The Eckart-Young theorem and Ky Fan's maximum principle :

Two sides of the same coin

By

Achiya Dax

Hydrological Service, Jerusalem , Israel

e-mail: dax20@water.gov.il

Outline †

* Orthogonal Quotients Matrices.

* The **Eckart – Young** Theorem.

* The Orthogonal Quotients Equality.

* **Ky Fan**'s extremum principles.

* Extended Extremum Principles.

† A. Dax, "On extremum properties of orthogonal quotients matrices", Lin. Alg. and its Applic., 432(2010), pp. 1234 – 1257.

Orthogonal Quotient Matrices

A a real m x n matrix

X_{m*} a real m x m* orthogonal matrix

$$X_{m*} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{m*}], \quad X_{m*}^T X_{m*} = I.$$

Y_{n*} a real n x n* orthogonal matrix

$$Y_{n*} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{n*}], \quad Y_{n*}^T Y_{n*} = I.$$

Then the m* x n* matrix

$$(X_{m*})^T A Y_{n*} = (x_i^T A y_j)$$

is called an "Orthogonal Quotient Matrix".

Rayleigh Quotient Matrices

S a symmetric n x n matrix

Y_k a real n x k orthogonal matrix

$$Y_k = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_k] \quad , \quad Y_k^T Y_k = I \,.$$

Then the k x k matrix

$$Y_k^T S \, Y_k = (\mathbf{y}_i^T S \, \mathbf{y}_j)$$

is called a "Rayleigh Quotient Matrix".

Matrices of this form plays important role in the **Rayleigh-Ritz** procedure and in **Krylov** subspace methods.

Example: **Lanczos Algorithm**

$$Y_k^{\mathbf{T}} S \, Y_k = T_k$$

$T_{\mathbf{k}}$ a real k x k tridiagonal matrix

$Y_{\mathbf{k}}$ a real n x k orthogonal matrix

The columns of $Y_{\mathbf{k}}$ form an orthonormal basis of a Krylov subspace,

$$\text{Span}\,\{\, \mathbf{x},\, S\mathbf{x},\, S^2\mathbf{x},\, \ldots,\, S^{k-1}\mathbf{x} \,\},$$

for some starting vector $\mathbf{x}$.

Ky Fan's Maximum Principle (1949)†

considers the problem of maximizing the trace of
a Rayleigh Quotient Matrix .

$$\max\{\mathrm{trace}(Y_k^T S Y_k) \mid Y_k \in \mathbf{Y}_k\}$$

S a symmetric positive semi-definite n x n matrix.

$\mathbf{Y}_k$ denotes the set of orthogonal n x k matrices.

† Emeritus professor **Ky Fan** died in Santa Barbara on March 22, 2010, at age 95.

The Eckart–Young Theorem (1936)

considers the approximation of one matrix by another matrix of lower rank

$$\textbf{minimize} \quad F(B) = \|A - B\|_F^2$$

$$\textbf{subject to} \quad \mathrm{rank}(B) \le k$$

(Also called the Schmidt-Mirsky Theorem.)

The Frobenius matrix norm

$A = (a_{ij})$ a real m x n matrix, $m \geq n$.

$$\|A\|_F = (\Sigma \Sigma (a_{ij})^2)^{1/2}$$

$$\|A\|_F^2 = \text{trace}(A^T A) = \text{trace}(A A^T)$$

$$= (\sigma_1)^2 + (\sigma_2)^2 + \ldots + (\sigma_n)^2 ,$$

where

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0 ,$$

denote the singular values of A .

The Singular Value Decomposition

$$A = U \Sigma V^T$$

$$\Sigma = \text{diag}\,\{\sigma_1, \sigma_2, \ldots, \sigma_p\} \;, \qquad p = \min\{m, n\}$$

$$U = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p] \;, \qquad U^T U = I$$

$$V = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p] \;, \qquad V^T V = I$$

$$A V = U \Sigma \qquad A^T U = V \Sigma$$

$$A \mathbf{v}_j = \sigma_j \mathbf{u}_j \;, \quad A^T \mathbf{u}_j = \sigma_j \mathbf{v}_j \qquad j = 1, \ldots, p\,.$$

Rank-k truncated SVD

$$T_k(A) = U_k D_k V_k^T$$

$$D_k = \text{diag}\{\sigma_1, \sigma_2, \ldots, \sigma_k\} ,$$

$$U_k = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k] , \qquad U_k^T U_k = I$$

$$V_k = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k] , \qquad V_k^T V_k = I$$

The matrix $T_k(A)$ is called a rank-k truncated SVD of A .

