

SHORT NORMAL PATHS AND SPECTRAL VARIATION¹

RAJENDRA BHATIA AND JOHN A. R. HOLBROOK

ABSTRACT. We introduce the notion of a “short normal path” between matrices S and T , that is, a continuous path from S to T consisting of normal matrices and having the same length as the straight line path from S to T . By this means we prove that for certain normal matrices S and T the eigenvalues of S and T may be paired in such a way that the maximum distance (in the complex plane) between the pairs is no more than the operator norm $\|S - T\|$. In particular, we generalize and provide a new approach to a recent result of Bhatia and Davis treating the case of unitary S and T .

1. Introduction. The basic aim of this work is to see how far the “normal path method” introduced in Bhatia [1] can advance our understanding of the spectral variation of normal matrices.

Let \mathbf{N} denote the set of normal matrices of a certain fixed size n ; if we wish to emphasize that we are speaking only of n -by- n matrices, we will write $\mathbf{N}(n)$. Thus $\mathbf{N}(n) =$ all complex n -by- n matrices N such that $N^*N = NN^*$, where M^* denotes the adjoint (conjugate transpose) of M . The norm of M as an operator on the Hilbert space \mathbf{C}^n is denoted here by $\|M\|$. We shall use several times the unitary invariance of this norm, that is, the relation $\|UMV\| = \|M\|$ for unitary U and V . We shall also routinely allow ourselves the convenience of speaking in terms of the operator corresponding to M , rather than the matrix M itself, and of changing bases where appropriate.

Consider two elements S and T of $\mathbf{N}(n)$; let $\lambda(1), \dots, \lambda(n)$ list the eigenvalues of S (repeating eigenvalues according to their multiplicities) and let $\mu(1), \dots, \mu(n)$ do the same for T . Since there is no canonical way to pair the eigenvalues of S with those of T , it is appropriate to define the “spectral distance” from S to T by

$$d(\text{Eig } S, \text{Eig } T) = \min_{\pi} \left(\max_k |\lambda(k) - \mu(\pi(k))| \right),$$

where π ranges over all permutations of $(1, 2, \dots, n)$. The following problem (raised explicitly in L. Mirsky [4]) has a considerable history: do we have

$$(1) \quad d(\text{Eig } S, \text{Eig } T) \leq \|S - T\|$$

for all S, T in \mathbf{N} ?

Received by the editors September 13, 1983 and, in revised form, July 11, 1984. Results from this paper were presented to the International Congress of Mathematicians (Warsaw, August 1983), to the Ontario Mathematics Meeting (Guelph, October 1983), and to the Canadian Mathematical Society (Montreal, December 1983).

1980 *Mathematics Subject Classification.* Primary 15A42, 15A57, 15A60; Secondary 53C99, 65F99.

Key words and phrases. Matrices, normal operator, spectrum, eigenvalue, spectral variation.

¹Much of this work was done while the authors were together at the University of Guelph. They are grateful to NSERC of Canada for its financial support through Operating Grant A8745.

The answer is “yes” in several interesting cases: H. Weyl noted this for the case of hermitian S and T in the course of [6]; V. S. Sunder showed in [5] that (1) holds whenever S is hermitian and T is skew-hermitian; recently Bhatia and Ch. Davis [2] have established (1) for unitary S and T . In each of these cases the argument involves finding a good matching of the eigenvalues of S and T , i.e., finding a good permutation π in the sense of our definition of $d(\text{Eig } S, \text{Eig } T)$.

In Bhatia [1] we find a different approach. There the matching is only implicit and the essential observation is that

$$(2) \quad d(\text{Eig } S, \text{Eig } T) \leq \text{length of any normal path joining } S \text{ and } T.$$

In §2 we give a different proof of (2) that removes certain technical restrictions on the method of [1]. This inequality clearly means, as noted in [1], that (1) holds for any S and T such that the line segment joining them consists entirely of normal matrices; Weyl’s result is, of course, included in this case. It will be useful to note that, assuming S and T are normal, the line segment joining S and T lies in \mathbf{N} exactly when $S - T$ is normal.

