抑制付きカーネル部分空間法

鷲沢　嘉一

† (独) 理化学研究所 脳科学総合研究センター　〒351–0198 埼玉県和光市広沢２－１
E-mail: †washizawa@brain.riken.jp

あらまし　CLAFIC 法などの主な部分空間法では，対象クラスの標本のみから入力パターンとクラスの類似度を測る関数を求める．このため，識別器の独立性が高いという利点があるが，複数のクラスに同じ特徴がある場合は，識別に有用な特徴が抽出できない．そこで，他のクラスの特徴を抑制する手法がいくつか提案されている．CLAFIC 法の他の拡張として，カーネルトリックを適用したカーネル部分空間法がある．カーネルトリックは標本を高次元空間へ写像し，問題を高次元の空間から標本数次元の双対問題へと変換することにより解を導く手法である．しかしながら，他のクラスの特徴を抑制する手法にカーネルトリックを適用する場合，全クラスの標本数の行列の演算が必要であり，クラス数が大きいと計算量が非常に大きくなる問題がある．本研究では，解の作用素の空間を制限することで計算量を削減し，現実的な抑制付きカーネル部分空間法を提案する．
キーワード　CLAFIC 法，カーネル部分空間法，カーネルトリック

Kernel Subspace methods with suppression

Yoshikazu WASHIZAWA

† Brain Science Institute, RIKEN
2–1, Hirosawa, Wako-shi, Saitama, 351-0198, Japan
E-mail: †washizawa@brain.riken.jp

Abstract　We propose efficient kernel subspace methods with supression. The experimental results demonstorate advantages of the proposed method.
Key words　CLAFIC, kernel subspace methods, kernel trick

1. はじめに

CLAFIC 法を始めとする部分空間法は，その類似度を測る類似度関数が各クラス標本のみから作成できるという特長がある．このため，文字認識や顔画像などのクラス数が非常に多い問題でも識別器の設計が容易に出来る．また，棄却判定やクラス数の増減への対応，１つのクラスを複数のサブクラスに分割するマルチテンプレート識別，１つの標本が複数のクラスに属するマルチクラス問題などにも柔軟に対応できる．

しかし，複数のクラスのパターンが同様の特徴を有する場合，類似度関数はその特徴を取り出してしまい，クラス間の有意な特徴差を取り出すことが出来ないことがある．このため，混合類似度法や相対 KL 変換法，競合部分空間を考慮した学習部分空間法 (LSM, learning subspace method) などの競合クラスの特徴を抑制しながら，対象クラスの特徴を取り出す手法がいくつか提案されている [1]〜[3]．混合類似度法は，対象クラスの類似クラスの平均パターンを用いる手法であり，相対 KL 変換法は，類似クラスの相関行列を用いる．競合クラスの特徴を抑制するための最も単純な手法として，部分空間の計算時に対象クラスの相関行列 R を固有値分解する代わりに，全クラスの相関行列 Q と正のパラメータ β を用いて，$R-\beta Q$ の固有空間を用いる手法が考えられる．この意味については次節で述べる．以降，この手法を抑制付き部分空間法と呼ぶ．

部分空間法の別の拡張として，カーネルトリックを用いたカーネル部分空間法がある [4], [5]．カーネルトリックは標本を高次元の特徴空間へ非線形写像し，問題を特徴空間の主問題から標本数空間の双対問題へ変換することにより，現実的な計算量で解を求める手法である [6]．CLAFIC 法にカーネルトリックを適用する場合は，各クラスの標本数の大きさの行列演算で解を求めることができる．しかしながら，抑制付き部分空間法では，対象クラスと競合クラス，あるいは全クラスの標本数の行列演算をする必要がある．このため，文字認識などのクラス数が非常に多い場合，計算コストが非現実的に大きくなる．例えば，アルファベットの認識で，各文字につき 1,000 文字の標本を用意した場合，カーネル部分空間法では，1,000x1,000 の行列の固有値分解と行列演算で類似度関数を設計することが出

来るが，抑制付きの場合は，26,000x26,000 の行列演算が必要となり，現実的な計算が難しい. [2], [7], [8] などは，抑制付きの部分空間法にカーネルトリックを適用したものであるが，あらかじめ，類似する標本を選んでおき，計算量を削減している．

本研究では，類似度関数の集合に制約を加えることにより，対象クラスの標本数の行列演算のみで設計できる抑制付きのカーネル部分空間法を提案する．また，手書き数字認識実験で提案手法の有効性を示す．

2. CLAFIC 法と抑制付き部分空間法

2.1 CLAFIC 法 [9], [10]

クラス数を c，入力次元数を d，入力パターンを $\boldsymbol{x}$ とすると CLAFIC 法の識別関数 $f(\boldsymbol{x})$ は，

$$f(\boldsymbol{x}) = \operatorname*{argmax}_{i=1,\ldots,c} g_i(\boldsymbol{x}) \quad (1)$$

$$g_i(\boldsymbol{x}) = \langle \boldsymbol{x}, P_i\boldsymbol{x} \rangle = \|P_i\boldsymbol{x}\|^2 \quad (2)$$

で与えられる．ここで，$\langle\cdot,\cdot\rangle$ は内積，$\|\cdot\|$ は l_2 ノルムを表し，$g_i(\boldsymbol{x})$ はクラス i の類似度関数，P_i はクラス i の Karhunen-Loéve(KL) 部分空間への正射影である．

