対角スケーリングと Wasserstein 空間

2016 年 9 月 12–13 日
北大ミニワークショップ「統計多様体の幾何学とその周辺 (8)」
清 智也　（東京大学）

※発表後に頂いたコメントなどは脚注に記しています．

0 Introduction

n 次正定値実対称行列全体を $\mathrm{Sym}^+(n)$ と表す．$S \in \mathrm{Sym}^+(n)$ に対し，両側から正定値の対角行列 D を掛けて別の行列 DSD を求める操作を一般に**対角スケーリング**と呼ぶ[1]．ここで D は S に依存してよい．典型例として，統計学では共分散行列 S に対して相関行列を求める操作

$$S_{ij} \mapsto \frac{S_{ij}}{\sqrt{S_{ii}}\sqrt{S_{jj}}} \tag{1}$$

がある．共分散行列に馴染みがなければ，$\mathbb{R}^n$ の基底 $x_1, \ldots, x_n \in \mathbb{R}^n$ に対するグラム行列 $S_{ij} = x_i^\top x_j$ を考えればよい．ここで $x_i^\top$ は x_i の転置を表す．

また別の対角スケーリングとして，次の定理が知られている．

定理 1 (Marshall & Olkin [4])**.** 任意の $S \in \mathrm{Sym}^+(n)$ に対して，

$$\sum_j D_i S_{ij} D_j = 1 \quad \forall i \tag{2}$$

を満たす正定値の対角行列 D が一意的に存在する．

式 (2) は，スケーリング後の行列 $B = DSD$ の行和（= 列和）が 1 になることを意味している．このように，行和が 1 となる正定値対称行列 B をここでは equisum 行列と呼ぼう．定理 1 の統計学への応用は [8] で与えられている[2]．

ところで，統計学やリスク解析の分野では確率分布を「座標軸ごとの単調変換」によって基準化するという考え方がある．その典型例はコピュラ (copula; たとえば [6]) であり，周辺分布が一様分布となるように確率分布を変換するというものである．この変換は，ガウス分布に制限すると，本質的には共分散行列から相関行列を求める操作，すなわち式 (1) の変換に対応している．こうして次の対応表が得られる．

共分散	確率分布
相関行列	コピュラ
equisum 行列	?

本研究の興味は「？」の部分であるが，その前に共分散行列の空間における対角スケーリングを幾何学的な観点から整理しておきたい．そのための道具が Wasserstein 空間である．

本稿の構成は以下の通りである．まず 1 節ではガウス分布に制限した Wasserstein 空間の基本的性質を復習する．続いて 2 節では対角スケーリングを Wasserstein 空間の枠組みで記述し，特に Marshall-Olkin の定理がエネルギー最小化として得られることを指摘する．最後に 3 節で，確率分布全体の空間に拡張した結果の概要を述べる．

---

[1] 正定値でない行列や，両側から別々の対角行列を掛ける場合もある．

[2] ただしそこでは equisum 行列のことを bi-unit 行列と呼んでいる．

1 Gaussian Wasserstein space

[9] を参考に，正定値対称行列の空間 $\mathcal{M}$ に Wasserstein 距離 d と Wasserstein 計量 g を別々に定義し，g に関する測地距離が d に一致することを確認する[3]．また，$\mathcal{M}$ 上のエネルギー関数[4]を定義し，その停留点と勾配流について簡単に述べる．

1.1 Wasserstein distance

$\mathcal{M} = \mathrm{Sym}^+(n)$ を n 次の正定値実対称行列全体とする．$S \in \mathcal{M}$ および正方行列 $A, B \in \mathbb{R}^{n\times n}$ に対して，$\langle A, B\rangle_S = \mathrm{tr}(ASB^\top)$, $\|A\|_S = \langle A, A\rangle_S^{1/2}$ とおく．それぞれ内積，ノルムの性質を満たす．

$S_0, S_1 \in \mathcal{M}$ に対して，

$$d(S_0, S_1) = \inf\left\{\|A - I\|_{S_0} \mid A \in \mathbb{R}^{n\times n},\quad AS_0A^\top = S_1\right\} \tag{3}$$

