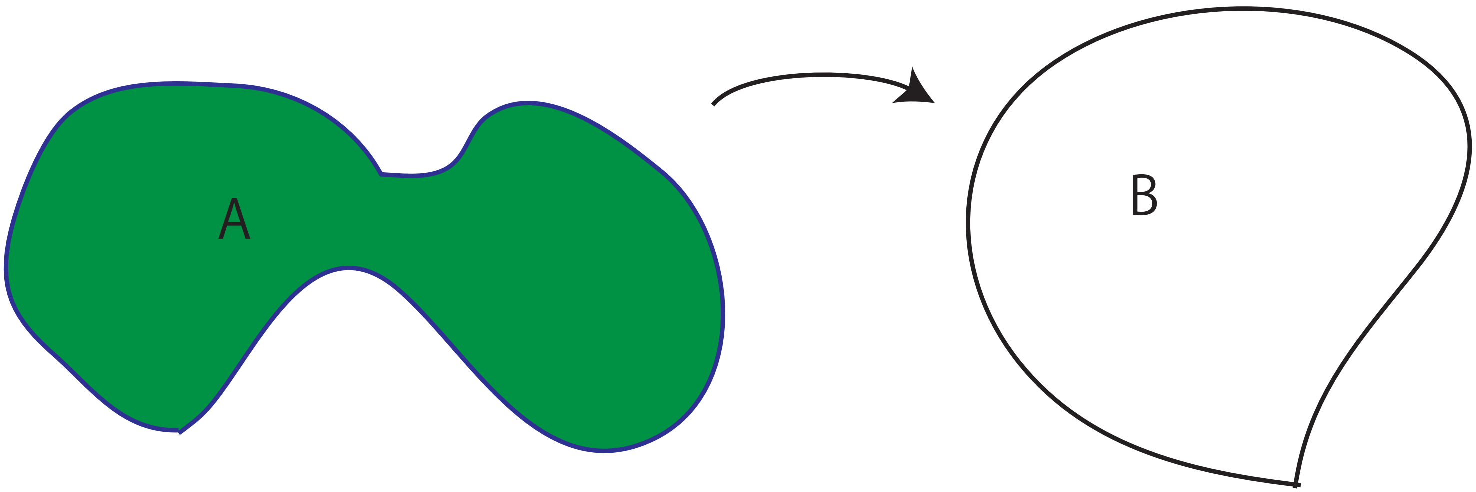
モンジュの最適輸送問題をめぐる話題について

會田 茂樹

モンジュの問題

下のように同じ面積の土地 A、B がある。A には芝生が生えているが、それを土地 B にはりかえたい。芝のはりかえには移動距離に依存したコストがかかるとき、どのようにはりかえるのが最良か。



モンジュの問題

ある砂山をそれと同じ体積の穴に移したい。砂の移動には移動距離に依存したコストがかかるとき、最適な移動のさせ方は何か？

歴史

これらの問題はフランスの数学者・工学者 Gaspard Monge(ガスパール・モンジュ) が 1781 年の論文の中で提唱した問題でモンジュの最適輸送問題と言う。

その後、Sudakov が問題解決を宣言した (1979 年) が、それには誤りがあった。Sudakov の論文が現れる以前にこの問題に関して重要なアイデアを出していたのは Leonid Kantorovich(1940 年代) であった。

そのアイデアに基づいて (ある場合ですが) 初めて厳密に解かれたのは 1987 年、Yann Brenier によると言えます。今回はこの問題に関してお話します。

歴史

この問題で使われる数学の概念は、面積、体積、積分などです。このモンジュの問題の解決には「ルベーグ積分」，「測度論」が必要です。

このルベーグ積分は Henri Lebesgue(アンリ・ルベーグ) が 1902 年の学位論文で創始し、20 世紀に発展したものです。高校や大学初年級で学ぶ微積分で出てくる積分 (リーマン積分) は 19 世紀のリーマンにより整備されたものですが、基本的にニュートン・ライプニッツ (17 世紀) の数学と言えます。

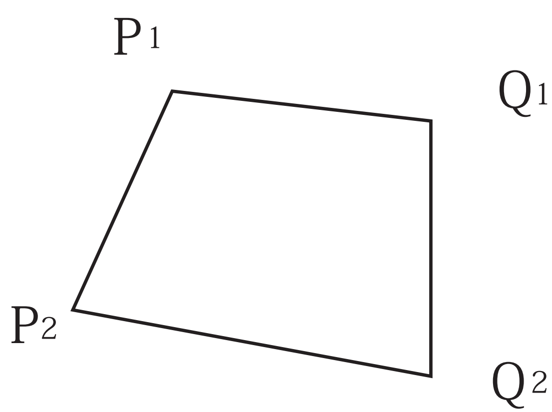
このような理論的準備が必要だったのが、この問題の難しさの一つと言えます。

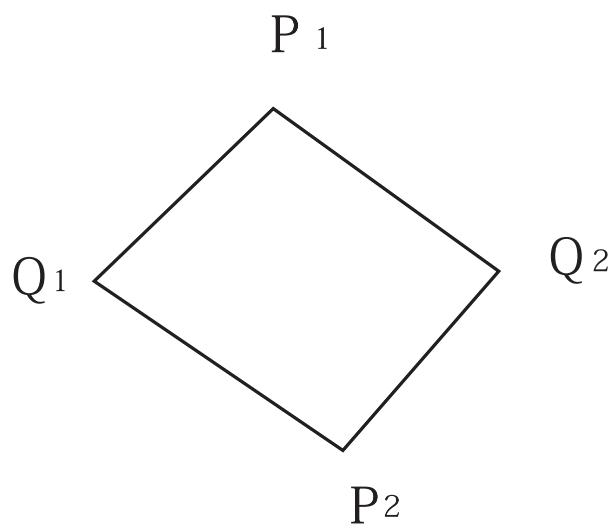
話の流れ

1. 離散的な場合のモンジュの問題
2. ヒッチコック型の輸送問題
3. モンジュの問題の定式化 (I)
4. モンジュの問題の定式化 (II)
5. モンジュ・カントロヴィッチの問題
6. モンジュの問題の解 (Brenier-McCann のアイデア)
7. 最適輸送写像の応用 (Warping map)
8. 確率分布全体の集合の幾何

離散的な場合:二点の移動

異なる二つの位置$\{P_1, P_2\}$に同じ重さの砂一粒ずつがあるとする。これを指定された位置$\{Q_1, Q_2\}$に移動させる。砂一粒の移動のコストはその距離に比例するとする。どのように移動させるとコストが最小になるか？





モンジュ性

$P_1 \longrightarrow Q_1$

$P_2 \longrightarrow Q_2$

が最小

離散的な場合:一般な場合

- 相異なる位置 $\{P_1, \ldots, P_n\}$ に同じ重さの砂粒がある。
- それをやはり相異なる位置 $\{Q_1, \ldots, Q_n\}$ に移動させる。P_i から Q_j への移動には P_i と Q_j の間の距離 $d(P_i, Q_j)$ に比例したコスト $cd(P_i, Q_j)$ がかかる。
- 最適な移動のさせ方は何か？

2 点の場合よりは複雑になるが、移動のさせ方は $n!$ 通りだからそれらの中で最小のものを探すことになる。

モンジュの問題はこれの連続版と言える。

この離散モンジュ問題の一般化にあたるヒッチコック型輸送問題を紹介する。

問題 1

- 工場 1, 工場 2 がありそれぞれ製品を 100 個、200 個生産する.
- それらを町 1 に 75 個, 町 2 に 225 個輸送したい.
- 各工場から各町への 1 個あたりの輸送費用は次の表のようにかかる. 輸送費用を最小にするには, どのように輸送すればよいか?

	町 1(75 個)	町 2(225 個)
工場 1(100 個)	2 円	6 円
工場 2(200 個)	1 円	4 円

⟹ 問題 1 の場合

解

- x :工場 1 から町 1 への配送量，y : 工場 2 から町 1 への配送量とする．
- $x+y=75,\ 0\le x\le 100,\ 0\le y\le 200$ の条件の下で総費用

$$\begin{aligned}
&\overset{(\mathrm{F1}\to\mathrm{C1})}{2x}+\overset{(\mathrm{F1}\to\mathrm{C2})}{6(100-x)}+\overset{(\mathrm{F2}\to\mathrm{C1})}{1\cdot y}+\overset{(\mathrm{F2}\to\mathrm{C2})}{4(200-y)}\\
&\quad=-4x-3y+1400\\
&\quad=-4x-3(75-x)+1400=-x+1175
\end{aligned}$$

を最小にしたい．

- $x=75,\ y=0$ のとき，最小値 1100 を取る．

行列

- 2 次正方行列 (2 × 2 行列): $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$

  a_{ij} : (i, j) 成分と呼ばれる。

- (m, n) 行列 ($m \times n$ 行列, m 行 n 列の行列)

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{i1} & \cdots & a_{ij} & \cdots & a_{in} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mj} & \cdots & a_{mn} \end{pmatrix}$$

ヒッチコック型の輸送問題

- 工場が m 個、町が n 個ある
- 工場 i の生産量は μ_i, 町 j の必要量が ν_j,
  $1 \leq i \leq m, 1 \leq j \leq n$
- $\displaystyle\sum_{i=1}^{m} \mu_i = \sum_{j=1}^{n} \nu_j$
- 工場 i から町 j への単位あたりの輸送コストは c_{ij}
- 工場 i から町 j への輸送量を x_{ij}

輸送コスト $\displaystyle C(X) = \sum_{j=1}^{n} \sum_{i=1}^{m} c_{ij} x_{ij}$ を $\displaystyle\sum_{j=1}^{n} x_{ij} = \mu_i, \ \sum_{i=1}^{m} x_{ij} = \nu_j$ の条件の下で最小とする X を求めよ．

解の存在

- 実行可能解の存在 $\sum_{j=1}^{n} x_{ij} = \mu_i, \sum_{i=1}^{m} x_{ij} = \nu_j$ をみたす x_{ij} を実行可能解という。
  北西隅ルールで実行可能解の存在はわかる．
- 最適解の存在
  最大・最小の存在定理
  コンパクト集合上で定義された連続関数には最大値・最小値をとるところがある.
  による．