P_i は，以下の最適化問題の最小解として特徴付けされる．

$$\begin{array}{ll} \displaystyle\min_{Y\in\mathbb{R}^{d\times d}} & \displaystyle\mathop{E}_{\boldsymbol{x}\in\Omega_i} \|\boldsymbol{x} - Y\boldsymbol{x}\|^2 \\ \text{subject to} & \operatorname{rank}(Y) \leqq r \end{array} \quad (3)$$

または，

$$\begin{array}{ll} \displaystyle\max_{Y\in\mathbb{R}^{d\times d}} & \displaystyle\mathop{E}_{\boldsymbol{x}\in\Omega_i} \|Y\boldsymbol{x}\|^2 \\ \text{subject to} & \operatorname{rank}(Y) \leqq r,\ Y^\top = Y,\ YY = Y \end{array} \quad (4)$$

ここで，Ω_i はクラス i のパターンの集合である．最適化問題 (4) の２番目と３番目の制約は Y が正射影であることを表している．解は，クラス i のパターン $\boldsymbol{x} \in \Omega_i$ の KL 部分空間への正射影となり，例えば相関行列 $R_i = E_{\boldsymbol{x}\in\Omega_i}\boldsymbol{x}\boldsymbol{x}^\top$ の大きい r 個の固有値に対応する固有ベクトルなどから求めることができる．

2.2 抑制付き部分空間法

CLAFIC 法は，対象クラスの特徴のみから類似度関数を計算する．このため，複数のクラスに同じ大きい特徴が含まれている場合，その特徴のみを取り出して，有意な特徴差を取り出さないという欠点がある．ここで，対象クラスに対して，同じ特徴を持つクラスを競合クラスと呼ぶ．これより，類似度関数の設計時に，競合クラスの特徴を抑制しながら，対象クラスの特徴を取り出すよう基準で設計すれば識別性能の向上が期待できる．

式 (2) の類似度関数は，射影ベクトルのノルムであるが，P_i は正射影であるため，投影距離を用いた類似度関数

$$g_i'(\boldsymbol{x}) = -\|\boldsymbol{x} - P_i\boldsymbol{x}\|^2 \quad (5)$$

も等価である．最適化問題 (3) より，行列 P_i は対象クラスのパターン $\boldsymbol{x}$ と $P_i\boldsymbol{x}$ の距離が平均的に小さくなるような行列を求めていることになる．競合クラスの特徴を抑制するには，競合クラスのパターン $\boldsymbol{y}$ と $P_i\boldsymbol{y}$ の距離が平均的に大きくなればよい．そこで，競合クラスのパターンの集合を Ψ とおき，以下の最適化問題を考える．

$$\begin{array}{ll} \displaystyle\min_{Y\in\mathbb{R}^{d\times d}} & \displaystyle\mathop{E}_{\boldsymbol{x}\in\Omega_i} \|\boldsymbol{x} - Y\boldsymbol{x}\|^2 - \beta \mathop{E}_{\boldsymbol{y}\in\Psi} \|\boldsymbol{y} - Y\boldsymbol{y}\|^2 \\ \text{subject to} & \operatorname{rank}(Y) \leqq r \end{array} \quad (6)$$

目的関数の第２項が競合クラスを抑制する項である．

最適化問題 (3) の目的関数を J_3 とおくと，

$$\begin{aligned} J_3 &= \mathop{E}_{\boldsymbol{x}\in\Omega_i} \|\boldsymbol{x} - Y\boldsymbol{x}\|^2 \\ &= \mathop{E}_{\boldsymbol{x}\in\Omega_i} \operatorname{Trace}[\boldsymbol{x}\boldsymbol{x}^\top - 2Y\boldsymbol{x}\boldsymbol{x}^\top + Y\boldsymbol{x}\boldsymbol{x}^\top Y^\top] \\ &= \operatorname{Trace}[R_i - 2YR_i + YR_iY^\top] \end{aligned}$$

となるのに対し，最適化問題 (6) の目的関数を J_5 とおくと同様に，

$$J_5 = \operatorname{Trace}[(R_i - \beta Q) - 2Y(R_i - \beta Q) + Y(R_i - \beta Q)Y^\top]$$

となる．ここで，Q は $\boldsymbol{y} \in \Psi$ の相関行列 $Q = \mathop{E}_{\boldsymbol{x}\in\Psi} \boldsymbol{y}\boldsymbol{y}^\top$ である．β は十分に小さく，$(R_i - \beta Q)$ の固有値はすべて非負であるとすると，最小解は $R_i - \beta Q$ 固有値分解して求まる固有空間への正射影となる．

2.3 正則化抑制付き部分空間法

抑制付き部分空間法を行う場合，$R - \beta Q$ が非負である必要があった．もし，$R - \beta Q$ に負の固有値がある場合には，最適化問題 (6) の目的関数が負の無限大となってしまい，類似度関数としての役割を果たさない．これを防ぐために正則化を導入することができる．部分空間法に正則化を導入した手法は，正則化２次識別法，ハイブリッド法と呼ばれる [11], [12].

