

共分散記述子のための計量学習と画像分類への応用

松澤 知己^{1,a)} レラトー レイサ² 加藤 毅¹

概要：近年，コンピュータビジョンの分野において，共分散記述子が強力な画像の表現として認識されるようになってきた．本研究では，共分散記述子による画像分類性能の向上をはかるため，訓練用データから正定値錐上の距離を学習する計量学習アルゴリズムを開発した．数値実験により学習の収束性を確認し，実データを使って，計量学習の効果を示す．

キーワード：共分散記述子，正定値錐，ダイクストラ法．

1. はじめに

近年，コンピュータビジョンの分野において，共分散記述子が強力な画像の表現として認識されるようになってきた．たとえば，拡散テンソルMRI [1], [2] では，組織内の異方性をパラメータに持つ拡散テンソルで組織における水の拡散を表現している．オプティカルフロー推定や動作切出しには，テクスチャや動作のような重要な画像の特徴を表現するために構造テンソルが良く用いられている [3]．共分散領域記述子は，テクスチャ分類 [4]，通行人検出 [5]，物体追跡 [6]，動作認識 [7]，顔認識 [8]，に応用されている．共分散記述子は正定値行列であることから，共分散記述子を使った解析には，リーマン多様体上での距離関数が好まれており，これまで多くの距離関数が試されてきた．例えばアファイン不変距離 [1] や，対数ユークリッド距離 [9]，コレスキュークリッド距離 [10]，Jダイバージェンス [11]，イエセンブレグマン LogDet ダイバージェンス [12] などがある．一方，通常のベクトル空間を使った画像分類では，データから距離計量を学習することにより画像分類性能が向上することが確かめられている [13], [14]．本研究では，共分散記述子による画像分類性能の向上をはかるため，訓練用データから正定値錐上の距離を学習するアルゴリズムを開発した．本稿では，そのアルゴリズムを提案する．

記法

- $n \times n$ の単位行列は \mathbf{I}_n で表す．ただし，下付き文字は省略する場合がある．
- $m \times n$ の零行列は $\mathbf{O}_{m \times n}$ で表す．ただし，下付き文字は省略する場合がある．
- 要素がすべて 1 である n 次元ベクトルは $\mathbf{1}_n$ のように表す．
- \mathbb{R}^n ， \mathbb{N}^n はそれぞれ n 次元の実数集合， n 次元の自然数集合を表し， $\mathbb{R}^{m \times n}$ は $m \times n$ の実行列の集合を表す．
- ある $n \in \mathbb{N}$ があつたとき， n 以下の自然数の集合を \mathbb{N}_n のように表す．
- \mathbb{R}_+^n はすべての要素が 0 以上である n 次元実ベクトルの集合を表す． \mathbb{R}_{++}^n はすべての要素が 0 より大きい n 次元実ベクトルの集合を表す．
- \mathbb{S}^n は $n \times n$ の対称行列の集合を表す． \mathbb{S}_+^n は $n \times n$ の半正定値対称行列集合を表し， \mathbb{S}_{++}^n は $n \times n$ の狭義正定値対称行列集合を表す．
- ある $\mathbf{X} \in \mathbb{R}^{n \times n}$ があつたとき， $\mathbf{X} \succ \mathbf{O}$ は \mathbf{X} が狭義正定値行列であることを示す．
- ある $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ があつたとき， $\langle \mathbf{x}, \mathbf{y} \rangle := \sum_{i=1}^n x_i y_i$ のように定義する．ただし， x_i と y_i はそれぞれ \mathbf{x} と \mathbf{y} の第 i 成分とする．
- ある $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{m \times n}$ があつたとき， $\langle \mathbf{X}, \mathbf{Y} \rangle := \sum_{i=1}^m \sum_{j=1}^n X_{i,j} Y_{i,j}$ のように定義する．ただし， $X_{i,j}$ と $Y_{i,j}$ はそれぞれ \mathbf{X} と \mathbf{Y} の第 (i, j) 成分とする．
- 行列 \mathbf{A} の転置行列は \mathbf{A}^\top ，逆行列は \mathbf{A}^{-1} のように表す．
- ある $n \times n$ の正方行列 \mathbf{X} があつたとき，トレースは $\text{tr}(\mathbf{X})$ のように表す．
- ある $\mathbf{x} = [x_1, \dots, x_n]^\top \in \mathbb{R}^n$ があつたとき， $\text{diag}(\mathbf{x})$ は

¹ 群馬大学理工学部，〒376-8515 群馬県桐生市天神町 1-5-1
School of Science and Technology, Gunma University
Tenjin-cho 1-5-1, Kiryu-shi, Gunma, 376-8515 Japan

² 産総研ゲノム情報センター，〒135-0064 東京都江東区青海
2-4-7 産総研 臨海副都心センター別館 (バイオ・IT 融合研究棟)
Computational Biology Research Center, AIST, AIST Tokyo
Waterfront Bio-IT Research Building 2-4-7 Aomi, Koto-ku,
Tokyo, 135-0064, Japan

^{a)} matsuzawa-tomoki@kato-lab.cs.gunma-u.ac.jp

対角成分に x_1, \dots, x_n をもつ $n \times n$ の対角行列を表す.

2. 正定値錐上のパラメトリックな距離計量関数

正定値行列間の距離を測るため、これまで様々な距離関数が検討されてきた [15]. たとえば、対数ユークリッド距離 (*Log-Euclidean distance*) は、2つの正定値行列 $\mathbf{X}_1, \mathbf{X}_2 \in \mathbb{S}_{++}^n$ に対して、

$$D_{\text{LE}}(\mathbf{X}_1, \mathbf{X}_2) := \|\log(\mathbf{X}_1) - \log(\mathbf{X}_2)\|_F$$

と定義される. ただし, \log は principal matrix logarithm 関数であり, 要素ごとに対数をとるものではないことに注意されたい. 冪ユークリッド距離 (*Power-Euclidean distance*) は

$$D_{\text{PE}}(\mathbf{X}_1, \mathbf{X}_2) := \|\mathbf{X}_1^p - \mathbf{X}_2^p\|_F$$

