ON THE DISTANCE BETWEEN NORMAL MATRICES

MATJAŽ OMLADIČ AND PETER ŠEMRL

(Communicated by Palle E. T. Jorgensen)

ABSTRACT. The least upper bound for the norm distance between two normal matrices is given in terms of their eigenvalues exclusively, thus solving a problem which appears to be long open.

1. INTRODUCTION

A classical result of Weyl [12] estimates the norm distance between two hermitian matrices A and B using their respective eigenvalues λ_i and μ_i, for $i = 1, 2, \ldots, n$:

$$(1) \qquad \min_{\sigma} \max_{j} |\lambda_j - \mu_{\sigma(j)}| \leq \|A - B\| \leq \max_{\sigma} \max_{j} |\lambda_j - \mu_{\sigma(j)}|,$$

where σ runs through all permutations of n indices. This estimate holds as well in the case that A, B, and $A - B$ are all normal, the left-hand estimate being shown by Bhatia [2] and the right-hand one by Sunder [10]. These estimates and their generalizations were studied by many authors and a number of interesting results was obtained. For example, the same lower bound for $\|A - B\|$ may be obtained when A and B are both unitary [4, 7], or when A is hermitian and B is skew-hermitian [11]. Many other related results may be found in [3]. It is also interesting to observe that many of the above assumptions on A and B, as well as their eigenvalues, are preserved under unitary similarity. This helps in estimating the distance between unitary orbits of two normal matrices.

In this paper we study the upper bound for this distance. Our result is final in the sense that we assume "the most general" suppositions on operators A and B that make sense in this setting, namely, we assume only that they are both normal. Our result is also exact in the sense that the obtained upper bound may be attained (even by 2×2 matrices). A starting motivation of ours might be the result of Hoffman and Wielandt [9]. Let us introduce the Hilbert-Schmidt norm of an $n \times n$ matrix A by $\|A\|_H = (\sum_{i,j=1}^{n} |a_{ij}|^2)^{1/2}$ and denote by $D(A)$ the diagonalization of a diagonalizable matrix A. They proved that for any

normal matrices A and B it holds that

$$\min_{W} \|D(A) - WD(B)W^{-1}\|_H \leq \|A - B\|_H$$

$$\leq \max_{W} \|D(A) - WD(B)W^{-1}\|_H,$$

where W runs over all permutation matrices. Using the obvious estimate $\|A\| \leq \|A\|_H \leq \sqrt{n}\|A\|$ we can easily obtain

$$\frac{1}{\sqrt{n}} \min_{\sigma} \max_{j} |\lambda_j - \mu_{\sigma(j)}| \leq \|A - B\|$$

(2)
$$\leq \sqrt{n} \max_{\sigma} \max_{j} |\lambda_j - \mu_{\sigma(j)}|.$$

This estimate is now valid for any two normal matrices. It turns out that the constant in the lower bound may be chosen independent of n [5, 6]. In the case that A is hermitian and B is skew-hermitian it was shown in [1] that constant $\sqrt{n}$ in the upper bound may be replaced by $\sqrt{2}$:

(3)
$$\|A - B\| \leq \sqrt{2} \max_{\sigma} \max_{j} |\lambda_j - \mu_{\sigma(j)}|.$$

Consider the matrices

$$A = \left\| \begin{matrix} 0 & 1 \\ 1 & 0 \end{matrix} \right\|, \qquad B = \left\| \begin{matrix} 0 & 1 \\ -1 & 0 \end{matrix} \right\|$$

to see that this estimate may not be improved even in two dimensions. The main purpose of this paper is to show that (3) holds for any two normal matrices A and B, independent of their dimensions. Our ideas are close to those, used independently by Davidson and Djoković and by Elsner, in their unpublished attempts to give a general proof of the left-hand part of inequality (1) (see [8], p. 193).