と定義する[5]．

注意 1. 平均 0，共分散行列 S のガウス分布を $N(0, S)$ と記す．$X_0 \sim N(0, S_0)$, $X_1 = AX_0$, $AS_0A^\top = S_1$ のとき，$X_1 \sim N(0, S_1)$ であり，$\|A - I\|_{S_0} = (\mathrm{E}[\|X_1 - X_0\|^2])^{1/2}$ となる．つまり，$d(S_0, S_1)$ は $N(0, S_0)$ を $N(0, S_1)$ に線形輸送するときのコストを表している．

補題 1. 式 (3) の inf は A が正定値対称行列のとき達成され，一意的に

$$A = S_0^{-1/2}(S_0^{1/2}S_1S_0^{1/2})^{1/2}S_0^{-1/2} \tag{4}$$

で与えられる[6]．また距離自体は $d(S_0, S_1) = \mathrm{tr}(S_0 - 2(S_0^{1/2}S_1S_0^{1/2})^{1/2} + S_1)$ となる．

Proof. $\|A-I\|_{S_0}^2 = \mathrm{tr}(S_1 - 2AS_0 + S_0)$ より $\mathrm{tr}(AS_0)$ を最大化すればよい．条件 $AS_0A^\top = S_1$ を満たす A は，直交行列 Q を用いて $S_0^{1/2}AS_0^{1/2} = (S_0^{1/2}S_1S_0^{1/2})^{1/2}Q$ と書ける．すると

$$\mathrm{tr}(AS_0) = \mathrm{tr}(S_0^{1/2}AS_0^{1/2}) = \mathrm{tr}((S_0^{1/2}S_1S_0^{1/2})^{1/2}Q)$$

となり，これは Q が単位行列のとき最大となる． □

命題 1. d は距離となる．

Proof. (i) $d(S_0, S_1) \geq 0$ は明らか．また補題より $d(S_0, S_1) = 0 \Leftrightarrow A = I \Leftrightarrow S_0 = S_1$.
(ii) $AS_0A^\top = S_1$ ならば $A^{-1}S_1A^{-\top} = S_0$ より，$\|A - I\|_{S_0} = \|I - A^{-1}\|_{S_1}$ となる．A について inf をとれば $d(S_0, S_1) = d(S_1, S_0)$.
(iii) $AS_0A^\top = S_1$, $BS_1B^\top = S_2$ とすれば $(BA)S_0(BA)^\top = S_2$ だから，

$$\begin{aligned} d(S_0, S_2) &\leq \|BA - I\|_{S_0} \\ &\leq \|BA - A\|_{S_0} + \|A - I\|_{S_0} \\ &= \|B - I\|_{S_1} + \|A - I\|_{S_0} \end{aligned}$$

が成り立つ．A, B について inf をとれば三角不等式を得る． □

d を **Wasserstein 距離**といい，$(\mathcal{M}, d)$ を **Gaussian Wasserstein 空間**という．

[3]本稿では平均ベクトルが 0 のガウス分布しか考慮せず，したがって共分散行列だけがパラメータとなる．
[4]正確には自由エネルギーという．
[5]d は次の意味で直交不変である：任意の直交行列 Q に対して $d(QS_0Q^\top, QS_1Q^\top) = d(S_0, S_1)$.
[6]式 (4) の右辺を S_0 と S_1 の幾何平均作用素という．

1.2 Wasserstein metric

多様体 $\mathcal{M} = \mathrm{Sym}^+(n)$ にリーマン計量を導入する．$S \in \mathcal{M}$ を中心とする局所座標を

$$\varphi_S(A) = ASA \in \mathcal{M}$$

によって定義する．ただし A は正定値対称行列を動くものとする．この写像は微分同相である．

φ_S の $A = I$ における微分は，

$$d\varphi_S|_{A=I}(X) = XS + SX, \quad X \in \mathrm{Sym}(n),$$

となる．この X を接ベクトルと考え，その全体を $T_S\mathcal{M}$ と記す．$X, Y \in T_S\mathcal{M}$ に対して

$$g_S(X,Y) = \langle X, Y\rangle_S = \mathrm{tr}(XSY)$$

と定義する．この計量 g を **Wasserstein 計量**と呼ぶ．

与えられた曲線 $\{S_t\}_{0\le t\le 1}$ に対し，時刻 t における接ベクトル X_t は

$$\frac{dS_t}{dt} = X_tS_t + S_tX_t$$

より定まる．この X_t を用いると，$\{S_t\}$ の長さは

$$L = L(\{S_t\}) = \int_0^1 \sqrt{g_{S_t}(X_t, X_t)}dt$$

で与えられる．S_0 と S_1 の間の測地距離 $d_g(S_0, S_1)$ とは，S_0 と S_1 をつなぐ曲線 S_t の長さの下限のことである．

補題 2. 任意の曲線 S_t に対し，$d(S_t, S_{t+dt}) = \sqrt{g_{S_t}(X_t, X_t)}dt + o(dt)$ が成り立つ．

