Vol.2010-CVIM-171 No.7
2010/3/18

頑健なヤコビ核主成分分析を用いた部分空間あてはめ

藤 木 淳[†] 赤 穂 昭太郎[†]

元来の核主成分分析を用いた部分空間あてはめ，すなわち特徴空間の距離に関する最小二乗法に基づいた部分空間あてはめは入力空間の距離の大小を反映していないため，特徴空間においては良いあてはめであっても，入力空間においては良いあてはめとは限らなかった．そこで筆者らは入力空間の計量に基づいた核主成分分析であるヤコビ核主成分分析を提案し，それを用いて部分空間をあてはめることによって入力空間における良いあてはめを実現した．しかし，入力空間の計量を考慮したため，入力空間における誤差に対する鋭敏性が元来の核主成分分析をもちいた部分空間あてはめに比べて高くなってしまった．そこで本稿では，ヤコビ核主成分分析を用いた部分空間あてはめ手法を頑健化する．具体的には，RANSAC とカイ二乗検定に基づいた例外値除去手法，対数双曲線余弦関数などを用いた M 推定，データ点と部分空間とのユークリッド距離の絶対値の和を最小化する最小絶対値法の枠組みを提案する．

Subspace fitting via robust Jacobian kernel PCA

Jun Fujiki[†] and Shotaro Akaho[†]

The subspace fitting method based on the original kernel principle component analysis (PCA), which minimizes the square distance in feature space, sometimes derives bad estimation because it does not reflect the metric on input space. Then authors proposed the subspace fitting method based on the kernel PCA based on the metric on input space, which is called Jacobian kernel PCA. However, Jacobian kernel PCA has the sensitivity against noise more than original kernel PCA because it directly reflects the metric on input space. Then in this paper, the authors propose robust subspace fitting methods based on Jacobian kernel PCA. such as the outlier detection method based on random sampling consensus (RANSAC) and χ^2 test, the robust method based on M-estimation and the robust method based on least absolute value estimation.

1. はじめに

観測データの構造の把握はデータ解析における基本的かつ重要な手順である．データの主構造の最も基本的な決定手法は**主成分分析**（**principal component analysis；PCA**）ある．PCA は，データを**ユークリッド**（**欧幾里得；欧氏**）空間の点とみなし，データの真値は欧氏空間内のアフィン空間上に分布するという仮定の下で有効な手法である．そして PCA ではデータと主構造とのずれである誤差を欧氏距離で測定し，誤差の二乗和を最小とする**最小二乗**（**least squares；LS**）基準を用いるが，これは誤差が等方正規分布に従うという仮定の基で最尤推定を与える．

しかしデータの真値や主構造がアフィン空間上に分布しない場合には PCA は不向きである．そこで**非線型主成分分析**（**nonlinear PCA；NLPCA**）が提案されてきた．NLPCA の考え方は，非線型構造をもつデータをアフィン空間上に分布させるために**特徴空間**（**feature space**）と呼ばれる高次元空間に**特徴写像**（**feature map**）と呼ばれる写像を行ない，その後に通常の PCA を行なう所にある．このとき，データの構造は特徴写像によってアフィン空間に写像される空間に限定されるため，データ構造の理解と特徴写像はほぼ等価である．よって特徴写像が理論的に既知の場合は問題がないが，未知の場合には，写像が先か構造が先かという鶏と卵の問題が生じる．しかしデータの構造を十分に近似できる写像が求まれば良いならば，多数の基底で表現される非常に高次元への写像を考えることにより鶏と卵の問題は近似的に解決される．

一般に特徴空間が高次元の場合，特徴空間における高次元ベクトルの計算量が多くなる．この問題を解決したのがカーネルトリック[1),24)] である．カーネルトリックとは，NLPCA に特徴写像の表現は不要で，特

† 産業技術総合研究所
National Institute of Advanced Industrial Science and Technology

徴空間での誤差を測るための計量（内積）を与える対称な正定値カーネル関数があれば十分であるという議論のすりかえのことである．この議論のすりかえにより，計算量が特徴写像の次元数ではなくデータ数に依存することとなり非常に高い次元への特徴写像を考えても計算量が破綻しにくくなる．そしてまたガウシアンカーネルのような無限次元空間への写像を扱うこともできる．このカーネルトリックを用いたNLPCAを**核主成分分析**（**kernel PCA；KPCA**）[4),21)] と呼ぶ．

ここでNLPCAにおける特徴写像の決定ややKPCAにおける核関数の決定は入力空間におけるデータ点にあてはめるべき曲線群や曲面群（を表現する特徴写像）を決定することと同じであり，NLPCAやKPCAによってあてはめられた曲線や曲面は，与えらえたデータの**主成分曲線**（**principle curve**）や**主成分曲面**（**principle surface**）である．

さて，KPCA[21)] は特徴空間における欧氏距離のLS基準によるあてはめ問題である．ゆえにあてはめ結果による入力データの推定値は特徴空間におけるあてはめ結果との最短距離であり，入力空間におけるあてはめ結果への最短距離とは限らない．そこでコンピュータビジョンでは入力空間における距離のLS基準によるあてはめ問題が研究されてきた．入力空間が欧氏空間の場合[2),5),16),18),20),23)] や球面の場合[10)] が提案され，そして一般的なリーマン空間の場合を核関数を用いて表現した**ヤコビ核主成分分析**（**Jacobian KPCA；JKPCA**）[11)] が提案された．

JKPCAは特徴写像を各観測データ点附近でJacobi行列を利用してアフィン写像に近似している．このアフィン写像により観測データ附近における入力空間と特徴空間の計量の対応を近似し，入力空間の誤差を特徴空間の誤差によって近似的に表現している．このことにより，入力空間のLS推定は特徴空間の重み付きLS推定により近似される．この近似は**等計量線**（**equi-metric curve**），つまり入力空間においてデータから等距離にある閉曲線の特徴写像による像を超楕円体によって近似していることに相当する．しかし一般に特徴写像は非線型写像であるから，この近似は誤差が十分に小さい範囲内でしか成立せず，**例外値**（**outlier**），つまり誤差の大きいデータに対しては近似誤差と実際の誤差の乖離が大きく，例外値を含む場合の推定結果は一般的に悪くなる．またLS推定自体モデルからデータの方向に過剰に適合し例外値が推定に悪影響を及ぼすことが知られている．

