情報論的学習理論テクニカルレポート 2009
Technical Report on Information-Based Induction Sciences 2009 (IBIS2009)

一次元正規分布のなす空間への曲線あてはめ

Curve fitting in the space of one-dimensional normal distribution

藤木 淳* 赤穂 昭太郎†
Jun Fujiki Shotaro Akaho

Abstract: We propose a method for extracting a one-dimensional structure from a set of parameters of one-dimensional normal distributions. In this paper, the one-dimensional structure is represented as a curve, which contains a linear parameter, on the manifold of one-dimensional normal distributions. And the fitting error is measured by metric tensor, which is the second order approximation of e-divergence and/or m-divergence. In this formulation, the estimation of curve fitting is represented by the framework of Jacobian kernel principal component analysis, which is the extension of kernel principal component.
Keywords: information geometry, manifold fitting, Euclideanization, kernel principal component analysis, Jacobian kernel

1 概要

1 次元正規分布の平均と分散の組がが複数個与えられたとき，それらは 2 次元空間をなすが，この 2 次元空間の中から構造を抽出するという問題を考える．2 次元空間の部分空間であるから，構造としては 0 次元空間である点及び 1 次元空間である曲線が考えられる．

0 次元構造は重心である．複数の 1 次元正規分布の重心として，複数の 1 次元正規分布から測った e-ダイバージェンスの和または m-ダイバージェンスの和を最小にする m-重心及び e-重心が提案されている [3]．1 次元構造は曲線であり，この構造を抽出する手法として，単純には**主成分分析（principle component analysis；PCA）**などが適用可能と考えられるが，PCA には以下のような問題点がある． PCA は平均と分散の組が従う分布が**ユークリッド（欧幾里得；欧氏）** 空間中の正規分布，つまり 1 次元正規分布の平均と分散の組が与える計量が一定であることを仮定しているが，1 次元正規分布の平均と分散の組が与える自然な計量は，情報幾何学的には一定ではない [4, 5]．また，PCA は線型構造は抽出できるが，非線型構造は抽出に不向きである．

そこで本稿では，1 次元正規分布の平均と分散をパラメータとする 2 次元空間の点列への曲線あてはめにより 1 次元構造を探すことを考える．その際，確率分布とあてはめ曲線の距離をダイバージェンスそのものではなく，ダイバージェンスを 2 次形式で近似したもので測る．このあてはめ問題は**核主成分分析（kernel PCA；KPCA）**[9] におけるあてはめ誤差を入力空間の計量で近似した**ヤコビ核主成分分析（Jacobian kernel PCA；JKPCA）** [8] の枠組みで捉えることができる．核関数であてはめ問題を表現することにより多項式核に対応する直線や 2 次曲線等の多項式曲線だけでなく，ガウス核などの様々な核関数に対応する曲線があてはめ可能となる．また，この結果はダイバージェンスを最小化することによって e-平坦空間や m-平坦空間をあてはめる手法 [3] を含めたダイバージェンスを最小化する曲線あてはめの初期値として利用することが可能である．

2 一次元正規分布のなす空間

2.1 4 つの座標系

本節では，一次元正規分布のなす多様体 $\mathcal{S}$ に 4 つの座標系を与える．最も基本的な座標系は，平均 μ，標準偏差 σ（> 0）に対して $\boldsymbol{\mu} = (\mu, \sigma)^\top$ となる座標系であり，一次元正規分布は

$$p(x; \boldsymbol{\mu}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\} \tag{1}$$

と表現できる．これは 2 次元平面の上半平面 $\sigma > 0$ が定義域である．

次の座標系は平均と分散の組 $\boldsymbol{\xi} = (\mu, V)^\top$（$V = \sigma^2$）

*産業技術総合研究所 脳神経情報研究部門, 〒 305-8568 茨城県つくば市梅園 1-1-1 中央第 2, e-mail jun-fujiki@aist.go.jp,
The National Institute of Advanced Industrial Science and Technology (AIST) Neuroscience Research Institute, Central 2, 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568
†e-mail s.akaho@aist.go.jp

で定義される座標系であり，1 次元正規分布は

$$p(x;\boldsymbol{\xi}) = \frac{1}{\sqrt{2\pi V}} \exp\left\{-\frac{(x-\mu)^2}{2V}\right\} \tag{2}$$

と表され，定義域は 2 次元平面の上半平面 $V > 0$ である．

さて，

$$p(x;\boldsymbol{\mu}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}x^2 + \frac{\mu}{\sigma^2}x - \frac{\mu^2}{2\sigma^2}\right\}$$

となるので，$F_1(x) = x^2, F_2(x) = x$ とし，

$$\theta_1 = -\frac{1}{2\sigma^2} = -\frac{1}{2V}, \quad \theta_2 = \frac{\mu}{\sigma^2} = \frac{\mu}{V}$$

とおくと，$\boldsymbol{\theta} = (\theta_1, \theta_2)^\top$ に対して

$$p(x;\boldsymbol{\theta}) = \exp\{\theta_1 F_1(x) + \theta_2 F_2(x) - \psi(\boldsymbol{\theta})\} \tag{3}$$

の形に表される．この $\boldsymbol{\theta}$ は 1 次元正規分布のなす多様体 $\mathcal{S}$ の局所座標系を定める助変数で e-座標系と呼ばれる．e-座標系の定義域は $\theta_1 < 0$ なる左半平面である．

