Journal of Computational Mathematics, Vol.9, No.4, 1991, 360–368.

EXTENSIONS OF THE KANTOROVICH INEQUALITY AND THE BAUER-FIKE INEQUALITY*1)

Sun Ji-guang
(*Computing Center, Academia Sinica, Beijing, China*)

Abstract

This paper proves a Kantorovich-type inequality on the matrix of the type

$$\frac{1}{2}\left(Q_1^H A Q_1 Q_1^H A^{-1} Q_1 + Q_1^H A^{-1} Q_1 Q_1^H A Q_1\right),$$

where A is an $n \times n$ positive definite Hermitian matrix and Q_1 is an $n \times m$ matrix with rank $(Q_1) = m$. The result is applied to get an extension of the Bauer-Fike inequality on condition numbers of similarities that block diagonalized matrices.

Let $A \in \mathbb{R}^{n\times n}$ (the set of complex $n \times n$ matrices), and let z_j, w_j be right and left eigenvectors of A corresponding to the eigenvalue λ_j, i.e.,

$$Az_j = \lambda_j z_j, \quad w_j^H A = \lambda_j w_j^H.$$

Define

$$s_j \equiv \cos\theta(z_j, w_j) = \frac{|w_j^H z_j|}{\|z_j\|_2 \|w_j\|_2},$$

where $\theta(z_j, w_j)$ denotes the angle between the one dimensional linear subspaces $\mathcal{R}(z_j)$ and $\mathcal{R}(w_j)$ spanned by z_j and w_j, respectively. Moreover, suppose that $Z, W \in \mathbb{C}^{n\times n}$ satisfy

$$W^H Z = I, \quad W^H A Z = \text{ diag } (\lambda_1, \cdots, \lambda_n), \tag{0.1}$$

and let

$$\kappa_2(A) \equiv \inf \|Z\|_2 \|Z^{-1}\|_2, \tag{0.2}$$

where $\|\cdot\|_2$ denotes the spectral norm, and the infimum is taken with respect to both matrices Z and W satisfying (0.1).

It is well known that if λ_j is a simple eigenvalue of A, then s_j is uniquely determined. Bauer and Fike [1] and Wilkinson [9] proved that the quantities s_j and $\kappa_2(A)$ give some measures of the sensitivity of the eigenvalues to perturbations of the elements of A, so s_j and $\kappa_2(A)$ are called condition numbers of the eigenvalues of A.

The condition numbers s_j and $\kappa_2(A)$ are related by the Bauer-Fike inequality[1]

$$\frac{1}{s_j} \le \frac{1}{2}\left(\kappa_2(A) + \frac{1}{\kappa_2(A)}\right). \tag{0.3}$$

This paper will give an extension of (0.3).

Suppose that $\mathcal{X}_1, \cdots, \mathcal{X}_r$ are linear subspaces of $\mathbb{C}^n$, and

$$\mathbb{C}^n = \mathcal{X}_1 \oplus \cdots \oplus \mathcal{X}_r, \quad \dim(\mathcal{X}_j) = m_j \quad \forall j. \tag{0.4}$$

* Received August 31, 1988.
1) The Project Supported by National Natural Science Foundation of China.

Let[3]

$$\mathcal{Y}_j = \bigcap_{\substack{k=1 \\ k\neq j}}^{r} \mathcal{X}_k^{\perp}, \quad j = 1, \cdots, r, \tag{0.5}$$

where $\mathcal{X}_k^{\perp}$ denotes the orthogonal complement subspace of $\mathcal{X}_k$ in $\mathbb{C}^n$. Obviously,

$$\dim (\mathcal{Y}_j) = m_j \quad \forall j, \quad \mathcal{Y}_1 \oplus \cdots \oplus \mathcal{Y}_r = \mathbb{C}^n. \tag{0.6}$$