Proof. $S_{t+dt} = (I + X_tdt + o(dt))S_t(I + X_tdt + o(dt))$ に注意すれば，

$$d(S_t, S_{t+dt}) = \|(I + X_tdt + o(dt)) - I\|_{S_t} = \|X_t\|_{S_t}dt + o(dt) = \sqrt{g_{S_t}(X_t, X_t)}dt + o(dt)$$

となる．なおこの評価は t について一様に成り立つ． □

定理 2. $d_g = d$ である．

Proof. $d_g \ge d$ の証明：S_0 と S_1 を結ぶ任意の曲線 S_t に対して，

$$\begin{aligned}
L(\{S_t\}) &= \lim_{m\to\infty} \sum_{i=0}^{m-1} \sqrt{g_{S_{i/m}}(X_{i/m}, X_{i/m})}\frac{1}{m} \quad \text{（リーマン積分）} \\
&= \lim_{m\to\infty} \sum_i d(S_{i/m}, S_{(i+1)/m}) \quad \text{（補題 2 より）} \\
&\ge d(S_0, S_1) \quad \text{（三角不等式）}
\end{aligned}$$

が成り立つ．曲線について inf をとれば $d_g(S_0, S_1) \ge d(S_0, S_1)$ を得る．

$d_g \le d$ の証明：$S_1 = (I + X)S_0(I + X)$, $I + X \in \mathrm{Sym}^+(n)$, となる X を用いて $S_t = (I + tX)S_0(I + tX)$ とおく．すると任意の $s, t \in [0,1]$ に対して

$$\begin{aligned}
d(S_s, S_t) &= \|(I + tX)(I + sX)^{-1} - I\|_{S_s} \quad [\because (I + tX)(I + sX)^{-1} \in \mathrm{Sym}^+(n)] \\
&= \|(I + tX) - (I + sX)\|_{S_0} \\
&= |t - s|\|X\|_{S_0} \\
&= |t - s|d(S_0, S_1)
\end{aligned}$$

が成り立つ．よって補題 2 から $\sqrt{g_{S_t}(X_t, X_t)}dt = d(S_t, S_{t+dt}) = d(S_0, S_1)dt$ であり，$L(\{S_t\})$ の定義から $L(\{S_t\}) = d(S_0, S_1)$ となる．特に $d_g(S_0, S_1) \le d(S_0, S_1)$ である． □

上の証明から分かる通り，$t \mapsto (I + tX)S_0(I + tX)$ が $(\mathcal{M}, g)$ の測地線となる．

1.3 Energy function(al)

K を半正定値対称行列とし，エネルギー関数を

$$E(S) = -\frac{1}{2}\log|S| + \frac{1}{2}\mathrm{tr}(KS), \quad S \in \mathcal{M},$$

とおく．これは K が正定値のときは $N(0,S)$ から $N(0,K^{-1})$ への Kullback-Leiblers ダイバージェンスと（K のみに依存する項を除いて）等しい．

$E(S)$ が $\mathrm{Sym}^+(n)$ 上の凸関数であることはよく知られている．実は次の結果が成り立つ．

定理 3. E は $(\mathcal{M}, g)$ の測地線に関して狭義凸である．

Proof. S_0 を固定し，測地線 $S_t = (I+tX)S_0(I+tX)$ を考えると

$$E(S_t) = -\frac{1}{2}\log|S_0| - \log|I+tX| + \frac{1}{2}\mathrm{tr}(K(I+tX)S_0(I+tX))$$

となる．この関数の 1 階，2 階微分を計算すると

$$\frac{d}{dt}E(S_t) = -\mathrm{tr}((I+tX)^{-1}X) + \mathrm{tr}(KXS_0(I+tX)),$$
$$\frac{d^2}{dt^2}E(S_t) = \mathrm{tr}((I+tX)^{-1}X(I+tX)^{-1}X) + \mathrm{tr}(KXS_0X) \geq 0$$

