

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?

1-2. Wasserstein 距離

1-3. 双対性 (duality)

2. 双対性について

2-1. Lagrange 未定乗数法.

2-2. $p=1$ のケース

3. ML との関連

3-1. GAN

3-2. VAE

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?



コスト $|\vec{x} - \vec{y}| = \sqrt{(x^u - y^u)(x_u - y_u)} \rightsquigarrow$ 一般に $|\vec{x} - \vec{y}|^p$ ($p \geq 1$)

OT問題 1

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

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

を求めよ

→ $\forall n^n$ が n ある

e.g.

Python Optimal Transport

確率分布で一般化

$$\left(\begin{array}{c} \cdot \\ \cdot \\ \cdot \\ \{ \vec{x}_i \} \end{array} \rightarrow \begin{array}{c} \cdot \\ \cdot \\ \cdot \\ \{ \vec{y}_i \} \end{array} \right) \Rightarrow \left(\begin{array}{c} \text{bell curve} \\ \text{bell curve} \\ p(\vec{x}) \end{array} \rightarrow \begin{array}{c} \text{bell curve} \\ \text{bell curve} \\ q(\vec{y}) \end{array} \right)$$

OT問題 2 (Monge)

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

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

を満たすもの中 $p \geq 1$ として

$$\min_T \left\{ \int |\underline{T}(\vec{x}) - \vec{x}|^p p(\vec{x}) d\vec{x} \right\} \quad \text{を求めよ}$$

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
 $T: \vec{x} \rightarrow \vec{y}$
 と表現不可

・ 良くない理由2 非線型性

$$R^2 = \int \sqrt{(T(\vec{x}) - \vec{y})^2}^p p(\vec{x}) d\vec{x}$$

$$\sqrt{z^p}$$

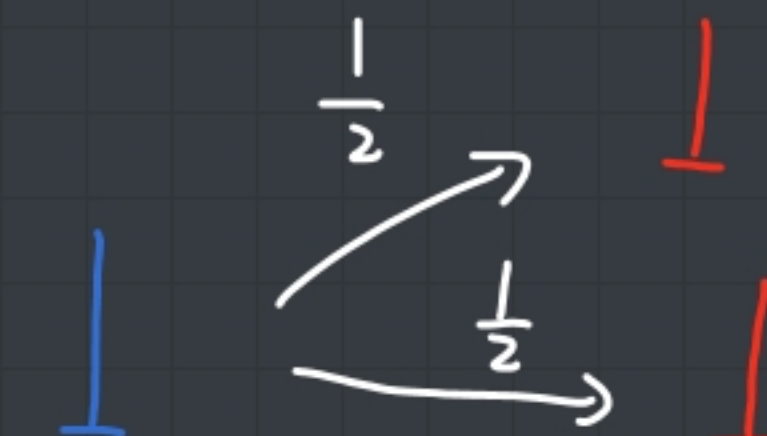
制約 δ 関数積分 $\rightarrow \frac{p(\vec{x})}{|\det \nabla T(\vec{x})|} \Big|_{T(\vec{x})=\vec{y}} = q(\vec{y})$

det が "いやだ"

より良い設定へ



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



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

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

constants $\left[\int \pi(\vec{x}, \vec{y}) d\vec{y} = p(\vec{x}) \text{ から } \int \pi(\vec{x}, \vec{y}) d\vec{x} = q(\vec{y}) \right]$

