Perturbation Theorems for the Generalized Eigenvalue Problem

Ludwig Elsner
Fakultät für Mathematik
Universität Bielefeld
Postfach 8640
4800 Bielefeld 1, Federal Republic of Germany

and

Ji-guang Sun*
Computing Center
Academia Sinica
Peking, China

Submitted by Shmuel Friedland

ABSTRACT

Given a matrix pair $Z=(A,B)$, the perturbation of its eigenvalues (α,β) is studied. Considering two pairs Z,W as points of the Grassmann manifold $G_{n,2n}$ and its eigenvalues as points in $G_{1,2}$, the projective complex plane, the distance of the spectra, measured in the chordal metric in $G_{1,2}$, is bounded by some distance of the matrix pairs in $G_{n,2n}$. Analogs of the Bauer-Fike theorem, Henrici's theorem, and the Hoffman-Wielandt theorem are obtained, from which the "classical" results can be derived.

0. INTRODUCTION

In this paper we shall treat the perturbation of the eigenvalues of the generalized eigenvalue problem

$$\beta Ax=\alpha Bx. \tag{0.1}$$

Here A, B are complex $n\times n$ matrices, x a nonzero vector, and $(\alpha,\beta)\neq(0,0)$ a pair of complex numbers.

*Visiting the University of Bielefeld. The work of this author was assisted by the Alexander von Humboldt Foundation of the Federal Republic of Germany.

52 Vanderbilt Ave., New York, NY 10017 0024-3795/82/070341-17$2.75

In the case of the classical eigenvalue problem $Ax = \lambda x$, there is a series of theorems which give bounds on the perturbation of the spectrum of A depending on A and the norm of the perturbation of A. For citing these results we introduce some notation. Let A and C have the eigenvalues λ_i and μ_i, $i = 1,\dots,n$, respectively. $S_A(C) = \max_i \min_j |\mu_i - \lambda_j|$ is the spectral variation of C with respect to A, and $v(A,C) = \min_\pi \max_i |\lambda_i - \mu_{\pi(i)}|$, where π runs through the permutations of $\{1,\dots,n\}$, is the eigenvalue variation of A and C [3]. Let $\| \ \|$ denote the usual Euclidean vector norm, $\| \ \|_2$ the spectral norm, and $\| \ \|_F$ the Euclidean matrix norm.

THEOREM 0.1 (Bauer and Fike [1]). *If A is diagonalizable, i.e. there exists a nonsingular $n \times n$-matrix T such that*

$$A = T\operatorname{diag}(\lambda_1,\dots,\lambda_n)T^{-1},$$

then

$$S_A(C) \leqslant \|T\|_2 \|T^{-1}\|_2 \|A - C\|_2.$$

THEOREM 0.2 (Henrici [3]). *If Δ is the $\| \ \|_2$-departure from normality of A and $\Delta \neq 0$, then*

$$S_A(C) \leqslant \frac{y}{g_n(y)} \|A - C\|_2, \qquad y = \frac{\Delta}{\|A - C\|_2},$$

where $g_n(x)$ is the inverse function of $x + x^2 + \cdots + x^n$ for nonnegative x.

If A is normal, then by Theorem 0.1

$$S_A(C) \leqslant \|A - C\|_2,$$

but we have an even stronger result.

THEOREM 0.3 (Hoffman and Wielandt [4]). *If A and C are both normal, then*

$$v(A,C) \leqslant \min_\pi \left(\sum_i |\lambda_i - \mu_{\pi(i)}|^2 \right)^{1/2} \leqslant \|A - C\|_F.$$

What are the corresponding results for the generalized eigenvalue problem (0.1)?

It is reasonable to consider an eigenvalue $(\alpha,\beta)\neq(0,0)$ as a point in the projective complex plane $G_{1,2}$ and measure the distance between two points in the chordal metric

$$\rho((\alpha,\beta),(\tilde{\alpha},\tilde{\beta}))=\frac{|\alpha\tilde{\beta}-\beta\tilde{\alpha}|}{\sqrt{|\alpha|^2+|\beta|^2}\sqrt{|\tilde{\alpha}|^2+|\tilde{\beta}|^2}} \tag{0.2}$$

(see e.g. [7], [8], [9]). If $Z=(A,B)$ is a pair of $n\times n$ matrices, identified with the $n\times 2n$ matrix Z, then obviously $TZ=(TA,TB)$ for a nonsingular $n\times n$ matrix T has the same eigenvalues and eigenvectors. Hence we consider $Z=(A,B)$ as an element of the Grassman manifold $G_{n,2n}$ of all n-dimensional subspaces of the $2n$-dimensional complex space C^{2n} by identifying Z with the linear subspace $L_Z=\{Z^Tx: x\in C^n\}$ spanned by the row vectors of Z (see e.g. [6], [12]). Here we use the obvious fact that for a regular pair (A,B) [i.e. $\det(A+\lambda B)\not\equiv 0$] the dimension of L_Z is n.

