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.

Orthogonal Quotient Matrices

$A$ a real $m \times n$ matrix

$X_{m^*}$ a real $m \times m^*$ orthogonal matrix

$X_{m^*} = [x_1, x_2, \ldots, x_{m^*}]$, $X_{m^*}^T X_{m^*} = I$.

$Y_{n^*}$ a real $n \times n^*$ orthogonal matrix

$Y_{n^*} = [y_1, y_2, \ldots, y_{n^*}]$, $Y_{n^*}^T Y_{n^*} = I$.

Then the $m^* \times n^*$ 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 $n \times n$ matrix

$Y_k$ a real $n \times k$ orthogonal matrix

$Y_k = [y_1, y_2, \ldots, y_k]$, $Y_k^T Y_k = I$.

Then the $k \times k$ matrix

$Y_k^T S Y_k = (y_i^T S y_j)$

is called a “Rayleigh Quotient Matrix”.

Matrices of this form play an important role in the Rayleigh-Ritz procedure and in Krylov subspace methods.
Example: Lanczos Algorithm

\[ Y_k^T S Y_k = T_k \]

- \( T_k \) a real \( k \times k \) tridiagonal matrix
- \( Y_k \) a real \( n \times k \) orthogonal matrix

The columns of \( Y_k \) form an orthonormal basis of a Krylov subspace,

\[ \text{Span} \{ x, Sx, S^2x, \ldots, S^{k-1}x \}, \]

for some starting vector \( x \).
Ky Fan’s Maximum Principle  (1949)†

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

\[
\max \left\{ \text{trace} \left( Y_k^T S Y_k \right) \mid Y_k \in Y_k \right\}
\]

\( S \) a symmetric positive semi-definite \( n \times n \) matrix.

\( Y_k \) denotes the set of orthogonal \( n \times k \) 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

\[
\begin{align*}
\text{minimize} & \quad F(B) = \|A - B\|_F^2 \\
\text{subject to} & \quad \text{rank}(B) \leq k
\end{align*}
\]

(Also called the Schmidt–Mirsky Theorem.)
The Frobenius matrix norm

\[ A = \begin{pmatrix} a_{ij} \end{pmatrix} \] a real \( m \times n \) matrix, \( m \geq n \).

\[ \| A \|_F = \left( \sum \sum (a_{ij})^2 \right)^{1/2} \]

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

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

where

\[ \sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0, \]

denote the singular values of \( A \).
The Singular Value Decomposition

\[ A = U \Sigma V^T \]

\[ \Sigma = \text{diag} \{ \sigma_1, \sigma_2, \ldots, \sigma_p \} \], \quad p = \min \{m, n\} 

\[ U = [u_1, u_2, \ldots, u_p] \], \quad U^T U = I 

\[ V = [v_1, v_2, \ldots, v_p] \], \quad V^T V = I 

\[ A V = U \Sigma \quad A^T U = V \Sigma \]

\[ A v_j = \sigma_j u_j \], \quad A^T u_j = \sigma_j v_j \quad j = 1, \ldots, p \]
**Rank-k truncated SVD**

\[ T_k(A) = U_k D_k V_k^T \]

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

\[ U_k = [u_1, u_2, \ldots, u_k] \quad U_k^T U_k = I \]

\[ V_k = [v_1, v_2, \ldots, v_k] \quad V_k^T V_k = I \]

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

\[
\text{minimize} \quad F(B) = \|A - B\|_F^2 \\
\text{subject to} \quad \text{rank}(B) \leq k
\]

(Also called the Schmidt–Mirsky Theorem.)
Unitarily Invariant matrix norms

\[ A = \left( \begin{array}{c c c}
        a_{ij}
      \end{array} \right) \quad \text{a real \ m \times \ n \ matrix}, \]

\[ X \quad \text{a real \ m \times m \ unitary \ matrix,} \quad X^T X = I. \]

\[ Y \quad \text{a real \ n \times n \ unitary \ matrix,} \quad Y^T Y = I. \]

\[ \|A\| = \|X^T A\| = \|A Y\| = \|X^T A Y\|. \]
Unitarily Invariant matrix norms

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

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

\[ 1 \leq p < \infty \]

\[ \|A\| = \sigma_1 + \sigma_2 + \ldots + \sigma_k \text{ Ky Fan k–norm, } k = 1, \ldots, n. \]

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

\[ \|A\|_2 = \sigma_1 = \max \{ \sigma_j \} \text{ the 2–norm.} \]

\[ j = 1, \ldots, n \]
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** \( F(B) = \|A - B\| \)