をみたすものを

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

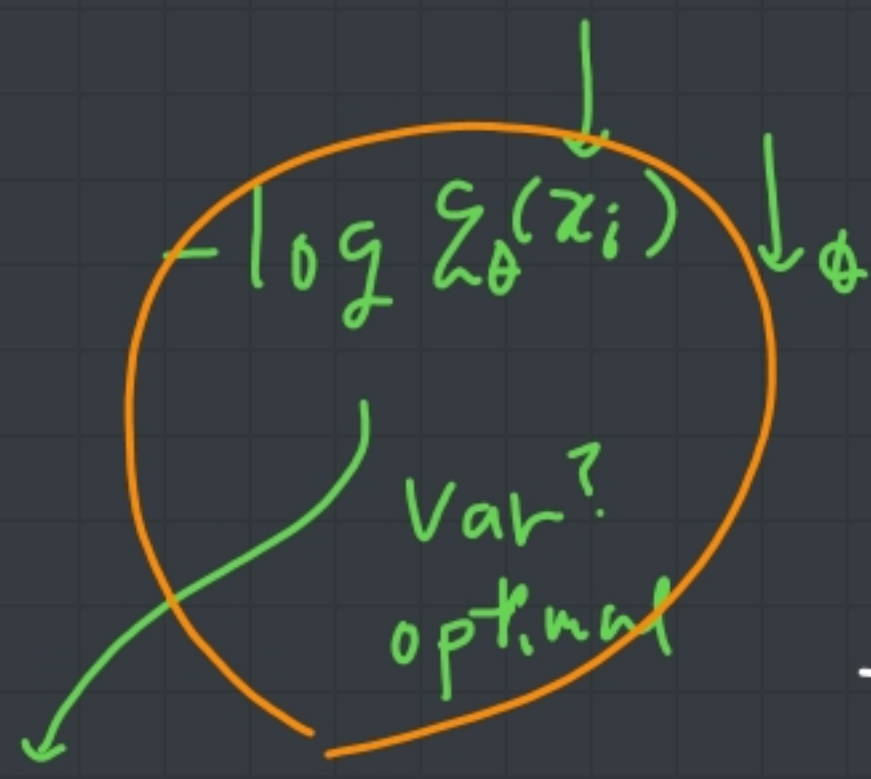
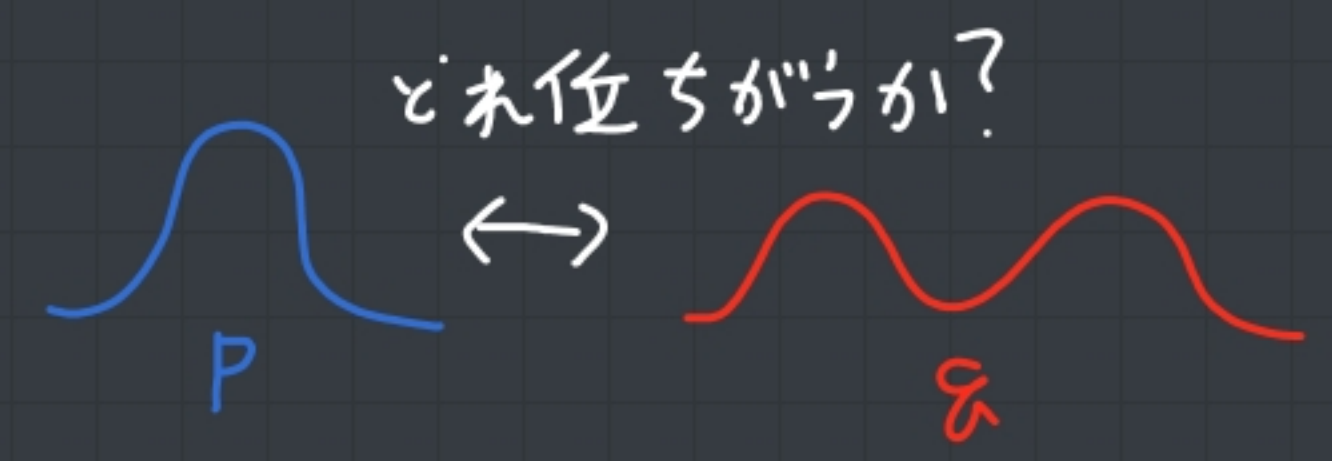
LP という
良いクラスの
問題!

1-2. Wasserstein 距離

確率分布の値のキョリ

f-divergence
 Fisher divergence
 diffusion model

Bregman



Kullback-Leibler "距離"

$$D_{KL}(P \parallel Q) = \int P(\vec{x}) d\vec{x} \cdot \log \frac{P(\vec{x})}{Q(\vec{x})}$$



p-Wasserstein 距離

$$D_{W_p}(P, Q) = \left(\min_{\pi} \int |\vec{x} - \vec{y}|^p \pi(\vec{x}, \vec{y}) d\vec{x} d\vec{y} \right)^{1/p}$$