If $Z=(A,B)$, $W=(C,D)$ are regular pairs, we use as a "distance" $d_2(Z,W)$ the "gap" between the corresponding subspaces L_Z and L_W. Here the "gap" is defined in the usual way as the norm of the difference of the orthogonal projectors [5, 8]. As those are given by

$$P_Z=Z^H(ZZ^H)^{-1}Z,\qquad P_W=W^H(WW^H)^{-1}W, \tag{0.3}$$

we have the metric

$$d_2(Z,W)=\|P_Z-P_W\|_2=\left\|Z^H(ZZ^H)^{-1}Z-W^H(WW^H)^{-1}W\right\|_2. \tag{0.4}$$

Besides this we use (see also [8]) the metric

$$d_F(Z,W)=\sqrt{\tfrac{1}{2}}\,\|P_Z-P_W\|_F. \tag{0.5}$$

Note that for $n=1$, the projective complex plane $G_{1,2}$, d_F and d_2 coincide with the chordal metric (0.2).

We shall show in the sequel that there are corresponding theorems replacing in (0.1)–(0.3) the distances between complex numbers and matrices by the chordal metric and d_2 or d_F respectively. For this purpose one has to define "diagonalizable," "normal," and "departure from normality" for matrix pairs appropriately.

In Section 1 we give the abovementioned definitions and some basic results. In the following sections we prove the corresponding "Bauer-Fike," "Henrici," and "Hoffman-Wielandt" theorems. That these generalizations are not just formal ones is shown in the last section. Here the theorems (0.1)–(0.3) are derived by a limiting argument from the corresponding results for matrix pairs. This papers is a continuation of [12] (by Sun alone), where some of the following results (Theorem 2.2 in a special case, and Theorem 4.1) have been proved in a different way.

1. DEFINITIONS AND BASIC RESULTS

Let $C_{n,m}$ denote the set of all $n \times m$ complex matrices.

DEFINITION 1. Let $A, B \in C_{n,n}$. A vector $x \in C_n$, $x \neq 0$, is an *eigenvector* of the matrix pair (A, B) corresponding to the *eigenvalue* (α, β), $\alpha, \beta \in C$, if

$$(\alpha, \beta) \neq (0,0) \quad \text{and} \quad \alpha Bx = \beta Ax.$$

DEFINITION 2. A matrix pair $Z = (A, B)$, $A, B \in C_{n,n}$ is called a *regular pair* if $\det(A + \lambda B) \not\equiv 0$.

It is easy to see that $\operatorname{rank}(A, B) = n$ in this case and that $AA^H + BB^H = ZZ^H$ is a positive definite $n \times n$ matrix. Here, as in the following, we consider Z as an element of $C_{n,2n}$.

DEFINITION 3. A regular matrix pair $Z = (A, B)$ is called *diagonalizable* if there exists a basis of C^n formed from eigenvectors of Z. It is called *normal* if there exists an orthonormal basis of eigenvectors.

For two regular pairs Z, W with eigenvalues (α_i, β_i), and (γ_i, δ_i), respectively, we define the *generalized spectral variation of W with respect to Z* by

$$S_Z(W) = \max_i \min_j \rho\big((\alpha_j, \beta_j), (\gamma_i, \delta_i)\big)$$

and the *generalized eigenvalue variation of Z and W* by

$$v(Z,W) = \min_\pi \max_i \rho\big((\alpha_i, \beta_i), (\gamma_{\pi(i)}, \delta_{\pi(i)})\big),$$

where π runs through all permutations of $\{1, \dots, n\}$.

THEOREM 1.1. *Suppose that $Z=(A,B)$ is a regular pair. Then it is diagonalizable if and only if there exist nonsingular $S,T\in C_{n,n}$ and diagonal matrices $\Lambda=\operatorname{diag}(\alpha_1,\dots,\alpha_n)$, $\Omega=\operatorname{diag}(\beta_1,\dots,\beta_n)$ such that*

$$SAT=\Lambda,\qquad SBT=\Omega. \tag{1.1}$$

Proof. If (1.1) holds, then obviously the ith column t_i of T satisfies $t_i\neq 0$ and $\beta_i At_i=\alpha_i Bt_i$. As Z is regular, $\det(\Lambda+\lambda\Omega)\not\equiv 0$ and hence $(\alpha_i,\beta_i)\neq(0,0)$. This shows that the t_i form a set of n linearly independent eigenvectors, and Z is diagonalizable.