となる．また $X \neq 0$ ならば不等号は $>$ となる．よって狭義凸である． □

一般に Wasserstein 空間の測地線に関する凸性を **displacement convexity** [5] と呼ぶ[7]．

補題 3. E が停留点を持つための必要十分条件は K が正定値となることである．また K が正定値でないとき，E は下に非有界である．

Proof. E を S の標準座標で微分すれば $(-S^{-1}+K)/2$ となる．よって K が正定値のとき，またそのときに限って $S = K^{-1}$ が停留点となる．

次に K が正定値でない（特に正則でない）とする．Q を直交行列とするとき，$K \mapsto QKQ^\top$，$S \mapsto QSQ^\top$ と置き換えても E は不変であるから，K は対角としても一般性を失わない．また正則でないから $K_{11} = 0$ としてよい．このとき

$$S_t = \begin{pmatrix} t & 0 \\ 0 & I_{n-1} \end{pmatrix}, \quad t > 0,$$

とおけば $E(S_t) = -(1/2)\log t + \mathrm{const.}$ となる．よって E は下に非有界である． □

定理 3 の証明から

$$\begin{aligned}\frac{d}{dt}E(S_t)|_{t=0} &= \mathrm{tr}(-X + KXS_0) \\ &= g_{S_0}(X, -S_0^{-1} + K)\end{aligned}$$

となる．よって E の勾配ベクトルは $\mathrm{grad}E(S) = -S + K \in T_S\mathcal{M}$ となる．

エネルギーの勾配流は，$X_t = -\mathrm{grad}E(S_t) = S_t^{-1} - K$，あるいは

$$\frac{dS_t}{dt} = S_tX_t + X_tS_t = 2I - KS_t - S_tK$$

で与えられる．

例 1. $K = 0$ のとき，$dS_t/dt = 2I$ となり，これを解くと $S_t = S_0 + 2tI$ となる．これは「拡散方程式の解」あるいは「Brown 運動」に対応する．

例 2. $K = I$ のとき，$dS_t/dt = 2(I - S_t)$ となり，これを解くと $S_t = e^{-2t}S_0 + (1-e^{-2t})I$ となる．これは「Ornstein-Uhlenbeck 過程」に対応する．

[7]いま考えている E はさらに，$E(ASA)$ とおいたときに $A \in \mathrm{Sym}^+(n)$ について凸となる．この性質は一般に generalized displacement convexity と呼ばれる [1].

2 Diagonal scaling

本節ではエネルギー関数の最小化によって，対角スケーリングに関する定理 1 が得られることを示す．

2.1 Diagonal restriction of Gaussian Wasserstein space

正の対角行列全体を $\mathcal{D}^+$ と表す．$R \in \mathcal{M}$ に対して，ファイバー $\mathcal{F}_R$ を

$$\mathcal{F}_R = \left\{DRD \mid D \in \mathcal{D}^+\right\} \subset \mathcal{M}$$

と定義する．これは R の可能な対角スケーリング全体を表す．

補題 4. 各ファイバー $\mathcal{F}_R$ は全測地的かつ平坦なリーマン部分多様体である．

Proof. $S_0, S_1 \in \mathcal{F}_R$ に対して，$S_0 = D_0RD_0, S_1 = D_1RD_1$ とおくと，$S_1 = (D_1D_0^{-1})R(D_1D_0^{-1})$ だから，S_0 と S_1 を結ぶ $\mathcal{M}$ の測地線は

$$\begin{aligned} S_t &= ((1-t)I + tD_1D_0^{-1})S_0((1-t)I + tD_1D_0^{-1}) \\ &= ((1-t)D_0 + tD_1)R((1-t)D_0 + tD_1) \end{aligned}$$

と書ける．これは $\mathcal{F}_R$ に含まれている．よって $\mathcal{F}_R$ は全測地的である．

また，Wasserstein 距離は $d(S_0, S_1)^2 = d(D_0RD_0, D_1RD_1)^2 = \sum_{i=1}^n (D_{0i} - D_{1i})^2 R_{ii}$ となるから平坦である．特に D はアファイン座標となっている． □