2. Preliminaries

Lemma 2.1. *Given arbitrary complex n-tuples $(\lambda_1, \ldots, \lambda_n)$ and $(\mu_1, \ldots, \mu_n)$, and a positive ε, there exist complex n-tuples $(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_n)$ and $(\tilde{\mu}_1, \ldots, \tilde{\mu}_n)$ such that*

   (i) *$|\lambda_i - \tilde{\lambda}_i| < \varepsilon$ and $|\mu_i - \tilde{\mu}_i| < \varepsilon$ for $i = 1, \ldots, n$.*
  (ii) *$\tilde{\lambda}_i \neq \tilde{\lambda}_j$, $\tilde{\mu}_i \neq \tilde{\mu}_j$, $\tilde{\lambda}_i \neq \tilde{\mu}_j$, and $\tilde{\lambda}_i \neq \tilde{\mu}_i$ for any indices $i \neq j$.*
 (iii) *The boundary of any circle in the complex plane contains at most three points from the set of $\tilde{\lambda}$'s and at most three points from $\tilde{\mu}$'s.*
 (iv) *The boundaries of any couple of concentric circles in the complex plane contain at most four points from the union of the sets of $\tilde{\lambda}$'s and $\tilde{\mu}$'s.*

Remark. Note that assertion (iii) of this lemma is implied by assertion (iv). However, in the proof we will need (iii) in order to get (iv).

Proof. First choose $\tilde{\lambda}$'s so that (the according parts of) (i), (ii), and (iii) are satisfied. Using an induction argument assume that for an index $k < n$ the

points $\widetilde{\mu}_i$ for $i = 1, \ldots, k$ were already chosen in such a way that in addition to the desired properties they also satisfy the following condition: the set of all couples of concentric circles such that the boundary of one of them contains at least two of $\widetilde{\lambda}$'s and the boundary of the other contains at least two of $\widetilde{\mu}$'s is finite. Consider the open disc of radius ε centered at μ_{k+1}, and exclude from it all points from the set of $\widetilde{\lambda}$'s and boundaries of all circles concentric to a circle through any three, respectively two, respectively one, of these points and going through one, respectively two, respectively three, of $\widetilde{\mu}$'s. Observe that any of these conditions excludes only a finite number of boundaries. Further, exclude from it all points from the set of $\widetilde{\mu}$'s. For any pair of indices $i, j \leq n$, $i \neq j$ let S_{ij} denote the line of points on the same distance from $\widetilde{\lambda}_i$ and $\widetilde{\lambda}_j$. In order to preserve the additional inductive assumption we finally exclude from the disc under consideration all the points symmetric to $\widetilde{\mu}$'s with respect to any of the lines S_{ij}. Since only a finite number of closed sets with empty interior was excluded from the starting open disc, the remaining set is nonempty. Choose $\widetilde{\mu}_{k+1}$ to be any of the points from it to get the lemma. $\square$

From now on let $\mathscr{U}(n)$ denote the set of all $n \times n$ unitary matrices.

Lemma 2.2. *If $U: \mathbb{R} \to \mathscr{U}(n)$ is a differentiable mapping with $U(0) = I$, then $U'(0)$ is skew-symmetric. If V is any skew-symmetric matrix, then there exists a differentiable mapping $U: \mathbb{R} \to \mathscr{U}(n)$ such that $U(0) = I$ and $U'(0) = V$.*

Proof. The assumption $U(t)U(t)^* = I$ implies $U'(t)U(t)^* + U(t)(U'(t))^* = 0$. Insert $t = 0$ here so that $U'(0)$ is skew-symmetric. If V is skew-symmetric, define $U(t) = \exp(Vt)$ to get a mapping with the desired properties. $\square$

Throughout the paper vectors will be viewed as columns. Thus, for a (column) vector x, x^* will denote the row of the conjugated entries of x.

Lemma 2.3. *Let A and B be $n \times n$ matrices such that $\max_{U \in \mathscr{U}(n)} \|A - UBU^*\| = \|A - B\|$. Let x be a vector of norm one such that $(A - B)(A - B)^* x = \|A - B\|^2 x$ and let $y = (A - B)^* x$. Then $\mathbb{R}e(x^*(VB - BV)y) = 0$ for every skew-symmetric matrix V.*

