Implementations of Fisher's Linear Discriminant Analysis from the Numerical Point of View

Pavel Schlesinger [1]
joint work with Jurjen Duintjer Tebbens [2]

[1]Institute of Formal and Applied Linguistics
Faculty of Mathematics and Physics
Charles University, Prague

[2]Institute of Computer Science
Academy of Sciences of the Czech Republic, Prague

CSDA 2005, Limassol
October 29, 2005

Outline

Outline

Motivation Example

- protein folding classification into 42 groups
- frequencies of 20 amino acids as predictors
  - 20 singles, 400 pairs, 8000 triples, . . .
- expensive data-collection
  - just hundreds of examples - 268
- classical $p \gg n$ issue in microarray, document classification etc.

Problems

- matrix storage
- computational cost
  - matrix multiplications, inversions
  - computing inner products
  - optimalization, QP, . . .
- . . . no straightforward using of $p < n$ methods

Motivation Example

- protein folding classification into 42 groups
- frequencies of 20 amino acids as predictors
  - 20 singles, 400 pairs, 8000 triples, ...
- expensive data-collection
  - just hundreds of examples - 268
- classical $p \gg n$ issue in microarray, document classification etc.

Problems

- matrix storage
- computational cost
  - matrix multiplications, inversions
  - computing inner products
  - optimalization, QP, ...
- ...no straightforward using of $p < n$ methods

Outline

Fisher's Criterion I

- variance . . . $\text{Var}(\boldsymbol{X}) = \boldsymbol{\Sigma} = \text{E}_{\boldsymbol{X}}\left[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^T\right]$
- decomposition (between- and within-variance)
  $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_B + \boldsymbol{\Sigma}_W$

$$\boldsymbol{\Sigma}_B = \sum_{k=1}^{K} \pi_k (\boldsymbol{\mu}_k - \boldsymbol{\mu})(\boldsymbol{\mu}_k - \boldsymbol{\mu})^T$$

$$\boldsymbol{\Sigma}_W = \sum_{k=1}^{K} \pi_k \text{E}_{\boldsymbol{X}|G}\left[(\boldsymbol{X}-\boldsymbol{\mu}_k)(\boldsymbol{X}-\boldsymbol{\mu}_k)^T\right]$$

- total variance after projection . . .
  $\text{Var}(\mathbf{c}^T\boldsymbol{X}) = \mathbf{c}^T\boldsymbol{\Sigma}_B\mathbf{c} + \mathbf{c}^T\boldsymbol{\Sigma}_W\mathbf{c}$
- Fisher's criterion

$$\max_{\mathbf{c}\in\mathbb{R}^p} \frac{\mathbf{c}^T\boldsymbol{\Sigma}_B\mathbf{c}}{\mathbf{c}^T\boldsymbol{\Sigma}_W\mathbf{c}} \quad \text{s. t.} \quad \mathbf{c} \neq \mathbf{0}$$

Fisher's Criterion II

- generalized eigenproblem with matrix pencil $(\mathbf{\Sigma}_B, \mathbf{\Sigma}_W)$
- (ordered) ν-eigenvectors give projected matrix $\mathbf{C} = \mathbf{C}_\nu$

1. estimate
   - $\mu, \mu_k, \mathbf{\Sigma}_B, \mathbf{\Sigma}_W$

$$\boldsymbol{B} \equiv \frac{(\boldsymbol{GM} - \mathbf{1}\bar{\mathbf{x}})^T(\boldsymbol{GM} - \mathbf{1}\bar{\mathbf{x}})}{g-1}, \quad \boldsymbol{W} \equiv \frac{(\mathrm{X} - \boldsymbol{GM})^T(\mathrm{X} - \boldsymbol{GM})}{n-g}$$

   - $\boldsymbol{B}, \boldsymbol{W} \in \mathbb{R}^{p \times p}$, $\boldsymbol{B}, \boldsymbol{W} \geq 0$
   - $\mathrm{rank}(\boldsymbol{B}) \leq K - 1$, $r = \mathrm{rank}(\boldsymbol{W}) \leq \min\{n, p\}$
2. solve $\boldsymbol{B}\mathbf{c} - \lambda\boldsymbol{W}\mathbf{c} = 0$ to obtain $\mathbf{C}$
   - find $K - 1$ largest eigenpairs
3. compute
   - all distances $||\widetilde{\mathbf{x}} - \widetilde{\mu}_k||^2 = ||\mathbf{x} - \mu_k||^2_{\mathbf{CC}^T}$
   - find group label through minimum

Fisher's Criterion II

- generalized eigenproblem with matrix pencil $(\boldsymbol{\Sigma}_B, \boldsymbol{\Sigma}_W)$
- (ordered) ν-eigenvectors give projected matrix $\mathbf{C} = \mathbf{C}_\nu$

1. estimate
   - $\boldsymbol{\mu}, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_B, \boldsymbol{\Sigma}_W$

$$\boldsymbol{B} \equiv \frac{(\boldsymbol{GM} - \mathbf{1}\bar{\boldsymbol{x}})^T(\boldsymbol{GM} - \mathbf{1}\bar{\boldsymbol{x}})}{g-1}, \quad \boldsymbol{W} \equiv \frac{(\mathrm{X} - \boldsymbol{GM})^T(\mathrm{X} - \boldsymbol{GM})}{n-g}$$

   - $\boldsymbol{B}, \boldsymbol{W} \in \mathbb{R}^{p \times p}$, $\boldsymbol{B}, \boldsymbol{W} \geq 0$
   - $\operatorname{rank}(\boldsymbol{B}) \leq K - 1$, $r = \operatorname{rank}(\boldsymbol{W}) \leq \min\{n, p\}$
2. solve $\boldsymbol{B}\mathbf{c} - \lambda\boldsymbol{W}\mathbf{c} = 0$ to obtain $\mathbf{C}$
   - find $K - 1$ largest eigenpairs
3. compute
   - all distances $||\widetilde{\boldsymbol{x}} - \widetilde{\boldsymbol{\mu}}_k||^2 = ||\boldsymbol{x} - \boldsymbol{\mu}_k||^2_{\mathbf{CC}^T}$
   - find group label through minimum

Outline

The $p \gg n$ case

- if $\boldsymbol{W}$ nonsingular $\Rightarrow$ transformation to a standard eigenproblem, e. g.

$$(\boldsymbol{W}^{-1}\boldsymbol{B} - \lambda\mathbf{I})\mathbf{c} = 0$$

- in the $p \gg n$ case $\boldsymbol{W}$ is singular $\Rightarrow$
  - transformation not possible
  - very challenging eigenproblem
    - may even happen that

$$\det(\boldsymbol{B} - \lambda\boldsymbol{W}) = 0 \qquad \forall\,\lambda \in \mathbb{C}\,!$$

pair $\{\boldsymbol{B}, \boldsymbol{W}\}$ is called singular matrix pencil

Outline

Underlying Linear Algebra

- Generalized Schur Decomposition [Moler, Stewart - 1973]:

$$\boldsymbol{Q}^T(\boldsymbol{B}-\boldsymbol{W})\boldsymbol{Z} = \boldsymbol{T} - \boldsymbol{S}$$