The message of the present paper is that there are significant cases where S and T can be joined by a path in \mathbf{N} that is “short”, that is, of length $\|S - T\|$, even though the “obvious” path of that length, namely the line segment from S to T , does not lie in \mathbf{N} . By a “plain” in \mathbf{N} we shall mean a subset F of \mathbf{N} such that any pair S, T in F may be joined by a short normal path lying in F . Within plains, clearly, (2) ensures that spectral variation is limited by (1). In §3 we show that CU , the set of scalar multiples of unitary matrices, is a plain. The corresponding application of (2) generalizes the result of Bhatia and Davis. It turns out that $\mathbf{N}(2)$ is also a plain. Nevertheless, our method does not always yield the best results on spectral variation. We are indebted to M.-D. Choi for an example of 3×3 normal matrices T and S that are not joined by any short normal path. Moreover, in Choi’s example one can take S to be hermitian and T skew-hermitian, so that Sunder’s theorem applies.

2. Normal paths and spectral variation. By a normal path (or curve) we mean a continuous function $\gamma: [a, b] \rightarrow \mathbf{N}$, defined on some real interval $[a, b]$. By the length of γ , $|\gamma|$, we mean the arclength with respect to the operator norm, that is,

$$|\gamma| = \sup \left\{ \sum \|\gamma(t_{k+1}) - \gamma(t_k)\| : a = t_0 < t_1 < \cdots < t_m = b \right\}.$$

In cases of interest to us, this length will be finite (the path will be “rectifiable”). We say such a path joins matrices S and T when $\gamma(a) = S$ and $\gamma(b) = T$.

In [1, see Theorem 3.3] Bhatia showed that (subject to certain technical restrictions), $d(\text{Eig } S, \text{Eig } T) \leq |\gamma|$ for any normal path joining S and T . This is the result that we mean to exploit in a new way in §3. First, however, we present a different and simpler proof of this inequality, which also removes the technical restrictions that seemed necessary in [1]. While the results of this section will be sufficient for our present purposes, the method of [1] (based on a moving decomposition of the tangent vector to the normal curve) entails other possibilities that should not be ignored. In particular, it applies, not just to the operator norm, but to a whole class of unitarily invariant matrix norms.

It is clear that the spectral distance $d(\text{Eig } S, \text{Eig } T)$ defines a pseudometric on \mathbf{N} , and standard results on spectral continuity (see, for example, T. Kato [3, Chapter 2, §5.1]) imply that

$$(3) \quad d(\text{Eig } T_k, \text{Eig } T) \rightarrow 0 \quad \text{whenever } \|T_k - T\| \rightarrow 0.$$

The only other ingredient we need is the following quantitative proposition on spectral variation around normal matrices. This proposition is well known, but since it plays a special role in our discussion we include the simple proof.

PROPOSITION 2.1. *Let N be normal and let A be any matrix. For each eigenvalue μ of A there is some eigenvalue λ of N such that $|\mu - \lambda| \leq \|A - N\|$.*

PROOF. Let u be a unit eigenvector of A corresponding to μ , and let u_1, \dots, u_n be an orthonormal basis of eigenvectors of N with corresponding eigenvalues $\lambda_1, \dots, \lambda_n$. Then

$$\begin{aligned} \|A - N\|^2 &\geq \|(A - N)u\|^2 = \|\mu u - \sum \lambda_k (u, u_k) u_k\|^2 \\ &= \|\sum (\mu - \lambda_k)(u, u_k) u_k\|^2 = \sum |\mu - \lambda_k|^2 |(u, u_k)|^2. \end{aligned}$$

Since $\sum |(u, u_k)|^2 = 1$, it is clear that for some $\lambda (= \lambda_k)$ we have $|\mu - \lambda| \leq \|A - N\|$. Q.E.D.

THEOREM 2.2. *Let γ be a (rectifiable) normal path joining the (normal) matrices S and T . Then $d(\text{Eig } S, \text{Eig } T) \leq |\gamma|$.*

PROOF. For convenience let us suppose that γ is parametrized on $[0, 1]$, and let us denote by γ_r that part of the curve which is parametrized on $[0, r]$. Let G be the set $\{r \in [0, 1]: d(\text{Eig } S, \text{Eig } \gamma(r)) \leq |\gamma_r|\}$; we wish to show that $1 \in G$.

