

Notes on Generalized Sliced Wasserstein Distance

Masanari Kimura
mkimura@ism.ac.jp

July 6, 2021

1 Wasserstein distances

$P_p(\Omega)$ を距離空間 (Ω, d) の上で定義される p 次モーメントが有限である Borel 確率測度の集合とする. また, $\mu \in P_p(X)$ および $\nu \in P_p(Y)$ をそれぞれ $X, Y \in \Omega$ の上で定義される確率測度とし, 対応する確率密度関数を I_μ, I_ν とする: $d\mu(x) = I_\mu(x)dx, d\nu(y) = I_\nu(y)dy$.

定義 1. (Wasserstein distance) ある $p \in [1, \infty]$ について, μ と ν の間の p -Wasserstein distance は最適輸送問題の解として定義される [Villani, 2008]:

$$W_p(\mu, \nu) := \left(\inf_{\gamma \in \Gamma(\mu, \nu)} \int_{X \times Y} d^p(x, y) d\gamma(x, y) \right)^{\frac{1}{p}}. \quad (1)$$

ここで $d^p(\cdot, \cdot)$ はコスト関数, $\Gamma(\mu, \nu)$ は全ての輸送計画の集合であり $\gamma \in \Gamma(\mu, \nu)$ は

$$\gamma(A \times Y) = \mu(A) \quad \text{for any Borel } A \subset X, \quad (2)$$

$$\gamma(X \times B) = \nu(B) \quad \text{for any Borel } B \subset Y \quad (3)$$

として与えられる.

Wasserstein distances の各性質については Appendix A に示す.

命題 2. p -Wasserstein distance は, $p \geq 2$ で以下の同値な表現を持つ:

$$W_p(\mu, \nu) = \left(\inf_{\varphi \in \Phi(\mu, \nu)} \int_X d^p(x, \varphi(x)) d\mu(x) \right)^{\frac{1}{p}}. \quad (4)$$

ここで $\Phi(\mu, \nu) = \{\varphi: X \rightarrow Y \mid \varphi_\# \mu = \nu\}$ であり, $\varphi_\# \mu$ は測度 μ の押し出しである:

$$\int_A d\varphi_\# \mu(y) = \int_{\varphi^{-1}(A)} d\mu(x) \quad \text{for any Borel subset } A \subset Y. \quad (5)$$

Proof. Brenier の定理 [Brenier, 1991] から直ちに従う (Appendix B). \square

命題 3. (Closed form of Wasserstein distances) 一次元連続確率測度を考える. I_μ と I_ν の累積分布関数をそれぞれ $F_\mu(x) = \mu((-\infty, x])$, $F_\nu(x) = \nu((-\infty, x])$ とする. このとき p -Wasserstein distance は以下の解析解を持つ:

$$W_p(\mu, \nu) = \left(\int_X d^p(x, F_\nu^{-1}(F_\mu(x))) d\mu(x) \right)^{\frac{1}{p}} = \left(\int_0^1 d^p(F_\mu^{-1}(z), F_\nu^{-1}(z)) dz \right)^{\frac{1}{p}}. \quad (6)$$

Proof. 最適輸送写像 $\varphi(x) = F_\nu^{-1}F_\mu(x)$ として一意に定まることから証明が得られる. \square

一次元確率分布において Wasserstein distances が解析解を持つという性質は非常に嬉しい. もし多次元確率分布を一次元表現にうまく変換できれば, Wasserstein distances を解析的に計算できるだろう, というのが sliced Wasserstein distances の基本的なアイディアである. 以下ではまず初めに Radon transform を導入し, それを用いて sliced Wasserstein distances を導出する.

2 Radon transform, sliced Wasserstein distances

以下の関数集合を考える:

$$L^1(\mathbb{R}^d) = \left\{ I : \mathbb{R}^d \rightarrow \mathbb{R} \mid \int_{\mathbb{R}^d} |I(x)| dx < \infty \right\}. \quad (7)$$

このとき Radon transform $\mathcal{R} : L^1(\mathbb{R}^d) \rightarrow L^1(\mathbb{R}^d \times \mathbb{S}^{d-1})$ は以下で定義される:

$$\mathcal{R}I(t, \theta) = \int_{\mathbb{R}^d} I(x) \delta(t - \langle x, \theta \rangle) dx. \quad (8)$$

ここで $(t, \theta) \in \mathbb{S}^{d-1}$ であり, $\mathbb{S}^{d-1} \subset \mathbb{R}^d$ は d 次元単位球を表す.

