社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

次元の縮退を利用した頑健な分布の推定法

岩村　雅一† 　大町真一郎† 　阿曽　弘具†

† 東北大学大学院工学研究科
〒 980-8579 仙台市青葉区荒巻字青葉 05

E-mail: †{masa,machi,aso}@aso.ecei.tohoku.ac.jp

あらまし　パターン認識では共分散行列の固有値と固有ベクトルを必要とすることが多いが，共分散行列を少数の学習用のサンプルから推定する場合には固有値の偏りによって認識性能が低下することが知られている．これに対して，小さな固有値が認識性能に悪影響を及ぼすとして認識に用いない手法や固有値展開後に固有値を補正する手法が提案されている．これらはいずれも固有値展開の後，偏った固有値に対して対処する手法であるが，固有値の偏りが固有値展開の際に発生することから，固有値に誤差が生じにくい工夫を固有値展開の前に施すことも可能であると考えられる．本論文では，最初に固有値の偏りが固有値展開で生じることを確認する．そして，固有値展開前に共分散行列を縮退させておくことで，固有値展開によって固有値が偏りにくくする手法を提案する．提案手法が標本共分散行列を用いた場合に比べて真の分布をより正しく推定し，認識性能を改善することを認識実験により確認した．

キーワード　固有値展開，固有値の偏り，次元の縮退

Robust Estimation of Distribution by Shrinkage Technique

Masakazu IWAMURA†, Shinichiro OMACHI†, and Hirotomo ASO†

† Graduate School of Engineering, Tohoku University
Aoba 05, Aramaki, Aoba-ku, Sendai-shi, 980–8579 Japan

E-mail: †{masa,machi,aso}@aso.ecei.tohoku.ac.jp

Abstract　Most pattern recognition applications require the eigenvalues and eigenvectors of the covariance matrix. It is well known that when the number of training samples is small, the eigenvalues of the covariance matrix contains bias, and the bias degrades recognition performance. There are some methods which ignore the small eigenvalues, or acquire better estimates of the covariance matrix by correcting the eigenvalues. Though all of these methods cope with the eigenvalues obtained after eigen decomposition, eigen decomposition seems to cause the biases of the eigenvalues. Therefore, it is worth trying to devise a method which avoids bias of eigenvalues. In this paper, it is confirmed that biases of the eigenvalues appear after eigen decomposition by experiments. Then a method of shrinking the covariance matrix before eigen decomposition for avoiding bias of the eigenvalues are proposed. The ability of the proposed method of estimating the true distribution more precisely than using the sample covariance matrix and of improving recognition performance is confirmed by the recognition experiments.

Key words　eigen decomposition, bias of eigenvalue, shrinkage of dimensionality

1. はじめに

パターン認識で用いられる主要な識別関数は共分散行列の逆行列を必要とするため，固有値展開によって共分散行列の固有値と固有ベクトルに分解し，固有値と固有ベクトルから逆行列を構成することが多い．ところが，学習用のサンプルから求めた標本共分散行列が推定誤差を含む場合には標本共分散行列の固有値と固有ベクトルが誤差を含み，大きな固有値がより大きく，小さな固有値がより小さく偏るため [1]，認識性能が低下することが知られている．

固有値の誤差（偏り）に対して，小さな固有値が認識性能に悪影響を及ぼすとして認識に用いない手法（文献 [2] ~ [5]）や，固有値展開後に固有値を補正する手法（文献 [6] ~ [10]）が提案されている．これらはいずれも誤差（偏り）を含んだ固有値に対して対策を施す手法である．

本論文では，最初に固有値の偏りがどこで生じるかを実験的に調査する．その結果，固有値の偏りは固有値展開で生じると考えられることから，従来のように固有値展開後に固有値を補正するのではなく，固有値展開の際に固有値が偏りにくくなるよう共分散行列を工夫する手法を提案する．具体的には，特徴量の次元数が大きくなると固有ベクトルの推定誤差が大きくなり，このことによって固有値が偏ると考えられることから，標本共分散行列を調べて，分布が等方性に近いと考えられる（超）平面を固有値展開前に縮退させる．縮退した共分散行列に対して固有値展開を行い，得られる縮退した固有値と固有ベクトルを元の次元数に復元する．認識実験により，提案手法が標本共分散行列を用いた場合に比べて真の分布をより正しく推定し，認識性能を改善することを確認する．