注意 2. $\mathcal{F}_R$ は幾何学的には簡単すぎるが，対角スケーリングと Gaussian Wasserstein 空間の相性の良さを表しているとも言える．

2.2 Diagonal restriction of energy function

K を半正定値行列とし，エネルギー関数

$$E(S) = -\frac{1}{2}\log|S| + \frac{1}{2}\mathrm{tr}(KS)$$

を考える．$S = DRD \in \mathcal{F}_R$ のとき

$$\begin{aligned} E(DRD) &= -\frac{1}{2}\log|R| - \log|D| + \frac{1}{2}\mathrm{tr}(KDRD) \\ &= -\frac{1}{2}\log|R| - \log|D| + \frac{1}{2}\sum_i\sum_j D_i(K_{ij}R_{ij})D_j \end{aligned}$$

となる．E の $\mathcal{F}_R$ への制限を $E|_{\mathcal{F}_R}$ と記す．

補題 5. $E|_{\mathcal{F}_R}$ は displacement convex である[8]．

Proof. E の displacement convexity と $\mathcal{F}_R$ の全測地性より言える． □

補題 6. $E|_{\mathcal{F}_R}$ が停留点を持つための必要十分条件は $K_{ii} > 0$ $(\forall i)$ である．また停留点を持たないときは下に非有界である．

[8] さらに generalized displacement convex であることも言える．

Proof. R_{ij} は正定値だからそのスペクトル分解を $R_{ij} = \sum_k \lambda_k q_{ik} q_{jk}$ とおき，最小固有値を $\lambda_* > 0$ とおく．任意の $x \in \mathbb{R}^n$ に対して

$$\begin{aligned}\sum_i \sum_j x_i (K_{ij} R_{ij}) x_j &= \sum_k \lambda_k \sum_{i,j} K_{ij} (x_i q_{ik})(x_j q_{jk}) \\ &\geq \lambda_* \sum_{i,j} K_{ij} x_i x_j \sum_k q_{ik} q_{jk} \\ &= \lambda_* \sum_i K_{ii} x_i^2\end{aligned}$$

という評価を得る．よって全ての i について $K_{ii} > 0$ ならば $K_{ij} R_{ij}$ は正定値行列となり，このとき E は停留点を持つ．一方，ある i について $K_{ii} = 0$ ならば，半正定値性より K の第 i 行，第 i 列は全て 0 となるので，目的関数に現れる D_i の項は $-\log D_i$ のみとなる．これは下に非有界であり，E は停留点を持たない． □

補題 7. $E|_{\mathcal{F}_R}$ の停留条件は $\sum_j D_i (K_{ij} R_{ij}) D_j = 1$ で与えられる．

Proof. $E(DRD)$ を D_i で偏微分すれば，停留条件は $-1/D_i + \sum_j (K_{ij} R_{ij}) D_j = 0$ となる． □

以上をまとめると次が言える．

定理 4. K を半正定値行列とし，$K_{ii} > 0$ $(\forall i)$ とする．このとき $\sum_j D_i (K_{ij} R_{ij}) D_j = 1$ を満たす $D \in \mathcal{D}^+$ が一意に存在する．

例 3. $K_{ij} = 1$ $(\forall i, j)$ とすれば定理 1 が得られる．

例 4. $K = I$ のとき $D_i R_{ii} D_i = 1$ となり，これは相関行列を求める対角スケーリングである．

$E|_{\mathcal{F}_R}$ の勾配流を求めてみる．$\mathcal{F}_R$ の曲線 $\{S_t = D_t R D_t\}$ の接ベクトル Δ_t は対角行列であり，$dS_t/dt = S_t \Delta_t + \Delta_t S_t$ と書ける．よって，(時刻の添字 t は省略して)

$$\begin{aligned}\frac{dE(S)}{dt} &= -\frac{1}{2}\mathrm{tr}(S^{-1}\frac{dS}{dt}) + \frac{1}{2}\mathrm{tr}(K\frac{dS}{dt}) \\ &= -\mathrm{tr}(\Delta) + \mathrm{tr}(KS\Delta) \\ &= -\sum_{i=1}^n S_{ii}\left(\frac{1}{S_{ii}} - \frac{(KS)_{ii}}{S_{ii}}\right)\Delta_i\end{aligned}$$