Proof. Let V be any skew-symmetric matrix. Choose a differentiable mapping $U(t)$ as in Lemma 2.2, and define a function $f: \mathbb{R} \to \mathbb{R}$ by

$$f(t) = \|(A - U(t)BU(t)^*)^* x\|^2, \qquad t \in \mathbb{R}.$$

As f has a maximum at $t = 0$, we get $f'(0) = 0$ which yields

$$0 = (((-VB - BV^*)(A - B)^* + (A - B)(-VB - BV^*)^*)x, x)$$
$$= ((BV - VB)y, x) + (x, (BV - VB)y),$$

and the lemma follows. $\square$

3. MAIN RESULTS

Theorem 3.1. *Let A and B be two normal $n \times n$ matrices and let their respective eigenvalues λ_i and μ_i for $i = 1, 2, \ldots, n$ satisfy conditions (ii) and (iv) of Lemma 2.1. Then $\|A - B\| \leq \sqrt{2} \max_\sigma \max_j |\lambda_j - \mu_{\sigma(j)}|$.*

Proof. Let us show even "more," namely, that

$$\max_{U \in \mathscr{U}(n)} \|A - UBU^*\| \leq \sqrt{2} \max_{\sigma} \max_{j} |\lambda_j - \mu_{\sigma(j)}|.$$

Assume with no loss of generality that A and B satisfy

$$\max_{U \in \mathscr{U}(n)} \|A - UBU^*\| = \|A - B\|.$$

We are going to show that A, B, A^*, and B^* as operators on a finite-dimensional Hilbert space have a common two-dimensional invariant subspace, and then apply the known results for two dimensions as given in the Introduction. If $n \leq 2$, we are done. Thus, assume $n > 2$, choose a vector x of norm one such that $(A - B)(A - B)^* x = \|A - B\|^2 x$, denote $\rho = \|A - B\|^2$ and $y = (A - B)^* x$, and observe that $\rho = \|y\|^2$. Use Lemma 2.3 to show that $\mathbb{R}e(x^*(VB - BV)y) = 0$ holds for every skew-symmetric matrix V. For any scalar λ and any vector e of norm one orthogonal to both x and y, we may apply this result to $V = \lambda x e^* - \bar{\lambda} e x^*$ in order to get $\mathbb{R}e(\lambda e^* By + \bar{\lambda} x^* Be x^* y) = 0$, which implies $e^* By + y^* x e^* B^* x = 0$ after choosing λ appropriately. As e was an arbitrary vector from the orthogonal complement to $\mathscr{L}\{x, y\}$, where $\mathscr{L}\{x, y\}$ denotes the linear span of vectors x, y, we have necessarily

$$(4) \qquad\qquad By + (y^* x)B^* x \in \mathscr{L}\{x, y\}.$$

Replace $A - B$ by $(A - B)^*$, x by $x_1 = y/\sqrt{\rho}$, and y by $y_1 = \sqrt{\rho} x$ in this consideration to obtain, also,

$$(5) \qquad\qquad \rho B^* x + (x^* y)By \in \mathscr{L}\{x, y\}.$$

Let us first consider the case where $\rho = (x^* y)(y^* x)$. This implies that x and y are linearly dependent. Let us introduce a unitary operator $U = 2xx^* - I$ and observe that $(A - UBU^*)^* x = (A - B)^* x - 2(xx^* - I)B^* x$. As $(A - B)^* x \in \mathscr{L}\{x\}$ and x is orthogonal to $(xx^* - I)B^* x$, it follows from $\|(A - B)^* x\| \geq \|(A - UBU^*)^* x\|$ that $(xx^* - I)B^* x = 0$. Therefore, x is an eigenvector of B corresponding to eigenvalue μ_1 (after a permutation of eigenvalues if necessary), and consequently, x is also an eigenvector of A corresponding to eigenvalue λ_1 (again after a possible permutation). Finally, we obtain $\|A - B\| = \|(A - B)^* x\| = |\lambda_1 - \mu_1| \leq \max_{\sigma} \max_{j} |\lambda_j - \mu_{\sigma(j)}|$.

