NORTH-HOLLAND

Real Perturbation Values and Real Quadratic Forms in a Complex Vector Space

Bo Bernhardsson and Anders Rantzer*
Department of Automatic Control
Lund Institute of Technology
Lund, Sweden

and

Li Qiu
Department of Electrical Engineering
Hong Kong University of Science and Technology
Hong Kong

Submitted by Roger A. Horn

ABSTRACT

A sequence of real numbers connected to a complex matrix is introduced. It is shown how these real perturbation values can be computed and that they have several properties similar to the singular values. The so-called real pseudospectra and real stability radii can be computed using the real perturbation values. The main result concerns the signature of real quadratic forms in complex vector spaces. © 1998 Elsevier Science Inc.

1. INTRODUCTION

For a linear transformation between the complexifications of two finite dimensional Euclidean spaces we introduce in Section 3 two sequences of

*The first two authors gratefully acknowledge support from contract M-MA 06513-305 of the Swedish Natural Science Research Council.

LINEAR ALGEBRA AND ITS APPLICATIONS 270:131–154 (1998)

0024-3795/98/$19.00
PII S0024-3795(97)00232-2

numbers, which we call *real perturbation values*, by modifying the usual definition of singular values in a way that takes the real structure into account. These definitions were motivated by the so-called *real stability radius problem* in control theory (see [3] and [11]) and in the computation of *real pseudospectra* in numerical analysis (see [13]). The main point turns out to be a result, proved in Section 4, on the signature of a quadratic form in a complex vector space, which may be of interest in other contexts as well. In an earlier version [2] of this paper the proof was based on the fairly complicated normal forms for pairs of Hermitian and complex symmetric matrices proved in [4] (see also [1, 6–9]). Following a suggestion by Lars Hörmander, we now use only normal forms for generic pairs. To make the presentation self-contained we give a short complete derivation of them, obtained together with him.

A flaw of the real perturbation values is that they are not continuous functions everywhere. The continuity properties are discussed in Section 5. Proposition 5.2 is joint work with Lars Hörmander.

2. SINGULAR VALUES

As a preliminary and to introduce notation we present the basic facts on singular values that lie behind the definition of real perturbation values and are needed for their study. This section can be ignored by readers familiar with the singular value decomposition and the rank approximation properties of singular values such as presented in e.g. [10].

Let H_1 and H_2 be two finite dimensional Hilbert spaces, and let $T : H_1 \to H_2$ be a linear map. In this section it does not matter if the scalars are real or complex. The operator $T^*T : H_1 \to H_1$ is then nonnegative and self-adjoint with rank equal to the rank r of T. Let $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r > 0$ be the positive eigenvalues of $(T^*T)^{1/2}$, and let $\varphi_1, \dots, \varphi_r$ be orthonormal eigenvectors with $T^*T\varphi_j = \sigma_j^2\varphi_j$. Then $\psi_j = T\varphi_j/\sigma_j$ are also orthonormal, and

$$
\begin{aligned}
T\varphi &= \sum_1^r \sigma_j(\varphi, \varphi_j)_{H_1}\psi_j, \qquad \varphi \in H_1, \\
T^*\psi &= \sum_1^r \sigma_j(\psi, \psi_j)_{H_2}\varphi_j, \qquad \psi \in H_2.
\end{aligned}
\tag{2.1}
$$

Thus the *singular values* $\sigma_j(T)$ of T are the same as those of T^*. We define $\sigma_j = 0$ when $j > r$. The maximum minimum principle for T^*T gives

$$\sigma_j(T) = \inf_{\operatorname{codim} W < j} \sup_{0 \neq \varphi \in W} \frac{\|T\varphi\|_{H_2}}{\|\varphi\|_{H_1}}, \tag{2.2}$$

$$\sigma_j(T) = \sup_{\dim W \geqslant j} \inf_{0 \neq \varphi \in W} \frac{\|T\varphi\|_{H_2}}{\|\varphi\|_{H_1}}. \tag{2.3}$$

From either (2.2) or (2.3) it follows at once that for every j

$$|\sigma_j(T_1) - \sigma_j(T_2)| \leqslant \|T_1 - T_2\| = \sigma_1(T_1 - T_2), \qquad T_1, T_2 \in \mathscr{L}(H_1, H_2). \tag{2.4}$$

More generally, it follows from (2.2) that

$$\sigma_j(T_1) \leqslant \sigma_k(T_2) + \sigma_l(T_1 - T_2), \qquad \text{if} \quad k + l = j + 1. \tag{2.5}$$

We can rewrite (2.2) in the form

$$\sigma_j(T) = \inf_{\operatorname{rank} S < j} \|T - S\|, \tag{2.2'}$$

for if $W = \operatorname{Ker} S$ then $\operatorname{codim} W < j$ and $\|T - S\|$ is at least equal to the norm of the restriction to W, hence $\|T - S\| \geqslant \sigma_j(T)$. There is equality when $S = PT$, where P is the orthogonal projection in H_2 on the space spanned by $\psi_1, \dots, \psi_{j-1}$; for $T - S$ is then obtained by dropping the first $j - 1$ terms in (2.1). Equivalently,

$$\sigma_j(T) = \inf\{\|\Delta\|;\ \Delta \in \mathscr{L}(H_1, H_2),\ \operatorname{rank}(T - \Delta) < j\}. \tag{2.2''}$$

This follows by just writing $\Delta = T - S$ in (2.2′). A similar formula follows from (2.3),

$$\sigma_j(T) = \left(\inf\{\|\Delta\|;\ \Delta \in \mathscr{L}(H_2, H_1),\ \dim \operatorname{Ker}(\operatorname{Id}_{H_1} - \Delta T) \geqslant j\}\right)^{-1}. \tag{2.3'}$$

If $\sigma_j(T) = 0$, i.e., rank $T < j$, then $\operatorname{rank}(\Delta T) < j$, so $\dim \operatorname{Ker}(\mathrm{Id}_{H_1} - \Delta T) < j$. The infimum in (2.3′) should then be interpreted as $+\infty$, and the reciprocal as 0. To prove (2.3′) we observe that if the kernel W of $S = \mathrm{Id}_{H_1} - \Delta T$ has dimension $\geqslant j$, then

$$\|\varphi\|_{H_1} = \|\Delta T\varphi\|_{H_1} \leqslant \|\Delta\| \, \|T\varphi\|_{H_2}, \qquad \varphi \in W,$$

so $\sigma_j(T) \geqslant 1/\|\Delta\|$ by (2.3). On the other hand, if we define $\Delta\psi_k = \varphi_k/\sigma_k(T)$, $k = 1,\ldots,j$, and $\Delta\psi = 0$ in the orthogonal space, then $\varphi_k - \Delta T\varphi_k = 0$, $k = 1,\ldots,j$, and $\|\Delta\| = 1/\sigma_j(T)$, so $\operatorname{rank}(\mathrm{Id}_{H_1} - \Delta T) \leqslant \dim H_1 - j$ and $\|\Delta\| = 1/\sigma_j(T)$; thus $\inf\|\Delta\| = 1/\sigma_j(T)$ as claimed in (2.3′).

In (2.3′) we may replace $\mathrm{Id}_{H_1} - \Delta T$ by $\mathrm{Id}_{H_2} - T\Delta$, for

$$\begin{gathered} \dim \operatorname{Ker}\left(\mathrm{Id}_{H_1} - ST\right) = \dim \operatorname{Ker}\left(\mathrm{Id}_{H_2} - TS\right), \\ T \in \mathscr{L}(H_1, H_2), \quad S \in \mathscr{L}(H_2, H_1). \end{gathered} \tag{2.6}$$

In fact, both sides are equal to the dimension of the kernel of

$$H_1 \oplus H_2 \ni (\varphi, \psi) \to (\varphi + S\psi, \psi + T\varphi) \in H_1 \oplus H_2,$$

which projects injectively to H_1 and H_2 with the kernels in (2.6) as range.

Another proof of (2.3′) follows from the following elementary lemma, which will be useful for later reference.