**subject to** \( \text{rank}(B) \leq k \),

for any unitarily invariant matrix norm.
Rank-\( k \) Matrices

\[
B = X_k R Y_k^T
\]

where

\( R \) is a \( k \times k \) matrix

\[
X_k = [x_1, x_2, \ldots, x_k], \quad X_k^T X_k = I
\]

\[
Y_k = [y_1, y_2, \ldots, y_k], \quad Y_k^T Y_k = I
\]
The Eckart–Young Problem

can be rewritten as

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

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

where \quad B = X_k R Y_k^T \quad \text{and} \quad R \text{ is a } k \times k \text{ matrix.}
**Theorem:** Let $X_k$ and $Y_k$ be a pair of orthogonal matrices as above, then the related “Orthogonal Quotients Matrix”

$$X_k^T A Y_k = \left( x_i^T A y_j \right)$$

solves the problem

$$\text{minimize } F(R) = \| A - X_k R Y_k^T \|_F$$
**Notation:**

$X_k$ - denotes the set of all real $m \times k$ orthogonal matrices $X_k$, 
$X_k = [x_1, x_2, \ldots, x_k]$,  
$X_k^T X_k = I$

$Y_k$ - denotes the set of all real $n \times k$ orthogonal matrices $Y_k$, 
$Y_k = [y_1, y_2, \ldots, y_k]$,  
$Y_k^T Y_k = I$
Corollary 1: The Eckart–Young Problem can be rewritten as

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

subject to $$X_k \in X_k$$ and $$Y_k \in Y_k$$,

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,

\[ X_k \in X_k \quad \text{and} \quad Y_k \in Y_k, \]

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

where \( R_k \) is the related orthogonal quotients matrix

\[ R_k = X_k^T A Y_k. \]
Corollary 2: The Eckart–Young Problem

\[
\text{minimize } F(X_k, Y_k) = \| A - X_k R_k Y_k^T \|_F^2 \\
\text{subject to } X_k \in \mathcal{X}_k \text{ and } Y_k \in \mathcal{Y}_k.
\]

is equivalent to

\[
\text{maximize } \| X_k^T A Y_k \|_F^2 \\
\text{subject to } X_k \in \mathcal{X}_k \text{ and } Y_k \in \mathcal{Y}_k.
\]

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.
Returning to symmetric matrices

Can we extend the Orthogonal Quotients Equality to Ky Fan extremum problems?
The Spectral Decomposition

\[ S = (s_{ij}) \quad \text{a symmetric positive semi-definite } n \times n \quad \text{matrix} \]

With eigenvalues \( \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0 \)

and eigenvectors \( v_1, v_2, \ldots, v_n \)

\[ S v_j = \lambda_j v_j, \quad j = 1, \ldots, n. \quad S V = V D \]

\[ V = [v_1, v_2, \ldots, v_n], \quad V^T V = V V^T = I \]

\[ D = \text{diag} \{ \lambda_1, \lambda_2, \ldots, \lambda_n \} \]

\[ S = V D V^T = \sum \lambda_j v_j v_j^T \]
Ky Fan’s **Maximum Principle**

$S$, a symmetric positive semi-definite $n \times n$ matrix

$Y_k$, the set of orthogonal $n \times k$ matrices

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

Solution is obtained for the Spectral matrix

$$V_k = [v_1, v_2, \ldots, v_k].$$

which is related to the largest $k$ eigenvalues.
Ky Fan’s **Minimum Principle**

$S$ a symmetric $n \times n$ matrix.

$Y_k$ the set of orthogonal $n \times k$ matrices.

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

Solution is obtained for the Spectral matrix

$$V_k = [v_{n-k+1}, \ldots, v_n],$$

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

\[ \text{trace}(S - Y_k S_k Y_k^T) = \text{trace}(S) - \text{trace}(S_k), \]

where \( \text{trace}(S_k) = \text{trace}(Y_k^T S Y_k) = \sum y_j^T S y_j . \)
Corollary 1: Ky Fan’s maximum problem

\[
\begin{align*}
\text{maximize} & \quad \text{trace}(Y_k S Y_k^T) \\
\text{subject to} & \quad Y_k \in \mathcal{Y}_k,
\end{align*}
\]

is equivalent to

\[
\begin{align*}
\text{minimize} & \quad \text{trace}(S - Y_k S_k Y_k^T) \\
\text{subject to} & \quad Y_k \in \mathcal{Y}_k.
\end{align*}
\]

The Spectral matrix \( V_k = [v_1, v_2, \ldots, v_k] \) 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

\[
\text{minimize } F(X_k, Y_k) = \|A - X_k R_k Y_k^T\|_F^2 \\
\text{subject to } X_k \in X_k \text{ and } Y_k \in Y_k.
\]

is equivalent to

\[
\text{maximize } \|X_k^T A Y_k\|_F^2 \\
\text{subject to } X_k \in X_k \text{ and } Y_k \in Y_k.
\]

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 derive both results?
Corollary 2: Ky Fan’s minimum problem

\[
\begin{align*}
\text{minimize} & \quad \text{trace} (Y_k S Y_k^T) \\
\text{subject to} & \quad Y_k \in \mathcal{Y}_k
\end{align*}
\]

is equivalent to

\[
\begin{align*}
\text{maximize} & \quad \text{trace} (S - Y_k^T S_k Y_k) \\
\text{subject to} & \quad Y_k \in \mathcal{Y}_k
\end{align*}
\]