Take matrices X_j, Y_j so that the columns of X_j, Y_j form orthonormal bases of $\mathcal{X}_j, \mathcal{Y}_j$, respectively. Since $(X_1, \cdots, X_r)$ and $(Y_1, \cdots, Y_r)$ are nonsingular $n \times n$ matrices, and

$$(Y_1, \cdots, Y_r)^H (X_1, \cdots, X_r) = \text{ diag } (Y_1^H X_1, \cdots, Y_r^H X_r),$$

the matrices $Y_j^H X_j$ are nonsingular. Define

$$\Theta(X_j, Y_j) \equiv \arccos(X_j^H Y_j Y_j^H X_j)^{\frac{1}{2}} > 0 \tag{0.7}$$

and

$$S_j \equiv \left\| [\cos \Theta(X_j, Y_j)]^{-1} \right\|^{-1}, \tag{0.8}$$

where $\| \cdot \|$ is any unitarily invariant norm, and $\Theta > 0 (\geq 0)$ denotes that Θ is a positive definite (semidefinite) Hermitian matrix. Especially, S_j will be written as $S_j^{(2)}$ or $S_j^{(F)}$ if we take the spectral norm $\| \cdot \|_2$ or the Frobenius norm $\| \cdot \|_F$ in (0.8), respectively.

The author [7] has proved that if $\mathcal{X}_j$ is an invariant right subspace of A corresponding to the semisimple eigenvalue λ_j of multiplicity m_j, then the quantity S_j^{-1} gives a measure of the sensitivity of the eigenvalue λ_j to perturbations of the elements of A.

The symbol $\mathcal{R}(\cdot)$ stands for the column space. Let

$$\mathcal{Z} = \left\{ Z \in \mathbb{C}^{n\times n} : Z = (Z_1, \cdots, Z_r), \quad Z_j \in \mathbb{C}^{n\times m_j}, \quad \mathcal{R}(Z_j) = \mathcal{X}_j \right\}, \tag{0.9}$$

and let

$$\kappa_2 \equiv \inf_{Z\in \mathcal{Z}} \|Z\|_2 \|Z^{-1}\|_2. \tag{0.10}$$

The Bauer-Fike inequality (0.3) has been extended by the author in the form ([7, Theorem 3.1])

$$\frac{1}{S_j^{(F)}} \leq \frac{\sqrt{m_j}}{2} (\kappa_2 + \frac{1}{\kappa_2}). \tag{0.11}$$

In this paper we shall give the following extension of (0.3).

Theorem 1. *Let $\mathcal{X}_1, \cdots, \mathcal{X}_r$ be linear subspaces of $\mathbb{C}^n$ satisfying (0.4). Let $\mathcal{Y}_1, \cdots, \mathcal{Y}_r$ be defined by (0.5), S_j by (0.8), and κ_2 by (0.10). Then*

$$\frac{1}{S_j^{(2)}} \leq \frac{1}{2} (\kappa_2 + \frac{1}{\kappa_2}), \quad j = 1, \cdots, r. \tag{0.12}$$

We can prove that inequalities (0.12) are equivalent to a result of Demmel [2] (a proof of the equivalence will be given in Appendix). We shall prove inequalities (0.12) by using a Kantorovich-type inequality stated in the following theorem.

Theorem 2. *Let $A \in \mathbb{C}^{n\times n}$ be any positive definite Hermitian matrix with the eigenvalues $\{\omega_j\}$ satisfying*

$$0 < l \leq \omega_n \leq \cdots \leq \omega_1 \leq L. \tag{0.13}$$

Further, let $Q_1 \in \mathbb{C}^{n\times m}$ and rank $(Q_1) = m$. Then

$$\frac{1}{2}(Q_1^H A Q_1 Q_1^H A^{-1} Q_1 + Q_1^H A^{-1} Q_1 Q_1^H A Q_1) \le \frac{(L+l)^2}{4Ll}(Q_1^H Q_1)^2. \tag{0.14}$$

We note that, for the special case $m = 1$, inequality (0.14) is the Kantorovich inequality (see [5, p.83]). Some extensions of the Kantorovich inequality have been made by Greub and Rheinboldt, Strang, Bloomfield and Watson, Knott, Khatri and Rao (see [6] and the references contained therein).

In the following we shall give proofs of Theorem 2 and Theorem 1, respectively in §1 and §2. In §3 we shall give another generalization of the Kantorovich inequality.