北西隅ルール

1. 工場 1 から可能な限りの量を町 1 に必要な範囲で送る．
2. まだ工場 1 に残量があれば町 2 が必要としている範囲で送る．それでもなお工場 1 に残量があれば町 3 へ送る．これを繰り返し，工場 1 の残量が 0 になるまで続ける．
3. 工場 1 から出荷した製品だけではナンバー i 以後の町の必要量が不足していれば，次は工場 2 から町 i へ可能な限り配送し，以下同様な方法で配送する．

$\Longrightarrow$ 実行可能解の存在

	町 1 (75 個)	町 2 (225 個)
工場 1 (100 個)	75 個	25 個
工場 2 (200 個)	0 個	200 個

最大・最小の存在定理

K を $\mathbb{R}^d$ の**コンパクト集合**とする．f を K を定義域とする連続関数とする．このとき，K のある点 P, Q で f は最大値，最小値を取る．

(注) 最大・最小の存在定理はユークリッド空間のコンパクト集合だけではなく、一般の位相空間 (**距離空間**あるいはより一般に位相が与えられた空間) のコンパクト集合上の連続関数に対して成立している. $\Longrightarrow$ MK **問題の解の存在** $\Longrightarrow$ $C(T)$ **とその定義域**

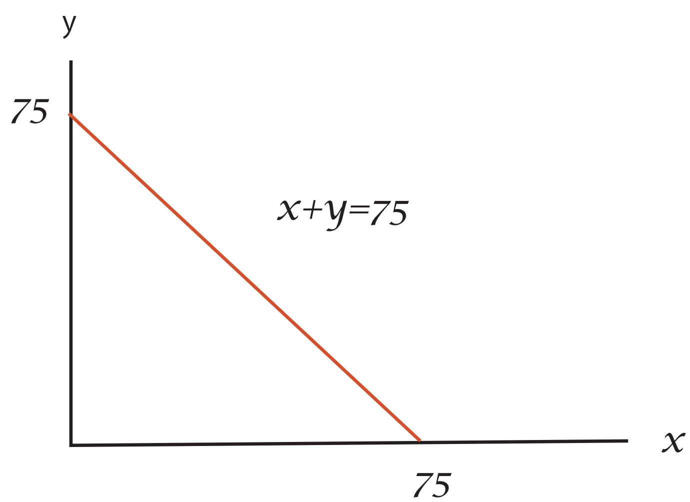
[問題1]の場合

[問題 1]のときは、

$$f(x, y) = -x + 1175$$

$$K = \left\{(x, y) \mid x + y = 75, 0 \leq x \leq 100, 0 \leq y \leq 200\right\}$$

に対して適用することになる。 $\Longrightarrow$ 最大・最小の存在定理



北西隅ルールに対する注意

- 北西隅ルールで実行可能解が作れる.
- 価格行列がモンジュ性を満たせば北西隅ルールで作った解は最適解になる.
- モンジュ性が無いときは、その得られた解をさらに改良して (例えば飛び石法と呼ばれる方法で) 最適解を作る.

行列$C=(c_{ij})$のモンジュ性

$1 \le i < r \le m, 1 \le j < s \le n$ をみたすすべての i, r, j, s について

$$c_{ij} + c_{rs} \le c_{is} + c_{rj}$$

$n = m = 2$ のときは、$c_{11} + c_{22} \le c_{12} + c_{22}$

この式は [問題 1] の価格行列で成立している。なぜなら

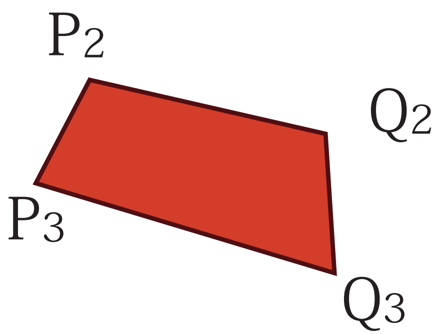
$$\begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} = \begin{pmatrix} 2 & 6 \\ 1 & 4 \end{pmatrix}$$

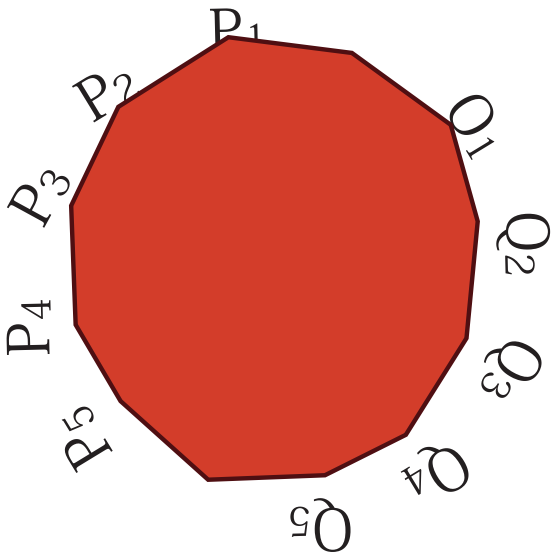
モンジュ性について

例えば下の凸多角形で

$$c_{ij} = d(P_i, Q_j)$$

とするとモンジュ性が成り立つ：





$$d(P_2, Q_2) + d(P_3, Q_3)$$
$$\leqq d(P_2, Q_3) + d(P_3, Q_2)$$

離散モンジュの問題との関連 (I)

- 相異なる位置 $\{P_1, \ldots, P_n\}$ に同じ重さの砂粒がある。
- それをやはり相異なる位置 $\{Q_1, \ldots, Q_n\}$ に移動させる。P_i から Q_j への移動には P_i と Q_j の間の距離 $d(P_i, Q_j)$ に比例したコスト $cd(P_i, Q_j)$ がかかる。
- 最適な移動のさせ方は何か？

という問題は $n = m$ の**ヒッチコック型の輸送問題**で

- $c_{ij} = cd(P_i, Q_j)$, $\mu_i = 1$ $(1 \le i \le n)$, $\nu_j = 1$ $(1 \le j \le n)$.
- $x_{ij} = 0$ または 1(⟸ この条件が加わる！)

の場合にあたる。これは線形計画法の問題の「割り当て問題」と呼ばれるものである。

離散モンジュの問題との関連 (II)

- $x_{ij} = 0$ または 1 の条件:　これは、砂粒が分割できないことを意味する。この条件が無い通常のヒッチコック型の問題は砂粒を分割して輸送するプランも含んでいるわけです。
- しかし、今の場合は**砂粒を分割しないで運ぶ範囲**に最小値を取るプランがあることがわかる。
- $\{P_i\}$ が n 個の工場, $\{Q_j\}$ が m 個 $(m > n)$ の町で P_i の生産量が 1, Q_j の必要量が $\frac{n}{m}$ のときは、当然生産量 1 を分割しなければならない。 $\Longrightarrow$ 存在定理のページ

モンジュの問題について

最初に述べたモンジュの問題は $\{P_1, \ldots, P_n\}$, $\{Q_1, \ldots, Q_n\}$ の n を限りなく大きくし、物質が連続的に分布している場合である。以下、この問題を考える。

まとめ

- 離散モンジュ問題はヒッチコック型の輸送問題の特別な問題である。
- 特別というのは、工場から町へ物質を分割して輸送する場合を考えないからである。
- ヒッチコック型の問題での実行可能解が物質を分割して運ぶ輸送プランのみであることもある。
- しかし、分割しない範囲で最適な輸送プランがあることもある。
- そのときは、離散モンジュの問題の解とヒッチコック型の輸送問題の解が一致している (Birkoff の定理)。

モンジュの問題の定式化 (I)

- $D_1, D_2 \subset \mathbb{R}^d$, 密度が一定 (1 とする) で連続的に分布している物質, かつ $|D_1| = |D_2|$
- $c(x, y)$: 位置 x の物質を位置 y に移動するのに要する単位質量あたりの価格

D_1 から D_2 への次の性質を持つ輸送写像を考える。

- $x \neq x' \implies T(x) \neq T(x')$,
  $T(D_1) = \{T(x) \mid x \in D_1\} = D_2$
- (MP) $U \subset D_1$ のとき $|T(U)| = |U|$.

この条件の下で $C(T) = \int_{D_1} c(x, T(x))dx$ を最小にする T を求めよ．(注)　MP=measure preserving=保測（質量保存）

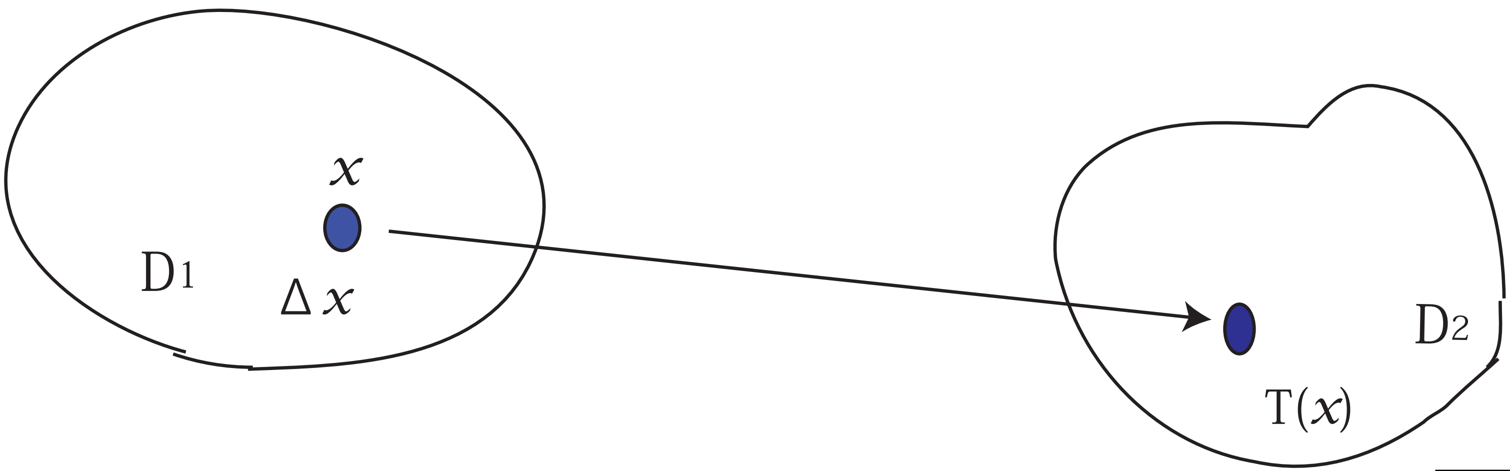
$C(T) = \int_{D_1} c(x, T(x))dx$ の説明

$\Delta x \subset D_1$ のとき、Δx を位置 $T(x)$ に移動させると

そのコストは $c(x, T(x))|\Delta x|$.

