

オンライン線形判別分析アルゴリズムの導出と これを用いた顔判別

平岡和幸 日台健一 濱平仁 溝口博 重原孝臣 三島健稔
埼玉大学 工学部

Derivation of Online Linear Discriminant Analysis and Its Application to Face Identification

K. Hiraoka, K. Hidai, M. Hamahira, H. Mizoguchi, T. Shigehara, and T. Mishima
Faculty of Engineering, Saitama University

Abstract: Linear Discriminant Analysis (LDA) has been used in various fields, e.g. image recognition problems. However, when we apply LDA to large size images with N pixels, we face the problem of heavy computational cost which arises from calculations on $N \times N$ matrices. Moreover, LDA has another weak point: when we add new images to the current image set, we have to execute almost whole calculations again from the beginning.

In the present paper, an online learning algorithm of LDA (OLDA) is derived and it is applied to face identification. By OLDA, the image recognition system can be updated with low computational cost when new additional images are presented. Hence OLDA has the ability of adaptation to the change of environment. OLDA also has the advantage that $N \times N$ matrices never appear in its calculation.

Key Words: linear discriminant analysis, online learning, face identification, steepest descent method, matrix dynamics

1 はじめに

ロボットが人間と共存するためには、人間と円滑にコミュニケーションを行う能力が必須である。そのため有用な技術の一つとして、顔画像認識が挙げられる。顔画像認識には、線形判別分析 (Linear Discriminant Analysis, LDA) が広く用いられている [8][9]。しかし、画素数 N が大きい画像に線形判別分析を適用しようとすると、 $N \times N$ の大規模行列を扱わなくてはならず、計算が困難になってしまう。また、線形判別分析には、新しいデータを追加しようとしたときに、一般化固有値問題をはじめから解き直さなくてはならないという欠点もある。

本研究では、オンライン型の線形判別分析アルゴリズムを導き、顔画像判別への適用例を示す。このアルゴリズムは、

1. $N \times N$ の大規模行列を扱う必要がない
2. 新しいデータの追加が、少ない計算量で可能である

という特長を持っている。

2 従来の線形判別分析法

2.1 判別の手続き

学習のための例題として、 n 枚の顔画像とそれぞれが誰の顔なのか提示されるとする。顔画像は、各画素の輝度を順に並べることにより、 N 次元縦ベクトル \mathbf{x} で表す。 N は 1 枚の画像に含まれる画素の数 (ピクセル数) である。第 t 枚目の顔画像を $\mathbf{x}(t)$ とし、それが誰の顔かをクラス $c(t)$ で表す。A さんの顔なら $c(t) = 1$ 、B さんの顔なら $c(t) = 2$ 、という具合である。判別しなくてはならない人数を M とすると、 $c(t)$ は、 $1, \dots, M$ のいずれかの値をとる。

このように与えられた例題に対し、全体の平均ベクトル $\bar{\mathbf{x}}$ と、各クラス c のクラス平均ベクトル $\bar{\mathbf{x}}^c$ を、

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{t=1}^n \mathbf{x}(t), \quad \bar{\mathbf{x}}^c = \frac{\sum_{t=1}^n \delta(c, c(t)) \mathbf{x}(t)}{\sum_{t=1}^n \delta(c, c(t))} \quad (1)$$

と計算する。ここに、

$$\delta(c, c(t)) = \begin{cases} 1 & (c = c(t)) \\ 0 & (c \neq c(t)) \end{cases} \quad (2)$$

とおく。

提示された顔画像ベクトル \mathbf{x} が誰の顔を判別する手続きは、次の通りである。 A を、ある $N \times L$ 行列 ($L < M, L \leq N$) としよう。この A を判別行列と呼ぶ。判別行列 A を使い、 N 次元の顔画像ベクトル \mathbf{x} を、 $\mathbf{y} = A^T \mathbf{x}$ と変換する。ここに T は転置を表す。この \mathbf{y} を特徴ベクトルと呼ぶ。提示された画像の特徴ベクトル \mathbf{y} と、前述のクラス平均の特徴ベクトル $\mathbf{y}^c = A^T \mathbf{x}^c$ との自乗距離 $d^c = \|\mathbf{y} - \mathbf{y}^c\|^2 = (\mathbf{y} - \mathbf{y}^c)^T (\mathbf{y} - \mathbf{y}^c)$ を求め、 d^c が最小となるクラス c を答える。

