数理解析研究所講究録
第 1916 巻 2014 年 172-185

最適輸送理論梗概

名古屋大学大学院多元数理科学研究科　高津　飛鳥*

Asuka TAKATSU
Graduate School of Mathematics, Nagoya University

0　ことはじめ

最適輸送理論/Wasserstein 幾何に関する参考文献を訊ねられたとき, 大抵は C. Villani の本 [14] を答えます. しかし著者も Preface で述べているようこの本では結果を一般の設定で完全に証明することを目指しているため, 特別な場合 (例えばユークリッド空間の場合) のみを知りたいときはやや煩雑に感じる議論が多々あります. また本は約 1000 ページあり, 読むのに気後れしてしまいます. そこで本稿では [14] における議論をユークリッド空間の場合に限って, 同作者の本 [13] を参考にしつつ, 差し出がましながら私見も交えつつ解説していきます. したがって本稿は [13, 14] を元にした概説であり, より完全な理論を求める人にはオリジナルを読むことをお勧めします. (概説なので 'である・だ' 調ではなく 'です・ます' 調にしました.) 本稿では完全なる最適輸送理論ではなく, あくまで触りとしての最適輸送理論を紹介していきたいと思います.

本稿の構成は以下の通りです: まず §1 では最適輸送問題とは何かを述べます. 平たく言うとこれは確率測度空間上の変分問題で, この変分問題の特別な場合は Wasserstein 距離関数と呼ばれる確率測度空間上の距離関数を導きます. 本稿ではこの特別な場合のみを議論していきます. §2 では変分問題の解の存在性を示します. 続く §3 ではこの特別な場合, Wasserstein 距離関数の定義といくつかの性質を紹介します. そして最後に §4 で, 正規分布族上における Wasserstein 幾何と情報幾何の違いを見て, 二つの幾何の関係を少し述べます.

1　最適輸送理論とは

最適輸送理論とは '物質をある場所から他の場所へ最小費用で移す' 理論です. (例えば, 18 世紀後半に G. Monge はどこかの土壌を削り取って運び, 城を守る盛り土を作ることを考えていたようです.) 例えば物質はユークリッド空間 $\mathbb{R}^d$ 上に存在するとします. また物質の質量は輸送の前後で不変とし, 簡単のため質量を 1 に正規化します. すると物質の分布を確率測度とみなすことができます. そこで最初の分布を μ, 最後の分布を ν で表すとし

*takatsu@math.nagoya-u.ac.jp

ます．大雑把に言うと，$d\mu(x)$ は位置 x にある物質の質量を表します．以下，断りが無い限り確率測度は常にボレルであるとします．(確率測度が分かり難い場合はルベーグ測度に関する密度関数，すなわち $\int_{\mathbb{R}^d} f dx = 1$ を満たす非負値関数 f を考え，$f(x)$ で位置 x にある物質の質量を表すとします．そして $d\mu(x) = f(x)dx$, $d\nu(y) = g(y)dy$ と理解します.) このとき 物質を μ を ν へ動かす輸送 T を考えます．輸送 T とは物質が位置 x から $y = T(x)$ に移ることなので，T は $\mathbb{R}^d$ から $\mathbb{R}^d$ 写像です．特に可測写像とします．(より厳密には，T は μ の台の上で定義されていれば十分です.) そして輸送 T により μ にある物質が全て ν に移るとします．このとき終点のある場所 A に運び込まれた物質は最初は $T^{-1}(A)$ にあったと考えられるので，輸送前の質量 $\mu[T^{-1}(A)]$ と輸送後の質量 $\nu[A]$ は等しいはずです．すなわち任意の可測集合 $A \subset \mathbb{R}^d$ に対し，$\nu[A] = \mu[T^{-1}(A)]$ が成立ちます．このようなとき，T は μ を ν に押出すと言い，$T_\sharp\mu = \nu$ と書き，そのような写像のなす集合を $\mathcal{T}(\mu, \nu)$ と書きます．また物質を位置 x から位置 y に運ぶのに費用 $c(x, y)$ が掛かるとします．すると c は $\mathbb{R}^d \times \mathbb{R}^d$ 上の非負実数値関数とみなせます．特に c は可測であるとします．位置 x にある物質の質量は $d\mu(x)$ なので位置 x の物質を $T(x)$ に全て移すには $c(x, T(x))d\mu(x)$ なる費用が掛かります．よって μ にある物質を ν に全て移すためには

$$C'(T) := \int_{\mathbb{R}^d} c(x, T(x)) d\mu(x)$$

なる総費用が掛かります．そして勿論，総費用は少ない方が良いので以下のような変分問題が考えられます．

問題 1.1 (*Monge*の問題) $\mathcal{T}(\mu, \nu)$ における $C'(T)$ の最小値および最小値を達成する元を求めよ．

Monge は元々 $c(x, y) = |x - y|$ の場合を考えていたそうですが，どのような費用関数を考えてもこの問題は実は well-defined ではありません．なぜなら μ を点 x に台を持つディラック測度 δ_x, ν を相異なる二点 y, z に台を持ち $\frac{1}{2}$ ずつ分布するディラック測度の和 $\frac{1}{2}(\delta_y + \delta_z)$ とすれば，$T_\sharp\mu = \nu$ を満たす可測写像 T は存在しない，つまり $\mathcal{T}(\mu, \nu) = \emptyset$ だからです．

そこで写像 T の代わりにカップリングを用います．$\mathbb{R}^d$ 上の確率測度 μ, ν に対し，μ, ν のカップリング π とは $\mathbb{R}^d \times \mathbb{R}^d$ 上の確率測度であり，その第一射影が μ, その第二射影が ν となる，すなわち可測集合 $A \subset \mathbb{R}^d$ に対し

$$\pi[A \times \mathbb{R}^d] = \mu[A], \qquad \pi[\mathbb{R}^d \times A] = \nu[A] \tag{1.1}$$

となるものです．

注記 1.2 条件 (1.1) は射影 $p_i : \mathbb{R}^d \times \mathbb{R}^d \ni (x_1, x_2) \mapsto x_i \in \mathbb{R}^d$ $(i = 1, 2)$ を使えば

$$p_{1\sharp}\pi = \mu, \qquad p_{2\sharp}\pi = \nu$$

と表されます．さらに，条件 (1.1) は $\mathbb{R}^d$ 上の任意の有界連続関数 h_1, h_2 に対して

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} (h_1(x) + h_2(y)) d\pi(x, y) = \int_{\mathbb{R}^d} h_1(x) d\mu(x) + \int_{\mathbb{R}^d} h_2(y) d\nu(y)$$