§1. Proof of Theorem 2

Decompose $A = U^H \Omega U$, where $U \in \mathbb{C}^{n\times n}$ is unitary, and $\Omega =$ diag $(\omega_1, \cdots, \omega_n)$. Let $U_1 = UQ_1$. Then the inequality (0.14) can be rewritten as

$$\frac{1}{2}(U_1^H \Omega U_1 U_1^H \Omega^{-1} U_1 + U_1^H \Omega^{-1} U_1 U_1^H \Omega U_1) \le \frac{(L+l)^2}{4Ll}(U_1^H U_1)^2. \tag{1.1}$$

Let $U_1^H = (v_1, \cdots, v_n)$. Then (1.1) is the following inequality:

$$H \equiv \frac{1}{2}\Big(\sum_{i=1}^n \omega_i v_i v_i^H \cdot \sum_{i=1}^n \frac{1}{\omega_i} v_i v_i^H + \sum_{i=1}^n \frac{1}{\omega_i} v_i v_i^H \cdot \sum_{i=1}^n \omega_i v_i v_i^H\Big) \le \frac{(L+l)^2}{4Ll}(U_1^H U_1)^2. \tag{1.2}$$

Obviously, we only need to consider the case of $L > l$. Following the way stated in [4] we prove inequality (1.2) as follows.

Let

$$\omega_i = L\varphi_i + l\psi_i, \quad \frac{1}{\omega_i} = \frac{\varphi_i}{L} + \frac{\psi_i}{l}, \quad i = 1, \cdots, n. \tag{1.3}$$

We get

$$\varphi_i, \psi_i \ge 0, \quad \varphi_i + \psi_i \le 1, \quad i = 1, \cdots, n. \tag{1.4}$$

Further, let

$$\Phi = \sum_{i=1}^n \varphi_i v_i v_i^H, \quad \Psi = \sum_{i=1}^n \psi_i v_i v_i^H. \tag{1.5}$$

Then

$$\Phi \ge 0, \quad \Psi \ge 0$$

and

$$0 \le \Phi + \Psi = \sum_{i=1}^n (\varphi_i + \psi_i) v_i v_i^H \le \sum_{i=1}^n v_i v_i^H = U_1^H U_1. \tag{1.6}$$

Substituting (1.3) into the left-hand side of (1.2), and using (1.5), we obtain

$$\begin{aligned}
H &= \frac{1}{2}\Big[\sum_{i=1}^{n}(L\varphi_i + l\psi_i)v_iv_i^H \cdot \sum_{i=1}^{n}(\frac{\varphi_i}{L} + \frac{\psi_i}{l})v_iv_i^H \\
&\quad + \sum_{i=1}^{n}(\frac{\varphi_i}{L} + \frac{\psi_i}{l})v_iv_i^H \cdot \sum_{i=1}^{n}(L\Phi + l\psi_i)v_iv_i^H\Big] \\
&= \frac{1}{2}\Big[(L\Phi + l\Psi)(\frac{1}{L}\Phi + \frac{1}{l}\Psi) + (\frac{1}{L}\Phi + \frac{1}{l}\Psi)(L\Phi + l\Psi)\Big] \\
&= (\Phi^2 + \Psi^2) + \frac{1}{2}(\frac{l}{L} + \frac{L}{l})(\Phi\Psi + \Psi\Phi) \\
&= (\Phi + \Psi)^2 + \frac{(L-l)^2}{2Ll}(\Phi\Psi + \Psi\Phi).
\end{aligned} \tag{1.7}$$

Since

$$2(\Phi\Psi + \Psi\Phi) \leq (\Phi + \Psi)^2,$$

from (1.7) and (1.6), we get

$$H \leq \Big[1 + \frac{(L-l)^2}{4Ll}\Big](\Phi + \Psi)^2 \leq \frac{(L+l)^2}{4Ll}(U_1^H U_1)^2.$$

Hence inequality (0.14) holds.

§2. Proof of Theorem 1

The symbol $\lambda(A)$ will be used to denote the set of the eigenvalues of a matrix A, and $\lambda_{\max}(A)$ the maximal eigenvalue of A if all the eigenvalues of A are real numbers.