一方，$\eta_i = \mathrm{E}_{\boldsymbol{\theta}}[F_i(\boldsymbol{x})]$ とおくと

$$\eta_1 = \mu^2 + \sigma^2 = \mu^2 + V, \quad \eta_2 = \mu$$

が成立し，$\boldsymbol{\eta} = (\eta_1, \eta_2)^\top$ は $\boldsymbol{\theta}$ に双対な座標系となり，これは m-座標系と呼ばれる．m-座標系の定義域は放物線の内部である $\eta_1 > {\eta_2}^2$ である．

今，$\boldsymbol{\theta}$ と $\boldsymbol{\eta}$ は座標変換によって全射となるので，その変換を $\boldsymbol{\theta}(\boldsymbol{\eta}), \boldsymbol{\eta}(\boldsymbol{\theta})$ と書くことにすると，e-座標と m-座標の間の座標変換は

$$\theta_1(\boldsymbol{\eta}) = -\frac{1}{2(\eta_1 - {\eta_2}^2)}, \quad \theta_2(\boldsymbol{\eta}) = \frac{\eta_2}{\eta_1 - {\eta_2}^2},$$
$$\eta_1(\boldsymbol{\theta}) = \left(\frac{\theta_2}{2\theta_1}\right)^2 - \frac{1}{2\theta_1}, \quad \eta_2(\boldsymbol{\theta}) = -\frac{\theta_2}{2\theta_1}$$

となる．他の座標変換も同様に計算できる．

2.2 測地線

さて，多様体 $\mathcal{S}$ の中で真っ直ぐな線は測地線と呼ばれるが，双対な座標系に対応して，双対な e-測地線と m-測地線が定義できる．e-測地線は e-座標系 $\boldsymbol{\theta}$ に対して線形な関係で結ばれる曲線であり，$\boldsymbol{\theta}_1$，$\boldsymbol{\theta}_2$ を通る e-測地線は $\boldsymbol{\theta}(t) = t\boldsymbol{\theta}_1 + (1-t)\boldsymbol{\theta}_2, t \in T \subset \mathbb{R}$ と助変数表示される．m-測地線は m-座標系 $\boldsymbol{\eta}$ に対して同様に定義される曲線 $\boldsymbol{\eta}(t) = t\boldsymbol{\eta}_1 + (1-t)\boldsymbol{\eta}_2$ である．本稿で考察する $\mathcal{S}$ は 2 次元多様体であるから，

$$\begin{aligned} \text{e-測地線} &: a\theta_1 + b\theta_2 + c = 0, \\ \text{m-測地線} &: p\eta_1 + q\eta_2 + r = 0 \end{aligned}$$

と表現できる．

多様体 $\mathcal{S}$ の部分空間には**平坦（flat）**という概念が定義される．もちろん平坦さも双対な構造に応じて e-平坦と m-平坦の 2 種類が定義できる．e-平坦な部分空間 $\mathcal{M}$ とは $\mathcal{M}$ 上の任意の 2 点を結ぶ e-測地線が $\mathcal{M}$ に含まれるものであり，m-平坦も m-測地線に関して同様に定義される．本稿では $\mathcal{S}$ の平坦な部分空間は測地線と点に限られる．

3 点のあてはめ

本稿の目的は複数の 1 次元正規分布に対して曲線あてはめにより 1 次元構造を抽出することであるが，その前に "重心" を求めることによって 0 次元構造を抽出する方法 [3] について述べておく．ここで複数の 1 次元正規分布の重心として，複数の 1 次元正規分布から測った e-ダイバージェンスの和を最小にする m-重心，及び複数の 1 次元正規分布から測った m-ダイバージェンスの和を最小にする e-重心の 2 通り考えることができる．

m-重心は e-ダイバージェンスの和を最小にすることから，θ 座標における算術平均となる．よって複数の 1 次元正規分布の θ 座標を $\{\boldsymbol{\theta}_{[d]}\}_{d=1}^D$ とし，m-重心の θ 座標を $\overline{\boldsymbol{\theta}}^{\mathrm{m}}$ とすると $\overline{\boldsymbol{\theta}}^{\mathrm{m}} = \frac{1}{D}\sum_{d=1}^D \boldsymbol{\theta}_{[d]}$ となる．同様に，e-重心は m-ダイバージェンスの和を最小にすることから，η 座標における算術平均となる．よって複数の 1 次元正規分布の η 座標を $\{\boldsymbol{\eta}_{[d]}\}_{d=1}^D$ とし，e-重心の η 座標を $\overline{\boldsymbol{\eta}}^{\mathrm{e}}$ とすると $\overline{\boldsymbol{\eta}}^{\mathrm{e}} = \frac{1}{D}\sum_{d=1}^D \boldsymbol{\eta}_{[d]}$ となる[1]

4 1 次元曲線のあてはめ

4.1 他の座標系における測地線の表現

e-測地線は座標 $\boldsymbol{\theta}$ における直線 $a\theta_1 + b\theta_2 + c = 0$ を表す．この直線の $\boldsymbol{\mu}$ における表現は $a - 2b\mu - 2c\sigma^2 = 0$ であるから，これは $\boldsymbol{\mu}$ 空間における放物線または直線を表わす．また，特徴写像

$$\boldsymbol{\mu} \mapsto \widetilde{\boldsymbol{\mu}}_{\boldsymbol{\theta}} = (\mu, \sigma^2, 1)^\top$$

により線型あてはめ $\boldsymbol{a}^\top \widetilde{\boldsymbol{\mu}}_{\boldsymbol{\theta}} = 0$ に帰着できる．同様に $\boldsymbol{\xi}$ 空間において e-測地線は直線を表わし，特徴写像

$$\boldsymbol{\xi} \mapsto \widetilde{\boldsymbol{\xi}}_{\boldsymbol{\theta}} = (\mu, V, 1)^\top$$