そこで最小二乗推定における例外値の影響を抑えるためにM推定量が提案された[15)]．M推定の基本的な考え方は，誤差分布が正規分布から少々ずれたとしても有効性がそれほど落ちないように例外値の影響を制御した推定量を作る所にある．例えば平均の推定において誤差がラプラス分布に従うと仮定すると，最尤推定量は標本平均ではなく標本中央値となるというように，直感的には，誤差分布がM推定量から導かれる“正規分布関数と似た”確率分布（M推定量によっては正規化できず確率分布と見做せない場合もある）に従うと仮定した場合の最尤推定と捉えられる．

本稿で提案する**頑健なヤコビ核主成分分析**（**Robust JKPCA；R-JKPCA**）はM推定量を基本としている．1つ目は**ランザック**（**random sample consensus；RANSAC**）[7)] と χ^2 検定の組合せによって例外値を除去し，許容値のみで推定を行なう[22)] 手法である．この推定に対応する確率分布は存在しないが，これも一種のM推定とみなせる．2つ目は評価関数を誤差の2乗和から誤差の対数双曲線余弦関数 $\log\cosh$ 和に拡張したものである．典型的なM推定の評価関数として，Huberの関数等数多く提案されているが，本稿では代表として対数双曲線余弦関数を用いる．これら2つをJKPCAと組み合わせることにより，頑健な曲線あてはめを実現する．

また，M推定の評価関数を誤差の絶対値和に拡張した**最小絶対値法**（**least absolute value estimation**）についても考察する．

2. ヤコビ核主成分分析

入力空間である m 次元空間 $\mathcal{R}$ の点列への非線型部分空間をあてはめる際，この点列を特徴空間である n 次元ヒルベルト空間 $\mathcal{H}$ に写像し，$\mathcal{H}$ における線型あてはめに帰着することが多い．このあてはめを $\mathcal{H}$ におけるLS基準で推定するのがKPCAである．しかし入力空間に自然な計量が定義されている場合に特徴空間におけるLS基準での推定は好ましくなく[3)]，入力空間の計量における近似LS基準で部分空間をあてはめるJKPCA[11)] が提案された．

2.1 入力空間の計量と特徴空間の計量

入力空間を m 次元リーマン空間 $\mathcal{R}$ とし，$\mathcal{R}$ 上の点 $\boldsymbol{x}$ におけるリーマン計量を $G_{\boldsymbol{x}}$ とする．また，観測データ点を $\{\boldsymbol{x}_{[d]}\}_{d=1}^{D}$ とする．JKPCAでは観測データ点 $\boldsymbol{x}_{[d]}$ 附近の空間を計量が $G_{\boldsymbol{x}_{[d]}}$ で一定であるアフィン空間で近似する．つまり $\boldsymbol{x}=\boldsymbol{x}_{[d]}+\delta\boldsymbol{x}$ なる点 $\boldsymbol{x}$ と観測データ点 $\boldsymbol{x}_{[d]}$ の距離 r^p を $(r^p)^2=(\delta\boldsymbol{x})^\top G_{\boldsymbol{x}_{[d]}}(\delta\boldsymbol{x})$ で近似する．点 $\boldsymbol{x}$ を n 次元ヒルベルト空間である特徴空間 $\mathcal{H}$ に特徴写像 $\boldsymbol{\phi}\colon\boldsymbol{x}\mapsto\boldsymbol{\phi}(\boldsymbol{x})$ で射影する．ここで特徴写像 $\boldsymbol{\phi}$ のヤコビ行列 $J_{\boldsymbol{\phi}}$ を用いると，特徴写

像は観測データ附近で $\boldsymbol{\phi}(\boldsymbol{x}) \approx \boldsymbol{\phi}_{[d]} + J_{\boldsymbol{\phi}_{[d]}}(\boldsymbol{x} - \boldsymbol{x}_{[d]})$（$\boldsymbol{\phi}_{[d]} = \boldsymbol{\phi}(\boldsymbol{x}_{[d]})$）と近似でき，

$$(r^p)^2 = (\delta\boldsymbol{\phi}(\boldsymbol{x}))^\top \mathcal{G}_{\boldsymbol{\phi}_{[d]}}\, \delta\boldsymbol{\phi}(\boldsymbol{x})$$

$$\left[\begin{array}{l} \boldsymbol{\phi}(\boldsymbol{x} + \delta\boldsymbol{x}) = \boldsymbol{\phi}(\boldsymbol{x}) + \delta\boldsymbol{\phi} \\ \mathcal{G}_{\boldsymbol{\phi}_{[d]}} = (J^+_{\boldsymbol{\phi}_{[d]}})^\top G_{\boldsymbol{x}_{[d]}} J^+_{\boldsymbol{\phi}_{[d]}} \ , \\ \mathcal{G}^{-1}_{\boldsymbol{\phi}_{[d]}} = J_{\boldsymbol{\phi}_{[d]}} G^{-1}_{\boldsymbol{x}_{[d]}} J^\top_{\boldsymbol{\phi}_{[d]}} \end{array}\right]$$

（X^+ は X のムーア・ペンローズ逆行列）と近似できる．

2.2 特徴空間における線型あてはめ

特徴空間のデータ $\boldsymbol{\phi}_{[d]} = \boldsymbol{\phi}(\boldsymbol{x}_{[d]})$ に対して $n-1$ 次元アフィン空間 $\boldsymbol{a}^\top\boldsymbol{\phi} + b = 0$ をあてはめるには，

$$\mathcal{E}(\boldsymbol{a}) = \sum_{d=1}^{D} \frac{(\boldsymbol{a}^\top\boldsymbol{\phi}_{[d]} + b)^2}{\boldsymbol{a}^\top \mathcal{G}^{-1}_{\boldsymbol{\phi}_{[d]}} \boldsymbol{a}} \tag{1}$$