と定義される. ただし, $p \in \mathbb{R}$ は定数である. コレスキユークリッド距離 (*Cholesky Euclidean distance*) [10] は,

$$D_{\text{PE}}(\mathbf{X}_1, \mathbf{X}_2) := \|\text{chol}(\mathbf{X}_1) - \text{chol}(\mathbf{X}_2)\|_F$$

と定義される. ただし, $\text{chol}(\cdot)$ は, コレスキ分解を表し, 正定値行列 \mathbf{X} に対して, $\text{chol}(\mathbf{X})\text{chol}(\mathbf{X})^\top = \mathbf{X}$ なる下三角行列で定義される. ここで, 写像 $\Phi: \mathbb{S}_{++}^n \rightarrow \mathbb{R}^{n \times n}$ を導入すると, これら3種類の距離計量関数は

$$D_{\Phi}(\mathbf{X}_1, \mathbf{X}_2) := \|\Phi(\mathbf{X}_1) - \Phi(\mathbf{X}_2)\|_F^2.$$

の二乗根としてまとめて表すことができる. つまり, 対数ユークリッド距離は $\Phi(\mathbf{X}) := \log(\mathbf{X})$ と設定することで得られる. 冪ユークリッド距離は $\Phi(\mathbf{X}) := \mathbf{X}^p$ と設定することで得られる. コレスキユークリッド距離は $\Phi(\mathbf{X}) := \text{chol}(\mathbf{X})$ と設定することで得られる.

本稿では, 次のパラメトリックな距離計量を導入する:

$$D_{\Phi}(\mathbf{X}_1, \mathbf{X}_2; \mathbf{W}) := \left\langle \mathbf{W}, (\Phi(\mathbf{X}_1) - \Phi(\mathbf{X}_2))(\Phi(\mathbf{X}_1) - \Phi(\mathbf{X}_2))^\top \right\rangle.$$

ただし, $\mathbf{W} \in \mathbb{S}_{++}^n$ はこの距離関数のパラメータである. もし \mathbf{W} が狭義正定値 (strictly positive definite) で, Φ が全単射なら, この距離関数 $D_{\Phi}(\cdot, \cdot; \mathbf{W})$ は次であらわされる距離の公理に従う:

- 非負性 (non-negativity):

$$D_{\Phi}(\mathbf{X}_1, \mathbf{X}_2; \mathbf{W}) \geq 0.$$

- 確定性 (identity of indiscernibles):

$$D_{\Phi}(\mathbf{X}_1, \mathbf{X}_2; \mathbf{W}) = 0 \text{ iff } \mathbf{X}_1 = \mathbf{X}_2.$$

- 対称性 (symmetry):

$$D_{\Phi}(\mathbf{X}_1, \mathbf{X}_2; \mathbf{W}) = D_{\Phi}(\mathbf{X}_2, \mathbf{X}_1; \mathbf{W}).$$

- 三角不等式 (triangle inequality):

$$D_{\Phi}(\mathbf{X}_1, \mathbf{X}_3; \mathbf{W}) \leq D_{\Phi}(\mathbf{X}_1, \mathbf{X}_2; \mathbf{W}) + D_{\Phi}(\mathbf{X}_2, \mathbf{X}_3; \mathbf{W}).$$

もし \mathbf{W} が特異行列なら, $D_{\Phi}(\cdot, \cdot; \mathbf{W})$ は擬似距離, すなわち確定性の性質が次に変わる: いかなる $\mathbf{X}_1 \in \mathbb{S}_{++}^n$ に対しても

- $D_{\Phi}(\mathbf{X}_1, \mathbf{X}_1; \mathbf{W}) = 0$,

がいえるが, 異なる正定値行列 \mathbf{X}_1 および \mathbf{X}_2 に対しても, $D_{\Phi}(\mathbf{X}_1, \mathbf{X}_2; \mathbf{W}) = 0$ となることを許す.

3. 正定値行列 \mathbf{W} の学習問題の定式化

距離計量関数のパラメータ $\mathbf{W} \in \mathbb{S}_{++}^n$ は, 解析タスクに適応するように決定する. 基本的なアイデアは, ベクトル間のマハラノビス距離のパラメータを最適化するために考案された Davis らの戦略 [13] と同じである. 多クラス分類問題を例にとって述べる. 訓練用データが $(\mathbf{X}_1, \omega_1), \dots, (\mathbf{X}_\ell, \omega_\ell) \in \mathbb{S}_{++}^n \times \mathbb{N}_{n_c}$ のように得られていたとする. ただし, \mathbf{X}_i は第 i 訓練用例題の共分散記述子で, ω_i はそのクラスラベルである. n_c はクラス数である. これら ℓ 個の訓練用例題に対し, K 個のペアを $(i_1, j_1), \dots, (i_K, j_K) \in \mathbb{N}_\ell \times \mathbb{N}_\ell$ を考え, 各ペアに対して次の制約を与える:

$$D_{\Phi}(\mathbf{X}_{i_k}, \mathbf{X}_{j_k}; \mathbf{W}) \begin{cases} \leq b_{\text{ub}} & \text{if } \omega_{i_k} = \omega_{j_k}, \\ \geq b_{\text{lb}} & \text{if } \omega_{i_k} \neq \omega_{j_k}, \end{cases} \quad (1)$$

ただし, b_{ub} および b_{lb} は, それぞれ, 同じクラス間の距離の上限, 異なるクラス間の距離の下限を定める定数である. ここで, y_k と b_k を

$$(y_k, b_k) := \begin{cases} (+1, b_{\text{ub}}) & \text{if } \omega_{i_k} = \omega_{j_k}, \\ (-1, b_{\text{lb}}) & \text{if } \omega_{i_k} \neq \omega_{j_k}, \end{cases}$$

とおくと, (1) は

$$y_k D_{\Phi}(\mathbf{X}_{i_k}, \mathbf{X}_{j_k}; \mathbf{W}) \leq y_k b_k,$$

のように簡潔に表すことができる. これら K 個の制約を満たす中で, KL ダイバージェンスの意味で単位行列に最も近い正定値行列 \mathbf{W} の値を決定することを考える. すなわち, 次の最適化問題を解くこととする:

$$\begin{aligned} \min \quad & \text{KL}(\mathbf{W}, \mathbf{I}) \quad \text{wrt } \mathbf{W} \in \mathbb{S}_{++}^n \\ \text{subject to} \quad & \forall k \in \mathbb{N}_K, \quad y_k D_{\Phi}(\mathbf{X}_{i_k}, \mathbf{X}_{j_k}; \mathbf{W}) \leq y_k b_k. \end{aligned} \quad (2)$$