2. 固有値の偏りに関する考察

サンプルから標本共分散行列の固有値と固有ベクトルを推定する過程は，主に標本共分散行列の推定と固有値展開の 2 つの段階に分けることができる．真の固有ベクトルの与え方を工夫することによって，標本固有値，標本固有ベクトルに生じる偏りがどの段階で生じるのか確認する．

2.1 固有値，固有ベクトルの推定手順

標本共分散行列の固有値と固有ベクトルを推定する手順について述べる．

2.1.1 標本共分散行列の推定

学習用の標本を $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ とし，d 次元ベクトル $\boldsymbol{X}_k$ を

$$\boldsymbol{X}_k = (X_{k1}, X_{k2}, \ldots, X_{kd})^{\mathrm{T}} \tag{1}$$

とおくと，標本共分散行列 $\hat{\boldsymbol{\Sigma}}$ は

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{k=1}^{n}(\boldsymbol{X}_k - \boldsymbol{\mu})(\boldsymbol{X}_k - \boldsymbol{\mu})^{\mathrm{T}} \tag{2}$$

で与えられる．ここで $\hat{\boldsymbol{\mu}}$ は標本平均ベクトル

$$\hat{\boldsymbol{\mu}} = \frac{1}{n}\sum_{k=1}^{n}\boldsymbol{X}_k \tag{3}$$

である．

2.1.2 固有値展開

標本共分散行列の第 i 固有値，固有ベクトルを $\hat{\lambda}_i$，$\hat{\boldsymbol{\phi}}_i$ とし，$\hat{\lambda}_i$，$\hat{\boldsymbol{\phi}}_i$ からなる行列を

$$\hat{\boldsymbol{\Lambda}} = \mathrm{diag}\left(\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_d\right) \tag{4}$$

$$\hat{\boldsymbol{\Phi}} = \left[\hat{\boldsymbol{\phi}}_1 \quad \hat{\boldsymbol{\phi}}_2 \quad \cdots \quad \hat{\boldsymbol{\phi}}_d\right] \tag{5}$$

とおく．このとき標本共分散行列 $\hat{\boldsymbol{\Sigma}}$ からその固有値，固有ベクトルを求める固有値展開は，

$$\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{\Phi}}\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{\Phi}}^{\mathrm{T}} \tag{6}$$

と表せる．

2.2 固有値に偏りが生じる処理の特定

2.2.1 特定方法

真の第 i 固有ベクトル $\boldsymbol{\phi}_i$ からなる行列 $\boldsymbol{\Phi}$ として，

$$\boldsymbol{\Phi} = \left[\boldsymbol{\phi}_1 \quad \boldsymbol{\phi}_2 \quad \cdots \quad \boldsymbol{\phi}_d\right] \equiv \boldsymbol{I} \tag{7}$$

を考え，$\boldsymbol{\Phi}$ をパラメータとして人工サンプルを作成する．このようなサンプルから求めた標本共分散行列の対角成分は，対応する固有値の推定値とみなせ，いつ固有値の偏りが生じるかによって次の状況が予想できる．

（1）偏りが標本共分散行列の推定で生じる場合

固有値展開を行う前の段階で，標本共分散行列の対角成分は大きな偏りを含む．

（2）偏りが固有値展開で生じる場合

固有値展開を行う前の段階では標本共分散行列の対角成分は大きな偏りを含まず，固有値展開を行った後に固有値が大きな偏りを含む．

	軸 1	軸 2
共分散行列の対角成分	0.987	0.993
固有値	1.38	0.599

表 1　固有値の偏り (次元数 2 , 分散 1 の等方性正規分布)

	軸 1	軸 2
共分散行列の対角成分	1.99	0.982
固有値	2.22	0.748

表 2　固有値の偏り (次元数 2 , 分散 2 , 1 の異方性正規分布)

2.2.2　調　　査

固有値がどの段階で偏るのかを調べるために，正規分布に従う人工サンプルを発生させ，人工サンプルから標本共分散行列を求め，固有値展開を行う．このとき，固有値展開の前後で標本共分散行列の対角成分または固有値に偏りが生じるかを調べる．サンプル数は 10 とする．実験は 1,000 回繰り返し，その平均を結果として示す．