Lemma 2.1. *Given linear transformations T_j: $H_0 \to H_j$, $j = 1, 2$, there exists a contraction $\Delta : H_1 \to H_2$ such that $\Delta T_1 = T_2$ if and only if $T_2^* T_2 \leqslant T_1^* T_1$.*

Proof. The necessity is obvious, for

$$\|T_2\varphi\|_{H_2} = \|\Delta T_1\varphi\|_{H_2} \leqslant \|T_1\varphi\|_{H_1}, \qquad \varphi \in H_0,$$

if such a Δ exists. Conversely, if $\|T_2\varphi\|_{H_2} \leqslant \|T_1\varphi\|_{H_1}$, $\varphi \in H_0$, then $T_1\varphi \mapsto T_2\varphi$ is a contraction defined on the range of T_1. It remains a contraction if it is extended to vanish on the orthogonal complement. ■

Let us now see how the lemma gives (2.3′). That a positive number σ is no greater than the number defined in (2.3′) means that $\dim \operatorname{Ker}(\mathrm{Id}_{H_1} - \Delta T)$

$\geqslant j$ for some $\Delta \in \mathscr{L}(H_2, H_1)$ with $\|\Delta\| \leqslant 1/\sigma$, that is, for some such Δ and some S with rank $\geqslant j$ we have $S - \Delta TS = 0$, that is, $\sigma \Delta TS = \sigma S$. By the lemma this is equivalent to $\sigma^2 S^*S \leqslant S^*T^*TS$, or equivalently $\|T\varphi\|_{H_2} \geqslant \sigma\|\varphi\|_{H_1}$ for all φ in the range of S, that is, a space of dimension $\geqslant j$. By (2.3) this is equivalent to $\sigma_j(T) \geqslant \sigma$, which proves (2.3′).

For a historical survey of singular values and rank approximation theorems see [12].

3. THE REAL PERTURBATION VALUES

We assume now that H_1 and H_2 are given as complexifications of real Hilbert spaces h_1 and h_2; thus $H_j = h_j \otimes_{\mathbf{R}} \mathbf{C}$. Then the set $\mathscr{L}(H_1, H_2)$ of linear transformations from H_1 to H_2 has a real linear subspace $\mathscr{L}^r(H_1, H_2)$ consisting of extensions of maps in $\mathscr{L}(h_1, h_2)$, and every $T \in \mathscr{L}(H_1, H_2)$ has a unique decomposition $T = \operatorname{Re} T + i \operatorname{Im} T$ with $\operatorname{Re} T$ and $\operatorname{Im} T \in \mathscr{L}^r(H_1, H_2)$. The same holds for $\mathscr{L}(H_2, H_1)$ and $\mathscr{L}^r(H_2, H_1)$. By analogy with (2.3′) and (2.2″) we introduce

$$\tau_k(T) = \left[\inf\left\{\|\Delta\|;\ \Delta \in \mathscr{L}^r(H_2, H_1),\ \dim \operatorname{Ker}\left(\operatorname{Id}_{H_1} - \Delta T\right) \geqslant k\right\}\right]^{-1}, \tag{3.1}$$

$$\tilde{\tau}_k(T) = \inf\{\|\Delta\|;\ \Delta \in \mathscr{L}^r(H_1, H_2),\ \operatorname{rank}(T - \Delta) < k\}. \tag{3.2}$$

In case there is no Δ with the required property we interpret the infimum as $+\infty$, which makes $\tau_k(T) = 0$ and $\tilde{\tau}_k(T) = +\infty$. By (2.6) the condition on Δ in (3.1) may be replaced by $\dim \operatorname{Ker}(\operatorname{Id}_{H_2} - T\Delta) \geqslant k$, which shows that $\tau_k(T^*) = \tau_k(T)$; it is obvious that $\tilde{\tau}_k(T^*) = \tilde{\tau}_k(T)$. It is also obvious that $\tau_k(T) \leqslant \sigma_k(T) \leqslant \tilde{\tau}_k(T)$.

The following theorem gives an approach to computing the real perturbation values defined by (3.1) and (3.2).

THEOREM 3.1. *With the preceding definitions we have*

$$\tau_k(T) = \inf_{\gamma \in (0,1]} \sigma_{2k}\left(\tilde{T}_\gamma\right), \tag{3.3}$$

$$\tilde{\tau}_k(T) = \sup_{\gamma \in (0,1]} \sigma_{2k-1}\left(\tilde{T}_\gamma\right), \tag{3.4}$$

where

$$\tilde{T}_\gamma = \begin{pmatrix} \operatorname{Re} T & -\gamma \operatorname{Im} T \\ \gamma^{-1} \operatorname{Im} T & \operatorname{Re} T \end{pmatrix} : h_1 \oplus h_1 \to h_2 \oplus h_2 .$$

The first step in the proof is a variant of Lemma 2.1:

LEMMA 3.2. *Let H_j, $j = 0, 1, 2$, be complexifications of real finite dimensional Hilbert spaces h_j. Given linear transformations $T_j : H_0 \to H_j$, $j = 1, 2$, there exists a contraction $\Delta \in \mathscr{L}^r(H_1, H_2)$ such that $\Delta T_1 = T_2$ if and only if, with block matrix notation,*

$$\begin{pmatrix} \operatorname{Re} T_2 & \operatorname{Im} T_2 \end{pmatrix}^* \begin{pmatrix} \operatorname{Re} T_2 & \operatorname{Im} T_2 \end{pmatrix} \leqslant \begin{pmatrix} \operatorname{Re} T_1 & \operatorname{Im} T_1 \end{pmatrix}^* \begin{pmatrix} \operatorname{Re} T_1 & \operatorname{Im} T_1 \end{pmatrix}, \tag{3.5}$$

or equivalently

$$\begin{pmatrix} T_2 & \bar{T}_2 \end{pmatrix}^* \begin{pmatrix} T_2 & \bar{T}_2 \end{pmatrix} \leqslant \begin{pmatrix} T_1 & \bar{T}_1 \end{pmatrix}^* \begin{pmatrix} T_1 & \bar{T}_1 \end{pmatrix}. \tag{3.5'}$$

Proof. Write the equation $\Delta T_1 = T_2$ as $\Delta(\operatorname{Re} T_1 \ \operatorname{Im} T_1) = (\operatorname{Re} T_2 \ \operatorname{Im} T_2)$; here $(\operatorname{Re} T_j \ \operatorname{Im} T_j)$ is a map from $h_0 \oplus h_0$ to h_j. By Lemma 2.1 (for spaces over the reals) we conclude that (3.5) is a necessary and sufficient condition for the existence of a contraction Δ. Since

$$\begin{pmatrix} T_j & \bar{T}_j \end{pmatrix} = \begin{pmatrix} \operatorname{Re} T_j & \operatorname{Im} T_j \end{pmatrix} \begin{pmatrix} \mathrm{Id}_{H_0} & \mathrm{Id}_{H_0} \\ i\,\mathrm{Id}_{H_0} & -i\,\mathrm{Id}_{H_0} \end{pmatrix},$$

the extension of (3.5) to the complexification is equivalent to (3.5′). ∎

Proof of Theorem 3.1. This proof relies on Theorem 4.1, which is stated and proved in Section 4. We shall first prove (3.3) and indicate afterwards the modifications required to prove (3.4). Let $0 < \tau \leqslant \tau_k(T)$. By (3.1) this means that one can find $S \in \mathscr{L}(H_1, H_1)$ of rank $\geqslant k$ and $\Delta \in \mathscr{L}^r(H_2, H_1)$ with $\|\Delta\| \leqslant \tau^{-1}$ such that $(\mathrm{Id}_{H_1} - \Delta T)S = 0$, that is, $\tau \Delta T S = \tau S$. By Lemma 3.2 this means precisely that

$$\tau^2 \begin{pmatrix} S & \bar{S} \end{pmatrix}^* \begin{pmatrix} S & \bar{S} \end{pmatrix} \leqslant \begin{pmatrix} TS & \overline{TS} \end{pmatrix}^* \begin{pmatrix} TS & \overline{TS} \end{pmatrix}. \tag{3.6}$$