を最小にする $\boldsymbol{a}$，b を求めれば良い[11]．ここでは

$$\widetilde{\boldsymbol{\phi}} = \begin{pmatrix} \boldsymbol{\phi} \\ 1 \end{pmatrix}, \quad \widetilde{\boldsymbol{a}} = \begin{pmatrix} \boldsymbol{a} \\ b \end{pmatrix}, \quad \widetilde{\mathcal{G}}_{\boldsymbol{\phi}_{[d]}} = \begin{pmatrix} \mathcal{G}_{\boldsymbol{\phi}_{[d]}} & \boldsymbol{0} \\ \boldsymbol{0}^\top & 0 \end{pmatrix}$$

と置いて $n+1$ 次元空間の原点を通る n 次元線型部分空間 $\widetilde{\boldsymbol{a}}^\top\widetilde{\boldsymbol{\phi}} = 0$ の推定問題に帰着させる，つまり

$$\mathcal{E}(\widetilde{\boldsymbol{a}}) = \sum_{d=1}^{D} \frac{\widetilde{\boldsymbol{a}}^\top \left[\widetilde{\boldsymbol{\phi}}_{[d]}\widetilde{\boldsymbol{\phi}}^\top_{[d]}\right] \widetilde{\boldsymbol{a}}}{\widetilde{\boldsymbol{a}}^\top \widetilde{\mathcal{G}}^+_{\boldsymbol{\phi}_{[d]}} \widetilde{\boldsymbol{a}}} \tag{2}$$

の最小にする $\widetilde{\boldsymbol{a}}$ を求めれば良く[2),16),23)]，式 (2) の最小化を核関数を用いて行なう．

2.3 核関数による表現

本稿では，核関数として微分可能なものを考え，核関数 $k(\boldsymbol{x}, \boldsymbol{y}) = \widetilde{\boldsymbol{\phi}}(\boldsymbol{x})^\top\widetilde{\boldsymbol{\phi}}(\boldsymbol{y})$ だけでなく，その微分

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

も用いる．これを**ヤコビ核**（**Jacobian kernel**）と呼ぶ．また $\boldsymbol{a}$ の存在範囲として $\widetilde{\boldsymbol{\phi}}_{[d]}$ の線型結合

$$\underset{(n\times 1)}{\boldsymbol{a}} = \sum_{p} \alpha_{[d]}\widetilde{\boldsymbol{\phi}}_{[d]} = \underset{(n\times D)}{\widetilde{\boldsymbol{\Phi}}}\ \underset{(D\times 1)}{\boldsymbol{\alpha}}$$

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

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

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

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

で定義すると，

$$\mathcal{K}_{[d]} = \widetilde{\boldsymbol{\Phi}}^\top \widetilde{\boldsymbol{\phi}}_{[d]}, \quad \boldsymbol{\mathcal{K}}_{[d]} = \widetilde{\boldsymbol{\Phi}}^\top J_{\widetilde{\boldsymbol{\phi}}_{[d]}}$$

が成立し，このとき式 (2) は，写像 $\boldsymbol{\phi}$ を含まない

$$\mathcal{E}'(\boldsymbol{\alpha}) = \sum_{d=1}^{D} \frac{\boldsymbol{\alpha}^\top \mathcal{K}_{[d]} \mathcal{K}^\top_{[d]} \boldsymbol{\alpha}}{\boldsymbol{\alpha}^\top \boldsymbol{\mathcal{K}}_{[d]} \widetilde{G}^+_{\boldsymbol{x}_{[d]}} \boldsymbol{\mathcal{K}}^\top_{[d]} \boldsymbol{\alpha}} \tag{3}$$

となり，これを最小にする $\boldsymbol{\alpha}$ を求めれば良い．ここで

$$\widetilde{G}_{\boldsymbol{x}_{[d]}} = \begin{pmatrix} G_{\boldsymbol{x}_{[d]}} & \boldsymbol{0} \\ \boldsymbol{0}^\top & 0 \end{pmatrix}$$

である．

2.4 解法とアルゴリズム

式 (3) の最小化アルゴリズム[2)] を紹介する．$\boldsymbol{\alpha}$ の近似値 $\widehat{\boldsymbol{\alpha}}$ が得られたとき，

$$\mu_{[d]} = \widehat{\boldsymbol{\alpha}}^\top \boldsymbol{\mathcal{K}}_{[d]} \widetilde{G}^+_{\boldsymbol{x}_{[d]}} \boldsymbol{\mathcal{K}}^\top_{[d]} \widehat{\boldsymbol{\alpha}}\ ,$$
$$\Lambda = \mathtt{diag}\left\{\mu_{[1]}, \ldots, \mu_{[D]}\right\}$$

とすると $\mathcal{E}'(\boldsymbol{\alpha}) \approx \boldsymbol{\alpha}^\top \mathcal{K}\Lambda^{-1}\mathcal{K}\boldsymbol{\alpha}$ と近似できるので，

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

（$\mathtt{MinEigenVec}[X]$ は正方行列 X の最小固有値に対応する単位固有ベクトル）となる．

以下のアルゴリズムにおいて右肩の $[k]$ によって k ステップ目の値を表すものとする．初期値は右肩が $[0]$ となる（初期値の設定は後述）．

(1) 初期値 $\mu^{[0]}_{[d]}$（$d = 1, \ldots, D$）の設定

(2) 収束するまで (a)，(b) を繰り返す

- **(a)** $\widehat{\boldsymbol{\alpha}}^{[k+1]} = \mathtt{MinEigenVec}[\mathcal{K}(\Lambda^{[k]})^{-1}\mathcal{K}]$
- **(b)** $\mu^{[k+1]}_{[d]} = (\widehat{\boldsymbol{\alpha}}^{[k+1]})^\top \boldsymbol{\mathcal{K}}_{[d]} G^+_{\boldsymbol{x}_{[d]}} \boldsymbol{\mathcal{K}}^\top_{[d]} (\widehat{\boldsymbol{\alpha}}^{[k+1]})$ で $\{\mu_{[d]}\}^D_{d=1}$ を更新．