ただし, $\text{KL}(\mathbf{X}, \mathbf{Y})$ は共分散行列を \mathbf{X} および \mathbf{Y} とする2つの多変量正規分布間の KL ダイバージェンスである. すなわち, 任意の正定値対称行列 \mathbf{X} および \mathbf{Y} に対して,

$$\begin{aligned} \text{KL}(\mathbf{X}, \mathbf{Y}) &:= \int \mathcal{N}(\mathbf{x}; \mathbf{0}, \mathbf{Y}) \log \frac{\mathcal{N}(\mathbf{x}; \mathbf{0}, \mathbf{Y})}{\mathcal{N}(\mathbf{x}; \mathbf{0}, \mathbf{X})} d\mathbf{x} \\ &= \frac{1}{2} (\text{tr}(\mathbf{Y}^{-1}\mathbf{X}) - n + \log \det(\mathbf{Y}) - \log \det(\mathbf{X})). \end{aligned}$$

と定義することが出来る．ところで，訓練用データセットによっては，(2) の制約を満たす \mathbf{W} はひとつも存在せず，実行不可能 (infeasible) になることもある．一方，ソフトマージンサポートベクトルマシンでは線形分離不可能にならないようにスラック変数を導入して問題を緩和している．このようなテクニックは機械学習においてしばしば用いられている．本研究でもスラック変数 $\xi = [\xi_1, \dots, \xi_K]^\top$ を導入して，(2) を次の問題に緩和する．

$$\begin{aligned} \min \quad & \text{KL}(\mathbf{W}, \mathbf{I}) + \sum_{k=1}^K c_k \text{KL}(\xi_k, 1) \\ \text{wrt} \quad & \mathbf{W} \in \mathbb{S}_{++}^n, \quad \xi \in \mathbb{R}_{++}^K, \\ \text{subject to} \quad & \forall k \in \mathbb{N}_K, \quad y_k D_{\Phi}(\mathbf{X}_{i_k}, \mathbf{X}_{j_k}; \mathbf{W}) \leq y_k b_k \xi_k. \end{aligned} \quad (3)$$

距離 $D_{\Phi}(\mathbf{X}_{i_k}, \mathbf{X}_{j_k}; \mathbf{W})$ をベクトル間のマハラノビス距離に置き換えると，Davis らの ITML [13] になる．文献 [16] でも同様な定式化をしているが，彼らの制約には b_k も y_k もなく， $\xi_{0,k} = y_k b_k$ のように定めてしまっている．しかし， $\xi_{0,k} > 0$ である必要があるので，[16] の変数の置き方には誤りがある．

4. 最適化算法

4.1 ブレグマン射影問題への帰着

本研究でも ITML [13] と同様，ダイクストラ法 (Dykstra algorithm) [17], [18] を使って，最適化問題 (3) を解く．ダイクストラ法は，複数の凸集合の共通集合へのユークリッド射影を求める方法として 1983 年に提案されたが [17]，その後，Censor & Reich [18] によって，ブレグマン射影 (Bregman projection) を行えるように拡張された．

本研究の学習問題 (3) も複数の凸集合の共通集合へのブレグマン射影になっていることを示す．ブレグマンダイバージェンスは，連続微分可能で狭義凸な (strictly convex) 実数値関数 φ から定義される関数 $D_{\varphi} : \text{dom}(\varphi) \times \text{ri}(\text{dom}(\varphi)) \rightarrow \mathbb{R}_+$ で

$$D_{\varphi}(\mathbf{x}, \mathbf{y}) = \varphi(\mathbf{x}) - \varphi(\mathbf{y}) - \langle \nabla \varphi(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle$$

と定義される．ただし， $\text{dom}(\varphi)$ は φ の定義域， $\text{ri}(\text{dom}(\varphi))$ は $\text{dom}(\varphi)$ の相対内部 (relative interior) を示す．

ここで， φ を $\mathbb{S}_{++}^n \times \mathbb{R}_{++}^K$ から \mathbb{R} への写像として，

$$\varphi(\mathbf{W}, \xi) := -\log \det(\mathbf{W}) - \sum_{k=1}^K c_k \log(\xi_k)$$

とおくと，そのブレグマンダイバージェンスは

$$D_{\varphi}((\mathbf{W}, \xi), (\mathbf{I}, \mathbf{1})) = \text{KL}(\mathbf{W}, \mathbf{I}) + \sum_{k=1}^K c_k \text{KL}(\xi_k, 1)$$

を満たすこと確かめられる．また (3) の制約から，凸集合

$$C_k := \{(\mathbf{W}, \xi) \in \mathbb{S}_{++}^n \times \mathbb{R}_{++}^K \mid y_k \langle \mathbf{A}_k, \mathbf{W} \rangle - y_k b_k \xi_k \leq 0\}$$

を作ると，その共通集合は，(3) の実行可能領域になる．ただし，

$$\mathbf{A}_k := (\Phi(\mathbf{X}_1) - \Phi(\mathbf{X}_2))(\Phi(\mathbf{X}_1) - \Phi(\mathbf{X}_2))^\top$$

である．よって，問題 (3) はダイクストラ法の対象とするブレグマン射影の問題となっていることが示された．

以降の議論は $\mathbf{A}_k \succ \mathbf{O}$ を仮定する．これは \mathbf{A}_k の対角成分に小さな正の数を加えることで簡単に実現できる．