が成立つことと同値です．これらの同値条件については [14, Chapter 1] にも書いてありますが，そこでは試験関数は $h_1 \in L^1(\mu), h_2 \in L^1(\nu)$ で与えられています．しかしこれは一般の確率空間 (X, μ) で考えているからであり，[13, p.18] にあるように X が完備可分距離空間で μ がボレル確率測度ならば試験関数として有界連続関数を選ぶことができます．

同様に，$T_\sharp\mu = \nu$ は $\mathbb{R}^d$ 上の任意の有界連続関数 h に対して

$$\int_{\mathbb{R}^d} h(T(x))d\mu(x) = \int_{\mathbb{R}^d} h(y)d\nu(y)$$

が成立つことと同値です．さらに μ, ν がルベーグ測度に絶対連続であり，その密度関数がそれぞれ f, g であるとします．そして T が適切な正則性-例えば，C^1-微分同相–を持つならば，T のヤコビ行列式 J_T に対し

$$f(x) = g(T(x))|J_T(x)|$$

という変数変換が殆ど至る所成立つこととも同値です．

例えば，μ, ν による直積測度 $\mu \times \nu$ は明らかに条件 (1.1) を満たすので，カップリングとなります．よって μ, ν のカップリングがなす空間 $\Pi(\mu, \nu)$ は空集合ではありません．また任意の $T \in \mathcal{T}(\mu, \nu)$ に対し，$\mathbb{R}^d$ 上の恒等写像 $\mathrm{id}_{\mathbb{R}^d}$ と T の直積写像

$$(\mathrm{id}_{\mathbb{R}^d}, T) : \mathbb{R}^d \ni x \mapsto (x, T(x)) \in \mathbb{R}^d \times \mathbb{R}^d$$

による μ の押出測度 $(\mathrm{id}_{\mathbb{R}^d}, T)_\sharp\mu$ は μ, ν のカップリングになっています．実際，$\mathbb{R}^d$ 上の任意の有界連続関数 h_1, h_2 に対して

$$\begin{aligned}\int_{\mathbb{R}^d\times\mathbb{R}^d} (h_1(x) + h_2(y))d(\mathrm{id}_{\mathbb{R}^d}, T)_\sharp\mu(x, y) &= \int_{\mathbb{R}^d} (h_1(x) + h_2(T(x)))d\mu(x) \\ &= \int_{\mathbb{R}^d} h_1(x)d\mu(x) + \int_{\mathbb{R}^d} h_2(y)d\nu(y)\end{aligned}$$

が成立つので，条件 (1.1) が成立ちます．よって

$$\mathcal{T}(\mu, \nu) \ni T \mapsto (\mathrm{id}_{\mathbb{R}^d}, T)_\sharp\mu \in \Pi(\mu, \nu) \tag{1.2}$$

なる関係があります．そこで写像 T の代わりにカップリング π を用いて条件を緩和し，問題 1.1 の代わりに次の問題を考えます:

問題 1.3 (*Monge–Kantorovich* の問題) $\Pi(\mu, \nu)$ において

$$C(\pi) := \int_{\mathbb{R}^d\times\mathbb{R}^d} c(x, y)d\pi(x, y)$$

の最小値および最小値を達成する元を求めよ．

このような定式化は 1940 年代に L. Kantorovich に提唱されました．(Kantorovich は Monge の問題を知らずにこのような定式化をしたそうです．また Kantorovich は線形計画法に大きな進展を与えています.)

2 問題 1.3の解の存在性

カップリングがなす空間 $\Pi(\mu,\nu)$ は空集合でないため $\Pi(\mu,\nu)$ 上で汎関数 C の解析が可能となり, 問題 1.3を考えることができます. ところが C が $\Pi(\mu,\nu)$ 上で有限の値を取るか, ましてや最小値を達成する元が存在するかどうかは分かりません. (最小値ではなく下限である可能性もあるからです.) しかし '適切' な条件下では, C は $\Pi(\mu,\nu)$ 上で有限値でありさらに最小値を達成します. その適切な条件を考えるために, 以下 $c(x,y)=|x-y|^2$ とし, μ,ν の 2 次モーメントは有限, すなわち

$$\int_{\mathbb{R}^d}|x|^2d\mu(x)<\infty,\qquad \int_{\mathbb{R}^d}|y|^2d\nu(y)<\infty$$

とします. するとこのとき $c(x,y)=|x-y|^2\le 2(|x|^2+|y|^2)$ より任意の $\pi\in\Pi(\mu,\nu)$ に対して

$$\int_{\mathbb{R}^d\times\mathbb{R}^d}c(x,y)d\pi(x,y)=2\left(\int_{\mathbb{R}^d}|x|^2d\mu(x)+\int_{\mathbb{R}^d}|y|^2d\nu(y)\right)<\infty$$

となるので, C は $\Pi(\mu,\nu)$ 上で有限値を取ります. さらに次が成立ち C は $\Pi(\mu,\nu)$ 上で最小値を達成します.

定理 2.1 任意の2次モーメントが有限な $\mathbb{R}^d$ 上の確率測度 μ,ν に対し, ある $\pi_\infty\in\Pi(\mu,\nu)$ が存在して,

$$\inf_{\pi\in\Pi(\mu,\nu)}\int_{\mathbb{R}^d\times\mathbb{R}^d}|x-y|^2d\pi(x,y)=\int_{\mathbb{R}^d\times\mathbb{R}^d}|x-y|^2d\pi_\infty(x,y)<\infty$$

が成立する.

この最小化子のことを最適輸送と呼びます. より一般の状況における最適輸送の存在定理は [14, Theorem 4.1] で見つけられます. また最適輸送の存在は保証されましたが, その一意性は保証されません. そして問題 1.3 の解が存在しても問題 1.1 の解が存在しない場合もあります. 証明の前に例をみてみましょう.