The Eckart–Young Theorem (1936)

says that the "truncated SVD" matrix

$$T_k(A) = U_k D_k V_k^T$$

solves the least norm problem

$$\textbf{minimize} \quad F(B) = \|A - B\|_F^2$$

$$\textbf{subject to} \quad \operatorname{rank}(B) \le k$$

(Also called the Schmidt–Mirsky Theorem.)

Unitarily Invariant matrix norms

$A = (a_{ij})$ a real m x n matrix,

X a real m x m unitary matrix, $X^T X = I$.

Y a real n x n unitary matrix, $Y^T Y = I$.

$$||A|| = ||X^T A|| = ||AY|| = ||X^T A Y||$$

Unitarily Invariant matrix norms

$$||A||_F = \left((\sigma_1)^2 + (\sigma_2)^2 + \ldots + (\sigma_n)^2 \right)^{1/2}$$ Frobenius

$$||A|| = \left((\sigma_1)^p + (\sigma_2)^p + \ldots + (\sigma_n)^p \right)^{1/2}$$ Schatten p-norm

$1 \le p < \infty$

$$||A|| = \sigma_1 + \sigma_2 + \ldots + \sigma_k$$ Ky Fan $k-$norm , $k = 1, \ldots, n$.

$$||A||_{tr} = \sigma_1 + \sigma_2 + \ldots + \sigma_n$$ the trace norm.

$$||A||_2 = \sigma_1 = \max_{j = 1, \ldots, n} \{ \sigma_j \}$$ the 2- norm.

Mirsky Theorem (1960)

says that the "truncated SVD" matrix

$$T_k(A) = U_k D_k V_k^T$$

solves the least norm problem

minimize $\quad F(B) = \| A - B \|$

subject to $\quad \text{rank}(B) \le k$,

for any unitarily invariant matrix norm.

Rank-k Matrices

$$B = X_k R Y_k^{\mathbf{T}}$$

where

R is a k x k matrix

$$X_k = [\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_k}] \quad , \quad X_k^T X_k = I$$

$$Y_k = [\mathbf{y_1}, \mathbf{y_2}, \ldots, \mathbf{y_k}] \quad , \quad Y_k^T Y_k = I$$

The Eckart – Young Problem

can be rewritten as

$$\textbf{minimize} \quad F(B) = \| A - X_k R Y_k^T \|_F^2$$

$$\textbf{subject to} \quad X_k^T X_k = I \quad \text{and} \quad Y_k^T Y_k = I \ ,$$

where $B = X_k R Y_k^T$ and R is a $k \times k$ matrix.

Theorem: Let X_k and Y_k be a pair of orthogonal matrices as obove, then the related

"Orthogonal Quotients Matrix"

$$X_k^T A Y_k = (x_i^T A y_j)$$

solves the problem

minimize $\quad F(R) = \| A - X_k R Y_k^T \|_F$

Notation:

$\mathbf{X}_k$ - denotes the set of all real m x k orthogonal matrices X_k ,

$$X_k = [\mathbf{x}_1 , \mathbf{x}_2 , \ldots , \mathbf{x}_k] \quad , \quad X_k^T X_k = I$$

$\mathbf{Y}_k$ - denotes the set of all real n x k orthogonal matrices Y_k ,

$$Y_k = [\mathbf{y}_1 , \mathbf{y}_2 , \ldots , \mathbf{y}_k] \quad , \quad Y_k^T Y_k = I$$

Corollary 1: The Eckart–Young Problem

can be rewritten as

$$\textbf{minimize} \;\; F(X_k, Y_k) = \| A - X_k R_k Y_k^T \|_F^2$$

$$\textbf{subject to} \quad X_k \in \mathbf{X}_k \;\; \text{and} \;\; Y_k \in \mathbf{Y}_k \;,$$

where R_k is the related Orthogonal Quotients Matrix

$$R_k = X_k^T A Y_k \; .$$

The Orthogonal Quotients Equality

For any pair of orthogonal matrices,

$$X_k \in \mathbf{X}_k \quad \text{and} \quad Y_k \in \mathbf{Y}_k \quad ,$$

$$\| A - X_k R_k Y_k^T \|_F^2 = \| A \|_F^2 - \| R_k \|_F^2$$

where R_k is the related orthogonal quotients matrix

$$R_k = X_k^T A Y_k \quad .$$

Corollary 2: The Eckart–Young Problem

$$\textbf{minimize} \;\; F(X_k, Y_k) = \|A - X_k R_k Y_k^T\|_F^2$$

$$\textbf{subject to} \quad X_k \in \mathbf{X}_k \;\; \text{and} \;\; Y_k \in \mathbf{Y}_k \; .$$

is equivalent to

$$\textbf{maximize} \;\; \|X_k^T A Y_k\|_F^2$$

$$\textbf{subject to} \quad X_k \in \mathbf{X}_k \;\; \text{and} \;\; Y_k \in \mathbf{Y}_k \; .$$