Suppose now that there exist n linearly independent eigenvectors t_i:

$$\alpha_i Bt_i=\beta_i At_i,\quad |\alpha_i|^2+|\beta_i|^2\neq 0,\qquad i=1,\dots,n. \tag{1.2}$$

Then the set of vectors

$$r_i=\begin{Bmatrix}\alpha_i^{-1}At_i & \text{if} & \alpha_i\neq 0,\\ \beta_i^{-1}Bt_i & \text{if} & \beta_i\neq 0,\end{Bmatrix}\qquad i=1,\dots,n \tag{1.3}$$

is well defined. If T and R are the matrices the columns of which are the t_i and the r_i respectively, then

$$AT=R\Lambda,\qquad BT=R\Omega \tag{1.4}$$

hold, as is seen columnwise from (1.2) and (1.3). From $Z=R(\Lambda T^{-1},\Omega T^{-1})$ and rank $Z=n$ we get that R is nonsingular. Setting $S=R^{-1}$ gives (1.1). ∎

Similarly one can prove

THEOREM 1.2. *A regular pair $Z=(A,B)$ is normal if and only if there exist a nonsingular matrix S, a unitary matrix U, and diagonal matrices $\Lambda=\operatorname{diag}(\alpha_i)$, $\Omega=\operatorname{diag}(\beta_i)$ such that*

$$SAU=\Lambda,\qquad SBU=\Omega \tag{1.5}$$

hold.

The following results refer to the metrics d_2, d_F defined in the introduction. Theorem 1.3 has been proved essentially in [11].

THEOREM 1.3. *For given regular matrix pairs Z and W define the $n \times n$ matrix*

$$H(Z,W) = I - (ZZ^H)^{-1/2} ZW^H (WW^H)^{-1} WZ^H (ZZ^H)^{-1/2}.$$

Then H is nonnegative definite and

$$d_2(Z,W) = \|H(Z,W)^{1/2}\|_2, \tag{1.6}$$

$$d_F(Z,W) = \|H(Z,W)^{1/2}\|_F. \tag{1.7}$$

Proof. Replacing Z, W by $(ZZ^H)^{-1/2}Z$ and $(WW^H)^{-1/2}W$ respectively, we may assume that $ZZ^H = WW^H = I$. The key observation is the following: For each number μ^2 the relation

$$(I - ZW^H WZ^H)Z - \mu^2 Z = Z\left[(Z^H Z - W^H W)^2 - \mu^2 I\right] \tag{1.8}$$

holds. If μ is an eigenvalue of $P_Z - P_W$ with eigenvector y, then by (1.8) μ^2 is an eigenvalue of $H(Z,W)$ with eigenvector $Zy \neq 0$. On the other hand, if μ^2 is an eigenvalue of $H(Z,W)$ with eigenvector $v = Zy_0$, then for all y in the span of y_0 and $L_Z^{\perp} = \{x: Zx = 0\}$ we have $HZy - \mu^2 Zy = 0$. Hence by (1.8) $ZKy = 0$, where $K = (Z^H Z - W^H W)^2 - \mu^2 I$. This shows that K maps the $n+1$-dimensional subspace span$(y_0, L_Z^{\perp})$ into $L_Z^{\perp}$.

Hence K is singular, and μ or $-\mu$ is an eigenvalue of $P_Z - P_W$. This implies (1.6).

To prove (1.7), we use the fact that

$$\begin{aligned} 2d_F^2(Z,W) &= \|P_Z - P_W\|_F^2 = \operatorname{tr}(Z^H Z - W^H W)^2 \\ &= \operatorname{tr}(Z^H Z Z^H Z + W^H W W^H W - Z^H Z W^H W - W^H W Z^H Z) \\ &= \operatorname{tr}(ZZ^H) + \operatorname{tr}(WW^H) - 2\operatorname{tr}(ZW^H WZ^H) \\ &= 2(n - \operatorname{tr} ZW^H WZ^H). \end{aligned}$$

On the other hand $\|H(Z,W)^{1/2}\|_F^2 = \operatorname{tr} H(Z,W) = n - \operatorname{tr} ZW^H WZ^H$. ■

Note that it might be easier to calculate $d_{2,F}(Z,W)$ via $H(Z,W)$, as $H \in C_{n,n}$, while $P_Z - P_W \in C_{2n,2n}$. It should also be noted that another proof of (1.6) is possible by using Theorem 6.34, p. 56 in [5] and observing that $\|H(Z,W)\|_2 = \|(I - P_Z)P_W\|_2^2 = \|(I - P_W)P_Z\|_2^2$.

For later use we derive

PROPOSITION 1.4. *Let* $Z=(A,B)$, $W=(C,D)$ *be two regular pairs,* (γ,δ) *an eigenvalue of* W *such that* $|\gamma|^2+|\delta|^2=1$, *and* x *with* $\|x\|=1$ *corresponding eigenvector. Then*

$$\|\delta Ax-\gamma Bx\|\leqslant d_2(Z,W)\cdot\|Z\|_2. \tag{1.9}$$

Proof. For convenience we assume $ZZ^H=WW^H=I$. Then from $\delta Cx-\gamma Dx=0$ we infer

$$\begin{aligned}\delta Ax-\gamma Bx&=\delta(A-ZW^HC)x-\gamma(B-ZW^HD)x\\&=(A-ZW^HC,B-ZW^HD)\begin{pmatrix}\delta x\\-\gamma x\end{pmatrix}=(Z-ZW^HW)\begin{pmatrix}\delta x\\-\gamma x\end{pmatrix}\\&=Z(P_Z-P_W)\begin{pmatrix}\delta x\\-\gamma x\end{pmatrix}\end{aligned}$$