  - $\boldsymbol{Q}$, $\boldsymbol{Z}$ orthogonal, $\boldsymbol{T}$, $\boldsymbol{S}$ upper triangular
  - singularity $\Rightarrow$ possible zeros on main diagonals of $\boldsymbol{T}$ and $\boldsymbol{S}$

1. $t_{ii} \neq 0 \neq s_{ii} \Rightarrow \det(\boldsymbol{T} - \frac{t_{ii}}{s_{ii}}\boldsymbol{S}) = 0$, $\frac{t_{ii}}{s_{ii}}$ …finite eigenvalue
2. $t_{ii} = 0$, $s_{ii} \neq 0 \Rightarrow \det(\boldsymbol{T} - 0 \cdot \boldsymbol{S}) = 0$, 0 …finite —"—
3. $t_{ii} \neq 0$, $s_{ii} = 0 \Rightarrow$ "$\det(\boldsymbol{T} - \infty\boldsymbol{S}) = 0$", "infinite" —"—
4. $t_{ii} = 0 = s_{ii} \Rightarrow \det(\boldsymbol{T} - \lambda\boldsymbol{S}) = 0$, $\forall\lambda \in \mathbb{C}$

Underlying Linear Algebra

- Generalized Schur Decomposition [Moler, Stewart - 1973]:

$$\boldsymbol{Q}^T(\boldsymbol{B}-\boldsymbol{W})\boldsymbol{Z} = \boldsymbol{T} - \boldsymbol{S}$$

  - $\boldsymbol{Q}$, $\boldsymbol{Z}$ orthogonal, $\boldsymbol{T}$, $\boldsymbol{S}$ upper triangular
  - singularity $\Rightarrow$ possible zeros on main diagonals of $\boldsymbol{T}$ and $\boldsymbol{S}$

1. $t_{ii} \neq 0 \neq s_{ii} \Rightarrow \det(\boldsymbol{T} - \frac{t_{ii}}{s_{ii}}\boldsymbol{S}) = 0$, $\frac{t_{ii}}{s_{ii}}$ ...finite eigenvalue
2. $t_{ii} = 0$, $s_{ii} \neq 0 \Rightarrow \det(\boldsymbol{T} - 0 \cdot \boldsymbol{S}) = 0$, 0 ...finite —"—
3. $t_{ii} \neq 0$, $s_{ii} = 0 \Rightarrow$ "$\det(\boldsymbol{T} - \infty \boldsymbol{S}) = 0$", "infinite" —"—
4. $t_{ii} = 0 = s_{ii} \Rightarrow \det(\boldsymbol{T} - \lambda \boldsymbol{S}) = 0$, $\forall \lambda \in \mathbb{C}$

- how to determine the $K-1$ largest eigenvalues???

Underlying Linear Algebra

- Generalized Schur Decomposition [Moler, Stewart - 1973]:

$$Q^T(B - W)Z = T - S$$

- Q, Z orthogonal, T, S upper triangular
- singularity $\Rightarrow$ possible zeros on main diagonals of T and S

1. $t_{ii} \neq 0 \neq s_{ii} \Rightarrow \det(T - \frac{t_{ii}}{s_{ii}} S) = 0$, $\frac{t_{ii}}{s_{ii}}$ ...finite eigenvalue
2. $t_{ii} = 0$, $s_{ii} \neq 0 \Rightarrow \det(T - 0 \cdot S) = 0$, 0 ...finite —"—
3. $t_{ii} \neq 0$, $s_{ii} = 0 \Rightarrow$ "$\det(T - \infty S) = 0$", "infinite" —"—
4. $t_{ii} = 0 = s_{ii} \Rightarrow \det(T - \lambda S) = 0, \forall \lambda \in \mathbb{C}$

- how to determine the $K - 1$ largest eigenvalues???

Underlying Linear Algebra

- Generalized Schur Decomposition [Moler, Stewart - 1973]:

$$\boldsymbol{Q}^T(\boldsymbol{B}-\boldsymbol{W})\boldsymbol{Z} = \boldsymbol{T} - \boldsymbol{S}$$

  - $\boldsymbol{Q}$, $\boldsymbol{Z}$ orthogonal, $\boldsymbol{T}$, $\boldsymbol{S}$ upper triangular
  - singularity $\Rightarrow$ possible zeros on main diagonals of $\boldsymbol{T}$ and $\boldsymbol{S}$

1. $t_{ii} \neq 0 \neq s_{ii} \Rightarrow \det(\boldsymbol{T} - \frac{t_{ii}}{s_{ii}}\boldsymbol{S}) = 0$, $\frac{t_{ii}}{s_{ii}}$ . . . finite eigenvalue
2. $t_{ii} = 0$, $s_{ii} \neq 0 \Rightarrow \det(\boldsymbol{T} - 0 \cdot \boldsymbol{S}) = 0$, 0 . . . finite —"—
3. $t_{ii} \neq 0$, $s_{ii} = 0 \Rightarrow$ "$\det(\boldsymbol{T} - \infty\boldsymbol{S}) = 0$", "infinite" —"—
4. $t_{ii} = 0 = s_{ii} \Rightarrow \det(\boldsymbol{T} - \lambda\boldsymbol{S}) = 0, \forall\lambda \in \mathbb{C}$

  - how to determine the $K-1$ largest eigenvalues???

Underlying Linear Algebra

- Generalized Schur Decomposition [Moler, Stewart - 1973]:

$$\boldsymbol{Q}^T(\boldsymbol{B}-\boldsymbol{W})\boldsymbol{Z} = \boldsymbol{T} - \boldsymbol{S}$$

  - $\boldsymbol{Q}, \boldsymbol{Z}$ orthogonal, $\boldsymbol{T}, \boldsymbol{S}$ upper triangular
  - singularity $\Rightarrow$ possible zeros on main diagonals of $\boldsymbol{T}$ and $\boldsymbol{S}$

1. $t_{ii} \neq 0 \neq s_{ii} \Rightarrow \det(\boldsymbol{T} - \frac{t_{ii}}{s_{ii}}\boldsymbol{S}) = 0$, $\frac{t_{ii}}{s_{ii}}$ ... finite eigenvalue
2. $t_{ii} = 0$, $s_{ii} \neq 0 \Rightarrow \det(\boldsymbol{T} - 0 \cdot \boldsymbol{S}) = 0$, 0 ... finite —"—
3. $t_{ii} \neq 0$, $s_{ii} = 0 \Rightarrow$ "$\det(\boldsymbol{T} - \infty\boldsymbol{S}) = 0$", "infinite" —"—
4. $t_{ii} = 0 = s_{ii} \Rightarrow \det(\boldsymbol{T} - \lambda\boldsymbol{S}) = 0, \forall\lambda \in \mathbb{C}$

  - how to determine the $K-1$ largest eigenvalues???

