# 最適輸送理論入門

田中 章詞　RIKEN AIP/iTHEMS

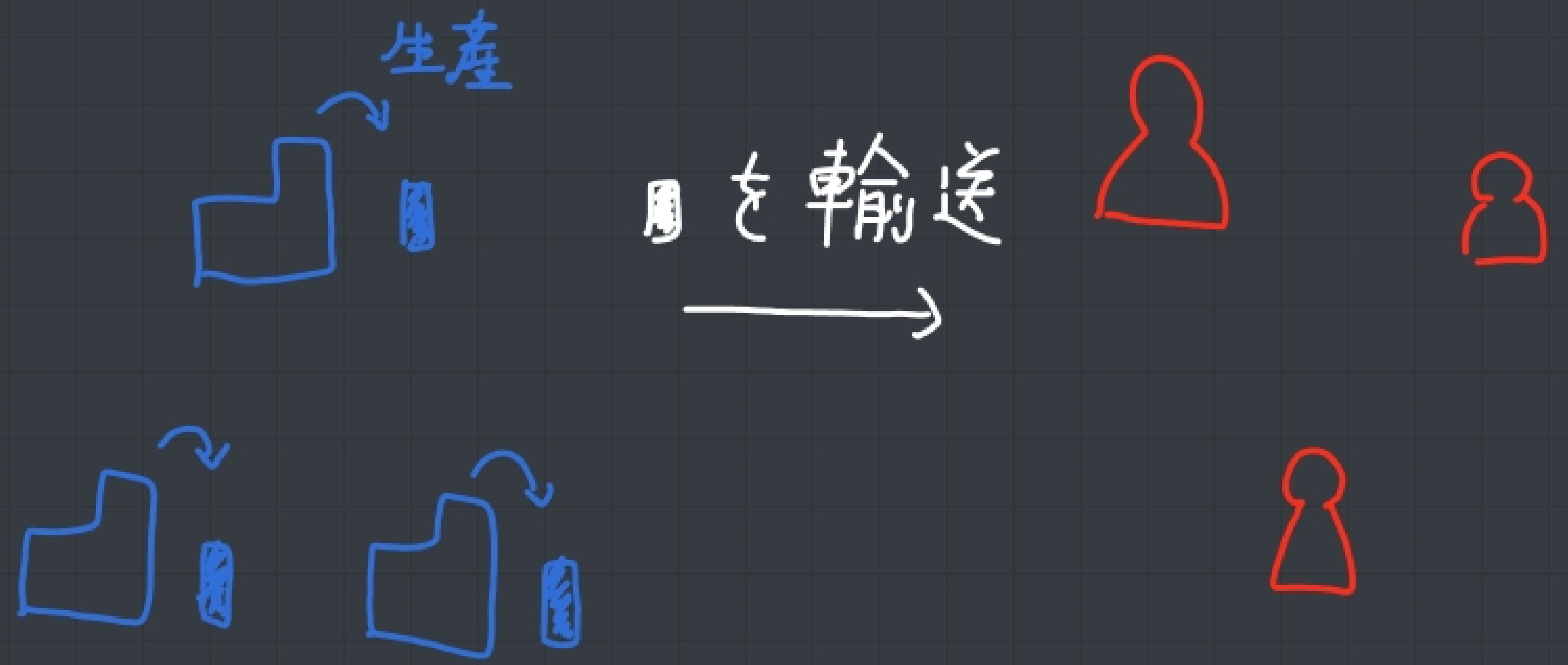
Ref: arXiv:1803.00567

# 内容

# 1 OTの概要

## 1-1 OTとは?

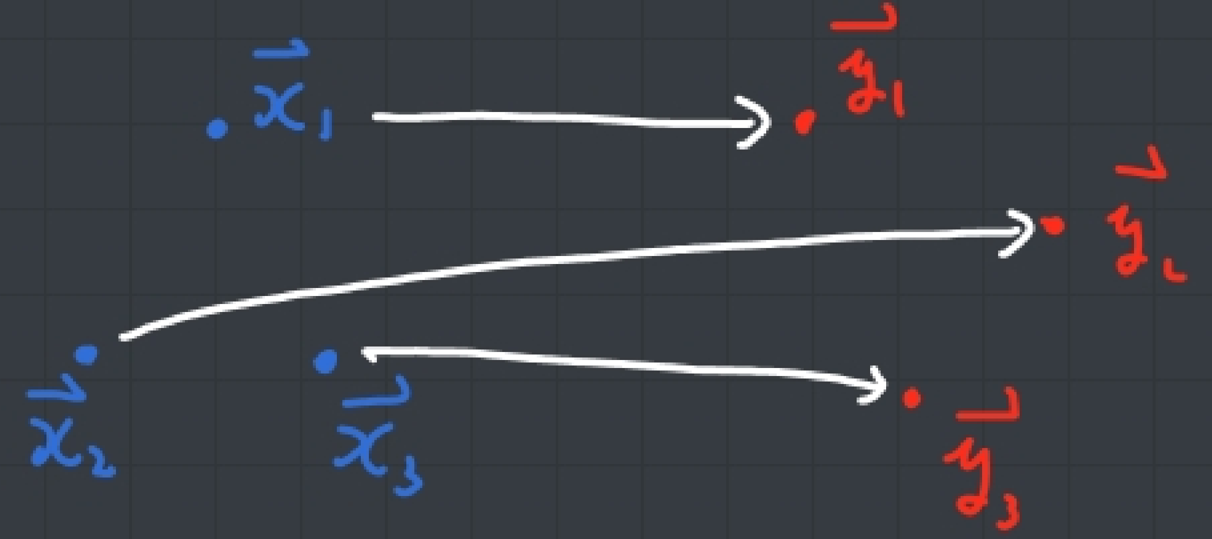
Optimal Transportの略




どのように輸送すればコスト最小になるか?

## ▨ 輸送／コスト？

輸送




$\vec{x}_i \to \vec{y}_j$ の関数

