Scientific
Research

From Eigenvalues to Singular Values: A Review

Achiya Dax
Hydrological Service, Jerusalem, Israel
Email: dax20@water.gov.il

ABSTRACT

The analogy between eigenvalues and singular values has many faces. The current review brings together several examples of this analogy. One example regards the similarity between Symmetric Rayleigh Quotients and Rectangular Rayleigh Quotients. Many useful properties of eigenvalues stem are from the Courant-Fischer minimax theorem, from Weyl's theorem, and their corollaries. Another aspect regards "rectangular" versions of these theorems. Comparing the properties of Rayleigh Quotient matrices with those of Orthogonal Quotient matrices illuminates the subject in a new light. The Orthogonal Quotients Equality is a recent result that converts Eckart-Young's minimum norm problem into an equivalent maximum norm problem. This exposes a surprising link between the Eckart-Young theorem and Ky Fan's maximum principle. We see that the two theorems reflect two sides of the same coin: there exists a more general maximum principle from which both theorems are easily derived. Ky Fan has used his extremum principle (on traces of matrices) to derive analog results on determinants of positive definite Rayleigh Quotients matrices. The new extremum principle extends these results to Rectangular Quotients matrices. Bringing all these topics under one roof provides new insight into the fascinating relations between eigenvalues and singular values.

1. Introduction

Let G be a real symmetric $n \times n$ matrix and let $\boldsymbol{x} = (x_1, \cdots, x_n)^{\mathrm{T}} \in \mathbb{R}^n$ be a given nonzero vector. Then the well-known Rayleigh Quotient is defined as

$$\rho = \rho(\boldsymbol{x}, G) = \boldsymbol{x}^{\mathrm{T}} G \boldsymbol{x} / \boldsymbol{x}^{\mathrm{T}} \boldsymbol{x}. \qquad (1.1)$$

One motivation behind this definition lies in the following observation. Let τ be a given real number. Then there exists an eigenvalue λ of G such that

$$|\lambda - \tau| \le \|G\boldsymbol{x} - \tau \boldsymbol{x}\|_2 / \|\boldsymbol{x}\|_2, \qquad (1.2)$$

and the value of τ that solves the minimum norm problem

$$\text{minimize } f(\tau) = \|G\boldsymbol{x} - \tau \boldsymbol{x}\|_2 \qquad (1.3)$$

is given by ρ. In other words, ρ provides an estimate for an eigenvalue corresponding to $\boldsymbol{x}$. Combining this estimate with the inverse iteration yields the celebrated Rayleigh Quotient Iteration. Other related features are the Courant-Fischer minimax inequalities, Weyl monotonicity theorem, and many other results that stem from

these observations. In particular, the largest and the smallest eigenvalues of G satisfy

$$\lambda_1 = \sup\left\{ \boldsymbol{x}^{\mathrm{T}} G \boldsymbol{x} / \boldsymbol{x}^{\mathrm{T}} \boldsymbol{x} \,\middle|\, \boldsymbol{x} \in \mathbb{R}^n, \, \boldsymbol{x} \ne \boldsymbol{0} \right\} \qquad (1.4)$$

and

$$\lambda_n = \inf\left\{ \boldsymbol{x}^{\mathrm{T}} G \boldsymbol{x} / \boldsymbol{x}^{\mathrm{T}} \boldsymbol{x} \,\middle|\, \boldsymbol{x} \in \mathbb{R}^n, \, \boldsymbol{x} \ne \boldsymbol{0} \right\} \qquad (1.5)$$

respectively. For detailed discussion of the Rayleigh Quotient and its properties see, for example, [1-41].

The question that initiates our study is how to extend the definition of Rayleigh Quotient in order to estimate a singular value of a general rectangular matrix, where the term "rectangular" means that the matrix is not necessarily symmetric or square. More precisely, let A be a real $m \times n$ matrix, $m \ge n$, and let $\boldsymbol{u} \in \mathbb{R}^m$ and $\boldsymbol{v} \in \mathbb{R}^n$ be a pair of nonzero vectors. Then we seek a scalar function of $A, \boldsymbol{u}$, and $\boldsymbol{v}$, $\sigma(\boldsymbol{u}, A, \boldsymbol{v})$ say, whose value approximates the "corresponding" singular value of A. The answer is given by the Rectangular Quotient,

$$\sigma = \sigma(\boldsymbol{u}, A, \boldsymbol{v}) = |\boldsymbol{u}^{\mathrm{T}} A \boldsymbol{v}| / (\|\boldsymbol{u}\|_2 \|\boldsymbol{v}\|_2), \qquad (1.6)$$

where $\|u\|_2 = \left(u^{\mathrm{T}}u\right)^{1/2}$ denotes the Euclidean vector norm. The justifications behind this definition are given in Section 3. It is demonstrated there that the properties of the Rectangular Quotient (1.6) resemble (extend) those of the Rayleigh Quotient (1.1). Indeed, as our review is aimed to show, the above similarity reflects a more general rule: the optimality properties of Orthogonal Quotient matrices resemble (extend) those of Rayleigh Quotient matrices.

Let $X_{\tilde{m}} = \left[x_1, \cdots, x_{\tilde{m}}\right]$ be a real $m \times \tilde{m}$ matrix with $\tilde{m}$ orthonormal columns. Let $Y_{\tilde{n}} = \left[y_1, \cdots, y_{\tilde{n}}\right]$ be a real $n \times \tilde{n}$ matrix with $\tilde{n}$ orthonormal columns. Then an $\tilde{m} \times \tilde{n}$ matrix of the form

$$Z = X_{\tilde{m}}^{\mathrm{T}} A Y_{\tilde{n}} \qquad (1.7)$$

is called an Orthogonal Quotient matrix. Note that

$$Z_{ij} = x_i^{\mathrm{T}} A y_j, \qquad (1.8)$$

so the entries of Z have the same absolute value as the corresponding rectangular quotients. Matrices of the form (1.7) can be viewed as "rectangular" Rayleigh Quotient matrices. The traditional definition of symmetric Rayleigh Quotient matrices refers to symmetric matrices of the form

$$Y_{\tilde{n}}^{\mathrm{T}} G Y_{\tilde{n}}, \qquad (1.9)$$

where G and $Y_{\tilde{n}}$ are defined as above, e.g., [20,30,36]. Symmetric Rayleigh Quotient matrices of this form are sometimes called sections. A larger class of Rayleigh Quotients matrices is considered in [34]. These matrices have the form

$$C^I W C, \qquad (1.10)$$

where W is a general (nonnormal) square matrix of order n. The $n \times k$ matrix C is assumed to be of full column rank, and the $k \times n$ matrix C^I denotes a left inverse of C. That is, a matrix satisfying $C^I C = I$. Matrices of the forms (1.9) and (1.10) play important roles in the Rayleigh-Ritz procedure and in Krylov subspace methods, e.g., [30,34]. In this context there is rich literature on residual bounds for eigenvalues and eigenspaces. See, for example, [19-21,30,32,34,36]. Another applications of Rayleigh Quotient matrices arise in optimization algorithms that try to keep their approximations on a specific Stiefel manifold, e.g., [6,7,9,35].

A third class of Rayleigh Quotient matrices is obtained from (1.7) by taking $\tilde{m} = \tilde{n} = k$. These matrices are involved in residual bounds for singular values and singular spaces, e.g., [2,21].

However, our review turns into different directions. It is aimed to explore the optimality properties of Orthogonal Quotients matrices. Comparing these properties with those of symmetric Rayleigh Quotients matrices reveals

highly interesting observations. At the heart of these observations stands a surprising relationship between Eckart-Young's minimum norm theorem [5] and Ky Fan's maximum principle [10].

The Eckart-Young theorem considers the problem of approximating one matrix by another matrix of a lower rank. The solution of this problem is also attributed to Schmidt [31]. See [17, pp.~137,138] and [33, p.~76]. The need for low-rank approximations of a matrix is a fundamental problem that arises in many applications, e.g., [3-5,8,14,15,18,33]. The maximum principle of Ky Fan considers the problem of maximizing the trace of a symmetric Rayleigh Quotient matrix. It is also a well-known result that has many applications, e.g., [1,10-12, 16,22,28]. Yet, so far, the two theorems have always been considered as independent and unrelated results which are based on different arguments. The Orthogonal Quotients Equality is a recent result that converts Eckart-Young's minimum norm problem into an equivalent maximum norm problem. This exposes a surprising similarity between the Eckart-Young theorem and Ky Fan's maximum principle. We see that the two theorems reflect two sides of the same coin: there exists a more general maximum rule from which both theorems are easily derived.

The plan of our review is as follows. It starts by introducing some necessary notations and facts. Then it turns to expose the basic properties of the Rectangular Quotient (1.6), showing that it solves a number of least norm problems that resemble (1.3). An error bound, similar to (1.2), enables us to bound the distance between σ and the closest singular value of A.

Another aspect of the analogy between eigenvalues and singular values is studied in Section 4, in which we consider "rectangular" versions of the Courant-Fischer minimax theorem and Weyl theorem. This paves the way for "traditional" proof of Eckart-Young theorem. Then, using Ky Fan's dominance theorem, it is easy to conclude Mirsky's theorem.

The relation between symmetric Rayleigh Quotient matrices and Orthogonal Quotients matrices is studied in Section 5. It is shown there that the least squares properties of Orthogonal Quotient matrices resemble those of symmetric Rayleigh-Quotient matrices. One consequence of these properties is the Orthogonal Quotients Equality, which is derived in Section 6. As noted above, this equality turns the Eckart-Young least squares problem into an equivalent maximum problem, which attempts to maximize the Frobenius norm of an Orthogonal Quotients matrix of the form (1.7).

The symmetric version of the Orthogonal Quotients Equality considers the problem of maximizing (or minimizing) traces of symmetric Rayleigh Quotients matrices. The solution of these problems is given by the celebrated

Ky Fan's Extremum Principles. The similarity between the maximum form of Eckart-Young theorem and Ky Fan's maximum principle suggests that both observations are special cases of a more general extremum principle. The derivation of this principle is carried out in Section 8. It is shown there that both results are easily concluded from the extended maximum principle.

The review ends by discussing some consequences of the extended principle. One consequence is a minimum-maximum equality that relates Mirsky's minimum norm problem with the extended maximum problem. A second type of consequence is about traces of Orthogonal Quotients matrices. The results of Ky Fan on traces of symmetric Rayleigh Quotients matrices [10] were extended in his latter papers [11,12] to products of eigenvalues and determinants. The new extremum principle enables the extension of these properties to Orthogonal Quotients matrices.

The current review brings together several old and new results. The "old" results come with appropriate references. In contrast, the "new" results come without references, as most of them are taken from a recent research paper [3] by this author. Yet the current paper derives a number of contributions which are not included in [3]. One contribution regards the extension of Theorem 11 behind the Frobenius norm. Another contribution is the minimum-maximum equality, which is introduced in Section 9. The main difference between [3] and this essay lies in their concept. The first one is a research paper that is aimed at establishing the extended extremum principle. The review exposes some fascinating features of the analogy between eigenvalues and singular values. For this purpose we present several apparently unrelated results. Putting all these topics under one roof gives a better insight into these relations. The description of the results concentrates on real valued matrices and vectors. This simplifies the presentation and helps to focus on the main ideas. The treatment of the complex-valued case should be quite obvious.

2. Notations and Basic Facts

In this section we introduce notations and facts which are needed for coming discussions. As before A denotes a real $m \times n$ matrix with $m \geq n$. Let

$$A = USV^{\mathrm{T}} \quad (2.1)$$

be an SVD of A, where $U = [u_1, \cdots, u_m]$ is an $m \times m$ orthogonal matrix, $V = [v_1, \cdots, v_n]$ is an $n \times n$ orthogonal matrix, and $S = \mathrm{diag}\{\sigma_1, \cdots, \sigma_n\}$ is an $m \times n$ diagonal matrix. The singular values of A are assumed to be nonnegative and sorted to satisfy

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0. \quad (2.2)$$

The columns of U and V are called left singular vectors and right singular vectors, respectively. These vectors are related by the equalities

$$A v_j = \sigma_j u_j \quad \text{and} \quad A^{\mathrm{T}} u_j = \sigma_j v_j, \quad j = 1, \cdots, n. \quad (2.3)$$

A further consequence of (2.1) is the equality

$$A = \sum_{j=1}^{n} \sigma_j u_j v_j^{\mathrm{T}}. \quad (2.4)$$

Moreover, let r denote the rank of A. Then, clearly,

$$\sigma_1 \geq \cdots \geq \sigma_r > 0 \quad \text{and} \quad \sigma_j = 0$$
$$\text{for} \quad j = r+1, \cdots, n. \quad (2.5)$$

So (2.4) can be rewritten as

$$A = \sum_{j=1}^{r} \sigma_j u_j v_j^{\mathrm{T}}. \quad (2.6)$$

Let the matrices

$$U_k = [u_1, \cdots, u_k] \in \mathbb{R}^{m \times k}$$
$$\text{and} \quad V_k = [v_1, \cdots, v_k] \in \mathbb{R}^{n \times k} \quad (2.7)$$

be constructed from the first k columns of U and V, respectively. Let $S_k = \mathrm{diag}\{\sigma_1, \cdots, \sigma_k\}$ be a $k \times k$ diagonal matrix. Then the matrix

$$T_k = U_k S_k V_k^{\mathrm{T}} = \sum_{j=1}^{k} \sigma_j u_j v_j^{\mathrm{T}} \quad (2.8)$$

is called rank-k Truncated SVD of A.

Let a_{ij}, u_{ij}, v_{ij}, denote the (i,j) entries of the matrices A, U, V, respectively. Then (2.4) indicates that

$$a_{ii} = \sum_{j=1}^{n} \sigma_j u_{ij} v_{ij} \quad \text{for } i = 1, \cdots, n, \quad (2.9)$$

and

$$\sum_{i=1}^{n} |a_{ii}| \leq \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_j |u_{ij}| \cdot |v_{ij}|$$
$$= \sum_{j=1}^{n} \sigma_j \sum_{i=1}^{n} |u_{ij}| \cdot |v_{ij}| \leq \sum_{j=1}^{n} \sigma_j. \quad (2.10)$$

where the last inequality follows from the Cauchy-Schwarz inequality and the fact that the columns of U and V have unit length.