により線型あてはめ $\boldsymbol{a}^\top \widetilde{\boldsymbol{\xi}}_{\boldsymbol{\theta}} = 0$ に帰着できる．そして $\boldsymbol{\eta}$ 空間における e-測地線は放物線または直線で，特徴写像

$$\boldsymbol{\eta} \mapsto \widetilde{\boldsymbol{\eta}}_{\boldsymbol{\theta}} = (\eta_1 - \eta_2^2, \eta_2, 1)^\top$$

により線型あてはめ $\boldsymbol{a}^\top \widetilde{\boldsymbol{\eta}}_{\boldsymbol{\theta}} = 0$ に帰着できる．

[1] Akaho[3] とは e-重心と m-重心の定義が逆であるが，e-平坦な空間へはデータから m-射影を行なうことが自然であるという観点から，本稿のように定義する．

m-測地線は座標 $\boldsymbol{\eta}$ における直線 $a\eta_1 + b\eta_2 + c = 0$ を表わす．この直線の $\boldsymbol{\mu}$ 空間における表現は

$$a(\mu^2 + \sigma^2) + b\mu + c = 0$$

と，円または直線である．よって特徴写像

$$\boldsymbol{\mu} \mapsto \widetilde{\boldsymbol{\mu}}_{\boldsymbol{\eta}} = (\mu, \mu^2 + \sigma^2, 1)^\top$$

により線型あてはめ $\boldsymbol{a}^\top \widetilde{\boldsymbol{\mu}}_{\boldsymbol{\eta}} = 0$ に帰着できる．同様に $\boldsymbol{\xi}$ 空間において m-測地線は μ^2 の係数が -1 の放物線，または直線を表わし，特徴写像

$$\boldsymbol{\xi} \mapsto \widetilde{\boldsymbol{\xi}}_{\boldsymbol{\eta}} = (\mu, \mu^2 + V, 1)^\top$$

によって，特徴空間における線型あてはめ $\boldsymbol{a}^\top \widetilde{\boldsymbol{\xi}}_{\boldsymbol{\eta}} = 0$ に帰着できる．そして $\boldsymbol{\theta}$ 空間において m-測地線は二次曲線を表わし，特徴写像

$$\boldsymbol{\theta} \mapsto \widetilde{\boldsymbol{\theta}}_{\boldsymbol{\eta}} = (2\theta_1 - \theta_2^2, \theta_1\theta_2, 1)^\top$$

により線型あてはめ $\boldsymbol{a}^\top \widetilde{\boldsymbol{\theta}}_{\boldsymbol{\eta}} = 0$ に帰着できる．

4.2 あてはめる曲線族

以上，あてはめる曲線族は全て $\boldsymbol{a}^\top F(\boldsymbol{\phi}) = 0$ の形である．よって以下，一般的に $\boldsymbol{\phi}$ 空間において線型助変数で記述される曲線族 $f(\boldsymbol{\phi}; \boldsymbol{a}) = \boldsymbol{a}^\top F(\boldsymbol{\phi}) = 0$ をあてはめる手法について述べる．

4.3 入力空間の計量

$\boldsymbol{\mu}$，$\boldsymbol{\xi}$，$\boldsymbol{\theta}$，$\boldsymbol{\eta}$ 空間における計量はそれぞれ，

$$G_{\boldsymbol{\mu}} = \frac{1}{\sigma^2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad G_{\boldsymbol{\xi}} = \frac{1}{2V^2}\begin{pmatrix} 2V & 0 \\ 0 & 1 \end{pmatrix},$$

$$G_{\boldsymbol{\theta}} = -\frac{1}{2\theta_1^3}\begin{pmatrix} \theta_2^2 - \theta_1 & -\theta_1\theta_2 \\ -\theta_1\theta_2 & \theta_1^2 \end{pmatrix},$$

$$G_{\boldsymbol{\eta}} = G_{\boldsymbol{\theta}}^{-1} = \frac{1}{(\eta_1 - \eta_2^2)^2}\begin{pmatrix} 1/2 & -\eta_2 \\ -\eta_2 & \eta_1 + \eta_2^2 \end{pmatrix}$$

であることが簡単に計算できる [4]．

4.4 ダイバージェンス

点 p を部分空間 $\mathcal{M}$ へ射影することを考える．射影方法として e-射影と m-射影が考えられる．

点 p から，$\mathcal{M}$ への m-射影を q とすると，q は m-ダイバージェンス

$$D(p||q) = \psi(\boldsymbol{\theta}(p)) + \phi(\boldsymbol{\eta}(q)) - \boldsymbol{\theta}(p)^\top \boldsymbol{\eta}(q)$$

を最小にする点 q となる．同様に，点 p から，$\mathcal{M}$ への e-射影を q とすると，q は e-ダイバージェンス

$$D(q||p) = \psi(\boldsymbol{\theta}(q)) + \phi(\boldsymbol{\eta}(p)) - \boldsymbol{\theta}(q)^\top \boldsymbol{\eta}(p)$$

を最小にする点 q となる．

このように，一般には e-ダイバージェンス，m-ダイバージェンスを用いた推定結果は異なる．しかし，微小距離に関しては 3 次以上の微小量を無視すると

$$\begin{aligned} D(p||p+dq) &\approx \frac{1}{2} d\boldsymbol{\eta}^\top G_{\boldsymbol{\eta}} d\boldsymbol{\eta} \\ &= \frac{1}{2} d\boldsymbol{\theta}^\top G_{\boldsymbol{\theta}} d\boldsymbol{\theta} \approx D(p+dp||p) \end{aligned}$$