The SVD matrices U_k and V_k solves both problems, giving the optimal values of $\sigma_{k+1}^2 + \dots + \sigma_n^2$ and $\sigma_1^2 + \sigma_2^2 + \dots + \sigma_k^2$, respectively.

Returning to symmetric matrices

Can we extend the

Orthogonal Quotients Equality

to Ky Fan extremum problems ?

The Spectral Decomposition

$S = (s_{ij})$ a symmetric positive semi-definite n x n matrix

With eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$

and eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$

$$S\,\mathbf{v}_j = \lambda_j \mathbf{v}_j, \quad \mathbf{j = 1, \ldots, n}. \qquad S\,V = V\,D$$

$$V = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n], \qquad V^T V = V\,V^T = I$$

$$D = \mathrm{diag}\,\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$$

$$S = V\,D\,V^T = \Sigma\, \lambda_j \mathbf{v}_j \mathbf{v}_j^T$$

Ky Fan's Maximum Principle

S a symmetric positive semi-definite n x n matrix

$\mathbf{Y}_k$ the set of orthogonal n x k matrices

$$\lambda_1 + \ldots + \lambda_k = \mathbf{max} \{ \mathbf{trace} (Y_k^T S Y_k) \mid Y_k \in \mathbf{Y}_k \}$$

Solution is obtained for the Spectral matrix

$$V_k = [\mathbf{v}_1 , \mathbf{v}_2 , \ \ldots \ , \mathbf{v}_k] \ .$$

which is related to the largest k eigenvalues.

Ky Fan's **Minimum** Principle

S a symmetric n x n matrix .

$\mathbf{Y}_k$ the set of orthogonal n x k matrices .

$$\lambda_{n-k+1} + \ldots + \lambda_n = \min \{ \mathrm{trace}(Y_k^T S Y_k) \mid Y_k \in \mathbf{Y}_k \}$$

Solution is obtained for the Spectral matrix

$$V_k = [\mathbf{v}_{n-k+1}, \ldots, \mathbf{v}_n] \ ,$$

which is related to the smallest k eigenvalues.

The Symmetric Quotients Equality

Given

S a symmetric nxn matrix

Y_k an orthogonal nxk matrix

$S_k = Y_k^T S Y_k$ the related "Rayleigh quotient matrix"

Then

$$\mathbf{trace}(S - Y_k S_k Y_k^T) = \mathbf{trace}(S) - \mathbf{trace}(S_k),$$

where $\mathbf{trace}(S_k) = \mathbf{trace}(Y_k^T S Y_k) = \Sigma\, \mathbf{y}_j^T S \mathbf{y}_j$.

Corollary 1: Ky Fan's maximum problem

$$\textbf{maximize} \quad \textbf{trace}(Y_k S Y_k^T)$$
$$\textbf{subject to} \quad Y_k \in \mathbf{Y}_k \ ,$$

is equivalent to

$$\textbf{minimize} \quad \textbf{trace}(S - Y_k S_k Y_k^T)$$
$$\textbf{subject to} \quad Y_k \in \mathbf{Y}_k \ .$$

The Spectral matrix $V_k = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k]$ solves both problems,

giving the optimal values of $\lambda_1 + \ldots + \lambda_k$

and $\lambda_{k+1} + \ldots + \lambda_n$, respectively.

Compare with the Eckart–Young Problems

minimize $F(X_k, Y_k) = \|A - X_k R_k Y_k^T\|_F^2$

subject to $X_k \in \mathbf{X}_k$ and $Y_k \in \mathbf{Y}_k$.

is equivalent to

maximize $\|X_k^T A Y_k\|_F^2$

subject to $X_k \in \mathbf{X}_k$ and $Y_k \in \mathbf{Y}_k$.

The SVD matrices U_k and V_k solves both problems,

giving the optimal values of $\sigma_{k+1}^2 + \ldots + \sigma_n^2$

and $\sigma_1^2 + \sigma_2^2 + \ldots + \sigma_k^2$, respectively.