例 2.2 任意の三点 $x,y,z\in\mathbb{R}^d$ に対し,

$$\mu:=\delta_x,\quad \nu:=\frac{1}{2}\left(\delta_y+\delta_z\right),\quad \pi_\infty:=\frac{1}{2}\left(\delta_{(x,y)}+\delta_{(x,z)}\right)=\mu\times\nu$$

とおきます. このとき $\Pi(\mu,\nu)=\{\pi_\infty\}$ なので, 問題 1.3 の最小化子は π_∞ です. そしてその最小値は

$$\inf_{\pi\in\Pi(\mu,\nu)}C(\pi)=\frac{1}{2}\left(|x-y|^2+|x-z|^2\right)$$

となります. 一方, $y\neq z$ を仮定すると前述の通り $\mathcal{T}(\mu,\nu)=\emptyset$ となるので, 問題 1.1 は well-defined ではありません.

例 **2.3** $\xi=(0,0),\eta=(1,0),\zeta=(0,1),\omega=(1,1)$ とし,

$$\mu=\frac{1}{2}(\delta_\xi+\delta_\omega),\qquad \nu=\frac{1}{2}(\delta_\eta+\delta_\zeta)$$

とします. また任意の $r\in[0,\frac{1}{2}]$ に対して

$$\pi_r:=r\cdot\delta_{(\xi,\eta)}+\left(\frac{1}{2}-r\right)\cdot\delta_{(\xi,\zeta)}+\left(\frac{1}{2}-r\right)\cdot\delta_{(\omega,\eta)}+r\cdot\delta_{(\omega,\zeta)} \tag{2.1}$$

は μ,ν のカップリングとなります. 逆に任意の μ,ν のカップリングは必ず (2.1) の形で表されます. そして

$$\begin{aligned}C(\pi_r)&=\int_{\mathbb{R}^2\times\mathbb{R}^2}|x-y|^2d\pi_r(x,y)\\&=r|\xi-\eta|^2+\left(\frac{1}{2}-r\right)|\xi-\zeta|^2+\left(\frac{1}{2}-r\right)|\omega-\eta|^2+r|\omega-\zeta|^2\\&=1\end{aligned}$$

なので, 任意のカップリングが最適輸送となります.

また, 問題 1.1 の解を考えてみましょう. μ,ν がディッラク測度の線型和で表されるので μ を ν に押出す写像は

$$\begin{cases}T_0(\xi)=\zeta,\\T_0(\omega)=\eta,\end{cases}\qquad\begin{cases}T_{1/2}(\xi)=\eta,\\T_{1/2}(\omega)=\zeta\end{cases}$$

の二つに限ります. このとき

$$\begin{aligned}C'(T_0)&=\int_{\mathbb{R}^2\times\mathbb{R}^2}|x-T_0(x)|^2d\mu(x)=\frac{1}{2}(|\xi-T_0(\xi)|^2+|\omega-T_0(\omega)|^2)=1,\\C'(T_{1/2})&=\int_{\mathbb{R}^2\times\mathbb{R}^2}|x-T_{1/2}(x)|^2d\mu(x)=\frac{1}{2}(|\xi-T_{1/2}(\xi)|^2+|\omega-T_{1/2}(\omega)|^2)=1\end{aligned}$$

となるので, T_0 も $T_{1/2}$ も問題 1.1 の最小値を与えます.

そして上の計算や $(\mathrm{id}_{\mathbb{R}^2},T_0)_\sharp\mu=\pi_0$, $(\mathrm{id}_{\mathbb{R}^2},T_{1/2})_\sharp\mu=\pi_{1/2}$ となる事実より,

$$\inf_{\pi\in\Pi(\mu,\nu)}C(\pi)=\inf_{T\in\mathcal{T}(\mu,\nu)}C'(T)=1$$

が成立ちます. よって問題 1.1 も問題 1.3 も最小化子を持ち, その最小値は一致します. 一方で任意の $r\in(0,\frac{1}{2})$ に対し, $(\mathrm{id}_{\mathbb{R}^2},T_r)_\sharp\mu=\pi_r$ となるような $T_r\in\mathcal{T}(\mu,\nu)$ は存在しないので, 問題 1.3 の任意の最小化子が問題 1.1 の最小化子となり得るわけではありません.

定理 2.1 の証明の鍵は
(1) $\Pi(\mu,\nu)$ は弱位相で点列コンパクト
(2) C は弱位相で下半連続

ということです．この事実を認めると，問題 1.3 の最小化列 $\{\pi_n\}_{n\in\mathbb{N}}$ に対して，(1) より弱収束部分列 $\{\pi_{n_k}\}_{k\in\mathbb{N}}$ とその弱極限 $\pi_\infty \in \Pi(\mu,\nu)$ が存在します．そして (2) より

$$\inf_{\pi\in\Pi(\mu,\nu)} C(\pi) \le C(\pi_\infty) \le \liminf_{k\to\infty} C(\pi_{n_k}) = \inf_{\pi\in\Pi(\mu,\nu)} C(\pi).$$

となり，π_∞ が最小化子であることが分かります．

ここで $\mathbb{R}^d$ 上の確率測度列 $\{\mu_n\}_{n\in\mathbb{N}}$ が $\mathbb{R}^d$ 上の確率測度 μ に弱収束するとは，$\mathbb{R}^d$ 上の任意の有界連続関数 h に対し

$$\lim_{n\to\infty}\int_{\mathbb{R}^d} h(x)d\mu_n(x) = \int_{\mathbb{R}^d} h(x)d\mu(x)$$

が成立つことです．そしてこの収束から導かれる位相を弱位相と呼びます．

証明．**(定理 2.1)** <u>(1) $\Pi(\mu,\nu)$ が弱位相で点列コンパクトであること</u>：任意の確率測度列 $\{\pi_n\}_{n\in\mathbb{N}} \subset \Pi(\mu,\nu)$ を考えます．まず，$\{\pi_n\}_{n\in\mathbb{N}}$ が弱収束部分列を持つことを示します．これは下で (証明抜きで) 述べる Prokhorov の定理より，任意の $\varepsilon > 0$ に対し，あるコンパクト集合 $K^\varepsilon \subset \mathbb{R}^d \times \mathbb{R}^d$ が存在して，任意の $\pi \in \Pi(\mu,\nu)$ に対して $\pi[(\mathbb{R}^d\times\mathbb{R}^d)\setminus K^\varepsilon] \le \varepsilon$ となることが必要十分です．