となり，e-ダイバージェンスと m-ダイバージェンスは等しくなる．本稿では，この 2 次形式をダイバージェンスの近似距離として用いることにする．

5 特徴空間における線型あてはめ

本節では，空間 $\boldsymbol{\phi}$ における計量が $G_{\boldsymbol{\phi}}$ であるときに，線型助変数で記述される曲線群

$$f(\boldsymbol{\phi}; \boldsymbol{a}) = \boldsymbol{a}^\top F(\boldsymbol{\phi}) = 0 \qquad (4)$$

をあてはめる手法について述べる．

例えば，一次元正規分布の集合に e-測地線をあてはめるには，$\boldsymbol{\mu}$ 空間において曲線族 $a\mu + b\sigma^2 + c = 0$ をあてはめれば良く，このあてはめを $\boldsymbol{\mu}$ 空間の計量に基づいて特徴空間 $\widetilde{\boldsymbol{\mu}}_{\boldsymbol{\theta}}$ において線型あてはめ $\boldsymbol{a}^\top \widetilde{\boldsymbol{\mu}}_{\boldsymbol{\theta}} = 0$ を行なえば良い．もちろん，$\boldsymbol{\xi}$ 空間の計量に基づいて特徴空間 $\widetilde{\boldsymbol{\xi}}_{\boldsymbol{\theta}}$ において線型あてはめを行っても良いし，$\boldsymbol{\eta}$ 空間の推定，$\boldsymbol{\theta}$ 空間の推定として考えても良い．

5.1 エネルギー関数

D 個の一次元正規分布が観測されたとし，それらの座標を $\{\boldsymbol{\phi}_{[d]}\}_{d=1}^{D}$ とする．そして $\widehat{\boldsymbol{\phi}}_{[d]}$ を $\boldsymbol{\phi}_{[d]}$ から式 (4) で表現される曲線へ下した垂線の足とする．また，$\widehat{\boldsymbol{\phi}}_{[d]} - \boldsymbol{\phi}_{[d]} = \delta\boldsymbol{\phi}_{[d]}$ とおく．このとき，

$$f(\widehat{\boldsymbol{\phi}}_{[d]}; \boldsymbol{a}) \approx f(\boldsymbol{\phi}; \boldsymbol{a}) + \frac{\partial f}{\partial \boldsymbol{\phi}}(\boldsymbol{\phi}_{[d]}; \boldsymbol{a})\, \delta\boldsymbol{\phi}_{[d]} = 0 \qquad (5)$$

が成立する．写像 $F: \boldsymbol{\phi} \mapsto F(\boldsymbol{\phi})$ ヤコビ行列を $J_F = \frac{\partial F}{\partial \boldsymbol{\phi}}$ とし，$\boldsymbol{\phi}_{[d]}$ における J_F の値を $J_{F_{[d]}}$ とすると，

$$\frac{\partial f}{\partial \boldsymbol{\phi}}(\boldsymbol{\phi}_{[d]}; \boldsymbol{a}) = \boldsymbol{a}^\top J_{F_{[d]}} \delta\boldsymbol{\phi}_{[d]}$$

により，式 (5) は $\boldsymbol{a}^\top J_{F_{[d]}}\, \delta\boldsymbol{\phi}_{[d]} = -\boldsymbol{a}^\top F_{[d]}$，つまり

$$\boldsymbol{a}^\top \left[J_{F_{[d]}}\, \delta\boldsymbol{\phi}_{[d]} \delta\boldsymbol{\phi}_{[d]}^\top J_{F_{[d]}}^\top \right] \boldsymbol{a} = \boldsymbol{a}^\top \left[F_{[d]}\, F_{[d]}^\top \right] \boldsymbol{a}, \qquad (6)$$

と変形できる．ここで $F_{[d]} = F(\boldsymbol{\phi}_{[d]})$ である．よって，データ点と曲線との距離 $r_{[d]}$ は $G_{[d]} = G_{\boldsymbol{\phi}_{[d]}}$ とおくと

$$\delta\boldsymbol{\phi}_{[d]}{}^\top G_{[d]}\, \delta\boldsymbol{\phi}_{[d]} = \delta F_{[d]}^\top \mathcal{G}_{[d]}\, \delta F_{[d]} \qquad (7)$$

($\mathcal{G}_{[d]} = J_{F_{[d]}}^{+\top} G_{[d]} J_{F_{[d]}}^{+}$）[2]の式 (6) の基での最小値となる．そしてそれはコーシー・シュワルツの不等式により，

$$r_{[d]}^2 = \frac{\boldsymbol{a}^\top \left[F_{[d]} F_{[d]}^\top\right] \boldsymbol{a}}{\boldsymbol{a}^\top \mathcal{G}_{[d]}^{+} \boldsymbol{a}} \tag{8}$$

と計算でき，$\boldsymbol{\phi}$ 空間の計量の基での曲線あてはめは，エネルギー関数

$$\mathcal{E}(\boldsymbol{a}) = \sum_{d=1}^{D} \frac{\boldsymbol{a}^\top \left[F_{[d]} F_{[d]}^\top\right] \boldsymbol{a}}{\boldsymbol{a}^\top \mathcal{G}_{[d]}^{+} \boldsymbol{a}} \tag{9}$$