コスト $|\vec{x}-\vec{y}| = \sqrt{(x^\mu - y^\mu)(x_\mu - y_\mu)}$ ⟶ 一般に $|\vec{x}-\vec{y}|^p$ $(p \geq 1)$

**OT問題 1**

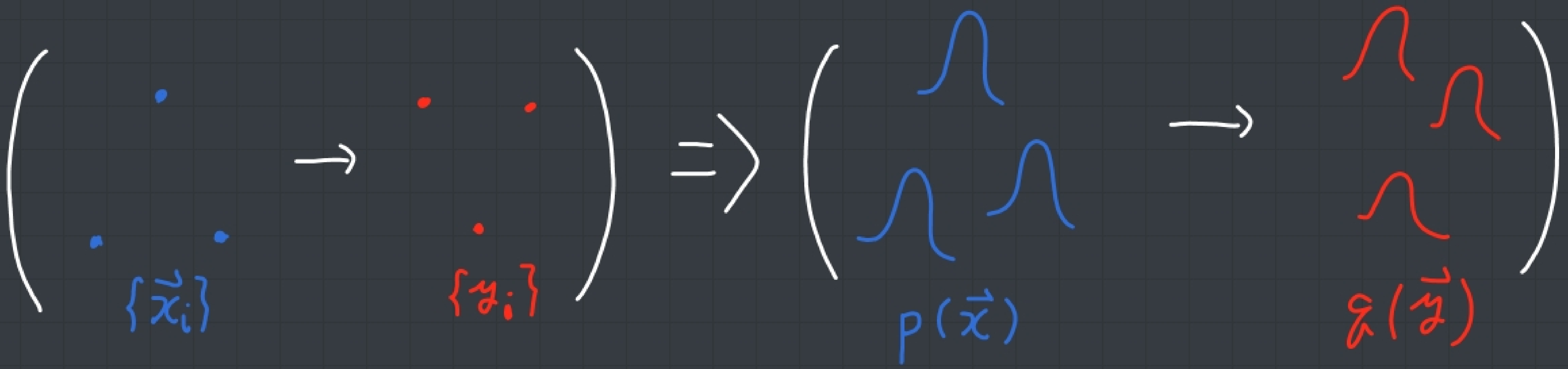
$\{\vec{x}_i\}_{i=1,2,\cdots,N}$ と $\{\vec{y}_i\}_{i=1,2,\cdots,N}$ に対して $p \geq 1$ として

$$\min_{\sigma \in S_N} \left\{ \frac{1}{N} \sum_{i=1}^{N} |\vec{x}_{\sigma(i)} - \vec{y}_i|^p \right\}$$