ここで原点を中心とする半径 $R > 0$ の閉球 B_R に対し $\lim_{R\to\infty}\mu[B_R] = 1$ が成立つので，任意の $\varepsilon > 0$ に対しあるコンパクト集合 $K_1^\varepsilon \subset \mathbb{R}^d$ が存在して $\mu[\mathbb{R}^d\setminus K_1^\varepsilon] \le \varepsilon/2$ が成立ちます．同様に任意の $\varepsilon > 0$ に対し，あるコンパクト集合 $K_2^\varepsilon \subset \mathbb{R}^d$ が存在して，$\nu[\mathbb{R}^d\setminus K_2^\varepsilon] \le \varepsilon/2$ も成立ちます．そこで $K^\varepsilon := K_1^\varepsilon \times K_2^\varepsilon \subset \mathbb{R}^d\times\mathbb{R}^d$ と定めれば，K^ε はコンパクトであり任意の $\pi \in \Pi(\mu,\nu)$ に対して

$$\pi[(\mathbb{R}^d\times\mathbb{R}^d)\setminus K^\varepsilon] \le \pi[(\mathbb{R}^d\setminus K_1^\varepsilon)\times\mathbb{R}^d] + \pi[\mathbb{R}^d\times(\mathbb{R}^d\setminus K_2^\varepsilon)] = \mu[\mathbb{R}^d\setminus K_1^\varepsilon] + \nu[\mathbb{R}^d\setminus K_2^\varepsilon] \le \varepsilon$$

が成立ちます．よって $\{\pi_n\}_{n\in\mathbb{N}}$ の弱収束部分列 $\{\pi_{n_k}\}_{k\in\mathbb{N}}$ が存在して，$\mathbb{R}^d\times\mathbb{R}^d$ 上の確率測度 π_∞ に収束します．

次に $\pi_\infty \in \Pi(\mu,\nu)$ を示します．ここで $\mathbb{R}^d$ 上の任意の有界連続関数 h_1, h_2 に対し，$h(x,y) := h_1(x) + h_2(y)$ もまた $\mathbb{R}^d\times\mathbb{R}^d$ 上の有界連続関数なので，

$$\begin{aligned}
\int_{\mathbb{R}^d\times\mathbb{R}^d} h(x,y)d\pi_\infty(x,y) &= \lim_{k\to\infty}\int_{\mathbb{R}^d\times\mathbb{R}^d} h(x,y)d\pi_{n_k}(x,y)\\
&= \lim_{k\to\infty}\int_{\mathbb{R}^d\times\mathbb{R}^d} (h_1(x)+h_2(y))d\pi_{n_k}(x,y)\\
&= \lim_{k\to\infty}\left(\int_{\mathbb{R}^d} h_1(x)d\mu(x) + \int_{\mathbb{R}^d} h_2(y)d\nu(y)\right)\\
&= \int_{\mathbb{R}^d} h_1(x)d\mu(x) + \int_{\mathbb{R}^d} h_2(y)d\nu(y)
\end{aligned}$$

が成立ちます．これより $\pi_\infty \in \Pi(\mu,\nu)$ が示され，$\Pi(\mu,\nu)$ が弱位相に関して点列コンパクトであることが示されました．

(2) $\mathbb{R}^d \times \mathbb{R}^d$ 上の確率測度がなす集合上で C が弱位相で下半連続であること: $\mathbb{R}^d\times\mathbb{R}^d$ 上の確率測度列 $\{\pi_n\}_{n\in\mathbb{N}}$ が $\mathbb{R}^d \times \mathbb{R}^d$ 上の確率測度 π_∞ に弱収束しているとします. このとき

$$C(\pi_\infty) \le \liminf_{n\to\infty} C(\pi_n)$$

が成立つことを示します. 任意の $k \in \mathbb{N}$ に対して,

$$c_k(x,y) := \min\{|x-y|^2, k\}$$

とおけば, $\{c_k\}_{k\in\mathbb{N}}$ は $\mathbb{R}^d \times \mathbb{R}^d$ 上の非負可測関数で, 任意の点 $(x,y) \in \mathbb{R}^d \times \mathbb{R}^d$ に対して

$$c_k(x,y) \le c_{k+1}(x,y) \le |x-y|^2 = \lim_{k\to\infty} c_k(x,y)$$

が成立ちます. また, c_k は $\mathbb{R}^d \times \mathbb{R}^d$ 上の有界連続関数でもありますので弱収束の性質より

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} c_k(x,y)d\pi_\infty(x,y) = \lim_{n\to\infty}\int_{\mathbb{R}^d\times\mathbb{R}^d} c_k(x,y)d\pi_n(x,y) = \liminf_{n\to\infty}\int_{\mathbb{R}^d\times\mathbb{R}^d} c_k(x,y)d\pi_n(x,y)$$

も成立ちます. よって, 単調収束定理より

$$\begin{aligned}
C(\pi_\infty) = \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2 d\pi_\infty(x,y) &= \lim_{k\to\infty}\int_{\mathbb{R}^d\times\mathbb{R}^d} c_k(x,y)d\pi_\infty(x,y) \\
&= \lim_{k\to\infty}\liminf_{n\to\infty}\int_{\mathbb{R}^d\times\mathbb{R}^d} c_k(x,y)d\pi_n(x,y) \\
&\le \lim_{k\to\infty}\liminf_{n\to\infty}\int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2 d\pi_n(x,y) \\
&= \liminf_{n\to\infty}\int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2 d\pi_n(x,y) = \liminf_{n\to\infty} C(\pi_n)
\end{aligned}$$

となり, 示されました. □

定理 2.4 (*Prokhorov* の定理) $\mathcal{P}$ を $\mathbb{R}^d$ 上の確率測度のなす空間の部分集合とする. このとき, $\mathcal{P}$ の弱位相に関する相対点列コンパクト性と, $\mathcal{P}$ の緊密性, すなわち任意の $\varepsilon > 0$ に対し, あるコンパクト集合 $K^\varepsilon \subset \mathbb{R}^d$ が存在し, 任意の $\mu \in \mathcal{P}$ に対し $\mu[\mathbb{R}^d \setminus K^\varepsilon] \le \varepsilon$ が成立つことは同値である.