Note first that, by the continuity of γ and of arclength, and by (3), G is closed. In particular $R \in G$, where $R = \sup G$. We claim that $R = 1$, for otherwise consider $N = \gamma(R)$, and let d be the minimum distance between distinct eigenvalues of N . Using again the continuity of γ and (3), we can find $r' \in (R, 1]$ such that $A = \gamma(r')$ satisfies $d(\text{Eig } N, \text{Eig } A) < d/2$. The matching of eigenvalues achieving $d(\text{Eig } N, \text{Eig } A)$ must then be such that each eigenvalue of A is paired with the closest eigenvalue of N . By Proposition 2.1 we conclude that $d(\text{Eig } N, \text{Eig } A) \leq \|N - A\|$. But then

$$d(\text{Eig } S, \text{Eig } \gamma(r')) \leq d(\text{Eig } S, \text{Eig } N) + d(\text{Eig } N, \text{Eig } A) \leq |\gamma_R| + \|N - A\|,$$

and since this last sum is certainly no greater than $|\gamma_{r'}|$ we have $r' \in G$, a contradiction. Q.E.D.

3. Construction of short normal paths.

THEOREM 3.1. *The set \mathbf{CU} , consisting of all scalar multiples of unitary matrices, is a plain (that is, every two elements can be joined by a short (normal) path lying in the set).*

PROOF. Any two elements N_0 and N_1 of \mathbf{CU} may be written in the form $N_0 = r_0 U_0$, $N_1 = r_1 U_1$, where the r_k are nonnegative reals and the U_k are unitary. Let K be a

skew-hermitian operator such that the unitary $U_1(U_0)^{-1}$ may be written as $\exp K$, and such that the eigenvalues of K lie in the interval $(-i\pi, i\pi]$.

With respect to an appropriate orthonormal basis of \mathbf{C}^n , K corresponds to $\text{diag}(i\theta_1, \dots, i\theta_n)$, where

$$(4) \quad |\theta_n| \leq \dots \leq |\theta_1| \leq \pi.$$

Since reduction to this form can be achieved by a unitary conjugation, and since norms and distances are all invariant under such conjugations, we can work with matrix representations of all operators with respect to this basis without changing any quantity of interest to us. In particular, we have

$$\|N_0 - N_1\| = \|r_0 - r_1 U_1(U_0)^{-1}\| = \|r_0 - r_1 \exp K\|,$$

the norm of a diagonal matrix with eigenvalues $r_0 - r_1 \exp(i\theta_k)$. Since we have (4) and r_0, r_1 are nonnegative, it follows that

$$(5) \quad \|N_0 - N_1\| = |r_0 - r_1 \exp(i\theta_1)|.$$

Parametrize the straight line from r_0 to $r_1 \exp(i\theta_1)$ in the form $r(t)\exp(it\theta_1)$, $0 \leq t \leq 1$. The right side of (5) is the length of this line, so that we also have

$$(6) \quad \begin{aligned} \|N_0 - N_1\| &= \int_0^1 |r'(t)\exp(it\theta_1) + r(t)i\theta_1 \exp(it\theta_1)| dt \\ &= \int_0^1 |r'(t) + r(t)i\theta_1| dt. \end{aligned}$$

Now define the operator-valued path $N(t) = r(t)\exp(tK)U_0$, $0 \leq t \leq 1$. Evidently $N(t)$ traces out a smooth curve in \mathbf{CU} , and it is easy to see that $N(0) = N_0$ and $N(1) = N_1$.

We claim that this curve is a short (normal) path. Indeed, its length L is given by

$$\begin{aligned} L &= \int_0^1 \|N'(t)\| dt \\ &= \int_0^1 \|r'(t)\exp(tK)U_0 + r(t)K \exp(tK)U_0\| dt, \end{aligned}$$

and since the norm is unitarily invariant we also have

$$L = \int_0^1 \|r'(t) + r(t)K\| dt.$$

Now the operator $r'(t) + r(t)K$ corresponds to a diagonal matrix with eigenvalues $r'(t) + r(t)i\theta_k$, so that its norm is the maximum modulus of these expressions. In view of (4) and the fact that $r'(t)$ and $r(t)$ are real, we have

$$(7) \quad L = \int_0^1 |r'(t) + r(t)i\theta_1| dt.$$

Equations (6) and (7) show that our path is indeed short. Q.E.D.

THEOREM 3.2. *Let S and T be normal operators with spectra lying on circles $C(S)$ and $C(T)$ respectively. If these circles are concentric then*

$$d(\text{Eig } S, \text{Eig } T) \leq \|S - T\|.$$

PROOF. If α is the common centre of the circles $C(S)$ and $C(T)$, then S and T lie in $\alpha + \mathbf{CU}$, clearly a (normal) plain by the last theorem. Our inequality follows from Theorem 2.2. Q.E.D.

