

## 常用 Fisher 判别函数的判别矩阵研究

程正东<sup>1,2</sup> 章毓晋<sup>3</sup> 樊祥<sup>1,2,4</sup> 朱斌<sup>1,2</sup>

**摘要** 在线性判别分析 (Linear discriminant analysis, LDA) 中, 比迹函数、比值函数和迹比函数是三种常用的 Fisher 判别函数, 每一个判别函数都可得到一个正交判别 (Orthogonal discriminant, OD) 矩阵和一个不相关判别 (Uncorrelated discriminant, UD) 矩阵. 本文的主要目的是对这 6 种判别矩阵的获取方法及其性质进行系统分析, 以期更清楚地认识它们的联系与区别. 当类内协方差阵非奇异时, 比迹、比值函数的判别矩阵和迹比函数的 OD 矩阵的获取方法及性质已有研究, 本文对迹比函数的 UD 矩阵的获取方法及性质进行了补充研究, 得到了迹比函数的 UD 矩阵与比迹、比值函数的 UD 矩阵是同一矩阵以及迹比函数的 UD 矩阵的判别函数值不超过它的 OD 矩阵的结论. 当类内协方差阵奇异时, 6 种判别矩阵的获取方法遇到了困难, 为克服这一困难, 本文首先用极限的思想重新定义了这三种判别函数, 然后采用求极限的方法得到了 6 种判别矩阵的获取方法. 从所得的获取方法可以看出, 当所需的判别向量均在类内协方差阵的零空间中时, 6 个判别矩阵是同一矩阵.

**关键词** Fisher 判别函数, 正交判别矩阵, 不相关判别矩阵, 极限

**DOI** 10.3724/SP.J.1004.2010.01361

### Study on Discriminant Matrices of Commonly-used Fisher Discriminant Functions

CHENG Zheng-Dong<sup>1,2</sup> ZHANG Yu-Jin<sup>3</sup> FAN Xiang<sup>1,2,4</sup> ZHU Bin<sup>1,2</sup>

**Abstract** In linear discriminant analysis (LDA), ratio-trace, ratio-value, and trace-ratio are three Fisher discriminant functions in common use. Each has one orthogonal discriminant (OD) matrix and one uncorrelated discriminant (UD) matrix. This paper aims to systematically analyze the obtained approach and properties of these six discriminant matrices and to more clearly recognize their relations and differences. When the within-class scatter matrix is nonsingular, the research on the discriminant matrices of ratio-trace and ratio-value functions and the OD matrix of trace-ratio function have already been done. This paper only discusses the obtained approach and properties of UD matrix of trace-ratio function, and obtains such conclusions that the UD matrix of trace-ratio function is equal to that of ratio-trace and ratio-value function, and the trace-ratio function value of UD matrix is no more than that of OD matrix. When the within-class scatter matrix is singular, it is difficult to obtain these six discriminant matrices. To overcome the difficulty, this paper introduces the limitation idea and redefines these three discriminant functions, then attains the obtained approach of the six discriminant matrices by calculating the limitations. From these obtained approaches, we can conclude that the six discriminant matrices are equal when the need discriminant vectors are all in the null space of the within-class scatter matrix.

**Key words** Fisher discriminant function, orthogonal discriminant matrix, uncorrelated discriminant matrix, limit

在模式识别中, 基于 Fisher 判别准则的线性判别分析 (Linear discriminant analysis, LDA) 是最为基本的有监督线性子空间方法. Fisher 判别准则是指对样本数据作线性投影, 使投影后数据的类间离散度达到最大而类内离散度达到最小, LDA 的任务就是寻找最佳投影. 寻找最佳投影的通用做法是先设置一个合适的目标函数, 然后最优化目标函数来得到最佳投影, 这样的目标函数常称为 Fisher 判

别函数. 目前人们已提出了多种行之有效的 Fisher 判别函数, 其中最为常用的三种判别函数如下:

1) 比迹函数<sup>[1]</sup>:

$$J_1(W) = \text{tr}((W^T S_w W)^{-1} (W^T S_b W))$$

2) 比值函数<sup>[2]</sup>:

$$J_2(\mathbf{w}) = \frac{\mathbf{w}^T S_b \mathbf{w}}{\mathbf{w}^T S_w \mathbf{w}}$$

3) 迹比函数<sup>[3]</sup>:

$$J_3(W) = \frac{\text{tr}(W^T S_b W)}{\text{tr}(W^T S_w W)}$$

其中,  $S_b, S_w$  分别表示样本数据的类间协方差阵与类内协方差阵,  $\mathbf{w}$  与  $W$  分别表示投影向量和投影矩阵. 最大化  $J_1(W)$  与  $J_3(W)$  得到的最佳投影矩阵  $W$  称为判别矩阵, 其列向量称为判别向量; 最大化  $J_2(\mathbf{w})$  可得到多个判别向量, 由它们可组成  $J_2(\mathbf{w})$

收稿日期 2009-10-23 录用日期 2010-03-30  
Manuscript received October 23, 2009; accepted March 30, 2010  
1. 脉冲功率激光技术国家重点实验室 (电子工程学院) 合肥 230037  
2. 合肥电子工程学院 合肥 230037 3. 清华大学电子工程系 北京 100084 4. 中国科学技术大学 合肥 230027  
1. State Key Laboratory of Pulsed Power Laser Technology (Electronic Engineering Institute), Hefei 230037 2. Electronic Engineering Institute of Hefei, Hefei 230037 3. Department of Electronic Engineering, Tsinghua University, Beijing 100084 4. University of Science and Technology of China, Hefei 230027

的判别矩阵. 需要注意的是, 本文中的判别向量都要求是单位向量.

在最大化 Fisher 判别函数时, 需要对判别矩阵  $W = (\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_d)$  附加一定的条件, 这是为了避免各个判别向量之间的夹角很小但判别函数值却很大的情形出现. 目前人们采用的附加条件主要有两种: 一种是要求各个判别向量相互正交<sup>[2]</sup>, 即满足  $\mathbf{w}_i^T \mathbf{w}_j = 0$  ( $i \neq j$ ), 称之为正交判别 (Orthogonal discriminant, OD) 矩阵, 这是从几何角度提出的附加条件; 另一种是要求样本经不同判别向量投影后的数据统计不相关<sup>[4]</sup>, 即满足  $\mathbf{w}_i^T S_t \mathbf{w}_j = 0$  ( $i \neq j$ ,  $S_t = S_w + S_b$ ), 称之为不相关判别 (Uncorrelated discriminant, UD) 矩阵, 这是从统计学角度提出的附加条件. 每一个 Fisher 判别函数都可得到一个 OD 矩阵和一个 UD 矩阵, 因此上述三个判别函数可得到 6 个判别矩阵. 在已有的文献中, 当类内协方差阵  $S_w$  非奇异时, 对  $J_1(W)$  的 UD 矩阵  $W_1^u$  与 OD 矩阵  $W_1^o$ ,  $J_2(\mathbf{w})$  的 UD 矩阵  $W_2^u$  与 OD 矩阵  $W_2^o$  以及  $J_3(W)$  的 OD 矩阵  $W_3^o$  的求取方法及性质都有讨论<sup>[1-6]</sup>, 但未见对  $J_3(W)$  的 UD 矩阵  $W_3^u$  的求取方法和性质进行讨论, 本文将补充相关内容.

上述三种判别函数的最大化方法各不相同, 比迹函数  $J_1(W)$  的最大化方法是广义特征值分解 (Generalized eigenvalue decomposition, GEVD)<sup>[1]</sup>, 比值函数  $J_2(\mathbf{w})$  的最大化方法是 Foley-Sammon 变换 (Foley-Sammon transformation, FST)<sup>[2, 4]</sup>, 而迹比函数  $J_3(W)$  的最大化方法是 Wang-Yan (WY) 算法<sup>[3]</sup>, 但这些方法只在  $S_w$  非奇异时有效. 当  $S_w$  奇异时, 为克服它们的不足, Chen 等<sup>[7]</sup> 提出了零空间 LDA (Null linear discriminant analysis, NLDA) 算法, Yang 等<sup>[8]</sup> 提出了双子空间算法. 不过, 这些算法既缺乏理论依据, 也没能解决  $S_w$  奇异时 UD 矩阵的求解问题. 本文引入极限的思想, 在  $S_w$  奇异时先用极限重新定义了上述三种判别函数, 然后通过求极限的办法得到了最优判别矩阵. 这样, 既解决了  $S_w$  奇异时 UD 矩阵的求解问题, 也从理论上证明了 NLDA 算法和双子空间算法的合理性.

## 1 最大化方法与判别矩阵回顾

设有  $N$  个训练样本, 它们分属  $c$  个类别, 用  $S_b$ ,  $S_w$  分别表示样本数据的类间协方差阵与类内协方差阵,  $S_t = S_w + S_b$  表示样本数据的总体协方差阵. 本节主要回顾当  $S_w$  非奇异时上述三种判别函数最大化方法和相应的判别矩阵, 以及当  $S_w$  奇异时的 NLDA 算法和双子空间算法.