証明は例えば [3, Theorem 8.6.2] で見つけられます. ([13, 14] に主張と参考文献はありますが証明はありません.)

このようにして問題 1.3 の最小化子の存在が示されましたが, 一方で例 2.3 のように解が一意となるとは限りません. また $\mathcal{T}(\mu,\nu) \neq \emptyset$ ならば (1.2) より,

$$\inf_{\pi\in\Pi(\mu,\nu)} C(\pi) \le \inf_{T\in\mathcal{T}(\mu,\nu)} C'(T) \tag{2.2}$$

が成立ちますが, 例 2.2 のように問題 1.3 の最小化子が存在しても問題 1.1 の最小化子として実現されるとは限りません. さらに問題 1.1 の最小化子が存在した場合に, (2.2) で等

号が成立つかも分かりません. そこで問題 1.3 の解が一意であり, さらに問題 1.1 の解を与え, (2.2) が等号になるための十分条件を (証明抜きで) 述べて本節を閉じたいと思います. 主張は本 [14] に掲載されている形とは異なる上に $\mathbb{R}^d$ 上で $c(x,y) = |x-y|^2$ の場合に限っています. より一般の設定下における結果は例えば [14, Theorem 5.10] で見つけられます.

主張を述べるために, 事実を紹介します. この事実のより一般の場合とその証明は, 例えば [14, Theorem 10.8] を参考にして下さい.

命題 2.5 (Rademacher の定理, cf.[13, 2.1.3]) φ を $\mathbb{R}^d$ 上の適正凸関数, すなわち $\mathbb{R}\cup\{\infty\}$ に値を取るが恒等的に $+\infty$ ではなく, そして任意の $x, y \in \mathbb{R}^d$ と $t \in [0,1]$ に対して

$$\varphi(tx + (1-t)y) \leq t\varphi(x) + (1-t)\varphi(y)$$

が成立つ, とする. このとき φ の勾配 $\nabla\varphi$ は殆ど至る所で定義される.

定理 2.6 ([13, Theorems 2.12, 2.16, 5.6]) μ, ν を $\mathbb{R}^d$ 上の二次モーメントが有限な確率測度とする.
(1) μ がルベーグ測度に対して絶対連続ならば, μ, ν の最適輸送 π が一意的に存在し, さらに $\mathbb{R}^d$ 上の適正凸下半連続関数 φ を用いて

$$\pi = (\mathrm{id}_{\mathbb{R}^d}, \nabla\varphi)_\sharp\mu$$

と表現される. ここで, $\nabla\varphi$ は μ に対して殆ど至る所一意的に存在し, $c(x,y) = |x-y|^2$ に対する問題 *1.1* の一意的な最小化子となる.
(2) $\mathbb{R}^d$ 上のある適正凸下半連続関数 φ が存在して $\nabla\varphi_\sharp\mu = \nu$ を満たせば, $\pi = (\mathrm{id}_{\mathbb{R}^d}, \nabla\varphi)_\sharp\mu$ は μ と ν の最適輸送である. さらに $\nabla\varphi$ は $c(x,y) = |x-y|^2$ に対する問題 *1.1* の最小化子である.
(3) μ がルベーグ測度に対して絶対連続とし, $T: \mathbb{R}^d \to \mathbb{R}^d$ を $c(x,y) = |x-y|^2$ に対する問題 *1.1* の最小化子とする.
このとき, $t \in [0,1]$ に対し,

$$T_t := (1-t)\,\mathrm{id}_{\mathbb{R}^d} + tT, \qquad \mu_t = T_{t\sharp}\mu$$

とすれば, $\pi_t := (\mathrm{id}_{\mathbb{R}^d}, T_t)_\sharp\mu \in \Pi(\mu, \mu_t)$ は最適輸送であり, $C(\pi_t) = t^2 C(\pi)$ が成立つ.

3 Wasserstein 幾何

以下, $\mathbb{R}^d$ 上の確率測度のなす集合を $\mathcal{P}(\mathbb{R}^d)$, ルベーグ測度に絶対連続な確率測度のなす集合を $\mathcal{P}^{\mathrm{ac}}(\mathbb{R}^d)$, 2 次モーメントが有限な確率測度のなす集合を $\mathcal{P}_2(\mathbb{R}^d)$ と書くことにします. そして $\mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d) := \mathcal{P}_2(\mathbb{R}^d) \cap \mathcal{P}^{\mathrm{ac}}(\mathbb{R}^d)$ とします. すると前節で見たように, 任意の $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ に対して

$$W_2(\mu,\nu) := \left(\inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2 d\pi(x,y)\right)^{\frac{1}{2}} = \left(\min_{\pi\in\Pi(\mu,\nu)} \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2 d\pi(x,y)\right)^{\frac{1}{2}}$$

は非負有限確定値になります. $\mathcal{P}_2(\mathbb{R}^d) \times \mathcal{P}_2(\mathbb{R}^d)$ 上の汎関数 W_2 は $(L^2\text{-})$Wasserstein 距離関数と呼ばれます. これは実際に $\mathcal{P}_2(\mathbb{R}^d)$ 上の距離関数です.

定理 3.1 (cf. [14, Definition 6.1]) $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ は距離空間である.

証明. (1) 非退化性: 任意の $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ に対し, $\pi := (\mathrm{id}_{\mathbb{R}^d}, \mathrm{id}_{\mathbb{R}^d})_\sharp \mu \in \Pi(\mu,\mu)$ なので

$$0 \le W_2(\mu,\mu) \le \left(\int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2 d\pi(x,y)\right)^{\frac{1}{2}} = \left(\int_{\mathbb{R}^d} |x-x|^2 d\mu(x)\right)^{\frac{1}{2}} = 0$$

となります. よって, $W_2(\mu,\mu)=0$ です. 逆にある $\mu,\nu \in \mathcal{P}_2(\mathbb{R}^d)$ が存在して $W_2(\mu,\nu)=0$ とすれば, ある $\pi \in \Pi(\mu,\nu)$ が存在して,