Underlying Linear Algebra

- Generalized Schur Decomposition [Moler, Stewart - 1973]:

$$\boldsymbol{Q}^T(\boldsymbol{B}-\boldsymbol{W})\boldsymbol{Z}=\boldsymbol{T}-\boldsymbol{S}$$

  - $\boldsymbol{Q}$, $\boldsymbol{Z}$ orthogonal, $\boldsymbol{T}$, $\boldsymbol{S}$ upper triangular
  - singularity $\Rightarrow$ possible zeros on main diagonals of $\boldsymbol{T}$ and $\boldsymbol{S}$

1. $t_{ii} \neq 0 \neq s_{ii} \Rightarrow \det(\boldsymbol{T}-\frac{t_{ii}}{s_{ii}}\boldsymbol{S})=0$, $\frac{t_{ii}}{s_{ii}}$ ...finite eigenvalue
2. $t_{ii}=0$, $s_{ii}\neq 0 \Rightarrow \det(\boldsymbol{T}-0\cdot\boldsymbol{S})=0$, 0 ...finite —"—
3. $t_{ii}\neq 0$, $s_{ii}=0 \Rightarrow$ "$\det(\boldsymbol{T}-\infty\boldsymbol{S})=0$", "infinite" —"—
4. $t_{ii}=0=s_{ii} \Rightarrow \det(\boldsymbol{T}-\lambda\boldsymbol{S})=0$, $\forall\lambda\in\mathbb{C}$
   - how to determine the $K-1$ largest eigenvalues???

Relation with FLDA

1. finite nonzero
   $(\boldsymbol{B} - \frac{t_{ii}}{s_{ii}}\boldsymbol{W})\mathbf{c} = 0 \quad \Rightarrow \quad \mathbf{c}^T\boldsymbol{B}\mathbf{c} = \frac{t_{ii}}{s_{ii}}\mathbf{c}^T\boldsymbol{W}\mathbf{c}$
   - $\lambda = \frac{t_{ii}}{s_{ii}}$ ... ratio of between- to within-variance
   - complement of null-spaces of $\boldsymbol{B}$, $\boldsymbol{W}$
2. finite zero
   $(\boldsymbol{B} - 0 \cdot \boldsymbol{W})\mathbf{c} = 0 \quad \Rightarrow \quad \mathbf{c}^T\boldsymbol{B}\mathbf{c} = 0$
   - opposite of FLDA aim
   - null-space of $\boldsymbol{B}$
3. infinite
   $\mathbf{c}^T\boldsymbol{B}\mathbf{c} = \lambda\mathbf{c}^T\boldsymbol{W}\mathbf{c}$ for $\lambda = \infty$ "$\Rightarrow$" $\mathbf{c}^T\boldsymbol{W}\mathbf{c} = 0$
   - wanted for FLDA, quality depends on $\mathbf{c}^T\boldsymbol{B}\mathbf{c}$
   - null-space of $\boldsymbol{W}$
4. any value is eigenvalue
   $\mathbf{c}^T\boldsymbol{B}\mathbf{c} = \lambda\mathbf{c}^T\boldsymbol{W}\mathbf{c} \quad \forall\lambda \in \mathbb{C}$ "$\Rightarrow$" $\mathbf{c}^T\boldsymbol{W}\mathbf{c} = 0 = \mathbf{c}^T\boldsymbol{B}\mathbf{c}$
   - uninteresting for FLDA
   - common null-space of $\boldsymbol{B}$, $\boldsymbol{W}$

Relation with FLDA

1. finite nonzero
   $(\boldsymbol{B} - \frac{t_{ii}}{s_{ii}}\boldsymbol{W})\mathbf{c} = 0 \quad \Rightarrow \quad \mathbf{c}^T\boldsymbol{B}\mathbf{c} = \frac{t_{ii}}{s_{ii}}\mathbf{c}^T\boldsymbol{W}\mathbf{c}$
   - $\lambda = \frac{t_{ii}}{s_{ii}}$ . . . ratio of between- to within-variance
   - complement of null-spaces of $\boldsymbol{B}$, $\boldsymbol{W}$
2. finite zero
   $(\boldsymbol{B} - 0 \cdot \boldsymbol{W})\mathbf{c} = 0 \quad \Rightarrow \quad \mathbf{c}^T\boldsymbol{B}\mathbf{c} = 0$
   - opposite of FLDA aim
   - null-space of $\boldsymbol{B}$
3. infinite
   $\mathbf{c}^T\boldsymbol{B}\mathbf{c} = \lambda\mathbf{c}^T\boldsymbol{W}\mathbf{c}$ for $\lambda = \infty$ " $\Rightarrow$ " $\mathbf{c}^T\boldsymbol{W}\mathbf{c} = 0$
   - wanted for FLDA, quality depends on $\mathbf{c}^T\boldsymbol{B}\mathbf{c}$
   - null-space of $\boldsymbol{W}$
4. any value is eigenvalue
   $\mathbf{c}^T\boldsymbol{B}\mathbf{c} = \lambda\mathbf{c}^T\boldsymbol{W}\mathbf{c} \quad \forall \lambda \in \mathbb{C}$ " $\Rightarrow$ " $\mathbf{c}^T\boldsymbol{W}\mathbf{c} = 0 = \mathbf{c}^T\boldsymbol{B}\mathbf{c}$
   - uninteresting for FLDA
   - common null-space of $\boldsymbol{B}$, $\boldsymbol{W}$

Relation with FLDA

1. finite nonzero
   $(\boldsymbol{B} - \frac{t_{ii}}{s_{ii}}\boldsymbol{W})\mathbf{c} = 0 \quad \Rightarrow \quad \mathbf{c}^T\boldsymbol{B}\mathbf{c} = \frac{t_{ii}}{s_{ii}}\mathbf{c}^T\boldsymbol{W}\mathbf{c}$
   - $\lambda = \frac{t_{ii}}{s_{ii}}$ ... ratio of between- to within-variance
   - complement of null-spaces of $\boldsymbol{B}$, $\boldsymbol{W}$
2. finite zero
   $(\boldsymbol{B} - 0 \cdot \boldsymbol{W})\mathbf{c} = 0 \quad \Rightarrow \quad \mathbf{c}^T\boldsymbol{B}\mathbf{c} = 0$
   - opposite of FLDA aim
   - null-space of $\boldsymbol{B}$
3. infinite
   $\mathbf{c}^T\boldsymbol{B}\mathbf{c} = \lambda\mathbf{c}^T\boldsymbol{W}\mathbf{c}$ for $\lambda = \infty$ " $\Rightarrow$ " $\mathbf{c}^T\boldsymbol{W}\mathbf{c} = 0$
   - wanted for FLDA, quality depends on $\mathbf{c}^T\boldsymbol{B}\mathbf{c}$
   - null-space of $\boldsymbol{W}$
4. any value is eigenvalue
   $\mathbf{c}^T\boldsymbol{B}\mathbf{c} = \lambda\mathbf{c}^T\boldsymbol{W}\mathbf{c}$ $\forall \lambda \in \mathbb{C}$ " $\Rightarrow$ " $\mathbf{c}^T\boldsymbol{W}\mathbf{c} = 0 = \mathbf{c}^T\boldsymbol{B}\mathbf{c}$
   - uninteresting for FLDA
   - common null-space of $\boldsymbol{B}$, $\boldsymbol{W}$