Next, assume that $\rho \neq (x^* y)(y^* x)$. In this case (4) and (5) imply that vectors By and $B^* x$ both belong to $\mathscr{L}\{x, y\}$. Denote by δ the coordinate of $B^* x$ corresponding to x with respect to the basis $\{x, y\}$, and write $C = A - \bar{\delta}$ and $D = B - \bar{\delta}$. Since clearly $\max_{U \in \mathscr{U}(n)} \|C - UDU^*\| = \|C - D\| = \|(C - D)^* x\|$, Lemma 2.3 may be applied for vectors x and y, and operators C and D. Thus, $Dy = \alpha x + \beta y$ and $D^* x = \gamma y$ for a certain choice of scalars α, β, and γ. From $x^* Dy = y^* D^* x$, it follows that

$$(6) \qquad\qquad \alpha + \beta(x^* y) = \rho \bar{\gamma}.$$

Choose $V = \lambda x y^* - \bar{\lambda} y x^*$ for a complex λ and apply Lemma 2.3 to show that $0 = \mathbb{R}e(\lambda y^* Dy - \rho \lambda x^* Dx)$ which implies $y^* Dy = \rho x^* Dx$. Consequently,

$\alpha(y^*x) + \rho\beta = y^*Dy = \rho x^*Dx = \rho\bar{\gamma}y^*x$, and (6) now forces $\beta = 0$. Thus, we obtain $Dy = \alpha x$, $Cy = (\rho + \alpha)x$, and therefore, also, $D^*x = \gamma y$, $C^*x = (1 + \gamma)y$. Hence,

$$(7) \qquad \begin{aligned} D^*Dy &= \alpha\gamma y, & C^*Cy &= (\rho + \alpha)(1 + \gamma)y, \\ DD^*x &= \alpha\gamma x, & CC^*x &= (\rho + \alpha)(1 + \gamma)x. \end{aligned}$$

For a set K of complex numbers and an arbitrary normal operator T, let $E_K(T)$ denote the spectral subspace of T corresponding to the set K, that is, the linear span of all eigenvectors of T corresponding to eigenvalues from K. Denote $L = \{\mu; |\mu - \bar{\delta}|^2 = \alpha\gamma\}$ and recall that $D = B - \bar{\delta}$ is normal to obtain from (7) that $\mathscr{L}\{x, y\} \subset E_{\{\alpha\gamma\}}(D^*D) \subset E_L(B)$. Similarly, the normality of $C = A - \bar{\delta}$ implies that $\mathscr{L}\{x, y\} \subset E_K(A)$, where $K = \{\lambda; |\lambda - \bar{\delta}|^2 = (\rho + \alpha)(1 + \gamma)\}$. Since vectors x and y are linearly independent, spaces $E_K(A)$ and $E_L(B)$ must be at least two-dimensional. On the other hand, the hypotheses (ii) and (iv) imply that the sum of the dimensions of these two spectral subspaces is at most four. This insures that $\mathscr{L}\{x, y\}$ is a spectral subspace for A, as well as a spectral subspace of B, and is consequently invariant under all the operators A, A^*, B, and B^*. □

Remark. Observe that in the course of the proof of Theorem 3.1 we actually proved the following:

Let A and B be two normal $n \times n$ matrices and let their respective eigenvalues λ_i and μ_i for $i = 1, 2, \ldots, n$ satisfy conditions (ii) and (iv) of Lemma 2.1. Moreover, assume that $\max_{U \in \mathscr{U}(n)} \|A - UBU^*\| = \|A - B\|$, let x be any nonzero vector such that $(A - B)(A - B)^*x = \|A - B\|^2 x$, and let $y = (A - B)^*x$. Then $\mathscr{L}\{x, y\}$ is a spectral subspace of both A and B. Therefore, it is invariant under all the operators A, A^*, B, and B^*.

Main Theorem. *Let A and B be two normal matrices with respective eigenvalues denoted by λ_i and μ_i for $i = 1, 2, \ldots, n$. Then*

$$\|A - B\| \leq \sqrt{2} \max_\sigma \max_j |\lambda_j - \mu_{\sigma(j)}|.$$

Corollary. *Let A and B be two normal matrices with respective eigenvalues denoted by λ_i and μ_i for $i = 1, 2, \ldots, n$. Then*

$$\max_{U \in \mathscr{U}(n)} \|A - UBU^*\| \leq \sqrt{2} \max_\sigma \max_j |\lambda_j - \mu_{\sigma(j)}|.$$