となる．よって勾配流は

$$\Delta_i = \frac{1}{S_{ii}} - \frac{(KS)_{ii}}{S_{ii}}$$

と書ける．あるいは

$$\frac{dS_{ij}}{dt} = S_{ij}\Delta_j + \Delta_i S_{ij} = \left(\frac{1}{S_{ii}} - \frac{(KS)_{ii}}{S_{ii}} + \frac{1}{S_{jj}} - \frac{(KS)_{jj}}{S_{jj}}\right) S_{ij}$$

で与えられる．$E|_{\mathcal{F}_R}$ の停留点では $(KS)_{ii} = 1$ となり，上式の右辺は 0 となる．

この勾配流の統計学的あるいは物理的な意味は今のところ明らかではない．少なくとも停留点を求めるアルゴリズムを一つ提供している[9]．

[9] ただしもともと第 1 象限上の滑らかな凸関数の最小化問題なので，汎用パッケージで容易に停留点を求めることができる．

3 A generalization of the Marshall-Olkin theorem

ここでは Gaussian に制限しない Wasserstein 空間を考え，定理 1 の関数空間版を述べる．形式的に議論し，証明は略す．Wasserstein 空間や最適輸送全般については [10] が参考になる．なお定理 1 の別の一般化が [2] に与えられている．

3.1 Wasserstein space not restricted to Gaussian

$\mathcal{P}_2(\mathbb{R}^n)$ を，$\mathbb{R}^n$ 上の確率測度のうち有限な 2 次モーメントを持ち，絶対連続で，かつ平均が 0 となるもの全体とする．平均が 0 というのは便宜上の仮定であり，本質的ではない．

$\mathcal{P}_2(\mathbb{R}^n)$ の距離を

$$d(\mu,\nu) = \inf\left\{\left(\int \|T(x)-x\|^2 d\mu\right)^{1/2} \mid T\sharp\mu = \nu\right\}$$

と定義する．ただし $T\sharp\mu$ は写像 T による測度 μ の押し出しを表す．$(\mathcal{P}_2(\mathbb{R}^n), d)$ を Wasserstein 空間という．上の inf は凸関数 φ の勾配 $T = \nabla\varphi$ で一意的に達成される（Brenier の定理）．

点 μ における接ベクトルは $X(x) = \nabla\varphi(x) - \mathrm{id}(x)$ と表される．情報幾何における接ベクトルの m-表現を X_{m} と表すとき，両者は

$$X_{\mathrm{m}} = \lim_{t\to 0}\frac{1}{t}\left\{(\mathrm{id}+tX)\sharp\mu - \mu\right\}$$

によって対応する．点 μ における接ベクトル X, Y に対し計量を

$$g_\mu(X,Y) = \int X(x)^\top Y(x) d\mu$$

と定義する [7]．μ と $\nu = (\nabla\varphi)\sharp\mu$ を結ぶ測地線は

$$\mu_t = [(1-t)\mathrm{id} + t\nabla\varphi]\sharp\mu = [\mathrm{id}+tX]\sharp\mu$$

で与えられる．測地線は displacement interpolation とも呼ばれる．

$\kappa : \mathbb{R}^n \to \mathbb{R}$ を広義凸関数とする．μ の密度を $\rho = d\mu/dx$ で表し，エネルギー汎関数

$$E(\mu) = \int \rho(x)\log\rho(x)dx + \int \rho(x)\kappa(x)dx \tag{5}$$

を考える．例えば $\kappa(x) = x^\top Kx/2$ かつ $\mu = N(0,S)$ ならば，$E(\mu)$ は 1.3 節で考えたエネルギー関数 $E(S)$ に一致する．

$E(\mu)$ は測地線に関して狭義凸関数，つまり displacement convex である．もし $Z = \int e^{-\kappa(x)}dx$ が有限ならば，$E(\mu)$ の停留点は $\rho(x) = e^{-\kappa(x)}/Z$ で与えられる．

$E(\mu)$ の勾配流を求めると，Fokker-Planck 方程式

$$\partial_t\rho = \nabla^\top(\nabla\rho + \rho\nabla\kappa)$$

が得られる（例えば [10] の 8, 9 章）．特に $\kappa = 0$ の場合は拡散方程式となる．

3.2 Coordinate-wise transformation

座標ごとの変換からなるファイバーを

$$\mathcal{F}_\mu = \left\{[(1-t)\mathrm{id} + t\nabla\varphi]\sharp\mu \mid \varphi(x) = \sum_{i=1}^n \varphi_i(x_i),\ \varphi_i : \mathrm{convex}\right\}$$