Relation with FLDA

1. finite nonzero
   $(\boldsymbol{B} - \frac{t_{ii}}{s_{ii}}\boldsymbol{W})\mathbf{c} = 0 \quad \Rightarrow \quad \mathbf{c}^T\boldsymbol{B}\mathbf{c} = \frac{t_{ii}}{s_{ii}}\mathbf{c}^T\boldsymbol{W}\mathbf{c}$
   - $\lambda = \frac{t_{ii}}{s_{ii}}$ ... ratio of between- to within-variance
   - complement of null-spaces of $\boldsymbol{B}$, $\boldsymbol{W}$
2. finite zero
   $(\boldsymbol{B} - 0 \cdot \boldsymbol{W})\mathbf{c} = 0 \quad \Rightarrow \quad \mathbf{c}^T\boldsymbol{B}\mathbf{c} = 0$
   - opposite of FLDA aim
   - null-space of $\boldsymbol{B}$
3. infinite
   $\mathbf{c}^T\boldsymbol{B}\mathbf{c} = \lambda\mathbf{c}^T\boldsymbol{W}\mathbf{c}$ for $\lambda = \infty$ " $\Rightarrow$ " $\mathbf{c}^T\boldsymbol{W}\mathbf{c} = 0$
   - wanted for FLDA, quality depends on $\mathbf{c}^T\boldsymbol{B}\mathbf{c}$
   - null-space of $\boldsymbol{W}$
4. any value is eigenvalue
   $\mathbf{c}^T\boldsymbol{B}\mathbf{c} = \lambda\mathbf{c}^T\boldsymbol{W}\mathbf{c} \quad \forall \lambda \in \mathbb{C}$ " $\Rightarrow$ " $\mathbf{c}^T\boldsymbol{W}\mathbf{c} = 0 = \mathbf{c}^T\boldsymbol{B}\mathbf{c}$
   - uninteresting for FLDA
   - common null-space of $\boldsymbol{B}$, $\boldsymbol{W}$

Relation with FLDA

1. finite nonzero
   $(\boldsymbol{B} - \frac{t_{ii}}{s_{ii}}\boldsymbol{W})\mathbf{c} = 0 \quad \Rightarrow \quad \mathbf{c}^T\boldsymbol{B}\mathbf{c} = \frac{t_{ii}}{s_{ii}}\mathbf{c}^T\boldsymbol{W}\mathbf{c}$
   - $\lambda = \frac{t_{ii}}{s_{ii}}$ ... ratio of between- to within-variance
   - complement of null-spaces of $\boldsymbol{B}$, $\boldsymbol{W}$
2. finite zero
   $(\boldsymbol{B} - 0 \cdot \boldsymbol{W})\mathbf{c} = 0 \quad \Rightarrow \quad \mathbf{c}^T\boldsymbol{B}\mathbf{c} = 0$
   - opposite of FLDA aim
   - null-space of $\boldsymbol{B}$
3. infinite
   $\mathbf{c}^T\boldsymbol{B}\mathbf{c} = \lambda\mathbf{c}^T\boldsymbol{W}\mathbf{c}$ for $\lambda = \infty$ " $\Rightarrow$ " $\mathbf{c}^T\boldsymbol{W}\mathbf{c} = 0$
   - wanted for FLDA, quality depends on $\mathbf{c}^T\boldsymbol{B}\mathbf{c}$
   - null-space of $\boldsymbol{W}$
4. any value is eigenvalue
   $\mathbf{c}^T\boldsymbol{B}\mathbf{c} = \lambda\mathbf{c}^T\boldsymbol{W}\mathbf{c} \quad \forall \lambda \in \mathbb{C}$ " $\Rightarrow$ " $\mathbf{c}^T\boldsymbol{W}\mathbf{c} = 0 = \mathbf{c}^T\boldsymbol{B}\mathbf{c}$
   - uninteresting for FLDA
   - common null-space of $\boldsymbol{B}$, $\boldsymbol{W}$

Relation with FLDA

1. finite nonzero
   $(\boldsymbol{B} - \frac{t_{ii}}{s_{ii}}\boldsymbol{W})\mathbf{c} = 0 \quad \Rightarrow \quad \mathbf{c}^T\boldsymbol{B}\mathbf{c} = \frac{t_{ii}}{s_{ii}}\mathbf{c}^T\boldsymbol{W}\mathbf{c}$
   - $\lambda = \frac{t_{ii}}{s_{ii}}$ ... ratio of between- to within-variance
   - complement of null-spaces of $\boldsymbol{B}$, $\boldsymbol{W}$
2. finite zero
   $(\boldsymbol{B} - 0 \cdot \boldsymbol{W})\mathbf{c} = 0 \quad \Rightarrow \quad \mathbf{c}^T\boldsymbol{B}\mathbf{c} = 0$
   - opposite of FLDA aim
   - null-space of $\boldsymbol{B}$
3. infinite
   $\mathbf{c}^T\boldsymbol{B}\mathbf{c} = \lambda\mathbf{c}^T\boldsymbol{W}\mathbf{c}$ for $\lambda = \infty$ " $\Rightarrow$ " $\mathbf{c}^T\boldsymbol{W}\mathbf{c} = 0$
   - wanted for FLDA, quality depends on $\mathbf{c}^T\boldsymbol{B}\mathbf{c}$
   - null-space of $\boldsymbol{W}$
4. any value is eigenvalue
   $\mathbf{c}^T\boldsymbol{B}\mathbf{c} = \lambda\mathbf{c}^T\boldsymbol{W}\mathbf{c} \quad \forall \lambda \in \mathbb{C}$ " $\Rightarrow$ " $\mathbf{c}^T\boldsymbol{W}\mathbf{c} = 0 = \mathbf{c}^T\boldsymbol{B}\mathbf{c}$
   - uninteresting for FLDA
   - common null-space of $\boldsymbol{B}$, $\boldsymbol{W}$

Relation with FLDA

1. finite nonzero
   $(\boldsymbol{B} - \frac{t_{ii}}{s_{ii}} \boldsymbol{W})\mathbf{c} = 0 \quad \Rightarrow \quad \mathbf{c}^T \boldsymbol{B} \mathbf{c} = \frac{t_{ii}}{s_{ii}} \mathbf{c}^T \boldsymbol{W} \mathbf{c}$
   - $\lambda = \frac{t_{ii}}{s_{ii}}$ ... ratio of between- to within-variance
   - complement of null-spaces of $\boldsymbol{B}$, $\boldsymbol{W}$
2. finite zero
   $(\boldsymbol{B} - 0 \cdot \boldsymbol{W})\mathbf{c} = 0 \quad \Rightarrow \quad \mathbf{c}^T \boldsymbol{B} \mathbf{c} = 0$
   - opposite of FLDA aim
   - null-space of $\boldsymbol{B}$
3. infinite
   $\mathbf{c}^T \boldsymbol{B} \mathbf{c} = \lambda \mathbf{c}^T \boldsymbol{W} \mathbf{c}$ for $\lambda = \infty$ " $\Rightarrow$ " $\mathbf{c}^T \boldsymbol{W} \mathbf{c} = 0$
   - wanted for FLDA, quality depends on $\mathbf{c}^T \boldsymbol{B} \mathbf{c}$
   - null-space of $\boldsymbol{W}$
4. any value is eigenvalue
   $\mathbf{c}^T \boldsymbol{B} \mathbf{c} = \lambda \mathbf{c}^T \boldsymbol{W} \mathbf{c} \quad \forall \lambda \in \mathbb{C}$ " $\Rightarrow$ " $\mathbf{c}^T \boldsymbol{W} \mathbf{c} = 0 = \mathbf{c}^T \boldsymbol{B} \mathbf{c}$
   - uninteresting for FLDA
   - common null-space of $\boldsymbol{B}$, $\boldsymbol{W}$