### 1.1 GEVD 与比迹函数的判别矩阵

最大化比迹函数  $J_1(W)$  就是作  $S_b$  相对于  $S_w$

的广义特征值分解, 即  $J_1(W)$  的判别矩阵是由广义特征方程<sup>[1]</sup>

$$S_b \mathbf{w} = \lambda S_w \mathbf{w} \quad \text{或} \quad S_w^{-1} S_b \mathbf{w} = \lambda \mathbf{w} \quad (1)$$

的特征向量构成的. 对此, 有

引理 1<sup>[9]</sup>. 若存在可逆矩阵  $P$ , 使得

$$\begin{aligned} P^T S_b P &= \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\} \\ P^T S_w P &= \text{diag}\{\mu_1, \mu_2, \dots, \mu_n\} \end{aligned} \quad (2)$$

则  $P$  的各个列就是  $S_b$  相对于  $S_w$  的广义特征向量,  $\lambda_i/\mu_i$ ,  $\lambda_2/\mu_2$ ,  $\dots$ ,  $\lambda_n/\mu_n$  是相应的广义特征值.

由于  $P^T S_t P = P^T S_b P + P^T S_w P$  为对角阵, 若按列分块  $P = (\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n)$ , 则  $\mathbf{p}_i S_t \mathbf{p}_j = 0$  ( $i \neq j$ ), 即由 GEVD 得到的判别矩阵是  $J_1(W)$  的 UD 矩阵. 于是,  $J_1(W)$  的 UD 矩阵  $W_1^u$  为

$$W_1^u = \left( \frac{\mathbf{p}_1}{\|\mathbf{p}_1\|}, \frac{\mathbf{p}_2}{\|\mathbf{p}_2\|}, \dots, \frac{\mathbf{p}_d}{\|\mathbf{p}_d\|} \right) \quad (3)$$

其中,  $\|\cdot\|$  表示向量的  $L_2$  范数,  $d$  为所需判别向量的个数,  $d \leq c - 1$ .

对  $W_1^u$  作 QR 分解  $W_1^u = QR$ ,  $Q$  为正交阵,  $R$  为上三角阵. 由于  $J_1(W)$  对可逆矩阵具有不变性, 因此  $Q$  就是  $J_1(W)$  的 OD 矩阵<sup>[5]</sup>, 即  $W_1^o = Q$ .

### 1.2 FST 与比值函数的判别矩阵

最大化比值函数  $J_2(\mathbf{w})$  的方法是作关于  $S_b$  相对于  $S_w$  的 Foley-Sammon 变换, 这是一种递推算法, 其最一般的形式为:

引理 2<sup>[10]</sup>. 设  $A$  和  $B$  为  $n$  阶对称阵,  $A \geq 0$ ,  $B > 0$ , 判别函数为  $J(\mathbf{w}) = (\mathbf{w}^T A \mathbf{w}) / (\mathbf{w}^T B \mathbf{w})$ , 则满足附加条件  $\mathbf{w}_i^T G \mathbf{w}_j = 0$  ( $i \neq j$ ,  $G > 0$ ) 的判别矩阵可由下述递推算法求得: 若已获得前  $i$  个判别向量, 则第  $i + 1$  个判别向量是广义特征方程  $P_i A \mathbf{w} = \lambda B \mathbf{w}$  的最大广义特征值所对应的单位特征向量, 其中

$$P_i = I - G D_i (D_i^T G B^{-1} G D_i)^{-1} D_i^T G B^{-1} \quad (4)$$

$D_0 = 0$ ,  $D_i = (\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_i)$ ,  $i = 1, 2, \dots, d - 1$ . 当  $S_w$  非奇异时, 令  $A = S_b$ ,  $B = S_w$ , 则<sup>[2, 4]</sup>

1) 当  $G = I$ , 即

$$P_i = I - D_i (D_i^T S_w^{-1} D_i)^{-1} D_i^T S_w^{-1}$$

时, 可得  $J_2(\mathbf{w})$  的 OD 矩阵  $W_2^o$ , 相应的变换称为 OFST (Orthogonal Foley-Sammon transformation);

2) 当  $G = S_t$ , 即

$$P_i = I - S_t D_i (D_i^T S_t S_w^{-1} S_t D_i)^{-1} D_i^T S_t S_w^{-1}$$

时, 可得  $J_2(\mathbf{w})$  的 UD 矩阵  $W_2^u$ , 相应的变换称为 UFST (Uncorrelated Foley-Sammon transformation).

### 1.3 WY 算法与迹比函数的判别矩阵

WY 算法是一种迭代算法, 它的理论基础是以下两个引理:

引理 3<sup>[3]</sup>. 设  $J_3(W)$  的 OD 矩阵为  $W_3^o$ , 则

$$\beta^* = \frac{\text{tr}((W_3^o)^T S_b W_3^o)}{\text{tr}((W_3^o)^T S_w W_3^o)} = \max_{\substack{\mathbf{w}_i^T \mathbf{w}_j = 0 \\ \|\mathbf{w}_i\|=1}} \left( \frac{\text{tr}(W^T S_b W)}{\text{tr}(W^T S_w W)} \right) \quad (5)$$

充要条件是

$$\text{tr}((W_3^o)^T (S_b - \beta^* S_w) W_3^o) = \max_{\substack{\mathbf{w}_i^T \mathbf{w}_j = 0 \\ \|\mathbf{w}_i\|=1}} \{\text{tr}[W^T (S_b - \beta^* S_w)]\} = 0 \quad (6)$$

引理 4<sup>[3]</sup>. 对任意的  $0 < \beta \leq \beta^*$ , 若

$$W_\beta = \arg \max_{\substack{\mathbf{w}_i^T \mathbf{w}_j = 0 \\ \|\mathbf{w}_i\|=1}} \{\text{tr}[W^T (S_b - \beta S_w) W]\}$$

使得

$$\tilde{\beta} = \frac{\text{tr}(W_\beta^T S_b W_\beta)}{\text{tr}(W_\beta^T S_w W_\beta)} \quad (7)$$

则  $\beta \leq \tilde{\beta} \leq \beta^*$ .

引理 3 指出  $J_3(W)$  的 OD 矩阵  $W_3^o$  是矩阵  $(S_b - \beta^* S_w)$  的前  $d$  个最大特征值对应的特征向量组成的列正交矩阵; 而引理 4 则给出了求取参数  $\beta^*$  的迭代方法, 由此可得获取  $W_3^o$  的 WY 迭代算法 (见算法 1).

算法 1 (获取  $W_3^o$  的 WY 迭代算法).

步骤 1. 初始化, 令  $\beta^{(0)} = 0$ ; 迭代次数  $t = 1, 2, \dots, T$ .

步骤 2. 对  $(S_b - \beta^{(t-1)} S_w)$  作特征值分解, 取前  $d$  个最大特征值对应的特征向量组成列正交矩阵  $W^{(t)}$ , 即

$$W^{(t)} = \max_{\substack{\mathbf{w}_i^T \mathbf{w}_j = 0 \\ \|\mathbf{w}_i\|=1}} \text{tr}\{W^T (S_b - \beta^{(t-1)} S_w) W\}$$

步骤 3.  $\beta^{(t)} = \frac{\text{tr}\{(W^{(t)})^T S_b (W^{(t)})\}}{\text{tr}\{(W^{(t)})^T S_w (W^{(t)})\}}$ .

步骤 4. 若  $|\beta^{(t-1)} - \beta^{(t)}| < \varepsilon$  ( $\varepsilon$  为给定的很小正数), 中止; 否则, 转到步骤 2; 于是,  $W_3^o = W^{(T)}$ ,  $\beta^* = \beta^{(T)}$ .

### 1.4 NLDA 算法

NLDA 算法<sup>[7]</sup> 只能获取零空间  $S_w^{-1}(0) = \{\mathbf{x} | S_w \mathbf{x} = 0\}$  中的判别向量, 它的判别函数为

$$J_b(\mathbf{w}) = \frac{\mathbf{w}^T S_b \mathbf{w}}{\mathbf{w}^T \mathbf{w}}, \quad \mathbf{w} \in S_w^{-1}(0) \quad (8)$$

最大化  $J_b(\mathbf{w})$  的方法<sup>[11]</sup> 是: 先对  $S_w$  作特征值分解以得到零特征值对应的特征向量组成的列正交矩阵  $Q$ , 然后对  $Q^T S_b Q$  作特征值分解以得到前  $d$  个最大特征值对应的特征向量组成的列正交矩阵  $V$ . 于是  $W = QV$  就是 NLDA 算法的判别矩阵, 显然它是 OD 矩阵.

### 1.5 双子空间算法

由于 NLDA 算法只能获取零空间  $S_w^{-1}(0)$  中的判别向量, 当零空间  $S_w^{-1}(0)$  的维数较小时, 其所得到的判别矩阵往往因为判别向量的个数太少而不能完成模式识别任务. 为克服这一不足, Yang 等提出了双子空间算法<sup>[8]</sup>, 即不仅要由 NLDA 算法得到  $S_w^{-1}(0)$  中的判别向量, 还要由 FST 得到  $S_w$  的主空间  $(S_w^{-1}(0))^\perp = \{\mathbf{x} | \mathbf{x}^T \mathbf{y} = 0, \mathbf{y} \in S_w^{-1}(0)\}$  中的判别向量, 这些判别向量合在一起组成判别矩阵. 双子空间算法分为两种: 一种用于求 OD 矩阵, 称为 NLDA + OFST 算法; 另一种试图用于求 UD 矩阵, 称为 NLDA + UFST, 但它得到的并不是 UD 矩阵.