The Eckart–Young Theorem

and Ky Fan’s maximum principle

are “two sides of the same coin”.

Is there an extended maximum principle

(for rectangular matrices) from which

one can derives both results ?

Corollary 2: Ky Fan's **minimum** problem

$$\textbf{minimize} \quad \textbf{trace}(Y_k S Y_k^T)$$

$$\textbf{subject to} \quad Y_k \in \mathbf{Y}_k \ ,$$

is equivalent to

$$\textbf{maximize} \quad \textbf{trace}(S - Y_k^T S_k Y_k)$$

$$\textbf{subject to} \quad Y_k \in \mathbf{Y}_k \ .$$

The matrix $V_k = [\mathbf{v}_{n-k+1}, \ldots, \mathbf{v}_n]$ solves both problems, giving the optimal value of $\lambda_{n-k+1} + \ldots + \lambda_n$ and $\lambda_1 + \ldots + \lambda_{n-k}$, respectively.

Can we extend Ky Fan’s principles

from eigenvalues of symmetric matrices

to singular values of rectangular matrices?

Notation: $1 \le m^* \le m$, $1 \le n^* \le n$,

$\mathbf{X}_{m^*}$ - denotes the set of all real $m \times m^*$ orthogonal matrices X_{m^*} ,

$$X_{m^*} = [\mathbf{x}_1 , \mathbf{x}_2 , \ldots , \mathbf{x}_{m^*}] \quad , \quad X_{m^*}^T X_{m^*} = I$$

$\mathbf{Y}_{n^*}$ - denotes the set of all real $n \times n^*$ orthogonal matrices Y_{n^*} ,

$$Y_{n^*} = [\mathbf{y}_1 , \mathbf{y}_2 , \ldots , \mathbf{y}_{n^*}] \quad , \quad Y_{n^*}^T Y_{*} = I$$

Reminder :

Given $X_{m*} \in \mathbf{X}_{m*}$ and $Y_{n*} \in \mathbf{Y}_{m*}$,

the **m* x n*** matrix

$$(X_{m*})^T A\, Y_{n*} = (\, x_i^T A y_j \,)$$

is called "**Orthogonal Quotients Matrix**".

Notations :

The **singular values** of the

Orthogonal Quotients Matrix

$$(X_{m^*})^T A \, Y_{n^*} = (x_i^T A y_j)$$

are denoted as

$$\eta_1 \geq \eta_2 \geq \dots \geq \eta_k \geq \mathbf{0} \, ,$$

where

$$\mathbf{k = min \{ m^*, n^* \}} \quad .$$

Question :

Which choice of orthogonal matrices

$$X_{m^*} \in \mathbf{X}_{m^*} \quad \text{and} \quad Y_{n^*} \in \mathbf{Y}_{m^*} \ ,$$

maximizes (or minimizes) the sum

$$(\eta_1)^p + (\eta_2)^p + \dots + (\eta_k)^p$$

where $p > 0$ is a given positive constant .

(Maximizing (or minimizing) the "Schatten p-norm" of $(X_{m^*})^T A Y_{n^*}$)

An Extended **Maximum** Principle:

The SVD matrices

$$U_{m^*} = [\mathbf{u}_1, \mathbf{u}_2, \ \ldots \ , \mathbf{u}_{m^*}] \quad \text{and} \quad V_{n^*} = [\mathbf{v}_1, \mathbf{v}_2, \ \ldots \ , \mathbf{v}_{n^*}]$$

solve the problem

$$\textbf{maximize} \quad F(X_{m^*}, Y_{n^*}) = (\eta_1)^p + (\eta_2)^p + \ldots + (\eta_k)^p$$

$$\textbf{subject to} \quad X_{m^*} \in \mathbb{X}_{m^*} \quad \text{and} \quad Y_{n^*} \in \mathbb{Y}_{n^*} \ ,$$

for any positive power $\mathbf{p > 0}$,

giving the optimal value of $(\sigma_1)^p + (\sigma_2)^p + \ldots + (\sigma_k)^p$.

An Extended Maximum Principle:

The SVD matrices

$U_{m^*} = [\mathbf{u}_1, \mathbf{u}_2, \ \ldots \ , \mathbf{u}_{m^*}]$ and $V_{n^*} = [\mathbf{v}_1, \mathbf{v}_2, \ \ldots \ , \mathbf{v}_{n^*}]$

solve the problem

maximize $F(X_{m^*}, Y_{n^*}) = \| (X_{m^*})^T A Y_{n^*} \|$

subject to $X_{m^*} \in \mathbf{X}_{m^*}$ and $Y_{n^*} \in \mathbf{Y}_{n^*}$,

for any unitarily invariant matrix norm.