Relation with FLDA

1. finite nonzero
   $(\boldsymbol{B} - \frac{t_{ii}}{s_{ii}}\boldsymbol{W})\mathbf{c} = 0 \quad \Rightarrow \quad \mathbf{c}^T\boldsymbol{B}\mathbf{c} = \frac{t_{ii}}{s_{ii}}\mathbf{c}^T\boldsymbol{W}\mathbf{c}$
   - $\lambda = \frac{t_{ii}}{s_{ii}}$ ... ratio of between- to within-variance
   - complement of null-spaces of $\boldsymbol{B}$, $\boldsymbol{W}$
2. finite zero
   $(\boldsymbol{B} - 0 \cdot \boldsymbol{W})\mathbf{c} = 0 \quad \Rightarrow \quad \mathbf{c}^T\boldsymbol{B}\mathbf{c} = 0$
   - opposite of FLDA aim
   - null-space of $\boldsymbol{B}$
3. infinite
   $\mathbf{c}^T\boldsymbol{B}\mathbf{c} = \lambda\mathbf{c}^T\boldsymbol{W}\mathbf{c}$ for $\lambda = \infty$ " $\Rightarrow$ " $\mathbf{c}^T\boldsymbol{W}\mathbf{c} = 0$
   - wanted for FLDA, quality depends on $\mathbf{c}^T\boldsymbol{B}\mathbf{c}$
   - null-space of $\boldsymbol{W}$
4. any value is eigenvalue
   $\mathbf{c}^T\boldsymbol{B}\mathbf{c} = \lambda\mathbf{c}^T\boldsymbol{W}\mathbf{c} \quad \forall \lambda \in \mathbb{C}$ " $\Rightarrow$ " $\mathbf{c}^T\boldsymbol{W}\mathbf{c} = 0 = \mathbf{c}^T\boldsymbol{B}\mathbf{c}$
   - uninteresting for FLDA
   - common null-space of $\boldsymbol{B}$, $\boldsymbol{W}$

Relation with FLDA

1. finite nonzero
   $(\boldsymbol{B} - \frac{t_{ii}}{s_{ii}}\boldsymbol{W})\mathbf{c} = 0 \quad \Rightarrow \quad \mathbf{c}^T\boldsymbol{B}\mathbf{c} = \frac{t_{ii}}{s_{ii}}\mathbf{c}^T\boldsymbol{W}\mathbf{c}$
   - $\lambda = \frac{t_{ii}}{s_{ii}}$ ... ratio of between- to within-variance
   - complement of null-spaces of $\boldsymbol{B}$, $\boldsymbol{W}$
2. finite zero
   $(\boldsymbol{B} - 0 \cdot \boldsymbol{W})\mathbf{c} = 0 \quad \Rightarrow \quad \mathbf{c}^T\boldsymbol{B}\mathbf{c} = 0$
   - opposite of FLDA aim
   - null-space of $\boldsymbol{B}$
3. infinite
   $\mathbf{c}^T\boldsymbol{B}\mathbf{c} = \lambda\mathbf{c}^T\boldsymbol{W}\mathbf{c}$ for $\lambda = \infty$ " $\Rightarrow$ " $\mathbf{c}^T\boldsymbol{W}\mathbf{c} = 0$
   - wanted for FLDA, quality depends on $\mathbf{c}^T\boldsymbol{B}\mathbf{c}$
   - null-space of $\boldsymbol{W}$
4. any value is eigenvalue
   $\mathbf{c}^T\boldsymbol{B}\mathbf{c} = \lambda\mathbf{c}^T\boldsymbol{W}\mathbf{c} \quad \forall \lambda \in \mathbb{C}$ " $\Rightarrow$ " $\mathbf{c}^T\boldsymbol{W}\mathbf{c} = 0 = \mathbf{c}^T\boldsymbol{B}\mathbf{c}$
   - uninteresting for FLDA
   - common null-space of $\boldsymbol{B}$, $\boldsymbol{W}$

Relation with FLDA

1. finite nonzero
   $(\boldsymbol{B} - \frac{t_{ii}}{s_{ii}} \boldsymbol{W})\mathbf{c} = 0 \quad \Rightarrow \quad \mathbf{c}^T \boldsymbol{B} \mathbf{c} = \frac{t_{ii}}{s_{ii}} \mathbf{c}^T \boldsymbol{W} \mathbf{c}$
   - $\lambda = \frac{t_{ii}}{s_{ii}}$ ... ratio of between- to within-variance
   - complement of null-spaces of $\boldsymbol{B}$, $\boldsymbol{W}$
2. finite zero
   $(\boldsymbol{B} - 0 \cdot \boldsymbol{W})\mathbf{c} = 0 \quad \Rightarrow \quad \mathbf{c}^T \boldsymbol{B} \mathbf{c} = 0$
   - opposite of FLDA aim
   - null-space of $\boldsymbol{B}$
3. infinite
   $\mathbf{c}^T \boldsymbol{B} \mathbf{c} = \lambda \mathbf{c}^T \boldsymbol{W} \mathbf{c}$ for $\lambda = \infty$ " $\Rightarrow$ " $\mathbf{c}^T \boldsymbol{W} \mathbf{c} = 0$
   - wanted for FLDA, quality depends on $\mathbf{c}^T \boldsymbol{B} \mathbf{c}$
   - null-space of $\boldsymbol{W}$
4. any value is eigenvalue
   $\mathbf{c}^T \boldsymbol{B} \mathbf{c} = \lambda \mathbf{c}^T \boldsymbol{W} \mathbf{c} \quad \forall \lambda \in \mathbb{C}$ " $\Rightarrow$ " $\mathbf{c}^T \boldsymbol{W} \mathbf{c} = 0 = \mathbf{c}^T \boldsymbol{B} \mathbf{c}$
   - uninteresting for FLDA
   - common null-space of $\boldsymbol{B}$, $\boldsymbol{W}$

Relation with FLDA

1. finite nonzero
   $(\boldsymbol{B} - \frac{t_{ii}}{s_{ii}}\boldsymbol{W})\mathbf{c} = 0 \quad \Rightarrow \quad \mathbf{c}^T\boldsymbol{B}\mathbf{c} = \frac{t_{ii}}{s_{ii}}\mathbf{c}^T\boldsymbol{W}\mathbf{c}$
   - $\lambda = \frac{t_{ii}}{s_{ii}}$ . . . ratio of between- to within-variance
   - complement of null-spaces of $\boldsymbol{B}$, $\boldsymbol{W}$
2. finite zero
   $(\boldsymbol{B} - 0 \cdot \boldsymbol{W})\mathbf{c} = 0 \quad \Rightarrow \quad \mathbf{c}^T\boldsymbol{B}\mathbf{c} = 0$
   - opposite of FLDA aim
   - null-space of $\boldsymbol{B}$
3. infinite
   $\mathbf{c}^T\boldsymbol{B}\mathbf{c} = \lambda\mathbf{c}^T\boldsymbol{W}\mathbf{c}$ for $\lambda = \infty$ " $\Rightarrow$ " $\mathbf{c}^T\boldsymbol{W}\mathbf{c} = 0$
   - wanted for FLDA, quality depends on $\mathbf{c}^T\boldsymbol{B}\mathbf{c}$
   - null-space of $\boldsymbol{W}$
4. any value is eigenvalue
   $\mathbf{c}^T\boldsymbol{B}\mathbf{c} = \lambda\mathbf{c}^T\boldsymbol{W}\mathbf{c} \quad \forall\lambda \in \mathbb{C}$ " $\Rightarrow$ " $\mathbf{c}^T\boldsymbol{W}\mathbf{c} = 0 = \mathbf{c}^T\boldsymbol{B}\mathbf{c}$
   - uninteresting for FLDA
   - common null-space of $\boldsymbol{B}$, $\boldsymbol{W}$