真の分布を (a) 次元数 2 , 分散 1 の等方性正規分布としたとき，(b) 次元数 2 , 分散 2 , 1 の異方性正規分布としたときの結果を表 1 , 2 に示す．(a) の場合 (表 1) は，固有値展開前の共分散行列の対角成分はほぼ正しく推定されているが，固有値展開後の固有値は軸 1 が大きく，軸 2 が小さく偏っている．(b) の場合 (表 2) についても，固有値展開前の共分散行列の対角成分はほぼ正しく推定されているが，固有値展開後の固有値は大きな固有値を持つ軸 1 が大きく，小さな固有値を持つ軸 2 が小さく偏っている．これらの結果は，固有値の偏りが固有値展開で生じることを示している．このことから，固有値展開前に共分散行列に工夫を施すことで頑健な分布の推定が可能になると考えられる．

3.　頑健な固有値の推定法

特徴量の次元数が大きく，固有値の大きさが同程度という状況では固有値の偏りが特に大きいと考えられる．本節ではこの状況に対処する頑健な固有値の推定法を提案する．

3.1　提案手法の概要

提案手法は，固有値展開を行う前に，標本共分散行列の中に含まれる分散 (固有値) の大きさが同程度の数本の軸 (多次元) を 1 本の軸 (1 次元) に縮退させ，固有値展開を行った後に元の次元数に復元する手法である．

軸を縮退させる際に，異なる大きさの固有値を同じ値にすることによる誤差が生じるが，元々分散の大きさが近い軸同士を縮退させているため，このような誤差はそれほど大きくないと考えられる．

3.1.1　共分散行列の縮退方法

3×3 の共分散行列を 2×2 に縮退させるときを例に，提案手法の共分散行列の縮退方法について述べる．

共分散行列は特徴ベクトルの各次元の相関を表す．n 本の 3 次元特徴ベクトルを

$$\boldsymbol{X}_k = (X_{k1}, X_{k2}, X_{k3})^{\mathrm{T}}, \qquad k = 1, \ldots, n \quad (8)$$

とおいたとき，各成分間 (X_{k1} と X_{k2} , X_{k1} と X_{k3} , X_{k2} と X_{k3}) にはそれぞれ相関があり，共分散行列はこの相関を表す．ここで 1 次元目 (X_{k1}) と 2 次元目 (X_{k2}) が作る標本共分散行列の固有値が同程度の大きさで，等方性の分布とみなして差し支えないとする．等方性の分布の分散を推定する場合，ベクトルの複数の成分を 1 次元のサンプルみなして，1 次元の分散と同様に計算するので，この場合も 1 次元目 (X_{k1}) と 2 次元目 (X_{k2}) の成分を 1 つの成分であるとして扱う．すなわち，1 次元目 (X_{k1}) と 2 次元目 (X_{k2}) を縮退させた共分散行列は，次のような $2n$ 本の 2 次元特徴ベクトル

$$\left\{ \begin{array}{c} \boldsymbol{X}_{2k-1} = (X_{k1}, X_{k3}) \\ \boldsymbol{X}_{2k} = (X_{k2}, X_{k3}) \end{array} \right\}, \quad k = 1, \ldots, n \quad (9)$$

の共分散行列を求めるのとほぼ同様に求める．ただし，縮退させた成分とその他の成分には相関 (共分散) があるので，X_{k1} と X_{k2} に相関がある場合は無相関にしてから縮退させ，無相関化のために用いた正規直交基底を固有ベクトルの一部と考える．

3.2　提案手法のアルゴリズム

3.2.1　準　　備

提案手法の具体的なアルゴリズムを述べる前に，アルゴリズムの説明に必要な記号，関数，用いるパラメータについて述べる．

元の共分散行列を $\boldsymbol{\Sigma}$ ($d \times d$ 行列) , $\boldsymbol{\Sigma}$ の縮退させる成分同士を無相関化した行列を $\boldsymbol{L}$, $\boldsymbol{L}$ を縮退させた行列を $\boldsymbol{L}'$ とする．$\boldsymbol{L}'$ は $m \times m$ 行列 ($m \leqq d$) で，m はアルゴリズムの途中で確定する．