4.2 ダイクストラ法の適用

問題 (3) にダイクストラ法を適用すると Algorithm 1 のようになる．このアルゴリズムは逐次的に 1 つの半空間へ

Algorithm 1 Dykstra algorithm for metric learning.

```

1: begin
2:  $\forall k \in \mathbb{N}_K : \alpha_k^0 := 0;$ 
3:  $\mathbf{W}_{0,K} := \mathbf{W}_0;$ 
4: for  $t = 1, 2, \dots$  do
5:    $\mathbf{W}_{t,0} := \mathbf{W}_{t-1,K};$ 
6:   for  $k = 1, \dots, K$  do
7:     Find  $\mathbf{Y} \in \mathbb{S}^n, \xi \in \mathbb{R}$  and  $\delta_k^{t-1/2} \in \mathbb{R}_+$  s.t. (4) is hold.
8:      $\delta_k^t := \max(\delta_k^{t-1/2}, -\alpha_k^{t-1}); \alpha_k^t := \alpha_k^{t-1} + \delta_k^t;$ 
9:      $\mathbf{W}_{t,k}^{-1} = \mathbf{W}_{t,k-1}^{-1} + \delta_k^t y_k \mathbf{A}_k;$  and  $\xi_{t,k}^{-1} = \xi_{t,k-1}^{-1} - \Delta \alpha_k^t y_k b_k / c_k;$ 
10:   end for
11: end for
12: end.

```

のブレグマン射影を行うものになっている．ブレグマン射影は次の連立方程式を同時に満たす $\mathbf{Y} \in \mathbb{S}^n, \xi \in \mathbb{R}$ および $\delta_k^{t-1/2} \in \mathbb{R}_+$ を見つけることで行うことができる：

$$\begin{aligned} \langle \mathbf{A}_k, \mathbf{Y} \rangle &= b_k \xi, \\ \mathbf{Y}^{-1} &= \mathbf{W}_{t,k-1}^{-1} + \delta_k^{t-1/2} y_k \mathbf{A}_k \succ \mathbf{O}, \\ \xi^{-1} &= \xi_{t,k-1}^{-1} - \Delta \alpha_k^{t-1/2} y_k b_k / c_k > 0. \end{aligned} \quad (4)$$

すなわち， $\mathbf{Y} \succ \mathbf{O}$ および $\xi > 0$ を保ったまま，1 変数関数

$$J_{t,k}(\delta) := \left\langle \mathbf{A}_k, (\mathbf{W}_{t,k-1}^{-1} + \delta y_k \mathbf{A}_k)^{-1} \right\rangle - \frac{b_k}{\xi_{t,k-1}^{-1} - \alpha_k^t y_k b_k / c_k} \quad (5)$$

がゼロになる δ を探す問題に帰着された．1 変数方程式の根を求めるのは容易であり，ニュートンラフソン法 (Newton-Raphson method) などの数値算法を使うことができる．しかし， $J_{t,k}(\delta)$ には $n \times n$ 行列の逆行列を含んでいるため，ナイーブに数値算法にかけてしまうと，計算負荷の高い $n \times n$ 行列の逆行列を取束するまで計算し続けることになる．すると，数値算法内部での反復回数を L とすると，1 回の半空間へのブレグマン射影に $O(Ln^3)$ かかることになる．[16] では，連立方程式 (4) が逆行列やスペクトラル分解

なしで解けるかのように書かれているが、これはウッドベリ-行列公式の誤用である。一方、本研究では、この射影を $O(n^3 + Ln)$ の計算量でできるアルゴリズムを発見した。さらに、連立方程式 (4) を満たす δ はユニークに存在することも分かった。以下に、そのアルゴリズムと証明を与える。

変数 $\mathbf{A}_k^{1/2}$, \mathbf{U} , \mathbf{D} , \mathbf{d} , を次のように定義する。 $\mathbf{A}_k^{1/2} \mathbf{A}_k^{1/2} = \mathbf{A}_k$ なる正定値行列 $\mathbf{A}_k^{1/2} \in \mathbb{S}_{++}^n$ は必ず存在し、 \mathbf{A}_k のスペクトラル分解から容易に求められる。その逆行列を $\mathbf{A}_k^{-1/2} \in \mathbb{S}_{++}^n$ と記す。 $\mathbf{A}_k^{-1/2} \mathbf{W}^{-1} \mathbf{A}_k^{-1/2}$ のスペクトラル分解を $\mathbf{UDU}^\top = \mathbf{A}_k^{-1/2} \mathbf{W}^{-1} \mathbf{A}_k^{-1/2}$ とする。ただし、 \mathbf{U} は正規直交行列、 $\mathbf{D} = \text{diag}\{d_1, \dots, d_n\}$ は $d_1 \geq \dots \geq d_n$ なる対角行列とする。これらを使うと、 $J_{t,k}(\delta)$ の第1項目は

$$\langle \mathbf{A}_k, (\mathbf{W}^{-1} + \delta y_k \mathbf{A}_k)^{-1} \rangle = \sum_{i=1}^n \frac{1}{d_i + y_k \delta} \quad (6)$$

と表すことができる。すなわち、式のまま計算すると $O(n^3)$ かかる計算が、 \mathbf{A}_k および $\mathbf{A}_k^{-1/2} \mathbf{W}^{-1} \mathbf{A}_k^{-1/2}$ のスペクトラル分解を予め求めておけば、 $O(n)$ で計算できることになる。 $\mathbf{A}_k^{-1/2}$ および $\mathbf{A}_k^{-1/2} \mathbf{W}^{-1} \mathbf{A}_k^{-1/2}$ のスペクトラル分解は $O(n^3)$ の計算量を伴うが、 $\mathbf{A}_k^{-1/2}$ はダイクストラ法を行う前に予め求めておけばよく、 $\mathbf{A}_k^{-1/2} \mathbf{W}^{-1} \mathbf{A}_k^{-1/2}$ のスペクトラル分解は1変数方程式 $J_{t,k}(\delta) = 0$ を解く前に求めておけばよい。2つの制約 $\xi > 0$ および $\mathbf{Y} \succ \mathbf{O}$ を満たす δ の範囲は、

$$I_{t,k} := \begin{cases} (-d_n, \frac{c_k}{b_k \xi_{t,k-1}}) & \text{for } y_k = +1, \\ (-\frac{c_k}{b_k \xi_{t,k-1}}, d_n) & \text{for } y_k = -1. \end{cases} \quad (7)$$