## 2 $S_w$ 非奇异时 $W_3^u$ 的获取方法及其性质

当  $S_w$  非奇异时, 在已有的文献中未见对  $J_3(W)$  的 UD 矩阵  $W_3^u$  的获取方法及其性质进行讨论, 本节将补充相关内容.

### 2.1 $W_3^u$ 的获取方法

迹比函数  $J_3(W)$  的 UD 矩阵  $W_3^u$  应为

$$W_3^u = \arg \max_{\substack{\mathbf{w}_i^T S_t \mathbf{w}_j = 0 \\ \|\mathbf{w}_i\|=1}} \left( \frac{\text{tr}(W^T S_b W)}{\text{tr}(W^T S_w W)} \right) \quad (9)$$

由于  $S_t$  正定, 存在可逆阵  $H$ , 使得  $H^T S_t H = I$ , 令  $\tilde{S}_b = H^T S_b H$ ,  $\tilde{S}_w = H^T S_w H$ ,  $\Phi = H^{-1} W$ , 则式 (9) 可写为

$$\tilde{\Phi} = \arg \max_{\substack{\phi_i^T \phi_j = 0 \\ \|\phi_i\|=1}} \left( \frac{\text{tr}(\Phi^T \tilde{S}_b \Phi)}{\text{tr}(\Phi^T \tilde{S}_w \Phi)} \right), \quad \Phi = (\phi_1, \phi_2, \dots, \phi_d) \quad (10)$$

由引理 3 和引理 4 知, 式 (10) 中的  $\tilde{\Phi}$  可由 WY 算法求得, 于是  $W_3^u = H \tilde{\Phi} R^{-1}$ , 其中  $R$  是对角阵, 其对角元是  $H \tilde{\Phi}$  的对应列的模长. 可见,  $J_3(W)$  的 UD 矩阵本质上也是由 WY 算法求得. 算法 2 给出了获取  $W_3^u$  的 WY 算法步骤.

算法 2 (获取  $W_3^u$  的 WY 算法).

步骤 1. 求取  $H$ , 使得  $H^T S_t H = I$ ;

步骤 2. 计算  $\tilde{S}_b = H^T S_b H$  和  $\tilde{S}_w = H^T S_w H$ ;

步骤 3. 对  $\tilde{S}_b$  与  $\tilde{S}_w$  实施获取 OD 矩阵的 WY 迭代算法, 得到矩阵  $\tilde{\Phi}$ ;

步骤 4. 对  $H\tilde{\Phi}$  的各个列实施单位化, 即得到  $J_1(W)$  的 UD 矩阵  $W_3^u$ .

### 2.2 $W_3^u$ 的性质

文献 [6] 证明了比值函数  $J_2(\mathbf{w})$  的 UD 矩阵  $W_2^u$  与比迹函数  $J_1(W)$  的 UD 矩阵  $W_1^u$  是相等的, 下面将证明迹比函数  $J_3(W)$  的 UD 矩阵  $W_3^u$  也等于比迹函数的 UD 矩阵  $W_1^u$ , 即有定理 1 成立.

**定理 1.**  $W_3^u = W_1^u = W_2^u$ .

**证明.** 设

$$\eta^* = \max_{\substack{\phi_i^T \phi_j = 0 \\ \|\phi_i\|=1}} \left( \frac{\text{tr}(\Phi^T \tilde{S}_b \Phi)}{\text{tr}(\Phi^T \tilde{S}_w \Phi)} \right) \quad (11)$$

由引理 3 知,

$$\tilde{\Phi} = \arg \max_{\substack{\phi_i^T \phi_j = 0 \\ \|\phi_i\|=1}} \{ \text{tr}[\Phi^T (\tilde{S}_b - \eta^* \tilde{S}_w) \Phi] \} \quad (12)$$

因此,  $\tilde{\Phi}$  是  $\tilde{S}_b - \eta^* \tilde{S}_w$  的前  $d$  个最大特征值对应的特征向量构成的列正交矩阵. 设  $\Psi$  是  $\tilde{S}_b - \eta^* \tilde{S}_w$  的所有特征向量构成的正交矩阵, 则有

$$\Psi^T (\tilde{S}_b - \eta^* \tilde{S}_w) \Psi = \Delta = \text{diag}\{\mu_1, \mu_2, \dots, \mu_n\} \quad (13)$$

令  $P = H\Psi$ , 则有

$$\begin{cases} P^T (S_b - \eta^* S_w) P = \Delta \\ P^T S_t P = P^T (S_b + S_w) P = I \end{cases} \quad (14)$$

从而

$$\begin{cases} P^T S_b P = \frac{I + \eta^* \Delta}{1 + \eta^*} \\ P^T S_w P = \frac{I - \Delta}{1 + \eta^*} \end{cases} \quad (15)$$

由引理 1 知,  $P$  的各个列是  $S_b$  相对于  $S_w$  的广义特征向量. 又  $\tilde{\Phi}$  是由  $\Psi$  的前  $d$  列构成, 因此  $J_3(W)$  的 UD 矩阵  $W_3^u$  也是由  $S_b$  相对于  $S_w$  的前  $d$  个最大广义特征值对应的单位特征向量构成的.  $\square$

定理 1 表明, 三种判别函数的 UD 矩阵尽管获取方法不同, 但所得的结果相同, 由此可将三种判别函数的 UD 矩阵统一记为  $W^u$ .

文献 [5] 证明了  $J_2(\mathbf{w})$  的 UD 向量的函数值不超过它的 OD 向量的函数值, 实际上  $J_3(W)$  也有类似的结论. 为了证明这一结论, 需要下面两个引理:

**引理 5**<sup>[12]</sup>. 设  $A = (a_{ij})$  为  $n$  阶正定阵,  $A^{-1} = (a^{ij})$ , 则  $a^{ii} \geq a_{ii}^{-1}$ .

**引理 6**<sup>[12]</sup>. 设  $A$  为  $n$  阶对称阵,  $U$  为其前  $d$  个最大特征值所对应的单位正交特征向量矩阵, 则对于任意  $n \times d$  列正交矩阵  $V$ , 有

$$\text{tr}(U^T A U) = \max_v \text{tr}(V^T A V) \quad (16)$$

**定理 2.**  $J_3(W_3^u) \leq J_3(W_3^o)$ .

**证明.** 设矩阵  $V$  是由  $S_b$  相对于  $S_w$  的所有单位广义特征向量组成的矩阵, 对应的特征值从大到小排列, 则  $V$  的前  $d$  列就是  $W_3^u$ , 由引理 1 不难得到

$$V^T (S_b - \beta^* S_w) V = \Lambda \quad (17)$$

其中,  $\Lambda$  是对角阵,  $\beta^*$  由式 (5) 定义. 对  $V$  实施 QR 分解, 即  $V = QR$ , 其中  $Q$  是列正交矩阵,  $R$  是上三角阵. 对  $R$  和其逆  $R^{-1}$  进行分块, 得

$$R = \begin{bmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{bmatrix}, R^{-1} = \begin{bmatrix} R_{11}^{-1} & -R_{11}^{-1} R_{12} R_{22}^{-1} \\ 0 & R_{22}^{-1} \end{bmatrix} \quad (18)$$

其中,  $R_{11}$  为  $d$  阶方阵. 由于  $V$  的各个列都是单位向量, 所以  $V^T V = R^T R$  的对角元全为 1, 从而  $R_{11}^T R_{11}$  的对角元也全为 1. 于是, 由式 (6) 和 (17), 得

$$\begin{aligned} 0 &= \text{tr}((W_3^o)^T (S_b - \beta^* S_w) W_3^o) = \\ &= \text{tr}((W_3^o)^T (V^T)^{-1} \Lambda V^{-1} W_3^o) = \\ &= \text{tr}((W_3^o)^T (Q^T R^T)^{-1} \Lambda (QR)^{-1} W_3^o) \geq \\ &= \text{tr}((I, 0) (R^T)^{-1} \Lambda R^{-1} (I, 0)^T) = \\ &= \text{tr}\left\{ \Lambda R^{-1} \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} (R^T)^{-1} \right\} = \\ &= \text{tr}\left\{ \Lambda \begin{bmatrix} R_{11}^{-1} (R_{11}^T)^{-1} & 0 \\ 0 & 0 \end{bmatrix} \right\} = \\ &= \text{tr}\{\Lambda_1 (R_{11}^T R_{11})^{-1}\} \geq \\ &= \text{tr}(\Lambda_1) = \text{tr}((W_3^u)^T (S_b - \beta^* S_w) W_3^u) \quad (19) \end{aligned}$$

其中,  $\Lambda_1$  是由  $\Lambda$  的前  $d$  个对角元素构成的对角阵. 式 (19) 中的第 1 个不等式利用引理 6 将  $W_3^o$  替换为  $Q(I, 0)^T$ , 第 2 个不等式利用了引理 5, 最后一个等式是因为  $W_3^u$  是由  $V$  的前  $d$  列构成. 因此

$$J_3(W_3^u) = \frac{\text{tr}((W_3^u)^T S_b W_3^u)}{\text{tr}((W_3^u)^T S_w W_3^u)} \leq \beta^* = J_3(W_3^o) \quad (20)$$

$\square$

## 3 $S_w$ 奇异时的判别矩阵

当  $S_w$  奇异时, 比迹函数  $J_1(W)$  与比值函数  $J_2(\mathbf{w})$  皆不能使用, 迹比函数  $J_3(W)$  在一定条件下也不能使用. 这时, 需要重新合理地定义这三个判别函数.