In particular, we obtain the following spectral variation result of Bhatia and Davis [2] by a method quite different from theirs.

COROLLARY 3.3. *If U and V are unitary, $d(\text{Eig } U, \text{Eig } V) \leq \|U - V\|$.*

It is a curious fact that $\mathbf{N}(2)$ is itself a plain. The example of Choi, mentioned in the introduction, dashes one's hope that $\mathbf{N}(n)$ might be a plain for some $n > 2$.

THEOREM 3.4. *$\mathbf{N}(2)$ is a plain.*

PROOF. Let S and T be elements of $\mathbf{N}(2)$. Since the spectra contain at most two points each, an elementary geometric construction shows that, except perhaps in the case where the spectra lie on parallel lines, there are concentric circles containing them. If α is the common centre, we have S and T contained in $\alpha + \mathbf{CU}$, which is a plain by Theorem 3.1, so that there is a short normal path joining S and T .

In the exceptional cases we may assume, by introducing an appropriate scalar multiplication, that the spectra of S and T lie on lines parallel to the real axis. Then the skew-hermitian part of $S - T$ is a scalar, so that $S - T$ is normal. As noted earlier, this ensures that the straight line from S to T lies in \mathbf{N} . Q.E.D.

Theorem 3.4 implies, of course, that (1) holds for all 2-dimensional operators. This fact in itself has long been known and may be established by various more direct arguments. For example, Proposition 2.1 may be applied symmetrically, that is, with $N = S$, $A = T$, and with $N = T$, $A = S$. When only two-point spectra are involved, this is enough to conclude that (1) holds.

It may have occurred to the reader that there is another natural way to deal with the exceptional cases in our proof of Theorem 3.4. One could consider perturbations of S or T that place the spectra on nonparallel lines and look for some sort of limit of the corresponding short normal paths as the perturbation is reduced to zero. More generally, one may ask whether the infimum of the lengths of normal paths joining a pair S and T is attained. This question is answered by the next proposition.

PROPOSITION 3.5. *If F is a closed set of matrices and*

$$L = \inf\{|\gamma| : \gamma \text{ is a path in } F \text{ joining } S \text{ and } T\},$$

then L is achieved, that is, there is a path of length L joining S and T in F .

PROOF. Let γ_k be a path from S to T lying in F and having length L_k , where $L_k \rightarrow L$ as $k \rightarrow \infty$.

We may reparametrize the paths with respect to arclength and by a change of scale we may assume that, for each k ,

$$\gamma_k : [0, 1] \rightarrow F \quad \text{with } \gamma_k(0) = S \text{ and } \gamma_k(1) = T,$$

and that, for each subinterval I of $[0, 1]$, $\gamma_{k,I}$, the part of γ_k parametrized on I , has length $|I|L_{ik}$.

After these modifications the functions γ_k are equicontinuous on $[0, 1]$; in fact,

$$\|\gamma_k(a) - \gamma_k(b)\| \leq |\gamma_{k,[a,b]}| = |b - a|L_k.$$

A standard argument of Arzela-Ascoli type implies that there is a subsequence of the γ_k converging uniformly to a path γ . Clearly this path will join S and T , lie in F , and have length L . Q.E.D.

REFERENCES

1. R. Bhatia, *Analysis of spectral variation and some inequalities*, Trans. Amer. Math. Soc. **272** (1982), 323–331.
2. R. Bhatia and Ch. Davis, *A bound for the spectral variation of a unitary operator*, Linear and Multilinear Algebra **15** (1984), 71–76.
3. T. Kato, *Perturbation theory for linear operators*, Springer-Verlag, Berlin and New York, 1980.
4. L. Mirsky, *Symmetric gauge functions and unitarily invariant norms*, Quart. J. Math. Oxford Ser. (2) **11** (1960), 50–59.
5. V. S. Sunder, *Distance between normal operators*, Proc. Amer. Math. Soc. **84** (1982), 483–484.
6. H. Weyl, *Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen*, Math. Ann. **71** (1912), 441–479.

CENTRE FOR ADVANCED STUDY IN MATHEMATICS, UNIVERSITY OF BOMBAY, BOMBAY 400 098, INDIA

DEPARTMENT OF MATHEMATICS AND STATISTICS, UNIVERSITY OF GUELPH, GUELPH, ONTARIO, N1G 2W1, CANADA