Another useful property regards the concepts of majorization and unitarily invariant norms. Recall that a matrix norm $\|\cdot\|$ on $\mathbb{R}^{m \times n}$ is called unitarily invariant if the equalities

$$\|A\| = \|X^{\mathrm{T}} A\| = \|AY\| = \|X^{\mathrm{T}} AY\| \quad (2.11)$$

are satisfied for any matrix $A \in \mathbb{R}^{m \times n}$, and any pair of unitary matrices $X \in \mathbb{R}^{m \times m}$ and $Y \in \mathbb{R}^{n \times n}$. Let B and

C be a given pair of $m \times n$ matrices with singular values

$$\beta_1 \geq \beta_2 \geq \cdots \geq \beta_n \geq 0$$
$$\text{and } \gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_n \geq 0,$$

respectively. Let $\beta = (\beta_1, \cdots, \beta_n)^T$ and $\gamma = (\gamma_1, \cdots, \gamma_n)^T$ denote the corresponding n-vectors of singular values. Then the weak majorization relation $\beta <_\omega \gamma$ means that these vectors satisfy the inequalities

$$\sum_{j=1}^{k} \beta_j \leq \sum_{j=1}^{k} \gamma_j \quad \text{for } k = 1, \cdots, n. \tag{2.12}$$

In this case we say that β is weakly majorized by γ, or that the singular values of B are weakly majorized by those of C. The Dominance Theorem of Ky Fan [11] relates these two concepts. It says that if the singular values of B are majorized by those of C then the inequality

$$\|B\| \leq \|C\| \tag{2.13}$$

holds for any unitarily invariant norm. For detailed proof of this fact see, for example, [1,11,17,22]. The most popular example of a unitarily invariant norm is, perhaps, the Frobenius matrix norm

$$\|A\|_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2 \right)^{1/2}, \tag{2.14}$$

which satisfies

$$\|A\|_F^2 = \text{trace}(A^T A) = \text{trace}(A A^T) = \sum_{j=1}^{n} \sigma_j^2. \tag{2.15}$$

Other examples are the Schatten p-norms,

$$\|A\| = \left(\sum_{j=1}^{n} \sigma_j^p \right)^{1/p}, \quad 1 \leq p < \infty, \tag{2.16}$$

and Ky Fan k-norms,

$$\|A\| = \sum_{j=1}^{k} \sigma_j, \quad k = 1, \cdots, n. \tag{2.17}$$

The trace norm,

$$\|A\|_{\text{tr}} = \sum_{j=1}^{n} \sigma_j \tag{2.18}$$

is obtained for $k = n$ and $p = 1$, while the spectral norm (the 2-norm)

$$\|A\|_2 = \sigma_1 = \max_j \sigma_j \tag{2.19}$$

corresponds to $k = 1$ and $p = \infty$.

Finally, let $\tilde{m}$ and $\tilde{n}$ be a pair of positive integers such that

$$1 \leq \tilde{m} \leq m \quad \text{and} \quad 1 \leq \tilde{n} \leq n.$$

Then

$$\mathbb{X}_{\tilde{m}} = \left\{ X_{\tilde{m}} \,\middle|\, X_{\tilde{m}} = [x_1, x_2, \cdots, x_{\tilde{m}}] \in \mathbb{R}^{m \times \tilde{m}} \text{ and } X_{\tilde{m}}^T X_{\tilde{m}} = I \right\}, \tag{2.20}$$

and

$$\mathbb{Y}_{\tilde{n}} = \left\{ Y_{\tilde{n}} \,\middle|\, Y_{\tilde{n}} = [y_1, y_2, \cdots, y_{\tilde{n}}] \in \mathbb{R}^{n \times \tilde{n}} \text{ and } Y_{\tilde{n}}^T Y_{\tilde{n}} = I \right\}, \tag{2.21}$$

denote the corresponding Stiefel manifolds. That is, $\mathbb{X}_{\tilde{m}}$ denotes the set of all real $m \times \tilde{m}$ matrices with orthonormal columns, while $\mathbb{Y}_{\tilde{n}}$ is the set of all real $n \times \tilde{n}$ matrices with orthonormal columns.

3. Rectangular Quotients

Let A be a real $m \times n$ matrix with $m \geq n$, and let $u = (u_1, \cdots, u_m)^T \in \mathbb{R}^m$ and $v = (v_1, \cdots, v_n)^T \in \mathbb{R}^n$ be a pair of nonzero vectors. To simplify the coming discussion we make the assumptions that $u^T A v \geq 0$, and that u and v are unit vectors. That is,

$$\|u\|_2 = (u^T u)^{1/2} = \left(\sum_{i=1}^{m} u_i^2 \right)^{1/2} = 1$$

$$\text{and } \|v\|_2 = (v^T v)^{1/2} = \left(\sum_{j=1}^{n} v_j^2 \right)^{1/2} = 1.$$

With these assumptions at hand the Rectangular Quotient (1.6) is reduced to the bilinear form

$$\sigma = \sigma(u, A, v) = u^T A v. \tag{3.1}$$

In this section we briefly derive the basic minimum norm properties that characterize this kind of bilinear forms. We shall start by noting that σ solves the one parameter minimization problem

$$\text{minimize } \alpha(\theta) = \|A - \theta u v^T\|_F^2. \tag{3.2}$$

This observation is a direct consequence of the equalities

$$\sum_{i=1}^{m} \sum_{j=1}^{n} (u_i v_j)^2 = \|u v^T\|_F^2 = \|u\|_2^2 \|v\|_2^2 = 1,$$

$$u^T A v = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} u_i v_j,$$

and

$$\|A - \theta u v^T\|_F^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} (a_{ij} - \theta u_i v_j)^2$$
$$= \|A\|_F^2 - 2\theta u^T A v + \theta^2. \tag{3.3}$$

Similar arguments show that $\sigma = u^T A v$ solves the least squares problems

$$\text{minimize } \beta(\theta) = \left\| A\boldsymbol{v} - \theta\boldsymbol{u} \right\|_2^2, \qquad (3.4)$$

and

$$\text{minimize } \gamma(\theta) = \left\| A^{\mathrm{T}}\boldsymbol{u} - \theta\boldsymbol{v} \right\|_2^2. \qquad (3.5)$$

Furthermore, substituting the optimal value of θ into (3.3) yields the Rectangular Quotient Equality

$$\left\| A - \left(\boldsymbol{u}^{\mathrm{T}}A\boldsymbol{v}\right)\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}} \right\|_F^2 = \left\| A \right\|_F^2 - \left(\boldsymbol{u}^{\mathrm{T}}A\boldsymbol{v}\right)^2, \qquad (3.6)$$

which means that solving the rank-one approximation problem

$$\text{minimize } f(\boldsymbol{x}, \boldsymbol{y}) = \left\| A - \boldsymbol{x}\boldsymbol{y}^{\mathrm{T}} \right\|_F^2 \qquad (3.7)$$

is equivalent to solving the maximization problem

$$\begin{aligned}\text{maximize } \quad & g(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{u}^{\mathrm{T}}A\boldsymbol{v} \\ \text{subject to } \quad & \left\| \boldsymbol{u} \right\|_2 = 1 \text{ and } \left\| \boldsymbol{v} \right\|_2 = 1. \end{aligned} \qquad (3.8)$$

Using the SVD of A the unit vectors in the last problem can be expressed in the form

$$\boldsymbol{u} = \sum_{i=1}^{m} \eta_i \boldsymbol{u}_i \text{ and } \boldsymbol{v} = \sum_{j=1}^{n} \xi_j \boldsymbol{v}_j, \text{ where } \sum_{i=1}^{m} \eta_i^2 = \sum_{j=1}^{n} \xi_j^2 = 1.$$

Therefore, since

$$\boldsymbol{u}_i^{\mathrm{T}}A\boldsymbol{v}_j = 0 \quad \text{for } i \neq j, \qquad (3.9)$$

the objective function of (3.8) satisfies

$$\left| \boldsymbol{u}^{\mathrm{T}}A\boldsymbol{v} \right| = \left| \sum_{j=1}^{n} \eta_j \xi_j \sigma_j \right| \leq \sigma_1 \sum_{j=1}^{n} \left| \eta_j \xi_j \right| \leq \sigma_1,$$

where the last inequality comes from the Cauchy-Schwarz inequality. Moreover, since $\boldsymbol{u}_1^{\mathrm{T}}A\boldsymbol{v}_1 = \sigma_1$, this pair of vectors solves (3.8), while σ_1 satisfies

$$\begin{aligned}\sigma_1 = \\ \max \left\{ \left| \boldsymbol{u}^{\mathrm{T}}A\boldsymbol{v} \right| \, \middle| \, \boldsymbol{u} \in \mathbb{R}^m, \left\| \boldsymbol{u} \right\|_2 = 1, \boldsymbol{v} \in \mathbb{R}^n, \left\| \boldsymbol{v} \right\|_2 = 1 \right\}. \end{aligned} \qquad (3.10)$$

The last result is analogous to (1.4). Yet, in contrast to (1.5), here the orthogonality relations (3.9) imply that

$$0 = \min \left\{ \left| \boldsymbol{u}^{\mathrm{T}}A\boldsymbol{v} \right| \, \middle| \, \boldsymbol{u} \in \mathbb{R}^m, \left\| \boldsymbol{u} \right\|_2 = 1, \boldsymbol{v} \in \mathbb{R}^n, \left\| \boldsymbol{v} \right\|_2 = 1 \right\}. \quad (3.11)$$

Further min-max properties of scalar rectangular quotients are obtained from the Courant-Fischer theorem, see the next section.

Another justification behind the proposed definition of the Rectangular Quotient comes from the observation that the Rayleigh quotient corresponding to the matrix $G = \begin{bmatrix} O & A \\ A^{\mathrm{T}} & O \end{bmatrix}$ and the vector $\boldsymbol{x} = \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix}$ is $\rho = \boldsymbol{u}^{\mathrm{T}}A\boldsymbol{v}$. Hence in this case the bound (1.2) implies the existence of a singular value of A, $\hat{\sigma}$, that satisfies

$$\left| \hat{\sigma} - \left| \boldsymbol{u}^{\mathrm{T}}A\boldsymbol{v} \right| \right|$$
$$\leq \left(\left\| A\boldsymbol{v} - \left| \boldsymbol{u}^{\mathrm{T}}A\boldsymbol{v} \right| \boldsymbol{u} \right\|_2 + \left\| A^{\mathrm{T}}\boldsymbol{u} - \left| \boldsymbol{u}^{\mathrm{T}}A\boldsymbol{v} \right| \boldsymbol{v} \right\|_2 \right) \Big/ \sqrt{2}. \qquad (3.12)$$

The last bound can be refined by applying the following retrieval rules, derived in [3]. Let $\boldsymbol{v} \in \mathbb{R}^n$ be a given unit vector that satisfies $A\boldsymbol{v} \neq \boldsymbol{0}$, and let

$$\boldsymbol{u} = A\boldsymbol{v} \big/ \left\| A\boldsymbol{v} \right\|_2$$

and

$$\sigma = \left\| A\boldsymbol{v} \right\|_2 = \boldsymbol{u}^{\mathrm{T}}A\boldsymbol{v}$$

provide the corresponding estimates of a left singular vector, and a singular value, respectively. Then

$$A\boldsymbol{v} - \sigma\boldsymbol{u} = A\boldsymbol{v} - \left\| A\boldsymbol{v} \right\|_2 \left(A\boldsymbol{v} \big/ \left\| A\boldsymbol{v} \right\|_2 \right) = \boldsymbol{0}$$

and (3.12) is reduced to

$$\left| \hat{\sigma} - \boldsymbol{u}^{\mathrm{T}}A\boldsymbol{v} \right| \leq \left\| A^{\mathrm{T}}\boldsymbol{u} - \left(\boldsymbol{u}^{\mathrm{T}}A\boldsymbol{v} \right)\boldsymbol{v} \right\|_2 \Big/ \sqrt{2}. \qquad (3.13)$$

Similarly let $\boldsymbol{u} \in \mathbb{R}^m$ be a given unit vector that satisfies $A^{\mathrm{T}}\boldsymbol{u} \neq \boldsymbol{0}$, and let

$$\boldsymbol{v} = A^{\mathrm{T}}\boldsymbol{u} \big/ \left\| A^{\mathrm{T}}\boldsymbol{u} \right\|_2$$

and

$$\sigma = \left\| A^{\mathrm{T}}\boldsymbol{u} \right\|_2 = \boldsymbol{u}^{\mathrm{T}}A\boldsymbol{v}$$

denote the corresponding estimates of a singular vector, and a singular value, respectively. Then here

$$A^{\mathrm{T}}\boldsymbol{u} - \left(\boldsymbol{u}^{\mathrm{T}}A\boldsymbol{v} \right)\boldsymbol{v} = \boldsymbol{0}$$

and (3.12) is reduced to

$$\left| \hat{\sigma} - \boldsymbol{u}^{\mathrm{T}}A\boldsymbol{v} \right| \leq \left\| A\boldsymbol{v} - \left(\boldsymbol{u}^{\mathrm{T}}A\boldsymbol{v} \right)\boldsymbol{u} \right\|_2 \Big/ \sqrt{2}. \qquad (3.14)$$

We shall finish this section by noting the difference between the Rectangular Quotient (1.6) and the Generalized Rayleigh Quotient (GRQ) proposed by Ostrowski [26]. Let W be a general (nonnormal) square matrix of order n and let $\boldsymbol{x}$ and $\boldsymbol{y}$ be two n-vectors that satisfy $\boldsymbol{x}^* \boldsymbol{y} \neq 0$ where $\boldsymbol{x}^*$ denotes the conjugate transpose of $\boldsymbol{x}$. Then the GRQ,

$$\omega = \omega(\boldsymbol{x}, W, \boldsymbol{y}) = \boldsymbol{x}^* W \boldsymbol{y} \big/ \left(\boldsymbol{x}^* \boldsymbol{y} \right), \qquad (3.15)$$

is aimed to approximate an eigenvalue of W that is "common" to $\boldsymbol{x}$ and $\boldsymbol{y}$. For detailed discussions of the GRQ and its properties see [25-27,29,40].