constraints OT 問題 3

④ 1次元 Gauss 分布



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

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

カンタンな計算により

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

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

・ $p=2$ のケースがカンタン

$$\int (x-y)^2 \pi(x, y) dx dy = \langle \underbrace{x^2 + y^2 - 2xy}_{\text{2次元ベクトルだけだよ}} \rangle_{\pi}$$

$$= \sigma_p^2 + \sigma_q^2 - 2k \quad \leftarrow \text{m/h}$$

$$\Downarrow \text{m/h}$$

$$\underline{(\sigma_p - \sigma_q)^2}$$

2次元ベクトルだけだよ $\pi = \mathcal{N}(\vec{0}, \Sigma = \begin{pmatrix} \sigma_p^2 & k \\ k & \sigma_q^2 \end{pmatrix})$

$\uparrow$ $\uparrow$
非正定値

$$|\Sigma| = (\sigma_p \sigma_q)^2 - k^2 \geq 0$$

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

1-3. 双対性 (duality)

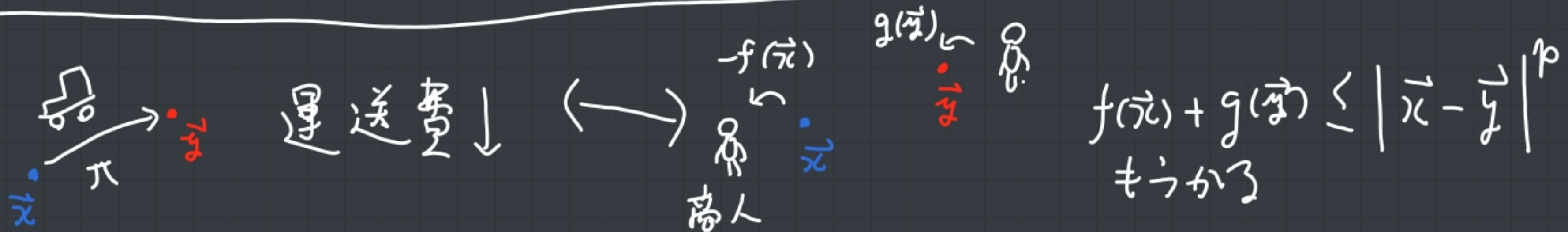
OT問題 3'