をみつけよ

→ ソルバーがある
e.g.
Python Optimal Transport

## 確率分布で一般化




**OT問題 2 (Monge)**

$p(\vec{x})$ と $q(\vec{y})$ に対して 関数 $T : \vec{x} \to \vec{y}$ のうち

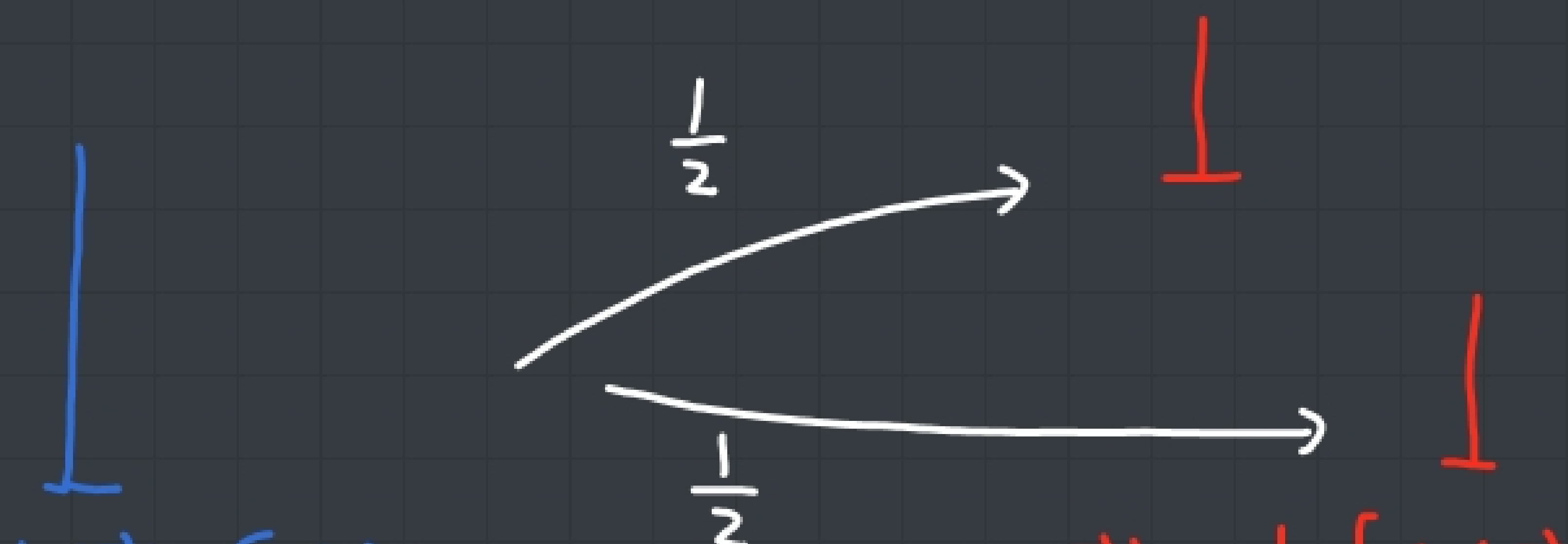
$$\int p(\vec{x})\, d\vec{x} \cdot \delta(\vec{y} - T(\vec{x})) = q(\vec{y})$$

をみたすものの中で $p \geq 1$ として

$$\min_{T} \left\{ \int |T(\vec{x}) - \vec{x}|^{p}\, p(\vec{x})\, d\vec{x} \right\}$$ をみつけよ

実は良い設定ではない

## ・良くない理由1　$T_*$ が存在しないことがある




$p(\vec{x}) = \delta(\vec{x})$　　$q(\vec{y}) = \frac{1}{2}\left\{\delta(\vec{y}-\vec{a}) + \delta(\vec{y}-\vec{b})\right\}$

1対2なので
$T : \vec{x} \to \vec{y}$
と表現不可

## ・良くない理由2　非線型性

$$\text{ロス} = \int \sqrt{(T(\vec{x}) - \vec{x})^2}^{\,p}\, p(\vec{x})\, d\vec{x}$$

$\sqrt{\quad}^{2}\,{}^{p}$

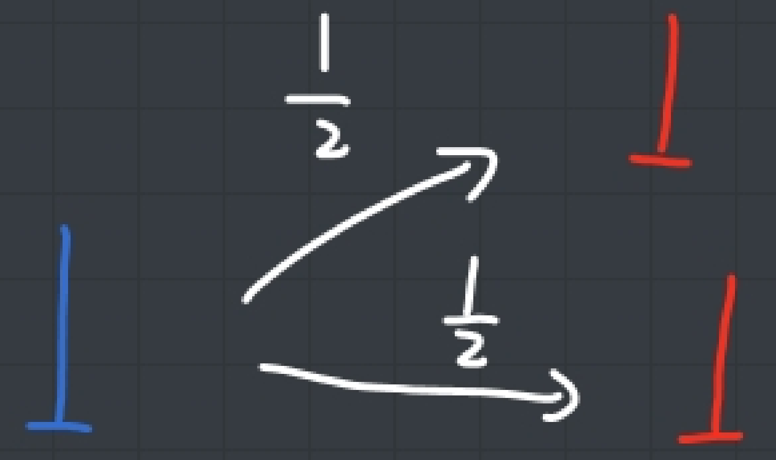
制約　δ関数積分 → $\left.\dfrac{p(\vec{x})}{|\det \nabla T(\vec{x})|}\right|_{T(\vec{x})=\vec{y}} = q(\vec{y})$

detがいやだ

## ■ より良い設定へ


$$\int \underline{\delta(\vec{y}-T(\vec{x}))}\, p(\vec{x})\, d\vec{x} = q(\vec{y})$$

$\Rightarrow$




$$\int \pi(\vec{y}|\vec{x})\, p(\vec{x})\, d\vec{x} = q(\vec{y})$$

### OT問題 3 (Monge-Kantorovich)

$p(\vec{x})$ と $q(\vec{y})$ に対して 同時確率 $\pi(\vec{x},\vec{y})$ のうち

constraints $\boxed{\int \pi(\vec{x},\vec{y})\, d\vec{y} = p(\vec{x}) \quad \text{かつ} \quad \int \pi(\vec{x},\vec{y})\, d\vec{x} = q(\vec{y})}$