4. From Eigenvalues to Singular Values

The connection between the singular values of A and the eigenvalues of the matrices $A^T A$, $A A^T$, and $\begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix}$, is quite straightforward. Indeed, many properties of

singular values are inherited from this connection. Yet, as our review shows, the depth of the relations is far beyond this basic connection. The Courant-Fischer minimax theorem and Weyl's theorem provide highly useful results on eigenvalues of symmetric matrices. See [1,30] for detailed discussions of these theorems and their consequences. Below we consider the adaption of these results when moving from eigenvalues to singular values. The adapted theorems are used to provide a "traditional" proof of the Eckart-Young theorem.

As before A denotes a real $m \times n$ matrix whose SVD is given by (2.1)-(2.7). The first pair of theorems provides useful "minimax" characterizations of singular values. In these theorems $\mathbb{V}_k$ denotes an arbitrary subspace of $\mathbb{R}^n$ that has dimension k. Similarly, $\mathbb{U}_\ell$ denotes an arbitrary subspace of $\mathbb{R}^m$ that has dimension ℓ.

Theorem 1 (Right Courant-Fischer Minimax Theorem) *The jth singular value of A satisfies*

$$\sigma_j = \max_{\mathbb{V}_j} \min \left\{ \|A\mathbf{v}\|_2 \,\middle|\, \mathbf{v} \in \mathbb{V}_j, \|\mathbf{v}\|_2 = 1 \right\} \qquad (4.1)$$

and

$$\sigma_j = \min_{\mathbb{V}_{j^*}} \max \left\{ \|A\mathbf{v}\|_2 \,\middle|\, \mathbf{v} \in \mathbb{V}_{j^*}, \|\mathbf{v}\|_2 = 1 \right\} \qquad (4.2)$$

where the integer j^* is defined by the equality

$$j + j^* = n + 1. \qquad (4.3)$$

(*The maximum in* (4.1) *is over all j dimensional subspaces $\mathbb{V}_j$ of $\mathbb{R}^n$. The minimum in* (4.2) *is over all j^* dimensional subspaces $\mathbb{V}_{j^*}$ of $\mathbb{R}^n$.) Moreover, the maximum in* (4.1) *is attained for $\mathbb{V}_j = \text{span}\{\mathbf{v}_1, \cdots, \mathbf{v}_j\}$, while the minimum in* (4.2) *is attained for*

$$\mathbb{V}_{j^*} = \text{span}\left\{ \mathbf{v}_j, \mathbf{v}_{j+1}, \cdots, \mathbf{v}_n \right\}.$$

Yet the solutions of both problems are not necessarily unique. □

Theorem 2 (Left Courant-Fischer Minimax Theorem) *The ith singular value of A satisfies*

$$\sigma_i = \max_{\mathbb{U}_i} \min \left\{ \|A^{\mathrm{T}}\mathbf{u}\|_2 \,\middle|\, \mathbf{u} \in \mathbb{U}_i, \|\mathbf{u}\|_2 = 1 \right\} \qquad (4.4)$$

and

$$\sigma_i = \min_{\mathbb{U}_{i^*}} \max \left\{ \|A^{\mathrm{T}}\mathbf{u}\|_2 \,\middle|\, \mathbf{u} \in \mathbb{U}_{i^*}, \|\mathbf{u}\|_2 = 1 \right\}, \qquad (4.5)$$

where the integer i^* is defined by the equality

$$i + i^* = m + 1. \qquad (4.6)$$

Moreover, the maximum in (4.4) *is attained for*

$$\mathbb{U}_i = \text{span}\left\{ \mathbf{u}_1, \cdots, \mathbf{u}_i \right\},$$

while the minimum in (4.5) *is attained for*

$$\mathbb{U}_{i^*} = \text{span}\left\{ \mathbf{u}_i, \mathbf{u}_{i+1}, \cdots, \mathbf{u}_m \right\}. \qquad □$$

Note that Theorem 2 is essentially Theorem 1 for A^{T}. The proof of Theorem 1 is based on the following idea. The condition (4.3) ensures the existence of a unit vector, $\mathbf{x}$, that belongs both to $\mathbb{V}_j$ and $\mathbb{V}_{j^*}$. Thus

$$\min \left\{ \|A\mathbf{v}\|_2 \,\middle|\, \mathbf{v} \in \mathbb{V}_j, \|\mathbf{v}\|_2 = 1 \right\} \le \|A\mathbf{x}\|_2$$

$$\le \max \left\{ \|A\mathbf{v}\|_2 \,\middle|\, \mathbf{v} \in \mathbb{V}_{j^*}, \|\mathbf{v}\|_2 = 1 \right\},$$

and

$$\sup_{\mathbb{V}_j} \min \left\{ \|A\mathbf{v}\|_2 \,\middle|\, \mathbf{v} \in \mathbb{V}_j, \|\mathbf{v}\|_2 = 1 \right\}$$

$$\le \inf_{\mathbb{V}_{j^*}} \max \left\{ \|A\mathbf{v}\|_2 \,\middle|\, \mathbf{v} \in \mathbb{V}_{j^*}, \|\mathbf{v}\|_2 = 1 \right\},$$

So the proof is concluded by verifying that equality holds when using the specified subspaces.

Let B denote another $m \times n$ real matrix and let

$$C = A - B \qquad (4.7)$$

denote the corresponding difference matrix. The singular values of B and C are denoted as

$$\beta_1 \ge \beta_2 \ge \cdots \ge \beta_n \ge 0 \quad \text{and} \quad \gamma_1 \ge \gamma_2 \ge \cdots \ge \gamma_n \ge 0, \quad (4.8)$$

respectively. The coming corollaries of Theorem 1 answer the question of how the rank of B affects the singular values of C.

Lemma 3 *Assume that $\text{rank}(B) = k$. In this case*

$$\gamma_1 \ge \sigma_{k+1}. \qquad (4.9)$$

Proof. Take $\mathbb{V}_{j^*} = \text{Null}(B)$. Then $j^* = n - k$ and $j = k + 1$. Consequently

$$\gamma_1 = \max \left\{ \|(A - B)\mathbf{v}\|_2 \,\middle|\, \mathbf{v} \in \mathbb{R}^n, \|\mathbf{v}\|_2 = 1 \right\}$$

$$\ge \max \left\{ \|(A - B)\mathbf{v}\|_2 \,\middle|\, \mathbf{v} \in \text{Null}(B), \|\mathbf{v}\|_2 = 1 \right\}$$

$$= \max \left\{ \|A\mathbf{v}\|_2 \,\middle|\, \mathbf{v} \in \text{Null}(B), \|\mathbf{v}\|_2 = 1 \right\} \ge \sigma_j = \sigma_{k+1},$$

where the last inequality follows from (4.2). □

Theorem 4 (Weyl) *Let B and C be as in* (4.7) - (4.8). *Then*

$$\beta_i + \gamma_j \ge \sigma_{i+j-1}, \qquad (4.10)$$

under the convention that $\sigma_\ell = 0$ when $\ell > n$.

Proof. Let B_{i-1} be a rank $i-1$ Truncated SVD of B, and let C_{j-1} to be a rank $j-1$ Truncated SVD of C. Then the largest singular value of $B - B_{i-1}$ is β_i, while the largest singular value of $C - C_{j-1}$ is γ_j. That is,

$$\beta_i = \max \left\{ \|(B - B_{i-1})\mathbf{v}\|_2 \,\middle|\, \mathbf{v} \in \mathbb{R}^n, \|\mathbf{v}\|_2 = 1 \right\} \qquad (4.11)$$

and

$$\gamma_j = \max \left\{ \|(C - C_{j-1})\mathbf{v}\|_2 \,\middle|\, \mathbf{v} \in \mathbb{R}^n, \|\mathbf{v}\|_2 = 1 \right\} \qquad (4.12)$$

Let the $m \times n$ matrix F be defined by the equality

$$F = B_{i-1} + C_{j-1}.$$

Then $\operatorname{rank}(F) \leq i + j - 2$, and

$$A - F = (B - B_{i-1}) + (C - C_{j-1}).$$

Hence from (4.11) and (4.12) we see that

$$\beta_i + \gamma_j \geq \max\left\{ \left\| (A - F)\boldsymbol{v} \right\|_2 \,\middle|\, \boldsymbol{v} \in \mathbb{R}^n, \|\boldsymbol{v}\|_2 = 1 \right\} \geq \sigma_{i+j-1},$$

where the last inequality follows from Lemma 3. $\quad\square$

Corollary 5 (Majorization) *Assume further that* $\operatorname{rank}(B) \leq k$, *which means* $\beta_i = 0$ *for* $i = k+1, \cdots, n$. *Then substituting* $i = k+1$ *in* (4.10) *gives*

$$\gamma_j \geq \sigma_{j+k} \quad \text{for } j = 1, \cdots, n-k. \tag{4.13}$$

In other words, if $\operatorname{rank}(B) \leq k$ *then the singular values of* C *majorize the singular values of* $A - T_k$. $\quad\square$

Recall that T_k denotes a rank-k Truncated SVD of A, as defined in (2.8). Roughly speaking the last corollary says that a rank-k perturbation of A may cause the singular values to "fall" not more than k "levels". The next corollary shows that they are unable to "rise" more than k levels.

Corollary 6 *Observe that* (4.7) *can be rewritten as* $A = C - (-B)$ *while* $-B$ *has the same singular values as* B. *Hence a further consequence of Theorem 4 is*

$$\beta_i + \sigma_j \geq \gamma_{i+j-1}. \tag{4.14}$$

Moreover, if $\operatorname{rank}(B) \leq k$ *then*

$$\sigma_j \geq \gamma_{j+k} \quad \text{for } j = 1, \cdots, n-k. \tag{4.15}$$

$\quad\square$

The next results provide useful bounds on the perturbed singular values.

Corollary 7 (Bounding and Interlacing) *Using* (4.10) *and* (4.14) *with* $i = 1$ *gives*

$$\sigma_j + \beta_1 \geq \gamma_j \geq \sigma_j - \beta_1, \quad j = 1, \cdots, n, \tag{4.16}$$

and

$$\left| \gamma_j - \sigma_j \right| \leq \beta_1, \quad j = 1, \cdots, n. \tag{4.17}$$

Furthermore, consider the special case when B *is a rank-one matrix. Then using* (4.13) *and* (4.15) *with* $k = 1$ *gives*

$$\sigma_{j-1} \geq \gamma_j \geq \sigma_{j+1}, \quad j = 1, \cdots, n, \tag{4.18}$$

where $\sigma_0 = +\infty$ and $\sigma_{n+1} = 0$. $\quad\square$

Theorem 8 (Eckart-Young) *Let* B *and* C *be as in* (4.7)-(4.8) *and assume that* $\operatorname{rank}(B) \leq k$. *Then*

$$\|A - B\|_F^2 \geq \sum_{i=k+1}^n \sigma_i^2. \tag{4.19}$$

Moreover, let T_k *be a rank-k Truncated SVD of* A, *as defined in* (2.8). *Then* T_k *solves the minimum norm problem*

$$\text{minimize} \quad F(B) = \|A - B\|_F^2 \tag{4.20}$$
$$\text{subject to} \quad \operatorname{rank}(B) \leq k,$$

giving the optimal value of

$$\|A - T_k\|_F^2 = \left\| \sum_{j=k+1}^n \sigma_j \boldsymbol{u}_j \boldsymbol{v}_j^{\mathrm{T}} \right\|_F^2 = \sum_{j=k+1}^n \sigma_j^2.$$

Proof. Using (4.13) we see that

$$\|A - B\|_F^2 = \|C\|_F^2 = \sum_{j=1}^n \gamma_j^2 \geq \sum_{j=1}^{n-k} \gamma_j^2 \geq \sum_{j=1}^{n-k} \sigma_{j+k}^2 = \sum_{i=k+1}^n \sigma_i^2.$$

$\quad\square$

The last theorem says that T_k is a best rank-k approximation of A, regarding the Frobenius norm. Observe that Lemma 3 proves a similar claim for the 2-norm. The next extension is due to Mirsky [23].

Theorem 9 (Mirsky) *Let* $\|\cdot\|$ *denote a unitarily invariant norm on* $\mathbb{R}^{m \times n}$. *Then the inequality*

$$\|A - B\| \geq \|A - T_k\|, \tag{4.21}$$

holds for any matrix $B \in \mathbb{R}^{m \times n}$ *such that* $\operatorname{rank}(B) \leq k$. *In other words,* T_k *solves the minimum norm problem*

$$\text{minimize} \quad \nu(B) = \|A - B\| \tag{4.22}$$
$$\text{subject to} \quad \operatorname{rank}(B) \leq k.$$

Proof. From Corollary 5 we see that the singular values of $A - B$ majorize those of $A - T_k$. Hence (4.21) is a direct consequence of Ky Fan Dominance Theorem. $\quad\square$

Another related problem is

$$\text{minimize} \quad \|C\|_F^2 \tag{4.23}$$
$$\text{subject to} \quad (A - C) \in \mathbb{R}_k^{m \times n},$$

where $\mathbb{R}_k^{m \times n}$ denotes the set of all real $m \times n$ matrices of rank k. Below we will show that the residual matrix

$$R_k = A - T_k = \sum_{j=k+1}^n \sigma_j \boldsymbol{u}_j \boldsymbol{v}_j^{\mathrm{T}} \tag{4.24}$$

solves this problem. In other words, R_k is the smallest perturbation that turns A into a rank-k matrix.

Theorem 10 *Let* B *and* C *be as in* (4.7) *and assume that*

$$k = \operatorname{rank}(B) < \operatorname{rank}(A) = r.$$

Then

$$\|C\|_F^2 \geq \|R_k\|_F^2 = \sum_{j=k+1}^r \sigma_j^2. \tag{4.25}$$

Proof. Using the Eckart-Young theorem we obtain

$$\|C\|_F^2 = \|A - B\|_F^2 \geq \|A - T_k\|_F^2 = \|R_k\|_F^2 = \sum_{j=k+1}^r \sigma_j^2.$$

$\quad\square$

Observe that R_k remains the solution of (4.23) when this problem is defined by any other unitarily invariant

norm. Note also that Total Least Squares problems give rise to a special form of (4.23) in which $k = r - 1$. In this case the solution matrix, $\tilde{C}_k$, is reduced to the rank-one matrix $\sigma_r \boldsymbol{u}_r \boldsymbol{v}_r^{\mathrm{T}}$, e.g., [14,15]. Further consequences of the Eckart-Young theorem are presented in Section 6.