Relation with FLDA

1. finite nonzero
   $(\boldsymbol{B} - \frac{t_{ii}}{s_{ii}} \boldsymbol{W})\mathbf{c} = 0 \quad \Rightarrow \quad \mathbf{c}^T \boldsymbol{B} \mathbf{c} = \frac{t_{ii}}{s_{ii}} \mathbf{c}^T \boldsymbol{W} \mathbf{c}$
   - $\lambda = \frac{t_{ii}}{s_{ii}}$ ... ratio of between- to within-variance
   - complement of null-spaces of $\boldsymbol{B}$, $\boldsymbol{W}$
2. finite zero
   $(\boldsymbol{B} - 0 \cdot \boldsymbol{W})\mathbf{c} = 0 \quad \Rightarrow \quad \mathbf{c}^T \boldsymbol{B} \mathbf{c} = 0$
   - opposite of FLDA aim
   - null-space of $\boldsymbol{B}$
3. infinite
   $\mathbf{c}^T \boldsymbol{B} \mathbf{c} = \lambda \mathbf{c}^T \boldsymbol{W} \mathbf{c} \quad \text{for} \quad \lambda = \infty \quad " \Rightarrow " \quad \mathbf{c}^T \boldsymbol{W} \mathbf{c} = 0$
   - wanted for FLDA, quality depends on $\mathbf{c}^T \boldsymbol{B} \mathbf{c}$
   - null-space of $\boldsymbol{W}$
4. any value is eigenvalue
   $\mathbf{c}^T \boldsymbol{B} \mathbf{c} = \lambda \mathbf{c}^T \boldsymbol{W} \mathbf{c} \quad \forall \lambda \in \mathbb{C} \quad " \Rightarrow " \quad \mathbf{c}^T \boldsymbol{W} \mathbf{c} = 0 = \mathbf{c}^T \boldsymbol{B} \mathbf{c}$
   - uninteresting for FLDA
   - common null-space of $\boldsymbol{B}$, $\boldsymbol{W}$

Introduction

- here only methods from R, Matlab
  - linked with LAPACK-libraries
  - all methods are backward stable
- common approaches . . . "eliminate" singularity
- slight modification of $\boldsymbol{W}$ while preserving crucial information . . . regularization
- most methods based on spectral decomposition

$$\boldsymbol{W} = \boldsymbol{Q}\,\mathrm{diag}(\lambda_1, \ldots, \lambda_r, 0, \ldots, 0)\boldsymbol{Q}^T, \quad \boldsymbol{Q}^T\boldsymbol{Q} = \mathbf{I},$$

where $r = \mathrm{rank}(\boldsymbol{W})$

Outline

Pseudoinverse

- let $\boldsymbol{\Lambda}_r = \mathrm{diag}(\lambda_1, \ldots, \lambda_r)$
- partition $\boldsymbol{Q} = (\boldsymbol{Q}_r, \boldsymbol{Q}_N)$, where $\boldsymbol{Q}_N$ spans null-space of $\boldsymbol{W}$
- transformation to standard eigenproblem with

$$\boldsymbol{W}^+ = \boldsymbol{Q}_r \boldsymbol{\Lambda}_r^{-1} \boldsymbol{Q}_r^T$$

Pseudoinverse - properties

+ smaller eigenproblem (dimension r)
+ no need to search for an optimal regularization parameter
– only finite eigenpairs (discards null-space of $\boldsymbol{W}$)

Pseudoinverse

- let $\boldsymbol{\Lambda}_r = \mathrm{diag}(\lambda_1, \ldots, \lambda_r)$
- partition $\boldsymbol{Q} = (\boldsymbol{Q}_r, \boldsymbol{Q}_N)$, where $\boldsymbol{Q}_N$ spans null-space of $\boldsymbol{W}$
- transformation to standard eigenproblem with

$$\boldsymbol{W}^+ = \boldsymbol{Q}_r \boldsymbol{\Lambda}_r^{-1} \boldsymbol{Q}_r^T$$

Pseudoinverse - properties

+ smaller eigenproblem (dimension r)
+ no need to search for an optimal regularization parameter
– only finite eigenpairs (discards null-space of $\boldsymbol{W}$)

Implementation I - Non-symmetric transformation

- solve standard eigenproblem

$$(\boldsymbol{Q}_r \boldsymbol{\Lambda}_r^{-1} \boldsymbol{Q}_r^T \boldsymbol{B} - \lambda \mathbf{I})\mathbf{c} = 0$$

- cost of nonsymmetric QR-method: $\pm 25p^3$ flops
- eigenvalues and -vectors can be ill-conditioned
- store several $p \times p$ matrices
- see e. g. [Cheng et al. - 1992]

Implementation II - Symmetric transformation

- solve standard eigenproblem

$$(\boldsymbol{\Lambda}_r^{-1/2}\boldsymbol{Q}_r^T\boldsymbol{B}\boldsymbol{Q}_r\boldsymbol{\Lambda}_r^{-1/2} - \lambda\mathbf{I})\mathbf{c}^* = 0,$$

$$\mathbf{c} = \frac{\boldsymbol{Q}_r\boldsymbol{\Lambda}_r^{1/2}\mathbf{c}^*}{\|\boldsymbol{Q}_r\boldsymbol{\Lambda}_r^{1/2}\mathbf{c}^*\|}$$

- cost of symmetric QR-method: $\pm 9p^3$ flops
- only eigenvectors can be ill-conditioned
- store $\boldsymbol{W}$, $\boldsymbol{Q} \in \mathbb{R}^{p \times p}$, but transformed eigenproblem is $r \times r$
- see e.g. [Krzanowski et al. - 1995]

Implementation III - SVD-implementation I

- exploits the given structure of $\boldsymbol{B}$ and $\boldsymbol{W}$, e. g.

$$\boldsymbol{W} = \left(\frac{X - \boldsymbol{GM}}{\sqrt{n-g}}\right)^T \frac{X - \boldsymbol{GM}}{\sqrt{n-g}} = \boldsymbol{Q\Lambda Q}^T$$

- use SVD instead of eigen decomposition
- cost: $\pm 4p^2n$ flops or $\pm 14pn^2$ flops for "economy size SVD"

Implementation III - SVD-implementation II

- only eigenvectors can be ill-conditioned
- storage: 1 $n \times p$ matrix, 1 $n \times r$ matrix
- here everywhere economy size SVD
- `lda()` function in R-environment with default parameters
  - pseudoinverse method
  - SVD implementation