であることを (6) および (5) から示すことができる。

解のユニーク性は次のように示すことができる。 $\delta_{\text{inf}} := \inf I_{t,k}$ および $\delta_{\text{sup}} := \sup I_{t,k}$ とおくと、 $y_k = +1$ の場合

$$\lim_{\delta \rightarrow \delta_{\text{inf}}+0} J_{t,k}(\delta) = +\infty, \quad \lim_{\delta \rightarrow \delta_{\text{sup}}-0} J_{t,k}(\delta) = -\infty,$$

かつ、 $\forall \delta \in I_{t,k}$, $\nabla J_{t,k}(\delta) < 0$ が成り立つ。 $y_k = -1$ の場合

$$\lim_{\delta \rightarrow \delta_{\text{inf}}+0} J_{t,k}(\delta) = -\infty, \quad \lim_{\delta \rightarrow \delta_{\text{sup}}-0} J_{t,k}(\delta) = +\infty,$$

かつ、 $\forall \delta \in I_{t,k}$, $\nabla J_{t,k}(\delta) > 0$ が成り立つ。よって、

$$\exists! \delta \in I_{t,k} \quad \text{s.t.} \quad J_{t,k}(\delta) = 0$$

が示された。

以上の議論を次の命題にまとめる。

Proposition 4.1. 非線形連立方程式 (4) の解はユニークに存在する。連立方程式を解く数値計算法の内部で L 回反復が行われるとすると、(4) の解は $O(n^3 + Ln)$ で見つけることができる。

一つの半空間へのブレグマン射影以外のステップはいずれの $O(n^2)$ で済む。 $L < n^2$ とすると、Algorithm 1 は1反復 $O(n^3)$ で計算できることになる。

4.3 乱択戦略

近年、最適化の手法として、確率的勾配法 [19], [20] が注目を集めているが、これらの文献で、例題をランダムに選ぶことが、規則的に選ぶより早く最適解に収束することが実験的に確かめられている [21]。ダイクストラ法の場合、半空間を周期的に (cyclic) 選んでブレグマン射影を行うものであるが、この半空間をランダムに選ぶこともできる。もしくは、1エポック (epoch) ごとに半空間の順番を無作為に入れ替えることもできる。本研究では、次節で、

- **Cyclic:** 周期的に半空間を選ぶ、
- **Perm:** エポックごとに半空間の順番を無作為に入れ替える、

• **Rand:** 毎回、半空間を無作為に選ぶ、
の3つの戦略を実験的に比較する。

5. 数値実験と画像分類への応用

本節では、提案した学習アルゴリズムの収束の速さを検証するため、および、画像分類性能を確認するため、数値実験を行った。

5.1 学習アルゴリズムの収束性の検証

学習アルゴリズムの収束性の評価するため、人工データセットを生成した。 $\mathbf{F}_k \in \mathbb{R}^{n \times n}$ の各要素を $[-0.5, +0.5]$ の範囲で一様分布で生成し、 $\mathbf{A}_k := \mathbf{F}_k \mathbf{F}_k^\top$ とした。 y_k は $+1$ と -1 をそれぞれ 50% の確率で無作為に決めた。 $\mathbf{b} = \mathbf{1}$ とし、 $K = 50$ とした。 $\mathbf{c} = \frac{1}{\lambda K}$ のように定め、 $\lambda = 10^{-2}, 10^{-3}, 10^{-4}$, $n = 10, 50, 100$ とし、Cyclic, Perm, Rand を比較した。

図1に各方法における収束の様子をプロットした。横軸はエポックを表す。 K 回半空間にブレグマン射影すると1エポックになる。縦軸は、エポックごとの目的関数の値と最適解との差を表す。この図より、Perm, Rand では λ の値にさほど影響を受けずに最適解に近づくことが見て取れる。一方Cyclicでは λ の値が小さくなるにつれて、つまり \mathbf{c} の値が大きくなるにつれて目的関数の値が最適解への収束が悪かった。

今度は、以上の実験を100回繰り返し、最適解に収束するまでのエポック数を測定した。双対ギャップが 10^{-5} 未満になったとき最適解に収束したとみなしている。ただし、最大エポック数は500回とし、その時点で双対ギャップが 10^{-5} に達していない場合でも反復を停止させた。すると、各条件で100個のエポック数が得られる。これを箱ひげ図にしたのが、図2である。この図からも、 λ の値が小さくなるにつれて、CyclicはPerm, Randに比べて収束が遅くな

ることが多かった。

5.2 画像分類への応用.

計量学習の効果を確認するために、2つの画像分類実験を行った。

テクスチャ認識. Brodatz texture dataset[22] はそれぞれ異なる 111 枚のテクスチャ画像で構成されている。すべての画像サイズは 640×640 で、またグレースケール化されている。我々は各画像を 4 つの領域に等分し、2 つの領域を訓練に、残りの領域を評価に用いた。

共分散記述子を計算するために、画像の各ピクセルから $[I_{xy}, |I_x|, |I_y|, |I_{xx}|, |I_{yy}|]$ のような特徴ベクトルを抽出した。ここで x, y はピクセルの座標であり、 I は画像自体を、 $|I_x|, |I_y|, \dots$ は勾配を表す。訓練用、テスト用の各領域から 128×128 サイズのパッチをランダムに 50 個取得し、共分散記述子を計算した。つまり、訓練用、評価用にそれぞれ $11,100 (= 111 \times 2 \times 50)$ 個の共分散記述子が得られたことになる。評価は k 近傍法で行う。具体的な評価の方法としては、まず 11,100 個の訓練用共分散記述子に対して、ある 1 つの評価用領域から計算された共分散記述子 $\mathbf{X}_{1,tst}, \dots, \mathbf{X}_{50,tst}$ のそれぞれの k 近傍 (合計 $11,100 \times k$ 個) を算出する。こうして得られたすべての近傍のなかで、最も多かった近傍のクラスラベルを予測ラベルとする。これを全ての評価用領域で行い、最終的な識別率を計算する。本実験では k 近傍法の k は 3 とした。また、Algorithm 1 の各パラメータはそれぞれ $K = 100 \times n_c$, $\mathbf{c} = \mathbf{1} \times \frac{1}{10^{-2} \times K}$, $b_{ub} = 0.05$, $b_{lb} = 0.95$ とした。