The product on the right-hand side is the operator

$$\begin{pmatrix} S^*T^*TS & S^*T^*\overline{TS} \\ \bar{S}^*\bar{T}^*TS & \bar{S}^*\bar{T}^*\overline{TS} \end{pmatrix} : H_1 \oplus H_1 \to H_1 \oplus H_1,$$

and replacing T by τ gives the operator in the left-hand side. If we set

$$A_\tau = T^*T - \tau^2 \, \mathrm{Id}_{H_1}, \qquad B_\tau = \bar{T}^*T - \tau^2 \, \mathrm{Id}_{H_1}, \tag{3.7}$$

then (3.6) can be written

$$\begin{pmatrix} S^*A_\tau S & S^*\bar{B}_\tau \bar{S} \\ \bar{S}^*B_\tau S & \bar{S}^*\bar{A}_\tau \bar{S} \end{pmatrix} \geqslant 0, \tag{3.6$'$}$$

or

$$\begin{pmatrix} S & 0 \\ 0 & \bar{S} \end{pmatrix}^* \begin{pmatrix} A_\tau & \bar{B}_\tau \\ B_\tau & \bar{A}_\tau \end{pmatrix} \begin{pmatrix} S & 0 \\ 0 & \bar{S} \end{pmatrix} \geqslant 0. \tag{3.6$''$}$$

Thus $0 < \tau \leqslant \tau_k(T)$ is equivalent to the existence of $S \in \mathscr{L}(H_1, H_1)$ of rank $\geqslant k$ such that (3.6″) is valid. By the equivalence of conditions (i) and (iv) in Theorem 4.1, to be proved later, we therefore conclude that

$$0 < \tau \leqslant \tau_k(T) \quad \Leftrightarrow$$

$$\begin{pmatrix} A_\tau & \bar{\beta}\bar{B}_\tau \\ \beta B_\tau & \bar{A}_\tau \end{pmatrix} \text{ has at least } 2k \text{ nonnegative eigenvalues if } |\beta| \leqslant 1. \tag{3.8}$$

Here it is not really important to allow complex values for β, for multiplication of β by a complex number of absolute value 1 gives a unitarily equivalent operator. It is therefore enough to take $\beta \in [-1, 0]$.

Next we prove that the condition in (3.8) is equivalent to

$$\sigma_{2k}(\tilde{T}_\gamma) \geqslant \tau, \qquad 0 < \gamma \leqslant 1. \tag{3.9}$$

First we observe that

$$\tilde{T}_\gamma = D_\gamma \begin{pmatrix} T & 0 \\ 0 & \bar{T} \end{pmatrix} E_\gamma, \qquad D_\gamma = \begin{pmatrix} i\gamma\,\mathrm{Id}_{H_2} & i\gamma\,\mathrm{Id}_{H_2} \\ \mathrm{Id}_{H_2} & -\mathrm{Id}_{H_2} \end{pmatrix},$$

$$E_\gamma = \frac{1}{2} \begin{pmatrix} \mathrm{Id}_{H_1}/i\gamma & \mathrm{Id}_{H_1} \\ \mathrm{Id}_{H_1}/i\gamma & -\mathrm{Id}_{H_1} \end{pmatrix}.$$

Equation (3.9) states that $\tilde{T}_\gamma^* \tilde{T}_\gamma - \tau^2\,\mathrm{Id}_{H_1 \oplus H_1}$ has at least $2k$ positive eigenvalues. After right and left multiplication by the inverse of E_γ and its adjoint this means that

$$\begin{pmatrix} T & 0 \\ 0 & \bar{T} \end{pmatrix}^* D_\gamma^* D_\gamma \begin{pmatrix} T & 0 \\ 0 & \bar{T} \end{pmatrix} - \tau^2 (E_\gamma E_\gamma^*)^{-1}$$

has at least $2k$ nonnegative eigenvalues. Here

$$D_\gamma^* D_\gamma = \begin{pmatrix} (\gamma^2+1)\,\mathrm{Id}_{H_2} & (\gamma^2-1)\,\mathrm{Id}_{H_2} \\ (\gamma^2-1)\,\mathrm{Id}_{H_2} & (\gamma^2+1)\,\mathrm{Id}_{H_2} \end{pmatrix},$$

$$(E_\gamma E_\gamma^*)^{-1} = \begin{pmatrix} (\gamma^2+1)\,\mathrm{Id}_{H_1} & (\gamma^2-1)\,\mathrm{Id}_{H_1} \\ (\gamma^2-1)\,\mathrm{Id}_{H_1} & (\gamma^2+1)\,\mathrm{Id}_{H_1} \end{pmatrix}.$$

If we divide by $\gamma^2 + 1$ and put $\beta = (\gamma^2 - 1)/(\gamma^2 + 1)$, the operator becomes

$$\begin{pmatrix} T^* & 0 \\ 0 & \bar{T}^* \end{pmatrix} \begin{pmatrix} \mathrm{Id}_{H_2} & \beta\,\mathrm{Id}_{H_2} \\ \beta\,\mathrm{Id}_{H_2} & \mathrm{Id}_{H_2} \end{pmatrix} \begin{pmatrix} T & 0 \\ 0 & \bar{T} \end{pmatrix} - \tau^2 \begin{pmatrix} \mathrm{Id}_{H_1} & \beta\,\mathrm{Id}_{H_1} \\ \beta\,\mathrm{Id}_{H_1} & \mathrm{Id}_{H_1} \end{pmatrix}$$

$$= \begin{pmatrix} A_\tau & \beta \bar{B}_\tau \\ \beta B_\tau & \bar{A}_\tau \end{pmatrix}, \tag{3.10}$$

which proves the equivalence of (3.8) and (3.9) and completes the proof of (3.3), apart from the proof of Theorem 4.1.

To prove (3.4) we first recall that by the definition of $\tilde{\tau}_k$ we have $\tau \geqslant \tilde{\tau}_k(T)$ if and only if $\operatorname{rank}(T - \Delta) < k$ for some $\Delta \in \mathscr{L}^r(H_1, H_2)$ with $\|\Delta\| \leqslant \tau$. The rank condition means that there is some $S \in \mathscr{L}(H_1, H_1)$ with $\operatorname{rank} S \geqslant \dim H_1 - (k-1)$ such that $(T - \Delta)S = 0$, that is, $(\Delta/\tau)S = TS/\tau$. By Lemma 3.2 this is equivalent to

$$\tau^2\begin{pmatrix} S & \bar{S} \end{pmatrix}^*\begin{pmatrix} S & \bar{S} \end{pmatrix} \geqslant \begin{pmatrix} TS & \overline{TS} \end{pmatrix}^*\begin{pmatrix} TS & \overline{TS} \end{pmatrix}. \qquad (3.11)$$

The calculations that proved the equivalence of (3.6) and (3.6″) show that (3.11) is equivalent to

$$\begin{pmatrix} S & 0 \\ 0 & \bar{S} \end{pmatrix}^* \begin{pmatrix} A_\tau & \bar{B}_\tau \\ B_\tau & \bar{A}_\tau \end{pmatrix} \begin{pmatrix} S & 0 \\ 0 & \bar{S} \end{pmatrix} \leqslant 0. \qquad (3.12)$$

Using Theorem 4.1 as before, we conclude that

$$\tau \geqslant \tilde{\tau}_k(T) \quad \Leftrightarrow$$

$$\begin{pmatrix} A_\tau & \bar{\beta}\bar{B}_\tau \\ \beta B_\tau & \bar{A}_\tau \end{pmatrix} \text{ has at least } 2[\dim H_1 - (k-1)]$$

$$\text{nonpositive eigenvalues if } |\beta| \leqslant 1. \qquad (3.13)$$

The proof of the equivalence of (3.8) and (3.9) shows that (3.13) means precisely that $\tilde{T}_\gamma^* \tilde{T}_\gamma - \tau^2 \operatorname{Id}_{H_1 \oplus H_1}$ has at least $2 \dim H_1 - [(2k-1) - 1]$ nonpositive eigenvalues, which by (2.2) means that $\sigma_{2k-1}(\tilde{T}_\gamma) \leqslant \tau$. The proof of (3.4) and Theorem 3.1 is now complete apart from the proof of Theorem 4.1. ■