2.5 初　期　値

上記アルゴリズムの初期値の設定方法として以下のような方法がある．

2.5.1 核主成分分析 (KPCA)

$\{\mu^{[0]}_{[d]} = 1\}^D_{d=1}$ の場合[2)] で

$$\mathcal{E}'(\boldsymbol{\alpha}) \approx \boldsymbol{\alpha}^\top \mathcal{K}^2 \boldsymbol{\alpha}$$

を最小化する $\widehat{\boldsymbol{\alpha}} = \mathtt{MinEigenVec}[\mathcal{K}^2]$ を初期値とする．

2.5.2 欧氏化 (Euclideanization)

欧氏化[8),11)] とは特徴写像によるデータ点附近の線分長の平均拡大率の逆数をデータ点に重みとして与える手法であり，

$$\mathcal{E}'(\boldsymbol{\alpha}) \approx \boldsymbol{\alpha}^\top \left(\mathcal{K} D^{\frac{1}{m}} \mathcal{K}\right) \boldsymbol{\alpha},$$
$$D = \mathrm{diag}\left\{\mathrm{Det}\,\mathcal{G}_{\boldsymbol{\phi}_{[1]}}, \ldots, \mathrm{Det}\,\mathcal{G}_{\boldsymbol{\phi}_{[D]}}\right\}$$
$$\mathrm{Det}\,\mathcal{G}_{\boldsymbol{\phi}_{[d]}} = \det\left\{J^+_{\boldsymbol{\phi}_{[d]}} (J^+_{\boldsymbol{\phi}_{[d]}})^\top\right\} \cdot \det\, G_{\boldsymbol{x}_{[d]}}$$

（$\mathrm{Det}\, X$ で X の 0 でない固有値の積を表す）を最小化する $\widehat{\boldsymbol{\alpha}} = \mathtt{MinEigenVec}[\mathcal{K} D^{\frac{1}{m}} \mathcal{K}]$ を初期値とする．

2.5.3 Taubin 法[23)]

各データ点で異なる計量を核空間で平均した $\mathrm{E}[G^{\mathrm{kernel}}_{[d]}] = \sum_{d=1}^{D} \boldsymbol{\mathcal{K}}_{[d]} G^{+}_{\boldsymbol{x}_{[d]}} \boldsymbol{\mathcal{K}}^{\top}_{[d]}$ で近似する手法であり，

$$\mathcal{E}'(\boldsymbol{\alpha}) \approx \frac{\boldsymbol{\alpha}^{\top}\mathcal{K}^2\boldsymbol{\alpha}}{\boldsymbol{\alpha}^{\top}\left(\mathrm{E}[G^{\mathrm{kernel}}_{[d]}]\right)\boldsymbol{\alpha}}$$

を最小化する $\widehat{\alpha}$，つまり一般化固有値問題

$$\mathcal{K}^2\boldsymbol{\alpha} = \lambda\left(\mathrm{E}[G^{\mathrm{kernel}}_{[d]}]\right)\boldsymbol{\alpha}$$

の最小固有値に対応する固有ベクトルを初期値とする．

2.6 ヤコビ核主成分分析

KPCA は $\mathcal{K}^2 = \mathcal{K}\mathcal{K}$ の固有ベクトルを固有値の大きい順に並べて新しい座標枠を作る手法であり，JKPCA は $\mathcal{K}(\Lambda^{[\infty]})^{-1}\mathcal{K}$ の固有ベクトルを固有値の大きい順に並べて新しい座標枠を作る手法である．よって特徴ベクトル $\widetilde{\boldsymbol{\phi}}_{[d]}$ に対して重み

$$(\mu^{[\infty]}_{[d]})^{-\frac{1}{2}} = ||\boldsymbol{\mathcal{K}}^{\top}_{[d]}\boldsymbol{\alpha}||^{-1}_{G^{+}_{\boldsymbol{x}_{[d]}}}$$

を与えたときの重みつき PCA である．この重みはデータ点 $\boldsymbol{x}_{[d]}$ 附近の計量と $\widetilde{\boldsymbol{\phi}}_{[d]}$ 附近の計量を比べ，$\widetilde{\boldsymbol{\phi}}_{[d]}$ 附近に局所的に入力空間の計量を反映させたものである．

3. ヤコビ核主成分分析の効果

本節では，ヤコビ核主成分分析の有用性を示すために人工データを用いた実験を行なった．

3.1 データの諸元

以下のデータ生成手法は本節の実験と第 7 節の実験で共通である．

放物線 $y = x^2$ 上の点を x 座標が区間 $[-3, 3]$ からの一様分布に従うように点を生成し，各点に平均 0，分散 0.18 のラプラス分布☆を曲線の法線方向に添加した人工データに 2 次の多項式核を用いて 2 次曲線をあてはめるシミュレーションを行なった．

生成した点のうち，曲線との最短距離が 0.3 未満のものを許容値，曲線との最短距離が 0.3 以上のものを例外値と決め，例外値の割合を汚染率と定義した．第 7 節で示した統計はこのように選んだデータセットを 50 組用いて作成した．

3.2 入力空間の計量を考える効果

KPCA（図では LS）と JKPCA（図では JK）の違いを明確にするため，汚染率 0% で実験を行なった．図 1 により，KPCA によるあてはめ結果は JKPCA によるあてはめ結果に比べて原点付近のあてはまりが悪いことがわかる． この理由は，多項式核に対応す