### 3.1 $S_w$ 奇异时判别函数的定义

本文采用极限的思想来定义这三个判别函数. 对  $S_w$  作特征值分解, 零特征值对应的特征向量组成

的列正交矩阵记为  $Q$ , 非零特征值对应的特征向量组成的列正交矩阵记为  $Q_1$ , 则

$$S_w = [Q, Q_1] \begin{bmatrix} 0 & 0 \\ 0 & \Lambda \end{bmatrix} \begin{bmatrix} Q^T \\ Q_1^T \end{bmatrix} = Q_1 \Lambda Q_1^T \quad (21)$$

其中,  $\Lambda$  是非零特征值构成的对角阵, 特征值从小到大排列. 现将  $S_w$  中的零特征值用参数  $\alpha (> 0)$  替代, 得到矩阵

$$S_\alpha = [Q, Q_1] \begin{bmatrix} \alpha I & 0 \\ 0 & \Lambda \end{bmatrix} \begin{bmatrix} Q^T \\ Q_1^T \end{bmatrix} = Q_1 \Lambda Q_1^T + \alpha Q Q^T \quad (22)$$

显然,  $S_\alpha$  非奇异; 而当  $\alpha \rightarrow 0$  时,  $S_\alpha \rightarrow S_w$ . 于是, 当  $S_w$  奇异时, 比迹函数、比值函数和迹比函数可分别定义为

$$\begin{cases} J_1(W) = \lim_{\alpha \rightarrow 0} J_1^\alpha(W) = \lim_{\alpha \rightarrow 0} \text{tr}[(W^T S_\alpha W)^{-1} (W^T S_b W)] \\ J_2(\mathbf{w}) = \lim_{\alpha \rightarrow 0} J_2^\alpha(\mathbf{w}) = \lim_{\alpha \rightarrow 0} \frac{\mathbf{w}^T S_b \mathbf{w}}{\mathbf{w}^T S_\alpha \mathbf{w}} \\ J_3(W) = \lim_{\alpha \rightarrow 0} J_3^\alpha(W) = \lim_{\alpha \rightarrow 0} \frac{\text{tr}(W^T S_b W)}{\text{tr}(W^T S_\alpha W)} \end{cases} \quad (23)$$

由于  $S_\alpha$  非奇异, 最大化  $J_1^\alpha(W)$ 、 $J_2^\alpha(\mathbf{w})$  和  $J_3^\alpha(W)$  可得相应的判别矩阵. 由于这些判别矩阵都是变量  $\alpha$  的函数, 因此不难看出, 当  $\alpha \rightarrow 0$  时,  $J_1^\alpha(W)$ 、 $J_2^\alpha(\mathbf{w})$  和  $J_3^\alpha(W)$  的判别矩阵的极限就是  $S_w$  奇异时  $J_1(W)$ 、 $J_2(\mathbf{w})$  和  $J_3(W)$  的判别矩阵.

### 3.2 $S_w$ 奇异时的 UD 矩阵

令  $S_{t_\alpha} = S_b + S_\alpha$ , 若非零向量  $\mathbf{w}_1, \mathbf{w}_2$  满足  $\mathbf{w}_1^T S_{t_\alpha} \mathbf{w}_2 = 0$ , 则称  $\mathbf{w}_1, \mathbf{w}_2$  关于  $S_{t_\alpha}$  共轭正交. 在式 (23) 中, 对判别函数  $J_1^\alpha(W)$  采用 GEVD 可得关于  $S_{t_\alpha}$  共轭正交判别矩阵  $W_1^{u\alpha}$ , 对  $J_2^\alpha(\mathbf{w})$  采用 UFST 可得关于  $S_{t_\alpha}$  共轭正交判别矩阵  $W_2^{u\alpha}$ , 对  $J_3^\alpha(W)$  采用算法 2 所给的 WY 算法可得关于  $S_{t_\alpha}$  共轭正交判别矩阵  $W_3^{u\alpha}$ . 由于  $S_{t_\alpha} \rightarrow S_t$ , 因此当  $\alpha \rightarrow 0$  时,  $W_1^{u\alpha}$ 、 $W_2^{u\alpha}$  和  $W_3^{u\alpha}$  的极限分别是  $S_w$  奇异时判别函数  $J_1(W)$ 、 $J_2(\mathbf{w})$  和  $J_3(W)$  的 UD 矩阵. 又由于  $S_\alpha$  非奇异, 由定理 1 知, 判别矩阵  $W_1^{u\alpha} = W_2^{u\alpha} = W_3^{u\alpha}$ , 从而它们的极限相同, 因此当  $S_w$  奇异时判别函数  $J_1(W)$ 、 $J_2(\mathbf{w})$  和  $J_3(W)$  的 UD 矩阵也是相同的. 记判别函数  $J_1^\alpha(W)$ 、 $J_2^\alpha(\mathbf{w})$  和  $J_3^\alpha(W)$  关于  $S_{t_\alpha}$  共轭正交的判别矩阵为  $W^{u\alpha}$ , 判别函数  $J_1(W)$ 、 $J_2(\mathbf{w})$  和  $J_3(W)$  的 UD 矩阵为  $W^u$ , 下面通过对  $W^{u\alpha}$  求极限来获取  $W^u$ .

采用式 (22) 记号, 令

$$H_\alpha = [Q, Q_1] \begin{bmatrix} \alpha^{-\frac{1}{2}} I_s & 0 \\ 0 & \Lambda^{-\frac{1}{2}} \end{bmatrix} \quad (24)$$

其中,  $I_s$  为  $s$  阶单位矩阵,  $s$  是零空间  $S_w^{-1}(0)$  的维数, 则

$$\begin{cases} H_\alpha^T S_\alpha H_\alpha = I \\ H_\alpha^T S_b H_\alpha = \begin{bmatrix} \alpha^{-1} Q^T S_b Q & \alpha^{-\frac{1}{2}} Q^T S_b Q_1 \Lambda^{-\frac{1}{2}} \\ \Lambda^{-\frac{1}{2}} Q_1^T S_b Q \alpha^{-\frac{1}{2}} & \Lambda^{-\frac{1}{2}} Q_1^T S_b Q_1 \Lambda^{-\frac{1}{2}} \end{bmatrix} \end{cases} \quad (25)$$

对  $H_\alpha^T S_b H_\alpha$  特征值分解, 使得

$$H_\alpha^T S_b H_\alpha = \begin{bmatrix} V_{\alpha 1} & V_{\alpha 2} \end{bmatrix} \begin{bmatrix} \alpha^{-1} \Delta_{\alpha 1} & 0 \\ 0 & \Delta_{\alpha 2} \end{bmatrix} \begin{bmatrix} V_{\alpha 1} & V_{\alpha 2} \end{bmatrix}^T \quad (26)$$

令  $\tilde{P}_\alpha = H_\alpha \begin{bmatrix} \alpha^{\frac{1}{2}} V_{\alpha 1} & V_{\alpha 2} \end{bmatrix}$ , 注意到  $S_{t_\alpha} = S_b + S_\alpha$ , 则有

$$\begin{cases} \tilde{P}_\alpha^T S_{t_\alpha} \tilde{P}_\alpha = \begin{bmatrix} \Delta_{\alpha 1} + \alpha I_s & 0 \\ 0 & \Delta_{\alpha 2} + I \end{bmatrix} \\ \tilde{P}_\alpha^T S_\alpha \tilde{P}_\alpha = \begin{bmatrix} \alpha I_s & 0 \\ 0 & I \end{bmatrix} \\ \tilde{P}_\alpha^T S_b \tilde{P}_\alpha = \begin{bmatrix} \Delta_{\alpha 1} & 0 \\ 0 & \Delta_{\alpha 2} \end{bmatrix} \end{cases} \quad (27)$$

由引理 1 知, 将  $\tilde{P}_\alpha$  的前  $d$  列单位化后组成的矩阵即  $W^{u\alpha}$ .