The proof of (3.3) also gives another characterization of $\tau_k(T)$, for we saw that $\tau \leqslant \tau_k(T)$ was equivalent to (3.6″), which, by Theorem 4.1 (iii), is equivalent to

$$(A_\tau \varphi, \varphi)_{H_1} + \operatorname{Re}\langle B_\tau \varphi, \varphi\rangle_{H_1} \geqslant 0, \qquad \varphi \in W,$$

where W is a complex subspace of H_1, of dimension $\geqslant k$. Explicitly this means that

$$(T\varphi, T\varphi)_{H_2} + \operatorname{Re}\langle T\varphi, T\varphi\rangle_{H_2} - \tau^2(\varphi, \varphi)_{H_1} - \tau^2 \operatorname{Re}\langle \varphi, \varphi\rangle_{H_1} \geqslant 0,$$

$$\varphi \in W.$$

Since $(\varphi, \varphi)_{H_1} + \mathrm{Re}\langle \varphi, \varphi\rangle_{H_1} = 2\|\mathrm{Re}\ \varphi\|^2_{H_1}$ and since there is an analogous identity in H_2, this means that $\|\mathrm{Re}(T\varphi)\|^2_{H_2} \geqslant \tau^2\|\mathrm{Re}\ \varphi\|^2_{H_1}$. Hence

$$\tau_k(T) = \sup_{\dim W \geqslant k} \inf_{\varphi \in W,\ \mathrm{Re}\ \varphi \neq 0} \frac{\|\mathrm{Re}(T\varphi)\|_{H_2}}{\|\mathrm{Re}\ \varphi\|_{H_1}}, \tag{3.1$'$}$$

where W is a *complex* subspace of H_1. This is a close analogue of (2.3).

We get a similar conclusion from the proof of (3.4), for it shows that $\tau \geqslant \tilde{\tau}_k(T)$ is equivalent to $\|\mathrm{Re}(T\varphi)\|^2_{H_2} \leqslant \tau^2\|\mathrm{Re}\ \varphi\|^2_{H_1}$ for every φ in a complex subspace W of H with $\operatorname{codim} W < k$. Hence we obtain an analogue of (2.2),

$$\tilde{\tau}_k(T) = \inf_{\operatorname{codim} W < k} \sup_{\varphi \in W,\ \mathrm{Re}\ \varphi \neq 0} \frac{\|\mathrm{Re}(T\varphi)\|_{H_2}}{\|\mathrm{Re}\ \varphi\|_{H_1}}, \tag{3.2$'$}$$

where W is a complex subspace of H_1.

4. REAL QUADRATIC FORMS IN A COMPLEX VECTOR SPACE

Let H be a finite dimensional complex vector space, and let Q be a real quadratic form in the underlying real vector space. There is a unique decomposition $Q = Q_0 + Q_1$ where Q_j are quadratic forms with $Q_j(iz) = (-1)^j Q_j(z)$; it is given by

$$Q_j(z) = \tfrac{1}{2}\left[Q(z) + (-1)^j Q(iz)\right], \qquad z \in H, \quad j = 0, 1.$$

The form Q_0 can be polarized to a Hermitian symmetric sesquilinear form $(z, w) \mapsto Q_0(z, w)$ which is linear in z and antilinear in w, $Q_0(z, z) = Q_0(z)$, and $Q_1(z) = \mathrm{Re}\ q(z)$ where q is a quadratic form with respect to the complex structure in H,

$$q(z) = Q_1(z) - iQ_1(\varepsilon z) = \frac{1}{2}\sum_0^3 \frac{Q(\varepsilon^j z)}{\varepsilon^{2j}}, \qquad \varepsilon = e^{\pi i/4}.$$

We can polarize q to a symmetric bilinear form $(z, w) \mapsto q(z, w)$ such that $q(z, z) = q(z)$.

If we identify H with the complexification of a real Hilbert space h, for example by introducing complex coordinates $z_1, \ldots, z_n$ identifying H with $\mathbf{C}^n$, then

$$Q_0(z,w) = (Az,w), \qquad \mathrm{Re}\, q(z,w) = \mathrm{Re}(Bz,\overline{w}) = \mathrm{Re}\langle Bz,w\rangle,$$

where $A^* = A$ and $B^T = B$. This notation is essential in conditions (i), (ii), (iv), (v) of the following theorem, while the others are expressed only in terms of $Q_0(\cdot,\cdot)$ and $q(\cdot,\cdot)$. The following theorem is a generalization of Theorem 2.1 in [5].

THEOREM 4.1. *The following conditions are equivalent*:

(i) *There exists a map* $S \in \mathscr{L}(H,H)$ *of rank* $\geqslant k$ *such that*

$$\begin{pmatrix} S & 0 \\ 0 & \overline{S} \end{pmatrix}^* \begin{pmatrix} A & \overline{B} \\ B & \overline{A} \end{pmatrix} \begin{pmatrix} S & 0 \\ 0 & \overline{S} \end{pmatrix} \geqslant 0. \tag{4.1}$$

(ii) *There exists a complex linear subspace* W *of* H *of dimension* $\geqslant k$ *such that*

$$(A\varphi,\varphi) + \left(\overline{B\overline{\psi}},\varphi\right) + \left(B\varphi,\overline{\psi}\right) + \left(\overline{A\overline{\psi}},\overline{\psi}\right) \geqslant 0, \qquad \varphi,\psi \in W, \tag{4.2}$$

or equivalently

$$(A\varphi,\varphi) + (A\psi,\psi) + 2\,\mathrm{Re}\langle B\varphi,\psi\rangle \geqslant 0, \qquad \varphi,\psi \in W. \tag{4.3}$$

(iii) *There exists a complex linear subspace* W *of* H *of dimension* k *such that*

$$(A\varphi,\varphi) + \mathrm{Re}\langle B\varphi,\varphi\rangle \geqslant 0, \qquad \varphi \in W, \tag{4.4}$$

or equivalently

$$|\langle B\varphi,\varphi\rangle| \leqslant (A\varphi,\varphi), \qquad \varphi \in W. \tag{4.5}$$

(iv) *The Hermitian operator*

$$\begin{pmatrix} A & \overline{\beta B} \\ \beta B & \overline{A} \end{pmatrix}$$

in $H \oplus H$ *has at least* $2k$ *nonnegative eigenvalues for every* $\beta \in \mathbf{C}$ *with* $|\beta| \leqslant 1$, *that is, the Hermitian form*

$$(A\varphi, \varphi) + \bar{\beta}(\bar{B}\psi, \varphi) + \beta(B\varphi, \psi) + (\bar{A}\psi, \psi), \qquad \varphi, \psi \in H, \quad (4.6)$$

has at least $2k$ *nonnegative eigenvalues when* $|\beta| \leqslant 1$.

(v) *The form* (4.6) *has at least* $2k$ *nonnegative eigenvalues when* $\beta \in [0, 1]$.

(vi) *The quadratic form*

$$(A\varphi, \varphi) + (A\psi, \psi) + 2\beta \operatorname{Re}\langle B\varphi, \psi\rangle, \qquad \varphi, \psi \in H, \qquad (4.7)$$

in $H \oplus H$ *considered as a real vector space has at least* $4k$ *nonnegative eigenvalues when* $\beta \in [0, 1]$.

(vii) *The quadratic form*

$$(A\varphi, \varphi) + \beta \operatorname{Re}\langle B\varphi, \varphi\rangle, \qquad \varphi \in H, \qquad (4.8)$$

in H *considered as a real vector space has at least* $2k$ *nonnegative eigenvalues if* $\beta \in [0, 1]$.

Proof. Let us first note a number of fairly trivial implications:

$$\text{(i)} \Leftrightarrow \text{(ii)} \Leftrightarrow \text{(iii)} \Rightarrow \text{(iv)} \Leftrightarrow \text{(v)} \Rightarrow \text{(vi)}, \qquad \text{(iii)} \Rightarrow \text{(vii)}.$$

Condition (ii) is just condition (i) with W equal to the range of S, and (4.3) implies (4.4) when we take $\varphi = \psi$. If we replace φ by $e^{i\theta}\varphi$ in (4.4), $\theta \in \mathbf{R}$, then (4.5) follows. From (4.5) we obtain $|\langle B(\varphi \pm \psi), \varphi \pm \psi\rangle| \leqslant (A(\varphi \pm \psi), \varphi \pm \psi)$, $\varphi, \psi \in W$, which implies $4|\langle B\varphi, \psi\rangle| \leqslant 2[A(\varphi, \varphi) + A(\psi, \psi)]$ if $\varphi, \psi \in W$, and proves (4.3) and (ii). From (ii) it follows that the form (4.6) with $\beta = 1$ is nonnegative when $\varphi \in W$ and $\bar{\psi} \in W$. Replacing ψ by $\beta\psi$, we conclude that the form (4.6) is also nonnegative for such φ, ψ when $|\beta| = 1$, and hence by convexity when $|\beta| \leqslant 1$, which proves (iv). That (iv) implies (v) is obvious, and the converse follows if ψ is replaced by $e^{i\theta}\psi$, $\theta \in \mathbf{R}$. As a quadratic form in $H \oplus H$ as a real vector space, the form (4.6) then has $\geqslant 4k$ nonnegative eigenvalues, which proves (vi). In the same way it is obvious that (iii) implies (vii).