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

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

をみたすものの中を

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



1次元 Gauss 分布 の例

$p=2$ のケース

Step 1 f にあたりをつける

「 σ^2 」が「出るはず」... $\rightarrow f(x) = \alpha \cdot x^2$ としてみよう α を動かす

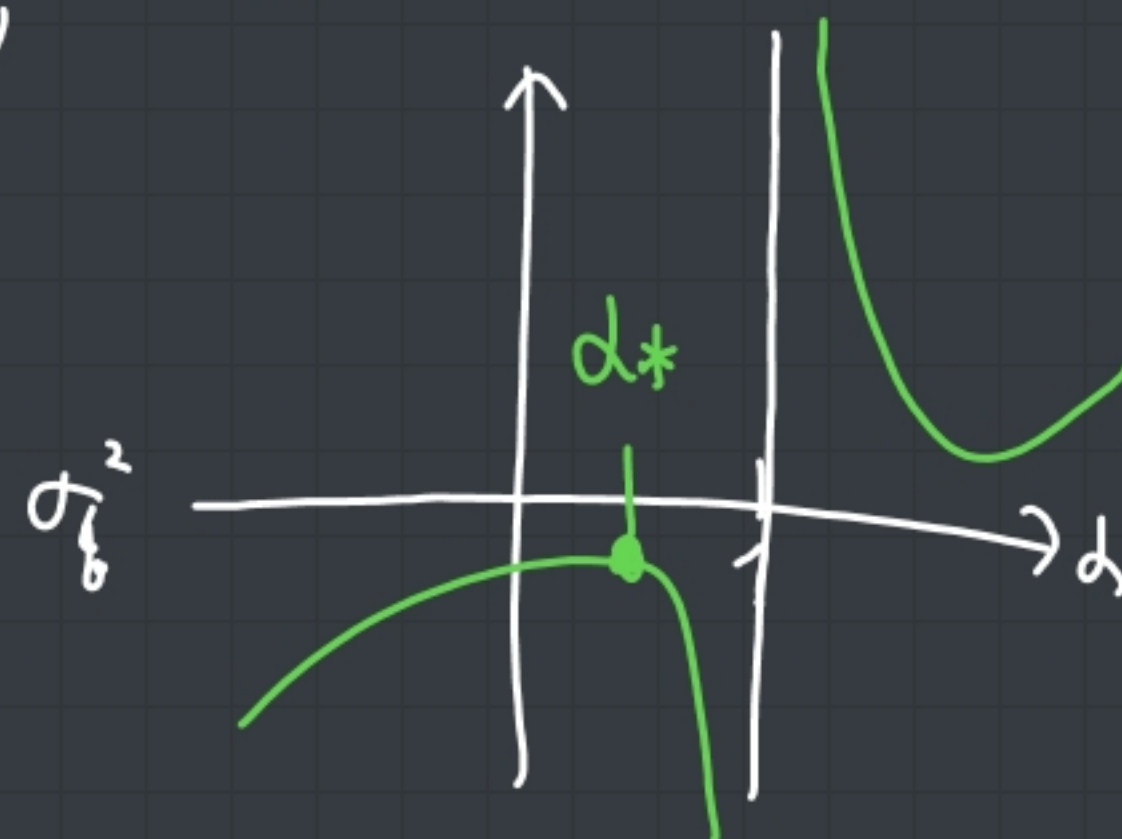
Step 2 最良のが「わかる」

$$g \text{ がよい} \rightarrow g(y) \leq (x-y)^2 - f(x) = \underbrace{(1-\alpha)}_{>0} \left(x - \frac{y}{1-\alpha}\right)^2 - \frac{y^2}{1-\alpha} + y^2$$

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

Step 3 $\alpha < 1$ まで max

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



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

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

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

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

$$\mathcal{L}(\pi, \underline{f}, \underline{g}) = \int |\underline{x} - \underline{y}|^p \pi(\underline{x}, \underline{y}) d\underline{x} d\underline{y} + \underbrace{\int \underline{f}(\underline{x}) \left\{ \underbrace{p(\underline{x})}_{\substack{0 \\ \wedge}} - \int \pi(\underline{x}, \underline{y}) d\underline{y} \right\} d\underline{x}}_{\substack{\infty \\ \downarrow}} + \underbrace{\int \underline{g}(\underline{y}) \left\{ \underline{q}(\underline{y}) - \int \pi(\underline{x}, \underline{y}) d\underline{x} \right\} d\underline{y}}_{\substack{0 \\ \vee}}$$

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

$$\stackrel{\substack{0 \\ \vee}}{=} \max_{f, g} \min_{\pi \geq 0} \mathcal{L}(\pi, f, g)$$

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

• $f^* = f^{**}$ $\min_{\pi} \max_{f, g} = \max_{f, g} \min_{\pi}$?