をみたすもので

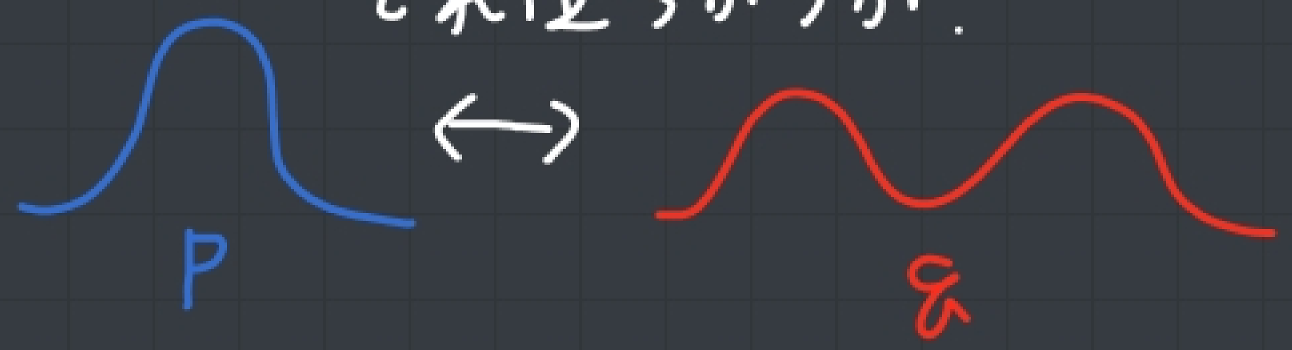
$$\min_{\pi} \left\{ \int |\vec{x}-\vec{y}|^p\, \pi(\vec{x},\vec{y})\, d\vec{x}\, d\vec{y} \right\}$$ をみつけよ

LPという良いクラスの問題!

# 1-2. Wasserstein 距離

■ 確率分布の間のキョリ

どれ位ちがうか?




$-\log q_\theta(x_i)$ $\downarrow\theta$

Var? optimal


Bregman

? ⊄ f-divergence

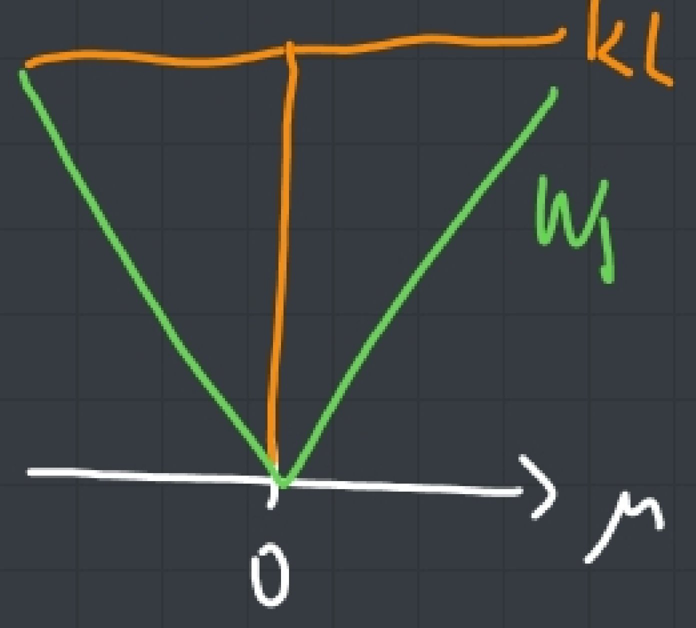
Fisher divergence

↑

diffusion model

Kullback-Leibler "距離"

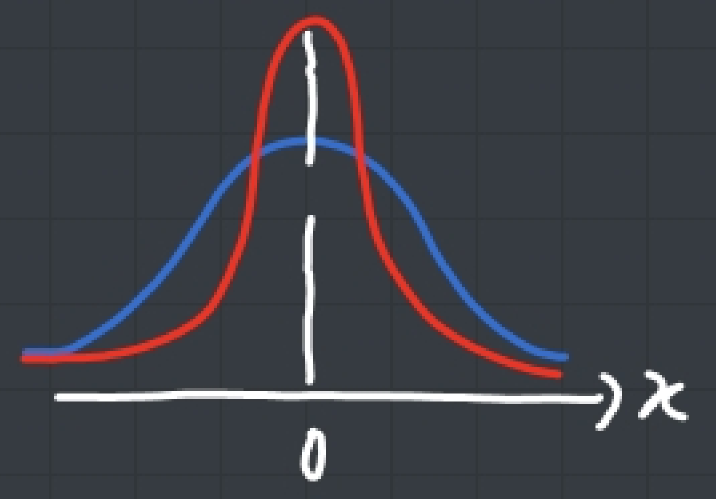
$$D_{KL}(p\|q)=\int p(\vec{x})\,d\vec{x}\cdot\log\frac{p(\vec{x})}{q(\vec{x})}$$




p-Wasserstein 距離

$$D_{W_p}(p,q)=\Big(\underbrace{\min_{\pi}\int|\vec{x}-\vec{y}|^p\,\pi(\vec{x},\vec{y})\,d\vec{x}\,d\vec{y}}_{\text{constraints OT 問題 3}}\Big)^{1/p}.$$