この手続きによる判別がうまくいくかどうかは、判別行列 A の選び方で決まる。次節のような手順で A を決定するのが、従来の線形判別分析法である。

2.2 判別行列の決定

クラス平均と全体平均との差 \mathbf{v}^c 、および、各例題とクラス平均との差 $\mathbf{w}(t)$ を、

$$\mathbf{v}^c = \bar{\mathbf{x}}^c - \bar{\mathbf{x}} \quad (3)$$

$$\mathbf{w}(t) = \mathbf{x}(t) - \bar{\mathbf{x}}^{c(t)} \quad (4)$$

とし、クラス間分散行列 B とクラス内分散行列 W を

$$B = \frac{1}{M} \sum_{c=1}^M \mathbf{v}^c \mathbf{v}^{cT} \quad (5)$$

$$W = \frac{1}{n} \sum_{t=1}^n \mathbf{w}(t) \mathbf{w}(t)^T \quad (6)$$

により求める。

分散行列 B, W に対し、方程式

$$B\mathbf{p} = \lambda W\mathbf{p} \quad (7)$$

$$\mathbf{p}^T W\mathbf{p} = 1 \quad (8)$$

を満たす実数 λ および N 次元ベクトル \mathbf{p} を、計量 W に関する B の一般化固有値および一般化固有ベクトルと呼ぶ。(7)(8) の解を、 λ が大きい順に並べて、

$$(\lambda_1, \mathbf{p}_1), \dots, (\lambda_N, \mathbf{p}_N), \quad \lambda_1 \geq \dots \geq \lambda_N \quad (9)$$

とおく。そして、 L 本目までの一般化固有ベクトルを並べた行列

$$P_L = (\mathbf{p}_1, \dots, \mathbf{p}_L) \quad (10)$$

を求め、 $A = P_L$ とするのが、従来の線形判別分析法である。

2.3 従来法の欠点

前節のように、判別行列 A を決定する問題は、一般化固有値問題 (7)(8) を解くことに帰着される。このため、画素数 N が大きい顔画像を判別に用いようとすると、 $N \times N$ の大規模行列 B, W に関する一般化固有値問題を解かねばならず、計算が困難になってしまう。また、新たに $(n+1)$ 枚目の例題 ($\mathbf{x}(n+1), c(n+1)$) が与えられたら、 B, W が変わって、一般化固有値問題を再度はじめから解き直さなくてはならない。それには $O(N^3)$ の計算量が必要になってしまう。このように、例題を逐次的に追加することが、従来法では効率良くできない。

3 オンライン線形判別分析

従来の線形判別分析法は、例題をすべてそろえてからそれらをまとめて処理するという、バッチ型のアルゴリズムであった。これに対し、本研究のアルゴリズムは、例題が一つ与えられるたびに判別行列 A を修正していくという、オンライン型のアルゴリズムになっている。 A を決めた後、提示された顔画像を判別する手順は、従来法と同じである。

ここでは具体的なアルゴリズムのみを示すことにし、その導出は付録にゆずる。

3.1 基本形

本節では、オンライン線形判別分析アルゴリズムの基本形を示す。実際には、この通りに実装するのではなく、計算の順序を工夫することで、計算量が大幅に削減される。この工夫については、次節で述べる。

毎時刻 $t = 1, 2, \dots$ ごとに、新しい例題 ($\mathbf{x}(t), c(t)$) が与えられるとする。与えられた例題に基づき、毎時刻、

以下のようにして変数を更新していく ($c = 1, \dots, M$),

$$t^c(t) = t^c(t-1) + \delta(c, c(t)) \quad (11)$$

$$\bar{\mathbf{x}}(t) = \left(1 - \frac{1}{t}\right) \bar{\mathbf{x}}(t-1) + \frac{1}{t} \mathbf{x}(t) \quad (12)$$

$$\bar{\mathbf{x}}^c(t) = \begin{cases} \left(1 - \frac{1}{t^c(t)}\right) \bar{\mathbf{x}}^c(t-1) + \frac{1}{t^c(t)} \mathbf{x}(t) & (c = c(t)) \\ \bar{\mathbf{x}}^c(t-1) & (c \neq c(t)) \end{cases} \quad (13)$$