Proof. As the Main Theorem and its Corollary are clearly equivalent, we may prove just the latter. But, clearly as well, we have $\max_{U \in \mathscr{U}(n)} \|A - UBU^*\| = \max_{U \in \mathscr{U}(n)} \|D(A) - UD(B)U^*\|$, while the right-hand side of this equation equals

$$\begin{aligned} f(\lambda_1, \ldots, \lambda_n, \mu_1, \ldots, \mu_n) \\ = \max_{U \in \mathscr{U}(n)} \|\operatorname{Diag}(\lambda_1, \ldots, \lambda_n) - U\operatorname{Diag}(\mu_1, \ldots, \mu_n)U^*\|. \end{aligned}$$

Thus, in order to get the theorem it suffices to show that f is a continuous function of λ's and μ's and then apply Lemma 2.1 and Theorem 3.1. To obtain the continuity of f, fix a set of λ's and μ's, choose a matrix $U_0 \in \mathscr{U}(n)$ such that the maximum in the definition of f is attained at U_0, and pick for any positive ε a set of $\widetilde{\lambda}$'s and $\widetilde{\mu}$'s such that $|\widetilde{\lambda}_i - \lambda_i| < \varepsilon$ and $|\widetilde{\mu}_i - \mu_i| < \varepsilon$. It follows that

$$
\begin{aligned}
f(\lambda_1, \ldots, &\lambda_n, \mu_1, \ldots, \mu_n) \\
&< \| \operatorname{Diag}(\widetilde{\lambda}_1, \ldots, \widetilde{\lambda}_n) - U_0 \operatorname{Diag}(\widetilde{\mu}_1, \ldots, \widetilde{\mu}_n) U_0^* \| + 2\varepsilon \\
&\leq f(\widetilde{\lambda}_1, \ldots, \widetilde{\lambda}_n, \widetilde{\mu}_1, \ldots, \widetilde{\mu}_n) + 2\varepsilon,
\end{aligned}
$$

and by symmetry $f(\widetilde{\lambda}_1, \ldots, \widetilde{\lambda}_n, \widetilde{\mu}_1, \ldots, \widetilde{\mu}_n) < f(\lambda_1, \ldots, \lambda_n, \mu_1, \ldots, \mu_n) + 2\varepsilon$, which finishes the proof. $\square$

REFERENCES

1. T. Ando and R. Bhatia, *Eigenvalue inequalities associated with the Cartesian decomposition*, Linear and Multilinear Algebra **22** (1987), 133–147.

2. R. Bhatia, *Analysis of spectral variation and some inequalities*, Trans. Amer. Math. Soc. **272** (1982), 323–332.

3. _____, *Perturbation bounds for matrix eigenvalues*, Pitman Res. Notes Math., vol. 162, 1987.

4. R. Bhatia and C. Davis, *A bound for the spectral variation of a unitary operator*, Linear and Multilinear Algebra **15** (1984), 71–76.

5. R. Bhatia, C. Davis, and P. Koosis, *An extremal problem in Fourier analysis with applications to operator theory*, J. Funct. Anal. **82** (1989), 138–150.

6. R. Bhatia, C. Davis, and A. McIntosh, *Perturbation of spectral subspaces and solution of linear operator equations*, Linear Algebra Appl. **52/53** (1983), 45–67.

7. R. Bhatia and J. A. R. Holbrook, *Short normal paths and spectral variation*, Proc. Amer. Math. Soc. **94** (1985), 377–382.

8. K. R. Davidson, *Finite dimension problems in operator theory*, Oper. Theory: Adv. Appl. **40** (1989), 187–201.

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

10. V. S. Sunder, *On permutations, convex hulls and normal operators*, Linear Algebra Appl. **48** (1982), 403–411.

11. _____, *Distance between normal operators*, Proc. Amer. Math. Soc. **84** (1982), 483–484.

12. H. Weyl, *Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen*, Math. Ann. **71** (1912), 441–479.

DEPARTMENT OF MATHEMATICS, E. K. UNIVERSITY OF LJUBLJANA, JADRANSKA 19, 61000 LJUBLJANA, YUGOSLAVIA