以下の最適化問題を考える．

$$\begin{array}{ll} \displaystyle\min_{Y\in\mathbb{R}^{d\times d}} & \displaystyle\mathop{E}_{\boldsymbol{x}\in\Omega_i} \|\boldsymbol{x} - Y\boldsymbol{x}\|^2 - \beta \mathop{E}_{\boldsymbol{y}\in\Psi} \|\boldsymbol{y} - Y\boldsymbol{y}\|^2 + \mu\|Y\|_F^2 \\ \text{subject to} & \operatorname{rank}(Y) \leqq r \end{array} \quad (7)$$

ここで，$\mu > 0$ は正則化パラメータ，$\|\cdot\|_F$ は Frobenius ノルムを表す．

正則化を用いる場合は，ランク制約がない場合でも類似度関数を構成することができる．ランク制約がない場合 ($r = d$ の場合)，最適化問題 (7) の目的関数 J_7 は

$$\begin{aligned} J_7 &= \operatorname{Trace}[R_i - YR_i - R_iY^\top + YR_iY^\top \\ &\qquad -\beta(Q - YQ - QY^\top + YQY^\top) + \mu YY^\top] \\ &= \operatorname{Trace}[Y(R_i - \beta Q + \mu I)Y^\top - Y(R_i - \beta Q) \\ &\qquad -(R_i - \beta Q)Y^\top + (R_i - \beta Q)] \end{aligned} \quad (8)$$

となる．ここで I は単位行列を表す．このとき，J_7 の Y における Z 方向への Gâteaux 微分は

$$\begin{aligned} &\delta J_7(Y; Z) \\ &= \lim_{\delta\to 0} \frac{1}{\delta} \operatorname{Trace}[J_7(Y + \delta Z) - J_7(Y)] \\ &= 2\lim_{\delta\to 0} \frac{1}{\delta} \operatorname{Trace}[(\delta Z)\Big((R_i - \beta Q + \mu I)Y^\top - (R_i - \beta Q)\Big)] \end{aligned}$$

となる．$R_i - \beta Q + \mu I$ が正則となるように β, μ をとれば，任意の方向 Z に対して，$\delta J_7(Y; Z) = 0$ となるのは，

$$Y = (R_i - \beta Q)(R_i - \beta Q + \mu I)^{-1} \tag{9}$$

となるときである．$\beta = 0$ のときは，Y は正則化２次識別法の解となる．

ランク制約がある場合は，式 (8) より，

$$\begin{aligned}
J_7 \;=\; & \mathrm{Trace}[\Big(Y(R_i - \beta Q + \mu I)^{1/2} \\
& -(R_i - \beta Q)(R_i - \beta Q + I)^{-1/2}\Big) \\
& \Big((R_i - \beta Q + \mu I)^{1/2} Y \\
& -(R_i - \beta Q + I)^{-1/2}(R_i - \beta Q)\Big)] \\
& -\mathrm{Trace}[(R_i - \beta Q)(R_i - \beta Q + \mu I)(R_i - \beta Q)] \\
& +\mathrm{Trace}[(R_i - \beta Q)] \\
= \; & \|Y(R_i - \beta Q + \mu I)^{\frac{1}{2}} - (R_i - \beta Q)(R_i - \beta Q + I)^{\frac{-1}{2}}\|_F^2 \\
& -\mathrm{Trace}[(R_i - \beta Q)(R_i - \beta Q + \mu I)(R_i - \beta Q)] \\
& +\mathrm{Trace}[(R_i - \beta Q)]
\end{aligned} \tag{10}$$

第２項と第３項は Y に関係しないため，第１項を J_7' とおく．

$$J_7' = \|Y(R_i - \beta Q + \mu I)^{\frac{1}{2}} - (R_i - \beta Q)(R_i - \beta Q + \mu I)^{-\frac{1}{2}}\|_F^2$$

ここで，$R - \beta Q$ の固有値分解を

$$R - \beta Q = \sum_{i=1}^{d} \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^\top \tag{11}$$

とし，固有値は降順に並んでいるとする．また，μ は最小固有値よりも大きくとる．

$$(R_i - \beta Q)(R_i - \beta Q + \mu I)^{-\frac{1}{2}} = \sum_{i=1}^{d} \frac{\lambda_i}{\sqrt{\lambda_i + \mu}} \boldsymbol{u}_i \boldsymbol{u}_i^\top \tag{12}$$

であり，$R_i - \beta Q + \mu I$ が正則であるため，J_7' が最小となるときは

$$\begin{aligned}
Y(R_i - \beta Q + \mu I)^{\frac{1}{2}} \;&=\; \sum_{i=1}^{r} \frac{\lambda_i}{\sqrt{\lambda_i + \mu}} \boldsymbol{u}_i \boldsymbol{u}_i^\top \\
Y \;&=\; \sum_{i=1}^{r} \frac{\lambda_i}{\lambda_i + \mu} \boldsymbol{u}_i \boldsymbol{u}_i^\top
\end{aligned} \tag{13}$$

となるときである．

3. カーネル部分空間法と抑制付きカーネル部分空間法

3.1 カーネル部分空間法

カーネル部分空間法では，入力 $\boldsymbol{x}$ は非線形写像 Φ により入力次元よりも高次元の空間 $\mathcal{F}$ に写像される．

$$\Phi : \mathbb{R} \to \mathcal{F} : \boldsymbol{x} \mapsto \Phi(\boldsymbol{x}) \tag{14}$$

このとき，写像 Φ を明示的に定義するのではなく，Mercer カーネルと呼ばれる以下の条件を満たす関数を用いて計算を行う．

$$k(\boldsymbol{x}_1, \boldsymbol{x}_2) = \langle \Phi(\boldsymbol{x}_1), \Phi(\boldsymbol{x}_2) \rangle \tag{15}$$