$\boldsymbol{\Sigma}$ の第 k 行 (第 k 列) を軸 k と呼ぶ．軸 k が縮退した結果，$\boldsymbol{L}'$ の第 i 行 (第 i 列) になることを，集合 $\mathcal{E}_i$ が軸 k を要素とすることで表す．初期状態では $\mathcal{E}_i = \{i\}$ $(i = 1, \ldots, d)$ のように，どの集合も要素を 1 つずつ持っているとする．

集合 $\mathcal{E}_i$ に属する軸 $\{i_1, \ldots, i_{|\mathcal{E}_i|}\}$ の x 番目の要素

i_x を返す関数を $\mathrm{ord}(i,x)$ とする．また，軸 k が属する集合 $\mathcal{E}_i$ の添字の番号 i を返す関数を $\mathrm{sind}(k)$，軸 k と軸 l が作る 2 次元共分散行列 $\begin{bmatrix}\sigma_{kk} & \sigma_{kl}\\ \sigma_{lk} & \sigma_{ll}\end{bmatrix}$ の 2 つの固有値の比を返す関数を $\mathrm{r}(k,l)$ とする．ただし，$\mathrm{r}(k,l) \geqq 1$ とする．

提案手法では 2 つのパラメータ α と β を用いる．α は統合する平面の候補を選ぶ際に使用し，β は統合を抑制する際に使用するパラメータである．これらのパラメータの詳細については次に示す提案手法のアルゴリズムの中で述べる．

3.2.2 提案手法のアルゴリズム

（ 1 ） 統合する平面の候補の選定標本共分散行列の $\frac{d(d-1)}{2}$ 個の各 2 次元平面（ 軸 e_i と e_j の組 ）に対して固有値の比 $\mathrm{r}(k,l)$ を計算し，$\mathrm{r}(k,l) \leqq \alpha$ を満たす k と l の組を「 統合候補平面 」とする．

（ 2 ） 統合する軸の決定

統合候補平面を構成する軸を k と l としたとき，k と l が属する集合 $\mathcal{E}_{\mathrm{sind}(k)}$ と $\mathcal{E}_{\mathrm{sind}(l)}$ を統合する．ただし，$a \in \mathcal{E}_{\mathrm{sind}(k)}$，$b \in \mathcal{E}_{\mathrm{sind}(l)}$ の組の中に，$\mathrm{r}(a,b) > \beta$ を満たす組がある場合は統合しない．これを全ての統合候補平面について，固有値の比 r が小さい順に評価する．

ここまでの処理で，縮退した共分散行列の大きさ m が確定する．最後に，$\{\mathcal{E}_i\}$ の添字の番号が 1 から m までになるように番号を付け直す（ 順番は任意 ）．

（ 3 ） $\boldsymbol{\Sigma}$ の部分共分散行列 $\tilde{\boldsymbol{\Sigma}}_{ii}$ の作成

$k = 1,\ldots,|\mathcal{E}_i|$，$l = 1,\ldots,|\mathcal{E}_j|$ とする．行列 $\tilde{\boldsymbol{\Sigma}}_{ij}$ は $\tilde{\boldsymbol{\Sigma}}_{ij}$ の kl 成分 $\tilde{\sigma}_{ij,kl}$ が

$$\tilde{\sigma}_{ij,kl} = \sigma_{\mathrm{ord}(i,k),\mathrm{ord}(j,l)} \tag{10}$$

で定まる $|\mathcal{E}_i| \times |\mathcal{E}_j|$ 行列とする．ただし，σ_{ij} は $\boldsymbol{\Sigma}$ の ij 成分である．

（ 4 ） 部分共分散行列 $\tilde{\boldsymbol{\Sigma}}_{ii}$ の要素の無相関化

$i = 1,\ldots,m$ について，$\tilde{\boldsymbol{\Sigma}}_{ii}$ の固有値，固有ベクトルからなる $|\mathcal{E}_i| \times |\mathcal{E}_i|$ 行列 $\boldsymbol{L}_{ii}$，$\boldsymbol{P}_{ii}$ を次式の固有値展開で求める．

$$\tilde{\boldsymbol{\Sigma}}_{ii} = \boldsymbol{P}_{ii}\boldsymbol{L}_{ii}\boldsymbol{P}_{ii}^{\mathrm{T}} \tag{11}$$