Before the proof we cite the Bendixson theorem (see [5, p.69]).

Bendixson Theorem. *Let $A \in \mathbb{C}^{n\times n}, \lambda(A) = \{\alpha_j\}$. Moreover, let*

$$B = \frac{A + A^H}{2}, \quad C = \frac{A - A^H}{2i}, \quad i = \sqrt{-1},$$

and let $\lambda(B) = \{\beta_j\}, \lambda(C) = \{\gamma_j\}$. Then

$$\min_k\{\beta_k\} \leq \ Re\ (\alpha_j) \leq \max_k\{\beta_k\}, \quad \min_k\{\gamma_k\} \leq \ Im\ (\alpha_j) \leq \max_k\{\gamma_k\}, \quad j = 1, \cdots, n.$$

The following result is a simple corollary of the Bendixson theorem.

Lemma 2.1. *Let $A > 0, B > 0$, and let*

$$\lambda(AB) = \{\lambda_j\}, \quad \lambda(\frac{AB + BA}{2}) = \{\mu_j\}.$$

Then

$$\min_k\{\mu_k\} \leq \lambda_j \leq \max_k\{\mu_k\}, \quad j = 1, \cdots, n.$$

Proof of Theorem 1. Let $Z = (Z_1, \cdots, Z_r)$ be any fixed matrix of $\mathcal{Z}$ with $Z_j \in \mathbb{C}^{n\times m_j} \forall j$, and let

$$W = (W_1, \cdots, W_r) = Z^{-H}, \quad W_j \in \mathbb{C}^{n\times m_j} \quad \forall\ j.$$

Further, let

$$X_j = Z_j(Z_j^H Z_j)^{-\frac{1}{2}}, \quad Y_j = W_j(W_j^H W_j)^{-\frac{1}{2}}, \quad j = 1, \cdots, r.$$

Then

$$\mathcal{R}(X_j) = \mathcal{X}_j, \quad \mathcal{R}(Y_j) = \mathcal{Y}_j$$

and

$$X_j^H X_j = Y_j^H Y_j = I^{(m_j)}, \quad j = 1, \cdots, r.$$

By the definition of $S_j^{(2)}$ (see (0.7) and (0.8)), we have

$$\begin{aligned} S_j^{(2)^{-1}} &= \left\| (X_j^H Y_j Y_j^H X_j)^{-\frac{1}{2}} \right\|_2 = \left\| \left[(Z_j^H Z_j)^{-\frac{1}{2}} (W_j^H W_j)^{-1} (Z_j^H Z_j)^{-\frac{1}{2}} \right]^{-\frac{1}{2}} \right\|_2 \\ &= \left\| \left[(Z_j^H Z_j)^{\frac{1}{2}} W_j^H W_j (Z_j^H Z_j)^{\frac{1}{2}} \right]^{\frac{1}{2}} \right\|_2 = \left\| (Z_j^H Z_j)^{\frac{1}{2}} W_j^H W_j (Z_j^H Z_j)^{\frac{1}{2}} \right\|_2^{\frac{1}{2}}. \end{aligned} \tag{2.1}$$

Let

$$E_j = (\underset{m_1}{0}, \cdots, \underset{m_{j-1}}{0}, I^{(m_j)}, \underset{m_{j+1}}{0}, \cdots, \underset{m_r}{0})^T.$$

Then $Z_j = ZE_j, W_j = WE_j$. Decompose $Z^H Z = U^H \Omega U$, where U is a unitary matrix, and

$$\Omega = \operatorname{diag}(\omega_1 \cdots \omega_n), \quad 0 < \omega_n \leq \cdots \leq \omega_1.$$