これが足されるので積分

$\int_{D_1} c(x, T(x))dx$ が総コストになる。 $\Longrightarrow$ モンジュの問題 (I)



コスト関数の例

$x = (x_1, \ldots, x_d)$, $y = (y_1, \ldots, y_d)$, $p > 0$ に対して

$$c_p(x, y) = |x - y|^p = \left\{ \sum_{i=1}^{d} |x_i - y_i|^2 \right\}^{p/2} .$$

$p = 2$ のときをこの講演では主に考える。$p = 1$ のときがモンジュの本来の問題と言えるが、それは、より難しい問題になる。Brenier の解は $p = 2$ の場合で、$p = 1$ の場合は Ambroiso, Evans, Gangbo らによりその後に解決された。

例

以下、平面上で例をあげる。平面の２点 x, y を

$$x = (x_1, x_2), y = (y_1, y_2)$$

と表す。x, y が x 軸、y 軸を表しているのではないことに注意してほしい。

図形の平行移動

(1) $x = (x_1, x_2), v = (v_1, v_2)$ に対し
$x + v = (x_1 + v_1, x_2 + v_2)$ と平行移動を定義。

(2) 平面上の図形 D を考える。

(3) $D_1 = D$, $D_2 = D + v := \{x + v \mid x \in D\}$ と定義。

(4) 平行移動の写像 $T : D_1 \to D_2$, $(T(x) = x + v)$
は上への一対一かつ MP を満たす。

T はコスト $c(x, y) = |x - y|^p = \{\sum_{i=1,2} |x_i - y_i|^2\}^{p/2}$ $(p \geq 1)$ の場合、最適輸送写像になる。$p < 1$ のときは一般にはそうではない。

一次変換と平行移動

$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ とおく。

- $c(x, y) = |x - y|^2$
- A の行列式 $\det A = a_{11}a_{22} - a_{12}a_{21} = \pm 1.$
- 図形 $D(= D_1$ とする$)$ に対して、
  $D_2 = \{Ax + v \mid x \in D\}.$

このとき、$T(x) = Ax + v$ は D_1 から D_2 への上への一対一写像で MP をみたす。これが最適輸送写像になるのは $a_{12} = a_{21},\ a_{11} > 0,\ \det A = 1$ のときに限る。 $\Longrightarrow$ 「一次変換と平行移動の例について」

合同な図形

- D_1, D_2: 合同な図形
- $c(x, y) = |x - y|^2$

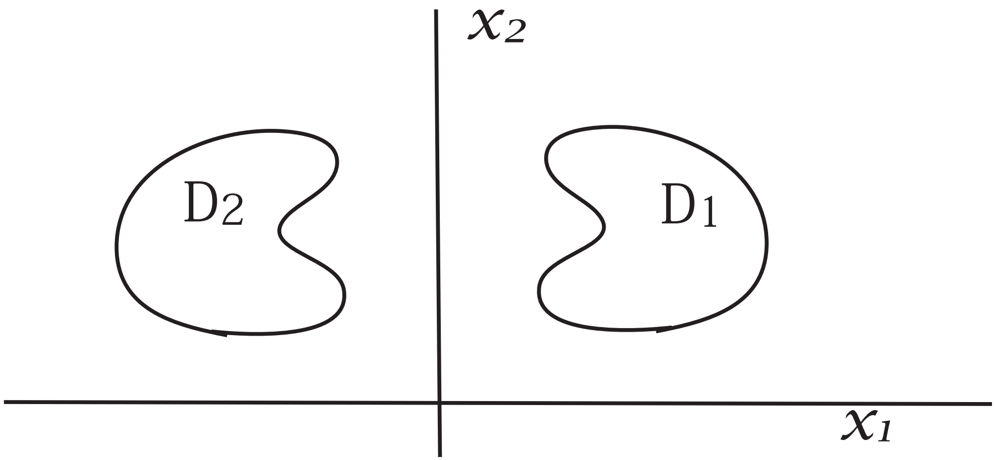
例えば D_2 は D_1 を原点のまわりに回転して得られるとする。この変換は最適輸送写像ではない。

$\Longrightarrow$ 回転行列について

$c(x,y)=|x-y|$のとき (I)

- D_1：平面上の図形でその第一座標 x_1 は $x_1>0$ である。$D_2=\{(-x_1,x_2)\mid(x_1,x_2)\in D_1\}$：$x_1=0$ に関して D_1 を反転したもの
- $T:D_1\to D_2$ を $T(x_1,x_2)=(-x_1,x_2)$ とおくと上への一対一と MP をみたす。

この写像 T は $c(x,y)=|x-y|$ のときの最適輸送写像。しかし $c(x,y)=|x-y|^2$ のときは違う。



$c(x, y) = |x - y|$のとき (II)

- D_1 : 平面上の図形
- $f(x_2)$: x_2 の滑らかな関数で $f(x_2) > 0$.
- $D_2 = \{(x_1 + f(x_2), x_2) \mid (x_1, x_2) \in D_1\}$

とおき、写像 $T : D_1 \to D_2$ を $T(x_1, x_2) = (x_1 + f(x_2), x_2)$ で定める。この T は上への一対一、MP を満たす。$c(x, y) = |x - y|$ のときに最適輸送写像になる。$c(x, y) = |x - y|^2$ のとき、最適輸送写像になるのは $f(x_2) =$ 定数 のときに限る。

モンジュの問題の定式化 (II)

以上は密度が一定の物質の輸送問題。これを場所によって密度が変わる物質の輸送問題に拡張する。そのため、確率分布 (確率測度) の概念を導入する。

確率分布とは?

確率分布

μ が**d 次元ユークリッド空間** $\mathbb{R}^d$ 上の確率分布 (確率測度) とは，$A \subset \mathbb{R}^d$ に対して確率 $\mu(A)$ を与えるものを言う。確率であるので、

(1) $\mu(A) \geq 0$

(2) $\mu(\mathbb{R}^d) = 1$

(3) $A \cap B = \emptyset$(排反事象) ならば
$\mu(A \cup B) = \mu(A) + \mu(B)$

を仮定する。

確率分布

代表的な例として

- 離散型確率分布
- 密度関数をもつ連続型確率分布

と呼ばれるものがある。

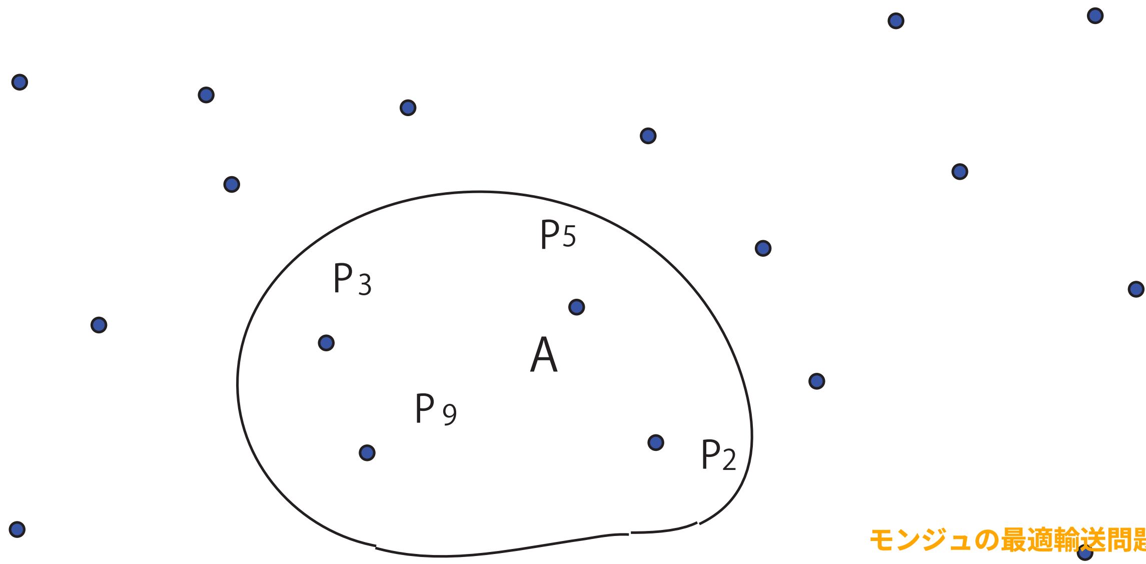
離散型確率分布

- $\mathbb{R}^d$ の中の相異なる m 個の点 $P_1, \ldots, P_m$ が存在して $\mu(\{P_i\}) = \mu_i$, $\sum_{i=1}^m \mu_i = 1$ となるもの．$A \subset \mathbb{R}^d$ に対しては

$$\mu(A) = \sum_{\{i \mid P_i \in A\}} \mu_i$$

と定める．下の絵の場合

$\mu(A) = \mu(\{P_2\}) + \mu(\{P_3\}) + \mu(\{P_5\}) + \mu(\{P_9\})$ である．



密度関数を持つ連続型確率分布

$f(x)$ を $\mathbb{R}^d$ 上の関数で

- すべての x について $f(x) \geq 0$
- $\displaystyle\int_{\mathbb{R}^d} f(x)dx = 1$

となるものとする．集合 $A \subset \mathbb{R}^d$ の確率を

$m_f(A) = \displaystyle\int_A f(x)dx$ と定めると m_f は確率分布になる．これを密度関数 $f(x)$ をもつ連続型確率分布と言う．

例１ 一様分布

D を $\mathbb{R}^d$ の部分集合で有界 (大きさが有限ということ) とする．$|A|$ で集合 A の体積を表す。

$$\mu(A) = \frac{|A \cap D|}{|D|}.$$

これは密度関数 $\dfrac{1_D(x)}{|D|}$ をもつ連続型確率分布. ただし

$$1_D = \begin{cases} 1 & (x \in D) \\ 0 & (x \notin D). \end{cases}$$

$\Longrightarrow$ Quadratic cost

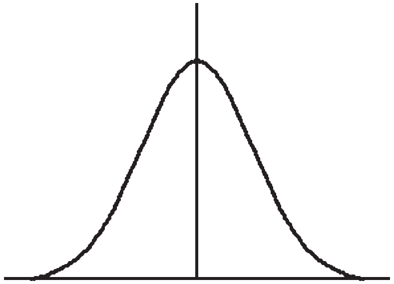
例２ 正規分布

$$\rho(x) = \frac{1}{(2\pi)^{\frac{d}{2}}} \exp\left(-\frac{|x|^2}{2}\right)$$

とすると

$$\int_{\mathbb{R}^d} \rho(x)dx = 1.$$

m_ρ を d 次元正規分布と言う．



$\Longrightarrow$ Quadratic cost

$\Longrightarrow$ **対数ソボレフ不等式のページ**

離散型でも密度を持つ連続型でもない確率分布 (I)