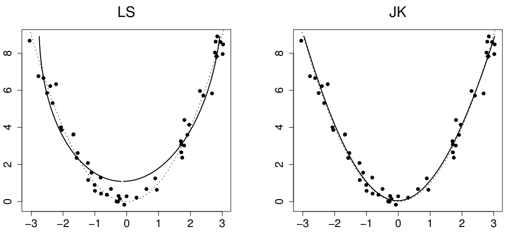


図 1 あてはめ結果：KPCA（左），JKPCA（右）

る特徴写像に存在する xy の項は，例えば $y \approx 0$ なら x のずれが非常に大きくても xy のずれは非常に小さくなるため，特徴空間における欧氏距離が小さくなってしまうからと考えられる．このことから，JKPCA のような入力空間の計量に基づくあてはめ手法が有効であると考えることができる．

4. 頑健なヤコビ核主成分分析

前節で述べたように，ヤコビ核主成分分析は入力空間の計量に基づくあてはめ手法のため，通常の核主成分分析を用いたあてはめに比べて例外値に過適合するという問題点がある．そこで本節では，ヤコビ核主成分分析の頑健化を試みる．

4.1 入力空間の近似誤差

JKPCA は入力空間における誤差 $r_{[d]}$ を

$$r^2_{[d]} \approx \frac{\boldsymbol{\alpha}^{\top}\mathcal{K}_{[d]}\mathcal{K}^{\top}_{[d]}\boldsymbol{\alpha}}{\boldsymbol{\alpha}^{\top}\boldsymbol{\mathcal{K}}_{[d]}\widetilde{G}^{+}_{\boldsymbol{x}_{[d]}}\boldsymbol{\mathcal{K}}^{\top}_{[d]}\boldsymbol{\alpha}} \tag{4}$$

（$\boldsymbol{\alpha}$ は JKPCA の収束時の値）と近似しており，この近似誤差

$$R_{[d]} = \sqrt{\frac{\boldsymbol{\alpha}^{\top}\mathcal{K}_{[d]}\mathcal{K}^{\top}_{[d]}\boldsymbol{\alpha}}{\boldsymbol{\alpha}^{\top}\boldsymbol{\mathcal{K}}_{[d]}\widetilde{G}^{+}_{\boldsymbol{x}_{[d]}}\boldsymbol{\mathcal{K}}^{\top}_{[d]}\boldsymbol{\alpha}}} \tag{5}$$

を判断基準として JKPCA を頑健化する．

4.2 RANSAC

RANSAC[7)] は，ランダムサンプリングに基づくロバストなモデル作成手法であり，

(1) 部分空間を形成する最小のデータ数を用いて部分空間をあてはめる．

(2) あてはめ結果に対して，全データの各々が許容値かどうかを検定し，許容値の個数を数える．

(3) ランダムランプリングを一定回数繰り返し，許容値の個数が最大となるものを選ぶ．

(4) 選ばれた個数が最大となる部分空間に基づき各々

☆ 平均 μ，分散 $2\lambda^{-2}$ のラプラス分布は $\frac{\lambda}{2}\exp(-\lambda|x-\mu|)$．

のデータを検定し，例外値を除去．

(5) 許容値に対して部分空間をあてはめ．

という手続きにより部分空間をあてはめる手法である．

データ量が十分にあるとし，汚染率（例外値の割合）が α，部分空間を作成するのに必要なデータ数を m 個であるとする．このとき m 個のデータが全て許容値である確率が β 以上となるためのランダムサンプリングの回数 k は

$$k \geq \frac{\log_{10}(1-\beta)}{\log_{10}\{1-(1-\alpha)^m\}} \ . \tag{6}$$

をみたす．

4.3 MAD 推定量

中央絶対偏差（**median absolute deviation；MAD**）とは，1 次元実データ $\{x_n\}_{n=1}^N$ に対して

$$\mathtt{mad}\,[x_n] = \mathtt{med}\left[\,|x_n - \mathtt{med}\,(x_n)|\,\right]$$

によって定義される確率分布の尺度助変数であり，$x_n \sim \mathtt{N}(0, \sigma^2)$ のとき，漸近的に

$$\sigma \simeq \frac{1}{\sqrt{\chi^2_{0.5}(1)}}\mathtt{mad}\,[x_n] \approx 1.4826\,\mathtt{med}\left[\,|x_n|\,\right]$$

となる[19]．ここで $\chi^2_\alpha(L)$ は自由度 L の χ^2 分布の $100\alpha\%$ 点である．

これと同様に，L 次元データ $\{\boldsymbol{x}_n\}_{n=1}^N$ が等方正規分布 $\boldsymbol{x}_n \sim \mathtt{N}(\mathbf{0}_L, \sigma^2 \mathrm{I}_L)$ に従うと仮定すると，漸近的に

$$\sigma \simeq \frac{1}{\sqrt{\chi^2_{0.5}(L)}}\mathtt{med}\left[||\boldsymbol{x}_n||\right] \tag{7}$$

が成立する．

本稿では m 次元入力空間を n 次元特徴空間に射影した後に n 次元特徴空間における k 次元あてはめ問題を考える．このときモデルからのずれである誤差ベクトルは $L = n - k$ 次元空間のベクトルとなる．しかし，入力空間の像は特徴空間内の m 次元以下の図形へと射影されるため，誤差ベクトルの分布には何らかの偏りが生ずる．しかし問題の単純化のため，誤差ベクトルの分布が等方正規分布 $\mathtt{N}(\mathbf{0}_L, \sigma^2 \mathrm{I}_L)$ に従うと仮定して，mad 推定量と σ の関係を利用し，誤差ベクトルのノルムの分散を推定する．

4.4 RANSAC と χ^2 検定

RANSAC と χ^2 検定による例外値除去手法[22] を JKPCA に以下のように適用する．本稿では，検定に用いる誤差分散の推定量 $\widehat{\sigma}^2$ の定め方も与える．

4.4.1 閾値となる誤差分散の決定

(1) RANSAC で部分空間のあてはめて残差の中央値を計算，という試行を式 (6) で定まる回数行なう．

(2) 全ての試行のうち中央値が最小となるものに対して式 (7) から誤差分散の推定量 $\widehat{\sigma}$ を求める．

4.4.2 Robust JKPCA（R-JKPCA）

(1) n 次元特徴ベクトルの集合 $\{\boldsymbol{\Phi}(\boldsymbol{x}_{[d]})\}_{d=1}^D$ からランダムに $\{\boldsymbol{\Phi}(\boldsymbol{y}_{[m]})\}_{m=1}$ を選び JKPCA により $\widetilde{\boldsymbol{\alpha}}$ を求め，全データに対して式 (5) の近似残差 $R_{[d]}$ を計算し，$R^2_{[d]} < \widehat{\sigma}^2$ なる個数を S とする．