$$\mathbf{v}^c(t) = \bar{\mathbf{x}}^c(t) - \bar{\mathbf{x}}(t) \quad (14)$$

$$\mathbf{w}(t) = \mathbf{x}(t) - \bar{\mathbf{x}}^{c(t)}(t) \quad (15)$$

$$B(t) = \frac{1}{M} \sum_{c=1}^M \mathbf{v}^c(t) \mathbf{v}^c(t)^T \quad (16)$$

そして、これらの変数を用いて、判別行列 A を更新する。

$$\begin{aligned} A(t) &= A(t-1) + \eta \left[B(t) A(t-1) \right. \\ &\quad - \frac{1}{2} B(t) A(t-1) A(t-1)^T \mathbf{w}(t) \mathbf{w}(t)^T A(t-1) \\ &\quad \left. - \frac{1}{2} \mathbf{w}(t) \mathbf{w}(t)^T A(t-1) A(t-1)^T B(t) A(t-1) \right] \end{aligned} \quad (17)$$

ここに、 η は、小さな正の数である。この η を学習係数と呼ぶ。

なお、初期値は、 $t^c(0) = 0$ 、 $\bar{\mathbf{x}}(0)$ と $\bar{\mathbf{x}}^c(0)$ は任意ベクトル、 $A(0)$ はフルランク ($\text{rank} A(0) = L$) の任意行列とする。

3.2 計算の工夫

前節の基本形通りに計算したのでは、 $N \times N$ 行列を扱う必要が生じてしまい、効率が悪くなる。実は、次のような手順に従えば、 $N \times N$ 行列を生成することなく、同じ計算を遂行することができる。

$t^c, \bar{\mathbf{x}}, \bar{\mathbf{x}}^c, \mathbf{v}^c, \mathbf{w}$ については、前節と同じように更新する。さらに、

$$\mathbf{y}^c(t) = A(t-1)^T \mathbf{v}^c(t) \quad (18)$$

$$\mathbf{z}(t) = A(t-1)^T \mathbf{w}(t) \quad (19)$$

$$F(t) = \frac{1}{M} \sum_{c=1}^M \mathbf{v}^c(t) \mathbf{y}^c(t)^T \quad (20)$$

$$g(t) = \frac{1}{M} \sum_{c=1}^M \mathbf{y}^c(t) (\mathbf{y}^c(t)^T \mathbf{z}(t)) \quad (21)$$

表 1: 線形判別分析の計算量の比較

[新しく追加された例題を学習する際の、計算量を示す。 N はデータベクトルのサイズ (画素数)、 M はクラス数 (判別すべき人数)、 L は抽出する特徴数。特に、画素数が大きい画像の場合、本研究の手法は有効である。さらなる改良策については、5 節参照。]

	time complexity	space (memory) complexity
従来法	$O(N^3)$	$O(N^2)$
[5][6]	$O(N^2)$	$O(N^2)$
本研究	$O(NML)$	$O(NM)$

を計算し、

$$\begin{aligned} A(t) &= A(t-1) \\ &\quad + \eta \left(F(t) - \frac{1}{2} F(t) \mathbf{z}(t) \mathbf{z}(t)^T - \frac{1}{2} \mathbf{w}(t) \mathbf{g}(t)^T \right) \end{aligned} \quad (22)$$

と判別行列 $A(t)$ を更新する。

$N \times N$ 行列 $B(t)$ を廃した点、および、計算の途中でも $N \times N$ 行列は生じない点に注意。

3.3 アルゴリズムの特徴

線形判別分析のオンライン化は、これまでもいくつか提案された [5][6]。しかし、それらには、 $N \times N$ 行列を保持・更新しなくてはならないという欠点がある。従来法、[5][6]、本研究のアルゴリズムについて、計算量を比較すると、表 1 のようになる。顔画像判別では、表中の、画素数 N 、判別すべき人数 M 、抽出する特徴数 L に関して、 $N \gg M > L$ となっていることが典型的である。このような場合、本研究の手法は効率が良い。