The matrix \( V_k = [v_{n-k+1}, \ldots, v_n] \) 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: \( 1 \leq m^* \leq m , \quad 1 \leq n^* \leq n , \)

\( X_{m^*} \) - denotes the set of all real \( m \times m^* \) orthogonal matrices \( X_{m^*} \),
\( X_{m^*} = [x_1, x_2, \ldots, x_{m^*}] \), \( X_{m^*}^T X_{m^*} = I \)

\( Y_{n^*} \) - denotes the set of all real \( n \times n^* \) orthogonal matrices \( Y_{n^*} \),
\( Y_{n^*} = [y_1, y_2, \ldots, y_{n^*}] \), \( Y_{n^*}^T Y_{n^*} = I \)
Reminder:

Given \( X_{m^*} \in X_{m^*} \) and \( Y_{n^*} \in Y_{m^*} \),

the \( m^* \times n^* \) 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

\[(X_{m*})^TAY_{n*} = (x_i^T Ay_j)\]

are denoted as

\[\eta_1 \geq \eta_2 \geq \ldots \geq \eta_k \geq 0,\]

where

\[k = \min \{ m^*, n^* \} .\]
Question:
Which choice of orthogonal matrices

\[ X_{m^*} \in X_{m^*} \quad \text{and} \quad Y_{n^*} \in Y_{m^*}, \]

maximizes (or minimizes) the sum

\[ (\eta_1)^p + (\eta_2)^p + \ldots + (\eta_k)^p \]

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

\[ U_{m^*} = [u_1, u_2, \ldots, u_{m^*}] \quad \text{and} \quad V_{n^*} = [v_1, v_2, \ldots, v_{n^*}] \]

solve the problem

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

\text{subject to} \quad X_{m^*} \in X_{m^*} \quad \text{and} \quad Y_{n^*} \in Y_{n^*},

for any positive power \( p > 0 \),

giving the optimal value of \( (\sigma_1)^p + (\sigma_2)^p + \ldots + (\sigma_k)^p \).
An Extended **Maximum Principle**:  

The SVD matrices  

\[ U_{m*} = [u_1, u_2, \ldots, u_{m*}] \quad \text{and} \quad V_{n*} = [v_1, v_2, \ldots, v_{n*}] \]

solve the problem  

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

subject to  

\( X_{m*} \in X_{m*} \quad \text{and} \quad Y_{n*} \in Y_{n*} \),

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 $p=1$ the SVD matrices

$$U_{m^*} = [u_1, u_2, \ldots, u_{m^*}] \quad \text{and} \quad V_{n^*} = [v_1, v_2, \ldots, v_{n^*}]$$

solve the “Rectangular Ky Fan problem”

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

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

\text{giving the optimal value of}

$$\sigma_1 + \sigma_2 + \ldots + \sigma_k.$$
Example 2: When \( p=2 \) the SVD matrices

\[
U_{m^*} = [u_1, u_2, \ldots, u_{m^*}] \quad \text{and} \quad V_{n^*} = [v_1, v_2, \ldots, v_{n^*}]
\]
solve the “rectangular Eckart–Young problem”

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

subject to \( X_{m^*} \in X_{m^*} \) and \( Y_{n^*} \in Y_{n^*} \),

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

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

subject to \( X_{m*} \in X_{m*} \) and \( Y_{n*} \in Y_{n*} \),

for any positive power \( p > 0 \).

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

\[
U = [u_1, u_2, \ldots, u_m] \text{ and } V = [v_1, v_2, \ldots, v_n].
\]
An Extended **Minimum** Principle:

Here we consider the problem

\[
\text{minimize } F(X_{m*}, Y_{n*}) = \|(X_{m*})^T A Y_{n*}\|
\]

subject to \( X_{m*} \in X_{m*} \) and \( Y_{n*} \in Y_{n*} \),

for any unitarily invariant matrix norm.

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

\[
U = [u_1, u_2, \ldots, u_m] \quad \text{and} \quad V = [v_1, v_2, \ldots, v_n].
\]
Summary †

* Orthogonal Quotients Matrices.
* The Eckart–Young Theorem.
* The Orthogonal Quotients Equality.
* Ky Fan’s extremum principles.
* Extended Extremum Principles.

The END

Thank You
Example 1*: When \( p = 1 \) and \( m^* = n^* = k \) the SVD matrices

\[
U_k = [u_1, u_2, \ldots, u_k] \quad \text{and} \quad V_k = [v_1, v_2, \ldots, v_k]
\]

solve the maximum trace problem

\[
\text{maximize } F(X_k, Y_k) = \text{trace} \left( (X_k)^T A Y_k \right)
\]

subject to \( X_k \in X_k \) and \( Y_k \in Y_k \),

giving the optimal value of

\[
\sigma_1 + \sigma_2 + \ldots + \sigma_k
\]

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