Radon transform は線形全単射であり [Natterer, 2001, Helgason, 2011], その逆写像 \mathcal{R}^{-1} は Fourier 変換 $\mathcal{F}\eta(\omega) = c|\omega|^{d-1}$ に対応するハイパスフィルタ $\eta(\cdot)$ を用いて

$$I(x) = \mathcal{R}^{-1}(\mathcal{R}I(t, \theta)) = \int_{\mathbb{S}^{d-1}} (\mathcal{R}I(\langle x, \theta \rangle, \theta) * \eta(\langle x, \theta \rangle)) d\theta \quad (9)$$

と書ける. このような逆写像は filtered back-projection と呼ばれる.

前述の通り, sliced Wasserstein distances のアイディアは以下の通りである:

1. 高次元確率分布の一次元表現を Radon transform によって獲得;
2. p -Wasserstein distance によって一次元表現を介して確率分布同士の距離を計算する.

これに基づいて, sliced p -Wasserstein distances は以下のように定義される.

定義 4. sliced Wasserstein distances ある $p \in [1, \infty]$ について, μ と ν の間の sliced p -Wasserstein distance は Radon transform \mathcal{R} を用いて

$$SW_p(I_\mu, I_\nu) := \left(\int_{\mathbb{S}^{d-1}} W_p^p(\mathcal{R}I_\mu(\cdot, \theta), \mathcal{R}I_\nu(\cdot, \theta)) d\theta \right)^{\frac{1}{p}}. \quad (10)$$

Wasserstein distances と同様に, sliced p -Wasserstein distances は正定値性, 対称性および三角不等式を満たすため距離関数となる [Bonnotte, 2013, Kolouri et al., 2016].

実用上は, MCMC などによって積分を有限和で近似することで計算される:

$$SW_p(I_\mu, I_\nu) \approx \left(\frac{1}{L} \sum_{l=1}^L W_p^p(\mathcal{R}I_\mu(\cdot, \theta_l), \mathcal{R}I_\nu(\cdot, \theta_l)) \right)^{\frac{1}{p}}. \quad (11)$$

高次元の場合, sliced Wasserstein distances は二つの確率測度の間の距離を小さく見積もってしまうことが知られている. この問題に対するヒューリスティックな対処法として, 以下の亜種が提案されている [Deshpande et al., 2019].

$$\max\text{-}SW_p(I_\mu, I_\nu) := \max_{\theta \in \mathbb{S}^{d-1}} W_p(\mathcal{R}I_\mu(\cdot, \theta), \mathcal{R}I_\nu(\cdot, \theta)). \quad (12)$$

3 Generalized Sliced Wasserstein Distance

上述の通り, sliced Wasserstein distances は高次元確率分布の一次元表現を線型写像によって獲得したのち, Wasserstein distances の解析解を用いて距離を計算していた. そこで, 確率分布の一次元表現を非線形変換によって計算するような sliced Wasserstein distances の一般化を考える.

3.1 Generalized Radon Transform

まず, $(X \subset \mathbb{R}^d) \times \mathbb{R}^n \setminus \{0\}$ の上のある関数 g を以下のように定義する.

1. g は C^∞ 級の実数値関数;
2. $\forall \lambda \in \mathbb{R}, g(x, \lambda\theta) = \lambda g(x, \theta)$;
3. $\forall (x \times \theta) \in X \times \mathbb{R}^n \setminus \{0\}, \frac{\partial g}{\partial x}(x, \theta) \neq 0$;
4. $\det\left(\left(\frac{\partial^2 g}{\partial x_i \partial \theta_j}\right)_{ij}\right) > 0$.

このとき, generalized Radon transform (GRT) \mathcal{G} は関数 g を用いて以下のように定義される [Beylkin, 1984, Denisyuk, 1994]:

$$\mathcal{G}I(t, \theta) = \int_{\mathbb{R}^d} I(x) \delta(t - g(x, \theta)) dx. \quad (13)$$

このとき Radon transform は GRT において $g(x, \theta) = \langle x, \theta \rangle$ とした特殊形になる.

3.2 Generalized Sliced-Wasserstein, Max-Generalized Sliced-Wasserstein Distances

元の sliced Wasserstein distances の Radon transform を GRT に差し替えることで、以下のように generalized sliced Wasserstein distances (GSW) が定義できる [Kolouri et al., 2019] :

$$GSW_p(I_\mu, I_\nu) := \left(\int_{\Omega_\theta} W_p^p(\mathcal{G}I_\mu(\cdot, \theta), \mathcal{G}I_\nu(\cdot, \theta)) d\theta \right)^{\frac{1}{p}}. \quad (14)$$

ここで Ω_θ は $g(\cdot, \theta)$ の実行可能パラメータのコンパクト集合. GSW もまた高次元空間において距離が低く見積もられる問題があるため、以下のように max-GSW を定義できる.

$$\text{max-GSW}_p := \max_{\theta \in \Omega_\theta} W_p(\mathcal{G}I_\mu(\cdot, \theta), \mathcal{G}I_\nu(\cdot, \theta)). \quad (15)$$

命題 5. GSW および max-GSW はどちらも $P_p(\Omega)$ の上の距離関数になる.