Then

$$W^H W = (Z^H Z)^{-1} = U^H \Omega^{-1} U.$$

Further, let $U_j = UE_j$. Then $U_j^H U_j = I^{(m_j)}$, and from (2.1) it follows that

$$\begin{aligned} S_j^{(2)^{-2}} &= \left\| (Z_j^H Z_j)^{\frac{1}{2}} W_j^H W_j (Z_j^H Z_j)^{\frac{1}{2}} \right\|_2 = \left\| (U_j^H \Omega U_j)^{\frac{1}{2}} U_j^H \Omega^{-1} U_j (U_j^H \Omega U_j)^{\frac{1}{2}} \right\|_2 \\ &= \lambda_{\max}\left((U_j^H \Omega U_j)^{\frac{1}{2}} U_j^H \Omega^{-1} U_j (U_j^H \Omega U_j)^{\frac{1}{2}} \right) = \lambda_{\max}(U_j^H \Omega U_j U_j^H \Omega^{-1} U_j). \end{aligned}$$

By Lemma 2.1,

$$S_j^{(2)^{-2}} \leq \lambda_{\max}\left(\frac{1}{2} \left[U_j^H \Omega U_j U_j^H \Omega^{-1} U_j + U_j^H \Omega^{-1} U_j U_j^H \Omega U_j \right] \right).$$

By Theorem 1,

$$S_j^{(2)^{-2}} \leq \frac{(\omega_1 + \omega_n)^2}{4\omega_1 \omega_n} = \left[\frac{1}{2} \left(\sqrt{\frac{\omega_1}{\omega_n}} + \sqrt{\frac{\omega_n}{\omega_1}} \right) \right]^2,$$

i.e.,

$$S_j^{(2)^{-1}} \leq \frac{1}{2} \left(\|Z\|_2 \|Z^{-1}\|_2 + \frac{1}{\|Z\|_2 \|Z^{-1}\|_2} \right). \tag{2.2}$$

Observe that $\tau + \tau^{-1}$ is a monotone increasing function for $\tau \geq 1$. Hence from (2.2) we get

$$\begin{aligned} S_j^{(2)^{-1}} &\leq \inf_{z \in Z} \left\{ \frac{1}{2} (\|Z\|_2 \|Z^{-1}\|_2 + \frac{1}{\|Z\|_2 \|Z^{-1}\|_2}) \right\} \\ &= \frac{1}{2} \left(\inf_{z \in Z} \|Z\|_2 \|Z^{-1}\|_2 + \frac{1}{\inf_{z \in Z} \|Z\|_2 \|Z^{-1}\|_2} \right) \\ &= \frac{1}{2} (\kappa_2 + \frac{1}{\kappa_2}), \quad j = 1, \cdots, r. \end{aligned}$$

This completes the proof.

Remark 2.2. Inequality (0.11) can be proved by applying Theorem 1. In fact, using the notation used in the proof of Theorem 1, by (0.8) we have

$$S_j^{(F)^{-1}} = \|(X_j^H Y_j Y_j^H X_j)^{-\frac{1}{2}}\|_F \leq \sqrt{m_j} \cdot \|(X_j^H Y_j Y_j^H X_j)^{-\frac{1}{2}}\|_2.$$

By (0.12) we get

$$\begin{aligned} S_j^{(F)^{-1}} &\leq \left(\frac{m_j(\omega_1+\omega_n)^2}{4\omega_1\omega_n}\right)^{\frac{1}{2}} = \frac{\sqrt{m_j}}{2}\left(\sqrt{\frac{\omega_1}{\omega_n}} + \sqrt{\frac{\omega_n}{\omega_1}}\right) \\ &= \sqrt{\frac{m_j}{2}}\left(\|Z\|_2\|Z^{-1}\|_2 + \frac{1}{\|Z\|_2\|Z^{-1}\|_2}\right). \end{aligned}$$

Hence inequality (0.11) holds.

§3. Another Generalization of the Kantorovich Inequality

Now we give another generalization of the Kantorovich inequality.