The essential contents of the theorem are therefore the implications

$$\text{(vi)} \Rightarrow \text{(ii)} \quad \text{and} \quad \text{(vii)} \Rightarrow \text{(ii)}. \qquad (4.9)$$

When proving them we may strengthen the hypotheses in (vi) and (vii) to assuming that there are $4k$ and $2k$ strictly positive eigenvalues, respectively; for this can be achieved by adding a small multiple of the identity to A. Then the hypotheses remain valid after a small perturbation of A and B, so it will be sufficient to study the generic case (see Lemma 4.2); for the set of all A, B for which (ii) holds is closed by the compactness of the set of subspaces of fixed dimension. We shall postpone the end of the proof of Theorem 4.1 until we have derived normal forms in the generic situation. ∎

In terms of complex coordinates $(z_1, \dots, z_n)$ in H we can write

$$Q_0(z) = \sum_{j,k=1}^{n} a_{jk}\bar{z}_j z_k, \quad a_{jk} = \bar{a}_{kj};$$

$$Q_1(z) = \operatorname{Re} \sum_{j,k=1}^{n} b_{jk} z_j z_k, \quad b_{jk} = b_{kj}.$$

Passing to new coordinates z' with $z_j = \sum_1^n T_{jk} z'_k$, we get for the corresponding matrices

$$A' = T^*AT, \qquad B' = T^tBT,$$

where T^t is the transpose of T and $T^* = \bar{T}^t$. If B is invertible it follows that

$$C' = \bar{T}^{-1}CT, \qquad \text{where} \quad C = \bar{B}^{-1}A, \quad C' = \bar{B}'^{-1}A'. \tag{4.10}$$

This implies that

$$\bar{C}'C' = T^{-1}\bar{C}CT, \tag{4.11}$$

which means that $\bar{C}C$ is the matrix of a complex linear transformation in H which is independent of the choice of coordinates.

LEMMA 4.2. *For a dense set of real quadratic forms Q in H, the matrices A and B are invertible and all eigenvalues of $\bar{C}C$ are simple.*

Proof. The matrices A and B are invertible if $\det A \det B \neq 0$, which is true on a dense set. The entries of $\bar{C}C$ are polynomials in the entries of the real and complex parts of B^{-1} and A, so the coefficients of $p(\lambda) = \det(\lambda \operatorname{Id}$

$-\overline{C}C) = \det(\lambda \operatorname{Id} - \overline{C}'C')$ are polynomials in them, and so is the discriminant of $p(\lambda)$. The eigenvalues are simple if the discriminant is nonzero. Now either the discriminant can be made nonzero by small perturbations in A and B, keeping $A^* = A$ and $B^t = B$, or else it is identically zero for all such A and B. However, it does not vanish identically, for if A and B are diagonal then $\det(\lambda \operatorname{Id} - \overline{C}C) = \prod(\lambda - |a_{jj}/b_{jj}|^2)$, so the discriminant is nonzero if $|a_{jj}/b_{jj}| \neq |a_{kk}/b_{kk}|$ when $j \neq k$. ■

The following lemma shows a normal form for a generic real quadratic form. The generic case is sufficient for our presentation, and a proof of Lemma 4.3 is included to make the presentation self-contained. For a complete treatment of the more difficult general case see [4, 6–9]. See also Chapter 4.6 in [10].

LEMMA 4.3. *If A and B are invertible and the eigenvalues of $\overline{C}C$ are simple, then the real eigenvalues are positive, the others occur in complex conjugate pairs, and the coordinates can be chosen so that*

$$Q(z) = \sum_1^r \left(\lambda_j |z_j|^2 + \operatorname{Re} z_j^2\right)$$

$$+ \sum_{j=r+1}^{r+s} \left[\lambda_j z_{2j-r-1}\bar{z}_{2j-r} + \bar{\lambda}_j z_{2j-r}\bar{z}_{2j-r-1}\right.$$

$$\left. + \operatorname{Re}\left(z_{2j-r-1}^2 + z_{2j-r}^2\right)\right]. \qquad (4.12)$$

Here λ_j^2, $j = 1, \dots, r$, are the positive eigenvalues of $\overline{C}C$, and λ_j^2, $j = r + 1, \dots, r + s$, are the eigenvalues of $\overline{C}C$ with positive imaginary part. The first (second) sum is to be omitted if $r = 0$ (if $s = 0$).

Proof. Let $z \neq 0$ be an eigenvector of $\overline{C}C$ with eigenvalue μ; thus $\overline{C}Cz = \mu z$. Then $C\overline{C}(Cz) = \mu Cz$; hence $\overline{C}\overline{C}\overline{z} = \bar{\mu}\overline{Cz}$, so $\overline{Cz}$ is an eigenvector with eigenvalue $\bar{\mu}$.

(i) If μ is real, then $\overline{Cz}$ must be a multiple of z; thus $\overline{Cz} = \lambda z$ for some $\lambda \in \mathbf{C}$, and $\overline{\lambda C z} = \mu z$, which implies $\mu = |\lambda|^2 > 0$. Since $Az = \overline{\lambda B z}$, we have

$$Q_0(z) = (Az, z) = \bar{\lambda}\,\overline{\langle Bz, z\rangle}. \qquad (4.13)$$

(ii) If $\operatorname{Im} \mu \neq 0$, then $\overline{Cz}$ is an eigenvector belonging to the eigenvalue $\bar{\mu}$. Let $\lambda^2 = \mu$, and set $\lambda \bar{w} = Cz$. Then also $\overline{Cw} = \lambda z$; thus

$$Az = \lambda \overline{Bw} \quad \text{and} \quad Aw = \overline{\lambda Bz}\,. \tag{4.14}$$

(iii) Let $\bar{C}Cz_j = \mu_j z_j$ and $\bar{C}Cz_k = \mu_k z_k$. Then

$$\mu_j\langle Bz_j, z_k\rangle = \langle \overline{AB}^{-1}Az_j, z_k\rangle = \langle z_j, \overline{AB}^{-1}Az_k\rangle$$
$$= \mu_k\langle z_j, Bz_k\rangle = \mu_k\langle Bz_j, z_k\rangle,$$

which proves that $\langle Bz_j, z_k\rangle = 0$ when $\mu_j \neq \mu_k$. Similarly,

$$\mu_j\big(Az_j, z_k\big) = \big(AB^{-1}\overline{AB}^{-1}Az_j, z_k\big) = \big(z_j, AB^{-1}\overline{AB}^{-1}Az_k\big) = \bar{\mu}_k\big(Az_j, z_k\big),$$

which proves that $(Az_j, z_k) = 0$ when $\mu_j \neq \bar{\mu}_k$. Thus the eigenvectors corresponding to real eigenvalues and the two dimensional spaces spanned by eigenvectors corresponding to complex conjugate eigenvalues of $\bar{C}C$ are mutually orthogonal with respect to the sesquilinear scalar product $(Az, w) = (z, Aw)$ and with respect to the bilinear scalar product $\langle Bz, w\rangle = \langle z, Bw\rangle$. It is therefore sufficient to examine the structure of these two kinds of spaces.

In case (i) above it follows now from the nondegeneracy of B that $\langle Bz, z\rangle \neq 0$. Replacing z by a multiple of z, we can then attain $\langle Bz, z\rangle = 1$, which by (4.13) implies that $(Az, z) = \bar{\lambda}$. Hence λ is real with $\lambda^2 = \mu$.