代表的な Mercer カーネルとして以下のものが知られている．

$$k(\boldsymbol{x}_1, \boldsymbol{x}_2) \;=\; \langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle \tag{16}$$

$$k(\boldsymbol{x}_1, \boldsymbol{x}_2) \;=\; (\langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle + c)^d \tag{17}$$

$$k(\boldsymbol{x}_1, \boldsymbol{x}_2) \;=\; \exp(-c\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^2) \tag{18}$$

ここで，実数 c と自然数 d はパラメータである．

L_i をクラス i の標本数とし，クラス i の標本を $\boldsymbol{x}_1^i, \ldots, \boldsymbol{x}_{L_i}^i$ と表す．高次元特徴空間 $\mathcal{F}$ に写像されたパターンの標本相関行列 R_i^Φ は

$$R_i^\Phi = \frac{1}{L_i} \sum_{j=1}^{L_i} \Phi(\boldsymbol{x}_j^i) \Phi(\boldsymbol{x}_j^i)^\top \tag{19}$$

で与えられる．カーネル関数に式 (18) の Gauss カーネルを用いた場合は，$\mathcal{F}$ が無限次元関数空間となり，$\Phi(\boldsymbol{x}_j^i)$ が関数となるため，転置を Neumann-Schatten 積 $\Phi(\boldsymbol{x}_j^i) \otimes \overline{\Phi(\boldsymbol{x}_j^i)}$ あるいはケットブラ表記 $|\Phi(\boldsymbol{x}_j^i)\rangle\langle\Phi(\boldsymbol{x}_j^i)|$ に読み替えればよい．これ以降，簡便のため，クラスを表す添え字 i は省略する．

CLAFIC 法では，相関行列 R を固有値分解して射影行列を求めたが，R^Φ は，非常に高次元であるため，現実的に固有値分解を計算することは難しい．そこで，Φ で写像された標本 $\Phi(\boldsymbol{x}_1), \ldots \Phi(\boldsymbol{x}_L)$ を並べた行列あるいは作用素 S を考える．

$$S = [\Phi(\boldsymbol{x}_1), \ldots, \Phi(\boldsymbol{x}_L)] = \sum_{j=1}^{L} \Phi(\boldsymbol{x}_j) \boldsymbol{e}_j^\top \in \mathcal{B}(\mathbb{R}^L, \mathcal{F}) \tag{20}$$

ここで，$\boldsymbol{e}_j$ は，第 j 要素が 1 でそれ以外が零の L 次元ベクトルを表し，$\mathcal{B}(\mathbb{R}^L, \mathcal{F})$ は，$\mathbb{R}^L$ から $\mathcal{F}$ への線形作用素全体を表す．S を用いると R^Φ は，

$$R^\Phi = \frac{1}{L} S S^\top \tag{21}$$

と表される．先ほどと同様に $\mathcal{F}$ が無限次元関数空間となる場合は，転置を随伴作用素 S^* に置き換えればよい．S の特異値分解を

$$S = \sum_{k=1}^{L'} \lambda_k \boldsymbol{u}_k \boldsymbol{v}_k^\top = U \Lambda V^\top \tag{22}$$

とおく．ここで，L' は S のランクであり，$U = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{L'}]$, $V = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{L'}]$ である．$SS^\top = U\Lambda^2 U^\top$ より，R^Φ の固有ベクトルは $\boldsymbol{u}_i$, $i = 1, \ldots, L'$ である．式 (22) の両辺の右から $V\Lambda^{-1}$ を掛けると

$$U \;=\; SV\Lambda^{-1} \tag{23}$$

$$\boldsymbol{u}_k \;=\; \frac{1}{\lambda_k} S \boldsymbol{v}_k \tag{24}$$

の関係が得られる．$\boldsymbol{v}_k$ は，カーネル Gram 行列 $S^\top S \in \mathbb{R}^{L \times L}$ の固有ベクトルであるため，標本数の大きさの行列を固有値分解すれば式 (24) の関係から，R^Φ の固有ベクトルが求められる．

カーネル Gram 行列の i,j 要素は $k(\boldsymbol{x}_i,\boldsymbol{x}_j)$ で計算ができる．

類似度関数は，

$$g(\boldsymbol{x}) = \|P\Phi(\boldsymbol{x})\|^2 \tag{25}$$

$$P = \sum_{k=1}^{r} \boldsymbol{u}_k\boldsymbol{u}_k^\top \tag{26}$$

であるため，式 (24) を代入すると

$$g(\boldsymbol{x}) = \langle \boldsymbol{h}(\boldsymbol{x}), \left(\sum_{k=1}^{r}\frac{1}{\lambda_k^2}\boldsymbol{v}_k\boldsymbol{v}_k^\top\right)\boldsymbol{h}(\boldsymbol{x})\rangle \tag{27}$$

$$= \sum_{k=1}^{r}\langle \boldsymbol{h}(\boldsymbol{x}), \frac{1}{\lambda_k}\boldsymbol{v}_k\rangle^2 \tag{28}$$

で与えられる．

ここで，$\boldsymbol{h}(\boldsymbol{x}) = S^\top\Phi(\boldsymbol{x}) = [k(\boldsymbol{x}_1,\boldsymbol{x}),\ldots,k(\boldsymbol{x}_L,\boldsymbol{x})]^\top \in \mathbb{R}^L$ は経験カーネル写像と呼ばれるベクトルである [6]．