5. Least Squares Properties of Orthogonal Quotients Matrices

The optimality properties of symmetric Rayleigh Quotient matrices form the basis of the celebrated Rayleigh-Ritz procedure, e.g., [30,34]. In this section we derive the corresponding properties of Orthogonal Quotients matrices. As we shall see, Orthogonal Quotient matrices extend symmetric Rayleigh Quotient matrices in the same way that Rectangular Quotients extend Rayleigh Quotients.

Theorem 11 *Let*

$$X_{\tilde{m}} = [\boldsymbol{x}_1, \cdots, \boldsymbol{x}_{\tilde{m}}] \in \mathbb{X}_{\tilde{m}} \quad and \quad Y_{\tilde{n}} = [\boldsymbol{y}_1, \cdots, \boldsymbol{y}_{\tilde{n}}] \in \mathbb{Y}_{\tilde{n}}$$

be a given pair of matrices with orthonormal columns, and let

$$Z = X_{\tilde{m}}^{\mathrm{T}} A Y_{\tilde{n}} \tag{5.1}$$

denote the corresponding orthogonal quotient matrix. Then Z solves the following three problems.

$$\begin{align} \text{minimize} \quad & \alpha(R) = \left\| A - X_{\tilde{m}} R Y_{\tilde{n}}^{\mathrm{T}} \right\|_F \\ \text{subject to} \quad & R \in \mathbb{R}^{\tilde{m} \times \tilde{n}}, \end{align} \tag{5.2}$$

$$\begin{align} \text{minimize} \quad & \beta(R) = \left\| A Y_{\tilde{n}} - X_{\tilde{m}} R \right\|_F \\ \text{subject to} \quad & R \in \mathbb{R}^{\tilde{m} \times \tilde{n}}, \end{align} \tag{5.3}$$

and

$$\begin{align} \text{minimize} \quad & \gamma(R) = \left\| A^{\mathrm{T}} X_{\tilde{m}} - Y_{\tilde{n}} R^{\mathrm{T}} \right\|_F \\ \text{subject to} \quad & R \in \mathbb{R}^{\tilde{m} \times \tilde{n}}. \end{align} \tag{5.4}$$

Proof. Completing the columns of $X_{\tilde{m}}$ to be an orthonormal basis of $\mathbb{R}^m$ gives an orthogonal $m \times m$ matrix, X_m, whose first $\tilde{m}$ columns are the columns of $X_{\tilde{m}}$. Similarly there exists an orthogonal $n \times n$ matrix, Y_n, whose first $\tilde{n}$ columns are the columns of $Y_{\tilde{n}}$. Therefore, since the Frobenius norm is unitarily invariant,

$$\begin{align} \alpha(R) &= \left\| X_m^{\mathrm{T}} \left(A - X_{\tilde{m}} R Y_{\tilde{n}}^{\mathrm{T}} \right) Y_n \right\|_F \\ &= \left\| X_m^{\mathrm{T}} A Y_n - R_\alpha \right\|_F, \end{align} \tag{5.5}$$

where

$$R_\alpha = X_m^{\mathrm{T}} X_{\tilde{m}} R Y_{\tilde{n}}^{\mathrm{T}} Y_n = \begin{bmatrix} R & 0 \\ 0 & 0 \end{bmatrix}. \tag{5.6}$$

The validity of the last equality is easily verified by

noting that the $m \times \tilde{m}$ matrix $X_m^{\mathrm{T}} X_{\tilde{m}}$ is composed of the first $\tilde{m}$ columns of the $m \times m$ identity matrix, while the $\tilde{n} \times n$ matrix $Y_{\tilde{n}}^{\mathrm{T}} Y_n$ is composed of the first $\tilde{n}$ rows of the $n \times n$ identity matrix. Note also that the corresponding principle submatrix of $X_m^{\mathrm{T}} A Y_n$ is $X_{\tilde{m}}^{\mathrm{T}} A Y_{\tilde{n}}$. Hence the choice $R = X_{\tilde{m}}^{\mathrm{T}} A Y_{\tilde{n}}$ minimizes $\alpha(R)$.

The other two problems are solved by similar arguments, using the equalities

$$\beta(R) = \left\| X_m^{\mathrm{T}} \left(A Y_{\tilde{n}} - X_{\tilde{m}} R \right) \right\|_F = \left\| X_m^{\mathrm{T}} A Y_{\tilde{n}} - R_\beta \right\|_F, \tag{5.7}$$

and

$$\gamma(R) = \left\| Y_n^{\mathrm{T}} \left(A^{\mathrm{T}} X_{\tilde{m}} - Y_{\tilde{n}} R^{\mathrm{T}} \right) \right\|_F = \left\| X_{\tilde{m}}^{\mathrm{T}} A Y_n - R_\gamma \right\|_F, \tag{5.8}$$

where

$$R_\beta = \begin{bmatrix} R \\ 0 \end{bmatrix} \in \mathbb{R}^{m \times \tilde{n}} \quad \text{and} \quad R_\gamma = [R, 0] \in \mathbb{R}^{\tilde{m} \times n}. \tag{5.9}$$

$\square$

Remark A further inspection of relations (5.7)-(5.9) indicates that Z solves problems (5.3) and (5.4) even if the Frobenius norm is replaced by any other unitarily invariant matrix norm. However the last claim is not valid for problem (5.2), since "punching" a matrix may increase its norm. To see this point consider the matrices

$A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$, whose trace norms are 2

and $\sqrt{5}$, respectively. That is, punching A increases its trace norm. Nevertheless, punching a matrix always reduces its Frobenius norm. Hence the proof of Theorem 11 leads to the following powerful results.

Corollary 12 *Let $\mathbb{P}^{\tilde{m} \times \tilde{n}}$ denote the set of all real $\tilde{m} \times \tilde{n}$ matrices that have a certain pattern of zeros. (For example, the set of all tridiagonal matrices.) Let the matrix $\tilde{Z} \in \mathbb{P}^{\tilde{m} \times \tilde{n}}$ be obtained from $X_{\tilde{m}}^{\mathrm{T}} A Y_{\tilde{n}}$ by setting zeros in the corresponding places. Then Theorem 11 remains valid when $\mathbb{R}^{\tilde{m} \times \tilde{n}}$ and $X_{\tilde{m}}^{\mathrm{T}} A Y_{\tilde{n}}$ are replaced by $\mathbb{P}^{\tilde{m} \times \tilde{n}}$ and $\tilde{Z}$, respectively.* $\square$

Corollary 13 *Assume that $\tilde{m} = \tilde{n} = k$ and let*

$$\mathbb{D}^{k \times k} = \left\{ D \mid D = \mathrm{diag}\{d_1, \cdots, d_k\} \in \mathbb{R}^{k \times k} \right\}$$

denote the set of all real diagonal $k \times k$ matrices. Let $X_k = [\boldsymbol{x}_1, \cdots, \boldsymbol{x}_k] \in \mathbb{X}_k$ and $Y_k = [\boldsymbol{y}_1, \cdots, \boldsymbol{y}_k] \in \mathbb{Y}_k$ be a given pair of matrices with orthonormal columns. Then the matrix

$$D_k = \mathrm{diag}\left\{ \boldsymbol{x}_1^{\mathrm{T}} A \boldsymbol{y}_1, \boldsymbol{x}_2^{\mathrm{T}} A \boldsymbol{y}_2, \cdots, \boldsymbol{x}_k^{\mathrm{T}} A \boldsymbol{y}_k \right\} \tag{5.10}$$

solves the following three problems.

$$\begin{align} \text{minimize} \quad & \alpha(D) = \left\| A - X_k D Y_k^{\mathrm{T}} \right\|_F \\ \text{subject to} \quad & D \in \mathbb{D}^{k \times k}, \end{align} \tag{5.11}$$

$$\begin{align} \text{minimize} \quad & \beta(D) = \left\| A Y_k - X_k D \right\|_F \\ \text{subject to} \quad & D \in \mathbb{D}^{k \times k}, \end{align} \tag{5.12}$$

and

$$\text{minimize} \quad \gamma(D) = \left\| A^{\mathrm{T}} X_k - Y_k D \right\|_F$$
$$\text{subject to} \quad D \in \mathbb{D}^{k \times k}. \tag{5.13}$$

$\square$

As in the scalar case, the residual functions, $\beta(R)$ and $\gamma(R)$, enable us to bound the "distance" between the singular values of R and A. Assume for a moment that $\tilde{m} = \tilde{n} = k$ and let $\rho_1, \cdots, \rho_k$ denote the singular values of R. Then there exists a permutation τ of $\{1, \cdots, n\}$ such that

$$\sum_{j=1}^{k} \left(\rho_j - \sigma_{\tau_j} \right)^2 \le \left(\beta^2(R) + \gamma^2(R) \right) \big/ 2, \tag{5.14}$$

see [2,21]. The relations (5.7)-(5.9) indicate that the minimal values of $\beta(\gamma)$ and $\gamma(R)$ equal the Frobenius norm of the corresponding off-diagonal blocks in the matrix $X_m^{\mathrm{T}} A Y_n$. The next observations regard the minimal value of $\alpha(R)$.

6. The Orthogonal Quotients Equality and Eckart-Young Theorem

In this section we derive the Orthogonal Quotients Equality and discuss its relation to the Eckart-Young theorem.

Theorem 14 (The Orthogonal Quotients Equality) *Let $X_{\tilde{m}} \in \mathbb{X}_{\tilde{m}}$ and $Y_{\tilde{n}} \in \mathbb{Y}_{\tilde{n}}$ be a given pair of matrices with orthonormal columns. Then*

$$\left\| A - X_{\tilde{m}} \left(X_{\tilde{m}}^{\mathrm{T}} A Y_{\tilde{n}} \right) Y_{\tilde{n}}^{\mathrm{T}} \right\|_F^2 = \|A\|_F^2 - \left\| X_{\tilde{m}}^{\mathrm{T}} A Y_{\tilde{n}} \right\|_F^2. \tag{6.1}$$

Proof. Following the proof of Theorem 11 we see that

$$\left\| A - X_{\tilde{m}} \left(X_{\tilde{m}}^{\mathrm{T}} A Y_{\tilde{n}} \right) Y_{\tilde{n}}^{\mathrm{T}} \right\|_F^2 = \left\| X_m^{\mathrm{T}} A Y_n - \tilde{R} \right\|_F^2, \tag{6.2}$$

where

$$\tilde{R} = X_m^{\mathrm{T}} X_{\tilde{m}} \left(X_{\tilde{m}}^{\mathrm{T}} A Y_{\tilde{n}} \right) Y_{\tilde{n}}^{\mathrm{T}} Y_n = \begin{pmatrix} Z & 0 \\ 0 & 0 \end{pmatrix}.$$

Therefore, since $Z = X_{\tilde{m}}^{\mathrm{T}} A Y_{\tilde{n}}$ is a principal submatrix of $X_m^{\mathrm{T}} A Y_n$,

$$\left\| X_m^{\mathrm{T}} A Y_n - \tilde{R} \right\|_F^2 = \left\| X_m^{\mathrm{T}} A Y_n \right\|_F^2 - \|Z\|_F^2$$
$$= \|A\|_F^2 - \left\| X_{\tilde{m}}^{\mathrm{T}} A Y_{\tilde{n}} \right\|_F^2.$$

Corollary 15 *Let $T = (t_{ij})$ be any $\tilde{m} \times \tilde{n}$ matrix whose entries satisfy the following rule: Either $t_{ij} = \boldsymbol{x}_i^{\mathrm{T}} A \boldsymbol{y}_j$ or $t_{ij} = 0$. In other words, T is obtained from $X_{\tilde{m}}^{\mathrm{T}} A Y_{\tilde{n}}$ by setting some entries to zero. Then*

$$\left\| A - X_{\tilde{m}} T Y_{\tilde{n}}^{\mathrm{T}} \right\|_F^2 = \|A\|_F^2 - \|T\|_F^2. \tag{6.3}$$

Corollary 16 (The Orthogonal Quotients Equality in Diagonal Form) *Let $X_k = [\boldsymbol{x}_1, \cdots, \boldsymbol{x}_k] \in \mathbb{X}_k$ and*

$Y_k = [\boldsymbol{y}_1, \cdots, \boldsymbol{y}_k] \in \mathbb{Y}_k$ *be a given pair of matrices with orthonormal columns. Then the diagonal matrix* (5.10) *satisfies*

$$\left\| A - X_k D_k Y_k^{\mathrm{T}} \right\|_F^2 = \|A\|_F^2 - \|D_k\|_F^2. \tag{6.4}$$

In vector notations the last equality takes the form

$$\left\| A - \sum_{j=1}^{k} \left(\boldsymbol{x}_j^{\mathrm{T}} A \boldsymbol{y}_j \right) \boldsymbol{x}_j \boldsymbol{y}_j^{\mathrm{T}} \right\|_F^2 = \|A\|_F^2 - \sum_{j=1}^{k} \left(\boldsymbol{x}_j^{\mathrm{T}} A \boldsymbol{y}_j \right)^2. \tag{6.5}$$

$\square$

Let us return now to consider the Eckart-Young problem (4.20). One way to express an $m \times n$ matrix, whose rank is at most k, is

$$B = X_k R Y_k^{\mathrm{T}} \tag{6.6}$$

where $X_k = [\boldsymbol{x}_1, \cdots, \boldsymbol{x}_k] \in \mathbb{X}_k$, $Y_k = [\boldsymbol{y}_1, \cdots, \boldsymbol{y}_k] \in \mathbb{Y}_k$ and $R \in \mathbb{R}^{k \times k}$. Alternatively we can write B in the form

$$B = X_k D Y_k^{\mathrm{T}} = \sum_{j=1}^{k} d_j \boldsymbol{x}_j \boldsymbol{y}_j^{\mathrm{T}}, \tag{6.7}$$

where $D = \text{diag}\{d_1, \cdots, d_k\}$ is a real diagonal $k \times k$ matrix. (The first form results from complete orthogonal decomposition of B, while the second form is obtained from the SVD.) Substituting (6.6) in the objective function of (4.20) results in the function $\alpha^2(R)$. Hence, by Theorem 11, there is no loss of generality in replacing R with $X_k^{\mathrm{T}} A Y_k$. Similarly D can be replaced with D_k, the solution of (5.11). These observations lead to the following conclusions.