The proof

is based on "rectangular" versions of

Cauchy Interlace Theorem ,

and

Poincare Separation Theorem .

The validity for Unitarily Invariant matrix norms follows from

Ky Fan Dominance Theorem .

Example 1 : When $\mathbf{p=1}$ the SVD matrices

$U_{m*} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{m*}]$ and $V_{n*} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n*}]$

solve the **"Rectangular Ky Fan problem"**

maximize $F(X_{m*}, Y_{n*}) = \eta_1 + \eta_2 + \ldots + \eta_k$

subject to $X_{m*} \in \mathbf{X}_{m*}$ and $Y_{n*} \in \mathbf{Y}_{n*}$,

giving the optimal value of

$$\sigma_1 + \sigma_2 + \ldots + \sigma_k \ .$$

Example 2 : When **p=2** the SVD matrices

$U_{m*} = [\mathbf{u}_1 , \mathbf{u}_2 , \ldots , \mathbf{u}_{m*}]$ and $V_{n*} = [\mathbf{v}_1 , \mathbf{v}_2 , \ldots , \mathbf{v}_{n*}]$

solve the **"rectangular Eckart–Young problem"**

maximize $\mathbf{F}(X_{m*}, Y_{n*}) = \| (X_{m*})^T A\, Y_{n*} \|_{\mathbf{F}}^{\mathbf{2}}$

subject to $X_{m*} \in \mathbb{X}_{m*}$ and $Y_{n*} \in \mathbb{Y}_{n*}$,

giving the optimal value of

$$(\sigma_1)^2 + (\sigma_2)^2 + \ldots + (\sigma_k)^2 \ .$$

An Extended Minimum Principle

Question: Can we prove a similar minimum principle ?

Answer: **Yes**, but building the solution matrices is more subtle .

The main difficulty here is to characterize cases in which the optimal value differs from zero.

An Extended Minimum Principle:

Here we consider the problem

$$\textbf{minimize } F(X_{m^*}, Y_{n^*}) = (\eta_1)^p + (\eta_2)^p + \ldots + (\eta_k)^p$$

$$\textbf{subject to} \quad X_{m^*} \in \mathbf{X}_{m^*} \text{ and } Y_{n^*} \in \mathbf{Y}_{n^*} ,$$

for any positive power $\mathbf{p > 0}$.

The solution matrices are obtained by deleting some columns from the SVD matrices

$$U = [\mathbf{u}_1 , \mathbf{u}_2 , \ldots , \mathbf{u}_m] \text{ and } V = [\mathbf{v}_1 , \mathbf{v}_2 , \ldots , \mathbf{v}_n] .$$

An Extended **Minimum** Principle:

Here we consider the problem

$$\textbf{minimize } F(X_{m*}, Y_{n*}) = \|(X_{m*})^T A Y_{n*}\|$$

$$\textbf{subject to} \quad X_{m*} \in \mathbf{X}_{\mathbf{m*}} \text{ and } Y_{n*} \in \mathbf{Y}_{\mathbf{n*}} ,$$

for any unitarily invariant matrix norm.

The solution matrices are obtained by deleting some columns from the SVD matrices

$$U = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m] \text{ and } V = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n].$$

Summary †

* Orthogonal Quotients Matrices.

* The **Eckart – Young** Theorem.

* The Orthogonal Quotients Equality.

* **Ky Fan**'s extremum principles.

* Extended Extremum Principles.

† A. Dax, "On extremum properties of orthogonal quotients matrices", Lin. Alg. and its Applic., 432(2010), pp. 1234 – 1257.

The END

Thank You

Example 1* : When $\mathbf{p = 1}$ and $\mathbf{m^* = n^* = k}$

the SVD matrices

$$U_k = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k] \text{ and } V_k = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k]$$

solve the maximum trace problem

$$\textbf{maximize } \mathbf{F}(X_k, Y_k) = \textbf{trace}\,(\,(X_k)^T A\, Y_k\,)$$

$$\textbf{subject to } X_k \in \mathbb{X}_k \text{ and } Y_k \in \mathbb{Y}_k ,$$

giving the optimal value of

$$\sigma_1 + \sigma_2 + \ldots + \sigma_k .$$

* See also Horn & Johnson, "Topics in Matrix Analysis", p. 195.