- 平面上の 2 点 P, Q を結ぶ線分 PQ を考える。
- $A \subset PQ$ に対して

$$\mu(A) = \frac{A\text{ の長さ}}{PQ\text{ の長さ}}$$

とする。

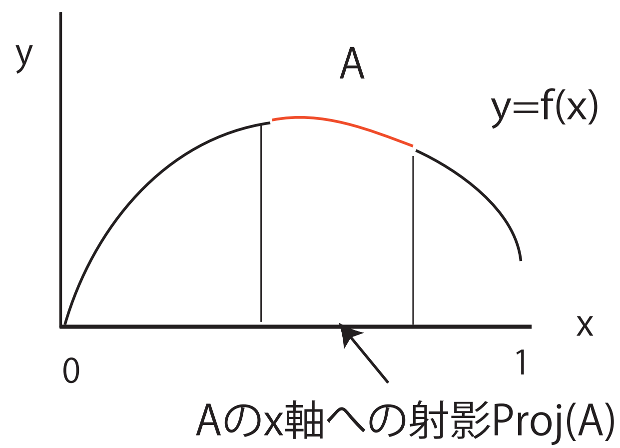
この確率分布は線分 PQ 上に一様に分布する確率分布である。

離散型でも密度を持つ連続型でもない確率分布 (II)

(x, y) 平面上で関数 $y = f(x)$ のグラフ

$$G(f) = \{(x, f(x)) \mid 0 \leq x \leq 1\}$$

を考える。$A \subset G(f)$ に対して A を x 軸に射影した図形 $\mathrm{Proj}(A)$ の長さを $\nu(A)$ とする。このν は $G(f)$ 上の確率分布になる。



$G(f)$ の長さで決まる確率とは違う！ $\Longrightarrow$ 像測度 (III)

測度による積分 (I)

- μ : $\mathbb{R}^d$ 上の確率分布 (確率測度)
  $\varphi(x)$: $\mathbb{R}^d$ 上の連続関数に対して
  積分 $\displaystyle\int_A \varphi(x)\mu(dx)$ ($A \subset \mathbb{R}^d$) を次のように定義する。

(1) A を小さい集合 A_i に分ける。

(2) 和 $\displaystyle\sum_i \varphi(P_i)\mu(A_i)$ を取る。P_i は A_i に属す点。

(3) (1) の分割を細かくしていくとき、(2) の総和がある一定の値に収束するとする。その極限値を $\displaystyle\int_A \varphi(x)\mu(dx)$ と定義する。

測度による積分 (II)

m_f：密度関数 $f(x)$ をもつ連続型確率分布つまり、$A \subset \mathbb{R}^d$ の確率が $\int_A f(x)dx$ であるような確率分布。

$$\int_{\mathbb{R}^d} \varphi(x) m_f(dx) = \int_{\mathbb{R}^d} \varphi(x) f(x) dx$$

測度による積分(III)

$\mu : \mu(P_i) = \mu_i \ (1 \le i \le m)$ となる離散型確率分布とする。($\sum_i^m \mu_i = 1$ に注意) このとき

$$\int_{\mathbb{R}^d} f(x)\mu(dx) = \sum_{i=1}^{m} f(P_i)\mu_i.$$

像測度 (I)

二つの集合 D_1, D_2 の体積に関するモンジュの問題を定式化したとき、写像 $T : D_1 \to D_2$ に

- $x \neq x' \Longrightarrow T(x) \neq T(x')$,
  $T(D_1) = \{T(x) \mid x \in D_1\} = D_2$
- (MP) $U \subset D_1$ のとき $|T(U)| = |U|$.

の条件をおいた。これを次のように緩めて考える：

- $V \subset D_2$ に対して $|V| = |T^{-1}(V)|$ ただし

$$T^{-1}(V) = \{x \in D_1 \mid T(x) \in V\}$$

と定める。

像測度 (II)

前ページの条件では T が上への一対一対応であることは必要無い。そこで, 一般に次の定義をおく。
$\mu : \mathbb{R}^d$ の確率分布
$T : \mathbb{R}^d$ から $\mathbb{R}^d$ への写像
に対して新しい確率分布 $T_\sharp \mu$ を

- $A \subset \mathbb{R}^d$ に対して

$$\left(T_\sharp \mu\right)(A) = \mu(T^{-1}(A)) \qquad A \subset \mathbb{R}^d$$

と定める。$T_\sharp \mu$ を μ の T による像測度と言う。$\left(T_\sharp \mu\right)(A)$ の値は A に輸送写像で持ち込まれる図形 $T^{-1}(A)$ の μ で計った質量である。

モンジュの問題の定式化 II

μ, ν を $\mathbb{R}^d$ の確率分布とする。次の性質をみたす写像 $T : \mathbb{R}^d \to \mathbb{R}^d$ を考える：

(MP) T による μ の像測度は ν である。
すなわち $A \subset \mathbb{R}^d$ に対して $\nu(A) = \mu(T^{-1}(A))$.
ただし $T^{-1}(A) = \left\{x \in \mathbb{R}^d \mid T(x) \in A\right\}$.

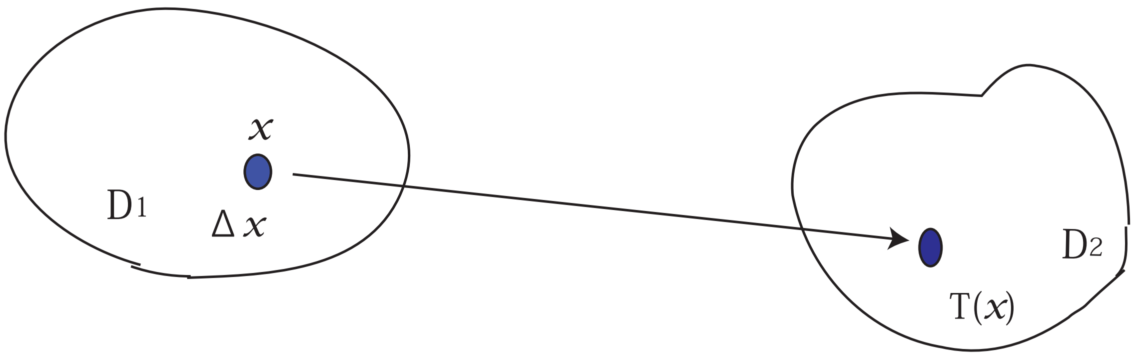
[モンジュの問題 (II)] 上記の性質をみたす T の中で $C(T) = \displaystyle\int_{\mathbb{R}^d} c(x, T(x))\mu(dx)$ を最小にする T を求めよ。$c(x, y)$ は x から y へ単位質量の物質を移動させるのに要するコスト。

$\int_{\mathbb{R}^d} c(x, T(x))\mu(dx)$ について

$\Delta x \subset \mathbb{R}^d$ のとき、Δx を位置 $T(x)$ に移動させると

そのコストは $c(x, T(x))\mu(\Delta x)$.

これが足されるので積分 $\int_{\mathbb{R}^d} c(x, T(x))\mu(dx)$
が総コストになる。 $\Longrightarrow$ モンジュの問題 (II)



モンジュの問題 (I) との関連

(1) D_1, D_2 を同じ体積をもつ集合

(2) μ: D_1 上の一様分布, ν : D_2 上の一様分布

のときを考えるとモンジュの問題 (I) の問題と同じになる。
ただし、

- 問題 (I)

(i) $x \neq x' \implies T(x) \neq T(x'), \quad T(D_1) = \{T(x) \mid x \in D_1\} = D_2$

(ii) $U \subset D_1$ のとき $|T(U)| = |U|$.

↓

- 問題 (II) $T_\sharp \mu = \nu$.

(MP) の言い換え

- $\mu = m_f, \nu = m_g$：密度をもつ分布
- 写像 T が滑らか

とする。(MP)(=保測) は次のように言い換えられる：

$$\det\left(DT(x)\right) g(T(x)) = f(x).$$

$n = 1, 2$ のときは $T'(x)g(T(x)) = f(x) \quad (n = 1)$,

$$\det\begin{pmatrix} \frac{\partial T_1}{\partial x_1}(x) & \frac{\partial T_1}{\partial x_2}(x) \\ \frac{\partial T_2}{\partial x_1}(x) & \frac{\partial T_2}{\partial x_2}(x) \end{pmatrix} g(T(x)) \quad = \quad f(x) \quad (n = 2)$$

(注) 一般に $\det\begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.$

関数 $C(T)$ とその定義域

以上の定式化で

- $\mathcal{D}$ = (MP) をみたす写像全体の集合
- $\mathcal{D}$ を定義域とする関数

$$C(T) = \int_{\mathbb{R}^d} c(x, T(x))\mu(dx)$$

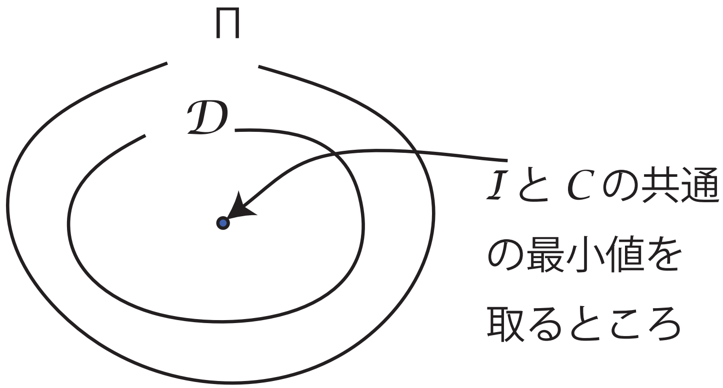
に対して最小値を取る写像が存在するか？という問題になる。しかし、$\mathcal{D}$ に入る要素があるか？は自明でない。

しかも**最大・最小の存在定理は使えない！**

$\mathcal{D}$: コンパクト集合、$C(T)$:連続関数でないから！

Kantorovichのアイデア(1942)