在式 (27) 中, 令  $\alpha \rightarrow 0$ , 得

$$\begin{cases} \tilde{P}^T S_t \tilde{P} = \begin{bmatrix} \Delta_1 & 0 \\ 0 & \Delta_2 + I \end{bmatrix} \\ \tilde{P}^T S_w \tilde{P} = \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix} \\ \tilde{P}^T S_b \tilde{P} = \begin{bmatrix} \Delta_1 & 0 \\ 0 & \Delta_2 \end{bmatrix} \end{cases} \quad (28)$$

其中,  $\tilde{P}_\alpha = H_\alpha \begin{bmatrix} \alpha^{\frac{1}{2}} V_{\alpha 1} & V_{\alpha 2} \end{bmatrix} \rightarrow \tilde{P} = [\tilde{P}_1, \tilde{P}_2]$ ,  $\Delta_{\alpha 1} \rightarrow \Delta_1$ ,  $\Delta_{\alpha 2} \rightarrow \Delta_2$ . 显然, 将  $\tilde{P}$  的前  $d$  列单位化后组成的矩阵就是所求的 UD 矩阵  $W^u$ .

进一步, 若令

$$V_{\alpha 1} = \begin{bmatrix} V_{\alpha 11} \\ V_{\alpha 21} \end{bmatrix}, \quad V_{\alpha 2} = \begin{bmatrix} V_{\alpha 12} \\ V_{\alpha 22} \end{bmatrix}$$

则由式 (25) 和 (26) 得

$$\alpha^{-1}V_{\alpha 11}\Delta_{\alpha 1}V_{\alpha 11}^T + V_{\alpha 12}\Delta_{\alpha 2}V_{\alpha 12}^T = \alpha^{-1}Q^T S_b Q \tag{29}$$

从而

$$Q^T S_b Q = \lim_{\alpha \rightarrow 0} (V_{\alpha 11}\Delta_{\alpha 1}V_{\alpha 11}^T + \alpha V_{\alpha 12}\Delta_{\alpha 2}V_{\alpha 12}^T) = \lim_{\alpha \rightarrow 0} V_{\alpha 11}\Delta_{\alpha 1}V_{\alpha 11}^T = V_{11}\Delta_1 V_{11}^T \tag{30}$$

其中,  $V_{\alpha 11} \rightarrow V_{11}$ . 式 (30) 说明  $V_{11}$  是由  $Q^T S_b Q$  的特征向量构成的正交矩阵, 而对角矩阵  $\Delta_1$  的对角元素就是  $Q^T S_b Q$  的特征值. 又

$$\begin{aligned} \tilde{P}_1 &= \lim_{\alpha \rightarrow 0} H_{\alpha}(\alpha^{\frac{1}{2}}V_{\alpha 1}) = \\ & \lim_{\alpha \rightarrow 0} [Q, Q^T] \begin{bmatrix} I_s & 0 \\ 0 & \alpha^{\frac{1}{2}}\Lambda^{-\frac{1}{2}} \end{bmatrix} \begin{bmatrix} V_{\alpha 11} \\ V_{\alpha 21} \end{bmatrix} = \\ & QV_{11} \end{aligned} \tag{31}$$

因此,  $\tilde{P}_1$  就是零空间  $S_w^{-1}(0)$  中由 NLDA 算法得到的判别矩阵.

若  $S_t$  非奇异, 注意到  $S_t = S_b + S_w$ , 由引理 1, 可得

$$\begin{cases} P^T S_t P = \begin{bmatrix} I_s & 0 \\ 0 & I \end{bmatrix} \\ P^T S_w P = \begin{bmatrix} 0 & 0 \\ 0 & I - \Delta \end{bmatrix} \\ P^T S_b P = \begin{bmatrix} I_s & 0 \\ 0 & \Delta \end{bmatrix} \end{cases} \tag{32}$$

对照式 (28) 与 (32), 并结合引理 1 知,  $\tilde{P}$  与  $P$  都是由  $S_b$  相对于  $S_t$  的广义特征向量组成的矩阵. 若记  $P = (P_1, P_2)$ , 则  $\tilde{P}_2$  与  $P_2$  的各个列单位化后二者相等, 从而  $\tilde{P}_2$  可由  $P_2$  替代, 即零空间  $S_w^{-1}(0)$  之外的 UD 矩阵可由  $S_b$  相对于  $S_t$  的 GEVD 获得. 但  $\tilde{P}_1$  不能由  $P_1$  替代, 这是因为式 (32) 中的  $P_1$  不唯一,  $\tilde{P}_1$  只是  $P_1$  的一个解, 直接作  $S_b$  相对于  $S_t$  的 GEVD 未必能得到  $\tilde{P}_1$ .

上述讨论说明, 通过取极限获取的  $S_w$  奇异时的 UD 矩阵由两部分构成: 一部分在零空间  $S_w^{-1}(0)$  中, 可由 NLDA 获得; 另一部分在零空间  $S_w^{-1}(0)$  之外, 可由  $S_b$  相对于  $S_t$  的 GEVD 获得. 基于此,  $S_w$  奇异时获取 UD 矩阵的方法称为 NLDA + GEVD 算法, 算法 3 列出了该算法的步骤.

**算法 3 ( $S_w$  奇异时获取 UD 矩阵的 NLDA + GEVD 算法).**

- 步骤 1. 对样本实施 PCA 降维, 使得  $S_t$  非奇异;
- 步骤 2. 特征值分解  $S_w$  得到  $Q$ , 再特征值分解  $Q^T S_b Q$ , 得到  $V_{11}$ , 记  $\tilde{P}_1 = QV_{11}$ ;
- 步骤 3. 计算  $S_b$  相对  $S_t$  的广义特征向量矩阵  $(P_1, P_2)$ ;

步骤 4. 用  $\tilde{P}_1$  替换  $P_1$ , 得到矩阵  $P = (\tilde{P}_1, P_2)$ , 再将  $P$  的前  $d$  个列单位化, 由它们即可组成 UD 矩阵  $W^u$ .

需要指出的是, NLDA + GEVD 算法与 NLDA + UFST 算法所得的结果不同, 这是因为 NLDA + UFST 算法是在  $S_w$  的主空间  $(S_w^{-1}(0))^{\perp}$  中求取 UD 向量, 由于主空间  $(S_w^{-1}(0))^{\perp}$  与零空间  $S_w^{-1}(0)$  正交, 从而零空间  $S_w^{-1}(0)$  中的 UD 向量与主空间  $(S_w^{-1}(0))^{\perp}$  中的 UD 向量正交, 不满足不相关性, 因此, NLDA + UFST 算法得到的不是 UD 矩阵.

**3.3  $S_w$  奇异时的 OD 矩阵**

当  $S_w$  奇异时, 尽管三种判别函数  $J_1(W)$ 、 $J_2(w)$  和  $J_3(W)$  的 UD 矩阵相同, 但它们的 OD 矩阵一般不同.

1) 对于  $J_1(W)$ , 将 NLDA + GEVD 算法得到的 UD 矩阵  $W^u$  进行 QR 分解, 即得它的 OD 矩阵  $W_1^o$ .

2) 对于  $J_2(w)$ ,  $J_2^{\alpha}(w)$  的 OD 矩阵  $W_2^{\alpha o} = (w_{21}^{\alpha o}, w_{22}^{\alpha o}, \dots, w_{2d}^{\alpha o})$  的极限  $W_2^o = (w_{21}^o, w_{22}^o, \dots, w_{2d}^o)$  即得到它的 OD 矩阵. 下面计算  $W_2^{\alpha o}$  的极限.

由式 (22) 易得,  $S_{\alpha} = \alpha Q Q^T + Q_1 \Lambda Q_1^T$ ,  $S_{\alpha}^{-1} = \alpha^{-1} Q Q Q^T + Q_1 \Lambda^{-1} Q_1^T$ ,  $Q^T Q_1 = 0$ ,  $\Lambda = Q_1^T S_w Q_1$ . 由 FST 知,  $w_{21}^{\alpha o}$  是矩阵  $S_{\alpha}^{-1} S_b$  的最大特征值对应的特征向量. 由于矩阵  $S_{\alpha}^{-1} S_b$  与矩阵  $\alpha S_{\alpha}^{-1} S_b$  具有相同的特征向量, 且当  $\alpha \rightarrow 0$  时,

$$\alpha S_{\alpha}^{-1} S_b = (Q Q^T + \alpha Q_1 \Lambda^{-1} Q_1^T) S_b \rightarrow Q Q^T S_b \tag{33}$$

因此,  $w_{21}^o$  是矩阵  $Q Q^T S_b$  的最大特征值对应的特征向量. 记  $Q^T S_b Q$  的最大特征值对应的特征向量为  $\nu_1$ , 则  $w_{21}^o = Q \nu_1$ .