## 例 1次元Gauss分布



$p = N(0, \sigma_p^2)$

$q = N(0, \sigma_q^2)$

カンタンな計算により

$$D_{KL}(p \| q) = \frac{1}{2}\left\{-\log\frac{\sigma_p^2}{\sigma_q^2} + \left(\frac{\sigma_p^2}{\sigma_q^2} - 1\right)\right\}$$

$D_{W_p}$ ?

$$\frac{(\mu_1 - \mu_2)^2}{\sigma^2}$$

・$p=2$ のケースがカンタン

$$\int (x-y)^2 \pi(x,y)\,dx\,dy = \langle x^2 + y^2 - 2xy \rangle_\pi$$

2次モーメントだけでよい

$$\pi = N\left(\vec{0}, \Sigma = \begin{pmatrix} \sigma_p^2 & k \\ k & \sigma_q^2 \end{pmatrix}\right)$$

半正定値

$$= \sigma_p^2 + \sigma_q^2 - 2k \xleftarrow{\min} |\Sigma| = (\sigma_p\sigma_q)^2 - k^2 \geq 0$$

$$\Downarrow \min$$

$$(\sigma_p - \sigma_q)^2$$

$$D_{W_2}(p, q) = |\sigma_p - \sigma_q|$$

# 1-3. 双対性 (duality)

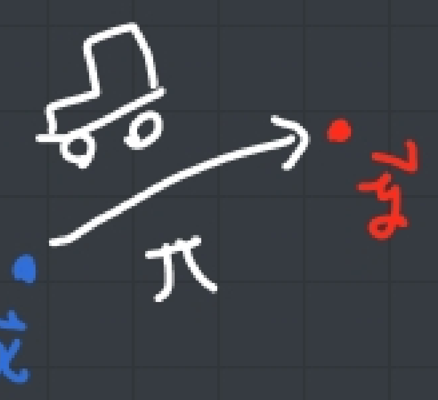
**OT問題 3'**

$p(\vec{x})$ と $q(\vec{y})$ に対して "ポテンシャル関数" $f(\vec{x}), g(\vec{y})$ で $p \geq 1$ として

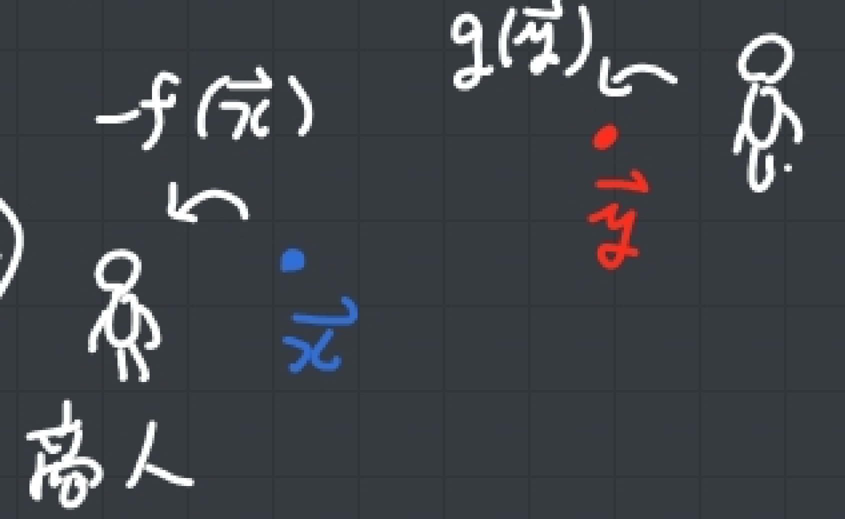
$$f(\vec{x}) + g(\vec{y}) \leq |\vec{x} - \vec{y}|^p$$

をみたすものの中で

$$\left( \max_{\substack{f,g \\ +\text{constraint}}} \left\{ \langle f(\vec{x}) \rangle_p + \langle g(\vec{y}) \rangle_q \right\} \right)^{1/p} = D_{W_p}(p, q)$$ となる