(2) **(1)** を反復して S を最大とする主成分部分空間を選び，近似残差を有意水準 $100(1-\gamma)\%$ で検定，つまり $R^2_{[d]} \geq \chi^2_\gamma(L) \cdot \widehat{\sigma}^2$ なるデータを例外値として除去．

(3) 許容値から JKPCA で主成分部分空間を推定．

5. M 推 定

M 推定は最も良く利用されるロバスト推定法の一つであり，二乗誤差の最小化 $\min \sum_{d=1}^D \frac{1}{2} R^2_{[d]}$ の代わりに，それを変形した評価基準を用いる．何故なら，二乗誤差は例外値の影響を大きく受けるという欠点があるため，その例外値からの寄与を制御したいからである．そこで，$\frac{1}{2}x^2$ の代わりに，$x = 0$ で唯一の最小値をもち $x > 0$ で単調増加であるような正定値の偶関数 $\rho(x)$ を誤差関数として用い，$\min \sum_{d=1}^D \rho(R_{[d]})$ を最小化させる．ここで二乗誤差は $\rho(x) = \frac{1}{2}x^2$ の場合であり，二乗誤差最小化は M 推定の一種であると言うことができる．

さて，評価関数 ρ において，各データの影響力は ρ の微分である**影響関数**（**influence function**）$\Psi(x) = \frac{\partial \rho(x)}{\partial x}$ によって与えられる．二乗誤差に対する ρ 及び Ψ は，図 2 の (a)，(b) のようになり，データの影響力はモデルからのずれに正比例して大きくなることがわかる．

これと，M 推定で代表的に用いられる評価関数の 1 つである対数双曲線余弦関数 $\log\cosh$ を比較する．このとき，二乗誤差の場合を 1 としたときの各データの重みは $w(x) = \frac{\Psi(x)}{x}$ によって表現でき，これを**重み関数**（**weight function**）と呼ぶ☆．

本稿では，Ψ が双曲線正接関数 $\tanh$ となる対数双曲線余弦関数として，そのマクローリン展開が $\frac{1}{2}x^2 + o(x^2)$ となるように定数倍した

$$\begin{aligned}\rho(x) &= \sigma^2 \log\cosh\frac{x}{\sigma} \\ &= \frac{1}{2}x^2 + \frac{1}{12\sigma^2}x^4 + o(x^4) \quad (\sigma \neq 0)\end{aligned}$$

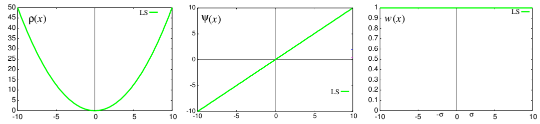
を用いることにする．このとき，

$$\Psi(x) = \sigma \tanh\frac{x}{\sigma}, \quad w(x) = \frac{\sigma}{x}\tanh\frac{x}{\sigma}$$

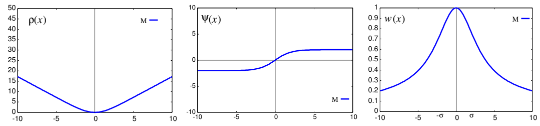
である．

☆ 本稿では，最小二乗推定の誤差関数を $\frac{1}{2}x^2$ としたので，重み関数は $w(x) = \frac{\Psi(x)}{x}$ と定義されるが，最小二乗推定の誤差関数を x^2 とすると重み関数は $w(x) = \frac{\Psi(x)}{2x}$ と定義される．

図 2 の (d)，(e)，(f) に対数双曲線余弦関数の ρ，Ψ 及び w を示す．データがモデルからあまりずれていない場合，データの影響力はモデルからのずれに正比例して大きくなるが，データがある程度モデルから離れるとその影響力は，双曲線正接関数の場合はほぼ一定の値と影響力が抑えられていることがわかる．



(a) $\rho(x) = x^2/2$　(b) $\Psi(x) = x$　(c) $w(x) = 1$



(d) $\rho(x) = \sigma^2 \log\cos\dfrac{x}{\sigma}$　(e) $\Psi(x) = \sigma\tanh\left(\dfrac{x}{\sigma}\right)$　(f) $w(x) = \dfrac{\sigma}{x}\tanh\left(\dfrac{x}{\sigma}\right)$

図 2　評価関数

6. 最小絶対値法

最小絶対値法（least absolute value estimation）とは，誤差の絶対値の和を最小化する推定量のことであり，**最小絶対偏差（least absolute deviations）**推定とも呼ばれる．これは M 推定において $\rho(x) = |x|$ としたものに対応する．このとき，影響関数は $\psi(x) = \mathrm{sign}(x)$（符号関数）となるが，この値は各々のデータがあてはめる空間の表側にあるか裏側にあるかのラベルと同等であり，このラベル付け問題を解くことと，最小絶対値法を解くことは本質的に同じである．このラベル付けを厳密に行なうには，全組合せである 2 のデータ数乗通り行なう必要があるが，このラベル付けは $x = 0$ で微分できない $\rho(x) = |x|$ を $|x| \approx \sqrt{x^2 + \varepsilon^2}$（$\varepsilon \approx 0$）と置き換えることによって近似的解くことが可能である．

7. 頑健なヤコビ核主成分分析の有効性の実験

本節では，頑健なヤコビ核主成分分析の有効性を示すために，

- 近似距離 $R_{[d]}$ を用いた RANSAC による許容値の選択を行なった（RS-）か否（無印）か．
- KPCA（LS）または JKPCA（JK）を用いて初期値を計算．
- このまま推定値とするか，特徴空間で M 推定を行なう（+M）か，入力空間で M 推定を行なう（+MJ）か．