オンライン線形判別分析アルゴリズムの理論的性質 (局所収束性など) は、[11] 参照。

4 顔画像判別への適用例

顔画像からの個人識別について、本アルゴリズムを適用した例を示す。

顔画像は、10 ピクセル \times 10 ピクセルとした。したがって、画素数 $N = 10 \times 10 = 100$ となる。各ピクセ

ルは -1 から $+1$ の 256 段階のグレースケール値を持つ (Fig. 1)。

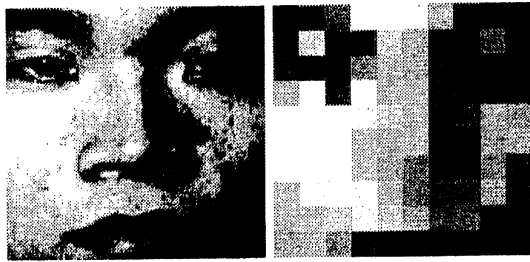


図 1: 実験に用いた画像の例 (左: 元画像、右: 10×10 ピクセル画像)

4.1 学習能力の確認

提案したアルゴリズムを、人物 A, B の 2 人の顔画像の判別問題に適用した。各人について 50 枚の顔画像を撮影し、実験用のサンプルとして使用した。これらの画像が、交互に与えられる：

$$\begin{aligned} \boldsymbol{x}(1) &= \text{image}_{A,1}, & \boldsymbol{x}(2) &= \text{image}_{B,1}, \\ \boldsymbol{x}(3) &= \text{image}_{A,2}, & \boldsymbol{x}(4) &= \text{image}_{B,2}, \dots, \\ \boldsymbol{x}(99) &= \text{image}_{A,50}, & \boldsymbol{x}(100) &= \text{image}_{B,50} \end{aligned}$$

線形判別分析で求める特徴数 (A の列数) は $L = 2$ とし、学習係数 η は 0.1 とした。

学習の結果、これらの画像の 97% を正しく判別できるようになった。結果を図 2 ($t = 100$ まで) に示す。ステップ t は、提示された画像の総枚数に一致する。判別性能は、10 ステップごとに評価した。評価には、全画像 100 枚を提示し、正しく判別する比率を調べた。なお、評価の間、学習は停止した。

4.2 環境変化

前節で述べたように、 $t = 100$ までは 2 人の人物 A, B ($c = 1, 2$) を提示した。 $t = 100$ において、突然、第 3 の人物 C を登場させ、 C に対応するクラス $c = 3$ を追加した。

$t = 100$ 以降、判別性能は人物 A, B, C の全画像 150 枚で評価した。判別性能は、環境変化により、一時的に約 80% まで低下した。

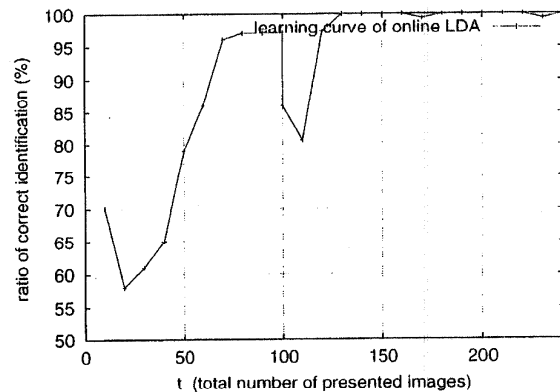
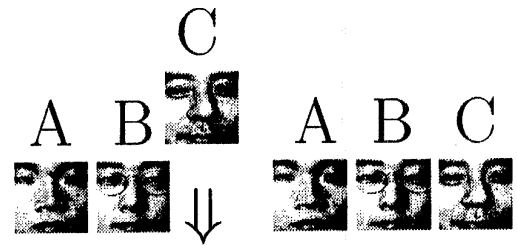


図 2: 実験で得られた学習曲線。 ($t = 100$ までは、2 人の人物 (A, B) の顔画像を交互に提示した。 $t = 100$ において、第 3 の人物 (C) のクラスを追加し、以降は 3 人の顔画像を順に提示した。本研究のアルゴリズムは、環境の変化に適応することができる。)