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2 d\pi(x,y) = 0$$

となります. そして $|x-y|^2$ の非負性より, π の台は対角集合 $\{(x,x) \mid x \in \mathbb{R}^d\}$ に含まれます. よって $\mathbb{R}^d$ 上の任意の有界連続関数 h に対して

$$\int_{\mathbb{R}^d} h(x)d\mu(x) = \int_{\mathbb{R}^d\times\mathbb{R}^d} h(x)d\pi(x,y) = \int_{\mathbb{R}^d\times\mathbb{R}^d} h(y)d\pi(x,y) = \int_{\mathbb{R}^d} h(y)d\nu(y)$$

が成立します. すなわち $\nu=\mu$ が成立ち, 非退化性が従います.
(2) 対称性: 写像 $R: \mathbb{R}^d\times\mathbb{R}^d \ni (x,y) \mapsto (y,x) \in \mathbb{R}^d\times\mathbb{R}^d$ を考えます. このとき

$$\Pi(\mu,\nu) \ni \pi \mapsto R_\sharp\pi \in \Pi(\nu,\mu)$$

は全単射です. よって $|x-y|^2 = |y-x|^2$ と併せて

$$\begin{aligned} W_2(\nu,\mu)^2 &= \inf_{\pi\in\Pi(\nu,\mu)} \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2 d\pi(x,y) \\ &= \inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2 d(R_\sharp\pi)(x,y) \\ &= \inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathbb{R}^d\times\mathbb{R}^d} |y-x|^2 d\pi(x,y) = W_2(\mu,\nu)^2 \end{aligned}$$

となり, 対称性が分かります.
(3) 三角不等式: 任意の $\mu_1,\mu_2,\mu_3 \in \mathcal{P}_2(\mathbb{R}^d)$ に対し, $\pi_{12} \in \Pi(\mu_1,\mu_2), \pi_{23} \in \Pi(\mu_2,\mu_3)$ を最適輸送とします. ここで以下で述べる貼合わせの補題を認めると, ある $\pi \in \mathcal{P}_2(\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d)$ が存在して, 任意の可測集合 $A \subset \mathbb{R}^d\times\mathbb{R}^d$ に対して,

$$\pi[A\times\mathbb{R}^d] = \pi_{12}[A], \qquad \pi[\mathbb{R}^d\times A] = \pi_{23}[A]$$

が成立ちます. よって $\pi_{13} \in \mathcal{P}(\mathbb{R}^d\times\mathbb{R}^d)$ を

$$\pi_{13}[A] := \pi[\{(x,y,z) \in \mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d \mid (x,z) \in A\}]$$

で定めれば, $\pi_{13} \in \Pi(\mu_1, \mu_3)$ です. ゆえに,

$$
\begin{aligned}
W_2(\mu_1, \mu_3) &\le \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} |x - z|^2 d\pi_{13}(x, z) \right)^{\frac{1}{2}} \\
&= \left(\int_{\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d} |x - y + y - z|^2 d\pi(x, y, z) \right)^{\frac{1}{2}} \\
&\le \left(\int_{\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 d\pi(x, y, z) \right)^{\frac{1}{2}} + \left(\int_{\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d} |y - z|^2 d\pi(x, y, z) \right)^{\frac{1}{2}} \\
&= \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 d\pi_{12}(x, y) \right)^{\frac{1}{2}} + \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} |y - z|^2 d\pi_{23}(y, z) \right)^{\frac{1}{2}} \\
&= W_2(\mu_1, \mu_2) + W_2(\mu_2, \mu_3)
\end{aligned}
$$

となり, 三角不等式が従います. □

以下に紹介する補題は, 共通する周辺分布を持つ二つのカップリングを共通部分を糊しろにして貼合わせる, と言うものです. 主張も証明も [13, Lemma 7.6] で見つけられます.

補題 3.2 (貼合わせの補題) μ_1, μ_2, μ_3 をそれぞれ完備可分距離空間 X_1, X_2, X_3 上の確率度とする. このとき任意のカップリング $\pi_{12} \in \Pi(\mu_1, \mu_2), \pi_{23} \in \Pi(\mu_2, \mu_3)$ に対し, ある $X_1 \times X_2 \times X_3$ 上の確率測度 π が存在して任意の可測集合 $A_{12} \subset X_1 \times X_2, A_{23} \subset X_2 \times X_3$ に対して

$$
\pi[A_{12} \times X_3] = \pi_{12}[A_{12}], \qquad \pi[X_1 \times A] = \pi_{23}[A_{23}]
$$

が成立する.

注記 3.3 より一般に, 任意の完備可分距離空間 (X, d) と $p \in [1, \infty)$ に対して

$$
\mathcal{P}_p(X) := \left\{ \mu \in \mathcal{P}(X) \;\middle|\; \inf_{x \in X} \int_X d(x, y)^p d\mu(y) < \infty \right\},
$$

$$
W_p(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \|d\|_{\mathrm{L}^p(\pi)} = \left(\inf_{\pi \in \Pi(\mu, \nu)} \int_{X \times X} d(x, y)^p d\pi(x, y) \right)^{\frac{1}{p}}
$$

を考えれば, $(\mathcal{P}_p(X), W_p)$ も完備可分距離空間となります. そしてこの距離空間を X 上の p 次の Wasserstein 空間と呼びます. そして W_p による収束は弱収束と p 次モーメントの収束の二条件の組合せと同値です. つまり, W_p で収束すれば必ず弱収束します. (詳しくは [14, Theorem 6.9] を参照にして下さい.)

X 上の Wasserstein 空間は X の距離の幾何反映します. それは

$$
(X, d) \ni x \mapsto \delta_x \in (\mathcal{P}_p(X), W_p)
$$

が等長埋込になっているからです. 実際, 任意の $x_0, y_0 \in X$ に対して $\Pi(\delta_{x_0}, \delta_{y_0}) = \{\delta_{(x_0, y_0)}\}$ なので,

$$
W_2(\delta_{x_0}, \delta_{y_0}) = \left(\int_{X \times X} d(x, y)^p d\delta_{(x_0, y_0)}(x, y) \right)^{\frac{1}{p}} = d(x_0, y_0)
$$

となります. 特に [2, p.10–11] で述べられているように, $(\mathcal{P}_p(X), W_p)$ はフィンスラー多様体のように振舞い, 特に $p=2$ のときはリーマン多様体のように振舞います. これは L^p-空間というノルム空間の中で L^2-空間のみが内積空間になることに関連します. そこで, リーマン幾何学的な側面を考察する際には $p=2$ を扱うのが適しています. 例えば, ある内部エネルギー汎関数の W_2 に関する振舞, 特に凸性をみることで, 微分構造を許容しない空間にリッチ曲率の下限 (と次元の上限) が定義できます. この'リッチ曲率が K 以上, 次元が N 以上' であるための条件は曲率次元条件 $\mathrm{CD}(K,N)$ という名前で呼ばれています. その定義は例えば [14, Definition 29.8] で見つけられますが, どのような背景の下この定義ができたのかは [14, Chapter 14] に書いてあります. ここでは $(\mathcal{P}_2(X), W_2)$ のリーマン構造や曲率次元条件に言及しませんが, $(\mathcal{P}_2(\mathbb{R}), W_2)$ の中で正規分布族がどのような振舞をするのかをみて, 情報幾何で現れるリーマン計量, フィッシャー計量とは違うことを考証したいと思います.