In case (ii) above we have $\langle z, Bw\rangle = 0$; hence $(z, Az) = 0$ and $(Aw, w) = 0$. Moreover,

$$\lambda\langle Bz, z\rangle = (z, Aw) = (Az, w) = \lambda\,\overline{\langle Bw, w\rangle}\,.$$

Since B is nondegenerate, we conclude that $\langle Bz, z\rangle \neq 0$, and we can normalize so that $\langle Bz, z\rangle = 1$, whence $\langle Bw, w\rangle = 1$, which implies $(Az, w) = \lambda$. In the basis w, z the matrices of the A and B therefore take the form

$$\begin{pmatrix} 0 & \lambda \\ \bar{\lambda} & 0 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

which completes the proof of the lemma. ∎

LEMMA 4.4. *If* $\operatorname{Im} \lambda > 0$, *then the quadratic form*

$$\lambda z_1 \bar{z}_2 + \bar{\lambda} z_2 \bar{z}_1 + \operatorname{Re}\left[\beta\left(z_1^2 + z_2^2\right)\right], \qquad z \in \mathbf{C}^2 \cong \mathbf{R}^4,$$

is positive definite in the subspace where $z_2 = iz_1$ *and negative definite in the subspace where* $z_2 = -iz_1$. *Thus the signature is* $2, 2$ *for arbitrary* $\beta \in \mathbf{C}$.

The proof is obvious.

End of proof of Theorem 4.1. What remains is to prove the implications (4.9) when B is the unit matrix and with $\operatorname{Im} \lambda_j > 0$ for $j = r+1, \dots, r+s$:

$$(Az, z) = \sum_1^r \lambda_j |z_j|^2 + \sum_{j=r+1}^{r+s} \left(\lambda_j z_{2j-r-1} \bar{z}_{2j-r} + \bar{\lambda}_j z_{2j-r} \bar{z}_{2j-r-1}\right). \quad (4.15)$$

In view of Lemma 4.4, hypothesis (vi) [or (vii)] remains valid if we restrict to the complex linear subspace where $z_{r+2j} = iz_{r+2j-1}$ for $j = 1, \dots, s$, and since the second sum in (4.15) is positive there, we need to prove the theorem only when

$$(Az, z) = \sum_1^r \lambda_j |z_j|^2, \qquad \langle Bz, z\rangle = \sum_1^r z_j^2, \qquad (4.16)$$

where $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_r$. We shall now prove the theorem in this case.

Explicitly the quadratic form (4.7) in $H \oplus H$, as a real vector space, is

$$\sum_{j=1}^r \Big\{\lambda_j\left[(\operatorname{Re} z_j)^2 + (\operatorname{Im} z_j)^2 + (\operatorname{Re} w_j)^2 + (\operatorname{Im} w_j)^2\right]$$
$$+2\beta(\operatorname{Re} z_j \operatorname{Re} w_j - \operatorname{Im} z_j \operatorname{Im} w_j)\Big\},$$

where each term has the eigenvalues $\lambda_j \pm \beta$ taken twice. The quadratic form (4.8) in H, as a real vector space, can be written

$$\sum_{j=1}^r \left\{\lambda_j\left[(\operatorname{Re} z_j)^2 + (\operatorname{Im} z_j)^2\right] + \beta\left[(\operatorname{Re} z_j)^2 - (\operatorname{Im} z_j)^2\right]\right\},$$

where each term has the eigenvalues $\lambda_j + \beta$. Both conditions (vi) and (vii) therefore mean that at least $2k$ of the eigenvalues $\lambda_j \pm \beta$ are nonnegative for every $\beta \in [0,1]$.

To make this condition explicit we let $\lambda_1, \dots, \lambda_l$ be the eigenvalues of A that are greater or equal to 1; for them we have $\lambda_j \pm \beta \geqslant 0$ when $\beta \in [0,1]$, which accounts for $2l$ nonnegative eigenvalues. Eigenvalues $\lambda \in [0,1)$ will always contribute an eigenvalue $\lambda + \beta \geqslant 0$, but the eigenvalue $\lambda - \beta$ becomes negative when $\beta > \lambda$. On the other hand, eigenvalues $\lambda \in [-1, 0)$ can contribute a nonnegative eigenvalue only when $\beta \geqslant -\lambda$. When $\beta = 0$ we must have $\lambda_1, \dots, \lambda_k \geqslant 0$. If $k > l$, then disappearing eigenvalues $\lambda_{k+1-\nu} - \beta$, $\nu = 1, \dots, k-l$, must be compensated by eigenvalues $\lambda_{k+\nu} + \beta$ that appear at least as early, that is,

$$\lambda_{k+1-\nu} \geqslant -\lambda_{k+\nu}, \qquad \nu = 1, \dots, k-l. \tag{4.17}$$

(Thus $2k - l \leqslant r$.) Since

$$\lambda_j\left(|z_j|^2 + |w_j|^2\right) + 2\,\mathrm{Re}\, z_j w_j \geqslant 0 \qquad \text{if} \quad \lambda_j \geqslant 1,$$

we need to examine only the case of pairs of eigenvalues with nonnegative sum as in (4.17). Simplifying notation, this means that we must examine

$$(Az, z) = \lambda_1 |z_1|^2 + \lambda_2 |z_2|^2, \qquad \langle Bz, w \rangle = z_1 w_1 + z_2 w_2$$

where $\lambda_1 + \lambda_2 \geqslant 0$. The condition (4.3) becomes

$$\lambda_1\left(|z_1|^2 + |w_1|^2\right) + \lambda_2\left(|z_2|^2 + |w_2|^2\right) + 2\,\mathrm{Re}(z_1 w_1 + z_2 w_2) \geqslant 0,$$
$$z, w \in W,$$

where W is a complex line in $\mathbf{C}^2$. This is true if $W = \{(z_1, z_2) \in \mathbf{C}^2;\ z_1 = iz_2\}$, since $\lambda_1 + \lambda_2 \geqslant 0$. The proof of Theorem 4.1 is now complete. ■

5. CONTINUITY PROPERTIES

By (2.4) the singular values $\sigma_k(T)$ are Lipschitz continuous functions of T, but the real perturbation values are not. That is caused by the fact that in (3.3) and (3.4) the infimum and supremum are taken over a noncompact set of parameter values γ, so it is only clear that τ_k is upper semicontinuous and that $\tilde{\tau}_k$ is lower semicontinuous. Although it follows at once from (2.4) that

$$|\tau_k(T) - \tau_k(T+E)| \leqslant \|E\| \quad \text{if} \quad E \in \mathscr{L}^r(H_1, H_2), \tag{5.1}$$

$$|\tilde{\tau}_k(T) - \tilde{\tau}_k(T+E)| \leqslant \|E\| \quad \text{if} \quad E \in \mathscr{L}^r(H_1, H_2), \tag{5.2}$$

the continuity properties with respect to the imaginary part of T are quite delicate. [When $\tilde{\tau}_k(T) = +\infty$, then (5.2) is only supposed to mean that $\tilde{\tau}_k(T+E) = +\infty$ too.] We will discuss only the continuity properties of $\tau_k(T)$.

We first study the limit when $\gamma \to 0$ of the singular values in (3.3) and (3.4). The following proposition is a special case of Lemma 5 in [11]. To make the presentation self-contained, we include a proof.

PROPOSITION 5.1. *Let*

$$\tilde{T}_\gamma = \begin{pmatrix} T_1 & -\gamma T_2 \\ \gamma^{-1} T_2 & T_1 \end{pmatrix} \in \mathscr{L}(h_1 \oplus h_1, h_2 \oplus h_2)$$

as in Theorem 3.1. *Then it follows when* $\gamma \to 0$ *that*

$$\sigma_j(\tilde{T}_\gamma) \to \begin{cases} \infty & \textit{if} \quad j \leqslant \operatorname{rank} T_2, \\ \|\hat{T}\| & \textit{if} \quad j = \operatorname{rank} T_2 + 1. \end{cases} \tag{5.3}$$

Here $\hat{T} = T_1|_{\operatorname{Ker} T_2} \oplus PT_1$, *where* P *is the orthogonal projection* $h_2 \to \operatorname{Ker} T_2^*$, *so* $(PT_1)^* = T_1^* P$ *has the same singular values as the restriction* $T_1^*|_{\operatorname{Ker} T_2^*}$.