References

- [Beylkin, 1984] Beylkin, G. (1984). The inversion problem and applications of the generalized radon transform. *Communications on pure and applied mathematics*, 37(5):579–599.
- [Bonnotte, 2013] Bonnotte, N. (2013). *Unidimensional and evolution methods for optimal transportation*. PhD thesis, Paris 11.
- [Brenier, 1991] Brenier, Y. (1991). Polar factorization and monotone rearrangement of vector-valued functions. *Communications on pure and applied mathematics*, 44(4):375–417.
- [Denisyuk, 1994] Denisyuk, A. (1994). Inversion of the generalized radon transform. *Translations of the American Mathematical Society-Series 2*, 162:19–32.
- [Deshpande et al., 2019] Deshpande, I., Hu, Y.-T., Sun, R., Pyrros, A., Siddiqui, N., Koyejo, S., Zhao, Z., Forsyth, D., and Schwing, A. G. (2019). Max-sliced wasserstein distance and its use for gans. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10648–10656.
- [Helgason, 2011] Helgason, S. (2011). *Integral Geometry and Radon Transform*. Springer.
- [Kolouri et al., 2019] Kolouri, S., Nadjahi, K., Simsekli, U., Badeau, R., and Rohde, G. (2019). Generalized sliced wasserstein distances. *Advances in Neural Information Processing Systems*, 32:261–272.

- [Kolouri et al., 2016] Kolouri, S., Zou, Y., and Rohde, G. K. (2016). Sliced wasserstein kernels for probability distributions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5258–5267.
- [Natterer, 2001] Natterer, F. (2001). *The mathematics of computerized tomography*. SIAM.
- [Villani, 2008] Villani, C. (2008). Optimal transport, old and new. notes for the 2005 saint-flour summer school. *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer.

A Properties of Wasserstein distances

命題 6. Wasserstein distance $W_p : P_p(X) \times P_p(X) \rightarrow [0, \infty)$ は P_p の計量である。

Proof. $W_p(\mu, \nu) \geq 0$ ($\forall \mu, \nu \in P_p(X)$) は明らかである。コスト関数 $c(x, y) = |x - y|^p$ および $\gamma \in \Gamma(\mu, \nu) \Leftrightarrow S_\# \gamma \in \Gamma(\mu, \nu)$ where ($S(x, y) = (y, x)$) より, W_p の対称性も得られる。もし $\mu = \nu$ のとき, $x = y$ で $\gamma(x, y) = \delta_x(y)\mu(x)$ をほとんど至る所

$$W_p(\mu, \nu) \leq \int_{X \times X} |x - y|^p d\gamma(x, y) = 0 \quad (16)$$

ととることができる。つぎに, もし $W_p(\mu, \nu) = 0$ のとき, ほとんど至る所 $x = y$ となる $\gamma \in \Gamma(\mu, \nu)$ が存在する。従って, 任意の関数 $f : X \rightarrow \mathbb{R}$ について,

$$\int_X f(x) d\mu(x) = \int_{X \times X} f(x) d\gamma(x, y) = \int_{X \times Y} f(y) d\gamma(x, y) = \int_Y f(y) d\nu(y) \quad (17)$$

が成り立つので, $\mu = \nu$ である。□

命題 7. すべての $p \in [1, +\infty)$ と任意の $\mu, \nu \in P_p(X)$ について, $W_p(\mu, \nu) \geq W_1(\mu, \nu)$ が成り立つ。さらに $X \subset \mathbb{R}^n$ が有界であるとき, $W_p(\mu, \nu) \leq \text{diam}(X)^{p-1} W_1(\mu, \nu)$ が成り立つ。

Proof. Jensen の不等式より, ある $\gamma \in \Gamma(\mu, \nu)$ について

$$\left(\int_{X \times X} |x - y|^p d\gamma(x, y) \right)^{\frac{1}{p}} \geq \int_{X \times X} |x - y| d\gamma(x, y) \quad (18)$$