運送費↓ ⟷




$f(\vec{x}) + g(\vec{y}) \leq |\vec{x} - \vec{y}|^p$ もうかる

# ■ 1次元Gauss分布での例

・p=2のケース

**Step1 fにあたりをつける**

「$\sigma^2$」が出るはず… → $f(x)=\alpha\cdot x^2$ としてみる　αを動かす

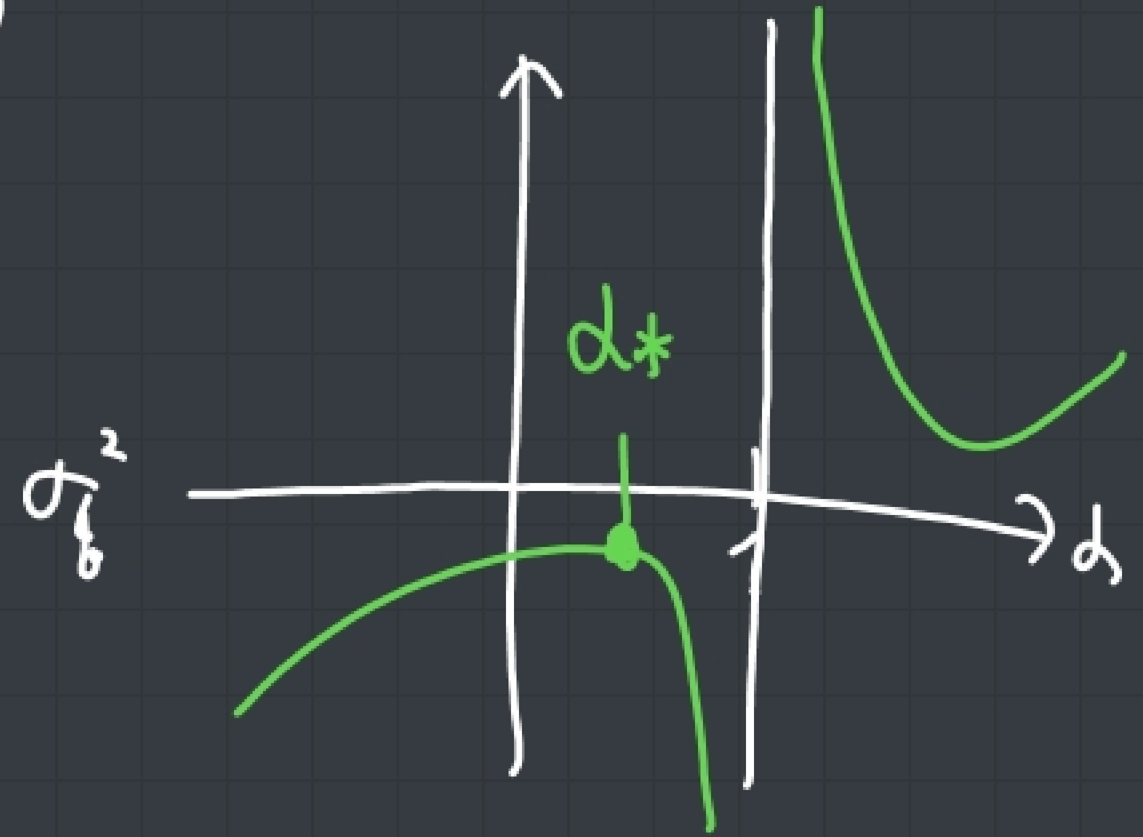
**Step2 最良のgが決まる**

g大がよい → $g(y)\le (x-y)^2-f(x) = \underset{>0}{(1-\alpha)}\left(x-\frac{y}{1-\alpha}\right)^2 \underbrace{-\frac{y^2}{1-\alpha}+y^2}$

$g^c(y)=\min_x\{\ \_\ \}=\ \_$

**Step3 α<1でmax**

$$F(\alpha)=\underbrace{\underbrace{\langle f(x)\rangle_p}_{\alpha\cdot\sigma_p^2}+\underbrace{\langle g^c(y)\rangle_q}_{-\frac{\sigma_q^2}{1-\alpha}+\sigma_q^2}}\overset{\alpha^*}{=}(\sigma_p-\sigma_q)^2$$




$$F'(\alpha)=\sigma_p^2-\frac{\sigma_q^2}{(1-\alpha)^2}$$

$$\frac{1}{1-\alpha_*}=\frac{\sigma_p}{\sigma_q}$$

$$\alpha_*=1-\frac{\sigma_q}{\sigma_p}$$

# 2. 双対性について

## 2-1. Lagrange未定乗数法

もとの問題のLagrange関数

$$\mathcal{L}(\pi, f, g) = \int |\vec{x}-\vec{y}|^p \pi(\vec{x},\vec{y})\,d\vec{x}\,d\vec{y} + \int \underline{f(\vec{x})}\left\{p(\vec{x}) - \int \pi(\vec{x},\vec{y})\,d\vec{y}\right\}d\vec{x}$$

$$+ \int \underline{g(\vec{y})}\left\{q(\vec{y}) - \int \pi(\vec{x},\vec{y})\,d\vec{x}\right\}d\vec{y}$$

$$\to \min_{\pi \geq 0} \max_{f,g} \mathcal{L}(\pi, f, g)$$

$$\overset{?}{=} \max_{f,g} \min_{\pi \geq 0} \mathcal{L}(\pi, f, g)$$

$$= \max_{f,g}\left\{ \langle f(\vec{x}) \rangle_p + \langle g(\vec{y}) \rangle_q + \min_{\pi \geq 0} \int \left\{ |\vec{x}-\vec{y}|^p - (f(\vec{x}) + g(\vec{y})) \right\} \pi(\vec{x},\vec{y})\,d\vec{x}\,d\vec{y} \right\}$$

- なぜ minmax = maxmin ?