4.3 適応能力の確認

$t = 100$ 以降は、人物 A, B, C それぞれ 50 枚ずつの画像を、順に提示した。

$$\begin{aligned} \boldsymbol{x}(101) &= \text{image}_{A,1}, & \boldsymbol{x}(102) &= \text{image}_{B,1}, \\ \boldsymbol{x}(103) &= \text{image}_{C,1}, & \boldsymbol{x}(104) &= \text{image}_{A,2}, \\ \boldsymbol{x}(105) &= \text{image}_{B,2}, & \boldsymbol{x}(106) &= \text{image}_{C,2}, \dots \end{aligned}$$

提案したアルゴリズムは、新しい環境に適応し、ほぼ 100% の判別率に回復した。

5 おわりに

オンライン線形判別アルゴリズムを導き、顔画像判別への適用例を示した。本アルゴリズムは、大規模行列を扱う必要がなく、少ない計算量で新しいデータの追加ができるという利点を持つ。今後の課題としては、

学習係数の決定法や、大域収束性の厳密な証明が挙げられる。計算量のさらなる削減も、今後の検討課題である。例えば、(18)(20)を

$$\mathbf{y}^c(t) = \begin{cases} A(t-1)^T \mathbf{v}^c(t) & (c = c(t)) \\ \mathbf{y}^c(t-1) & (c \neq c(t)) \end{cases} \quad (23)$$

$$F(t) = F(t-1) + \frac{1}{M} \mathbf{v}^{c(t)}(t) \left(\mathbf{y}^{c(t)}(t) - \mathbf{y}^{c(t)}(t-1) \right)^T \quad (24)$$

のように変更すれば、表1中のtime complexityを $O(NM)$ に下げることができる。学習係数 η が十分小さければ、このように変更しても、結果に大きな影響はないと期待される。実際に判別性能がどの程度変化するかは、実験によって確認する必要がある。

本研究の一部は、文部省科学研究費助成金(基盤B)により行われた。This work has been partly supported by CREST of JST (Japan Science and Technology) 279102.

参考文献

- [1] E. Oja, "A simplified neuron model as a principal component analyzer", J. Math. Biol., Vol. 15, pp. 267-273, 1982.
- [2] T. D. Sanger, "Optimal unsupervised learning in a single-layer linear feedforward neural network", Neural Networks, Vol. 2, No. 6, pp. 459-473, 1989.
- [3] K. Hornik and C.-M. Kuan, "Convergence analysis of local feature extraction algorithms", Neural Networks, Vol. 5, No. 2, pp. 229-240, 1992.
- [4] W.-Y. Yan, U. Helmke, and J. B. Moore, "Global analysis of Oja's flow for neural networks", IEEE Tr. NN, Vol. 5, No. 5, pp. 674-683, 1994.
- [5] J. Mao and A. K. Jain, "Artificial neural networks for feature extraction and multivariate data projection", IEEE Tr. NN, Vol. 6, No. 2, pp. 296-317, 1995.
- [6] C. Chatterjee and V. P. Roychowdhury, "On self-organizing algorithms and networks for class-separability features", IEEE Tr. NN, Vol. 8, No. 3, pp. 663-678, 1997.

- [7] M. B. Nevel'son and R. Z. Has'minskiĭ, *Stochastic Approximation and Sequential Estimation*, Nauka, 1968 (in Russian; translated into Japanese by T. Kitagawa and K. Tajima, 1983).
- [8] T. Kurita and S. Hayamizu, "Gesture Recognition using HLAC Features of PARCOR Images and HMM based Recognizer", Proc. 3rd int. conference on automatic face and gesture recognition, April, 1998, pp. 422-427.
- [9] W. Zhao, et al., "Discriminant Analysis of Principal Components for Face Recognition", Proc. 3rd int. conference on automatic face and gesture recognition, 1998, pp. 422-427. pp. 336-341.
- [10] K. Fukunaga, *Statistical Pattern Recognition*, Academic Press, New York, 1989.
- [11] 平岡 和幸, 濱平 仁 "逐次学習型の線形判別分析アルゴリズムについて— 学習アルゴリズムの提案と局所収束性の証明 —", 信学技報 NC99-73, pp. 85-92, 1999.