- $\mathcal{D}$ を定義域とする関数 C は扱いにくい。
- 定義域を広げ、関数をより広い範囲 Π で定義した新しい関数 $I : \Pi \to \mathbb{R}$ を考える。I の最小値を見つける。
- 最後に、その最小値がもとの定義域 $\mathcal{D}$ に属すことを言えばよい。



$$
\begin{aligned}
I &: \Pi \to \mathbb{R} \\
C &: \mathcal{D} \to \mathbb{R}, \quad I|_{\mathcal{D}} = C
\end{aligned}
$$

Monge-Kantorovichの問題

$$\Pi(\mu,\nu) = \left\{\pi \,\middle|\, \begin{array}{l} \pi \text{は} \mathbb{R}^d \times \mathbb{R}^d \text{上の確率分布で} \\ \mathbb{R}^d \text{の任意の集合} A, B \text{について} \\ \pi(A \times \mathbb{R}^d) = \mu(A), \pi(\mathbb{R}^d \times B) = \nu(B)^\star \end{array}\right\}$$

$$I(\pi) = \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x,y)\pi(dxdy) \quad (\pi \in \Pi(\mu,\nu))$$

$\Pi(\mu,\nu)$ の中で$I(\pi)$を最小にするものを求めよ。

(注)　★ :π の**周辺分布**が μ,ν ということ。

$\Pi(\mu,\nu) \neq \emptyset$

μ, ν に対して次の性質をもつ確率測度 π_* の存在を示すことができる。任意の $A \subset \mathbb{R}^d, B \subset \mathbb{R}^d$ に対して

$$\pi_*(A \times B) = \mu(A)\nu(A).$$

したがって、$\pi_* \in \Pi$. この π_* を $\mu \times \nu$ と書き、μ と ν の直積確率測度と言う。

ヒッチコック型輸送問題との関連 (I)

- $\mu : \mu(\{P_i\}) = \mu_i \quad \sum_{i=1}^{m} \mu_i = 1$
- $\nu : \nu(\{Q_j\}) = \nu_j \quad \sum_{j=1}^{n} \nu_j = 1$

のとき、Π は $\{(P_i, Q_j)\}$ の mn 個の点の上に確率がある離散型分布でその確率 $\pi_{ij} = \pi(\{(P_i, Q_j)\})\ (1 \leq i \leq m, 1 \leq j \leq n)$ が

$$\sum_{i=1}^{m} \pi_{ij} = \nu_j, \quad \sum_{j=1}^{n} \pi_{ij} = \mu_i$$

をみたすものである。

ヒッチコック型輸送問題との関連 (II)

また、$c_{ij} = c(P_i, Q_j)$ とおくと

$$I(\pi) = \sum_{1 \le i \le m, 1 \le j \le n} c_{ij}\pi_{ij}.$$

これを $\sum_{i=1}^{m} \pi_{ij} = \nu_j, \quad \sum_{j=1}^{n} \pi_{ij} = \mu_i$ の条件の下で最小にするわけだから**ヒッチコック型の問題**になる。直積確率測度 $\mu \times \nu$ は $\pi_{ij} = \mu_i \nu_j$ となる (単純な配送プラン！)。

モンジュの問題との関連

- モンジュの問題は非線形な輸送写像に対する最小化問題
- Monge-Kantorovich 問題 (=MK 問題) は積空間 $\mathbb{R}^d \times \mathbb{R}^d = \mathbb{R}^{2d}$ 上の確率分布全体で定義された線形関数に対する最小化問題

です。写像の問題がどうして確率分布の問題になるのか？その橋渡しが必要です。それは

輸送写像 T のグラフ $G(T)$ 上の像測度

によりなされる。

像測度 (III)

一般に写像 $T : \mathbb{R}^d \to \mathbb{R}^d$ があるとき、$n = 1$ のときと同様グラフ $G(T) = \left\{(x, T(x)) \in \mathbb{R}^d \times \mathbb{R}^d \mid x \in \mathbb{R}^d\right\}$ が考えられる.
$\mu : \mathbb{R}^d$ 上の確率分布のとき, 写像
$I \times T : x \to (x, T(x)) \in G(T)$ による μ の像測度 $(I \times T)_\sharp \mu$ は次のようになる. $A \subset G(T)$ に対して

$$\left((I \times T)_\sharp \mu\right)(A) = \mu\left(\left\{x \in \mathbb{R}^d \mid (x, T(x)) \in A\right\}\right).$$

$n = 1$ で μ が $[0, 1]$ 上の一様分布のとき、これは、前、定義した (x, y) 平面のグラフ$G(f)$ 上の確率分布と同じである。

$G(T)$ 上の像測度の性質

- $T : \mathbb{R}^d \to \mathbb{R}^d$
- μ : $\mathbb{R}^d$ 上の確率分布 μ,　　$T_\sharp \mu = \nu$

とする (すなわち、$T \in \mathcal{D}$)。このとき
$\pi_T := (I \times T)_\sharp \mu : G(T) (\subset \mathbb{R}^d \times \mathbb{R}^d)$ 上の確率分布は
以下の性質をみたす：

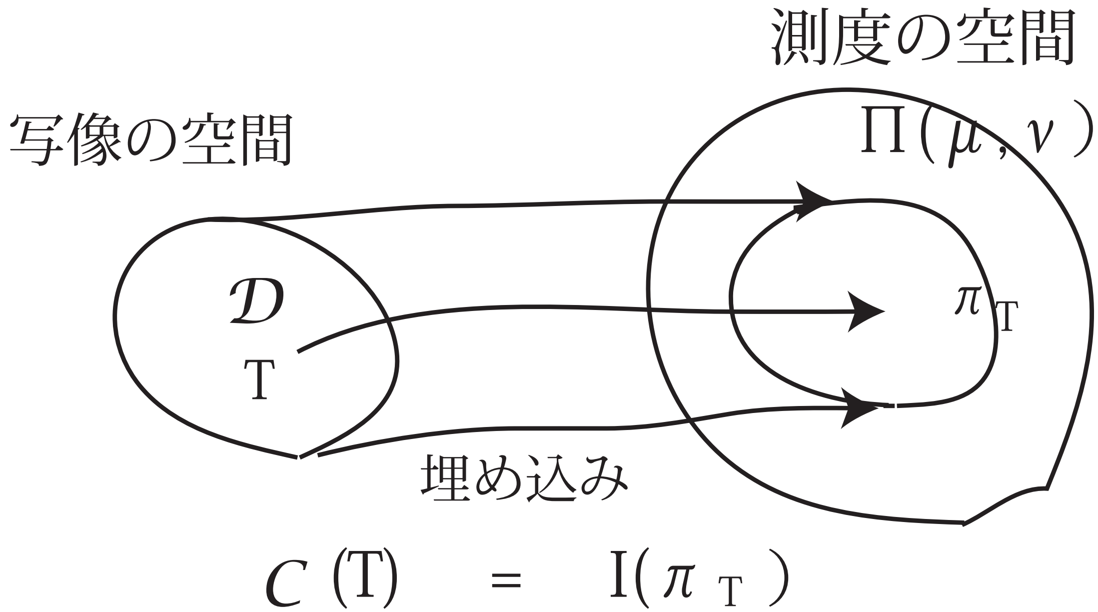
(1) $\pi_T(A \times \mathbb{R}^d) = \mu(A),\ \pi_T(\mathbb{R}^d \times B) = \nu(B)$

(2) $$\int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y)\pi_T(dx, dy) = \int_{\mathbb{R}^d} c(x, T(x))\mu(dx) = C(T)$$

$\mathcal{D}$ の $\Pi(\mu,\nu)$ への埋め込み

$$T(\in \mathcal{D}) \to \pi_T \in \Pi(\mu,\nu)$$

の一対一写像で T を $\Pi(\mu,\nu)$ へ埋め込む。



$c(x,y)=|x-y|^2$ の場合

以下、簡単のため、
$x=(x_1,\ldots,x_d)$, $y=(y_1,\ldots,y_d)$ に対し

$$c(x,y)=|x-y|^2 = \sum_{i=1}^{d}|x_i-y_i|^2$$

$$\int_{\mathbb{R}^d}|x|^2\mu(dx) < \infty \quad \int_{\mathbb{R}^d}|x|^2\nu(dx) < \infty$$

の場合のみを考える。この条件は、前にあげた**標準正規分布**や**有界な集合上の一様分布**はみたしている。

注意

$$\boxed{\int_{\mathbb{R}^d} |x|^2 \mu(dx) < \infty, \quad \int_{\mathbb{R}^d} |x|^2 \nu(dx) < \infty}$$

ならば

$\pi \in \Pi(\mu, \nu)$ のとき

$$\begin{aligned} I(\pi) &= \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 \pi(dxdy) \\ &\leq \int_{\mathbb{R}^d \times \mathbb{R}^d} \left(2|x|^2 + 2|y|^2\right) \pi(dxdy) \\ &= 2 \int_{\mathbb{R}^d} |x|^2 \mu(dx) + 2 \int_{\mathbb{R}^d} |y|^2 \nu(dy) < \infty \end{aligned}$$

MK 問題の解の存在

- $\Pi(\mu, \nu)$ はweak convergence(弱収束) の位相でコンパクト集合になる。
- $I : \Pi(\mu, \nu) \to \mathbb{R}$ は弱収束の位相で連続である。

したがって、**最大・最小の存在定理**により最小値を取る確率分布 $\pi \in \Pi(\mu, \nu)$ が存在する。しかし、前ページの条件 (２次モーメントの存在) だけでは解が $\mathcal{D}$ に属すとは言えません。そもそも μ を ν に移す輸送写像が存在しないこともある。**離散モンジュの問題との関連 (II) のページ**を参照。

最適解が$\mathcal{D}$に属すための条件

定理 (Yann Brenier)

μ,ν は密度関数を持つ連続型確率分布で、２次のモーメントが存在するとする。このとき、I の最小値を取るπについて写像$T\in\mathcal{D}$が存在して$\pi=(I\times T)_\sharp\mu$ と書ける。すなわちこのT は$c(x,y)=|x-y|^2$ のときのモンジュの問題の解でもある。また、T はμ の確率密度関数が正である範囲で定義されている写像である。

証明のアイデア (I)