$i \neq j$ については，$\boldsymbol{L}_{ij}$ を $\boldsymbol{L}_{ij} = \boldsymbol{P}_{ii}^{\mathrm{T}}\tilde{\boldsymbol{\Sigma}}_{ij}\boldsymbol{P}_{jj}$ で定まる $|\mathcal{E}_i| \times |\mathcal{E}_j|$ 行列とする．$\boldsymbol{L}_{ij}$ は，$i = j$ のときは対角行列になるが，$i \neq j$ のときは必ずしも 0 にならない．

このとき，$\boldsymbol{P}_{ii}$ を対角成分に持つ $d \times d$ 行列 $\boldsymbol{P}$ を

$$\boldsymbol{P} = \begin{bmatrix}\boldsymbol{P}_{11} & & 0\\ & \ddots & \\ 0 & & \boldsymbol{P}_{mm}\end{bmatrix} \tag{12}$$

とおくと，$\boldsymbol{P}$ は正規直交基底である．さらに，$\tilde{\boldsymbol{\Sigma}}$，$\boldsymbol{L}$ を

$$\tilde{\boldsymbol{\Sigma}} = \begin{bmatrix}\tilde{\boldsymbol{\Sigma}}_{11} & \ldots & \tilde{\boldsymbol{\Sigma}}_{1m}\\ \vdots & \ddots & \vdots\\ \tilde{\boldsymbol{\Sigma}}_{m1} & \ldots & \tilde{\boldsymbol{\Sigma}}_{mm}\end{bmatrix} \tag{13}$$

$$\boldsymbol{L} = \begin{bmatrix}\boldsymbol{L}_{11} & \ldots & \boldsymbol{L}_{1m}\\ \vdots & \ddots & \vdots\\ \boldsymbol{L}_{m1} & \ldots & \boldsymbol{L}_{mm}\end{bmatrix} \tag{14}$$

とおくと，$\boldsymbol{L} = \boldsymbol{P}^{\mathrm{T}}\tilde{\boldsymbol{\Sigma}}\boldsymbol{P}$ を満たす．

（ 5 ） $\boldsymbol{L}$ の縮退

$\boldsymbol{L}$ を縮退させて $\boldsymbol{L}'$ を作成する．$i = 1,\ldots,m$ について以下の処理を行う．ただし，l_{ijkl} は $\boldsymbol{L}_{ij}$ の kl 成分を表す．

（ a ） $\boldsymbol{L}'$ の対角成分

$$l'_{ii} = \frac{1}{|\mathcal{E}_i|}\sum_{k=1}^{|\mathcal{E}_i|} l_{iikk} \tag{15}$$

（ b ） $\boldsymbol{L}'$ の非対角成分

$$l'_{ij} = \frac{1}{|\mathcal{E}_i||\mathcal{E}_j|}\sum_{k=1}^{|\mathcal{E}_i|}\sum_{l=1}^{|\mathcal{E}_j|} l_{ijkl} \tag{16}$$

（ 6 ） 縮退した行列 $\boldsymbol{L}'$ の固有値展開

固有値展開により，$\boldsymbol{L}'$ の固有値，固有ベクトルからなる $m \times m$ 行列 $\boldsymbol{\Lambda}' = \mathrm{diag}\,(\lambda'_1, \lambda'_2, \ldots, \lambda'_m)$，$\boldsymbol{\Psi}'$ を次式の固有値展開で求める．

$$\boldsymbol{L}' = \boldsymbol{\Psi}'\boldsymbol{\Lambda}'\boldsymbol{\Psi}'^{\mathrm{T}} \tag{17}$$

（ 7 ） $\boldsymbol{L}$ の固有値，固有ベクトルの復元

固有値展開の結果得られる m 次元の固有値と固有ベクトルを以下の手順で d 次元に復元する．復元された固有ベクトル $\boldsymbol{\Phi}$ は $\boldsymbol{L}$ の固有ベクトルに相当する．

（ a ） 固有値 $\boldsymbol{\Lambda}$ の復元

復元後の第 k 固有値を λ_k とする．$k = 1,\ldots,d$ について，次式のように定める．

$$\lambda_k = \frac{\lambda'_{\mathrm{sind}(k)}}{\sqrt[d]{|\boldsymbol{\Sigma}|}} \tag{18}$$