を最小化する $\boldsymbol{a}$ を求めることによって実現できる．

5.2 解法とアルゴリズム

式 (9) のようなレイリー商の和を最小化する問題はコンピュータビジョンにおいて深く研究されており，多くの手法が提案されている（例えば [6, 10]）．本稿で与えたレイリー商の和は 1 次元正規分布のなす空間のダイバージェンスの近似に基づくので，レイリー商の和の最小化を高精度に実現する必要がないため，Akaho[1] によって提案された最小化手法を紹介する．

なお，ここで $f(\phi; \boldsymbol{a})$ は十分に滑らかであるとする，つまり $\boldsymbol{a}$ が微小変化したとき，$\partial f/\partial \boldsymbol{\phi}$ も十分小さいものとする．

今，$\boldsymbol{a}$ の近似値 $\widehat{\boldsymbol{a}}$ が得られたとし，$\lambda_{[d]}$ 及び Λ を

$$\lambda_{[d]} = \widehat{\boldsymbol{a}}^\top \mathcal{G}_{[d]}^{+} \widehat{\boldsymbol{a}}, \quad \Lambda = \mathrm{diag}\{\lambda_{[1]}, \ldots, \lambda_{[D]}\},$$

と定義し，$\mathcal{F} = (F_{[1]} \quad \cdots \quad F_{[D]})$ とおくと，エネルギー関数は

$$\mathcal{E}(\boldsymbol{a}) \approx \boldsymbol{a}^\top \left[\mathcal{F}\Lambda^{-1}\mathcal{F}^\top\right] \boldsymbol{a}$$

と近似できるので，

$$\boldsymbol{a} = \texttt{UnitMinEigenVec}\left[\mathcal{F}\Lambda^{-1}\mathcal{F}^\top\right]$$

が成立[3]する．

次のアルゴリズムにおいて，右上の添字 $[k]$ は k 番目のステップにおける値を表す．特に初期値は [0] で表す．初期値の与え方は後に述べる．

(1) 初期値 $\{\lambda_{[d]}^{[0]}\}_{d=1}^{D}$ を計算．

(2) (a) 及び (b) を収束するまで繰り返し：

- **(a)** $\widehat{\boldsymbol{a}}^{[k+1]} := \texttt{UnitMinEigenVec}[\mathcal{F}(\Lambda^{[k]})^{-1}\mathcal{F}^\top]$
- **(b)** $\lambda_{[d]}^{[k+1]} := (\widehat{\boldsymbol{a}}^{[k+1]})^\top \mathcal{G}_{[d]}^{+} (\widehat{\boldsymbol{a}}^{[k+1]})$.

[2] X^+ で X のムーア・ペンローズ逆行列を表わす．

[3] $\texttt{UnitMinEigenVec}[X]$ で X の最小固有値に対応する単位固有ベクトルを表わす．

6 計量のユークリッド化（欧氏化）

本節では提案手法の初期値としての計量の（0 階）欧氏化を述べる．Akaho[1] は通常の LS 解を初期値とした．通常の LS 解は計量の変換を考慮しない故，射影空間の計量テンソルが単位行列であるとの仮定の基で助変数を推定する．推定値は $\{\lambda_{[d]}^{[0]} = 1\}_{d=1}^{D}$ と置いたエネルギー関数の近似式 $\mathcal{E}(\boldsymbol{a}) \approx \boldsymbol{a}^\top[\mathcal{F}\mathcal{F}^\top]\boldsymbol{a}$ から求められる

$$\widehat{\boldsymbol{a}}^{[1]} = \texttt{UnitMinEigenVec}[\mathcal{F}\mathcal{F}^\top]$$

であり，これが通常の LS 解である．しかし通常の LS 解は初期値として良くないことがある．そこで超球面データに対する**計量の欧氏化（Euclideanization of metric）** が提案された [7].

欧氏化の概念は入力空間の計量をなるべく保つための射影空間における計量の調整である．その意味で先に述べたで曲線のあてはめ手法も一種の計量の欧氏化である．そこでこれらの欧氏化を区別するために，Fujiki and Akaho[7] によって提案された欧氏化を計量の 0 階欧氏化と呼び，本稿で提案した欧氏化を計量の 1 階欧氏化と呼ぶことにする．

当初, 計量の 0 階欧氏化は超球面の**等方向射影（equidirectional projection；EDP）**の場合にのみ提案された [7] が，これは線分長の変化を体積要素の拡大率を利用して調整する手法であり，その概念は，写像により n 次元体積要素が k 倍に拡大されたとすると，線分長は $k^{\frac{1}{n}}$ 倍拡大されると期待できる．そこで入力空間の最小二乗法は射影空間における重み附き最小二乗法で近似できる．このとき各データに附する重みは $k^{-\frac{2}{n}}$ となる．

さて $\boldsymbol{\phi}$ 空間をを射影した空間を $F(\boldsymbol{\phi}) \in \mathbb{R}^N$ とし，その射影のヤコビ行列を $J_F = \frac{\partial F}{\partial \boldsymbol{\phi}}$ とする．このとき写像 $\boldsymbol{\phi} \mapsto F(\boldsymbol{\phi})$ における 2 次元体積（面積）の拡大率は次のように計算される：

$$\begin{aligned}\left|\frac{\partial F}{\partial \phi_1} d\phi_1 \wedge \frac{\partial F}{\partial \phi_2} d\phi_2\right| &= \sqrt{\det\left(J_F^\top J_F\right)}\,|d\phi_1 \wedge d\phi_2| \\ &= \frac{1}{\sqrt{\mathrm{Det}\ \mathcal{G}}}\end{aligned}$$