令  $D_{\alpha 1} = w_{21}^{\alpha o}$ ,  $D_1 = w_{21}^o$ , 则  $w_{22}^{\alpha o}$  是矩阵  $S_{\alpha}^{-1}(I - D_{\alpha 1}(D_{\alpha 1}^T S_{\alpha}^{-1} D_{\alpha 1})^{-1} D_{\alpha 1}^T S_{\alpha}^{-1}) S_b$  的最大特征值对应的特征向量. 由于

$$\begin{aligned} \alpha S_{\alpha}^{-1}(I - D_{\alpha 1}(D_{\alpha 1}^T S_{\alpha}^{-1} D_{\alpha 1})^{-1} D_{\alpha 1}^T S_{\alpha}^{-1}) S_b &= \\ (I - (Q Q^T + \alpha Q_1 \Lambda^{-1} Q_1^T) D_{\alpha 1} (D_{\alpha 1}^T (Q Q^T + \alpha Q_1 \Lambda^{-1} Q_1^T) D_{\alpha 1})^{-1} D_{\alpha 1}) (\alpha S_{\alpha}^{-1} S_b) &\rightarrow \\ (I - Q Q^T D_1 (D_1^T Q Q^T D_1)^{-1} D_1^T) Q Q^T S_b &= \\ (Q Q^T - Q \nu_1 \nu_1^T Q^T) S_b \end{aligned} \tag{34}$$

因此,  $w_{22}^o$  是矩阵  $(Q Q^T - Q \nu_1 \nu_1^T Q^T) S_b$  的最大特征值对应的特征向量. 若记  $Q^T S_b Q$  的第二大特征值对应的特征向量为  $\nu_2$ , 则不难得到  $w_{22}^o = Q \nu_2$ . 依此类推, 设  $Q^T S_b Q$  的第  $i$  大特征值对应的特征向量为  $\nu_i$ , 则有

$$w_{2i}^o = \lim_{\alpha \rightarrow 0} w_{2i}^{\alpha o} = Q \nu_i, \quad i = 1, 2, \dots, s \tag{35}$$

即  $S_w$  奇异时, 比值函数  $J_2(w)$  的前  $s$  个 OD 向量

就是零空间  $S_w^{-1}(0)$  中由 NLDA 算法得到的 OD 向量, 其中  $s$  是零空间  $S_w^{-1}(0)$  的维数.

令  $D_{\alpha s} = (\mathbf{w}_{21}^{\alpha}, \mathbf{w}_{22}^{\alpha}, \dots, \mathbf{w}_{2s}^{\alpha})$ ,  $D_s = (\mathbf{w}_{21}^{\alpha}, \mathbf{w}_{22}^{\alpha}, \dots, \mathbf{w}_{2s}^{\alpha})$ , 则  $\mathbf{w}_{2(s+1)}^{\alpha}$  是特征方程 (36) 的最大特征值对应的特征向量

$$(I - D_{\alpha s}(D_{\alpha s}^T S_{\alpha}^{-1} D_{\alpha s})^{-1} D_{\alpha s}^T S_{\alpha}^{-1}) S_b \mathbf{w}_{2(s+1)}^{\alpha} = \lambda_{\alpha} S_{\alpha} \mathbf{w}_{2(s+1)}^{\alpha} \quad (36)$$

两边同时令  $\alpha \rightarrow 0$ , 得

$$Q_1 Q_1^T S_b \mathbf{w}_{2(s+1)}^0 = \lambda Q_1 \Lambda Q_1^T \mathbf{w}_{2(s+1)}^0 \quad (37)$$

记  $\boldsymbol{\nu}_{s+1}$  为  $Q_1^T S_b Q_1$  相对于  $Q_1^T S_w Q_1$  的最大广义特征值对应的特征向量, 则  $\mathbf{w}_{2(s+1)}^0 = Q_1 \boldsymbol{\nu}_{s+1}$ .

令  $D_{\alpha(s+1)} = (D_{\alpha s}, \mathbf{w}_{2(s+1)}^{\alpha})$ ,  $D_{s+1} = (D_s, \mathbf{w}_{2(s+1)}^0)$ , 则  $\mathbf{w}_{2(s+2)}^{\alpha}$  是特征方程 (38) 的最大特征值对应的特征向量

$$(I - D_{\alpha(s+1)}(D_{\alpha(s+1)}^T S_{\alpha}^{-1} D_{\alpha(s+1)})^{-1} D_{\alpha(s+1)}^T S_{\alpha}^{-1}) \times S_b \mathbf{w}_{2(s+2)}^{\alpha} = \lambda_{\alpha} S_{\alpha} \mathbf{w}_{2(s+2)}^{\alpha} \quad (38)$$

对其两边取极限, 得

$$\lim_{\alpha \rightarrow 0} [(I - D_{\alpha(s+1)}(D_{\alpha(s+1)}^T S_{\alpha}^{-1} D_{\alpha(s+1)})^{-1} \times D_{\alpha(s+1)}^T S_{\alpha}^{-1}) S_b \mathbf{w}_{2(s+2)}^{\alpha}] = \lim_{\alpha \rightarrow 0} [\lambda_{\alpha} S_{\alpha} \mathbf{w}_{2(s+2)}^{\alpha}] \quad (39)$$

$$\lim_{\alpha \rightarrow 0} [(I - D_{s+1}(D_{s+1}^T S_{\alpha}^{-1} D_{s+1})^{-1} \times D_{s+1}^T S_{\alpha}^{-1}) S_b \mathbf{w}_{2(s+2)}^{\alpha}] = \lim_{\alpha \rightarrow 0} [\lambda_{\alpha} S_{\alpha} \mathbf{w}_{2(s+2)}^{\alpha}] \quad (40)$$

$$\lim_{\alpha \rightarrow 0} [(I - Q Q^T - Q_1 \boldsymbol{\nu}_{s+1} (\boldsymbol{\nu}_{s+1}^T \Lambda^{-1} \boldsymbol{\nu}_{s+1})^{-1}) \times \boldsymbol{\nu}_{s+1}^T \Lambda^{-1} Q_1^T] S_b \mathbf{w}_{2(s+2)}^{\alpha}] = \lim_{\alpha \rightarrow 0} [\lambda_{\alpha} S_{\alpha} \mathbf{w}_{2(s+2)}^{\alpha}] \quad (41)$$

$$(Q_1 Q_1^T - Q_1 \boldsymbol{\nu}_{s+1} (\boldsymbol{\nu}_{s+1}^T \Lambda^{-1} \boldsymbol{\nu}_{s+1})^{-1}) \boldsymbol{\nu}_{s+1}^T \Lambda^{-1} Q_1^T \times S_b \mathbf{w}_{2(s+2)}^0 = \lambda_{s+2} Q_1 \Lambda Q_1^T \mathbf{w}_{2(s+2)}^0 \quad (42)$$

记  $\boldsymbol{\nu}_{s+2}$  为  $(I - \boldsymbol{\nu}_{s+1} (\boldsymbol{\nu}_{s+1}^T (Q_1^T S_w Q_1)^{-1} \boldsymbol{\nu}_{s+1})^{-1} \boldsymbol{\nu}_{s+1}^T) \times (Q_1^T S_w Q_1)^{-1} Q_1^T S_b Q_1$  相对于  $Q_1^T S_w Q_1$  的最大广义特征值对应的特征向量, 则  $\mathbf{w}_{2(s+2)}^0 = Q_1 \boldsymbol{\nu}_{s+2}$ .

类似地, 记  $\boldsymbol{\nu}_{s+i}$  为  $(I - H_i (H_i^T (Q_1^T S_w Q_1)^{-1} \times H_i)^{-1} H_i^T (Q_1^T S_w Q_1)^{-1}) Q_1^T S_b Q_1$  相对于  $Q_1^T S_w Q_1$  的最大广义特征值对应的特征向量, 其中  $H_i = (\boldsymbol{\nu}_{s+1}, \dots, \boldsymbol{\nu}_{s+i-1})$ , 则

$$\mathbf{w}_{2(s+i)}^0 = \lim_{\alpha \rightarrow 0} \mathbf{w}_{2(s+i)}^{\alpha} = Q_1 \boldsymbol{\nu}_{s+i}, \quad i = 1, 2, \dots, d - s \quad (43)$$

不难看出,  $\boldsymbol{\nu}_{s+i}$  是对判别函数

$$J_{2Q_1}(\boldsymbol{\nu}) = \frac{\boldsymbol{\nu}^T Q_1^T S_b Q_1 \boldsymbol{\nu}}{\boldsymbol{\nu}^T Q_1^T S_w Q_1 \boldsymbol{\nu}} \quad (44)$$

实施 OFST 得到的第  $i$  个向量, 因此在零空间  $S_w^{-1}(0)$  之外的判别向量是由 OFST 得到的; 由于  $Q$  与  $Q_1$  正交, 所以这些判别向量都在主空间  $(S_w^{-1}(0))^{\perp}$  中.

由上述讨论可知,  $J_{2\alpha}(\mathbf{w})$  的 OD 矩阵的极限就是由 NLDA + OFST 算法获得的 OD 矩阵, 这实际上从理论上说明了双子空间算法的合理性.