and hence (1.9). ∎

Finally we shall need

PROPOSITION 1.5. *For* $A,C\in C_{n,n}$ *and* $r>0$ *let* $Z_r=(A,rI)$, $W_r=(C,rI)$. *Then*

$$\lim_{r\to\infty} rd_i(Z_r,W_r)=\|A-C\|_i,\qquad i=2,F. \tag{1.10}$$

Proof. By explicit calculation one gets

$$\lim_{r\to\infty} r\left(Z_r^H(Z_rZ_r^H)^{-1}Z_r-W_r^H(W_rW_r^H)^{-1}W_r\right)=\begin{pmatrix}0 & A^H-C^H\\ A-C & 0\end{pmatrix}$$

(1.10) follows now from (0.3), (0.4), and (0.5). ∎

The following result helps to clarify the relation between the metric d_2 and the spectral norm $\|\ \|_2$.

PROPOSITION 1.6. *For regular pairs Z and W*

$$d_2(Z,W) = \min\{\|(ZZ^H)^{-1/2}Z - TW\|_2, T \in C_{n,n}\}. \tag{1.11}$$

Proof. We may assume $ZZ^H = I$. It is well known (see e.g. [5], [8]) that

$$d_2(Z,W) = \|P_Z - P_W\|_2$$
$$= \max\left(\max_{\substack{x \in L_Z \\ \|x\|=1}} \min_{y \in L_W} \|x - y\|, \max_{\substack{y \in L_W \\ \|y\|=1}} \min_{x \in L_Z} \|x - y\| \right).$$

Using the fact that there exists a unitary transformation U mapping L_Z onto L_W and L_W onto L_Z, we see that both maxmin expressions above coincide and hence

$$d_2(Z,W) = \max_{\substack{x \in L_Z \\ \|x\|=1}} \min_{y \in L_W} \|x - y\|$$
$$= \max_{\substack{z \in C_n \\ \|z\|=1}} \min_{w \in C_n} \|Z^T z - W^T w\|.$$

Specifying $w = Tz$, we have $d_2(Z,W) \leqslant \|Z^T - W^T T\|_2$ and

$$d_2(Z,W) \leqslant \inf\{\|Z - TW\|_2, T \in C_{n,n}\}. \tag{1.12}$$

On the other hand, as $\{x \in C_{2n}, \|x\| = 1\} \supset \{Z^H y, y \in C_n, \|y\| = 1\}$,

$$\|Z^H Z - W^H(WW^H)^{-1}W\|_2 \geqslant \max_{\|y\|=1} \{\|[Z^H Z - W^H(WW^H)^{-1}W] Z^H y\|$$
$$= \|Z^H - W^H(WW^H)^{-1}WZ^H\|_2 = \|Z - T_0 W\|_2,$$

where $T_0 = ZW^H(WW^H)^{-1}$. This together with (1.12) shows (1.11). ∎

In particular,

$$d_2(Z,W) \leqslant \|(ZZ^H)^{-1/2}(Z - W)\|_2 \leqslant \|(ZZ^H)^{-1/2}\|_2 \|Z - W\|_2. \tag{1.13}$$

2. THE BAUER-FIKE THEOREM

THEOREM 2.1. *Let* $Z=(A,B)$ *be a diagonalizable pair, and*

$$SAT = \operatorname{diag}(\alpha_i) = \Lambda, \qquad SBT = \operatorname{diag}(\beta_i) = \Omega, \tag{2.1}$$

with $S, T \in C_{n,n}$ *nonsingular. Let* W *be a regular pair. Then*

$$S_Z(W) \leqslant \|T^{-1}\|_2\|T\|_2 d_2(Z,W). \tag{2.2}$$

Proof. We may assume $ZZ^H = I$ and (γ,δ) an eigenvalue of W, $|\gamma|^2+|\delta|^2=1$. Then according to Proposition 1.4 for the normalized eigenvector x of W corresponding to (γ,δ), the inequality

$$\|\delta Ax - \gamma Bx\| \leqslant d_2(Z,W) \tag{2.3}$$

holds. On the other hand, from (2.1) and $ZZ^H = I$, one has $I = AA^H + BB^H = S^{-1}(\Lambda T^{-1}T^{-H}\Lambda^H + \Omega T^{-1}T^{-H}\Omega^H)S^{-H}$, and hence $y^H SS^H y = y^H[\Lambda(T^HT)^{-1}\Lambda^H + \Omega(T^HT)^{-1}\Omega^H]y \leqslant \|(T^HT)^{-1}\|_2 y^H(\Lambda\Lambda^H + \Omega\Omega^H)y$ for all $y \in C_n$, or

$$y^H(SS^H)^{-1}y \geqslant \|T^{-1}\|_2^{-2} y^H(\Lambda\Lambda^H + \Omega\Omega^H)^{-1}y. \tag{2.4}$$