以下の証明のアイデアは R.J.McCann による。

(1) μ_n : 離散型確率分布で相違なる n 個の点に確率 $\frac{1}{n}$ があるものを取る。

(2) さらに $n \to \infty$ のとき μ_n は μ に収束するとする。

(3) ν_n も同様に離散的で $\nu_n \to \nu$ となるものとする。

(4) μ_n, ν_n に対する MK 問題の解を π_n とする。

(5) π_n の極限 π が μ, ν の MK 問題の解であることを示す。

(6) π がある写像 T のグラフの上に載っていることを示す。

証明のアイデア (II)

(6) で鍵になるのは次の事実。

• $\{P_1, \ldots, P_n\}$: μ_n が分布している点
$\{Q_1, \ldots, Q_n\}$: ν_n が分布している点
とし、$P_i \to Q_i$ $(1 \le i \le n)$ が最適輸送写像とする。$\mathbb{R}^d \times \mathbb{R}^d$ の部分集合

$$S = \{(P_i, Q_i) \mid 1 \le i \le n\}$$

を定義する。このとき、$\pi_n(\{(P_i, Q_i)\}) = \frac{1}{n}$, $\pi_n(S) = 1$ かつ

- (S の Cyclically Monotonicity) 任意の $\{(x_i, y_i)\}_{i=0}^k \subset S$ に対して ($x_{k+1} = x_0$ とする, k は任意の自然数)

$$\sum_{i=0}^{k} (y_i, x_{i+1} - x_i) \le 0.$$

Cyclically Monotonicity

Cyclically Monotonicity=巡回的単調性は次の自明な式と同値：

$\{x_1, \ldots, x_k\}$：$\{P_1, \ldots, P_k\}$を並べかえたもの

$\{y_1, \ldots, y_k\}$：$\{Q_1, \ldots, Q_k\}$を並べかえたもの

とすると

$$\sum_{i=1}^{k} |P_i - Q_i|^2 \le \sum_{i=1}^{k} |x_i - y_i|^2.$$

巡回的単調性と凸関数 (I)

$y = f(x)$ $(-\infty < x < \infty)$ を $f''(x) > 0$ となる関数とする。

$$S = \{(x, f'(x)) \mid -\infty < x < +\infty\}$$

と定める。このとき、S は巡回的単調である。すなわち数列 $\{x_i \mid 0 \le i \le k\}$ に対し $x_{k+1} = x_0$ とおくと

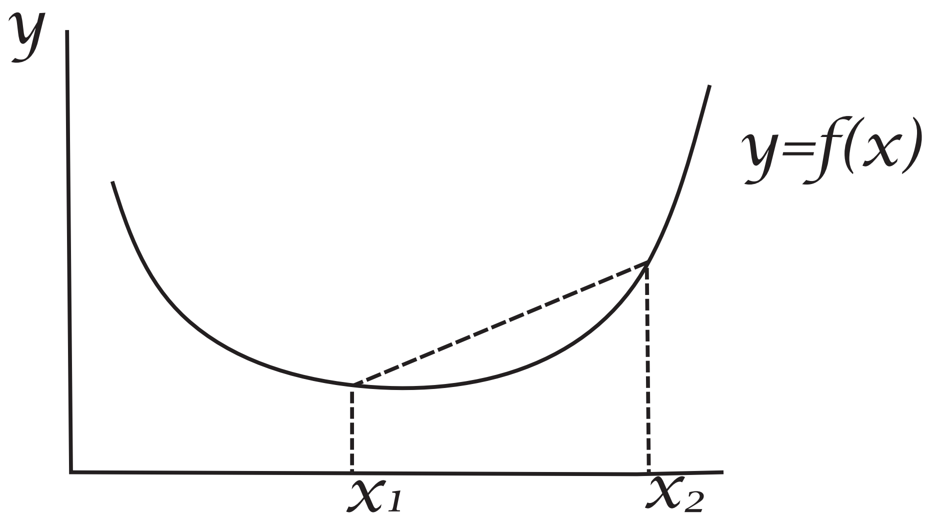
$$\sum_{i=0}^{k} f'(x_i)\,(x_{i+1} - x_i) \le 0.$$

これは高校の範囲内の知識で示せる。

巡回的単調性と凸関数 (II)

$f''(x) > 0$ をみたす関数は次の凸性をみたす。
$0 < \lambda < 1$ のとき

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2).$$



T の求め方

π_n の測度が巡回的単調な集合の上に載っていることからその極限 π も巡回的単調な集合の上に分布している確率とわかる。さらに

定理 (Rockafeller) 巡回的単調な集合はある凸関数 ϕ の微分の関数のグラフ

$$G(\nabla\phi) = \{(x, \nabla\phi(x)) \mid x \in \mathbb{R}^d\}$$

の上にある。
したがって、π の分布している集合がやはり凸関数の微分のグラフ上にあることがわかる。$T = \nabla\phi$ である。

$\mathbb{R}^d$ 上の関数の微分

一変数関数 $y = f(x)$ の凸性は述べたが、$\mathbb{R}^d$ 上の関数 $\phi(x)$ $(x = (x_1, \ldots, x_d))$ の凸性、微分は次のように定義される。

- ϕ が凸： 任意の $x, y \in \mathbb{R}^d$, $0 < \lambda < 1$ に対して
  $\phi\,(\lambda x + (1-\lambda)y) \le \lambda\phi(x) + (1-\lambda)\phi(y)$
- ϕ の微分 (グラジエント)：

$$\nabla\phi(x) = \left(\frac{\partial\phi}{\partial x_1}(x), \ldots, \frac{\partial\phi}{\partial x_d}(x)\right).$$

Brenier の定理の補足 (I)

- μ の確率密度関数を $f(x)$,
  ν の確率密度関数を $g(x)$
- $D = \{x \in \mathbb{R}^d \mid f(x) > 0\}$,
  $D' = \{x \in \mathbb{R}^d \mid g(x) > 0\}$

Brenier の定理の T は D から D' への写像で $\mathbb{R}^d$ で定義された凸関数 $\phi(x)$ を用いて

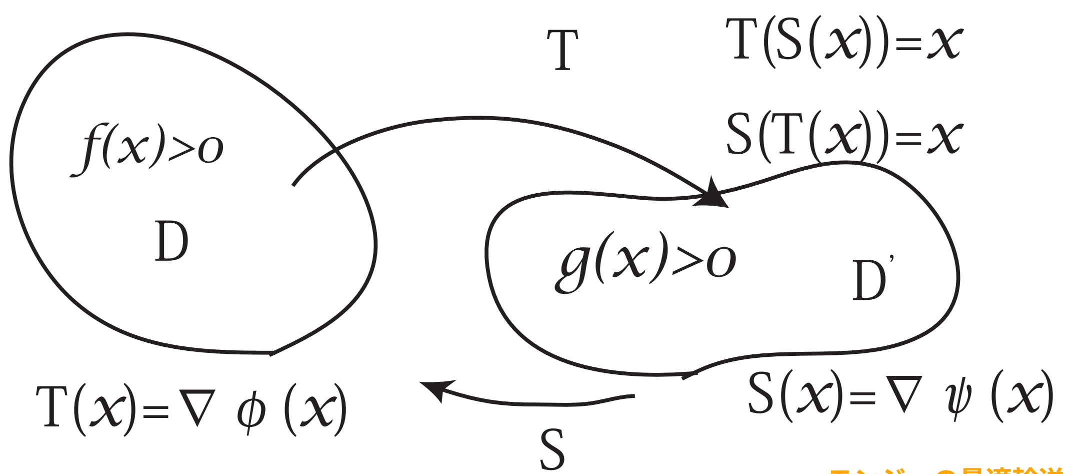
$$T(x) = \nabla\phi(x), \quad x \in D$$

と書ける。

Brenierの定理の補足 (II)

同様に ν を μ に写す最適輸送写像 S も存在する。
この S について：

- S は D' から D への写像である。
- S もある凸関数の微分になる。
- T と S は互いに逆写像である。



一次変換と平行移動の例について

- B : 2×2 行列で $\det B \neq 0$(すなわち逆行列が存在する) とする。
- v : 2 次元ベクトル
- $\phi(x) = \frac{1}{2}(Bx, Bx) + (v, x)$ とおく $(x \in \mathbb{R}^2)$.

この ϕ は凸関数で $T(x) = \nabla\phi(x) = {}^tBBx + v$.

$A = {}^tBB = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ とおくと

$\det A > 0, a_{11} > 0, a_{12} = a_{21}$ が成立する。この $T(x)$ が以前、「一次変換と平行移動」に出て来たものである。

回転行列について

- $A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$.
- $T(x) = Ax \ (x \in \mathbb{R}^2)$.

これは原点のまわりの角度 θ の回転を表すが、この行列の成分が $a_{12} = a_{21}, a_{11} > 0$ をみたすのは $\sin\theta = 0, \cos\theta = 1$ の単位行列に限る。したがって、それ以外は $T(x)$ は最適輸送写像ではない。

$\Longrightarrow$ 合同な図形

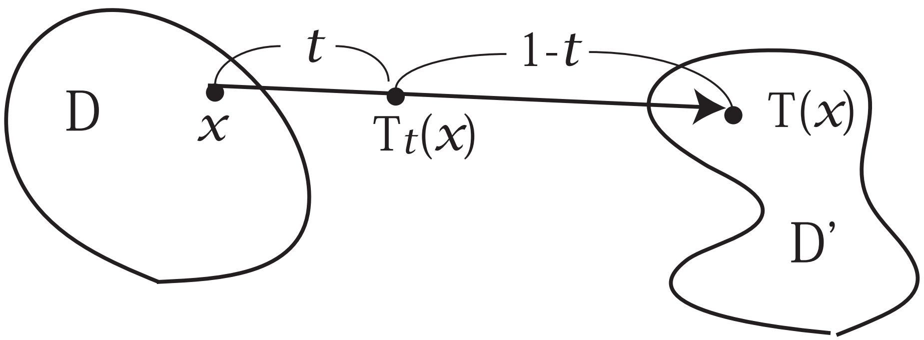
最適輸送写像の性質

μ を ν に写す Brenier の写像 T は次の性質を持つ：

(1) $T_t(x) = (1-t)x + tT(x)$ と定義する。

(2) $(T_t)_\sharp\mu = \mu_t$ とおく。($\mu_0 = \mu, \mu_1 = \nu$ である)。

定理 (McCann) T_t は μ を μ_t に写す最適輸送写像である。



⟹ リーマン多様体としての $\mathcal{P}(\mathbb{R}^d)$ のページ

最適輸送写像の応用 (I)