射影作用素 P は，以下の最適化問題で特徴付けをすることができる．

$$\begin{array}{ll} \min\limits_{Y\in\mathcal{B}(\mathcal{F})} & \dfrac{1}{L}\sum\limits_{j=1}^{L}\|\Phi(\boldsymbol{x}_j) - Y\Phi(\boldsymbol{x}_j)\|^2 \\ \text{subject to} & \mathrm{rank}(Y) \leqq r,\ \mathcal{N}(Y) \supset \mathcal{R}(S)^\perp \end{array} \tag{29}$$

ここで，$\mathcal{N}(\cdot)$ は核空間，$\mathcal{R}(\cdot)$ は値域を表す．

3.2 抑制付きカーネル部分空間法

抑制付きカーネル部分空間法は，最適化問題 (7) に非線形写像 $\Phi(\cdot)$ を適用した以下の最適化問題

$$\begin{array}{ll} \min\limits_{Y\in\mathcal{B}(\mathcal{F})} & \dfrac{1}{L}\sum\limits_{j=1}^{L}\|\Phi(\boldsymbol{x}_j) - Y\Phi(\boldsymbol{x}_j)\|^2 \\ & -\beta\dfrac{1}{M}\sum\limits_{j=1}^{M}\|\Phi(\boldsymbol{y}_j) - Y\Phi(\boldsymbol{y}_j)\|^2 \\ & +\mu\|Y\|_F^2 \\ \text{subject to} & \mathrm{rank}(Y) \leqq r, \mathcal{N}(Y) \supset \mathcal{R}(T)^\perp \end{array} \tag{30}$$

の解で与えられる．ここで，

$$T = [\Phi(\boldsymbol{x}_1),\ldots,\Phi(\boldsymbol{x}_L)\Phi(\boldsymbol{y}_1),\ldots,\Phi(\boldsymbol{y}_M)]$$

$$= \sum_{j=1}^{L}\Phi(\boldsymbol{x}_j)\boldsymbol{e}_j^\top + \sum_{j=1+L}^{L+M}\Phi(\boldsymbol{y}_{j-L})\boldsymbol{e}_j^\top, \tag{31}$$

$\boldsymbol{e}_j \in \mathbb{R}^{L+M}$ である．

これをカーネル部分空間法と同様に計算すると，高次元特徴空間の主問題は，対象クラスの標本数 L と競合クラスの標本数 M の重複を省いた和，N 次元の双対問題となる．前述の通り，クラス数が非常に大きい (数十〜数百) 場合は，L が現実的な大きさ ($\sim 10,000$) でも N がその数十倍から数百倍となり，計算ができない．このため，解の範囲を Y の値域の直交補空間 $\mathcal{R}(Y)$ が対象クラスの標本が高次元特徴空間 $\mathcal{F}$ で張る空間 $\mathcal{R}(S)$ の直交補空間を含むような作用素であり，かつ，Y の値域が $\mathcal{R}(S)$ に含まれるような作用素に限り，最適化問題を解く．

$$\begin{array}{ll} \min\limits_{Y\in\mathcal{B}(\mathcal{F})} & \dfrac{1}{L}\sum\limits_{j=1}^{L}\|\Phi(\boldsymbol{x}_j) - Y\Phi(\boldsymbol{x}_j)\|^2 \\ & -\beta\dfrac{1}{M}\sum\limits_{j=1}^{M}\|\Phi(\boldsymbol{y}_j) - Y\Phi(\boldsymbol{y}_j)\|^2 \\ & +\mu\|Y\|_F^2 \\ \text{subject to} & \mathrm{rank}(Y) \leqq r,\ \mathcal{N}(Y) \supset \mathcal{R}(S)^\perp,\ \mathcal{R}(Y) \subset \mathcal{R}(S) \end{array} \tag{32}$$

ここで，S は式 (20) で定義される対象クラスのみの標本を並べた作用素である．抑制の目的は，Y が抽出する対象クラスの特徴の中に対象クラス以外にも含まれる特徴があるとき，その特徴を抑制することであるため，解の空間を限っても抑制の効果が期待できる．また，非線形写像の次元が小さい場合や，原空間と高次元特徴空間の直積空間への写像

$$\Phi' : \mathbb{R}^d \to \mathcal{F}\times\mathbb{R}^d,\ \boldsymbol{x} \mapsto [\Phi(x)^\top\boldsymbol{x}^\top]^\top$$

$$k'(\boldsymbol{x}_1,\boldsymbol{x}_2) = \langle\Phi'(\boldsymbol{x}_1),\Phi'(\boldsymbol{x}_2)\rangle = k(\boldsymbol{x}_1,\boldsymbol{x}_2) + \langle\boldsymbol{x}_1,\boldsymbol{x}_2\rangle$$

を用いれば，対象クラスの標本が張る空間と標本全体が張る空間の重なりが大きくなり，抑制の効果が大きくなると考えられる．

最適化問題 (32) の解を求める．

［補題 1］ 制約条件 $\mathcal{N}(Y) \subset \mathcal{R}(S)^\perp$ は，Y がある作用素 $A \in \mathcal{B}(\mathbb{R}^L,\mathcal{F})$ を用いて，$Y = AS^\top$ と表されることと同値である．

(証明) (i)『任意の作用素 $A \in \mathcal{B}(\mathbb{R}^L,\mathcal{F})$, $Y \in \mathcal{B}(\mathcal{F})$, $S \in \mathcal{B}(\mathbb{R}^L,\mathcal{F})$ に対して，$Y = AS^\top$ ならば $\mathcal{N}(Y) \supset \mathcal{R}(S)^\perp$ である．』は，$\mathcal{R}(S)^\perp = \mathcal{N}(S^\top)$ より，自明である．