Outline

Epsilon-perturbation I

- consider $\boldsymbol{Q}\boldsymbol{\Lambda}_\varepsilon\boldsymbol{Q}^T$ instead of $\boldsymbol{Q}_r\boldsymbol{\Lambda}_r\boldsymbol{Q}_r^T$,

  $$\text{where } \boldsymbol{\Lambda}_\varepsilon = \operatorname{diag}(\lambda_1, \ldots, \lambda_r, \varepsilon, \ldots, \varepsilon)$$

  see e. g. [Cheng et al. - 1992]
- transformation of modified generalized eigenproblem

  $$(\boldsymbol{B} - \lambda\boldsymbol{Q}\boldsymbol{\Lambda}_\varepsilon\boldsymbol{Q}^T)\mathbf{c} = 0$$

Epsilon-perturbation - properties

+ no exclusion of any null-spaces
– large eigenproblem (dimension p)
– need for regularization parameter ε
– too small $\varepsilon \Rightarrow$ ill-conditioned eigenproblem

Epsilon-perturbation I

- consider $\boldsymbol{Q\Lambda}_\varepsilon \boldsymbol{Q}^T$ instead of $\boldsymbol{Q}_r \boldsymbol{\Lambda}_r \boldsymbol{Q}_r^T$,

$$\text{where } \boldsymbol{\Lambda}_\varepsilon = \text{diag}(\lambda_1, \ldots, \lambda_r, \varepsilon, \ldots, \varepsilon)$$

  see e. g. [Cheng et al. - 1992]
- transformation of modified generalized eigenproblem

$$(\boldsymbol{B} - \lambda \boldsymbol{Q\Lambda}_\varepsilon \boldsymbol{Q}^T)\mathbf{c} = 0$$

Epsilon-perturbation - properties

+ no exclusion of any null-spaces
– large eigenproblem (dimension p)
– need for regularization parameter ε
– too small $\varepsilon \Rightarrow$ ill-conditioned eigenproblem

Epsilon-perturbation II

- nonsymmetric & symmetric implementations: as before, only forming transformed eigenproblem is little more expensive
- SVD implementation: as before, BUT
  - store $p \times p$ matrix $\boldsymbol{Q}$
  - cannot exploit economy SVD: $\pm 4p^2n$

Outline

Null-space I

- vectors $\mathbf{c}$ in the null-space of $\boldsymbol{W}$ satisfy $\mathbf{c}^T \boldsymbol{W} \mathbf{c} = 0$
- if $\mathbf{c}^T \boldsymbol{B} \mathbf{c}$ is large $\Rightarrow$ can be used for projection
- find largest eigenvalues of $\boldsymbol{B}$ in null-space of $\boldsymbol{W}$ by solving

$$(\boldsymbol{Q}_N^T \boldsymbol{B} \boldsymbol{Q}_N - \mathbf{I})\mathbf{c} = 0$$

see e. g. [Guo et al. - 2003]

The Null-space method - properties

+ does not need a regularization parameter
– solves an eigenproblem of dimension $p - r$
– discards finite eigenvalues of original problem

Null-space I

- vectors $\mathbf{c}$ in the null-space of $\boldsymbol{W}$ satisfy $\mathbf{c}^T \boldsymbol{W} \mathbf{c} = 0$
- if $\mathbf{c}^T \boldsymbol{B} \mathbf{c}$ is large $\Rightarrow$ can be used for projection
- find largest eigenvalues of $\boldsymbol{B}$ in null-space of $\boldsymbol{W}$ by solving

$$(\boldsymbol{Q}_N^T \boldsymbol{B} \boldsymbol{Q}_N - \mathbf{I})\mathbf{c} = 0$$

see e. g. [Guo et al. - 2003]

The Null-space method - properties

+ does not need a regularization parameter
– solves an eigenproblem of dimension $p - r$
– discards finite eigenvalues of original problem

Null-space II

- symmetric implementation: as for Epsilon perturbation, but storage and cost a little cheaper
- SVD
  - storage: 1 $p \times (p - r)$ matrix
  - cost of 1 full SVD: $4p^2n$

Outline

Common Null-space Elimination I

- ...numerical analysists recommend [Parlett - 1998]
- common null-space of $\boldsymbol{B}$, $\boldsymbol{W}$ $\Rightarrow$ ill-posed eigenproblem
- project onto complement of common null-space
- $\boldsymbol{B}x = 0 \quad \wedge \quad \boldsymbol{W}x = 0 \quad \Leftrightarrow \quad (\boldsymbol{B} + \boldsymbol{W})x = 0$, because $\boldsymbol{B}, \boldsymbol{W} \geq 0$
- compute spectral decomposition $\boldsymbol{B} + \boldsymbol{W}$
- let $\boldsymbol{P}$ contain eigenvectors for non-zero eigenvalues
- solve the projected problem

$$(\boldsymbol{P}^T\boldsymbol{B}\boldsymbol{P} - \lambda\boldsymbol{P}^T\boldsymbol{W}\boldsymbol{P})\mathbf{c}^* = 0, \quad \mathbf{c} = \boldsymbol{P}\mathbf{c}^*$$

- vector selection by considering $\mathbf{c}^T\boldsymbol{B}\mathbf{c}$, $\mathbf{c}^T\boldsymbol{W}\mathbf{c}$

Common Null-space Elimination II

Common Null-space Elimination - properties

+ problem smaller than n ($\ll p$)
+ no exclusion of null-space of $\boldsymbol{W}$

Implementation with QZ:

- storage: only $p \times n$ matrices
- cost: order $pn^2 + n^3$
- both eigenvalues and -vectors can be ill-conditioned

Common Null-space Elimination II

Common Null-space Elimination - properties

+ problem smaller than n ($\ll p$)
+ no exclusion of null-space of $\boldsymbol{W}$

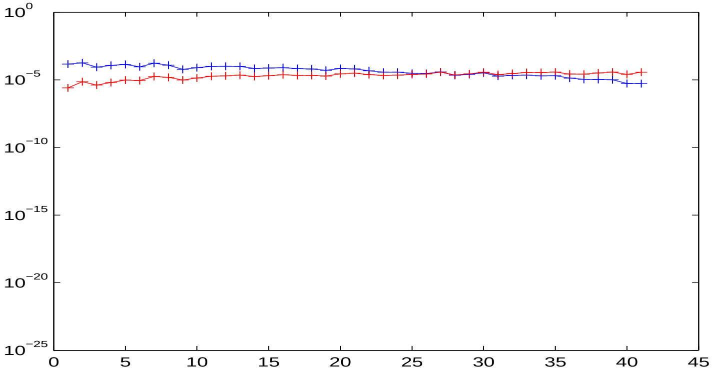
Implementation with QZ:

- storage: only $p \times n$ matrices
- cost: order $pn^2 + n^3$
- both eigenvalues and -vectors can be ill-conditioned

Numerical Experiment

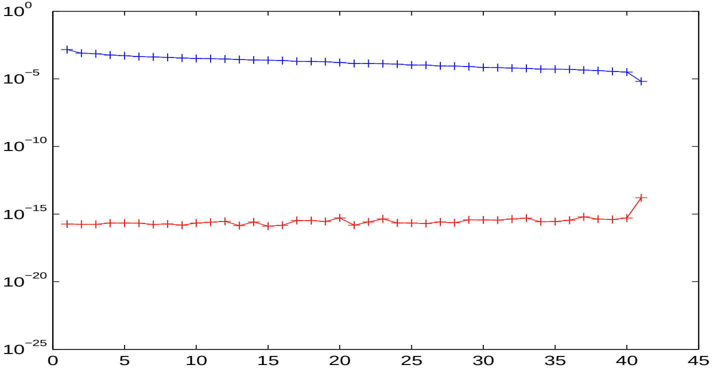
- protein fold classification problem
- sample size $n = 268$
    - 143 in training set
    - 125 in test set
- $p = 400$ predictors
- $K = 42$ classes