（ b ） 固有ベクトルの復元

固有ベクトルは正規直交基底であるので，直交性と正規性を満たすことが必要である．最初に 1 次独立な基底 $\boldsymbol{\Psi}$ を作成し，その後 Gram-Schmidt の直交化で正規直交基底 $\boldsymbol{\Psi}_{\mathrm{GS}}$ を求める．

i. 1 次独立な基底 $\boldsymbol{\Psi}$ の作成

$k,l = 1,\ldots,d$ について，ψ_{kl} を以下のように定め

る．これにより $\mathrm{sind}(p) \neq \mathrm{sind}(q)$ を満たす p，q について，ψ_p と ψ_q は直交する．また，どのベクトルも 1 次従属にならない．

A. $\boldsymbol{\Psi}'$ の対角成分（$\mathrm{sind}(k) = \mathrm{sind}(l)$ のとき）

$$\psi_{kl} = \begin{cases} \psi'_{\mathrm{sind}(k),\mathrm{sind}(l)}, & \text{for } k = l \\ 0, & \text{for } k \neq l \end{cases} \tag{19}$$

B. $\boldsymbol{\Psi}'$ の非対角成分（$\mathrm{sind}(k) \neq \mathrm{sind}(l)$ のとき）

$$\psi_{kl} = \frac{\psi'_{\mathrm{sind}(k),\mathrm{sind}(l)}}{\sqrt{|\mathcal{E}_{\mathrm{sind}}(k)||\mathcal{E}_{\mathrm{sind}}(l)|}} \tag{20}$$

ii. $\boldsymbol{\Psi}$ の正規直交化

$\boldsymbol{\Psi}$ に Gram-Schmidt の直交化を適用して得られる行列を $\boldsymbol{\Psi}_{\mathrm{GS}}$ とする．

（8）$\boldsymbol{\Sigma}$ の固有ベクトル $\boldsymbol{\Phi}$ の計算

$\boldsymbol{\Lambda}$，$\boldsymbol{L}$，$\boldsymbol{\Sigma}$ の関係は，

$$\begin{aligned} \boldsymbol{\Lambda} &\sim \boldsymbol{\Psi}_{\mathrm{GS}}^{\mathrm{T}} \boldsymbol{L} \boldsymbol{\Psi}_{\mathrm{GS}} \\ &= \boldsymbol{\Psi}_{\mathrm{GS}}^{\mathrm{T}} \boldsymbol{P}^{\mathrm{T}} \boldsymbol{\Sigma} \boldsymbol{P} \boldsymbol{\Psi}_{\mathrm{GS}} \end{aligned} \tag{21}$$

と書けるので，$\boldsymbol{\Phi}$ を次式で求める．

$$\boldsymbol{\Phi} = \boldsymbol{P} \boldsymbol{\Psi}_{\mathrm{GS}} \tag{22}$$

4. 提案手法の性能評価

4.1 共分散行列の推定能力

提案手法で求めた固有値と固有ベクトルが表す分布が，標本共分散行列の固有値と固有ベクトルが表す分布に比べて真の分布との距離が近くなっていることを Kullback divergence [11] を用いて確認する．

Kullback divergence は，$f(\boldsymbol{x})$，$g(\boldsymbol{x})$ を確率分布とすると次式で定義される．

$$D_{\mathrm{KL}} = \int \{f(\boldsymbol{x}) - g(\boldsymbol{x})\} \log \frac{f(\boldsymbol{x})}{g(\boldsymbol{x})} d\boldsymbol{x} \tag{23}$$

$f(\boldsymbol{x})$，$g(\boldsymbol{x})$ がともに正規分布で，$f(\boldsymbol{x})$ が平均 $\boldsymbol{\mu}_1$，共分散行列 $\boldsymbol{\Sigma}_1$ によって $g(\boldsymbol{x})$ が平均 $\boldsymbol{\mu}_2$，共分散行列 $\boldsymbol{\Sigma}_2$ によって定義される場合の Kullback divergence は次式で与えられる．

$$\begin{aligned} D_{\mathrm{KL}} = &\frac{1}{2} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^{\mathrm{T}} \left(\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1}\right) (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \\ &+ \mathrm{tr}\left[\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_2 + \boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_1 - 2\boldsymbol{I}\right] \end{aligned} \tag{24}$$