(ii)『任意の作用素 $Y \in \mathcal{B}(\mathcal{F})$, $S \in \mathcal{B}(\mathbb{R}^L,\mathcal{F})$ に対して，$\mathcal{N}(Y) \supset \mathcal{R}(S)^\perp$ ならば，$Y = AS^\top$ を満たす作用素 $A \in \mathcal{B}(\mathbb{R}^L,\mathcal{F})$ が存在する．』を証明する．X の Moore-Penrose 一般逆作用素を $X^\dagger$ で表す．$A = Y(S^\top)^\dagger$ と置けば $AS^\top = Y(S^\top)^\dagger S^\top = YP_{\mathcal{R}(S)} = Y(I - P_{\mathcal{R}(S)^\perp})$ である．ここで，$P_{\mathcal{R}(S)}$, $P_{\mathcal{R}(S)^\perp}$ はそれぞれ，$\mathcal{R}(S)$, $\mathcal{R}(S)^\perp$ への正射影を表す．$\mathcal{N}(Y) \supset \mathcal{R}(S)^\perp$ より，$YP_{\mathcal{R}(S)^\perp} = 0$ であるため，$AS^\top = Y$ である．

(i), (ii) より，補題 1 が証明できる．(証明終)

［補題 2］ 制約条件 $\mathcal{R}(Y) \subset \mathcal{R}(S)$ は，Y がある作用素 $B \in \mathcal{B}(\mathcal{F},\mathbb{R}^L)$ を用いて，$Y = SB$ と表されることと同値である．

証明は補題 1 とほぼ同様であるため，省略する．

補題 1, 2 より，Y は，行列 $X \in \mathbb{R}^{L\times L}$ を用いて $Y = SXS^\top$ と表すことができる．最適化問題 (32) の目的関数を J と置くと

$$
\begin{aligned}
J \ = \ & \frac{1}{L}\sum_{j=1}^{L}\Big(k(\boldsymbol{x}_j,\boldsymbol{x}_j) - \langle S^\top\Phi(\boldsymbol{x}_j), XS^\top\Phi(\boldsymbol{x}_j)\rangle \\
& -\langle XS^\top\Phi(\boldsymbol{x}_j), S^\top\Phi(\boldsymbol{x}_j)\rangle \\
& +\langle XS^\top\Phi(\boldsymbol{x}_j), S^\top SXS^\top\Phi(\boldsymbol{x}_j)\rangle\Big) \\
& \frac{\beta}{M}\sum_{j=1}^{M}\Big(k(\boldsymbol{x}_j,\boldsymbol{x}_j) - \langle S^\top\Phi(\boldsymbol{y}_j), XS^\top\Phi(\boldsymbol{y}_j)\rangle \\
& -\langle XS^\top\Phi(\boldsymbol{y}_j), S^\top\Phi(\boldsymbol{y}_j)\rangle \\
& +\langle XS^\top\Phi(\boldsymbol{y}_j), S^\top SXS^\top\Phi(\boldsymbol{y}_j)\rangle\Big) \\
& +\mu\|SXS^\top\|_F^2
\end{aligned}
$$

$K_x = S^\top S \in \mathbb{R}^{L\times L}$, $S_y = [\Phi(\boldsymbol{y}_1),\ldots,\Phi(\boldsymbol{y}_M)] \in \mathcal{B}(\mathbb{R}^L,\mathcal{F})$, $K_y = S^\top S_y \in \mathbb{R}^{L\times M}$ とおくと

$$
\begin{aligned}
J \ = \ & \frac{1}{L}\mathrm{Trace}[K_x - K_xXK_xK_xX^\top K_x - K_xXK_xK_xX^\top K_x \\
& +K_x^\top X^\top K_xXK_x] - \frac{\beta}{M}\mathrm{Trace}[S_y^\top S_y - K_y^\top XK_y \\
& -K_y^\top X^\top K_y + K_y^\top X^\top K_xXK_y] + \mathrm{Trace}[XK_xX^\top K_x^\top] \\
= \ & \mathrm{Trace}[K_x^{\frac{1}{2}}X(\frac{1}{L}K_xK_x^\top - \frac{\beta}{M}K_yK_y^\top + \mu K_x)X^\top K_x^{\frac{1}{2}} \\
& -X(\frac{1}{L}K_xK_x^\top - \frac{\beta}{M}K_yK_y^\top) \\
& -(\frac{1}{L}K_xK_x^\top - \frac{\beta}{M}K_yK_y^\top)X^\top + \frac{1}{L}K_x - \frac{\beta}{M}S_y^\top S_y]
\end{aligned} \tag{33}
$$

ここで，$K_{xy} = \frac{1}{L}K_xK_x^\top - \frac{\beta}{M}K_yK_y^\top$ とおく．

a） ランク制約がない場合，すなわち $r = L$ である場合

J の X における Z 方向の Gâteaux 微分 $\delta J(X;Z)$ は

$$
\begin{aligned}
\delta J(X;Z) \ = \ & \lim_{\delta\to 0}\frac{1}{\delta}\mathrm{Trace}[K_x^{\frac{1}{2}}\delta Z(K_{xy}+\mu K_x)X^\top K_x^{\frac{1}{2}} \\
& +K_x^{\frac{1}{2}}X(K_{xy}+\mu K_x)\delta Z^\top K_x^{\frac{1}{2}} - \delta ZK_{xy} \\
& -K_{xy}\delta Z^\top] \\
= \ & \lim_{\delta\to 0}\frac{2}{\delta}\mathrm{Trace}[\big(K_xX(K_{xy}+\mu K_x) - K_{xy}\big)\delta Z^\top]
\end{aligned}
$$