Proof. When $j \leqslant \operatorname{rank} T_2$ the result follows from the fact that $\sigma_j(\gamma \tilde{T}_\gamma) \to \sigma_j(T_2) > 0$ as $\gamma \to 0$. Assume therefore that $\operatorname{rank} T_2 = j - 1$. Choose a linear map $G : h_1 \to (\operatorname{Ker} T_2)^\perp$ such that $T_2 G\varphi + T_1\varphi = PT_1\varphi$ for all $\varphi \in h_1$. The subspaces $W = \operatorname{Ker} T_2 \oplus h_1$ and $\{\varphi = (\varphi_1 + \gamma G\varphi_2, \varphi_2); (\varphi_1, \varphi_2) \in W\}$ both have codimension $j-1$. Hence we get from (2.2)

$$\begin{aligned} \sigma_j(\tilde{T}_\gamma) &\leqslant \sup_{0 \neq \varphi \in W} \frac{\|\tilde{T}_\gamma(\varphi_1 + \gamma G\varphi_2, \varphi_2)\|}{\|(\varphi_1 + \gamma G\varphi_2, \varphi_2)\|} \\ &= \sup_{0 \neq \varphi \in W} \frac{\|(T_1(\varphi_1 + \gamma G\varphi_2) - \gamma T_2\varphi_2, PT_1\varphi_2)\|}{\|(\varphi_1 + \gamma G\varphi_2, \varphi_2)\|} \\ &\to \sup_{0 \neq \varphi \in W} \frac{\|(T_1\varphi_1, PT\varphi_2)\|}{\|\varphi\|} = \|\hat{T}\| \quad \text{when} \quad \gamma \to 0. \end{aligned}$$

If $\operatorname{Ker} T_2 \neq \emptyset$, then we can find $\varphi_1 \in \operatorname{Ker} T_2$ such that $\|T_1\varphi_1\| = \|T_1|_{\operatorname{Ker} T_2}\|$ and $\|\varphi_1\| = 1$. The subspace $V = \{\varphi_0 + c\varphi_1;\ \varphi_0 \in (\operatorname{Ker} T_2)^{\perp}\}$ of h_1 has dimension j. Hence we get from (2.3)

$$\sigma_j\big(\tilde{T}_\gamma\big) \geqslant \inf_{\varphi\in V;\ \|\varphi\|=1} \tilde{T}_\gamma(\varphi, 0) \geqslant \inf_{\varphi\in V;\ \|\varphi\|=1} \big\|\big(T_1(\varphi_0 + c\varphi_1), \gamma^{-1}T_2\varphi_0\big)\big\|$$

$$\geqslant \inf_{\varphi\in V;\ \|\varphi\|=1} \max\big(c\|T_1\varphi_1\| - \|T_1\|\,\|\varphi_0\|, \gamma^{-1}\sigma_{j-1}(T_2)\|\varphi_0\|\big).$$

Since $c = (1 - \|\varphi_0\|^2)^{1/2} \geqslant 1 - \|\varphi_0\|$, we have

$$\sigma_j\big(\tilde{T}_\gamma\big) \geqslant \inf_{\|\varphi_0\|\leqslant 1} \max\big(\|T_1|_{\operatorname{Ker} T_2}\| - 2\|T_1\|\,\|\varphi_0\|, \gamma^{-1}\sigma_{j-1}(T_2)\|\varphi_0\|\big)$$

$$\geqslant \frac{\|T_1|_{\operatorname{Ker} T_2}\|}{1 + 2\|T_1\|\gamma/\sigma_{j-1}(T_2)} \to \|T_1|_{\operatorname{Ker} T_2}\| \qquad \text{when} \quad \gamma \to 0.$$

This bound if obvious if $\operatorname{Ker} T_2 = \emptyset$, and by applying it to T^* we get $\underline{\lim}_{\gamma\to 0}\ \sigma_j(\tilde{T}_\gamma) \geqslant \|\hat{T}\|$. ■

PROPOSITION 5.2. *$\tau_k(T)$ is continuous at $T = T_1 + iT_2$ when* rank $T_2 \geqslant 2k - 1$.

Proof. Let $S = T + E$ where $\operatorname{rank} T_2 \geqslant 2k - 1$ and $E \in \mathscr{L}(H_1, H_2)$. Since $T \to \tau_k(T)$ is upper semicontinuous, it is sufficient to find a good lower estimate of $\sigma_{2k}(\tilde{S}_\gamma)$ when E is small. We immediately get $\sigma_{2k}(\tilde{S}_\gamma) \geqslant \sigma_{2k}(\tilde{T}_\gamma) - \|\tilde{E}_\gamma\|$. We have $\|\gamma\tilde{E}_\gamma\| \leqslant \|E\|$ for $0 < \gamma \leqslant 1$, for if $\varphi, \psi \in h_1$ and $\|\varphi\|^2 + \|\psi\|^2 = 1$ then

$$\big\|\gamma\tilde{E}_\gamma(\varphi, \psi)\big\|^2 = \|\gamma E_1\varphi - \gamma^2 E_2\psi\|^2 + \|E_2\varphi + \gamma E_1\psi\|^2$$

$$\leqslant \|E_1\varphi - E_2(\gamma\psi)\|^2 + \|E_2\varphi + E_1(\gamma\psi)\|^2$$

$$\leqslant \|\tilde{E}_1\|^2\big(\|\varphi\|^2 + \|\gamma\psi\|^2\big) \leqslant \|E\|^2.$$

We therefore have

$$\sigma_{2k}\left(\tilde{S}_\gamma\right) \geqslant \sigma_{2k}\left(\tilde{T}_\gamma\right) - \frac{\|E\|}{\gamma}. \tag{5.4}$$

If $\sigma_{2k}(T_2) > 0$, then $\sigma_{2k}(S_2) > 0$ if $\|E\| < \sigma_{2k}(T_2)$. Hence $\sigma_{2k}(\tilde{S}_\gamma) \to \infty$ when $\gamma \to 0$. The infimum $\tau_k(T+E)$ is therefore attained for some $\gamma_0 \in (0, 1]$. We have $\gamma_0^{-1}\sigma_{2k}(S_2) \leqslant \sigma_{2k}(\tilde{S}_{\gamma_0}) = \tau_k(T+E)$, since $\gamma_0^{-1}S_2$ is obtained from $\tilde{S}_{\gamma_0}$ by a restriction followed by a projection in the range. This gives

$$\tau_k(T+E) = \sigma_{2k}\left(\tilde{S}_{\gamma_0}\right) \geqslant \tau_k(T) - \frac{\|E\|\tau_k(T+E)}{\sigma_{2k}(S_2)},$$

which after rearranging and using $\sigma_{2k}(S_2) \geqslant \sigma_{2k}(T_2) - \|E\| > 0$ gives

$$\tau_k(T+E) \geqslant \tau_k(T) - \frac{\|E\|\tau_k(T)}{\sigma_{2k}(T_2)}.$$

This proves continuity if $\operatorname{rank} T_2 \geqslant 2k$.

Now assume that $\operatorname{rank} T_2 = 2k-1$. It is necessary to improve the lower bound (5.4) for small $\gamma > 0$. Put $a = \sigma_{2k}(\tilde{S}_\gamma)$. From (2.2) there exists a subspace W with $\operatorname{codim} W = 2k-1$ such that

$$\|S_1\varphi_1 - \gamma S_2\varphi_2\|^2 + \|\gamma^{-1}S_2\varphi_1 + S_1\varphi_2\|^2 \leqslant a^2\|\varphi\|^2,$$
$$\varphi = (\varphi_1, \varphi_2) \in W. \tag{5.5}$$

This gives

$$\|S_2\varphi_1\| \leqslant \gamma(\|S_2\| + a)\|\varphi\|, \qquad \varphi \in W,$$

and hence

$$\|T_2\varphi_1\| \leqslant \left[\gamma(\|S_2\| + a) + \|E\|\right]\|\varphi\|, \qquad \varphi \in W.$$