と定義する．また式 (5) の汎関数 E を $\mathcal{F}_\mu$ に制限したものを $E|_{\mathcal{F}_\mu}$ と記す．

このとき Marshall & Olkin の定理の関数空間版は次のように述べられる．

定理 5 (In preparation). $\kappa(x)=(\sum_i x_i)^2/2$ とし，$\mu\in\mathcal{P}_2(\mathbb{R}^n)$ は次の正則条件 (i), (ii) を満たすとする：(i) μ の密度関数 ρ はいたるところ正である．(ii) 各ペア $i\neq j$ に対して $\iint \rho_{ij}(x_i,x_j)^2/\{\rho_i(x_i)\rho_j(x_j)\}dx_idx_j<\infty$ である[10]．ただし ρ_{ij} は ρ の 2 次元周辺密度を表し，ρ_i は 1 次元周辺密度を表す．

このとき，汎関数 $E|_{\mathcal{F}_\mu}$ の停留点 $\nu\in\mathcal{F}_\mu$ が一意に存在する．また ν は次の条件を満たす：

$$\int f(x_i)\left(\sum_{j=1}^n x_j\right)d\nu=\int f'(x_i)d\nu,\quad \forall i=1,\ldots,n,\quad \forall f\in C_b^1(\mathbb{R}). \tag{6}$$

ここで $C_b^1(\mathbb{R})$ は C^1 級かつ微分が有界な関数全体を表すものとする．

注意 3. $n=1$ の場合，式 (6) は Stein の等式として確率統計の分野でよく知られた式となり（例えば [3]），特に停留点 ν は標準正規分布となる．

注意 4. $\kappa(x)=\sum_i(x_i^2/2)$ の場合は，コピュラと本質的に同じものが停留点として得られる．「本質的に」と言っているのは，周辺分布が一様分布でなく標準正規分布になるためである．

最後に，一般の κ に対して勾配流の方程式は

$$\partial_t\rho_i=\partial_i(\partial_i\rho_i+\rho_i\mathrm{E}_\rho[\nabla\kappa(x)|x_i]),\quad i=1,\ldots,n,$$

で与えられる．ただし ρ_i は ρ の第 i 周辺密度であり，$\mathrm{E}_\rho[\cdot|x_i]$ は ρ に関する条件付き期待値を表す．この勾配流の統計学的，物理的意味は不明である[11]．

References

[1] Ambrosio, L., Gigli, N., and Savaré, G. (2005). *Gradient Flows – in Metric Spaces and in the Space of Probability Measures*, Birkhäuser.

[2] Borwein, J. M., Lewis, A. S., and Nussbaum, R. D. (1994). Entropy minimization, DAD problems, and doubly stochastic kernels, *J. Funct. Anal.*, **123**, 264–307.

[3] Chen, L. H. Y., Goldstein, L., and Shao, Q. (2011). *Normal Approximation by Stein's Method*, Springer.

[4] Marshall, A. W., Olkin, I., (1968). Scaling of matrices to achieve specified row and column sums. *Numer. Math.*, 12, 83–90.

[5] McCann, R. J. (1997). A convexity principle for interacting gases, *Adv. Math.*, **128** (1), 153–179.

[6] Nelsen, R. B. (2006). *An Introduction to Copulas*, 2nd ed., Springer.

[7] Otto, F. (2001). The geometry of dissipative evolution equations: the porous medium equation, *Comm. Partial Differential Equations*, **26**, 101–174.

[8] Sei, T. (2016). An objective general index for multivariate ordered data, *J. Multivariate Anal.*, **147**, 247–264.

[9] Takatsu, A. (2010). Wasserstein geometry of Gaussian measures, *Osaka J. Math.*, **48** (4), 1005–1026.

[10] Villani, C. (2003). *Topics in Optimal Transportation*, American Mathematical Society.

[10]条件 (ii) は，2 次元周辺コピュラ密度が 2 乗可積分という条件と同値である．

[11]ピンと来た方は是非ご一報ください．sei (@) mist.i.u-tokyo.ac.jp