任意の Z に対して，$\delta J(X;Z) = 0$ となるのは，

$$
K_xX(K_{xy}+\mu K_x) = K_{xy} \tag{34}
$$

が成り立つときである．$\mathcal{R}(K_x) \subset \mathcal{R}(K_y)$ より，$\mathcal{R}(K_x) = \mathcal{R}(K_y)$ である．また，$\mathcal{N}(K_{xy}+\mu K_x) \subset \mathcal{N}(K_{xy})$ であるため，式 (34) は解を持ち，その解の 1 つは

$$
X = K_x^\dagger K_{xy}(K_{xy}+\mu K_x)^\dagger \tag{35}
$$

で与えられる．特に，K_x, $K_{xy}+\mu K_x$ が正則となる場合 Moore-Penrose 一般逆行列は逆行列に置き換わる．

b） ランク制約がある場合

Mercer カーネルを用いた場合，同一の標本がない限り，K_x は正定値である．以降，$(K_{xy}+\mu K_x)$ が正定値となるように β, μ が選ばれているものとする．

式 (33) で J の X に関係のない項を除いたものを J' とすると

$$
\begin{aligned}
J' \ = \ & \mathrm{Trace}[\big(K_x^{\frac{1}{2}}X(K_{xy}+\mu K_x)^{\frac{1}{2}} \\
& -K_x^{-\frac{1}{2}}K_{xy}(K_{xy}+\mu K_x)^{-\frac{1}{2}}\big) \\
& \big((K_{xy}+\mu K_x)^{\frac{1}{2}}X^\top K_x^{\frac{1}{2}} - (K_{xy}+\mu K_x)^{-\frac{1}{2}}K_{xy}K_x^{-\frac{1}{2}}\big)] \\
= \ & \|K_x^{\frac{1}{2}}X(K_{xy}+\mu K_x)^{\frac{1}{2}} - K_x^{-\frac{1}{2}}K_{xy}(K_{xy}+\mu K_x)^{-\frac{1}{2}}\|_F^2
\end{aligned} \tag{36}
$$

式 (36) より，問題は行列の最良近似問題となる．ここで，$K_x^{-\frac{1}{2}}K_{xy}(K_{xy}+\mu K_x)^{-\frac{1}{2}}$ の特異値分解を

$$
K_x^{-\frac{1}{2}}K_{xy}(K_{xy}+\mu K_x)^{-\frac{1}{2}} = \sum_{i=1}^{L}\lambda_i\boldsymbol{u}_i\boldsymbol{v}_i^\top \tag{37}
$$

とおくと $K_x^{\frac{1}{2}}X(K_{xy}+\mu K_x)^{\frac{1}{2}}$ のランクは r 以下であるため

$$
K_x^{\frac{1}{2}}X(K_{xy}+\mu K_x)^{\frac{1}{2}} = \sum_{i=1}^{r}\lambda_i\boldsymbol{u}_i\boldsymbol{v}_i^\top \tag{38}
$$

を満たす X が存在するときは，上式を満たすとき J' が最小となる．

$$
\mathcal{R}(K_x^{\frac{1}{2}}) \supset \mathcal{R}\Big(\sum_{i=1}^{r}\lambda_i\boldsymbol{u}_i\boldsymbol{v}_i^\top\Big) \tag{39}
$$

$$
\mathcal{N}\Big((K_{xy}+\mu K_x)^{\frac{1}{2}}\Big) \subset \mathcal{N}\Big(\sum_{i=1}^{r}\lambda_i\boldsymbol{u}_i\boldsymbol{v}_i^\top\Big) \tag{40}
$$

が満たされるため，行列方程式 (38) は解を持ち，その解の 1 つは

$$
X = K_x^{-\frac{1}{2}}\Big(\sum_{i=1}^{r}\lambda_i\boldsymbol{u}_i\boldsymbol{v}_i^\top\Big)(K_{xy}+\mu K_x)^{-\frac{1}{2}} \tag{41}
$$

と与えられる．

特異値分解の性質を用いると

$$
\boldsymbol{v}_i \ = \ \frac{1}{\lambda_i}(K_{xy}+\mu K_x)^{-\frac{1}{2}}K_{xy}K_x^{-\frac{1}{2}}\boldsymbol{u}_i \tag{42}
$$

であり，これを に代入すると

$$
X = K_x^{-\frac{1}{2}}\Big(\sum_{i=1}^{r}\boldsymbol{u}_i\boldsymbol{u}_i^\top\Big)K_x^{-\frac{1}{2}}K_{xy}(K_{xy}+\mu K_x)^{-1} \tag{43}
$$

で与えられる．ここで，$\boldsymbol{u}_i$ は，$K_x^{-\frac{1}{2}}K_{xy}(K_{xy}+\mu K_x)^{-1}K_{xy}K_x^{-\frac{1}{2}}$ の降順に並べた固有値に対応する固有ベクトルである．