Theorem 17 (Equivalent Formulations of the Eckart-Young Problem) *There is no loss of generality in writing the Eckart-Young problem* (4.20) *in the forms*

$$\text{minimize} \quad F(X_k, Y_k) = \left\| A - X_k \left(X_k^{\mathrm{T}} A Y_k \right) Y_k^{\mathrm{T}} \right\|_F^2$$
$$\text{subject to} \quad X_k \in \mathbb{X}_k \text{ and } Y_k \in \mathbb{Y}_k, \tag{6.8}$$

or

$$\text{minimize} \quad F(X_k, Y_k) = \left\| A - X_k D_k Y_k^{\mathrm{T}} \right\|_F^2$$
$$\text{subject to} \quad X_k \in \mathbb{X}_k \text{ and } Y_k \in \mathbb{Y}_k. \tag{6.9}$$

Moreover, both problems are solved by the SVD matrices U_k and V_k. (See (2.1)-(2.7) *for the definition of these matrices.)* $\square$

The objective functions of the last problems form the left sides of the Orthogonal Quotients Equalities

$$\left\| A - X_k \left(X_k^{\mathrm{T}} A Y_k \right) Y_k^{\mathrm{T}} \right\|_F^2 = \|A\|_F^2 - \left\| X_k^{\mathrm{T}} A Y_k \right\|_F^2 \tag{6.10}$$

and

$$\left\| A - X_k D_k Y_k^{\mathrm{T}} \right\|_F^2 = \|A\|_F^2 - \|D_k\|_F^2. \tag{6.11}$$

These relations turn the minimum norm problems (6.8)-(6.9) into equivalent maximum norm problems.

Theorem 18 (Maximum Norm Formulations of the Eckart-Young Problem) *The Eckart-Young problems* (6.8) *and* (6.9) *are equivalent to the problems*

$$\text{maximize} \quad \left\|X_k^{\mathrm{T}} A Y_k\right\|_F^2 \tag{6.12}$$
$$\text{subject to} \quad X_k \in \mathbb{X}_k \ \text{ and } \ Y_k \in \mathbb{Y}_k,$$

and

$$\text{maximize} \quad \sum_{j=1}^{k}\left(\boldsymbol{x}_j^{\mathrm{T}} A \boldsymbol{y}_j\right)^2 \tag{6.13}$$
$$\text{subject to} \quad X_k \in \mathbb{X}_k \ \text{ and } \ Y_k \in \mathbb{Y}_k,$$

respectively. The SVD matrices U_k *and* V_k *solve both problems.* $\quad\square$

Let $\tilde{\sigma}_j = \tilde{\sigma}_j\left(X_k^{\mathrm{T}} A Y_k\right)$, $j = 1, \cdots, k$, denote the singular values of the orthogonal quotients matrix $X_k^{\mathrm{T}} A Y_k$. Then, clearly,

$$\left\|X_k^{\mathrm{T}} A Y_k\right\|_F^2 = \tilde{\sigma}_1^2 + \cdots + \tilde{\sigma}_k^2, \tag{6.14}$$

and the Eckart-Young problem (6.12) can be rewritten as

$$\text{maximize} \quad \tilde{\sigma}_1^2 + \cdots + \tilde{\sigma}_k^2 \tag{6.15}$$
$$\text{subject to} \quad X_k \in \mathbb{X}_k \ \text{ and } \ Y_k \in \mathbb{Y}_k.$$

In the next sections we consider extended problems of this type. The key for solving the extended problems lies in the properties of symmetric orthogonal quotients matrices.

7. The Symmetric Quotients Equality and Ky Fan's Extremum Principles

Let $G = \left(g_{ij}\right)$ be a real $n \times n$ symmetric matrix with spectral decomposition

$$G = Q \Lambda Q^{\mathrm{T}}, \tag{7.1}$$

where $\Lambda = \text{diag}\left\{\lambda_1, \cdots, \lambda_n\right\}$ is a diagonal $n \times n$ matrix, and $Q = \left[\boldsymbol{q}_1, \cdots, \boldsymbol{q}_n\right]$ is an orthogonal $n \times n$ matrix, $Q^{\mathrm{T}} Q = Q Q^{\mathrm{T}} = I$. It is assumed that the eigenvalues of G are sorted to satisfy

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n. \tag{7.2}$$

Let $Y_k = \left[\boldsymbol{y}_1, \cdots, \boldsymbol{y}_k\right] \in \mathbb{Y}_k$ be a given $n \times k$ matrix with orthonormal columns and let

$$D_k = \text{diag}\left\{\boldsymbol{y}_1^{\mathrm{T}} G \boldsymbol{y}_1, \cdots, \boldsymbol{y}_k^{\mathrm{T}} G \boldsymbol{y}_k\right\}$$

denote the $k \times k$ diagonal matrix which forms the diagonal of $Y_k^{\mathrm{T}} G Y_k$. Recall that

$$\text{trace}\left(G\right) = \sum_{j=1}^{n} g_{jj} = \sum_{j=1}^{n} \lambda_j,$$

is invariant under (orthogonal) similarity transformations. Hence by following the proof of (6.1) we obtain the following results.

Theorem 19 (The Symmetric Quotients Equality) *Using the above notations,*

$$\text{trace}\left(G - Y_k\left(Y_k^{\mathrm{T}} G Y_k\right) Y_k^{\mathrm{T}}\right)$$
$$= \text{trace}\left(G\right) - \text{trace}\left(Y_k^{\mathrm{T}} G Y_k\right) \tag{7.3}$$

and

$$\text{trace}\left(G - Y_k D_k Y_k^{\mathrm{T}}\right)$$
$$= \text{trace}\left(G\right) - \text{trace}\left(D_k\right), \tag{7.4}$$

where

$$\text{trace}\left(Y_k^{\mathrm{T}} G Y_k\right) = \text{trace}\left(D_k\right) = \sum_{j=1}^{k} \boldsymbol{y}_j^{\mathrm{T}} G \boldsymbol{y}_j.$$

$\quad\square$

Note that Theorem 19 remains valid when G is replaced by any real $n \times n$ matrix. The role of symmetry becomes prominent in problems that attempt to maximize or minimize $\text{trace}\left(Y_k^{\mathrm{T}} G Y_k\right)$, as considered by Ky Fan [10]. In our notations Ky Fan's problems have the form

$$\text{maximize} \quad \text{trace}\left(Y_k^{\mathrm{T}} G Y_k\right)$$
$$\text{subject to} \quad Y_k \in \mathbb{Y}_k \tag{7.5}$$

and

$$\text{minimize} \quad \text{trace}\left(Y_k^{\mathrm{T}} G Y_k\right)$$
$$\text{subject to} \quad Y_k \in \mathbb{Y}_k. \tag{7.6}$$

The solution of these problems lies in the following well-known properties of symmetric matrices, e.g., [16,30,40].

Theorem 20 (Cauchy Interlace Theorem) *Let the* $k \times k$ *matrix* $\tilde{G}$ *be obtained from* G *by deleting* $\ell = n - k$ *rows and the corresponding* ℓ *columns. Let*

$$\tilde{\lambda}_1 \geq \tilde{\lambda}_2 \geq \cdots \geq \tilde{\lambda}_k$$

denote the eigenvalues of $\tilde{G}$. *Then*

$$\lambda_j \geq \tilde{\lambda}_j \quad \text{for } j = 1, \cdots, k,$$
$$\text{and} \ \ \tilde{\lambda}_{k+1-i} \geq \lambda_{n+1-i} \quad \text{for } i = 1, \cdots, k. \tag{7.7}$$

In particular, when $k = n - 1$,

$$\lambda_1 \geq \tilde{\lambda}_1 \geq \lambda_2 \geq \tilde{\lambda}_2 \geq \lambda_3 \geq \cdots \geq \lambda_{n-1} \geq \tilde{\lambda}_{n-1} \geq \lambda_n. \tag{7.8}$$

$\quad\square$

Corollary 21 (Poincaré Separation Theorem) *Let* $Y_k \in \mathbb{Y}_k$ *be a given* $n \times k$ *matrix with orthonormal columns, and let* $Y_n \in \mathbb{Y}_n$ *be an orthogonal* $n \times n$ *matrix, whose first* k *columns are the columns of* Y_k. *Then* $\tilde{G} = Y_k^{\mathrm{T}} G Y_k$ *is obtained from* $Y_n^{\mathrm{T}} G Y_n$ *by deleting the last* $\ell = n - k$ *rows and the last* ℓ *columns. Therefore, since* $Y_n^{\mathrm{T}} G Y_n$ *has the same eigenvalues as* G, *the eigenvalues of* $\tilde{G}$ *satisfy* (7.7) *and* (7.8).

Corollary 22 (Ky Fan's Extremum Principles) *Consider the spectral decomposition* (7.1)-(7.2) *and let the matrix* $Q_k = \left[\boldsymbol{q}_1, \cdots, \boldsymbol{q}_k\right] \in \mathbb{Y}_k$ *be constructed from the*

first k columns of Q. *Then* Q_k *solves* (7.5), *giving the optimal value of* $\sum_{j=1}^{k} \lambda_j$. *That is,*

$$\sum_{j=1}^{k} \lambda_j = \max \left\{ \text{trace}\left(Y_k^{\mathrm{T}} G Y_k\right) \Big| Y_k \in \mathbb{Y}_k \right\}. \qquad (7.9)$$

The minimum trace problem (7.6) *is solved by the matrix*

$$\hat{Q}_k = \left[\boldsymbol{q}_{n+1-k}, \cdots, \boldsymbol{q}_n\right] \in \mathbb{Y}_k,$$

which is composed of the last k columns of Q. The optimal value of (7.6) *is, therefore,* $\sum_{i=1}^{k} \lambda_{n+1-i}$. *That is,*

$$\sum_{i=1}^{k} \lambda_{n+1-i} = \min \left\{ \text{trace}\left(Y_k^{\mathrm{T}} G Y_k\right) \Big| Y_k \in \mathbb{Y}_k \right\}. \qquad (7.10)$$

$\square$

The symmetric quotients equality (7.3) means that Ky Fan's problems, (7.5) and (7.6), are equivalent to the problems

$$\begin{aligned} \text{minimize} \quad & \text{trace}\left(G - Y_k \left(Y_k^{\mathrm{T}} G Y_k\right) Y_k^{\mathrm{T}}\right) \\ \text{subject to} \quad & Y_k \in \mathbb{Y}_k, \end{aligned} \qquad (7.11)$$

and

$$\begin{aligned} \text{maximize} \quad & \text{trace}\left(G - Y_k \left(Y_k^{\mathrm{T}} G Y_k\right) Y_k^{\mathrm{T}}\right) \\ \text{subject to} \quad & Y_k \in \mathbb{Y}_k, \end{aligned} \qquad (7.12)$$

respectively. Note the remarkable similarity between Eckart-Young problems (6.8) and (6.12), and Ky Fan problems (7.11) and (7.5), respectively.

A further insight is gained by considering the case when G is positive semidefinite. In this case the spectral decomposition (7.1)-(7.2) coincides with the SVD of G and the $k \times k$ matrix $Y_k^{\mathrm{T}} G Y_k$ is also positive semidefinite. Let

$$\tilde{\lambda}_1 \geq \tilde{\lambda}_2 \geq \cdots \geq \tilde{\lambda}_k \geq 0, \qquad (7.13)$$

denote the eigenvalues (the singular values) of this matrix. Then here the interlacing relations (7.7) imply majorization relations between the singular values of $Y_k^{\mathrm{T}} G Y_k$ and the singular values of the matrices $Q_k^{\mathrm{T}} G Q_k$ and $\hat{Q}_k^{\mathrm{T}} G \hat{Q}_k$. Consequently, for any unitarily invariant norm on $\mathbb{R}^{k \times k}$, the matrix Q_k solves the problem

$$\begin{aligned} \text{maximize} \quad & \left\| Y_k^{\mathrm{T}} G Y_k \right\| \\ \text{subject to} \quad & Y_k \in \mathbb{Y}_k, \end{aligned} \qquad (7.14)$$

while $\hat{Q}_k$ solves the problem

$$\begin{aligned} \text{minimize} \quad & \left\| Y_k^{\mathrm{T}} G Y_k \right\| \\ \text{subject to} \quad & Y_k \in \mathbb{Y}_k. \end{aligned} \qquad (7.15)$$

In the next section we move from symmetric orthogonal quotients matrices to rectangular ones. In this case singular values take the role of eigenvalues. Yet, as we shall see, the analogy between the two cases is not always straightforward.

8. Maximizing (Minimizing) Norms of Orthogonal Quotients Matrices

Let us return to consider orthogonal quotient matrices of the form (5.1). Define

$$k = \min\{\tilde{m}, \tilde{n}\},$$

and let

$$\tilde{\sigma}_1 \geq \tilde{\sigma}_2 \geq \cdots \geq \tilde{\sigma}_k \geq 0$$

denote the singular values of the orthogonal quotients matrix $X_{\tilde{m}}^{\mathrm{T}} A Y_{\tilde{n}}$. We shall start by investigating the problems

$$\begin{aligned} \text{maximize} \quad & F_p\left(X_{\tilde{m}}, Y_{\tilde{n}}\right) = \sum_{j=1}^{k} \tilde{\sigma}_j^p \\ \text{subject to} \quad & X_{\tilde{m}} \in \mathbb{X}_{\tilde{m}} \text{ and } Y_{\tilde{n}} \in \mathbb{Y}_{\tilde{n}} \end{aligned} \qquad (8.1)$$

and

$$\begin{aligned} \text{minimize} \quad & F_p\left(X_{\tilde{m}}, Y_{\tilde{n}}\right) = \sum_{j=1}^{k} \tilde{\sigma}_j^p \\ \text{subject to} \quad & X_{\tilde{m}} \in \mathbb{X}_{\tilde{m}} \text{ and } Y_{\tilde{n}} \in \mathbb{Y}_{\tilde{n}} \end{aligned} \qquad (8.2)$$

where p is a given positive real number, $0 < p < \infty$. Yet the coming results enable us to handle a larger family of objective functions. Perhaps the more interesting problems of this type occur when $p = 1$ and $p = 2$. In these cases the objective function is reduced to

$$F_1\left(X_{\tilde{m}}, Y_{\tilde{n}}\right) = \sum_{j=1}^{k} \tilde{\sigma}_j, \qquad (8.3)$$

and

$$F_2\left(X_{\tilde{m}}, Y_{\tilde{n}}\right) = \sum_{j=1}^{k} \tilde{\sigma}_j^2 = \left\| X_{\tilde{m}}^{\mathrm{T}} A Y_{\tilde{n}} \right\|_F^2, \qquad (8.4)$$

respectively. In particular, when $p = 2$ and $\tilde{m} = \tilde{n} = k$ problem (8.1) coincides with the Eckart-Young problem (6.12). The solution of (8.1) and (8.2) is based on "rectangular" extensions of Theorems 20 and 21. The first theorem is due to Thompson [37]. We outline its proof to clarify its close relation to Cauchy Interlace Theorem.