Now let $\varphi_1 = \varphi_{10} + \varphi_{11}$, where φ_{11} is the component of φ_1 orthogonal to $\operatorname{Ker} T_2$. Since $\sigma_{2k-1}(T_2)$ is the smallest nonzero singular value of T_2, we have

$$\|\varphi_{11}\| \leqslant \frac{\|T_2\varphi_1\|}{\sigma_{2k-1}(T_2)} \leqslant \delta\|\varphi\|, \qquad \varphi \in W, \tag{5.6}$$

where $\delta = [\gamma(\|S_2\| + a) + \|E\|]/\sigma_{2k-1}(T_2)$. When $\delta < 1$ the map

$$W \ni \varphi \to (\varphi_{10}, \varphi_2) \in \operatorname{Ker} T_2 \oplus h_1$$

is invertible, since the dimensions of W and $\operatorname{Ker} T_2 \oplus h_1$ are equal and since

$$\|\varphi\| \leqslant \|(\varphi_{10}, \varphi_2)\| + \delta\|\varphi\|,$$

which gives

$$\|\varphi\| \leqslant \frac{\|(\varphi_{10}, \varphi_2)\|}{1-\delta}. \tag{5.7}$$

If we take $\varphi_2 = 0$ then (5.5) gives

$$\|T_1\varphi_1\| \leqslant (a + \|E\|)\|\varphi_1\|, \qquad \varphi \in W,$$

and with (5.6) and (5.7) we therefore have

$$\|T_1\varphi_{10}\| \leqslant \frac{(a + \|E\| + \delta\|T_1\|)\|\varphi_{10}\|}{1-\delta}, \qquad \varphi_{10} \in \operatorname{Ker} T_2.$$

This gives a lower bound of a in terms of $\|T_1|_{\operatorname{Ker} T_2}\|$. Similar calculations for T^* give the same estimate with $\|T_1^*|_{\operatorname{Ker} T_2^*}\|$. With $\hat{T}$ defined as in Proposition 5.1 we therefore get

$$(1-\delta)\|\hat{T}\| \leqslant a + \|E\| + \delta\|T_1\|.$$

We conclude that

$$a \geqslant (1-\hat{\delta})\|\hat{T}\| - \|E\| - \hat{\delta}\|T_1\|,$$

with $\hat{\delta} = [\gamma(\|S_2\| + \|\hat{T}\|) + \|E\|]/\sigma_{2k-1}(T_2)$. This bound is obvious when $\hat{\delta} \geqslant 1$ or $a \geqslant \|\hat{T}\|$ and has been proved in the other case. With the bound (5.4) we obtain

$$\begin{aligned} &\inf_{\gamma \in (0,1]} \sigma_{2k}\big(\tilde{S}_\gamma\big) \\ &\quad \geqslant \inf_{\gamma \in (0,1]} \sigma_{2k}\big(\tilde{T}_\gamma\big) - \sup_{\gamma} \min\Big(\|E\|/\gamma,\ \hat{\delta}\big(\|T_1\| + \|\hat{T}\|\big) + \|E\|\Big). \end{aligned}$$

The minimum can be bounded from above with

$$c(T, E) = \left(\frac{2\|E\|\big(\|T_2 + E\| + \|\hat{T}\|\big)\|T_1\|}{\sigma_{2k-1}(T_2)} \right)^{1/2} + \frac{2\|T_1\|\,\|E\|}{\sigma_{2k-1}(T_2)} + \|E\|,$$

which gives

$$\tau_k(T + E) \geqslant \tau_k(T) - c(T, E), \tag{5.8}$$

where $c(T, E) \to 0$ when $E \to 0$. This proves lower semicontinuity and hence continuity. ■

REMARK 1. Note that lower estimates of the form (5.8) can be transformed into upper estimates of the form

$$\tau_k(T + E) \leqslant \tau_k(T) + c(T + E, -E).$$

It is easy to see that the proof gives Lipschitz continuity when rank $T_2 \geqslant 2k$.

REMARK 2. It is easy to check that with

$$T = \begin{pmatrix} i\,\mathrm{Id}_{2k-2} & 0 \\ 0 & 1 + i\varepsilon \end{pmatrix}$$

we have $\tau_k(T) = 1$ if $\varepsilon = 0$ and $\tau_k(T) = 0$ if $\varepsilon \neq 0$. This gives an example of discontinuity with rank $T_2 = 2k - 2$.

6. CONCLUDING REMARKS

In numerical analysis it is of interest to compute so-called pseudospectra (also called spectral value sets); see [13]. For a given $\varepsilon > 0$ the ε-pseudospectrum of a matrix $A \in \mathbf{R}^{n \times n}$ is a region of the complex plane defined as

$$\mathrm{sp}_\varepsilon(A) = \bigcup_{\substack{E \in \mathbf{C}^{n \times n} \\ \|E\| < \varepsilon}} \sigma(A + E),$$

where $\sigma(A + E)$ denotes the spectrum of $A + E$. From (2.3′) and (2.2″) it follows that if complex perturbations E are allowed, then

$$\mathrm{sp}_\varepsilon(A) = \left\{z;\ \sigma_1\left((z\,\mathrm{Id} - A)^{-1}\right) > \varepsilon^{-1}\right\} = \left\{z;\ \sigma_n(z\,\mathrm{Id} - A) < \varepsilon\right\}.$$

In some situations it might however be natural to consider only real perturbations E. It follows now from the definitions that the *real pseudospectrum*

$$\mathrm{sp}_{R,\varepsilon}(A) = \bigcup_{\substack{E \in \mathbf{R}^{n\times n} \\ \|E\| < \varepsilon}} \sigma(A + E)$$

is given by

$$\mathrm{sp}_{R,\varepsilon}(A) = \left\{z;\ \tau_1\left((z\,\mathrm{Id} - A)^{-1}\right) > \varepsilon^{-1}\right\} = \left\{z;\ \tilde{\tau}_n(z\,\mathrm{Id} - A) < \varepsilon\right\}.$$

The *real stability radius* of a stable matrix A is given by

$$r_R(A) = \min\{\|E\|,\ E \in R^{n\times n};\ A + E \text{ unstable}\},$$

where "stable" means that all eigenvalues are in a prescribed open region C_g of the complex plane. The real stability radius is given by the largest ε such that $\mathrm{sp}_{R,\varepsilon}(A)$ is contained in C_g. It can be computed as

$$r_R(A) = \inf_{z \in \partial C_g} \tilde{\tau}_n(z\,\mathrm{Id} - A)$$

where ∂C_g denotes the boundary of the stability region.

REFERENCES

1 B. Bernhardsson, Simultaneous Block-Diagonalization of One Hermitian and One Symmetric Form, Internal Report TFRT-7520, Dept. of Automatic Control, Lund, Sweden, 1994.

2 B. Bernhardsson, A. Rantzer, and L. Qiu, On real perturbation values, presented at 6th ILAS Conference, Chemnitz, 1996.

3 B. Bernhardsson, A. Rantzer, and L. Qiu, A summary on the real stability radius and real perturbation values, in *Hurwitz Centennial on Stability Theory*, Birkhäuser, 1995.

4 J. B. Ermolaev, The simultaneous reduction of symmetric and Hermitian forms, *Izv. Vyss. Zaved. Mat.* 21:10–23 (1961).

5 C. H. Fitzgerald and R. A. Horn, On the structure of Hermitian-symmetric inequalities, *J. London Math. Soc. (2)* 15:419–430 (1977).

6 J. Haantjes, Klassifikation der antilinearen Transformationen, *Math. Ann.* 112:98–106 (1936).

7 Y. P. Hong and R. A. Horn, On the reduction of pairs of Hermitian or symmetric matrices to diagonal form by congruence, *Linear Algebra Appl.* 72:213–226 (1986).

8 Y. P. Hong and R. A. Horn, A canonical form for matrices under consimilarity, *Linear Algebra Appl.* 102:143–168 (1988).

9 Y. P. Hong and R. A. Horn, A characterization of unitary congruence, *Linear and Multilinear Algebra* 25:105–119 (1989).

10 R. A. Horn and C. A. Johnson, *Matrix Analysis*, Cambridge U.P., 1985.

11 L. Qiu, B. Bernhardsson, A. Rantzer, E. Davison, P. Young, and J. Doyle, A formula for computation of the real stability radius, *Automatica* 31:879–890 (1995).

12 G. W. Stewart, On the early history of the singular value decomposition, *SIAM Rev.* 35:551–566 (1993).

13 N. Trefethen, Pseudospectra of matrices, in *Numerical Analysis 1991*, (D. F. Griffiths and G. A. Watson, eds.), Longam, 1992.

Received 25 March 1996; final manuscript accepted 13 April 1997