Theorem 3.1. *Let $A \in \mathbb{C}^{n\times n}$ be nonsingular with the singular values $\{\sigma_i\}$ satisfying*

$$0 < l \leq \sigma_n \leq \cdots \leq \sigma_1 \leq L.$$

Further, let $P_1, Q_1 \in \mathbb{C}^{n\times m}$ and $P_1^H P_1 = Q_1^H Q_1 = I^{(m)}$. Then

$$\frac{1}{2}(P_1^H A Q_1 Q_1^H A^{-1} P_1 + P_1^H A^{-H} Q_1 Q_1^H A^H P_1) \leq \frac{(L+l)^2}{4Ll} I^{(m)}. \tag{3.1}$$

Proof. Assume that $A = U\Sigma V^H$ is the singular value decomposition of A, where U, V are unitary matrices, and $\Sigma = \text{diag}\ (\sigma_1, \cdots, \sigma_n)$. Let

$$U_1 = U^H P_1, \quad V_1 = V^H Q_1.$$

Then $U_1^H U_1 = V_1^H V_1 = I^{(m)}$, and the left-hand side of inequality (3.1) can be rewritten as

$$H \equiv \frac{1}{2}(U_1^H \Sigma V_1 V_1^H \Sigma^{-1} U_1 + U_1^H \Sigma^{-1} V_1 V_1^H \Sigma U_1).$$

Further, let

$$\begin{aligned} &U_1^H = (z_1, \cdots, z_n), \quad V_1^H = (w_1, \cdots, w_n), \\ &\sigma_i = L\varphi_i + l\psi_i, \quad \frac{1}{\sigma_i} = \frac{\varphi_i}{L} + \frac{\psi_i}{l}, \quad i = 1, \cdots, n, \\ &\Phi_0 = \text{ diag } (\varphi_1, \cdots, \varphi_n), \Psi_0 = \text{ diag } (\psi_1, \cdots, \psi_n) \end{aligned}$$

and

$$\Phi = U_1^H \Phi_0 V_1, \quad \Psi = U_1^H \Psi_0 V_1.$$

Then we have

$$\Phi_0, \Psi_0 \geq 0, \quad \Phi_0 + \Psi_0 \leq I^{(n)}$$

and

$$\begin{aligned}
H &= \frac{1}{2}\Big(\sum_{i=1}^{n}\sigma_i z_i w_i^H \sum_{i=1}^{n}\frac{1}{\sigma_i}w_i z_i^H + \sum_{i=1}^{n}\frac{1}{\sigma_i}z_i w_i^H \sum_{i=1}^{n}\sigma_i w_i z_i^H\Big)\\
&= \frac{1}{2}\Big[\sum_{i=1}^{n}(L\varphi_i + l\psi_i)z_i w_i^H \sum_{i=1}^{n}(\frac{\varphi_i}{L}+\frac{\psi_i}{l})w_i z_i^H\\
&+\sum_{i=1}^{n}(\frac{\varphi_i}{L}+\frac{\psi_i}{l})z_i w_i^H \sum_{i=1}^{n}(L\varphi_i + l\psi_i)w_i z_i^H\Big]\\
&= \frac{1}{2}\Big[(L\Phi + l\Psi)(\frac{1}{L}\Phi^H + \frac{1}{l}\Psi^H) + (\frac{1}{L}\Phi + \frac{1}{l}\Psi)(L\Phi^H + l\Psi^H)\Big]\\
&= \Phi\Phi^H + \Psi\Psi^H + \frac{1}{2}(\frac{l}{L}+\frac{L}{l})(\Phi\Psi^H + \Psi\Phi^H)\\
&= (\Phi+\Psi)(\Phi+\Psi)^H + \frac{(L-l)^2}{2Ll}(\Phi\Psi^H + \Psi\Phi^H).
\end{aligned} \tag{3.2}$$

Since

$$2(\Phi\Psi^H + \Psi\Phi^H) \le (\Phi+\Psi)(\Phi+\Psi)^H,$$

it follows from (3.2) that

$$H \le (1 + \frac{(L-l)^2}{4Ll})(\Phi+\Psi)(\Phi+\Psi)^H = \frac{(L+l)^2}{4Ll}(\Phi+\Psi)(\Phi+\Psi)^H. \tag{3.3}$$

Observe that

$$(\Phi+\Psi)(\Phi+\Psi)^H = U_1^H(\Phi_0+\Psi_0)V_1V_1^H(\Phi_0+\Psi_0)U_1 \le I^{(m)};$$

hence from (3.3) we get

$$H \le \frac{(L+l)^2}{4Ll}I^{(m)}.$$

This proves inequality (3.1).