Theorem 23 (A Rectangular Cauchy Interlace Theorem) *Let the $\tilde{m} \times \tilde{n}$ matrix $\tilde{A}$ be obtained from A by deleting m' rows and n' columns of A. That is, $\tilde{m} + m' = m$ and $\tilde{n} + n' = n$. Define*

$$k = \min\{\tilde{m}, \tilde{n}\}$$

and let

$$\tilde{\sigma}_1 \geq \tilde{\sigma}_2 \geq \ldots \geq \tilde{\sigma}_k \geq 0$$

denote the singular values of $\tilde{A}$. Then

$$\sigma_j \geq \tilde{\sigma}_j \quad \text{for } j = 1, \cdots, k. \qquad (8.5)$$

Furthermore, the number of positive singular values of $\tilde{A}$ is bounded from below by

$$\ell = r - (m' + n'),$$

where $r = \text{rank}(A)$. Consequently $\ell \leq k$, and if $\ell \geq 1$ the first ℓ singular values of $\tilde{A}$ satisfy the lower bounds

$$\tilde{\sigma}_{\ell+1-i} \geq \sigma_{r+1-i} \quad \text{for } i = 1, \cdots, \ell. \qquad (8.6)$$

Proof. The proof is by induction on ℓ', where $\ell' = m' + n'$ is the overall number of deleted rows and columns. For $\ell' = 1$ there are two cases to consider. Assume first that $\tilde{A}$ is obtained by deleting one row of A. Then using Theorem 20 with $G = AA^{\mathrm{T}}$ and $\tilde{G} = \tilde{A}\tilde{A}^{\mathrm{T}}$ gives the desired results. The second possibility is that $\tilde{A}$ is obtained by deleting one column of A. In this case Theorem 20 is used with $G = A^{\mathrm{T}}A$ and $\tilde{G} = \tilde{A}^{\mathrm{T}}\tilde{A}$. Similar arguments enable us to complete the induction step. $\qquad\square$

Observe that the bounds (8.5) and (8.6) are "strict" in the sense that these bounds can be satisfied as equalities. Take, for example, a diagonal matrix.

Corollary 24 (A Rectangular Poincaré Separation Theorem) *Consider the $\tilde{m} \times \tilde{n}$ matrix $\tilde{A} = X_{\tilde{m}}^{\mathrm{T}} A Y_{\tilde{n}}$, where $X_{\tilde{m}} \in \mathbb{X}_{\tilde{m}}$ and $Y_{\tilde{n}} \in \mathbb{Y}_{\tilde{n}}$. Let*

$$\tilde{\sigma}_1 \geq \tilde{\sigma}_2 \geq \cdots \geq \tilde{\sigma}_k \geq 0$$

denote the singular values of $\tilde{A}$, where $k = \min\{\tilde{m}, \tilde{n}\}$. Then

$$\sigma_j \geq \tilde{\sigma}_j \quad \text{for } j = 1, \cdots, k. \qquad (8.7)$$

Furthermore, define $\ell = r - (m' + n')$ where $r = \text{rank}$ (A), $m' = m - \tilde{m}$, and $n' = n - \tilde{n}$. Then $\ell \leq k$ and if $\ell \geq 1$ the first ℓ singular values of $\tilde{A}$ satisfy the lower bounds

$$\tilde{\sigma}_{\ell+1-i} \geq \sigma_{r+1-i} \quad \text{for } i = 1, \cdots, \ell. \qquad (8.8)$$

Proof. Let the matrix $X_m \in \mathbb{X}_m$ be obtained by completing the columns of $X_{\tilde{m}}$ to be an orthonormal basis of $\mathbb{R}^m$. Let the matrix $Y_n \in \mathbb{Y}_n$ be obtained by completing the columns of $Y_{\tilde{n}}$ to be an orthonormal basis of $\mathbb{R}^n$. Then the $m \times n$ matrix $X_m^{\mathrm{T}} A Y_n$ has the same singular values as A, and $\tilde{A}$ is obtained from $X_m^{\mathrm{T}} A Y_n$ be deleting the last m' rows and the last n' columns.

Corollary 25 *Using the former notations,* $\qquad\square$

$$\sigma_j^p \geq \tilde{\sigma}_j^p \quad \text{for } j = 1, \cdots, k, \qquad (8.9)$$

and

$$\sum_{j=1}^{k} \sigma_j^p \geq \sum_{j=1}^{k} \tilde{\sigma}_j^p. \qquad (8.10)$$

Furthermore, if $\ell \geq 1$ then

$$\tilde{\sigma}_{\ell+1-i}^p \geq \sigma_{r+1-i}^p \quad \text{for } i = 1, \cdots, \ell, \qquad (8.11)$$

and

$$\sum_{j=1}^{\ell} \tilde{\sigma}_j^p \geq \sum_{i=1}^{\ell} \sigma_{r+1-i}^p. \qquad (8.12)$$

Similar inequalities hold when the power function $f(\theta) = \theta^p$ is replaced by any other real valued function which is increasing in the interval $[0, \infty)$.

Theorem 26 (A Rectangular Maximum Principle) *Let the $m \times \tilde{m}$ matrix $U_{\tilde{m}} = [u_1, \cdots, u_{\tilde{m}}]$ be constructed from the first $\tilde{m}$ columns of U, and let the $n \times \tilde{n}$ matrix $V_{\tilde{n}} = [v_1, \cdots, v_{\tilde{n}}]$ be constructed from the first $\tilde{n}$ columns of V. (Recall that U and V form the SVD of A, see (2.1)-(2.8).) Then this pair of matrices solves the maximum problem (8.1), giving the optimal value of $\sum_{j=1}^{k} \sigma_j^p$. That is,*

$$\sum_{j=1}^{k} \sigma_j^p = \max\left\{ \sum_{j=1}^{k} \tilde{\sigma}_j^p \,\middle|\, X_{\tilde{m}} \in \mathbb{X}_{\tilde{m}}, Y_{\tilde{n}} \in \mathbb{Y}_{\tilde{n}} \right\}. \qquad (8.13)$$

However, the solution matrices are not necessarily unique.

Proof. The proof is a direct consequence of (8.10) and the fact that $U_{\tilde{m}}^{\mathrm{T}} A V_{\tilde{n}}$ is a diagonal $\tilde{m} \times \tilde{n}$ matrix whose diagonal entries are $\sigma_j, j = 1, \cdots, k$. $\qquad\square$

Corollary 27 (A rectangular Ky Fan Maximum Principle) *Consider the special case when $p = 1$. In this case*

$$\sum_{j=1}^{k} \sigma_j = \max\left\{ \sum_{j=1}^{k} \tilde{\sigma}_j \,\middle|\, X_{\tilde{m}} \in \mathbb{X}_{\tilde{m}}, Y_{\tilde{n}} \in \mathbb{Y}_{\tilde{n}} \right\}, \qquad (8.14)$$

and the optimal value is attained for the matrices $U_{\tilde{m}}$ and $V_{\tilde{n}}$.

Corollary 28 *Consider the special case when $p = 2$. In this case*

$$\sum_{j=1}^{k} \sigma_j^2 = \max\left\{ \left\| X_{\tilde{m}}^{\mathrm{T}} A Y_{\tilde{n}} \right\|_F^2 \,\middle|\, X_{\tilde{m}} \in \mathbb{X}_{\tilde{m}}, Y_{\tilde{n}} \in \mathbb{Y}_{\tilde{n}} \right\}, \qquad (8.15)$$

and the optimal value is attained for the matrices $U_{\tilde{m}}$ and $V_{\tilde{n}}$. Furthermore, if $\tilde{m} = \tilde{n} = k$ then (8.15) is reduced to (6.12). This gives an alternative way to prove the Eckart-Young theorem.

Theorem 29 (A Rectangular Minimum Principle) *Let the $m \times \tilde{m}$ matrix $\hat{U}_{\tilde{m}} = [u_{m'+1}, \cdots, u_m]$ be obtained from U by deleting the first m' columns of U. Let the $n \times \tilde{n}$ matrix $\hat{V}_{\tilde{n}}$ be obtained from V by deleting n' columns in the following way: If $m' \geq \tilde{n}$ then $\hat{V}_{\tilde{n}}$ is composed from the first $\tilde{n}$ columns of V. Otherwise, when $m' < \tilde{n}$, the first m' columns of $\hat{V}_{\tilde{n}}$ are the first m' columns of V, and the rest columns of $\hat{V}_{\tilde{n}}$ are the*

last $\tilde{n} - m'$ *columns of* V. *Then the matrices* $\hat{U}_{\tilde{m}}$ *and* $\hat{V}_{\tilde{n}}$ *solve the minimum problem* (8.2). *The optimal value of* (8.2) *depends on the integer* $\ell = r - (m' + n')$. *If* $\ell \leq 0$ *the optimal value equals zero. Otherwise, when* $\ell \geq 1$, *the optimal value equals* $\sum_{i=1}^{\ell} \sigma_{r+1-i}^{p}$.

Proof. Let the $m \times \tilde{m}$ matrix M be obtained from the $m \times m$ identity matrix, I, by deleting the first m' columns of I. Then, clearly, $\hat{U}_{\tilde{m}} = UM$. Similarly, define N to be an $n \times \tilde{n}$ matrix such that $\hat{V}_{\tilde{n}} = VN$. That is, N is obtained from the $n \times n$ identity matrix by deleting the corresponding columns. With these notations at hand (2.1) implies the equalities

$$U_{\tilde{m}}^{\mathrm{T}} A V_{\tilde{n}} = M^{\mathrm{T}} U^{\mathrm{T}} A V N = M^{\mathrm{T}} S N.$$

So the matrix $U_{\tilde{m}}^{\mathrm{T}} A V_{\tilde{n}}$ is obtained from S by deleting the corresponding m' rows and n' columns.

Observe that the remaining nonzero entries of $M^{\mathrm{T}} S N$ are the singular values of this matrix. Note also that the rule for deleting rows and columns from S is aimed to make the size of the remaining nonzero entries as small as possible: The product $M^{\mathrm{T}} S$ deletes the first m' rows of S, which contain the largest m' singular values. Then the product $(M^{\mathrm{T}} S) N$ annihilates the next n' largest singular values. The remaining nonzero entries of S are, therefore, the smallest that we can get. The number of positive singular values in $U_{\tilde{m}}^{\mathrm{T}} A V_{\tilde{n}}$ is, clearly, $\max\{0, \ell\}$. The optimality of our solution stems from (8.8) and (8.12). $\qquad\square$

Another pair of matrices that solve (8.2) is gained by reversing the order in which we delete rows and columns from S: Start by deleting the first n' columns of S, which contain the n' largest singular values. Then delete the m' rows of S that contain the next m' largest singular values.

Corollary 30 (A rectangular Ky Fan Minimum Principle) *Consider the special case when* $p = 1$ *and* $\ell \geq 1$. *In this case*

$$\sum_{i=1}^{\ell} \sigma_{r+1-i} = \min\left\{ \sum_{j=1}^{k} \tilde{\sigma}_j \,\middle|\, X_{\tilde{m}} \in \mathbb{X}_{\tilde{m}}, Y_{\tilde{n}} \in \mathbb{Y}_{\tilde{n}} \right\}, \quad (8.16)$$

and the optimal value is attained for the matrices $\hat{U}_{\tilde{m}}$ *and* $\hat{V}_{\tilde{n}}$.

Corollary 31 *Consider the special case when* $p = 2$ *and* $\ell \geq 1$. *In this case*

$$\sum_{i=1}^{\ell} \sigma_{r+1-i}^{2} = \min\left\{ \left\| X_{\tilde{m}}^{\mathrm{T}} A Y_{\tilde{n}} \right\|_{F}^{2} \,\middle|\, X_{\tilde{m}} \in \mathbb{X}_{\tilde{m}}, Y_{\tilde{n}} \in \mathbb{Y}_{\tilde{n}} \right\}, \quad (8.17)$$

and the optimal value is attained for the matrices $\hat{U}_{\tilde{m}}$ *and* $\hat{V}_{\tilde{n}}$.

The next theorem extends our results to arbitrary unitarily invariant norms.