3) 对于  $J_3(W)$ , 当  $d > \dim(S_w^{-1}(0))$  时, 可由 WY 算法直接得到 OD 矩阵  $W_3^{\alpha}$ . 当  $d \leq \dim(S_w^{-1}(0))$  时, 不难看出  $J_3(W)$  的函数值最大可以达到  $+\infty$ . 因此, 若记

$$\beta^{\alpha} = \frac{\text{tr}((W_3^{\alpha})^T S_b W_3^{\alpha})}{\text{tr}((W_3^{\alpha})^T S_{\alpha} W_3^{\alpha})} = \max_{\substack{\mathbf{w}_i^T \mathbf{w}_j = 0 \\ \|\mathbf{w}_i\| = 1}} \left( \frac{\text{tr}(W^T S_b W)}{\text{tr}(W^T S_{\alpha} W)} \right) \quad (45)$$

则当  $\alpha \rightarrow 0$  时,  $\beta^{\alpha} \rightarrow +\infty$ . 由引理 3 知,  $W_3^{\alpha}$  是矩阵  $(S_b - \beta^{\alpha} S_{\alpha})$  前  $d$  个最大特征值对应的特征向量. 由于矩阵  $(S_b - \beta^{\alpha} S_{\alpha})$  与矩阵  $(S_b / \beta^{\alpha} - S_{\alpha})$  具有相同的特征向量, 且当  $\alpha \rightarrow 0$  时,  $(S_b / \beta^{\alpha} - S_{\alpha}) \rightarrow -S_w$ , 因此  $W_3^{\alpha}$  中的  $d$  个判别向量  $\mathbf{w}_{31}^{\alpha}, \mathbf{w}_{32}^{\alpha}, \dots, \mathbf{w}_{3d}^{\alpha}$  全部落在零空间  $S_w^{-1}(0)$  中. 这与文献 [13] 结论相同, 但零空间  $S_w^{-1}(0)$  中的列正交矩阵并不唯一,  $W_3^{\alpha}$  只是其中的一个解.

令  $W_3^{\alpha} = QV$ , 其中  $Q$  由式 (21) 定义,  $V = (\boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_d)$  是列正交矩阵. 下面证明  $V$  是矩阵  $Q^T S_b Q$  的前  $d$  个最大特征值对应的特征向量. 首先

$$\alpha \beta^{\alpha} \rightarrow \text{tr}((W_3^{\alpha})^T S_b W_3^{\alpha}) = \eta \quad (46)$$

若令  $\mathbf{w}_{3i}^{\alpha} = \mathbf{w}_{3i}^0 + \boldsymbol{\gamma}_{i\alpha}$ , 则

$$\boldsymbol{\gamma}_{i\alpha} \rightarrow 0, \quad \beta^{\alpha} \boldsymbol{\gamma}_{i\alpha} \rightarrow 0 \quad (47)$$

其次

$$(S_b - \beta^{\alpha} S_{\alpha}) \mathbf{w}_{3i}^{\alpha} = \lambda_{i\alpha} \mathbf{w}_{3i}^{\alpha} \quad (48)$$

$$(S_b - \beta^{\alpha} S_{\alpha})(Q\boldsymbol{\nu}_i + \boldsymbol{\gamma}_{i\alpha}) = \lambda_{i\alpha}(Q\boldsymbol{\nu}_i + \boldsymbol{\gamma}_{i\alpha}) \quad (49)$$

两边左乘  $Q^T$ , 并将式 (22) 代入式 (49), 得

$$Q^T [S_b - \beta^{\alpha} (\alpha Q Q^T + Q_1 \Lambda Q_1^T)] (Q\boldsymbol{\nu}_i + \boldsymbol{\gamma}_{i\alpha}) = \lambda_{i\alpha} Q^T (Q\boldsymbol{\nu}_i + \boldsymbol{\gamma}_{i\alpha}) \quad (50)$$

化简, 得

$$Q^T S_b Q \boldsymbol{\nu}_i - (\alpha \beta^{\alpha} \boldsymbol{\nu}_i + \beta^{\alpha} Q^T \boldsymbol{\gamma}_{i\alpha}) = \lambda_{i\alpha} \boldsymbol{\nu}_i + (Q^T \boldsymbol{\gamma}_{i\alpha}) \quad (51)$$

令  $\alpha \rightarrow 0$ , 得

$$Q^T S_b Q \nu_i = (\lambda_i + \eta) \nu_i \quad (52)$$

由此说明, 当  $d \leq \dim(S_w^{-1}(0))$  时,  $J_3(W)$  的 OD 矩阵就是由 NLDA 算法得到的判别矩阵.

综合第 3.2 节与第 3.3 节关于三种判别函数在  $S_w$  奇异时的判别矩阵的讨论结果, 不难得到:

**定理 3.** 当判别向量的个数  $d \leq \dim(S_w^{-1}(0))$  时, 三种判别函数的 6 个判别矩阵是同一个矩阵, 它就是由 NLDA 算法得到的判别矩阵.

特别地, 当样本被 PCA 降维至  $N - 1$  时, 零空间  $S_w^{-1}(0)$  中含有  $c - 1$  个判别向量, 若取  $d = c - 1$ , 则 NLDA 算法就是唯一最优算法.

### 4 实验

本文将分别在 Yale 人脸库和 Feret 人脸库<sup>[14]</sup>上验证文中的结论. Yale 人脸库含有 15 个人的 165 幅图像, 每人有 11 幅图像, 图像均为 100 像素  $\times$  100 像素. 实验时, 每人随机取 5 幅共 75 幅图像作为训练集, 余下 90 幅图像作为测试集. Feret 数据库中的一个子库包含 1010 人, 每人 2 幅图像, 只取前 200 人共 400 幅图像用于实验, 实验前手工标定两眼中心进行校正, 并将每幅图像裁剪成 128 像素  $\times$  128 像素. 实验时, 随机取 100 人的 200 幅图像作为训练集, 余下 100 人的 200 幅图像作为测试集. 每个实验都做 10 轮, 取 10 轮结果的平均作为最后的结果. 在实施 LDA 之前先用 PCA 将样本降维, 记样本数为  $N$ , 降维后样本的维数为  $n$ , 取投影轴的个数  $d = \text{rank}(S_b)$ . 采用最近邻分类器, 以余弦距离作为距离度量.

#### 4.1 对定理 1 的验证

欲验证  $J_3(W)$  的 UD 矩阵  $W_3^u$  与  $J_1(W)$  的 UD 矩阵  $W_1^u$  和  $J_2(w)$  的 UD 矩阵  $W_2^u$  相同, 只需验证矩阵  $W_3^u$  与  $W_1^u, W_2^u$  的距离为零. 在 Yale 人脸库上进行实验, 取  $n = 20, 30, 40, 50, 60$  等 5 个不同的值, 这时矩阵  $S_w$  非奇异, 表 1 给出了实验结果, 表 1 中的  $\|\cdot\|$  表示矩阵的 Frobenius 范数. 由表 1 可见, 尽管矩阵  $W_3^u$  与  $W_1^u, W_2^u$  的距离不为零, 但距离值已非常小, 考虑到计算误差, 可以认为三种判别函数的 UD 矩阵是相同的.

#### 4.2 对定理 2 的验证

为验证  $J_3(W_3^u) \leq J_3(W_3^o)$ , 依然在 Yale 人脸库上进行实验. 取  $n = 20, 30, 40, 50, 60$  等 5 个不同的值, 实验结果见表 2. 由表 2 可见,  $J_3(W)$  的 UD 矩阵的函数值不超过它的 OD 矩阵的函数值.

表 1 Yale 人脸库上  $W_1^u, W_2^u, W_3^u$  之间的距离  
Table 1 The distances among  $W_1^u, W_2^u$ , and  $W_3^u$  on Yale face database

$J_3(W)$	$\ W_3^u - W_1^u\ $	$\ W_3^u - W_2^u\ $
$n = 20$	$6.9417 \times 10^{-14}$	$5.8693 \times 10^{-14}$
$n = 30$	$6.8088 \times 10^{-14}$	$6.7884 \times 10^{-14}$
$n = 40$	$1.2418 \times 10^{-13}$	$1.7141 \times 10^{-13}$
$n = 50$	$1.7973 \times 10^{-13}$	$2.1578 \times 10^{-13}$
$n = 60$	$3.9103 \times 10^{-11}$	$1.5181 \times 10^{-11}$

表 2  $W_3^u$  与  $W_3^o$  在 Yale 人脸库上述比函数值的比较  
Table 2 Comparison of the trace-ratio function values between  $W_3^u$  and  $W_3^o$  on Yale face database

$J_3(W)$	$W_3^u$	$W_3^o$
$n = 20$	3.7852	5.2126
$n = 30$	9.2748	11.2145
$n = 40$	16.3735	19.8646
$n = 50$	31.4948	37.6231
$n = 60$	75.1909	90.1838

#### 4.3 NLDA + GEVD 算法与 NLDA + UFST 算法的比较

若要对 NLDA + GEVD 算法得到的判别矩阵  $W^u$  与 NLDA + UFST 算法得到的判别矩阵  $W^y$  进行比较, 实际上只需比较它们的比值函数值即可. 在 Yale 人脸库上以  $n = N - 10$  为例进行实验, 此时  $S_w$  奇异, 有 5 个判别向量在零空间  $S_w^{-1}(0)$  中, 余下的 9 个判别向量在零空间  $S_w^{-1}(0)$  外, 它们的比值函数值见表 3. 从表 3 可以看出, 在零空间  $S_w^{-1}(0)$  中两种算法得到的 5 个判别向量是一样的, 这是因为它们都是由 NLDA 算法得到的. 但在零空间  $S_w^{-1}(0)$  之外两种算法得到的 9 个判别向量是不同的, 这是因为 NLDA + GEVD 算法得到的判别向量满足不相关性, 而 NLDA + UFST 算法得到的判别向量不满足不相关性, 即  $W^u$  是 UD 矩阵, 而  $W^y$  不是.