ここで $\mathcal{G} = J_F^{+\top} G_{\boldsymbol{\phi}} J_F^{+}$ であり，$\mathrm{Det}\, X$ は非負対称行列 X の非零固有値の積[4]を表すものとする．よって 2 次元体積は $(\mathrm{Det}\ \mathcal{G})^{-1/2}$ 倍拡大されるので線分の長さは $(\mathrm{Det}\ \mathcal{G})^{-1/4}$ 倍拡大されると期待できる．よって入力空間の LS 法は射影空間において，重み $(\mathrm{Det}\ \mathcal{G})^{1/2}$ を附した重み附き LS 法となる．

[4] データが誤差を含む場合，零固有値が非零固有値となることがある．この場合は最大 2 固有値の積を計算すれば良い．何故なら，一般に $\mathcal{G}$ の階数は 2 だからである．

なお，この射影による微小距離の変化は既に述べたように

$$
\begin{aligned}
r^2 &= d\boldsymbol{\phi}^\top G_{\boldsymbol{\phi}} d\boldsymbol{\phi} \\
&= dF^\top (J_F^{+T} G_{\boldsymbol{\phi}} J_F^{+}) dF = dF^\top \mathcal{G}\, dF
\end{aligned}
$$

だから 0 階欧氏化は微小距離の変換規則から簡単に計算できる．

ここで射影空間における 3 種類の距離を比較してみる．第一は普通の最小二乗法で用いられる距離であり，

$$
r^2 = \frac{\{f(\boldsymbol{\phi};\boldsymbol{a})\}^2}{\boldsymbol{a}^\top \boldsymbol{a}} = \frac{\{f(\boldsymbol{\phi};\boldsymbol{a})\}^2}{\boldsymbol{a}^\top \mathrm{I}_N\, \boldsymbol{a}}
$$

である．第二の距離は 0 階欧氏化であり，

$$
r^2 = \frac{\{f(\boldsymbol{\phi};\boldsymbol{a})\}^2}{\boldsymbol{a}^\top \{(\mathrm{Det}\ \mathcal{G}^+)^{1/n}\, \mathrm{I}_N\} \boldsymbol{a}},
$$

である．そして最後の距離は本稿で提案する 1 階欧氏化であり，

$$
r^2 = \frac{\{f(\boldsymbol{\phi};\boldsymbol{a})\}^2}{\boldsymbol{a}^\top \mathcal{G}^+\, \boldsymbol{a}}
$$

である．この 3 種類の距離において分子は共通であり分母のみが異なることに注意する．普通の最小二乗法における射影空間の計量行列は単位行列である．0 階欧氏化における射影空間の計量行列はスカラー行列（比率は場所によって異なる）である．そして 1 階欧氏化において射影空間の計量行列は一般的な非負対称行列である．なお 0 階欧氏化は 1 階欧氏化の近似として捉えることができ，n 次元超曲面[5] $F(\boldsymbol{\phi}) \in \mathbb{R}^N$ において

$$
\mathcal{G}^+ \approx (\mathrm{Det}\ \mathcal{G}^+)^{1/n}\, \mathrm{I}_N
$$

と近似されている．

以上のことから，0 階欧氏化は，特徴空間の LS よりも提案手法である 1 階欧氏化に近い近似であると考えることができ，1 階欧氏化の初期値としてより適切であると考えられる．

7 核関数による表現

以上に述べたあてはめ手法は特徴空間がリーマン空間である場合の手法であるが，近年，核関数を利用することによって，リーマン空間の手法をヒルベルト空間に拡張することが行なわれている [9, 11].

核関数を用いた表現により，直線あてはめは 1 次多項式カーネル，2 次曲線あてはめは 2 次多項式カーネル，というように n 次曲線のあてはめを n 次多項式カーネルを用いることによって実現できる．また，**ガウス核（Gaussian kernel）**を用いることにより，ヒルベルト空間におけるあてはめも行なうことができる．このようなカーネルに基づいて定義される曲線族を**核曲線（kernel curve）**と呼ぶことにする．

なお，このあてはめ問題は標本特徴空間という空間を考えることにより，標本特徴空間という空間における線型あてはめ問題に帰着できることがわかり，カーネルトリックとは，ヒルベルト空間を標本特徴空間へ射影することによって有限次元の問題に帰着していることがわかる．

7.1 ヤコビ核主成分分析

本小節では，ヤコビ核主成分分析 [8] の枠組みであてはめ問題を核関数で表現する．

本稿では微分可能な核関数 $k(\boldsymbol{x},\boldsymbol{y}) = F(\boldsymbol{x})^\top F(\boldsymbol{y})$ 及びその微分である**ヤコビ核（Jacobian kernel）**[8]

$$
\underset{(m\times 1)}{\boldsymbol{k}(\boldsymbol{x},\boldsymbol{y})} = \frac{\partial k(\boldsymbol{x},\boldsymbol{y})}{\partial \boldsymbol{x}^\top} = J_F(\boldsymbol{x})^\top F(\boldsymbol{y})
$$

を用いる．また $\boldsymbol{a}$ の存在範囲として $F_{[d]}$ の線型結合

$$
\underset{(n\times 1)}{\boldsymbol{a}} = \sum_p \alpha_{[d]} F_{[d]} = \underset{(n\times D)}{\boldsymbol{F}}\ \underset{(D\times 1)}{\boldsymbol{\alpha}}
$$