Now

$$\begin{aligned}
&(\delta Ax - \gamma Bx)^H(\delta Ax - \gamma Bx)\\
&\quad = (T^{-1}x)^H(\delta\Lambda - \gamma\Omega)^H S^{-H}S^{-1}(\delta\Lambda - \gamma\Omega)T^{-1}x\\
&\quad \geqslant \|T^{-1}\|_2^{-2}(T^{-1}x)^H(\delta\Lambda - \gamma\Omega)^H(\delta\Lambda - \gamma\Omega)(\Lambda\Lambda^H + \Omega\Omega^H)^{-1}T^{-1}x\\
&\quad \geqslant \|T^{-1}\|_2^{-2}\|T^{-1}x\|^2 \min_i \frac{|\delta\alpha_i - \gamma\beta_i|^2}{|\alpha_i|^2 + |\beta_i|^2}\\
&\quad \geqslant \|T^{-1}\|_2^{-2}\|T\|_2^{-2} \min_i \rho^2((\delta,\gamma),(\alpha_i,\beta_i)).
\end{aligned}$$

This together with (2.3) gives (2.2). ■

In the case of a normal pair Z, the matrix T in (2.1) can be chosen as unitary (see Theorem 1.2); hence we get at once

THEOREM 2.2. *Let $Z=(A,B)$ be a normal pair and W a regular pair. Then*

$$S_Z(W) \leqslant d_2(Z,W). \tag{2.5}$$

Theorem 2.1 can also be applied to the definite generalized eigenvalue problem (see [10], [13], [14]): We call $Z=(A,B)$ a definite pair if A, B are Hermitian and

$$c(A,B)=\min\{|x^H(A+iB)x| : \|x\|=1\} > 0. \tag{2.6}$$

Here the following holds:

THEOREM 2.3. *Let $Z=(A,B)$ be a definite pair and W a regular pair. Then*

$$S_Z(W) \leqslant \frac{\|Z\|_2}{c(A,B)} d_2(Z,W). \tag{2.7}$$

Proof. According to [10, Corollary 2.3] there exists a nonsingular matrix Q such that

$$Q^H A Q = \Lambda = \operatorname{diag}(\alpha_i), \qquad Q^H B Q = \Omega = \operatorname{diag}(\beta_i) \tag{2.8}$$

and $\alpha_i^2+\beta_i^2=1$, $i=1,\ldots,n$. From (2.6) and (2.8) we get

$$\frac{x^H x}{x^H Q^H Q x} \cdot \left| \frac{x^H(\Lambda+i\Omega)x}{x^H x} \right| = \left| \frac{x^H Q^H (A+iB) Q x}{x^H Q^H Q x} \right| \geqslant c(A,B),$$

and hence

$$\frac{\|Qx\|^2}{\|x\|^2} \leqslant \frac{1}{c(A,B)} \left| \frac{x^H(\Lambda+i\Omega)x}{x^H x} \right|,$$

which implies

$$\|Q\|_2 \leqslant c(A,B)^{-1/2}. \tag{2.9}$$

Furthermore, from (2.8) and $\Lambda^2 + \Omega^2 = I$,

$$I = Q^H(AQQ^HA + BQQ^HB)Q,$$

$$Q^{-H}Q^{-1} = AQQ^HA + BQQ^HB,$$

and hence

$$\|Q^{-1}\|_2^2 \leqslant \|Q\|_2^2\|A^2 + B^2\|_2 \leqslant c(A,B)^{-1}\|Z\|_2^2. \tag{2.10}$$

(2.9), (2.10), and (2.2) yield now (2.7). ∎

Let us remark that in [10] and [14] bounds for $v(Z,W)$ are derived, assuming that W is a definite pair, too. It is therefore not possible to compare the two results, even if one uses (1.13).

3. THE HENRICI THEOREM

For easy quotation we formulate Proposition 3.1, the proof of which follows exactly the proof of Henrici's Theorem 0.2 in [3] and is therefore omitted.

PROPOSITION 3.1. *Let $T = \Delta + M$ be a nonsingular upper triangular matrix $\in C_{n,n}$ with diagonal Δ and strictly upper triangular part M. Let $m = \|M\|_2 \neq 0$ and $\|T^{-1}\|_2^{-1} \leqslant r$. Then*

$$\min|t_{ii}| \leqslant \frac{m}{g_n(m/r)}, \tag{3.1}$$

where g_n is the inverse function of $x + x^2 + \cdots + x^n$ for $x \geqslant 0$.