前ページで 2 点 x と $T(x)$ を $t : (1-t)$ の比で内分する点 $T_t(x)$ を取るとこれがやはり最適輸送写像であることがわかった。

$$\boxed{\begin{array}{lcl} \mu & : & D_1\text{上の一様分布}, \\ \nu & : & D_2\text{上の一様分布} \end{array}}$$

のとき $T_0(D_1) = D_1, T_1(D_1) = D_2$.
したがって、

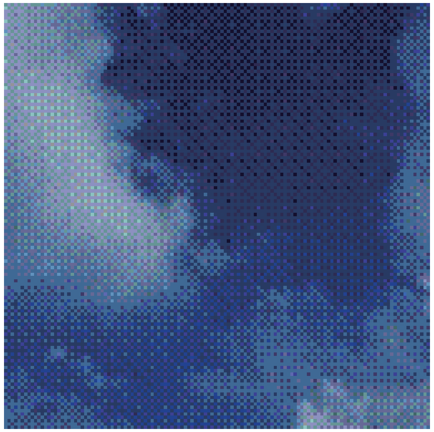
$$\boxed{T_t(D_1)\text{ は }D_1\text{ と }D_2\text{ を補間する図形}}$$

と言える。

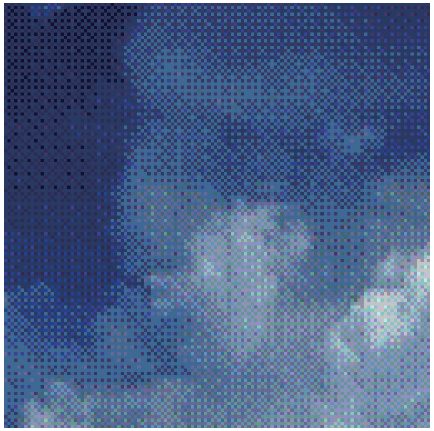
最適輸送写像の応用 (II)

ある異なる時刻の画像が与えられたときその二つの画像を連続的に画像で補間したいことがある。画像データを全時刻得ていないような場合。最適輸送写像はそのような問題に応用されている。
Allen R. Tannenbaum(Georgia Tech University) の web page
http://www.bme.gatech.edu/groups/bil/
にある例をお見せします。

空の画像



始め



終わり

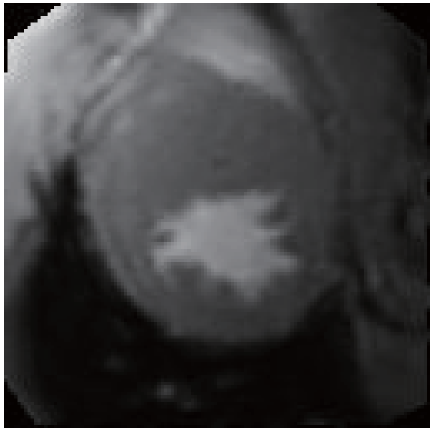
炎の画像


始め


終わり

心臓の画像

始め



終わり

ある不等式

$\Phi(m)$ を $\mathbb{R}$ 上の関数で $\Phi(m_0) = 0$, $\Phi(m) > 0$ $(m \neq m_0)$ とする $(m_0 \in \mathbb{R})$

$$\boxed{\begin{array}{l}(\star)\ 2\Phi(m) \leq \Phi'(m)^2\ \forall m \in \mathbb{R} \\ \Longrightarrow |m - m_0|^2 \leq 2\Phi(m)\ \forall m \in \mathbb{R}\end{array}}$$

$\Phi(m) = \frac{1}{2}|m - m_0|^2$ で等号が成立する。また、$\mathbb{R}^n$ 上の関数 $\Phi(m)$ $(m \in \mathbb{R}^n)$ が $\Phi(m_0) = 0$, $\Phi(m) > 0$ $(m \neq m_0)$ とする $(m_0 \in \mathbb{R}^n)$. このとき、

$$\boxed{\begin{array}{l}(*)\ 2\Phi(m) \leq |\nabla\Phi(m)|^2\ \forall m \in \mathbb{R}^n \\ \Rightarrow |m - m_0|^2 \leq 2\Phi(m)\ \forall m \in \mathbb{R}^n\end{array}}$$

対数ソボレフ不等式

$f(x) : \mathbb{R}^d$ 上の関数で $f(x) > 0$, $\int_{\mathbb{R}^d} f(x) m_\rho(dx) = 1$ とする。次の不等式 (対数ソボレフ不等式) が成立する：

$$\int_{\mathbb{R}^d} f(x) \log f(x) m_\rho(dx) \leq \frac{1}{2} \int_{\mathbb{R}^d} \frac{|\nabla f(x)|^2}{f(x)} m_\rho(dx).$$

$m_\rho(dx)$ は**標準正規分布**. この不等式は場の量子論の研究と関連して Leonard Gross により 1975 年に発見された不等式である。

$\Longrightarrow$ Stam の不等式

輸送コスト不等式

定理 (Talagrand, 1996)

m_f :密度関数 $f(x)$ をもつ確率分布で

$\int_{\mathbb{R}^d} |x|^2 f(x)dx < \infty$ とする。

$$W_2(m_f, m_\rho) = \left\{\inf\{I(\pi) \mid \pi \in \Pi(m_f, m_\rho)\}\right\}^{1/2}$$

とおくと

$$W_2(m_f, m_\rho)^2 \le 2 \int_{\mathbb{R}^d} f(x) \log\left(\frac{f(x)}{\rho(x)}\right) dx.$$

確率分布の空間上の解析

$$\mathcal{P}(\mathbb{R}^d) = \left\{ m_f \,\middle|\, m_f \text{ は確率密度関数 } f \text{ をもつ } \mathbb{R}^d \text{ 上の連続型確率分布で } \int_{\mathbb{R}^d} |x|^2 f(x)dx < \infty \right\}$$

$$\Phi(m_f) = H(m_f \mid m_\rho) = \int_{\mathbb{R}^d} f(x) \log\left(\frac{f(x)}{\rho(x)}\right) dx$$

とおくと (★), (*) のように

対数ソボレフ不等式 (LSI) $\Longrightarrow$ 輸送コスト不等式 (TI)

が Otto-Villani らにより示されている。時間があればこの話を最後にします。

リーマン多様体

集合 M がリーマン多様体であるとは：

- M は局所的にはユークリッド空間の開集合と微分同相である
- M 上の滑らかな曲線 $c(t)$ に対して接ベクトル $c'(t)$ とその大きさ $|c'(t)|$ が定まるもの

接ベクトルの大きさを定めるものをリーマン計量という。

例

ユークリッド空間

- $M = \mathbb{R}^d$
- $c(t) = (c_1(t), \ldots, c_d(t))$ に対して、

$$
\begin{aligned}
c'(t) &= (c'_1(t), \ldots, c'_d(t)) \\
|c'(t)| &= \left\{ \sum_{i=1}^{d} c'_i(t)^2 \right\}^{1/2}
\end{aligned}
$$

例

球面

- $M = S^{d-1} = \{x = (x_1, \ldots, x_d) \mid \sum_{i=1}^{d} x_i^2 = 1\}$
- $c(t) = (c_1(t), \ldots, c_d(t)) \subset S^{d-1} \subset \mathbb{R}^d$ に対して

$$\begin{aligned} c'(t) &= (c_1'(t), \ldots, c_d'(t)) \in T_{c(t)}S^{d-1} \\ &= c(t)\text{ における }S^{d-1}\text{ の接平面} \subset \mathbb{R}^d \\ |c'(t)| &= \left\{ \sum_{i=1}^{d} c_i'(t)^2 \right\}^{1/2} \end{aligned}$$

リーマン多様体上の距離と測地線

- p, q: リーマン多様体 M の２点
- p, q の間の距離を

$$d(p,q) = \min\left\{ \int_0^1 |c'(t)| dt \;\middle|\; c(t) \text{ は } M \text{ 上の } c(0)=p,\ c(1)=q \text{ を満たす曲線} \right\}$$

と定義する。min は右辺の集合の中の最小値を取ることを意味する。$l(c) = \int_0^1 |c'(t)| dt$ は曲線 c の長さである。

リーマン多様体上の距離と測地線

$d(p,q)$ の右辺の性質をみたす曲線の中で最小値を取る曲線を最短測地線と言う。

グラジエントベクトル場

M がリーマン多様体のとき、関数 $\Phi : M \to \mathbb{R}$ に対して、グラジエントベクトル場 $\mathrm{grad}\Phi(m)$ が定まる。微分方程式

$$\begin{aligned} \frac{d}{dt}m(t) &= -\mathrm{grad}\Phi(m(t)) \\ m(0) &= m_0 \end{aligned}$$

の解をグラジエントフローと言う。

- $\Phi(m(t))$ は t の減少関数である。
- Φ が m_0 で最小値を取るなら、$m(t)$ は $t \to \infty$ で m_0 に収束することが期待できる。

リーマン多様体上での不等式(*)

M をリーマン多様体とすると (*) の一般化として次が成立する：$\Phi(m)$ は M 上の関数で $\Phi(m_0) = 0$, $\Phi(m) > 0$ $(m \neq m_0)$ を満たすならば (+ ある仮定で)

$$
\boxed{\begin{aligned} 2\Phi(m) \leq |\mathrm{grad}\Phi(m)|^2 \ \forall m \in M \\ \Rightarrow d(m, m_0)^2 \leq 2\Phi(m) \forall m \in M \end{aligned}}
$$

証明には Φ によるグラジエントフローが使われる。

リーマン多様体としての$\mathcal{P}(\mathbb{R}^d)$ (I)

ここまで考えてきたリーマン多様体は有限次元のリーマン多様体。しかし、“無限次元”のリーマン多様体の定義もある。

例えば：
ヒルベルト多様体, etc

$\mathcal{P}(\mathbb{R}^d)$ はそのような無限次元リーマン多様体のようなものでは無い。しかし、形式的には、リーマン多様体と思えて次がわかる：

リーマン多様体としての$\mathcal{P}(\mathbb{R}^d)$ (II)

- $W_2(m_f, m_g)$ は $\mathcal{P}(\mathbb{R}^d)$ 上の距離(Wasserstein distance) を定めることが知られているが、この距離はリーマン多様体としての距離 $d(m_f, m_g)$ と一致する。
- **「最適輸送写像の性質」**で述べた $\{\mu_t\}_{0\le t\le 1}$ は μ と ν を結ぶ最短測地線である。