一般物体認識. さらに我々は ETH-80 dataset[23] を用いて一般物体認識実験も行った。ETH-80 dataset は 8 つのカテゴリを持つデータセットである。各カテゴリは 10 個のオブジェクトで構成されており、それぞれのオブジェクトは 41 枚の画像を含んでいる。我々は各オブジェクトごとランダムに 20 枚の画像を訓練用に選択し、残りは評価に用いた。

テクスチャ認識とは異なり、本実験では 1 枚の画像から 1 つの共分散記述子を計算した。その際に画像の各ピクセルから得る特徴は $[x, y, R_{xy}, G_{xy}, B_{xy}, |I_x|, |I_y|, |I_{xx}|, |I_{yy}|]$ のように定義した。ETH-80 dataset はカラー画像で構成されているため、 R_{xy}, G_{xy}, B_{xy} は画像の各色成分を、 I はグレースケール化した画像を表している。この特徴を用いて、各画像を共分散記述子に変換し、 k 近傍法を用いて分類実験を行った。実験に必要な各パラメータは $\mathbf{c} = \mathbf{1} \times \frac{1}{10^{-3} \times K}$ とし、この他はテクスチャ認識で用いたものと同様とした。

比較手法. 本研究では 4 種類の Φ : $\Phi(\mathbf{X}) = \mathbf{X}$, $\Phi(\mathbf{X}) = \log m(\mathbf{X})$, $\Phi(\mathbf{X}) = \mathbf{X}^{\frac{1}{2}}$, $\Phi(\mathbf{X}) = \text{chol}(\mathbf{X})$ を用いて、以下のような 8 つの方法で比較実験を行った。ただし、 $D_{\Phi}(\mathbf{X}_1, \mathbf{X}_2; \mathbf{I}) = D_{\Phi}(\mathbf{X}_1, \mathbf{X}_2)$ となることに注意され

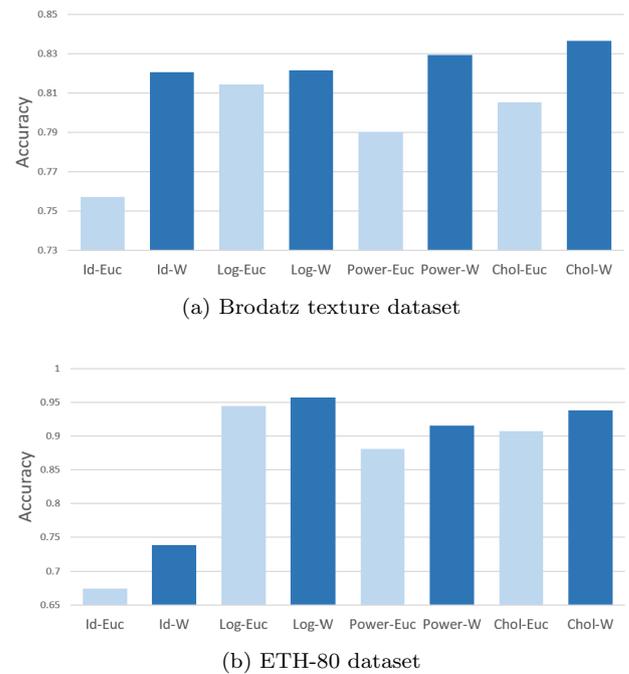


図 3 Categorization Performance.

たい。

- **Id-Euc:** $\Phi(\mathbf{X}) = \mathbf{X}$ かつ $\mathbf{W} = \mathbf{I}$.
- **Id-W:** $\Phi(\mathbf{X}) = \mathbf{X}$ かつ \mathbf{W} を学習.
- **Log-Euc:** $\Phi(\mathbf{X}) = \log m(\mathbf{X})$ かつ $\mathbf{W} = \mathbf{I}$.
- **Log-W:** $\Phi(\mathbf{X}) = \log m(\mathbf{X})$ かつ \mathbf{W} を学習.
- **Power-Euc:** $\Phi(\mathbf{X}) = \mathbf{X}^{\frac{1}{2}}$ かつ $\mathbf{W} = \mathbf{I}$.
- **Power-W:** $\Phi(\mathbf{X}) = \mathbf{X}^{\frac{1}{2}}$ かつ \mathbf{W} を学習.
- **Chol-Euc:** $\Phi(\mathbf{X}) = \text{chol}(\mathbf{X})$ かつ $\mathbf{W} = \mathbf{I}$.
- **Chol-W:** $\Phi(\mathbf{X}) = \text{chol}(\mathbf{X})$ かつ \mathbf{W} を学習.

実験結果. 図 3 に、テクスチャ認識と一般物体認識上での実験結果を示す。縦軸は無作為に 5 回実験を繰り返して得た平均識別率である。実験の結果、いずれのデータセットでも計量学習ありの方が計量学習をしないときに比べて、識別率が向上することがわかった。特に、テクスチャ認識では提案法である **Chol-W** が、一般物体認識では提案法である **Log-W** が最も高い識別率を示した。

6. 結論

本研究の成果をまとめると次のようになる：

- 共分散記述子の距離関数として、正定値錐上における新たなパラメトリックな距離計量関数を提案した。
- 距離計量関数のパラメータを、距離の公理を保持したまま、学習するアルゴリズムを提案した。
- 学習アルゴリズムに乱択戦略を導入すると、収束が速くなることを実験的に示した。
- 画像分類への応用として、テクスチャ認識と一般物体認識の 2 つの応用に適用し、計量学習の有効性を確認した。