From Theorem 3.1 we get the following corollary immediately.

Corollary 3.2. *Let $A \in \mathbb{C}^{n\times n}$ be as in Theorem 3.1, and let $p_1, q_1 \in \mathbb{C}^n$ with $\|p_1\|_2 = \|q_1\|_2 = 1$. Then*

$$\mathrm{Re}\,(p_1^H A q_1 q_1^H A^{-1} p_1) \le \frac{(L+l)^2}{4Ll}.$$

§4. Appendix

Let $Z = (z_1, \cdots, z_r) \in \mathcal{Z}$, $Z_j \in \mathbb{C}^{n\times m_j}$ and $\mathcal{R}(Z_j) = \mathcal{X}_j$ for $j = 1, \cdots, r$, where $\mathcal{X}_j$ are defined by (0.4). Further, let

$$X_j = Z_j(Z_j^H Z_j)^{-\frac{1}{2}}, \quad \hat{X}_j = (X_1, \cdots, X_{j-1}, X_{j+1}, \cdots, X_r), \quad X_j' = \hat{X}_j(\hat{X}_j^H\hat{X}_j)^{-\frac{1}{2}},$$

and define

$$\theta_j \equiv \arccos\left\|X_j^H X_j'\right\|_2, \quad j = 1, \cdots, r. \tag{4.1}$$

Demmel [2] has proved that the inequalities

$$\csc\theta_j + \sqrt{\csc^2\theta_j - 1} \le \kappa_2, \quad j = 1, \cdots, r \tag{4.2}$$

hold.

Now we prove that inequalities (2) and (0.12) are equivalent.

First of all we prove that the relations

$$\frac{1}{S_j^{(2)}} = \csc\theta_j, \quad j = 1, \cdots, r \tag{4.3}$$

hold. Without loss of generality we may consider $j = 1$. Let $2m_1 \le n$. By [7, Theorem 1.1] and the invertibility of the matrix (X_1, X_1'), there are unitary matrices $Q \in \mathbb{C}^{n\times n}, U_1 \in \mathbb{C}^{m_1\times m_1}$ and $U_1' \in \mathbb{C}^{(n-m_1)\times(n-m_1)}$ such that

$$QX_1U_1 = \begin{array}{cc} \begin{pmatrix} I \\ 0 \\ 0 \end{pmatrix} & \begin{array}{l} m_1 \\ n-2m_1 \\ m_1 \end{array} \\ m_1 & \end{array}, \quad QX_1'U_1' = \begin{array}{cc} \begin{pmatrix} \Gamma & 0 \\ 0 & I \\ \Sigma & 0 \end{pmatrix} & \begin{array}{l} m_1 \\ n-2m_1 \\ m_1 \end{array} \\ m_1 \quad n-2m_1 & \end{array}, \tag{4.4}$$

where

$$\Gamma = \operatorname{diag}(\gamma_1, \cdots, \gamma_{m_1}), \quad \Sigma = \operatorname{diag}(\sigma_1, \cdots, \sigma_{m_1}),$$

and

$$\sigma_1 \ge \cdots \ge \sigma_{m_1} > 0, \quad 0 \le \gamma_1 \le \cdots \le \gamma_{m_1}, \quad \sigma_j^2 + \gamma_j^2 = 1, \quad j = 1, \cdots, r.$$

Let

$$X = (X_1, X_1'), \quad W = X^{-H} = (W_1, W_1'), \quad W_1 \in \mathbb{C}^{n\times m_1}, \tag{4.5}$$

and let

$$\mathcal{Y}_1 = \mathcal{R}(W_1).$$

Then it is easy to verify that

$$\mathcal{Y}_1 = \bigcap_{k=2}^{r} \mathcal{X}_k^{\perp}.$$

Further, let

$$Y_1 = W_1(W_1^H W_1)^{-\frac{1}{2}}.$$

Then, by (0.8)

$$\frac{1}{S_1^{(2)}} = \left\| [\cos\Theta(X_1, Y_1)]^{-1} \right\|_2 = \left\| (X_1^H Y_1 Y_1^H X_1)^{-\frac{1}{2}} \right\|_2. \tag{4.6}$$