A 付録: オンライン線形判別分析アルゴリズムの導出

A.1 解くべき問題

まず、 $d^c = (\mathbf{x} - \mathbf{x}^c)^T A A^T (\mathbf{x} - \mathbf{x}^c)$ である。したがって、 Q を任意の直交行列とすると、判別行列 $A = P_L$ のかわりに $\tilde{A} = P_L Q$ を用いても、自乗距離 d^c の値は変化しない。判別ではこの d^c が最小となるクラス c を答えるのだから、 d^c が変化しないということは、判別結果に影響が出ないことを意味する。よって、線形判別分析を行うためには、 P_L 自体を求めなくても、 $A = P_L Q$ (Q は直交行列)と表される行列 A を一つ求めれば十分である。

A.2 ポテンシャル関数

$N \times L$ 行列 A に対し、ポテンシャル関数 $-\phi(A)$ を

$$\phi(A) = \text{Tr} \left[A^T B A \left(I - \frac{1}{2} A^T W A \right) \right] \quad (25)$$

と定義する。ここに I は単位行列である。

Theorem 1 $\text{rank} B \geq L$, $\text{rank} W = N$ を仮定する。このとき、 $\phi(A)$ が最大になる A は、 $A = P_L Q$ (Q は直交行列)と表される。□

Proof: A.5 節を参照。 ■

したがって、線形判別分析は、 $\phi(A)$ が最大となるような行列 A を求める問題に帰着される。

A.3 最急降下法

$\phi(A)$ の最大化 ($-\phi(A)$ の最小化) を、最急降下法で行うことを考えよう。すなわち、 A の (i, j) 成分を a_{ij} として、

$$a_{ij} \leftarrow a_{ij} + \eta' \frac{\partial \phi(A)}{\partial a_{ij}} \quad (26)$$

という更新を繰り返すことにより、 $\phi(A)$ の最大化をめざす。ここに、 η' は小さな正の数である。これを具体的に計算し、

$$A \leftarrow A + \eta \left(BA - \frac{1}{2} BAA^T W A - \frac{1}{2} WAA^T B A \right) \quad (27)$$

という更新則を得る。ここに、 $\eta = 2\eta'$ とおいた。

A.4 確率近似法

式 (27) 中に現れる分散行列 W は、 $\mathbf{w}(t)\mathbf{w}(t)^T$ の平均値である。これを、瞬間値 $\mathbf{w}(t)\mathbf{w}(t)^T$ でおきかえれば、本研究のアルゴリズム (17) が得られる。このような置きかえは、確率近似法の一般論により、正当化される [7]。

A.5 Theorem 1 の証明

δJ を、 $\text{Tr}(\delta J^T \delta J) \ll 1$ なる行列とすると、

$$\phi(A(I + \delta J)) - \phi(A) \quad (28)$$

$$\approx 2\text{Tr} \left[\delta J^T A^T B A \left(I - \frac{1}{2} A^T W A \right) \right] - \text{Tr} \left[A^T B A A^T W A \delta J \right] \quad (29)$$

$$= \text{Tr} \left[A^T B A (\delta J + \delta J^T) \right] - \text{Tr} \left[A^T W A A^T B A (\delta J + \delta J^T) \right] \quad (30)$$

これが任意の δJ で 0 となるためには、

$$A^T B A = (A^T W A)(A^T B A) = (A^T B A)(A^T W A) \quad (31)$$

でなくてはならない。なぜならば、一般に、「対称行列 X, Y が、任意の対称行列 Z に対して $\text{Tr}(XZ) = \text{Tr}(YZ)$ なら、 $X = Y$ 」が言えるからである。

以上から、 $\phi(A)$ の極値においては、(31) が成り立つ。このとき特に、次式が成り立っている。

$$\phi(A) = \frac{1}{2} \text{Tr}(A^T B A) \quad (32)$$

ところで、(31) が成り立っていれば、 $J_B \equiv A^T B A$ と $J_W \equiv A^T W A$ は同時対角化可能である。よって、適当な直

交行列 R をとり、 $\tilde{A} = AR$ を A とおきなおせば、始めから J_B, J_W は対角と仮定して一般性を失わない。更に (31) から

$$J_B = \text{diag}(\beta_1, \dots, \beta_K, 0, \dots, 0) \quad (33)$$

$$J_W = \text{diag}(1, \dots, 1, \omega_{K+1}, \dots, \omega_L) \quad (34)$$

という形に限定することができる。ここに、 $\beta_1 > \dots > \beta_K > 0$ とする。また、 $\tilde{A} = \sqrt{W}A$ を A とおきなおすことにより、はじめから $W = I$ と仮定して一般性を失わない。