Theorem 32 *Let* $\|\cdot\|$ *be a unitarily invariant norm on*

$\mathbb{R}^{\tilde{m} \times \tilde{n}}$. *Then the matrices* $U_{\tilde{m}}$ *and* $V_{\tilde{n}}$, *which solve* (8.1), *also solve the problem*

$$\text{maximize} \quad F(X_{\tilde{m}}, Y_{\tilde{n}}) = \left\| X_{\tilde{m}}^{\mathrm{T}} A Y_{\tilde{n}} \right\|$$
$$\text{subject to} \quad X_{\tilde{m}} \in \mathbb{X}_{\tilde{m}} \text{ and } Y_{\tilde{n}} \in \mathbb{Y}_{\tilde{n}}. \quad (8.18)$$

Similarly the matrices $\hat{U}_{\tilde{m}}$ *and* $\hat{V}_{\tilde{n}}$, *which solve* (8.2), *also solve the problem*

$$\text{minimize} \quad F(X_{\tilde{m}}, Y_{\tilde{n}}) = \left\| X_{\tilde{m}}^{\mathrm{T}} A Y_{\tilde{n}} \right\|$$
$$\text{subject to} \quad X_{\tilde{m}} \in \mathbb{X}_{\tilde{m}} \text{ and } Y_{\tilde{n}} \in \mathbb{Y}_{\tilde{n}}. \quad (8.19)$$

Proof. From (8.7) we see that the singular values of $X_{\tilde{m}}^{\mathrm{T}} A Y_{\tilde{n}}$ are majorized by those of $U_{\tilde{m}}^{\mathrm{T}} A Y_{\tilde{n}}$. This shows that

$$\left\| U_{\tilde{m}}^{\mathrm{T}} A V_{\tilde{n}} \right\| \geq \left\| X_{\tilde{m}}^{\mathrm{T}} A Y_{\tilde{n}} \right\|, \quad (8.20)$$

which proves the first claim. Similarly, (8.8) means that the singular values of $\hat{U}_{\tilde{m}}^{\mathrm{T}} A \hat{V}_{\tilde{n}}$ are majorized by those of $X_{\tilde{m}}^{\mathrm{T}} A Y_{\tilde{n}}$. This shows that

$$\left\| X_{\tilde{m}}^{\mathrm{T}} A Y_{\tilde{n}} \right\| \geq \left\| \hat{U}_{\tilde{m}}^{\mathrm{T}} A \hat{V}_{\tilde{n}} \right\|, \quad (8.21)$$

which proves the second claim. $\qquad\square$

9. A Minimum-Maximum Equality

The Orthogonal Quotients Equality (6.1) connects the Eckart-Young minimum problem with an equivalent maximum problem. The validity of this equality depends on specific properties of the Frobenius matrix norm. The question raised in this section is whether it is possible to extend this equality to other unitarily invariant matrix norms. In other words, Mirsky's minimum norm problem (4.22) is related to the maximum norm problem (8.18) The next theorems answer this question when using the Shatten p-norm (2.16) in its power form,

$$\|A\|^p = \sum_{j=1}^{n} \sigma_j^p, \quad 1 \leq p < \infty. \quad (9.1)$$

Theorem 33 (A Minimum-Maximum Equality) *Assume that* $\tilde{m} = \tilde{n} = k$. *In this case the power function* (9.1) *satisfies the equality*

$$\min\left\{ \left\| A - \hat{X}_k \left(\hat{X}_k^{\mathrm{T}} A \hat{Y}_k \right) \hat{Y}_k^{\mathrm{T}} \right\|^p \,\middle|\, \hat{X}_k \in \mathbb{X}_k, \hat{Y}_k \in \mathbb{Y}_k \right\}$$
$$+ \max\left\{ \left\| \tilde{X}_k^{\mathrm{T}} A \tilde{Y}_k \right\|^p \,\middle|\, \tilde{X}_k \in \mathbb{X}_k, \tilde{Y}_k \in \mathbb{Y}_k \right\} = \|A\|^p. \quad (9.2)$$

Proof. The optimal value of the minimized term is given by Theorem 9 (Mirsky's theorem), and this value equals $\sum_{j=k+1}^{n} \sigma_j^p$. The optimal value of the other problem is determined by the maximum principle (8.13), and this value equals $\sum_{j=1}^{k} \sigma_j^p$. $\qquad\square$

If $p=1$ the power function (9.1) coincides with the trace norm (2.18). In this case (9.2) yields the following elegant result.

$$\min\left\{\left\|A-\hat{X}_k\left(\hat{X}_k^{\mathrm{T}}A\hat{Y}_k\right)\hat{Y}_k^{\mathrm{T}}\right\|_{tr}\middle|\hat{X}_k\in\mathbb{X}_k,\hat{Y}_k\in\mathbb{Y}_k\right\}$$
$$+\max\left\{\left\|\tilde{X}_k^{\mathrm{T}}A\tilde{Y}_k\right\|_{tr}\middle|\tilde{X}_k\in\mathbb{X}_k,\tilde{Y}_k\in\mathbb{Y}_k\right\}=\|A\|_{tr}. \qquad (9.3)$$

We have seen that Mirsky's minimization problem (4.22) and the maximum problem (8.18) share a common feature: The optimal values of both problems are obtained for the SVD matrices, U_k and V_k. This observation enables us to extend the minimum-maximum equality to other unitarily invariant norms. Consider, for example, the spectral norm (2.19). In this case (9.3) is replaced with

$$\min\left\{\left\|A-\hat{X}_k\left(\hat{X}_k^{\mathrm{T}}A\hat{Y}_k\right)\hat{Y}_k^{\mathrm{T}}\right\|_2\middle|\hat{X}_k\in\mathbb{X}_k,\hat{Y}_k\in\mathbb{Y}_k\right\}$$
$$+\max\left\{\left\|\tilde{X}_k^{\mathrm{T}}A\tilde{Y}_k\right\|_2\middle|\tilde{X}_k\in\mathbb{X}_k,\tilde{Y}_k\in\mathbb{Y}_k\right\}=\sigma_1+\sigma_{k+1}. \qquad (9.4)$$

Recall further that the jth columns of the SVD matrices form a pair of singular vectors that correspond to σ_j. Indeed, it is this property that ensures the equality in (9.2). This paves the way for another variant of the Orthogonal Quotients Equality.

Theorem 34 *Let the matrices* $X_k=[\boldsymbol{x}_1,\cdots,\boldsymbol{x}_k]\in\mathbb{X}_k$ *and* $Y_k=[\boldsymbol{y}_1,\cdots,\boldsymbol{y}_k]\in\mathbb{Y}_k$ *be composed from pairs of singular vectors of* A. *That is,* $X_k^{\mathrm{T}}X_k=Y_k^{\mathrm{T}}Y_k=I$, *and for* $j=1,\cdots,k$, *the jth columns of these matrices satisfy:* $A\boldsymbol{y}_j=\left(\boldsymbol{x}_j^{\mathrm{T}}A\boldsymbol{y}_j\right)\boldsymbol{x}_j$, $A^{\mathrm{T}}\boldsymbol{x}_j=\left(\boldsymbol{x}_j^{\mathrm{T}}A\boldsymbol{y}_j\right)\boldsymbol{y}_j$, *and the rectangular quotient* $\boldsymbol{x}_j^{\mathrm{T}}A\boldsymbol{y}_j$ *is a singular value of* A. *In this case the power function* (9.1) *satisfies the equality*

$$\left\|A-X_k\left(X_k^{\mathrm{T}}AY_k\right)Y_k^{\mathrm{T}}\right\|^p+\left\|X_k^{\mathrm{T}}AY_k\right\|^p=\|A\|^p. \qquad (9.5)$$

Proof. The term $\left\|X_kAY_k\right\|^p$ consists of k powers of singular values, while the other term consists of the rest $n-k$ powers. $\square$

10. Traces of Rectangular Matrices

We have seen that Ky Fan's extremum principles maximize and minimize traces of symmetric Rayleigh Quotient matrices. In this section we bring analog results in terms of rectangular matrices. For this purpose we define the trace of a rectangular $m\times n$ matrix as

$$\mathrm{trace}(A)=\sum_{i=1}^{q}a_{ii} \qquad (10.1)$$

where $q=\min\{m,n\}$. With this definition at hand the new problems to solve are

$$\text{maximize}\quad \mathrm{trace}\left(X_{\tilde{m}}^{\mathrm{T}}AY_{\tilde{n}}\right)$$
$$\text{subject to}\quad X_{\tilde{m}}\in\mathbb{X}_{\tilde{m}}\text{ and }Y_{\tilde{n}}\in\mathbb{Y}_{\tilde{n}} \qquad (10.2)$$

and

$$\text{minimize}\quad \mathrm{trace}\left(X_{\tilde{m}}^{\mathrm{T}}AY_{\tilde{n}}\right)$$
$$\text{subject to}\quad X_{\tilde{m}}\in\mathbb{X}_{\tilde{m}}\text{ and }Y_{\tilde{n}}\in\mathbb{Y}_{\tilde{n}} \qquad (10.3)$$

Using (2.10) we see that

$$-\sum_{j=1}^{n}\sigma_j\le\mathrm{trace}(A)\le\sum_{j=1}^{n}\sigma_j$$

and

$$-\sum_{j=1}^{k}\sigma_j\le-\sum_{j=1}^{k}\tilde{\sigma}_j\le\mathrm{trace}\left(X_{\tilde{m}}^{\mathrm{T}}AY_{\tilde{n}}\right)\le\sum_{j=1}^{k}\tilde{\sigma}_j\le\sum_{j=1}^{k}\sigma_j. \qquad (10.4)$$

On the other hand, the matrices $U_{\tilde{m}}$ and $V_{\tilde{n}}$ that solve (8.1) satisfy

$$\mathrm{trace}\left(U_{\tilde{m}}AV_{\tilde{n}}\right)=\sum_{j=1}^{k}\sigma_j \qquad (10.5)$$

and

$$\mathrm{trace}\left(-U_{\tilde{m}}AV_{\tilde{n}}\right)=-\sum_{j=1}^{k}\sigma_j, \qquad (10.6)$$

which leads to the following conclusions.

Corollary 35 *The matrices* $U_{\tilde{m}}$ *and* $V_{\tilde{n}}$ *solve* (10.2) *giving the optimal value of* $\sum_{j=1}^{k}\sigma_j$. *That is,*

$$\sum_{j=1}^{k}\sigma_j=\max\left\{\mathrm{trace}\left(X_{\tilde{m}}^{\mathrm{T}}AY_{\tilde{n}}\right)\middle|X_{\tilde{m}}\in\mathbb{X}_{\tilde{m}},Y_{\tilde{n}}\in\mathbb{Y}_{\tilde{n}}\right\}. \qquad (10.7)$$

Corollary 36 *The matrices* $-U_{\tilde{m}}$ *and* $V_{\tilde{n}}$ (*or* $U_{\tilde{m}}$ *and* $-V_{\tilde{n}}$) *solve* (10.3) *giving the optimal value of* $-\sum_{j=1}^{k}\sigma_j$. *That is*

$$-\sum_{j=1}^{k}\sigma_j=\min\left\{\mathrm{trace}\left(X_{\tilde{m}}^{\mathrm{T}}AY_{\tilde{n}}\right)\middle|X_{\tilde{m}}\in\mathbb{X}_{\tilde{m}},Y_{\tilde{n}}\in\mathbb{Y}_{\tilde{n}}\right\}. \qquad (10.8)$$

Finally we note that when $\tilde{m}=\tilde{n}=k$ the matrix $X_{\tilde{m}}^{\mathrm{T}}AY_{\tilde{n}}$ turns to be a square matrix and Corollary 35 is reduced to the following known result.

$$\sum_{j=1}^{k}\sigma_j=\max\left\{\mathrm{trace}\left(X_k^{\mathrm{T}}AY_k\right)\middle|X_k\in\mathbb{X}_k,Y_k\in\mathbb{Y}_k\right\}, \qquad (10.9)$$

e.g., [17, p.~195], [22, p.~515], [24].

11. Products of Eigenvalues versus Products of Singular Values

Ky Fan has used his extremum principles (Corollary 22) to derive analog results on determinants of positive semidefinite Rayleigh Quotient matrices (see below). In this section the new principle (Theorem 32) is used to extend these results to Orthogonal Quotient matrices. The interest in these problems stems from the following properties of symmetric matrices. Let G be a real sym-

metric positive semidefinite $n \times n$ matrix with eigenvalues

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0.$$

Let $Y_k \in \mathbb{Y}_k$ be an arbitrary $n \times k$ matrix with orthonormal columns, and let

$$\tilde{\lambda}_1 \geq \tilde{\lambda}_2 \geq \cdots \geq \tilde{\lambda}_k \geq 0$$

denote the eigenvalues of the $k \times k$ matrix $Y_k^T G Y_k$. Then, clearly,

$$\det\left(Y_k^T G Y_k\right) = \prod_{j=1}^{k} \tilde{\lambda}_j, \qquad (11.1)$$

while Corollary 21 implies the inequalities

$$\prod_{i=1}^{k} \lambda_{n+1-i} \leq \prod_{j=1}^{k} \tilde{\lambda}_j \leq \prod_{j=1}^{k} \lambda_j. \qquad (11.2)$$

Let the matrices Q_k and $\hat{Q}_k$ be defined as in Corollary 22. Then the eigenvalues of the matrices $Q_k^T G Q_k$ and $\hat{Q}_k^T G \hat{Q}_k$ are

$$\lambda_1 \geq \cdots \geq \lambda_k, \text{ and } \lambda_{n+1-k} \geq \cdots \geq \lambda_n,$$

respectively. Hence from (11.2) we see that

$$\prod_{j=1}^{k} \lambda_j = \max\left\{\det\left(Y_k^T G Y_k\right) = \prod_{j=1}^{k} \tilde{\lambda}_j \middle| Y_k \in \mathbb{Y}_k\right\} \qquad (11.3)$$

and

$$\prod_{i=1}^{k} \lambda_{n+1-i} = \min\left\{\det\left(Y_k^T G Y_k\right) = \prod_{j=1}^{k} \tilde{\lambda}_j \middle| Y_k \in \mathbb{Y}_k\right\}, \quad (11.4)$$

where optimal values are attained for the matrices Q_k and $\hat{Q}_k$, respectively. A further strengthening of (11.4) is gained by applying Hadamard determinant theorem, which says that the determinant of a symmetric positive semidefinite matrix, $Y_k^T G Y_k$, is smaller than the product of its diagonal entries. That is,

$$\det\left(Y_k^T G Y_k\right) \leq \prod_{j=1}^{k} \mathbf{y}_j^T G \mathbf{y}_j, \qquad (11.5)$$

where $\mathbf{y}_j$ denotes the jth column of Y_k. Combining (11.5) with (11.1) and (11.2) gives the inequality

$$\prod_{i=1}^{k} \lambda_{n+1-i} \leq \prod_{j=1}^{k} \mathbf{y}_j^T G \mathbf{y}_j, \qquad (11.6)$$

for any matrix $Y_k = [\mathbf{y}_1, \cdots, \mathbf{y}_k] \in \mathbb{Y}_k$. Also, as we have seen, equality holds in (11.6) when $Y_k = \hat{Q}_k$. This brings us to the following observation of Ky Fan [12].