の $2\times2\times3 = 12$ 通りによるあてはめ結果を比較した．例えば RANSAC による許容値の選択を行ない KPCA を用いて初期値を計算した後に入力空間で M 推定を行なった結果は “RS-LS+MJ” のように表される．

7.1 あてはめられた 2 次曲線の比較

本小節では，汚染率が 10%，30%，50% の場合にあてはめられた曲線を 12 通り（汚染率が 50% の場合は 6 通り）に対して比較する．

まず汚染率が 10% の場合の結果を図 3 に示す．1 段目は LS，2 段目は JK，3 段目は RS-LS，4 段目は RS-JK で，1 列目は無印，2 列目は +M，3 列目は +MJ である． 汚染率が 10% の場合，KPCA は RANSAC

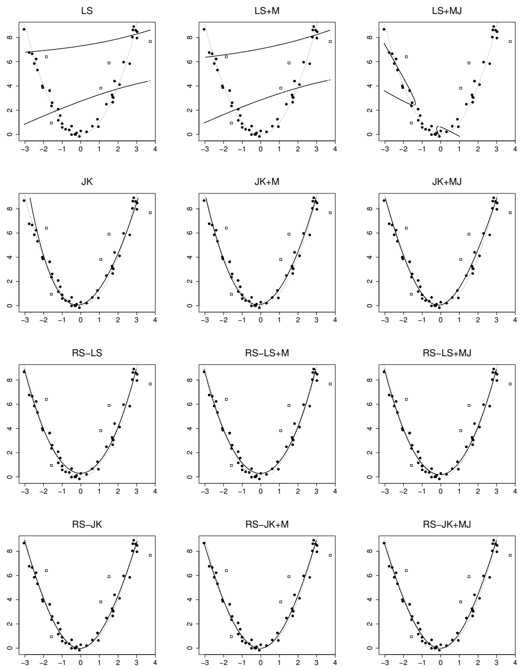


図 3　汚染率 10% のあてはめ結果

を行なわないと破綻するが，KPCA は RANSAC を行なわないと破綻するが，JKPCA は RANSAC を行なわなくとも破綻していないことがわかる．また，M 推定や RANSAC を用いることにより，JKPCA のデータへの過適合が軽減されていることもわかる．

次に汚染率が 30% の場合の結果を図 4 に示す．1 段目は LS，2 段目は JK，3 段目は RS-LS，4 段目は RS-JK で，1 列目は無印，2 列目は +M，3 列目は +MJ である． 汚染率が 30% の場合，KPCA は原点付近の誤差を許容するため RANSAC を行なったとしても破綻することがわかる（原点付近を除けば悪くないあて

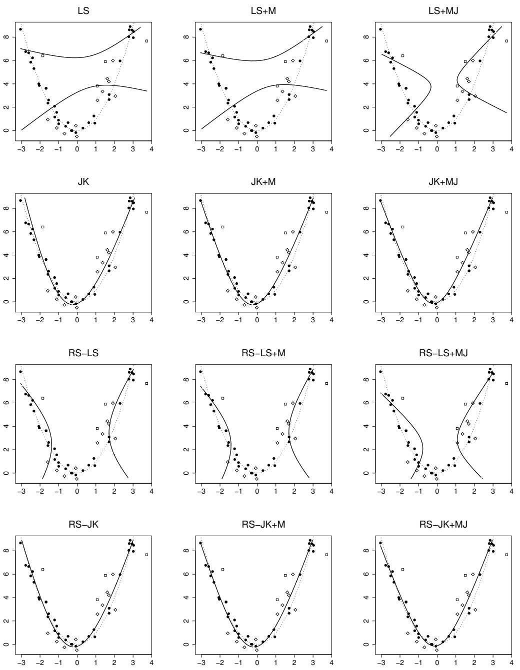


図 4 汚染率 30% のあてはめ結果

はめ結果である）．また，JKPCA は RANSAC を行なわなくとも破綻していないことから，RANSAC の効果よりも入力空間の計量を考慮することの重要性がわかる．

次に汚染率が 50% の場合の結果（LS は除く）を図 5 に示す．1 段目は JK，2 段目は RS-JK で，1 列目は無印，2 列目は +M，3 列目は +MJ である． 汚染

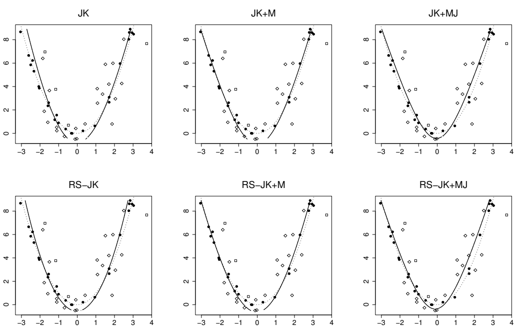


図 5 汚染率 50% のあてはめ結果

率が 50% の場合，JKPCA を用いると RANSAC はあまり効果がないことがわかる．そして近似距離 $R_{[d]}$ に基づく M 推定が推定結果を少し改善していることがわかる．

7.2 あてはめ誤差の比較

汚染率が 30% の場合に対して，許容値 1 個あたりの入力空間におけるあてはめ誤差 $\frac{1}{D}\sum_{d=1}^{D} R_{[d]}$ を比較した． 図 6 により，やはり RANSAC の効果より

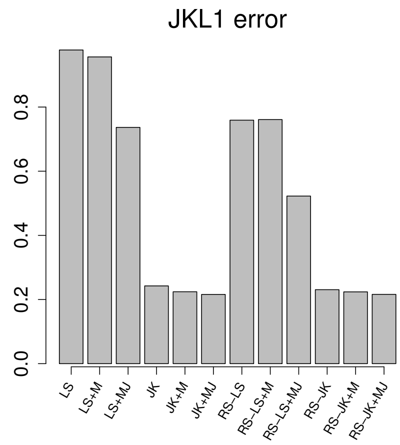


図 6 あてはめ誤差（近似ユークリッド距離）の比較

も JKPCA を用いることの効果の方が大きいことがわかる．

また，汚染率に対する許容値 1 個あたりの入力空間におけるあてはめ誤差の変化を，JKPCA に対して調べた．これは 50 データセットの平均値である．図 7 の 1 行目は JK，2 行目は JK+MJ であり，1 列目は RANSAC 無，2 列目は RANSAC 有である． 図 7