As a measure for the departure of a regular matrix pair $Z = (A,B)$ from normality we introduce the following function:

$$m: C_{n,2n} \to \mathfrak{R}_+ = \{x \in \mathbf{R}, x \geqslant 0\}.$$

Let us denote by $\mathcal{M}_Z$ the set of all matrix pairs (T,U) such that

$$T \text{ is nonsingular}, \qquad U \text{ is unitary}, \tag{3.2a}$$

$$TAU \text{ and } TBU \text{ are upper triangular}, \tag{3.2b}$$

$$|(TAU)_{ii}|^2 + |(TBU)_{ii}|^2 = 1, \qquad i = 1,\dots,n. \tag{3.2c}$$

It has been shown in [7] that $\mathcal{M}_Z \neq \varnothing$. For $(T,U) \in \mathcal{M}_Z$ define

$$\mu(T,U) = \|(TAU - \operatorname{diag}(TAU),\ TBU - \operatorname{diag}(TBU))\|_2.$$

Here we use the notation $\operatorname{diag}(A)$ for the diagonal matrix formed from the diagonal elements of A. We call

$$m(Z) = \min\{\mu(T,U) : (T,U) \in \mathcal{M}_Z\} \tag{3.3}$$

the departure of Z from normality. This is justified, because one can show by using compactness arguments, that there exists $(T_0, U_0) \in \mathcal{M}_Z$ such that

$$\mu(T_0,U_0) \leqslant \mu(T,U) \qquad \forall (T,U) \in \mathcal{M}_Z.$$

It is obvious from (3.2) that $m(Z)$ depends only on L_Z and that $m(Z) = 0$ if and only if Z is a normal pair.

We are now able to formulate the analogue of Henrici's Theorem 0.2:

THEOREM 3.1. *Let Z, W be regular pairs, $m(Z)$ the departure of Z from normality, and $m(Z) \neq 0$. Then*

$$S_Z(W) \leqslant \frac{y}{g_n(y)}[1 + m(Z)]d_2(Z,W), \qquad y = \frac{m(Z)}{[1+m(Z)]d_2(Z,W)}. \tag{3.4}$$

Proof. By eventually multiplying $Z = (A,B)$ from the left by a suitable $n \times n$ matrix, which doesn't affect $S_Z(W)$, $m(Z)$, and y, we may assume

$$A = (D_A + M_A)U, \qquad B = (D_B + M_B)U, \tag{3.5}$$

U unitary, $D_A = \operatorname{diag}(\alpha_i)$, $D_B = \operatorname{diag}(\beta_i)$, $|D_A|^2 + |D_B|^2 = I$, M_A and M_B strictly upper triangular, and $\|(M_A, M_B)\|_2 = m(Z)$. For an eigenvalue (γ, δ) of W,

$|\gamma|^2+|\delta|^2=1$, consider

$$H=\delta(D_A+M_A)-\gamma(D_B+M_B).$$

Then $HU=\delta A-\gamma B$. Obviously it is only necessary to consider the case that (γ,δ) is not an eigenvalue of Z. Then H^{-1} exists and $\|H^{-1}\|_2^{-1}=\|(HU)^{-1}\|_2^{-1}=\|(\delta A-\gamma B)^{-1}\|_2^{-1}\leqslant\|\delta Ax-\gamma Bx\|$ for the normalized eigenvector x of W corresponding to (γ,δ). Using Proposition 1.4, we get

$$\begin{aligned}\|H^{-1}\|_2^{-1}\leqslant\|(A,B)\|_2 d_2(Z,W)&\leqslant\left[\|(D_A,D_B)\|_2+\|(M_A,M_B)\|_2\right]d_2(Z,W)\\&=\left[1+m(Z)\right]d_2(Z,W).\end{aligned}\tag{3.6}$$

Applying now Proposition 3.1 to H, we get

$$\min|\delta\alpha_i-\gamma\beta_i|\leqslant\frac{\mu}{g_n(\mu/r)},\tag{3.7}$$

where $\mu=\|\delta M_A-\gamma M_B\|_2\leqslant m(Z)$, $r=(1+m(Z))d_2(Z,W)$. Observing that the (α_i,β_i) are the normalized eigenvalues of Z, we get from (3.7) the desired result (3.4). ∎

It is well known (see [3]) that the right-hand side of (3.1) tends to r if m tends to zero. Hence we get Theorem 2.2 from (3.4) for $m(Z)\to 0$.

4. THE HOFFMAN-WIELANDT THEOREM

THEOREM 4.1. *Let Z, W be two normal pairs with eigenvalues (α_i,β_i) and (γ_i,δ_i), $i=1,\dots,n$, respectively. There exists a permutation π of $\{1,\dots,n\}$ such that*

$$v(Z,W)\leqslant\left(\sum_{i=1}^{n}\rho^2\big((\alpha_i,\beta_i),(\gamma_{\pi(i)},\delta_{\pi(i)})\big)\right)^{1/2}\leqslant d_F(Z,W).\tag{4.1}$$