Pseudoinverse



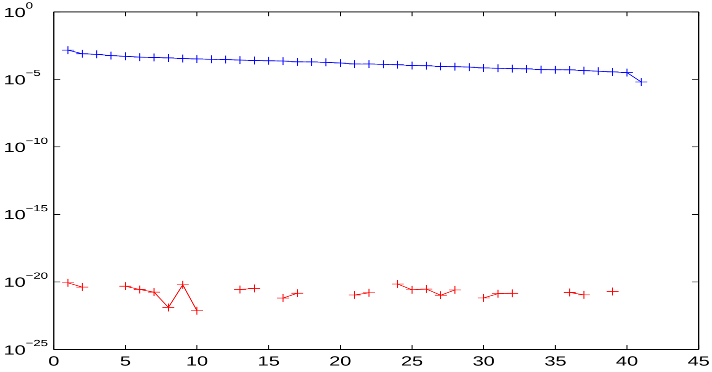
blue curve: $\mathbf{c}^T \boldsymbol{B} \mathbf{c}$, red curve: $\mathbf{c}^T \boldsymbol{W} \mathbf{c}$

Epsilon-perturbation



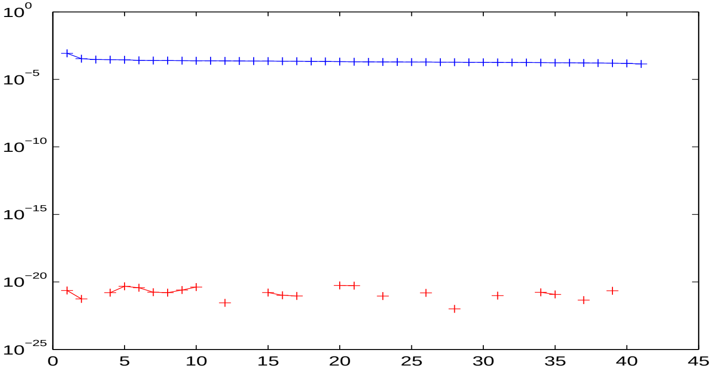
blue curve: $\mathbf{c}^T \boldsymbol{B} \mathbf{c}$, red curve: $\mathbf{c}^T \boldsymbol{W} \mathbf{c}$

Null-Space



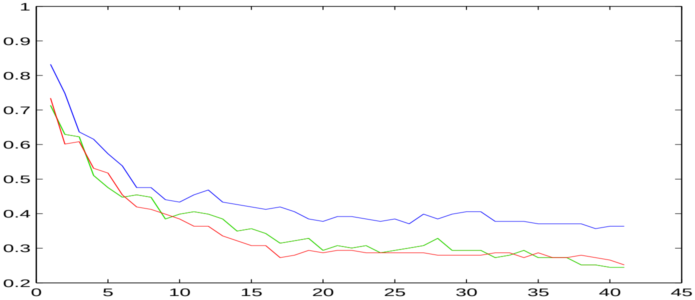
blue curve: $\mathbf{c}^T \boldsymbol{B} \mathbf{c}$, red curve: $\mathbf{c}^T \boldsymbol{W} \mathbf{c}$

Common Null-Space Elimination



blue curve: $\mathbf{c}^T \boldsymbol{B} \mathbf{c}$, red curve: $\mathbf{c}^T \boldsymbol{W} \mathbf{c}$

Error Rates of Classifi ers



blue curve: Pseudoinverse, green curve: both Epsilon and Null-space, red curve: Common Null-space elimination

- two middle methods: error rate of 24%
- can compete with Support Vector Machine strategy 23.2% (other over 28.8%, see e. g. [Markowetz et al. - 2003])

Conclusions

- Generalized Schur decomposition as a theoretical tool for better understanding of FLDA
  - covers all proposed methods
- Common-null space elimination
  - seems most appropriate, not only theoretically, but also experimentally and numerically
- in addition we found...
  - pseudoinverse method (R) for $p \gg n$ needs not be the best at all!
  - exploitation of structure with SVD

Open Questions & Future work

- choice and influence of regularization parameter in perturbation technique
- main challenge for nearby future is extension to very large problems
    - sparsity!!!
    - exploitation of special structure
- generalization to nonlinear cases
    - kernels method for Fisher [Baudat, Anouar - 2000]

References

G. Baudat and F. Anouar:
Generalized discriminant analysis using a kernel approach.
Neural Computation, 12(10):2385–404(1), 2000.

Y. Cheng, Y. Zhuang and J. Yang:
Optimal Fisher discriminant analysis using the rank decomposition.
Pattern Recognition, 25(1):101–111, 1992.

Y.-F. Guo, S.-J. Li, J.-Y. Yang, T.-T. Shu and L.-D. Wu:
A generalized Foley-Sammon transform based on generalized fi sher discriminant criterion and its application to face recognition.
Pattern Recognition Letters, 24:147–158, 2003.

W. J. Krzanowski, P. Jonathan, W .V .McCarthy and M .R. Thomas:
Discriminant analysis with singular covariance matrices: methods and applications to spectroscopic data.
Applied Statistics, 44:101–115, 1995.

F. Markowetz, L. Edler and M. Vingron:
Support Vector Machines for Protein Fold Class Prediction.
Biometrical Journal, 45(3):377–389, 2003.

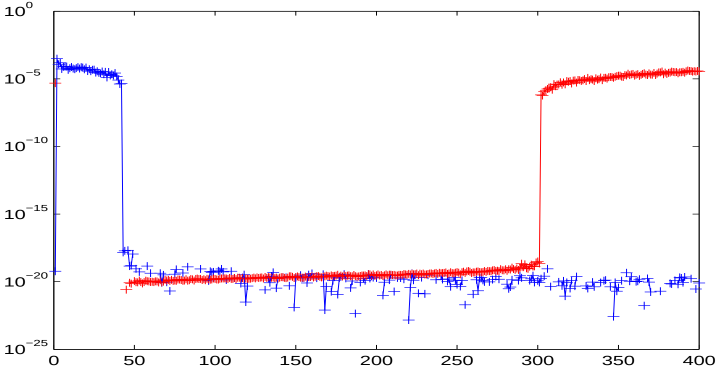
C. B. Moler and G. W. Stewart:
An algorithm for generalized matrix eigenvalue problems.
SIAM Journal on Numerical Analysis, 10:241–256, 1973.

B. N. Parlett:
The symmetric Eigenvalue Problem.
SIAM, Philadelphia, 1998.

Thank you for your attention!

This work is supported partly by the Program Information Society under project 1ET400300415 and the MSMT CR Project LC536

Generalized Schur Decomposition (by QZ)



blue curve: diagonal of $\boldsymbol{T}(\boldsymbol{B})$, red curve: diagonal of $\boldsymbol{S}(\boldsymbol{W})$