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

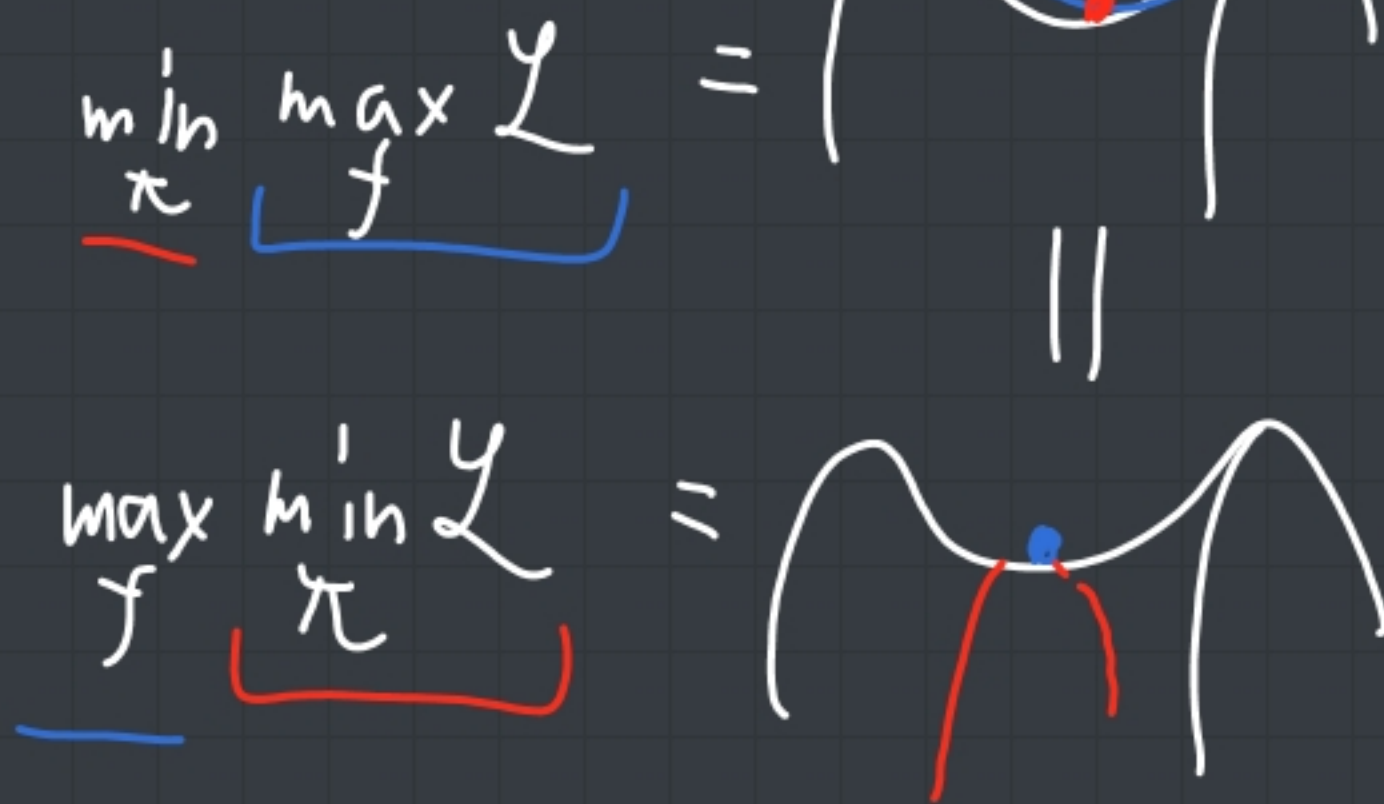
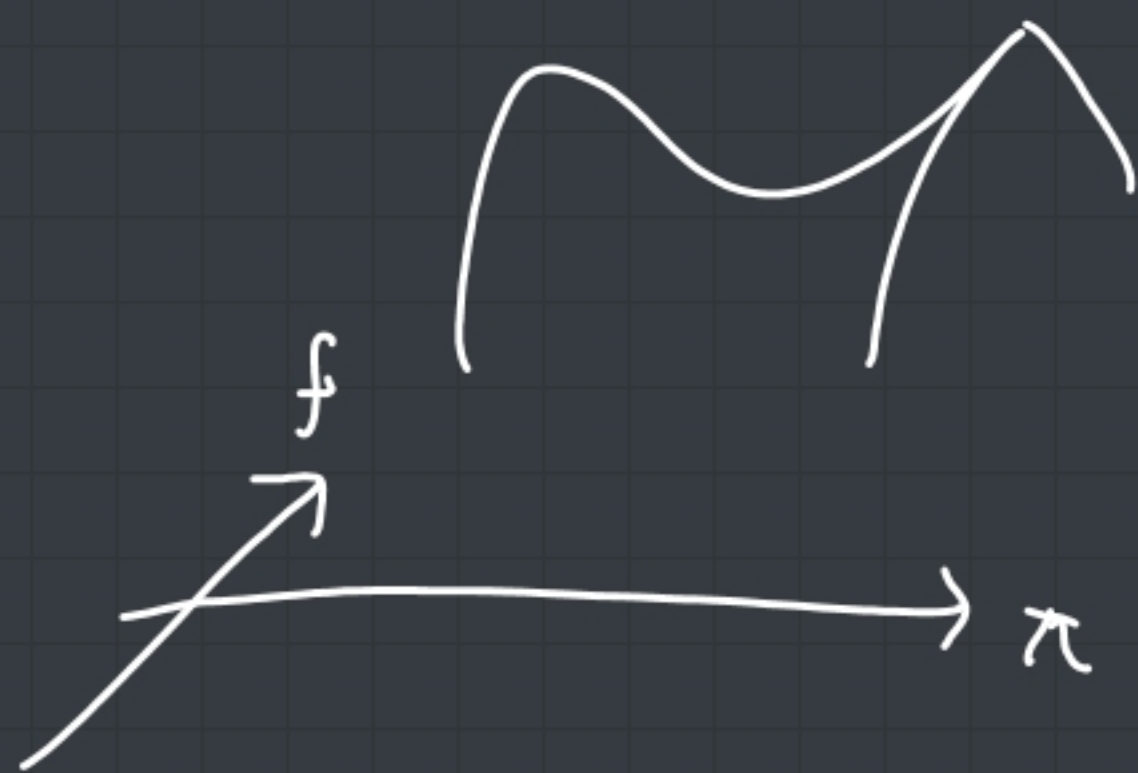
$$\Rightarrow \mathcal{L}(\pi, f, g) = \sum_{i,j} c_{ij} \pi_{ij} + \sum_i f_i \{p_i - \sum_j \pi_{ij}\} + \sum_j g_j \{x_j - \sum_i \pi_{ij}\}$$