［系 1］ $r = L$ かつ，K_x, $K_{xy}+\mu K_x$ が正則である場合，式 (35) と式 (b)) は等価である．

$Y = SXS^\top$ より，類似度関数は

$$
\begin{aligned}
g(\boldsymbol{x}) \ = \ & -\|\Phi(\boldsymbol{x}) - Y\Phi(\boldsymbol{x})\|^2 \\
= \ & k(\boldsymbol{x},\boldsymbol{x}) - 2\langle\boldsymbol{h}(\boldsymbol{x}), X\boldsymbol{h}(\boldsymbol{x})\rangle \\
& +\langle\boldsymbol{h}(\boldsymbol{x}), X^\top K_xX\boldsymbol{h}(\boldsymbol{x})\rangle
\end{aligned} \tag{44}
$$

となる．

X の計算は，$K_yK_y^\top$ の計算の $L\times M$ 行列の積の計算の他は，すべて L 次の正方行列の積，逆，固有値分解で計算ができる．

表 1　手書き数字識別実験結果

手法	パラメータ	誤識別率 ± 標準偏差 [%]
CLAFIC	$r = 26$	5.03 ± 0.10
SPCA	$r = 31$, $\mu = 10^{-2.8}$, $\beta = 10^{-3}$	4.88 ± 0.10
KPCA	$r = 155$	3.57 ± 0.12
SKPCA	$\beta = 10^{-1.8}$, $\mu = 8 \times 10^{-4}$, $r = 260$	3.51 ± 0.11

4. 実　験

手書き数字のデータベース MNIST を用いて性能の比較実験を行った．MNIST の学習用標本 60,000 サンプル中から無作為に 5,000 サンプルを学習用に抽出し，残りの 55,000 サンプルを検定用に使用した．このランダムサンプリングを 20 回行い，誤識別率の平均と標準偏差を比較した．実験では，CLAFIC 法，正則化抑制付き部分空間法 (SPCA, 式 (13))，カーネル部分空間法 (KPCA)，抑制付きカーネル部分空間法 (SKPCA) の４つの手法を比較した．カーネル関数は，式 (18) の Gaussian カーネル，$c = 1$ を用いた．競合クラスの標本 $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_M$ には対象クラス以外のすべての標本を用いた．表 1 に結果を示す．パラメータはいくつかの候補から誤識別率が最小となるものを選んだ．KPCA と提案法の SKPCA では，Student の片側 t 検定において，5.2%の有意水準で提案法の優位性が示された．

5. ま と め

抑制付きカーネル部分空間法に対して，解の存在範囲を限定することで，対象クラスの標本数の大きさの行列演算で設計ができる手法を提案した．提案法は，行列計算が対象クラスの標本数の大きさであるためクラス数が非常に大きな場合でも現実的に設計を行うことができるという利点がある．手書き数字識別実験結果は，提案法が通常のカーネル部分空間法よりも低い誤識別率を示した．これは解の空間を限っても，抑制の効果があるためであると考えられる．

今後の課題として，パラメータ選択に関する議論，棄却やクラス数の増減への対応などが挙げられる．

謝　辞

本研究は，日本学術振興会科学研究費補助金 No. 19700153(若手 B) の補助を受けた．

文　献

[1] 飯島泰蔵：“パターン認識理論”, 森北出版 (1989).
[2] 鷲沢, 疋田, 田中, 山下：“カーネル相対主成分分析による多クラスパターン認識”, 第二回 FIT（情報科学技術フォーラム）情報レターズ, pp. 207–208 (2003).
[3] 池野, 山下, 小川：“相対 KL 変換法によるパターン認識”, 信学論 (D-II), **J80-D-II**, 2, pp. 541–547 (1997).
[4] 前田, 村瀬：“カーネル非線形部分空間法によるパターン認識”, 信学論 (D-II), **J82-D-II**, 4, pp. 600–612 (1999).
[5] 津田：“ヒルベルト空間における部分空間法”, 信学論 (D-II), **J82-D-II**, 4, pp. 592–599 (1999).
[6] B. Schölkopf, S. Mika, C. Burges, P. Knirsch, K.-R. Müller, G. Rätsch and A. Smola: “Input space vs. feature space in kernel-based methods”, IEEE Transactions on Neural Networks, **10**, 5, pp. 1000–1017 (1999).
[7] Y. Washizawa, K. Hikida, T. Tanaka and Y. Yamashita: “Kernel relative principal component analysis for pattern recognition”, Proc. of Joint IAPR Iinternational Workshops on Syntactical and Structural Pattern Recognition and Statistical Pattern Recognition (SSPR/SPR 2004), pp. 1105–1113 (2004).
[8] Y. Washizawa and Y. Yamashita: “Kernel projection classifiers with suppressing features of other classes”, Neural Computation, **18**, 8, pp. 1932–1950 (2006).
[9] S. Watanabe and N. Pakvasa: “Subspace method in pattern recognition”, Proc. 1st Int. J. Conf on Pattern Recognition, Washington DC, pp. 25–32 (1973).
[10] E. Oja, 小川, 佐藤：“パターン認識と部分空間法”, 産業図書 (1986).
[11] 鷲沢：“正則化を用いた２次識別器 -部分空間法との比較-”, 第 10 回　画像の認識・理解シンポジウム予稿集, pp. 510–515 (2007).
[12] Y. Washizawa: “Regularization vs. rank reduction in quadratic classifiers”, Proc. of ACCV 2007 workshop subspace 2007, pp. 108–115 (2007).