#### 4.4 6 种判别矩阵的判别性能比较

由定理 1 和第 4.1 节的结论知, 三种判别函数实际上只得到了 1 个 UD 矩阵和 3 个 OD 矩阵共 4 个不同的判别矩阵. 为比较它们的判别性能, 分别在 Yale 人脸库和 Feret 人脸库上进行实验. 实验时, 让  $n$  从 1 变化到  $N - 1$ , 记录下正确识别率, 结果见图 1 和图 2. 由图 1 和图 2 可见, 三种 OD 矩阵的判别性能较接近, 但 UD 矩阵的判别性能与它们有较大差别. 由定理 2 知, UD 矩阵的函数值不超过 OD 矩

阵, 虽然在 Yale 人脸库上 UD 矩阵的判别性能弱于 OD 矩阵, 但在 Feret 人脸库上 UD 矩阵的判别性能却强于 OD 矩阵, 这说明判别矩阵的函数值并不能准确反映判别性能.

表 3 NLDA + GEVD 算法与 NLDA + UFST 算法的比较  
Table 3 Comparison between NLDA + GEVD algorithm and NLDA + UFST algorithm

$J_2(\mathbf{w}_i)$ 值	$W^u$	$W^y$
$i = 1$	43 793 492.7781	43 793 492.7781
$i = 2$	21 469 744.1039	21 469 744.1039
$i = 3$	16 020 838.0020	16 020 838.0020
$i = 4$	12 857 366.4798	12 857 366.4798
$i = 5$	8 894 362.4360	8 894 362.4360
$i = 6$	3 260 446.3555	3 260 975.4491
$i = 7$	695 461.5540	695 750.6067
$i = 8$	774.5882	1 138.5426
$i = 9$	350.5667	451.1253
$i = 10$	163.2153	194.5261
$i = 11$	87.4238	107.0164
$i = 12$	54.6687	66.9784
$i = 13$	32.0990	39.9374
$i = 14$	19.3317	23.0465
正确识别率 (%)	97.8889	97.7778

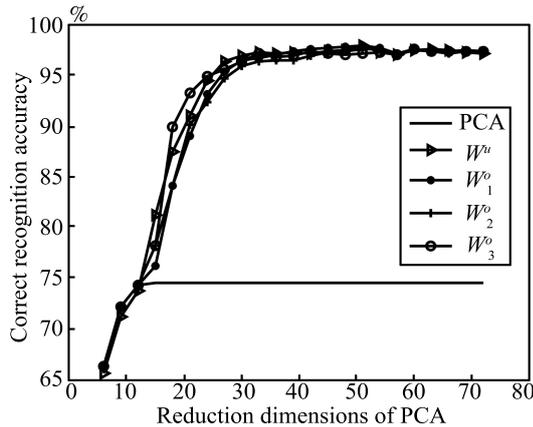


图 1 判别矩阵在 Yale 人脸库上的判别性能比较  
Fig. 1 Comparison of discriminant performances of discriminant matrixes on Yale face database

### 5 总结

本文对比迹函数  $J_1(W)$ 、比值函数  $J_2(\mathbf{w})$  和迹比函数  $J_3(W)$  的 6 种判别矩阵进行了系统分析. 当

$S_w$  非奇异时, 讨论了它们的区别与联系; 当  $S_w$  奇异时, 采用极限的思想讨论了它们的获取方法.

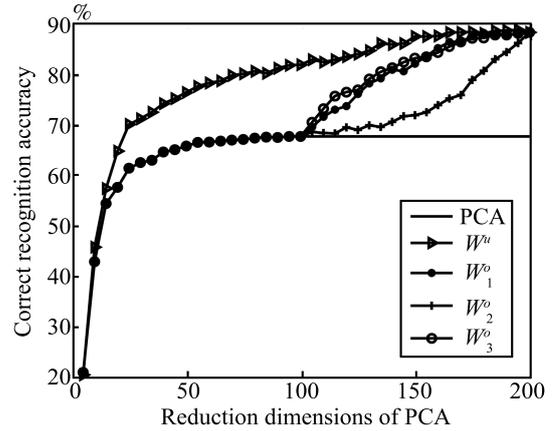


图 2 判别矩阵在 Feret 人脸库上的判别性能比较  
Fig. 2 Comparison of discriminant performances of discriminant matrixes on Feret face database

### References

- 1 Rao C R. The utilization of multiple measurements in problems of biological classification. *Journal of the Royal Statistical Society, Series B*, 1948, **10**(2): 159–203
- 2 Foley D H, Sammon J W. An optimal set of discriminant vectors. *IEEE Transactions on Computers*, 1975, **24**(3): 281–289
- 3 Wang H, Yan S C, Xu D, Tang X O, Huang T. Trace ratio vs. ratio trace for dimensionality reduction. In: *Proceedings of the Conference on Computer Vision and Pattern Recognition*. Minneapolis, USA: IEEE, 2007. 1–8
- 4 Jin Z, Yang J Y, Hu Z S, Lou Z. Face recognition based on uncorrelated discriminant transformation. *Pattern Recognition*, 2001, **34**(7): 1405–1416
- 5 Cheng Zheng-Dong, Zhang Yu-Jin, Fan Xiang. A new generalized form of fisher linear discriminant function. *Pattern Recognition and Artificial Intelligence*, 2009, **22**(2): 176–181  
(程正东, 章毓晋, 樊祥. Fisher 线性鉴别函数的一种新推广形式. *模式识别与人工智能*, 2009, **22**(2): 176–181)
- 6 Yang Jian, Yang Jing-Yu, Liu Ning-Zhong. Theory and algorithm of uncorrelated optimal discriminant analysis. *Journal of Nanjing University of Science and Technology*, 2002, **26**(2): 179–182  
(杨健, 杨静宇, 刘宁钟. 统计不相关最优鉴别分析的理论及算法. *南京理工大学学报*, 2002, **26**(2): 179–182)
- 7 Chen L F, Liao H Y M, Ko M T, Lin J C, Yu G J. A new LDA-based face recognition system which can solve the small sample size problem. *Pattern Recognition*, 2000, **33**(10): 1713–1726
- 8 Yang J, Yang J Y. An optimal FLD algorithm for facial feature extraction. In: *Proceedings of the Intelligent Robots and Computer Vision XX: Algorithms, Techniques, and Active Vision*. Boston, USA: SPIE, 2001. 438–444
- 9 Zhang S, Sim T. Discriminant subspace analysis: a Fukunaga-Koontz approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, **29**(10): 1732–1745

- 10 Zheng W M, Zhao L, Zou C R. An efficient algorithm to solve the small sample size problem for LDA. *Pattern Recognition*, 2004, **37**(5): 1077–1079
- 11 Guo Y F, Wu L D, Lu H, Feng Z, Xue X Y. Null Foley-Sammon transform. *Pattern Recognition*, 2006, **39**(11): 2248–2251
- 12 Wang Song-Gui, Wu Mi-Xia, Jia Zhong-Zhen. *Matrix Inequalities (Second Edition)*. Beijing: Science Press, 2006. 148–170  
(王松桂, 吴密霞, 贾忠贞. 矩阵不等式 (第二版). 北京: 科学出版社, 2006. 148–170)
- 13 Liu J, Chen S C, Tan X Y. A study on three linear discriminant analysis based methods in small sample size problem. *Pattern Recognition*, 2008, **41**(1): 102–116
- 14 Phillips P J, Moon H, Rizvi S A, Rauss P J. The FERET evaluation methodology for face recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, **22**(10): 1090–1104



程正东 合肥电子工程学院博士. 主要研究方向为模式识别和图像处理. 本文通信作者.  
E-mail: czd06@mails.tsinghua.edu.cn  
(**CHENG Zheng-Dong** Ph. D. at Electronic Engineering Institute of Hefei. His research interest covers pattern recognition and image processing.

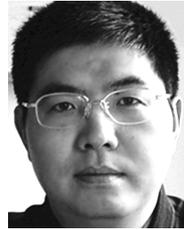
Corresponding author of this paper.)



章毓晋 清华大学电子工程系教授. 主要研究方向为图像工程 (图像处理、图像分析、图像理解及其技术应用).

E-mail: zhangyj@ee.tsinghua.edu.cn

(**ZHANG Yu-Jin** Professor in the Department of Electronic Engineering, Tsinghua University. His research interest covers image engineering which includes image processing, image analysis, and image understanding, as well as their applications.)



樊祥 合肥电子工程学院教授. 主要研究方向为光电对抗和信号处理.

E-mail: fanxianglxl@163.com

(**FAN Xiang** Professor at Electronic Engineering Institute of Hefei. His research interest covers electro-optical countermeasures and signal processing.)



朱斌 合肥电子工程学院博士. 主要研究方向为光电信号处理和红外图像处理. E-mail: zhubin0201@163.com

(**ZHU Bin** Ph. D. at Electronic Engineering Institute of Hefei. His research interest covers photoelectric signal processing and infrared image processing.)