Theorem 37 (Ky Fan)

$$\prod_{i=1}^{k} \lambda_{n+1-i} = \min\left\{\prod_{j=1}^{k} \mathbf{y}_j^T G \mathbf{y}_j \middle| Y_k \in \mathbb{Y}_k\right\}, \qquad (11.7)$$

and the optimal value is attained for $\hat{Q}_k$. $\qquad \square$

Let us return now to consider rectangular orthogonal quotients matrices. Using the notations of Section 8, the problems that we want to solve are

$$\text{maximize} \quad P\left(X_{\tilde{m}}, Y_{\tilde{n}}\right) = \prod_{j=1}^{k} \tilde{\sigma}_j \qquad (11.8)$$

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

and

$$\text{minimize} \quad P\left(X_{\tilde{m}}, Y_{\tilde{n}}\right) = \prod_{j=1}^{k} \tilde{\sigma}_j \qquad (11.9)$$

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

(Compare with (8.1) and (8.2), respectively.) Let the matrices $U_{\tilde{m}}$ and $V_{\tilde{n}}$ be defined as in Theorem 26. Using (2.1) one can verify that $P\left(U_{\tilde{m}}, V_{\tilde{n}}\right) = \prod_{j=1}^{k} \sigma_j$, which is the maximal possible value, see (8.7). This brings us to the following conclusions.

Theorem 38

$$\prod_{j=1}^{k} \sigma_j = \max\left\{\prod_{j=1}^{k} \tilde{\sigma}_j \middle| X_{\tilde{m}} \in \mathbb{X}_{\tilde{m}}, Y_{\tilde{n}} \in \mathbb{Y}_{\tilde{n}}\right\}, \qquad (11.10)$$

and the optimal value is attained for $U_{\tilde{m}}$ and $V_{\tilde{n}}$. $\qquad \square$

Corollary 39 *If $\tilde{m} = \tilde{n} = k$ then*

$$\prod_{j=1}^{k} \sigma_j = \det\left(U_k^T A V_k\right)$$

$$= \max\left\{\det\left(X_k^T A Y_k\right) \middle| X_k \in \mathbb{X}_k, Y_k \in \mathbb{Y}_k\right\}, \qquad (11.11)$$

where U_k and V_k are defined in (2.7). $\qquad \square$

The solution of (11.9) is found by following the notations and the proof of Theorem 29.

Theorem 40 *The matrices $\hat{U}_{\tilde{m}}$ and $\hat{V}_{\tilde{n}}$ solve (11.9). The number of positive singular values of the matrix $\hat{U}_{\tilde{m}}^T A \hat{V}_{\tilde{n}}$ is $\ell = r - m' - n'$. If $\ell < k$ then*

$$0 = \min\left\{\prod_{j=1}^{k} \tilde{\sigma}_j \middle| X_{\tilde{m}} \in \mathbb{X}_{\tilde{m}}, Y_{\tilde{n}} \in \mathbb{Y}_{\tilde{n}}\right\}. \qquad (11.12)$$

Otherwise, when $\ell = k$,

$$\prod_{i=1}^{\ell} \sigma_{r+1-i} = \min\left\{\prod_{j=1}^{k} \tilde{\sigma}_j \middle| X_{\tilde{m}} \in \mathbb{X}_{\tilde{m}}, Y_{\tilde{n}} \in \mathbb{Y}_{\tilde{n}}\right\}. \quad (11.13)$$

$$\square$$

Recall that $m \geq n$. Hence the equality $\ell = k$ is possible only when $r = n$. If $m > n$ then the equalities $\ell = k$ and $r = n$ imply $m' = 0$ and $\tilde{m} = m$. Otherwise, when $m = n = r$, the equality $\ell = k$ implies that either $m' = 0$ or $n' = 0$.

Corollary 41 *If $\tilde{m} = \tilde{n}$ then*

$$\left|\det\left(\hat{U}_{\tilde{m}} A \hat{V}_{\tilde{n}}\right)\right|$$

$$= \min\left\{\left|\det\left(X_{\tilde{m}}^T A Y_{\tilde{n}}\right)\right| \middle| X_{\tilde{m}} \in \mathbb{X}_{\tilde{m}}, Y_{\tilde{n}} \in \mathbb{Y}_{\tilde{n}}\right\}. \qquad (11.14)$$

12. Concluding Remarks

According to an old adage, the whole can sometimes be much more than the sum of its parts. The Rayleigh-Ritz procedure, the Eckart-Young theorem, and Ky Fan maximum principle are fundamental results that have several applications. The observation that these topics are closely related is new and surprising. It illuminates these issues in a new light.

The extended maximum principle is a powerful tool that has important consequences. In particular we see that both Eckart-Young's maximum problem and Ky Fan's maximum problem are special cases of this observation. The minimum-maximum theorem connects the extended maximum problem with Mirsky's minimum norm problem.

The review provides a second look at results of Ky Fan that consider eigenvalues of symmetric Rayleigh Quotient matrices. It extends these results to "rectangular" versions that consider singular values of Orthogonal Quotients matrices. The proofs illustrate the usefulness of Ky Fan's dominance theorem. With this theorem at hand Mirsky's theorem is easily derived from Weyl theorem. Similarly, it helps to establish the extended extremum principle.

REFERENCES

[1] R. Bhatia, "Matrix Analysis," Springer, New York, 1997. http://dx.doi.org/10.1007/978-1-4612-0653-8

[2] X. Chen and W. Li, "On the Rayleigh Quotient for Singular Values," *Journal of Computational Mathematic*, Vol. 25, 2007, pp. 512-521.

[3] A. Dax, "On Extremum Properties of Orthogonal Quotient Matrices," *Linear Algebra and its Applications*, Vol. 432, No. 5, 2010, pp. 1234-1257. http://dx.doi.org/10.1016/j.laa.2009.10.034

[4] J. W. Demmel, "Applied Numerical Linear Algebra," SIAM, Philadelphia, 1997. http://dx.doi.org/10.1137/1.9781611971446

[5] G. Eckart and G. Young, "The Approximation of One Matrix by Another of Lower Rank," *Psychometrika*, Vol. 1, No. 3, 1936, pp. 211-218. http://dx.doi.org/10.1007/BF02288367

[6] A. Edelman, T. Arias and S. Smith, "The Geometry of Algorithms with Orthogonality Constraints," *The SIAM Journal on Matrix Analysis and Applications*, Vol. 20, No. 2, 1998, pp. 303-353. http://dx.doi.org/10.1137/S0895479895290954

[7] A. Edelman and S. Smith, "On Conjugate Gradient-Like Methods for Eigen-Like Problems," *BIT Numerical Mathematics*, Vol. 36, No. 3, 1996, pp. 494-508. http://dx.doi.org/10.1007/BF01731929

[8] L. Elden, "Matrix Methods in Data Mining and Pattern Recognition," SIAM, Philadelphia, 2007. http://dx.doi.org/10.1137/1.9780898718867

[9] L. Elden and H. Park, "A Procrust *S* Problem on the Stiefel Manifold," Technical Report, Department of Mathematics, Linköping University, Linköping, 1997.

[10] K. Fan, "On a theorem of Weyl Concerning Eigenvalues of Linear Transformations I," *Proceedings of the National Academy of Sciences*, *USA*, Vol. 35, No. 11, 1949, pp. 652-655. http://dx.doi.org/10.1073/pnas.35.11.652

[11] K. Fan, "Maximum Properties and Inequalities for the Eigenvalues of Completely Continuous Operators," *Proceedings of the National Academy of Sciences*, *USA*, Vol. 37, No. 11, 1951, pp. 760-766. http://dx.doi.org/10.1073/pnas.37.11.760

[12] K. Fan, "A Minimum Property of the Eigenvalues of a Hermitian Transformation," *The American Mathematical Monthly*, Vol. 60, No. 1, 1953, pp. 48-50. http://dx.doi.org/10.2307/2306486

[13] E. Fischer, "Concerning Quadratic Forms with Real Coefficients," *Monatshefte für Mathematik und Physik*, Vol. 16, 1906, pp. 234-249.

[14] G. H. Golub and C. Van Loan, "An Analysis of the Total Least Squares Problem," *SIAM Journal on Numerical Analysis*, Vol. 17, No. 6, 1980, pp. 883-893. http://dx.doi.org/10.1137/0717073

[15] G. H. Golub and C. F. Van Loan, "Matrix Computations," Johns Hopkins University Press, Baltimore, 1983.

[16] R. A. Horn and C. R. Johnson, "Matrix Analysis," Cambridge University Press, 1985. http://dx.doi.org/10.1017/CBO9780511810817

[17] R. A. Horn and C. R. Johnson, "Topics in Matrix Analysis," Cambridge University Press, Cambridge, 1991. http://dx.doi.org/10.1017/CBO9780511840371

[18] R. M. Johnson, "On a Theorem Stated by Eckart and Young," *Psychometrica*, Vol. 28, No. 3, 1963, pp. 259-263. http://dx.doi.org/10.1007/BF02289573

[19] W. Kahan, B. N. Parlett and E. Jiang, "Residual Bounds on Approximate Eigensystems of Nonnormal Matrices," *SIAM Journal on Numerical Analysis*, Vol. 19, No. 3, 1982, pp. 470-484.

[20] R.-C. Li, "On Eigenvalues of a Rayleigh Quotient Matrix," *Linear Algebra and Its Applications*, Vol. 169, 1992, pp. 249-255. http://dx.doi.org/10.1016/0024-3795(92)90181-9

[21] X. G. Liu, "On Rayleigh Quotient Theory for the Eigenproblem and the Singular Value Problem," *Journal of Computational Mathematics*, Supplementary Issue, 1992, pp. 216-224.

[22] A. W. Marshall and I. Olkin, "Inequalities: Theory of Majorization and Its Applications," Academic Press, New York, 1979.

[23] L. Mirsky, "Symmetric Gauge Functions and Unitarily Invariant Norms," *Quarterly Journal of Mathematics*, Vol. 11, No. 1, 1960, pp. 50-59. http://dx.doi.org/10.1093/qmath/11.1.50

[24] J. von Neumann, "Some Matrix-Inequalities and Metrization of Matrix-Space," *Tomsk. Univ. Rev.*, Vol. 1, 1937, pp. 286-300. (In: A. H. Taub, Ed., *John von Neumann Collected Works*, Vol. IV, Pergamon, Oxford, 1962, pp.

205-218.)

[25] D. P. O'Leary and G. W. Stewart, "On the Convergence of a new Rayleigh Quotient Method with Applications to Large Eigenproblems," TR-97-74, Institute for Advanced Computer Studies, University of Maryland, College Park, 1997.

[26] A. M. Ostrowski, "On the Convergence of the Rayleigh Quotient Iteration for the Computation of the Characteristic Roots and Vectors. III (Generalized Rayleigh Quotient and Characteristic Roots with Linear Elementary Divisors," *Archive for Rational Mechanics and Analysis*, Vol. 3, No. 1, 1959, pp. 325-340. http://dx.doi.org/10.1007/BF00284184

[27] A. M. Ostrowski, "On the Convergence of the Rayleigh Quotient Iteration for the Computation of the Characteristic Roots and Vectors. IV (generalized Rayleigh Quotient for Nonlinear Elementary Divisors," *Archive for Rational Mechanics and Analysis*, Vol. 3, No. 1, 1959, pp. 341-347. http://dx.doi.org/10.1007/BF00284185

[28] M. L. Overton and R. S. Womersley, "On the Sum of the Largest Eigenvalues of a Symmetric Matrix," *SIAM Journal on Numerical Analysis*, Vol. 13, No. 1, 1992, pp. 41-45. http://dx.doi.org/10.1137/0613006

[29] B. N. Parlett, "The Rayleigh Quotient Iteration and Some Generalizations for Nonnormal Matrices," *Mathematics of Computation*, Vol. 28, No. 127, 1974, pp. 679-693. http://dx.doi.org/10.1090/S0025-5718-1974-0405823-3

[30] B. N. Parlett, "The Symmetric Eigenvalue Problem," Prentice-Hall, Englewood Cliffs, 1980.

[31] E. Schmidt, "Zur Theorie der Linearen und Nichtlinearen Integralgleichungen. I Teil. Entwicklung Willkürlichen Funktionen nach System Vorgeschriebener," *Mathematische Annalen*, Vol. 63, No. 4, 1907, pp. 433-476. http://dx.doi.org/10.1007/BF01449770

[32] G. W. Stewart, "Two Simple Residual Bounds for the Eigenvalues of Hermitian Matrices," *SIAM Journal on Matrix Analysis and Applications*, Vol. 12, No. 2, 1991, pp. 205-208. http://dx.doi.org/10.1137/0612016

[33] G. W. Stewart, "Matrix Algorithms," Vol. I: Basic Decompositions, SIAM, Philadelphia, 1998.

[34] G. W. Stewart, "Matrix Algorithms," Vol. II: Eigensystems, SIAM, Philadelphia, 2001.

[35] E. Stiefel, "Richtungsfelder und Fernparallelismus in *n*-Dimensionalen Mannigfaltigkeiten," *Commentarii Mathematici Helvetici*, Vol. 8, No. 1, 1935-1936, pp. 305-353.

[36] J. G. Sun, "Eigenvalues of Rayleigh Quotient Matrices," *Numerische Mathematik*, Vol. 59, No. 1, 1991, pp. 603-614. http://dx.doi.org/10.1007/BF01385798

[37] R. C. Thompson, "Principal Submatrices IX: Interlacing Inequalities for Singular Values of Submatrices," *Linear Algebra and its Applications*, Vol. 5, 1972, pp. 1-12.

[38] R. C. Thompson, "The Behavior of Eigenvalues and Singular Values under Perturbations of Restricted Rank," *Linear Algebra and Its Applications*, Vol. 13, No. 1-2, 1976, pp. 69-78. http://dx.doi.org/10.1016/0024-3795(76)90044-6

[39] H. Weyl, "Das Asymptotische Verteilungsgesetz der Eigenwerte Linearer Partieller Differentialgleichungen (Mit Einer Anwendung auf die Theorie der Hohlraumstrahlung," *Mathematische Annalen*, Vol. 71, No. 4, 1912, pp. 441-479. http://dx.doi.org/10.1007/BF01456804

[40] J. H. Wilkinson, "The Algebraic Eigenvalue Problem," Clarendon Press, Oxford, 1965.

[41] F. Zhang, "Matrix Theory: Basic Results and Techniques," Springer-Verlag, New York, 1999. http://dx.doi.org/10.1007/978-1-4757-5797-2