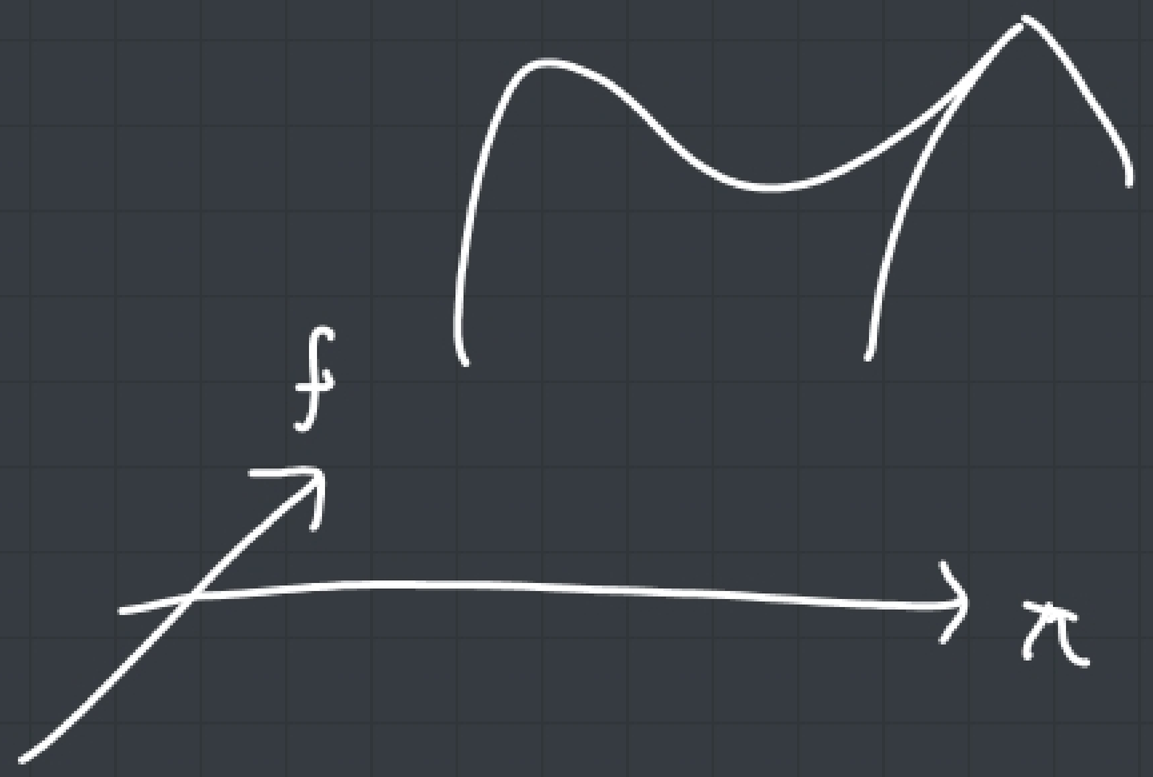
$\vec{x}, \vec{y} \Rightarrow i, j$

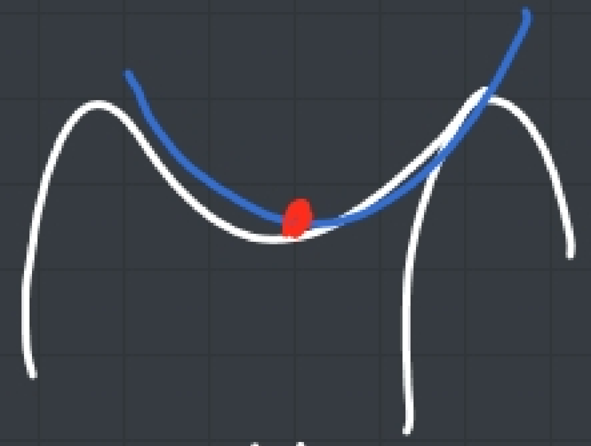
$\rightarrow \mathcal{L}(\pi, f, g) = \sum_{ij} c_{ij}\pi_{ij} + \sum_{i} f_i \left\{ p_i - \sum_{j} \pi_{ij} \right\} + \sum_{j} g_j \left\{ q_j - \sum_{i} \pi_{ij} \right\}$