だけを考える[6]．ここで $D \times D$ 行列 $\mathcal{K}$ を

$$
(\mathcal{K})_{ij} = k(\boldsymbol{x}_{[i]},\boldsymbol{x}_{[j]}),\quad \mathcal{K} = \begin{pmatrix} \mathcal{K}_{[1]} & \cdots & \mathcal{K}_{[D]} \end{pmatrix}
$$

で定義し，また，$D \times m$ 行列 $\boldsymbol{\mathcal{K}}_{[d]}$ を

$$
\begin{aligned}
\boldsymbol{k}_{[i][j]} &= \boldsymbol{k}(\boldsymbol{x}_{[i]},\boldsymbol{x}_{[j]}), \\
\boldsymbol{\mathcal{K}}_{[d]} &= \begin{pmatrix} \boldsymbol{k}_{[d][1]} & \cdots & \boldsymbol{k}_{[d][D]} \end{pmatrix}^\top
\end{aligned}
$$

で定義すると，

$$
\mathcal{K}_{[d]} = \boldsymbol{F}^\top F_{[d]},\quad \boldsymbol{\mathcal{K}}_{[d]} = \boldsymbol{F}^\top J_{F_{[d]}}
$$

が成立する．このとき，式 (9) について

$$
\begin{aligned}
\boldsymbol{a}^\top \left[F_{[d]} F_{[d]}^\top \right] \boldsymbol{a} &= \boldsymbol{\alpha}^\top \mathcal{K}_{[d]} \mathcal{K}_{[d]}^\top \boldsymbol{\alpha}, \\
\boldsymbol{a}^\top \mathcal{G}_{[d]}^+ \boldsymbol{a} &= \boldsymbol{\alpha}^\top \boldsymbol{\mathcal{K}}_{[d]} G_{[d]}^{-1} \boldsymbol{\mathcal{K}}_{[d]}^\top \boldsymbol{\alpha}
\end{aligned}
$$

である．このとき式 (9) は，写像 F を含まない

$$
\mathcal{E}'(\boldsymbol{\alpha}) = \sum_{d=1}^{D} \frac{\boldsymbol{\alpha}^\top \mathcal{K}_{[d]} \mathcal{K}_{[d]}^\top \boldsymbol{\alpha}}{\boldsymbol{\alpha}^\top \boldsymbol{\mathcal{K}}_{[d]} G_{[d]}^{-1} \boldsymbol{\mathcal{K}}_{[d]}^\top \boldsymbol{\alpha}} \tag{10}
$$

[5] $F(\boldsymbol{\phi})$ は S^n の像であるから，$F(\boldsymbol{\phi})$ は射影空間上の n 次元超曲面を表す．よって式 (10) の正確な表現は次のようになる：$\mathrm{rank}\,\mathcal{G}^+ = n$ であるから，非負対称行列 $\mathcal{G}^+$ の特異値分解は $U \mathrm{diag}\,\{d_1,\cdots,d_n\} U^\top$ となり，0 階欧氏化による $\mathcal{G}^+$ の近似は $(d_1 \cdots d_n)^{1/n} UU^\top$ となる．つまり $F(\boldsymbol{\phi})$ の接空間への正射影と拡大の合成変換となる．

[6] 赤穂 [2] は $\boldsymbol{a} = \boldsymbol{\Phi\alpha} + \sum_{d=1}^{D} \boldsymbol{\beta}_{[d]}^\top \left(\frac{\partial F(\boldsymbol{x})}{\partial \boldsymbol{x}} \right)_{[d]}^\top$ の範囲で $\boldsymbol{a}$ を探索し，そのため $\boldsymbol{K}(\boldsymbol{x},\boldsymbol{y}) = J_F(\boldsymbol{x})^\top J_F(\boldsymbol{y})$ も定義される．この核関数を**計量核（metric kernel）**と名付ける．

となり，これを最小にする $\boldsymbol{\alpha}$ を求めることとなる．

この最小化問題は式 (9) の最小化と同様に解ける．

今，$\boldsymbol{\alpha}$ の近似値 $\widehat{\boldsymbol{\alpha}}$ が得られたとし，$\lambda_{[d]}$ 及び Λ を

$$\begin{aligned}\lambda_{[d]} &= \widehat{\boldsymbol{\alpha}}^\top \boldsymbol{\mathcal{K}}_{[d]} G_{[d]}^{-1} \boldsymbol{\mathcal{K}}_{[d]}^\top \widehat{\boldsymbol{\alpha}}, \\ \Lambda &= \mathrm{diag}\{\lambda_{[1]}, \ldots, \lambda_{[D]}\},\end{aligned}$$

と定義すると，エネルギー関数は

$$\mathcal{E}'(\boldsymbol{\alpha}) \approx \boldsymbol{\alpha}^\top \left[\mathcal{K}\Lambda^{-1}\mathcal{K}\right] \boldsymbol{\alpha}$$

と近似できるので，

$$\boldsymbol{\alpha} = \mathtt{UnitMinEigenVec}\left[\mathcal{K}\Lambda^{-1}\mathcal{K}\right]$$

が成立する．よってアルゴリズムは以下の通り：

(1) 初期値 $\{\lambda_{[d]}^{[0]}\}_{d=1}^{D}$ を計算[7]．

(2) (a) 及び (b) を収束するまで繰り返し：

- **(a)** $\widehat{\boldsymbol{\alpha}}^{[k+1]} := \mathtt{UnitMinEigenVec}[\mathcal{K}(\Lambda^{[k]})^{-1}\mathcal{K}]$
- **(b)** $\lambda_{[d]}^{[k+1]} := (\widehat{\boldsymbol{\alpha}}^{[k+1]})^\top \boldsymbol{\mathcal{K}}_{[d]} G_{[d]}^{-1} \boldsymbol{\mathcal{K}}_{[d]}^\top (\widehat{\boldsymbol{\alpha}}^{[k+1]})$.