参考文献

- [1] Pennec, X., Fillard, P. and Ayache, N.: A Riemannian framework for tensor computing, *International Journal of Computer Vision*, Vol. 66, No. 1, pp. 41–66 (2006).
- [2] Fletcher, P. T. and Joshi, S.: Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors, *Computer Vision and Mathematical Methods in Medical and Biomedical Image Analysis*, Springer, pp. 87–98 (2004).
- [3] Goh, A. and Vidal, R.: Clustering and dimensionality reduction on Riemannian manifolds, *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, IEEE, pp. 1–7 (2008).
- [4] Tuzel, O., Porikli, F. and Meer, P.: Region covariance: A fast descriptor for detection and classification, *Computer Vision–ECCV 2006*, Springer, pp. 589–600 (2006).
- [5] Tuzel, O., Porikli, F. and Meer, P.: Pedestrian detection via classification on riemannian manifolds, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 30, No. 10, pp. 1713–1727 (2008).
- [6] Zhang, S., Kasiviswanathan, S., Yuen, P. C. and Harandi, M.: Online dictionary learning on symmetric positive definite manifolds with vision applications, *The Twenty-Ninth AAAI Conf. Artificial Intelligence, AAAI, Austin, Texas* (2015).
- [7] Guo, K., Ishwar, P. and Konrad, J.: Action recognition from video using feature covariance matrices, *Image Processing, IEEE Transactions on*, Vol. 22, No. 6, pp. 2479–2494 (2013).
- [8] Harandi, M. T., Sanderson, C., Hartley, R. and Lovell, B. C.: Sparse coding and dictionary learning for symmetric positive definite matrices: A kernel approach, *Computer Vision–ECCV 2012*, Springer, pp. 216–229 (2012).
- [9] Arsigny, V., Fillard, P., Pennec, X. and Ayache, N.: Log-Euclidean metrics for fast and simple calculus on diffusion tensors, *Magnetic resonance in medicine*, Vol. 56, No. 2, pp. 411–421 (2006).
- [10] Dryden, L., Koloydenko, A. and Zhou, D.: Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging, *Annals of Applied Statistics*, pp. – (2009).
- [11] Wang, Z. and Vemuri, B. C.: An affine invariant tensor dissimilarity measure and its applications to tensor-valued image segmentation, *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, Vol. 1, IEEE, pp. I-228 (2004).
- [12] Cherian, A., Sra, S., Banerjee, A. and Papanikolopoulos, N.: Efficient similarity search for covariance matrices via the Jensen-Bregman LogDet divergence, *Computer Vision (ICCV), 2011 IEEE International Conference on*, IEEE, pp. 2399–2406 (2011).
- [13] Davis, J. V., Kulis, B., Jain, P., Sra, S. and Dhillon, I. S.: Information-theoretic metric learning, *Proceedings of the 24th international conference on Machine learning, ICML '07*, New York, NY, USA, ACM, pp. 209–216 (2007).
- [14] Weinberger, K. Q. and Saul, L. K.: Distance Metric Learning for Large Margin Nearest Neighbor Classification, *J. Mach. Learn. Res.*, Vol. 10, pp. 207–244 (2009).
- [15] Jayasumana, S., Hartley, R., Salzmann, M., Li, H. and Harandi, M. T.: Kernel Methods on the Riemannian Manifold of Symmetric Positive Definite Matrices., *CVPR, IEEE*, pp. 73–80 (2013).
- [16] Huang, Z., Wang, R., Shan, S., Li, X. and Chen, X.: Log-Euclidean Metric Learning on Symmetric Positive Definite Manifold with Application to Image Set Classification, *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, pp. 720–729 (2015).
- [17] Dykstra, R. L.: An Algorithm for Restricted Least Squares Regression, *Journal of the American Statistical Association*, Vol. 78, No. 384, pp. 837–842 (1983).
- [18] Censor, Y. and Reich, S.: The Dykstra algorithm with Bregman projections, *Communications in Applied Analysis*, Vol. 2, pp. 407–419 (1998).
- [19] Bottou, L.: Large-scale machine learning with stochastic gradient descent, *Proceedings of COMPSTAT'2010*, Springer, pp. 177–186 (2010).
- [20] Johnson, R. and Zhang, T.: Accelerating stochastic gradient descent using predictive variance reduction, *Advances in Neural Information Processing Systems*, pp. 315–323 (2013).
- [21] Defazio, A. J., Caetano, T. S. and Domke, J.: Finito: A faster, permutable incremental gradient method for big data problems, *arXiv preprint arXiv:1407.2710* (2014).
- [22] Randen, T. and Husoy, J. H.: Filtering for texture classification: A comparative study, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 21, No. 4, pp. 291–310 (1999).
- [23] Leibe, B. and Schiele, B.: Analyzing appearance and contour based methods for object categorization, *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, Vol. 2, IEEE, pp. II-409 (2003).