ここでは平均が同じで分散のみ異なる正規分布に従うサンプルについて実験するので，式(24)で $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ とした

$$D'_{\mathrm{KL}} = \mathrm{tr}\left[\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_2 + \boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_1 - 2\boldsymbol{I}\right] \tag{25}$$

で評価を行う．

	Kullback divergence
標本共分散行列	10044
提案手法の固有値，固有ベクトル	4371

表 3　真の分布との距離

実験では，正規分布に従う人工特徴量を用いる．最初に文字画像の特徴量を作成し，文字画像の特徴量から推定した分布パラメータを用いて人工特徴量を作成する．まず，NIST Special Database 19 [12] の数字サンプルを 64×64 の大きさに非線形正規化 [13] した後，196 次元の方向線素特徴量 [14] を抽出し，文字画像の特徴量とする．次に，クラス i の平均ベクトル $\boldsymbol{\mu}_i$ と共分散行列 $\boldsymbol{\Sigma}_i$ をクラス毎に文字画像の特徴量 36,000 文字分から推定する．人工サンプルはこの推定値を真の値として，正規分布 $\mathrm{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ に従うように乱数を用いて作成する．このように作成すると，同じクラスの文字画像の特徴量と人工特徴量は平均がほぼ等しく，分布の形が異なる特徴量になる．

提案手法のパラメータは $\alpha = 1.1$，$\beta = 5$ を用いた．10 字種の Kullback divergence の平均を表 3 に示す．提案手法で求めた固有値，固有ベクトルが表す分布標本共分散行列の固有値と固有ベクトルが表す分布に比べて真の分布との距離が近くなっていることが確認できた．

4.2 認識性能

提案手法の有効性を確認するため，認識実験を行い，既存の固有値を補正する手法との比較から提案手法の有効性を確認する．

実験は各 100 回行い，得られた認識率を平均する．学習サンプル数は 200 個とし，テストサンプルには学習用とは異なるサンプルを 1,000 個ずつ用いる．

提案手法（“Prop”）の比較対象としては，通常の固有値展開で得られた固有値，固有ベクトルを用いた場合（“Orig”）と，この固有値を以下の手法で補正したものを用いる．補正は，James-Stein 推定量（“JS”）[6]，Stein 推定量（“Stein”）[7]（損失関数 $L_2(\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = \mathrm{tr}(\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1} - \boldsymbol{I})^2$ に対する）Haff 推定量（“Haff”）[8], [15]，酒井らの手法（“RQDF”）[9] を用いる．提案手法のパラメータ α と β は予備実験の結果，最も認識率が高かった値を用いる．すなわち，文字画像の特徴量の場合は $\alpha = 1.$，$\beta = 17$，人工特徴量の場合は $\alpha = 1.1$，$\beta = 5$ を用いる．識別関数とし

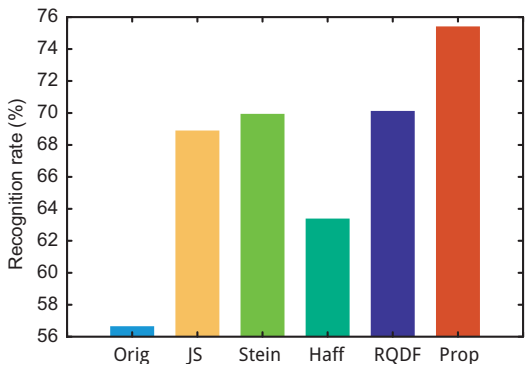

図 1　認識実験の結果 (文字画像)

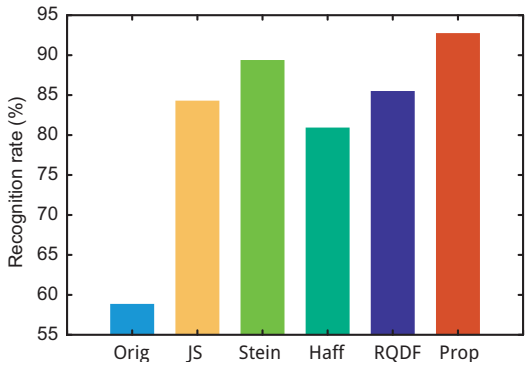

図 2　認識実験の結果 (人工サンプル)