8 核関数と標本特徴空間

先程の $\boldsymbol{F} = (F_{[1]} \quad \cdots \quad F_{[D]})$ と特徴写像 $\boldsymbol{\phi} \mapsto F(\boldsymbol{\phi})$ を用いて**標本特徴写像（sample feature map）**を

$$\mathcal{K} : \boldsymbol{\phi} \mapsto \mathcal{K}(\boldsymbol{\phi}) = \boldsymbol{F}^\top F(\boldsymbol{\phi}) \in \mathbb{R}^D$$

によって定義し，$\mathcal{K}(\boldsymbol{\phi})$ の属する空間 $\mathbb{R}^D$ を**標本特徴空間（sample feature space）**と呼ぶことにする．このとき，標本特徴写像のヤコビ行列が $\boldsymbol{\mathcal{K}} = \boldsymbol{F}^\top J_F$ となることに注意すると，標本特徴空間での SLS 基準による超平面 $\boldsymbol{\alpha}^\top \mathcal{K}(\boldsymbol{\phi}) = 0$ のあてはめ問題は式 (10) の最小化となる．

つまり，核関数を用いたあてはめ問題の書換えは，特徴空間におけるあてはめ問題を標本特徴空間におけるあてはめ問題へと書換えたことに対応することがわかる．この書換えによってヒルベルト空間の問題を，標本数で定まる有限次元の問題に帰着していることがカーネルトリックの一つの意味であると考えることができる．

なお，標本特徴空間における欧氏化が核関数を用いた場合の欧氏化に対応することとなり，1 階欧氏化である

$$r^2 = \frac{(\boldsymbol{\alpha}^\top \mathcal{K}_{[d]})^2}{\boldsymbol{\alpha}^\top \boldsymbol{\mathcal{K}}_{[d]} G_{[d]}^{-1} \boldsymbol{\mathcal{K}}_{[d]}^\top \boldsymbol{\alpha}}$$

をスカラー行列で近似した

$$r^2 = \frac{(\boldsymbol{\alpha}^\top \mathcal{K}_{[d]})^2}{\boldsymbol{\alpha}^\top \{(\mathrm{Det}\ \boldsymbol{\mathcal{K}}_{[d]} G_{[d]}^{-1} \boldsymbol{\mathcal{K}}_{[d]}^\top)^{1/n}\ \mathrm{I}_D\} \boldsymbol{\alpha}},$$

が 0 階欧氏化となる．

これらを初期値としてレイリー商の和を最小とする $\boldsymbol{\alpha}$ を求めれば良い．

9 おわりに

本稿では，一次元正規分布の集合に線型助変数で記述される曲線群をダイバージェンスの和の近似値を最小化するようにあてはめる手法を提案した．提案手法は近似距離の最小化であるから，データの構造，すなわちあてはめるべき曲線族を決定する特徴写像 $F(\phi)$ がデータに相応しくない場合はあてはまりが悪くなるため，（非線型 PCA 一般に言えることであるが）特徴写像の選び方が重要な問題となる．

[7] この結果は核主成分分析 [9] を行なったことに相当する．

References

[1] S. Akaho, “Curve fitting that minimizes the mean square of perpendicular distances from sample points,” In Proc. of SPIE93, Vision Geometry II, Vol.2060, (1993).

[2] 赤穂昭太郎, “入力空間でのマージンを最大化するサポートベクタマシン,” 信学論 D-II, vol. J86-D-II, no. 7, pp. 934-942, 2003.

[3] S. Akaho, “The e-PCA and m-PCA: dimension reduction by information geometry,” In Proc. of Int. Joint Conf. on Neural Networks (IJCNN2004), pp. 129-134, 2004.

[4] S. Amari, *Differential Geometrical Methods in Statistics*, Springer-Verlag, 1985.

[5] S. Amari, H. Nagaoka, *Methods of Information Geometry*, AMS and Oxford university press, 2000.

[6] W. Chojnacki, M. J. Brooks, A. van den Hengel and D. Gawley, “On the fitting of surfaces to data with covariances,” IEEE Trans. Patt. Anal. Mach. Intell., vol.22, no.11, pp.1294.1303, Nov. 2000.

[7] J. Fujiki and S. Akaho, “Small hypersphere fitting by Spherical Least Square,” In Proc. of ICONIP05, pp.439-444, (2005).

[8] 藤木淳，赤穂昭太郎, 入力空間での計量に基づいた核主成分分析, 信学技報, vol. 108, no. 327, pp. 69-74, Nov. 2008.

[9] B. Schölkopf, A. Smola and K.-R. Müller, “Nonlinear component analysis as a kernel eigenvalue problem,” Neural Computation, vol. 10, pp. 1299-1319, 1998.

[10] G. Taubin, “Estimation of planar curves, surfaces and, non-planar space curves defined by implicit equations with applications to edge and rage image segmentation,” IEEE Trans. Patt. Anal. Mach. Intell., vol.13, no.11, pp.1115.1138, 1991.

[11] 津田宏治, “ヒルベルト空間における部分空間法,” 電子情報通信学会論文誌 D-II, vol.J82-D-II, no.4, pp.592-599, Apr. 1999.