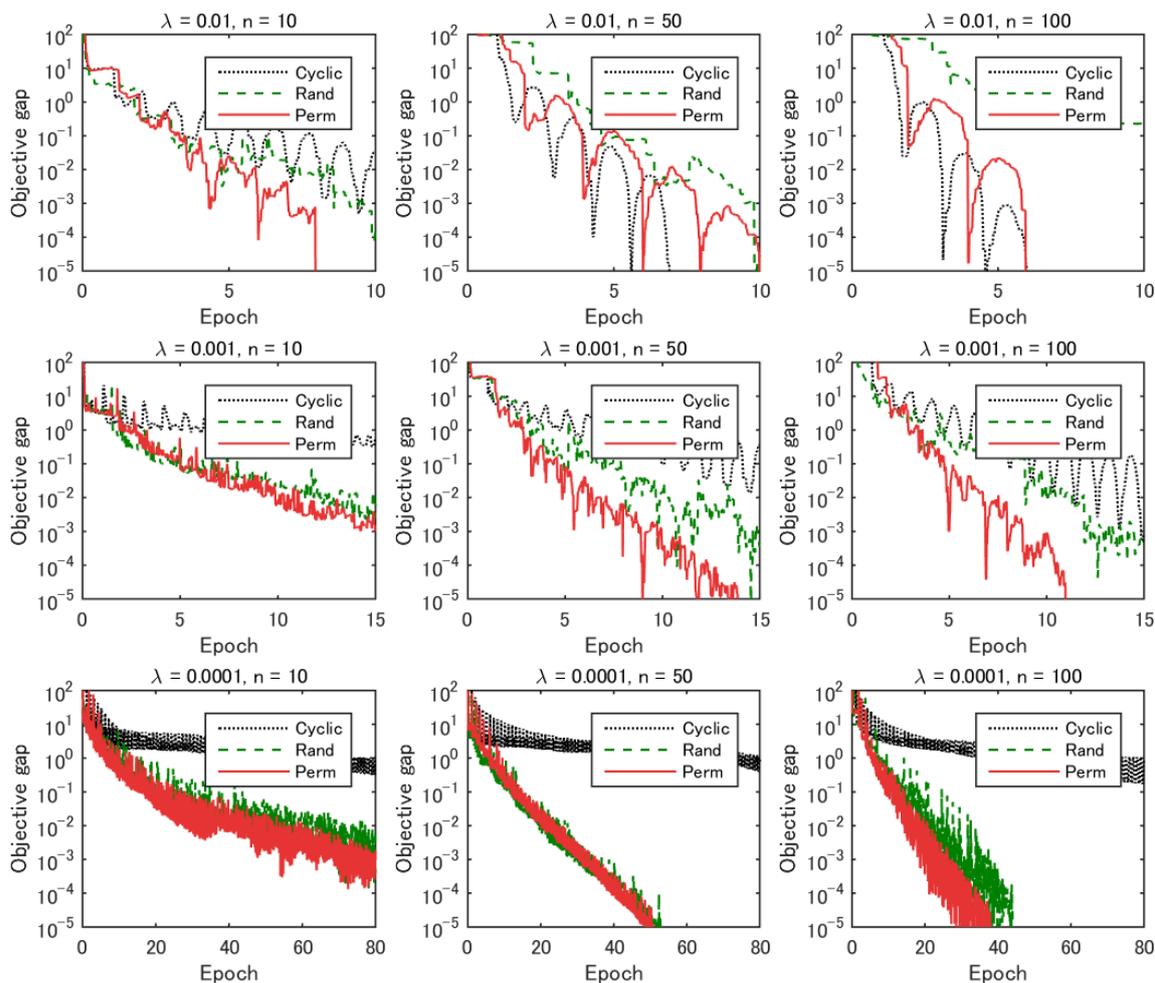


図 1 目的関数が収束する様子。プレグマン射影を行う半空間の選び方における目的関数の収束性の違いを評価した。ただし、**Cyclic** は周期的に半空間を選択，**Perm** はエポックごとに半空間の順番を無作為に入れ替えて選択，**Rand** は毎回半空間を無作為に選択するとした。横軸はエポックを表す。K 回半空間にプレグマン射影すると 1 エポックになる。また、縦軸は、エポックごとの目的関数の値と最適解との差を表す。

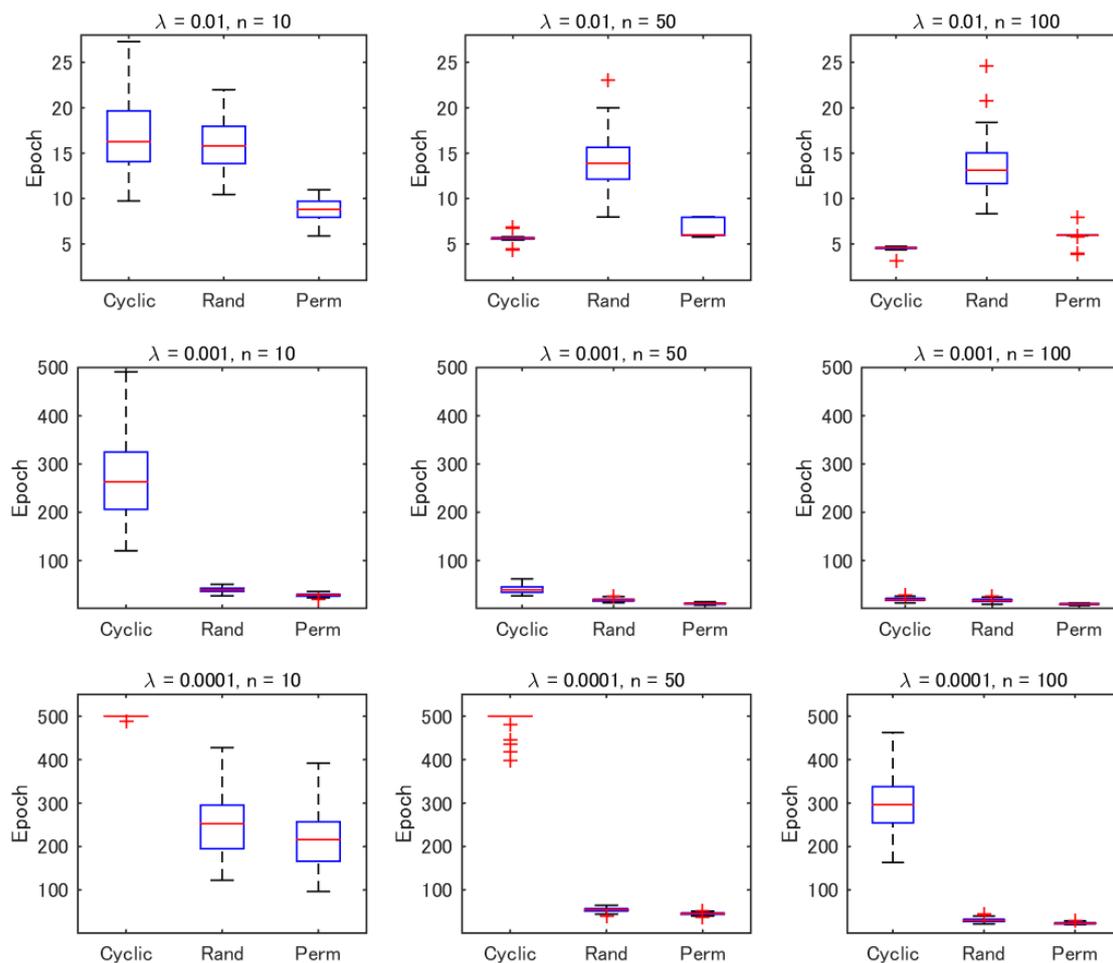


図 2 双対ギャップが 10^{-5} 以下に収束するまでの反復数. 人工データセットを用いた実験を 100 回繰り返し, 各手法が最適解に収束するまでのエポック数を測定した. 双対ギャップが 10^{-5} 未満になったとき最適解に収束したとみなした. ただし, 最大エポック数は 500 回とし, その時点で双対ギャップが 10^{-5} に達していない場合でも反復を停止させた. こうして, 各条件で得られた 100 個のエポック数を箱ひげ図で表した.