4 正規分布族上の幾何

任意の $m \in \mathbb{R}$, $\sigma > 0$ に対して, $N(m, \sigma^2)$ で平均が m, 分散が σ^2 である $\mathbb{R}$ 上の正規分布を表すとします. すなわち, $N(m,\sigma^2)$ とはルベーグ測度に対する密度関数が

$$\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$$

である実数上の確率測度です. (そして任意の $m \in \mathbb{R}$, $\sigma > 0$ に対して, $N(m,\sigma^2) \in \mathcal{P}_2^{\mathrm{ac}}(\mathbb{R})$ が成立ちます.) 正規分布のなす空間を $\mathcal{N}$ とすれば,

$$\mathcal{N} \ni N(m,\sigma^2) \mapsto (m,\sigma) \in \mathbb{R} \times \mathbb{R}_{>0} \tag{4.1}$$

なる自然な対応により, $\mathcal{N}$ は上半平面に同相です.

(1) フィッシャー計量: フィッシャー計量は適切な条件を満たすパラメータ付けられた確率測度族上に定義される情報幾何において取り扱われるリーマン計量です. (ここでは詳細に触れませんが, 例えば [1] を参考にして下さい.) 正規分布族 $\mathcal{N}$ 上でフィッシャー計量は

$$ds^2 = \frac{dm^2 + 2d\sigma^2}{\sigma^2}$$

で与えられます. すなわち, $(\mathcal{N}, ds^2)$ は断面曲率が $-\frac{1}{2}$ である定曲率多様体です.

(1)Wasserstein 距離について: 任意の $m_1, m_2 \in \mathbb{R}$, $\sigma_1, \sigma_2 > 0$ に対して, $N(m_1, \sigma_1^2)$ と $N(m_2, \sigma_2^2)$ の最適輸送を求めしょう. 例えば, 実数上の確率測度間に対する最適輸送は累積分布関数を使って表される事実 ([13, Theorem 2.18]) を使っても良いですし, そうでなくても一般次元の正規分布族間の最適輸送は既知です (例えば [4, 5, 6, 9] 参照). これらの事実により,

$$T(x) = \frac{\sigma_2}{\sigma_1}(x - m_1) + m_2, \quad \pi = (\mathrm{id}_{\mathbb{R}}, T)_\sharp N(m_1, \sigma_1^2)$$

とおくと $\pi \in \Pi(N(m_1,\sigma_1^2), N(m_2,\sigma_2^2))$ は最適輸送になっています. そこで

$$
\begin{aligned}
W_2(N(m_1,\sigma_1^2), N(m_2,\sigma_2^2))^2 &= \int_{\mathbb{R}} |x-T(x)|^2 dN(m_1,\sigma_1^2)(x) \\
&= \int_{\mathbb{R}} \left| \frac{\sigma_1-\sigma_2}{\sigma_1}(x-m_1) + (m_1-m_2) \right|^2 dN(m_1,\sigma_1^2)(x) \\
&= (m_1-m_2)^2 + (\sigma_1-\sigma_2)^2
\end{aligned}
$$

が成立ちます. よって, (4.1) の対応は等長的になっていることが分かります.

より一般に, $\mathbb{R}^d$ 上の正規分布族を考えると, $\mathbb{R}^d$ 上の正規分布は平均ベクトル $m \in \mathbb{R}^d$ と共分散行列 $V \in \mathrm{Sym}_+(d,\mathbb{R})$ でパラメータ付けられます. ここで $\mathrm{Sym}_+(d,\mathbb{R})$ は d 次の正定値実対称行列のなす集合です. そこで $N(m,V)$ で平均ベクトルが m, 共分散行列が V である $\mathbb{R}^d$ 上の正規分布を表すとします. このとき $m_1, m_2 \in \mathbb{R}^d$, $V_1, V_2 \in \mathrm{Sym}_+(d,\mathbb{R})$ に対し

$$
X := V_2^{\frac{1}{2}} \left(V_2^{\frac{1}{2}} V_1 V_2^{\frac{1}{2}} \right)^{-\frac{1}{2}} V_2^{\frac{1}{2}} \in \mathrm{Sym}_+(d,\mathbb{R}), \quad T(x) = X(x-m_1) + m_2
$$

とおけば $T_\sharp N(m_1,V_1) = N(m_2,V_2)$ となります. また T は凸関数の勾配として再現されるので, 定理 2.6 より $\pi = (\mathrm{id}_{\mathbb{R}^d}, T)_\sharp N(m_1,V_1) \in \Pi(N(m_1,V_1), N(m_2,V_2))$ が最適輸送であることが分かります. (これは元々 [4, 5, 6, 9] で示されています.) ゆえに

$$
\begin{aligned}
W_2(N(m_1,V_1), N(m_2,V_2))^2 &= \int_{\mathbb{R}^d} |x-T(x)|^2 dN(m_1,V_1)(x) \\
&= |m_1-m_2|^2 + \mathrm{Tr}(V_1) + \mathrm{Tr}(V_2) - 2\mathrm{Tr}\left(\left(V_2^{\frac{1}{2}} V_1 V_2^{\frac{1}{2}} \right)^{\frac{1}{2}} \right)
\end{aligned}
\tag{4.2}
$$

が成立ちます. さらに $t \in [0,1]$ に対して

$$
m_t = (1-t)m_1 + tm_2, \qquad V_t := [(1-t)I_d + tX] \cdot V_1 \cdot [(1-t)I_d + tX] \in \mathrm{Sym}_+(d,\mathbb{R})
$$