Observe that, from (4) and (5)

$$W = (X_1, X_1')^{-H} = \left[Q^H \left(\begin{array}{c:cc} I & \Gamma & 0 \\ \hdashline 0 & 0 & I \\ 0 & \Sigma & 0 \end{array} \right) \begin{pmatrix} U_1^H & 0 \\ 0 & U_1'^H \end{pmatrix} \right]^{-H}$$

$$= Q^H \left(\begin{array}{c:cc} I & 0 & 0 \\ \hdashline 0 & 0 & I \\ -\Sigma^{-1}\Gamma & \Sigma^{-1} & 0 \end{array} \right) \begin{pmatrix} U_1^H & 0 \\ 0 & U_1'^H \end{pmatrix}.$$

Therefore

$$W_1 = Q^H \begin{pmatrix} I \\ 0 \\ -\Sigma\Gamma \end{pmatrix} U_1^H, \quad Y_1 = Q^H \begin{pmatrix} \Sigma \\ 0 \\ -\Gamma \end{pmatrix} U_1^H. \tag{4.7}$$

On the one hand, substituting (4) and (7) into (6) we get

$$\frac{1}{S_1^{(2)}} = \left\|(U_1\Sigma^2U_1^H)^{-\frac{1}{2}}\right\|_2 = \|\Sigma^{-1}\|_2 = \frac{1}{\sigma_{m_1}}, \tag{4.8}$$

and on the other hand, by (1) and (4),

$$\cos\theta_1 = \|X_1^H X_1'\|_2 = \|U_1^H(\Gamma, 0)U_1'^H\|_2 = \gamma_{m_1},$$

and so

$$\csc\theta_1 = \frac{1}{\sqrt{1-\cos^2\theta_1}} = \frac{1}{\sigma_{m_1}}. \tag{4.9}$$

Equalities (8) and (9) give (3) for $j = 1$ and $2m_1 \le n$. With the same argument we can prove that the relations (3) hold for $2m_1 > n$.

By (3) we can rewrite inequalities (2) as

$$\frac{1}{S_j^{(2)}} + \sqrt{(\tfrac{1}{S_j^{(2)}})^2 - 1} \le \kappa_2, \quad j = 1, \cdots, r. \tag{4.10}$$

Moreover, inequalities (0.12) can be rewritten as

$$\kappa_2^2 - \frac{2\kappa_2}{S_j^{(2)}} + 1 \ge 0, \quad j = 1, \cdots, r. \tag{4.11}$$

It is obvious that inequalities (10) and (11) are equivalent, and so inequalities (2) and (0.12) are equivalent.

References

[1] F.L. Bauer and C. T. Fike, Norms and exclusion theorems, *Numer. Math.*, **2** (1960), 137–141.

[2] J. Demmel, The condition number of equivalunce transformations that block diagonalize matrix pencils, in Matrix Pencils (B. Kågström and A. Ruhe. Eds.), Lecture Notes in Mathematics 973, Springer, 1982, 2–16.

[3] L. Elsner, Block scaling with optimal Euclidean condition, *Linear Algebra and Appl.*, **58** (1984), 69–73.

[4] P. Henrici, Two remarks on the Kantorovich inequality, *Amer. Math. Monthly*, **68** (1961), 904–906.

[5] A.S. Householder, The Theory of Matrices in Numerical Analysis, Blaisedell, New York, 1964.

[6] C.G. Khatri and C.R. Rao, Some extensions of the Kantorovich inequality and statistical applications, *J. Multivariate Analysis*, **11** (1981), 498–505.

[7] J.G. Sun, The stability of orthogonal projections, *J. Graduate School*, **1** (1984), 123–133.

[8] J.G. Sun, Condition numbers of semisimple multiple eigenvalues, submitted to *Math. Numer. Sinica*.

[9] J.H. Wilkinson, The Algebraic Eigenvalue Problem, Oxford University Press, 1965.