Proof. Using the notation $Z=(A,B)$, $W=(C,D)$,

$$SAU_1=\Lambda=\operatorname{diag}(\alpha_i),\qquad SBU_1=\Omega=\operatorname{diag}(\beta_i),\qquad \Lambda\Lambda^H+\Omega\Omega^H=I$$

$$TCU_2=\Gamma=\operatorname{diag}(\gamma_i),\qquad TDU_2=\Delta=\operatorname{diag}(\delta_i),\qquad \Gamma\Gamma^H+\Delta\Delta^H=I$$

where S, T nonsingular, U_1, U_2 unitary, we get after some manipulations

$$\begin{aligned}
P_Z - P_W &= Z^H(ZZ^H)^{-1}Z - W^H(WW^H)^{-1}W \\
&= \begin{pmatrix} U_1 & 0 \\ 0 & U_1 \end{pmatrix}\left[\begin{pmatrix} \Lambda^H\Lambda & \Lambda^H\Omega \\ \Omega^H\Lambda & \Omega^H\Omega \end{pmatrix}\right. \\
&\quad \left. - \begin{pmatrix} V & 0 \\ 0 & V \end{pmatrix}\begin{pmatrix} \Gamma^H\Gamma & \Gamma^H\Delta \\ \Delta^H\Gamma & \Delta^H\Delta \end{pmatrix}\begin{pmatrix} V^H & 0 \\ 0 & V^H \end{pmatrix}\right]\begin{pmatrix} U_1^H & 0 \\ 0 & U_1^H \end{pmatrix} \\
&= \tilde{U}(\tilde{\Lambda} - \tilde{V}\tilde{\Gamma}\tilde{V}^H)\tilde{U}^H, \qquad \text{where} \quad V = U_1^H U_2,
\end{aligned} \tag{4.2}$$

and the definitions of $\tilde{U}, \tilde{\Lambda}, \tilde{V}, \tilde{\Gamma}$ are according to the preceding expression, e.g.

$$\tilde{U} = \begin{pmatrix} U & 0 \\ 0 & U \end{pmatrix}, \qquad \tilde{\Lambda} = \begin{pmatrix} \Lambda^H\Lambda & \Lambda^H\Omega \\ \Omega^H\Lambda & \Omega^H\Omega \end{pmatrix}, \qquad \text{etc.}$$

Hence, as $\tilde{U}$ is unitary,

$$\begin{aligned}
2d_F^2(Z,W) &= \|P_Z - P_W\|_F^2 = \|\tilde{\Lambda} - \tilde{V}\tilde{\Gamma}\tilde{V}^H\|_F^2 \\
&= \operatorname{tr}(\tilde{\Lambda} - \tilde{V}\tilde{\Gamma}\tilde{V}^H)(\tilde{\Lambda}^H - \tilde{V}\tilde{\Gamma}^H\tilde{V}^H) \\
&= \operatorname{tr}(\tilde{\Lambda}\tilde{\Lambda}^H + \tilde{V}\tilde{\Gamma}\tilde{\Gamma}^H\tilde{V}^H - \tilde{V}\tilde{\Gamma}\tilde{V}^H\tilde{\Lambda}^H - \tilde{\Lambda}\tilde{V}\tilde{\Gamma}^H\tilde{V}^H) = f(V).
\end{aligned} \tag{4.3}$$

Here we remark that for the case $V = P =$ permutation matrix we have, after some calculations,

$$\begin{aligned}
f(P) &= \|\tilde{\Lambda} - \tilde{P}\tilde{\Gamma}\tilde{P}^H\|_F^2 = \left\|\begin{pmatrix} \Lambda^H\Lambda - P^T\Gamma^H\Gamma P & \Lambda^H\Omega - P^T\Lambda\Gamma^H P \\ \Omega^H\Lambda - P^T\Lambda^H\Gamma P & \Omega^H\Omega - P^T\Delta^H\Delta P \end{pmatrix}\right\|_F^2 \\
&= \sum_{i=1}^{n} \left\|\begin{pmatrix} \alpha_i\bar{\alpha}_i - \gamma_{\pi(i)}\bar{\gamma}_{\pi(i)} & \bar{\alpha}_i\beta_i - \bar{\gamma}_{\pi(i)}\delta_{\pi(i)} \\ \alpha_i\bar{\beta}_i - \gamma_{\pi(i)}\bar{\delta}_{\pi(i)} & \beta_i\bar{\beta}_i - \delta_{\pi(i)}\bar{\delta}_{\pi(i)} \end{pmatrix}\right\|_F^2 \\
&= 2\sum_{i=1}^{n} |\alpha_i\delta_{\pi(i)} - \beta_i\gamma_{\pi(i)}|^2 \\
&= 2\sum_{i=1}^{n} \rho^2\big((\alpha_i, \beta_i), (\gamma_{\pi(i)}, \delta_{\pi(i)})\big).
\end{aligned} \tag{4.4}$$