さらに i, j なく g なく $C = P = 1$

$$\begin{aligned} \Rightarrow \mathcal{L}(\pi, f) &= \pi + f \{1 - \pi\} \\ &= \pi + f - f\pi \end{aligned}$$

$\rightarrow$ strong duality

$\pi^* + f^* \leq$



2-2 $p=1$ の π - π

$$\left| \min_z F(z) - \min_u G(u) \right| \quad (*)$$

$$\leq \max_z |F(z) - G(z)|$$

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 $\geq$ rhs

$$\langle h(\vec{x}) \rangle_P - \langle h(\vec{y}) \rangle_Q$$

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

$$\downarrow \max_h$$

$$\text{rhs} \leq \text{lhs}$$

lhs $\leq$ rhs $g^c(\vec{x}) = \min_{\vec{y}} \{ |\vec{x} - \vec{y}| - g(\vec{y}) \}$

$$1. \quad f(\vec{x}) \leq \underline{g^c(\vec{x})} \leq \{ |\vec{x} - \vec{x}| - g(\vec{x}) \} \leq -g(\vec{x})$$

$$2. \quad |g^c(\vec{x}) - g^c(\vec{y})| = \left| \min_{\vec{z}} \{ |\vec{x} - \vec{z}| - g(\vec{z}) \} - \min_{\vec{z}} \{ |\vec{y} - \vec{z}| - g(\vec{z}) \} \right|$$

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

$$\underline{lhs} \leq \underline{g^c} \leq \text{rhs}$$

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

$$\text{lhs} = \text{dual} \leq \text{rhs}$$

3. MLとの関連



3-1. GAN

arXiv:1406.2661 z"は

$$\min_{\mathcal{Q}} \max_{0 \leq h \leq 1}$$

$$\left\{ \langle \log h(\vec{x}) \rangle_P + \langle \log \{1 - h(\vec{y})\} \rangle_Q \right\}$$

$$D_{KL}(P \parallel \frac{P+Q}{2}) + D_{KL}(Q \parallel \frac{P+Q}{2}) - 2 \log 2$$

arXiv:1312.6114



3-2. VAE

$$\min_Q \min_{\pi} \{ \dots \}$$

→ VAE-like objective

WVAE

arXiv:1711.01558

$$\min_Q D_W(P, Q)$$

→ WGAN

arXiv:1701.07875

$$\max_{h \text{ Lip}} \{ \langle h \rangle - \langle h \rangle \}$$

→ 反論(?)

arXiv:1710.08446

→ z"のGAN?

why WVAE?

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

$$\approx \min_{\pi} \langle 1 + D(\pi, \mu) \rangle$$

$p=2 \Rightarrow$ Monge-Ampere
eq

$$\pi(x, z) = \underbrace{\pi(z|x)}_{\text{latent var model}} p(x)$$

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

$$\int \underbrace{\pi(x, z)}_{\parallel} dx = \boxed{q(y)}$$

$$\int \underbrace{q(y|z)}_{\parallel} \left(\int r(z|x) p(x) dx \right) dz$$

latent var model

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

$$\int q(\vec{z}) \underbrace{q(\vec{x}|\vec{z})} dz = q(\vec{x})$$