さらに $i, j \rightarrow$ なし. $g$ なし. $c = p = 1$

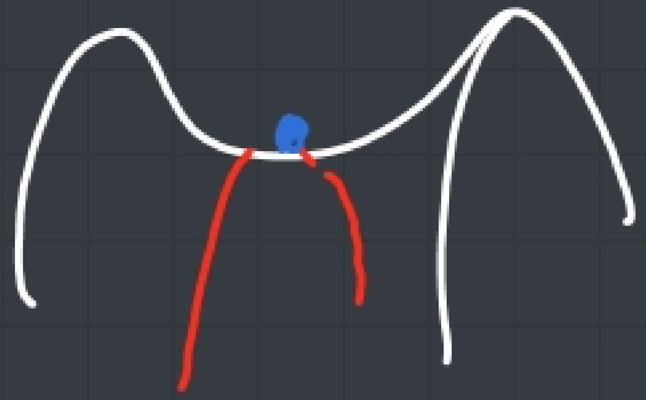
$$\rightarrow \mathcal{L}(\pi, f) = \pi + f\{1 - \pi\}$$
$$= \pi + f - f\pi$$

$\rightarrow$ strong duality でけんさく




$\min_{\pi} \max_{f} \mathcal{L} =$ 

$\parallel$

$\max_{f} \min_{\pi} \mathcal{L} =$ 

## 2-2 $p=1$ のケース

$$|\min_z F(z) - \min_u G(u)| \leq \max_z |F(z) - G(z)| \quad (*)$$

**Kantorovich-Rubinstein duality**

$$\min_\pi \left\{ \langle |\vec{x} - \vec{y}| \rangle_\pi \right\} = \max_{h:\text{Lipschitz}} \left\{ \langle h(\vec{x}) \rangle_p - \langle h(\vec{y}) \rangle_q \right\}$$

$$\langle f(\vec{x}) \rangle_p + \langle g(\vec{y}) \rangle_q \leq \langle g^c(\vec{x}) \rangle_p - \langle g^c(\vec{y}) \rangle_q$$

(prf)

**lhs ≥ rhs**

$$\langle h(\vec{x}) \rangle_p - \langle h(\vec{y}) \rangle_q$$

$$= \int \left\{ \underbrace{h(\vec{x}) - h(\vec{y})}_{\leq |\vec{x} - \vec{y}|} \right\} \pi_*(\vec{x}, \vec{y})\, d\vec{x}\, d\vec{y} \leq \text{lhs}$$

$\downarrow \max_h$

rhs ≤ lhs

**lhs ≤ rhs**

$$g^c(\vec{x}) = \min_{\vec{y}} \left\{ |\vec{x} - \vec{y}| - g(\vec{y}) \right\}$$

1. $f(\vec{x}) \leq g^c(\vec{x}) \leq \left\{ \cancel{|\vec{x} - \vec{x}|} - g(\vec{x}) \right\} \leq -g(\vec{x})$

2. $|g^c(\vec{x}) - g^c(\vec{y})| = \left| \min_z \{ |\vec{x} - \vec{z}| - g(\vec{z}) \} - (\vec{x} \text{ を } \vec{y} \text{ に}) \right|$

$$\leq \max_z \underbrace{\left| |\vec{x} - \vec{z}| - |\vec{y} - \vec{z}| \right|}_{\leq |\vec{x} - \vec{y}|} \quad (\triangle\text{不等式})$$

lhs ≤ $g^c$ ≤ rhs

$\downarrow$ $\boxed{\max_{f,g}}$

lhs = dual ≤ rhs

# 3. MLとの関連

## 3-1. GAN

arXiv:1406.2661 では

$$\min_{q} \max_{0\le h\le 1} \overbrace{\left\{ \langle \log h(\vec{x}) \rangle_{p} + \langle \log\{1-h(\vec{y})\} \rangle_{q} \right\}}^{D_{KL}\left(p \,\|\, \frac{p+q}{2}\right) + D_{KL}\left(q \,\|\, \frac{p+q}{2}\right) - 2\log 2}$$

$$\min_{q} \underbrace{D_{W_1}(p, q)}_{\max_{h:\mathrm{Lip}} \{\langle h \rangle - \langle h \rangle\}} \to \text{WGAN}$$

arXiv:1701.07875

→ 反論(?)

arXiv:1710.08446

→ どのGAN?

arXiv:1312.6114

↓

## 3-2. VAE

$$\min_{q} \min_{\pi} \{ \cdots \}$$

⟶ VAE-likeなobjective　WVAE　arXiv:1711.01558

## why WVAE?

$$D^p_{W_p}(p, q) = \min_\pi \langle |\vec{x} - \vec{y}|^p \rangle_\pi$$

$$\pi(x, y) = \underline{\pi(y|x)}\, p(x)$$

リアンザッツ

$$\int q(y|z)\, r(z|x)\, dz$$

$$\gtrsim \min_r \langle | \quad | + D(q, \ ) \rangle$$

$p = 2 \Rightarrow$ Monge-Ampere eq

latent var model

$$\vec{z} \xrightarrow{q(\vec{x}|\vec{z})} \vec{x}$$

$$\boxed{\int q(\vec{z})\, q(\vec{x}|\vec{z})\, d\vec{z} = q(\vec{x})}$$

$$\int \pi(x, y)\, dx = \boxed{q(y)}$$

$$\parallel$$

$$\int q(y|z) \left( \int r(z|x)\, p(x)\, dx \right) dz$$