Here $\pi(i)$ is defined by P via $P^T\operatorname{diag}(t_j)P = \operatorname{diag}(t_{\pi(i)})$. It is easy to see that the diagonal entries of $(\tilde{\Lambda} - \tilde{V}\tilde{\Gamma}\tilde{V}^H)(\tilde{\Lambda}^H - \tilde{V}\tilde{\Gamma}^H\tilde{V}^H)$ depend linearly on the numbers $|v_{ik}|^2$, $i, k = 1,\dots,n$, where $V = (v_{ik})$. Defining $\hat{W} = (\hat{w}_{ik})$ by $\hat{w}_{ik} = |v_{ik}|^2$, $\hat{W}$ is a bistochastic matrix and we have

$$f(V) = l(\hat{W}) \tag{4.5}$$

with a linear functional l. As the bistochastic matrices form a convex polyhedron the vertices of which are the permutation matrices, there exists a permutation matrix P such that $l(\hat{W}) \geqslant l(P)$. Hence

$$2d_F^2(Z,W) = f(V) = l(\hat{W}) \geqslant l(P) = f(P).$$

Using (4.4) we get the second inequality in (4.1). As the first one is evident, the proof is completed. ∎

5. FINAL REMARKS

Despite of the great formal similarity, the "classical" Theorems 0.1, 0.2 and 0.3 are not special cases of the corresponding results for the generalized eigenvalue problem, namely Theorems 2.1, 3.1, and 4.1. It is, however, possible to derive the former theorems from the letter ones by a limiting procedure.

We treat first Theorem 0.1. If A is diagonalizable, $T^{-1}AT = \Lambda = \operatorname{diag}(\lambda_i)$, then the matrix pair $A_r = (A, rI)$ is diagonalizable, too, with eigenvalues (λ_i, r). If μ is an eigenvalue of C, then (μ, r) is an eigenvalue of $C_r = (C, rI)$. From (2.2) we get

$$\min_i \left| \frac{r\lambda_i - r\mu}{\sqrt{|\mu|^2 + |r|^2}\sqrt{|\lambda_i|^2 + |r|^2}} \right| \leqslant \|T\|_2 \|T^{-1}\|_2 \, d_2(A_r, C_r). \tag{5.1}$$

Multiplying (5.1) by r, using Proposition 1.5 and considering $r \to \infty$ gives

$$\min_i |\lambda_i - \mu| \leqslant \|A - C\|_2 \|T\|_2 \|T^{-1}\|_2$$

i.e. Theorem 0.1.

Similarly we get Theorem 0.2 from Theorem 3.1. Here we have to use the fact that if Δ is the departure from normality of the matrix $A \in C_{n,n}$, then

$$m(A, rI) \leqslant \frac{\Delta}{r}.$$

Using this inequality in (3.4) leads to Theorem 0.2.

When deriving the classical Hoffman-Wielandt Theorem 0.3 from (4.1), one has to use the fact that there is a fixed permutation π and a sequence $\{r_j\} \to \infty$ such that

$$\sum_{i=1}^{n} \rho^2\big((\lambda_i, r_j), (\mu_{\pi(i)}, r_j)\big) \leqslant d_F^2(A_{r_j}, C_{r_j}).$$

Again, multiplying by r_j, considering $j \to \infty$, and using Proposition 1.5 for the Euclidean norm gives Theorem 0.3.

It should be mentioned that this limiting procedure was used previously by Stewart [10].

REFERENCES

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

2 L. Elsner, Einschließungssätze für Eigenwerte nichtnormaler Matrizen. Funktionalanalysis, Approximationstheorie, Numerische Mathematik, ISNM 7, Birkhäuser, Basel, 1967, pp. 185–193.

3 P. Henrici, Bounds for iterates, inverses, spectralvariation and fields of values of non-normal matrices, *Numer. Math.* 4:24–39 (1962).

4 A. J. Hoffman and H. W. Wielandt, The variation of the spectrum of a normal matrix, *Duke Math. J.* 20:37–39 (1953).

5 T. Kato, *Perturbation Theory for Linear Operators*, Springer, New York, 1966.

6 Y. Matushima, *Differentiable Manifolds*, New York, 1972.

7 G. W. Stewart, On the sensitivity of the eigenvalue problem $Ax = \lambda Bx$, *SIAM J. Numer. Anal.* 9:669–689 (1972).

8 G. W. Stewart, Error and perturbation bounds for subspaces associated with certain eigenvalue problems, *SIAM Rev.* 15:727–764 (1973).

9 G. W. Stewart, Gershgorin theory for the generalized eigenvalue problem $Ax = \lambda Bx$, *Math. Comp.* 29:600–606 (1975).

10 G. W. Stewart, Perturbation bounds for the definite generalized eigenvalue problem, *Linear Algebra Appl.* 23:69–85 (1979).

11 Ji-guang Sun, Invariant subspaces and generalized invariant subspaces, (II). Perturbation theorems, *Math. Numer. Sinica* 2:113–123 (1980).

12 Ji-guang Sun, A theorem on the perturbation of generalized eigenvalues, Report at the Conference on Numerical Mathematics, Guangzhou, China, Dec. 1979.
13 Ji-guang Sun, The perturbation bounds for eigenspaces of definite matrix pairs. I. The $\sin\theta$ theorems, II. The $\sin 2\theta$ theorems, submitted for publication.
14 Ji-guang Sun, A note on Stewart's theorem for definite matrix pairs, to appear.

Received 31 December 1981