とおけば, 定理 2.6 より $\{N(m_t,V_t)\}_{t\in[0,1]}$ が $N(m_1,V_1)$ から $N(m_2,V_2)$ への Wasserstein 距離関数に関する最短線になっています. そして $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ の最短線は分岐しない ([14, Corollary 7.32]) ので, Wasserstein 空間 $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ の中で正規分布族は凸集合になっています. そこで Wasserstein 距離関数を正規分布族に自然に制限することが出来ます. 空間 $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ は無限次元であるためにその解析は難しいのですが, 正規分布族は有限次元なのでその性質を仔細に検証できます. (いくつかの性質は [11] にまとめてあります. 例えば, (4.2) より $\mathbb{R}^d$ 上の正規分布族は Wasserstein 距離関数に関して平坦でないことが直ちに分かりますが, [11] では Wasserstein 距離関数に関して正規分布族が (Alexandrov 空間として) 非負曲率であることにも言及しています.)

ただ Wasserstein 空間が無限次元で解析が困難だと言っても, ある程度はその性質が知られています. 例えば $(\mathcal{P}_2(\mathbb{R}), W_2)$ は '平坦' であり, $d \geq 2$ に対して $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ は錐空間になっています. (平坦性については [10, §4.5], 錐空間については [12] を参照にして下さい.)

また, 情報幾何 (フィッシャー計量) と Wasserstein 幾何が異なることは述べましたが, 全くもって相関が無いわけではありません. 例えば, $\mu, \nu \in \mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$ に対して μ が ν にも絶対連続であるとき

$$H_\nu(\mu) := \int_{\mathbb{R}^d} \frac{d\mu}{d\nu} \log\left(\frac{d\mu}{d\nu}\right) d\nu$$

なる量を μ の ν を参照にした相対エントロピーと呼びます. そして情報幾何の文脈では, 相対エントロピーは一般化されたピタゴラスの定理 ([1, 定理 3.4]) を満たすため, 距離関数の二乗のように振舞うと考えられています. このとき相異なる距離関数, Wasserstein 距離と相対エントロピーの平方根, があれば比べたくなるのが人情で例えば標準正規分布 $N(0, I_d)$ と任意の $\mu \in \mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$ に対し

$$W_2(\mu, N(0, I_d)) \leq \sqrt{2H_{N(0,I_d)}(\mu)}$$

が成立つことが知られています. この不等式はタラグランド不等式と呼ばれる輸送不等式の一つです. ここで輸送不等式とは大雑把に言って, 最小総費用 (W_2) をエネルギー差 (H) で上から評価する不等式です. 詳しくは [14, Chapter 22] を参考にして下さい.

例えば, 費用関数は距離関数の二乗のままで, 相対エントロピーをブレグマンダイバージェンスに取り替えてみましょう. ここで正数上の C^1-狭義凸関数 U に附随するブレグマンダイバージェンス H^U とは $\mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d) \times \mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$ 上の非負値汎関数で, $\mu, \nu \in \mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$ に対して μ の ν を参照にしたブレグマンダイバージェンスは

$$H_\nu^U(\mu) = \int_{\mathbb{R}^d} \left[U\left(\frac{d\mu}{dx}\right) - U\left(\frac{d\nu}{dx}\right) - U'\left(\frac{d\nu}{dx}\right)\left(\frac{d\mu}{dx} - \frac{d\nu}{dx}\right)\right] dx$$

と定義されます. 例えば, $U(r) = r\log(r)$ とすれば, 標準正規分布を参照にした相対エントロピー $H_{N(0,I_d)}$ と標準正規分布を参照にしたブレグマンダイバージェンス $H^U_{N(0,I_d)}$ は一致します. よってブレグマンダイバージェンスは相対エントロピーの一般化とみなさせます. さらにブレグマンダイバージェンスは一般化されたピタゴラスの定理を満たすので, 情報幾何の文脈では距離関数の二乗のように振舞うとみなせます. (ブレグマンダイバージェンスについては例えば [7] やそこで挙げられている文献を参考にして下さい.) そして適切な条件下において, Wasserstein 距離関数とブレグマンダイバージェンスの平方根を比較する輸送不等式が成立つことが知られています. (例えば, [8, Theorem 6.3] を参照にして下さい.)

このようにして二つの幾何, Wasserstein 幾何と情報幾何, の様相は異なりますが全く関連がないわけでなく, 輸送不等式を通して二つの幾何を比べることが可能です.

References

[1] 甘利俊一, 長岡浩司, **情報幾何の方法**, 岩波講座 応用数学, 岩波書店, 1993.

[2] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, second ed., Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2008.

[3] P. Billingsley, *Probability and measure*, third ed., Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons Inc., New York, 1995, A Wiley-Interscience Publication.

[4] D. C. Dowson and B. V. Landau, *The Fréchet distance between multivariate normal distributions*, J. Multivariate Anal. **12** (1982), no. 3, 450–455.

[5] C R. Givens and R. M. Shortt, *A class of Wasserstein metrics for probability distributions*, Michigan Math. J. **31** (1984), no. 2, 231–240.

[6] M. Knott and C. S. Smith, *On the optimal mapping of distributions*, J. Optim. Theory Appl. **43** (1984), no. 1, 39–49.

[7] J. Naudts, *Generalised thermostatistics*, Springer-Verlag, 2011.

[8] S. Ohta and A. Takatsu, Displacement convexity of generalized relative entropies. II, Comm. Anal. Geom. **21** (2013), 687–785,

[9] I. Olkin and F. Pukelsheim, *The distance between two random vectors with given dispersion matrices*, Linear Algebra Appl. **48** (1982), 257–263.

[10] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations **26** (2001), no. 1-2, 101–174.

[11] A. Takatsu, *Wasserstein geometry of Gaussian measures*, Osaka J. Math. **48** (2011), no. 4, 1005–1026.

[12] A. Takatsu and T. Yokota, *Cone structure of L^2-Wasserstein spaces*, J. Topol. Anal. **04**(2012), no. 02, 237–253.

[13] C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003.

[14] C. Villani, *Optimal transport, old and new*, Grundlehren der Mathematischen Wissenschaften, vol. 338, Springer-Verlag, 2009.