ij} \\ &= \sum_{i} d(x_i, x_i)^p a_i \\ &= 0 \quad (\because \text{非退化性}) \end{aligned} \tag{19}$$

となる.

また，距離の公理から導かれる $d(x_i, x_j) \geq 0$ より，
$C(x_i, x_j) \geq 0 \quad \forall i, j$ が成り立つ．これと，$P_{ij} \geq 0 \quad \forall i, j$ より，
最適輸送問題の目的関数は常に非負であるので，この $\boldsymbol{P}$ が最適解である．すなわち，$W_p(\alpha, \beta) = 0$ である.

以上 (i),(ii) より，$W_p(\alpha, \beta) = 0 \iff \alpha = \beta$ が示された. $\square$

ワッサースタイン距離-対称性の証明-

2.$W_p(\alpha) = W_p(\beta) \qquad \forall \alpha, \beta$ を示す.

proof.

$d(x_i, x_j) = d(x_j, x_i) \qquad \forall x_i, x_j \in \mathcal{X}$ より明らか. $\square$

ワッサースタイン距離-三角不等式の証明-

3.$W_p(\alpha,\beta)+W_p(\beta,\gamma)\geq W_p(\alpha,\gamma) \qquad \forall\alpha,\beta,\gamma$ を示す．

proof.

$\boldsymbol{R} \stackrel{\text{def}}{=} \boldsymbol{P}^*\,\mathrm{diag}(1/\boldsymbol{b})\boldsymbol{Q}^* \in \mathbb{R}^{n\times n}_{\geq 0}$ とする．ただし，$1/\boldsymbol{b}$ は成分ごとに逆数をとり，0 の逆数は 0 であると定義する．

このとき，

$$\begin{aligned}\boldsymbol{R}\mathbf{1}_n &= \boldsymbol{P}^*\,\mathrm{diag}(1/\boldsymbol{b})\boldsymbol{Q}^*\mathbf{1}_n \\ &= \boldsymbol{P}^*\,\mathrm{diag}(1/\boldsymbol{b})\boldsymbol{b} \\ &= \boldsymbol{P}^*\mathbf{1}_{\boldsymbol{b}\geq 0}{}^{*6} \\ &= \boldsymbol{P}^*\mathbf{1}_n \\ &= \boldsymbol{a}\end{aligned} \tag{20}$$

が成り立つ．同様に，$\boldsymbol{R}^\top\mathbf{1}_n=\boldsymbol{c}$ なので，$\boldsymbol{R}\in\mathcal{U}(\boldsymbol{a},\boldsymbol{c})$ となる．

*6 i 次元目の要素は $b_i>0$ のとき 1,otherwise 0 である指示ベクトル

proof.(続き)

$$
\begin{aligned}
W_p(\alpha,\gamma) &= \min_{\boldsymbol{P}\in\mathcal{U}(\boldsymbol{a},\boldsymbol{c})}\left(\sum_{i,k} C(x_i,x_k)P_{ik}\right)^{1/p} \\
&= \min_{\boldsymbol{P}\in\mathcal{U}(\boldsymbol{a},\boldsymbol{c})}\left(\sum_{i,k} d(x_i,x_k)^p P_{ik}\right)^{1/p} \\
&\leq \left(\sum_{i,k} d(x_i,x_k)^p R_{ik}\right)^{1/p} \qquad (\because \boldsymbol{R}\in\mathcal{U}(\boldsymbol{a},\boldsymbol{c})) \\
&= \left(\sum_{i,k} d(x_i,x_k)^p \sum_j P^*_{ij}\frac{1}{b_j}Q^*_{jk}\right)^{1/p} \qquad (\because \boldsymbol{R}\text{ の定義より}) \\
&= \left(\sum_{i,j,k} d(x_i,x_k)^p P^*_{ij}\frac{1}{b_j}Q^*_{jk}\right)^{1/p}
\end{aligned}
$$

proof.(続き)

$$
\begin{aligned}
W_p(\alpha,\gamma) &\le \left(\sum_{i,j,k} d(x_i,x_k)^p P^*_{ij}\frac{1}{b_j}Q^*_{jk}\right)^{1/p} \\
&\le \left(\sum_{i,j,k}\big(d(x_i,x_j)+d(x_j,x_k)\big)^p P^*_{ij}\frac{1}{b_j}Q^*_{jk}\right)^{1/p} \quad (\because \text{三角不等式}) \\
&\le \left(\sum_{i,j,k} d(x_i,x_j)^p P^*_{ij}\frac{1}{b_j}Q^*_{jk}\right)^{1/p} \\
&\quad + \left(\sum_{i,j,k} d(x_j,x_k)^p P^*_{ij}\frac{1}{b_j}Q^*_{jk}\right)^{1/p}
\end{aligned}
$$

(∵ ミンコフスキーの不等式 *7)

*7 $1 \le p$ のとき，$\|\boldsymbol{x}+\boldsymbol{y}\|_p \le \|\boldsymbol{x}\|_p + \|\boldsymbol{y}\|_p$ が成り立つ．ここで，$\|\boldsymbol{x}\|_p = (\sum_i |x_i|^p)^{1/p}$

proof.(続き)

$$
\begin{aligned}
W_p(\alpha,\gamma) &\leq \left(\sum_{i,j,k} d(x_i,x_j)^p P^*_{ij}\frac{1}{b_j}Q^*_{jk}\right)^{1/p} \\
&\quad + \left(\sum_{i,j,k} d(x_j,x_k)^p P^*_{ij}\frac{1}{b_j}Q^*_{jk}\right)^{1/p} \\
&= \left(\sum_{i,j} d(x_i,x_j)^p P^*_{ij}\frac{1}{b_j}\sum_k Q^*_{jk}\right)^{1/p} \\
&\quad + \left(\sum_{j,k} d(x_j,x_k)^p Q^*_{jk}\frac{1}{b_j}\sum_i P^*_{ij}\right)^{1/p} \\
&= \left(\sum_{i,j} d(x_i,x_j)^p P^*_{ij}\right)^{1/p} + \left(\sum_{j,k} d(x_j,x_k)^p Q^*_{jk}\right)^{1/p} \\
&= W_p(\alpha,\beta) + W_p(\beta,\gamma)
\end{aligned}
$$

$\square$