[ア] さて、まず $A^T A = I$ の場合を考えよう。 $A^T A = I$ の条件のもと、 $2\phi(A) = \text{Tr}(A^T B A)$ の最大値を、Lagrange 未定係数法により求める。 $\Gamma = (\gamma_{ij})$ を未定係数として、

$$\begin{aligned} \mathcal{L}(A, \Gamma) &\equiv \text{Tr}(A^T B A) - \sum_{i,j} \gamma_{ij} \left((A^T A)_{ij} - \delta(i, j) \right) \\ &= \text{Tr}(A^T B A) - \text{Tr}[\Gamma(A^T A - I)] \end{aligned} \quad (35)$$

に対し、 $\partial \mathcal{L} / \partial A = 2BA - 2A\Gamma$ が 0 となるには、 $BA = A\Gamma$ でなくてはならない。条件 $A^T A = I$ とあわせると、 A は、 $A = PQ$ (Q は直交行列) と表されることがわかる。ここに、 $N \times L$ 行列 P は、一般化固有ベクトル $\mathbf{p}_1, \dots, \mathbf{p}_N$ から相異なる L 本を選び、並べた行列である。このとき、

$$2\phi(A) = \text{Tr}(P^T B P) \quad (36)$$

であることから、「 $A^T A = I$ の条件のもと $\phi(A)$ を最大化する A は、 $A = P_L Q$ (Q は直交行列) と表される」を得る。

[イ] 次に、 $A^T A = J_W \neq I$ かつ $\omega_{K+1} \dots \omega_L \neq 0$ の場合を考える。このときは、 $\tilde{A} = A J_W^{-1/2}$ とおけば、 $\tilde{A}^T \tilde{A} = I$ であり、しかも

$$L(\tilde{A}) = \text{Tr} \left(J_W^{-1/2} J_B J_W^{-1/2} \right) \quad (37)$$

$$= \text{Tr} (J_B J_W^{-1}) = \text{Tr} J_B = L(A) \quad (38)$$

となっている。したがって、この場合は、[ア] に帰着される。そして結局、[イ] のような A では、 $L(A)$ は最大値となっていないことが [ア] からわかる。

[ウ] 最後に、 $A^T A = J_W \neq I$ かつ $\omega_{K+1} \dots \omega_L = 0$ の場合を考える。今、

$$\omega_{K+1} \omega_{K+2} \dots \omega_H \neq 0 \quad (39)$$

$$\omega_{H+1} = \omega_{H+2} = \dots = \omega_L = 0 \quad (40)$$

だったとしよう。このときは、

$$G \equiv \text{diag}(1, \dots, 1, \omega_{K+1}^{-1/2}, \dots, \omega_H^{-1/2}, 0, \dots, 0) \quad (41)$$

として $\tilde{A} = AG$ とおけば、 $\tilde{A}^T \tilde{A} = \text{diag}(1, \dots, 1, 0, \dots, 0)$ であり、しかも [イ] のときと同様に $L(\tilde{A}) = L(A)$ となっている。特に、この \tilde{A} は、 $\tilde{A} = (\tilde{\mathbf{a}}_1, \dots, \tilde{\mathbf{a}}_H, \mathbf{0}, \dots, \mathbf{0})$ という形である。さて、 $\text{rank} B \geq L$ の仮定から、

$$\tilde{\mathbf{a}}_i^T \mathbf{c} = 0 \quad (i = 1, \dots, H) \quad (42)$$

$$\mathbf{c}^T \mathbf{c} = 1, \quad \mathbf{c}^T B \mathbf{c} > 0 \quad (43)$$

を満たすベクトル \mathbf{c} が存在する。この \mathbf{c} を使って $N \times L$ 行列 $C = (\mathbf{0}, \dots, \mathbf{0}, \mathbf{c})$ を作れば、

$$L(\tilde{A} + C) = L(\tilde{A}) + \mathbf{c}^T B \mathbf{c} > L(\tilde{A}) \quad (44)$$

となる。したがって、[ウ] のような A では、 $L(A)$ は最大値になっていない。

[ア] [イ] [ウ] をあわせれば Theorem 1 が結論される。 ■