が成り立つので, $W_p(\mu, \nu) \geq W_1(\mu, \nu)$ が証明される。

次に X が有界であるとする, $\forall x, y \in X$ について

$$|x - y|^p \leq \left(\max_{w, z \in X} |w - z|^{p-1} \right) |x - y| = \left(\text{diam}(X) \right)^{p-1} |x - y|. \quad (19)$$

従って,

$$\int_{X \times X} |x - y|^p d\gamma(x, y) \leq \left(\text{diam}(X) \right)^{p-1} \int_{X \times X} |x - y| d\gamma(x, y) \quad (20)$$

であることから, $W_p(\mu, \nu) \leq \text{diam}(X)^{p-1} W_1(\mu, \nu)$ が証明される。□

B Brenier's theorem

定理 8. Brenier's theorem ある $p \geq 2$ について $\mu \in P_p(X)$, $\nu \in P_p(Y)$ とする. このとき, コスト関数 $c(x, y) = \frac{1}{2}|x - y|^2$ に対して最適な輸送計画 $\gamma^\dagger \in \Gamma(\mu, \nu)$ は一意に定まり, 以下で与えられる:

$$\gamma^\dagger = \varphi_\# \mu, \quad (21)$$

ここで φ は μ -a.e. で定義される凸関数.

Proof. まず, 解の存在を示す. γ^\dagger を Kantorovich の最適輸送問題の最小化解とする.

$$\gamma^\dagger(A \times B) = \int_A \gamma^\dagger(B|x) d\mu(x)$$

と書くと, $\partial\varphi(x) = \{\nabla\varphi(x)\}$ for L -a.e. $x \in X$ であるので,

$$\text{supp}(\gamma^\dagger(\cdot|x)) \subset \{\nabla\varphi(x)\} \quad \text{for } \mu\text{-a.e. } x \in X$$

となる. このことから, $\gamma^\dagger(\cdot|x) = \partial\varphi(x)$ for μ -a.e. $x \in X$ もわかる. 以上より,

$$\gamma^\dagger = \varphi_\# \mu$$

と書ける最適輸送解が存在し,

$$\begin{aligned} \nu(B) &= \gamma^\dagger(\mathbb{R}^n \times B) \\ &= \varphi_\# \mu(\mathbb{R}^n \times B) \\ &= \mu((\text{Id} \times \nabla\varphi)^{-1}(\mathbb{R}^n \times B)) \\ &= \mu(\{x \mid (\text{Id} \times \nabla\varphi)(x) \in \mathbb{R}^n \times B\}) \\ &= \mu(\{x \mid \nabla\varphi(x) \in B\}) \\ &= (\nabla\varphi)_\# \mu(B) \end{aligned}$$

となることが示される. 次に, このような最適輸送解の一意性を示す. $\bar{\varphi}$ を別の凸関数とし, $\bar{\varphi}_\# \mu = \nu$ となると仮定する. $\bar{\varphi}_\# \mu$ もまた最適輸送解となることから,

$$\int_X \bar{\varphi} d\mu + \int_Y \bar{\varphi}^* d\nu = \int_X \varphi d\mu + \int_Y \varphi^* d\nu$$

である. これを変形すると,

$$\begin{aligned} \int_{X \times Y} \bar{\varphi}(x) + \bar{\varphi}^*(y) d\gamma^\dagger(x, y) &= \int_{X \times Y} \varphi(x) + \varphi^*(y) d\gamma^\dagger(x, y) \\ &= \int_{X \times Y} x \cdot y d\gamma^\dagger(x, y) \\ &= \int_{X \times Y} x \cdot y d\varphi_\# \mu(x, y) \\ &= \int_X x \cdot \nabla\varphi(x) d\mu(x) \end{aligned}$$

が得られる．同様に，

$$\int_{X \times Y} \bar{\varphi}(x) + \bar{\varphi}^*(y) \, d\gamma^\dagger(x, y) = \int_X \bar{\varphi}(x) + \bar{\varphi}^*(\nabla\varphi(x)) d\mu(x).$$

したがって，

$$\int_X \left\{ \bar{\varphi}(x) + \bar{\varphi}(\nabla\varphi(x)) - x \cdot \nabla\varphi(x) \right\} d\mu(x) = 0.$$

特に， $\bar{\varphi}(x) + \bar{\varphi}^*(\nabla\varphi(x)) - x \cdot \nabla\varphi(x) = 0$ for μ -a.e. x であり， $\nabla\varphi(x) \in \partial\bar{\varphi}(x)$ for μ -a.e. x であるので， $\nabla\varphi(x) = \nabla\bar{\varphi}(x)$ が示された． \square