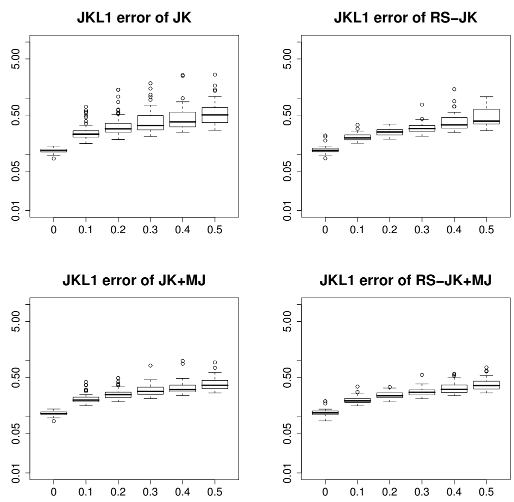


図 7 あてはめ誤差（近似ユークリッド距離）と汚染率

からわかるように，やはり RANSAC の効果よりも，

JKPCA の効果のが高いことがわかる．そして，両方を組合せることにより，若干ではあるが，より良い推定結果が得られていることがわかる．

8. ま と め

本稿では，ヤコビ核主成分分析の頑健化手法を提案した．本稿で与えた例外値が持つ誤差はそれほど大きくなかったためか，RANSAC の効果があまり出なかったが，誤差を平面上の一様分布にしたがってランダムに与えると，図 8 のように RANSAC の効果が出るようになる． 今後は誤差のもつ分布について検討し，頑健性の評価を行ないたい．

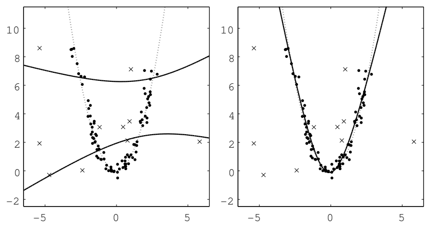


図 8　JKPCA（左）と R-JKPCA（右）(100 点，汚染率 10%)

参 考 文 献

1) M. Aizerman, É. Braverman and L. Rozonoér, "Theoretical foundations of the potential function method in pattern regognition learning," Automation and Remote Control, **25**:821-837, 1964.
2) S. Akaho, "Curve fitting that minimizes the mean square of perpendicular distances from sample points," SPIE, Vision Geometry II, 1993.
3) 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, **J86-D-II**(7):934-942, 2003.
4) 赤穂昭太郎, "カーネル多変量解析-非線形データ解析の新しい展開-," 岩波書店, 2008.
5) W. Chojnacki, M. j. Brooks, A. van den Hangel and D. Gawley, "On the fitting of surface to data with covariances," IEEE TPAMI, **22**(11):1294-1303, 2000.
6) C.Ding, D.Zhou, X.He and H.Zha, R_1-PCA: Rotational invariant L_1-norm principal component analysis for robust subspace factorizaion, In Proc. of ICML2006.
7) M.A.Fischer and R.C.Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," Comm. ACM,**24**:381-395, 1981
8) 藤木淳, 赤穂昭太郎, "球面上の点列への小円あてはめ〜カメラ運動の平滑化に向けて〜," 信学技報 PRMU2004-149:91-96, 2004(12).
9) J. Fujiki and S. Akaho, "Spherical PCA with Euclideanization," Subspace 2007(ACCV07).
10) 藤木淳, 赤穂昭太郎, "球面最小二乗法による球面上の曲線あてはめ," Subspace 2008(MIRU2008).
11) 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, Nov.2008.
12) 藤木淳, 赤穂昭太郎, 日野英逸, 村田昇, "頑健なヤコビ核主成分分析に向けて," 信学技報, Mar.2009.
13) N. H. Gray, P. A. Geiser and J. R. Geiser, "On the least-squares fit of small and great circles to spherically projected orientation data," Mathematical Geology, **12**(3):173-184, 1980.
14) R.Hartley and A.Zisserman. *Multiple view geometry in computer vision.* Cambridge University, Cambridge, 2nd edition, 2003.
15) P. J. Huber, "Robust estimation of a location parameter," Ann. Math. Stat., **35**:73-101, 1964.
16) 金谷健一, 菅谷保之, "幾何学的当てはめの厳密な最尤推定の統一的計算法," 情処研報, 2008-CVIM-164-3:17-24, 2008.
17) Q.Ke and T.Kanade, "Robust L_1 norm factorization in the presence of outliers and missing data by alternative convex programming," In Proc. of CVPR2004, pp.592-599, 2004.
18) Y. Leeden and P. Meer, "Heteroscedastic regression in computer vision: problems with bilinear constraint," IJCV, **37**(2):127-150, 2000.
19) R. J. Rousseeuw and A. M. Leroy, Robust Regression and Outlier Detection, John Wiley & Sons, NY, 1987.
20) P.D.Sampson, "Fitting conic sections to 'very scattered' data: an iterative refinement of the Bookstein algorithm," Comput. Vision, Graphics, and Image Processing, **18**:97-108, 1982.
21) B.Schölkopf, A.Smola and K.-R.Müller, "Nonlinear component analysis as a kernel eigenvalue problem," Neural Computation, **10**:1299-1319, 1998.
22) Y.Sugaya and K.Kanatani, "Outlier removal for motion tracking by subspace separation," IEICE Trans. Inf.&Syst., Jun.2003.
23) G.Taubin, "Estimation of planar curves, surfaces, and nonplanar space curves defined by implicit equations with applicatons to edge and range image segmentation," IEEE TPAMI, **13**(11):1115-1138, 1991.
24) V.Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, 1995.