リーマン多様体としての $\mathcal{P}(\mathbb{R}^d)$ (III)

$g \in \mathcal{P}(\mathbb{R}^d)$ ($g(x) > 0\ \forall x$ とする) を固定する。
相対エントロピー
$\Phi(m_f) = H(m_f|m_g) = \int_{\mathbb{R}^d} F(x)\log F(x) m_g(dx)$ について次が成立する。ただし、$F(x) = \dfrac{f(x)}{g(x)}$ である。
すなわち、$F(x)$ は m_f の m_g をもとに考えた密度関数である。

(1) $\Phi(m_f) \geq 0$

(2) $\Phi(m_f) = 0 \iff f = g$

リーマン多様体としての$\mathcal{P}(\mathbb{R}^d)$ (IV)

次の同値性が示せる ($F(x) = \frac{f(x)}{g(x)}$ である)。

$(*1)$ (LSI) $2\alpha\Phi(m_f) \le \int_{\mathbb{R}^d} \frac{|\nabla F(x)|^2}{F(x)} dm_g(x) \quad \forall f$

$(*2)$ $2\alpha\Phi(m_f) \le |\mathrm{grad}\Phi(m_f)|^2 \quad \forall f$

$(\star 1)$ (TI) $W_2(m_f, m_g)^2 \le \frac{2}{\alpha}\Phi(F) \quad \forall F$

$(\star 2)$ $d(m_f, m_g)^2 \le \frac{2}{\alpha}\Phi(F) \quad \forall F$

$(*1) \Longleftrightarrow (*2)$, $(\star 1) \Longleftrightarrow (\star 2)$

有限次元の類推から、$(*2)$ が成立すると仮定すると $(\star 2)$ が示せそうである。実際それは正しい。

$\mathcal{P}(\mathbb{R}^d)$ 上のグラジエントフロー

Φ によるグラジエントフローの方程式

$$\begin{aligned} \frac{d}{dt}m(t) &= -\mathrm{grad}\Phi(m(t)) \\ m(0) &= m_{h_0 g} \end{aligned}$$

の解を $m(t) = m_{h(t)g}$ と書いたとき $h(t,x)$ は次の熱方程式に従う：

$$\begin{aligned} \frac{\partial h}{\partial t}(t,x) &= \Delta h(x) - (\nabla U(x), \nabla h(x)) \\ h(0,x) &= h_0(x) \end{aligned}$$

ただし、$U(x) = -\log g(x)$ としている。このことを用いて、$\inf \nabla^2 U(x) > -\infty$ のとき (LSI)→(TI) が示せる。

非線形偏微分方程式への応用

熱方程式は従来、$L^2(\mathbb{R}^d, m_g)$ 上の汎関数

$$E(f) = \frac{1}{2}\int_{\mathbb{R}^d} |\nabla f(x)|^2 m_g(dx)$$

によるグラジエントフローととらえられてきた。今回の新しい見方は Felix Otto によるもの (1998 年ぐらい) で、その拡張も含めて現在発展中の分野である。

数学用語

- d 次元ユークリッド空間 $\mathbb{R}^d$：
  d 個の実数の組みからなる集合
  $\mathbb{R}^d = \{(x_1, \ldots, x_d) \mid x_i \in \mathbb{R}\}$ のこと。$d = 1$ のときは数直線、$d = 2$ のときは平面、$d = 3$ のときは空間を表している。数学では $d \geq 4$ に対しても考える。

  $\Longrightarrow$ 最大・最小の存在定理　$\Longrightarrow$ 確率分布

- 連続関数:
  x の関数 $y = f(x)$ が連続関数であるとは、x が連続的に動くとき、それとともに $f(x)$ も連続的に動くような関数を言う。正確には ε-δ 論法で定式化され、理系の学部では通常大学 1 年次に勉強することになります。

数学用語

- 部分集合 $K \subset \mathbb{R}^d$ がコンパクト集合：
  A は $\mathbb{R}^d$ の部分集合で大きさが有限 (非常に大きな円、球を考えるとその中に入ってしまう) かつ、その境界も A に含まれているような集合. 例えば

$$\{(x,y) \mid x^2+y^2 \le 1\} = \text{半径 1 の円板}$$
$$\{(x,y) \mid |x| \le 1, |y| \le 1\} = \text{一辺の長さ 2 の正方形}$$

  など.

$$\{(x,y) \mid x+y \le 1, x > 0, y \ge 0\}$$
$$= \text{三角形で } y \text{ 軸上にある辺が欠けているもの}$$

  はコンパクト集合ではない. $\Longrightarrow$ 最適解の存在

数学用語

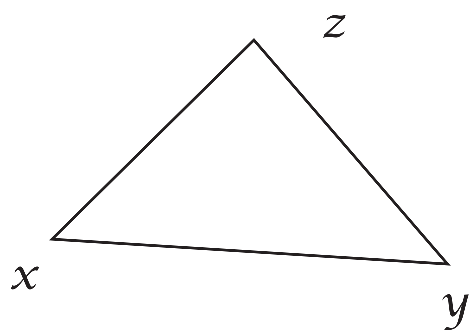
- 最大・最小の存在定理

「コンパクト集合上の連続関数は最大値・最小値を取る」と主張する定理。
コンパクト性が無いと最大値・最小値を取らないことがあることに注意して下さい。

例　開区間 $(0,1) = \{x \mid 0 < x < 1\}$ 上の連続関数 $f(x) = x$.

数学用語

- **距離空間**　　集合 X が距離空間であるとは, X の要素 x, y の間に距離 $d(x, y)$ が定まり、次の性質が成立しているときに言う：
  - すべての x, y について $d(x, y) \geq 0$
  - $d(x, y) = 0 \Longleftrightarrow x = y$
  - (三角不等式)　すべての x, y, z について $d(x, z) \leq d(x, y) + d(y, z)$



$\Longrightarrow$ 最大・最小の存在定理

数学用語

積分 $\int_A f(x)dx$ を次のように定義する。$f(x)$ は x が原点から離れて行くと十分早く 0 に収束する関数とする。

(1) A を小さい集合 A_i に分ける。

(2) 和 $\sum_i f(P_i)|A_i|$ を取る。P_i は A_i に属す点, $|A_i|$ は A_i の体積。

(3) (1) の分割をどんどん細かくして行くとき、(2) の和がある一定の値に収束するとする。その極限値を $\int_A f(x)dx$ と定義する。

$\Longrightarrow \int_{D_1} c(x, T(x))dx$

$\Longrightarrow$ 密度関数をもつ連続型確率分布の定義

数学用語

Birkoff の定理

$n \times n$ 行列 $X = (x_{ij})$ が bistochastic matrix(両側確率行列) とはすべての $1 \le i \le n, 1 \le j \le n$ について

$$\sum_{i=1}^{n} x_{ij} = 1, \qquad \sum_{j=1}^{n} x_{ij} = 1.$$

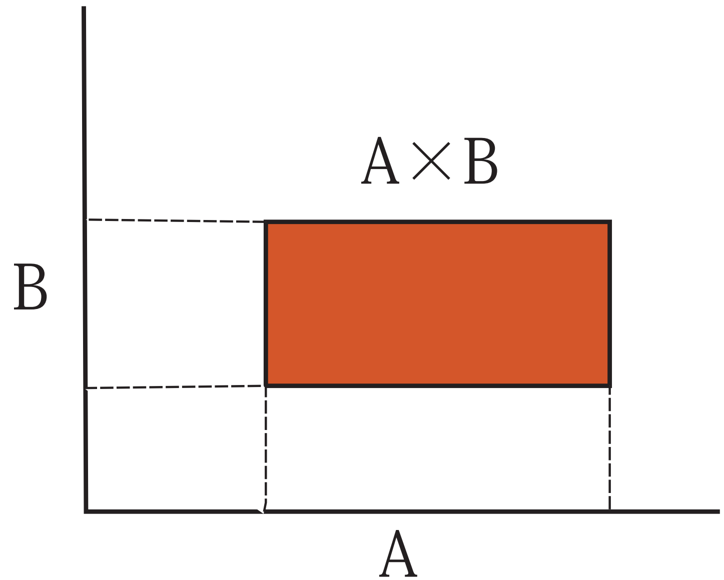
両側確率行列全体の集合上で定義された関数 $C(X) = \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}x_{ij}$ は、適当な $\{1, \ldots, n\}$ の並べかえ $\{k_1, \ldots, k_n\}$ が存在して $x_{ik_i} = 1$ となるような確率行列で最大・最小を取る。 $\Longrightarrow$ 離散モンジュの問題 (II)

数学用語

- **積集合** $A \subset \mathbb{R}^d, B \subset \mathbb{R}^d$ に対して

$$A \times B = \{(x, y) \mid x \in A, y \in B\}.$$

$A = \mathbb{R}^d, B = \mathbb{R}^d$ のときの積空間 $\mathbb{R}^d \times \mathbb{R}^d$ は $\mathbb{R}^{2d}$ と同じになる。 $\Longrightarrow$ MK **問題のページ**



数学用語

- 周辺分布　　積空間

$\mathbb{R}^{2d} = \{(x, y) \mid x \in \mathbb{R}^d, y \in \mathbb{R}^d\}$ 上の確率分布 π に対して

$$
\begin{aligned}
\pi_1(A) &= \pi\left(A \times \mathbb{R}^d\right) \quad (A \subset \mathbb{R}^d) \\
\pi_2(B) &= \pi\left(\mathbb{R}^d \times B\right) \quad (B \subset \mathbb{R}^d)
\end{aligned}
$$

を π の周辺分布と言う。MK 問題の条件を言い換えると π の最初の d 次元空間と後の d 次元空間の周辺分布がそれぞれ μ, ν であることである。$\Longrightarrow$ MK **問題の定式化**

数学用語

Stam は 1959 年に次の不等式を証明した。

$$\boxed{\mathcal{N}(f)I(f) \geq d.}$$

ここで

$$\mathcal{N}(f) = \frac{\exp\left(-\frac{2}{d}\int_{\mathbb{R}^d} f(x)\log f(x)dx\right)}{2\pi e},$$

$$I(f) = \int_{\mathbb{R}^d} \frac{|\nabla f(x)|^2}{f(x)} dx.$$

$\mathcal{N}(f)$: Schannon's entropy power functional

$I(f)$: f の Fisher information

と呼ばれる。 $\Longrightarrow$ 対数ソボレフ不等式のページ