て 2 次識別関数を用いたときの認識結果を図 1，2 に示す．提案手法を用いた場合，既存の手法に比べて良い認識率を得ていて，提案手法が認識に有効な固有値，固有ベクトルを推定していることがわかる．

5. ま と め

パターン認識では共分散行列の固有値と固有ベクトルを必要とすることが多いが，共分散行列を少数の学習用のサンプルから推定する場合には固有値の偏りによって認識性能が低下することが知られているが，これまでに小さな固有値を認識に用いない手法や固有値展開後に固有値を補正する手法が提案されている．これらはいずれも固有値展開の後，偏った固有値に対して対処する手法である．

本論文では，固有値の偏りが固有値展開で生じることを確認し，固有値展開前に共分散行列を縮退させておくことで，固有値展開によって固有値が偏りにくくする手法を提案した．提案手法が標本共分散行列に比べて真の分布をより正しく推定し，認識性能を改善することを認識実験により確認した．

提案手法の導出過程で，分布を特定するような仮定を一切用いていない．そのため，幅広い応用が期待できる．一方，提案手法の 2 つのパラメータは真の分布の違いを吸収しているものと考えられるが，この定め方を導くのは今後の課題である．

文　献

[1] 竹下鉄夫，木村文隆，三宅康二，“マハラノビス距離の推定誤差に関する考察，” 信学論 (D)，vol.J70-D, no.3，pp.567–573，Mar.，1987．

[2] F. Kimura, K. Takashina, S. Tsuruoka and Y. Miyake, “Modified quadratic discriminant functions and the application to Chinese character recognition,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol.9, no.1, pp.149–153, Jan., 1987.

[3] 加藤寧，安倍正人，根元義章，“改良型マハラノビス距離を用いた高精度な手書き文字認識，” 信学論 (D-II)，vol.J79-D-II, no.1，pp.45–52，Jan.，1996．

[4] F. Sun, S. Omachi and H. Aso, “Precise selection of candidates for handwritten character recognition using feature regions,” IEICE Trans. Inf. & Syst., vol.E79-D, no.5, pp.510–515, May, 1996.

[5] J. H. Friedman, “Regularized discriminant analysis,” Journal of American Statistical Association, vol.84, no.405, pp.165–175, 1989.

[6] W. James and C. Stein, Estimation with quadratic loss, Proc. 4th Berkeley Symp. on Math. Statist. and Prob., pp.361–379, 1961.

[7] C. Stein, Estimation of a covariance matrix, Rietz Lecture, 39th Annual Meeting IMS, Atlanta, GA, 1975.

[8] L. R. Haff, “Empirical bayes estimation of the multivariate normal covariance matrix,” Annals of Statistics, vol.8, pp.586–597, 1980.

[9] 酒井充，米田政明，長谷博行，丸山博，直江美知子，“固有値の偏り補正に基づく 2 次識別関数，” 信学論 (D-II)，vol.J82-D-II, no.4，pp.631–640，Apr.，1999．

[10] 岩村雅一，大町真一郎，阿曽弘具，“標本共分散行列の固有ベクトルを用いた真のマハラノビス距離の推定法，” 信学論 (D-II)，vol.J86-D-II, no.1，pp.22–31，Jan.，2003．

[11] 高木幹雄，下田陽久，画像解析ハンドブック，pp.656–658，東京大学出版会，東京，Jan.，1991．

[12] P. J. Grother, NIST special database 19 — handprinted forms and characters database, Technical report, National Institute of Standards and Technology, Mar., 1995.

[13] 山田博三，斉藤泰一，山本和彦，“線密度イコライゼーション—相関法のための非線形正規化法，” 信学論 (D)，vol.J67-D, no.11，pp.1379–1383，Nov.，1984．

[14] 孫寧，田原透，阿曽弘具，木村正行，“方向線素特徴量を用いた高精度文字認識，” 信学論 (D-II)，vol.J74-D-II, no.3，pp.330–339，Mar.，1991．

[15] S. P. Lin and M. D. Pelman, A Monte Carlo comparison of four estimators for a covariance matrix, Vol. 6, pp.411–429, Elsevier Science Publishing Company, Inc., New York, 1985.