Spectral Variation, Normal Matrices, and Finsler Geometry

RAJENDRA BHATIA

How did two matrix-theorists who had never worked together before come to prove a theorem which has had consequences throughout the field and beyond? I will try to put together the personal and the mathematical sides of the Hoffman-Wielandt Theorem, its prehistory, and attempts (both successful and unsuccessful) to generalise it.

> *Wielandt was really trying to do the thing for operator norms and the Frobenius norm was his second choice.*

Thus begins Alan Hoffman's commentary on his joint paper with Helmut Wielandt [HW], one of the best known in linear algebra. The paper is less than three pages long and, of a piece with that brevity, Hoffman's commentary consists of just one paragraph. It continues,

> *In fact, he had a proof of HW with a constant bigger than 1 in front. It was quite lovely, involving a path in matrix space, and I hope someone else has found a use for that method. Since linear programming was in the air at the National Bureau of Standards in those days, it was natural for us to discover the proof that appeared in the paper. The most difficult task was convincing each other that something this short and simple was worth publishing. In fact, we padded it with a new proof of the Birkhoff theorem on doubly stochastic matrices. I think the reason for the theorem's popularity is the publicity given it by Wilkinson in his book on the algebraic eigenvalue problem (J. H. Wilkinson, The Algebraic Eigenvalue Problem, Clarendon Press, Oxford, 1965).*

In this article I will explain what it was that Wielandt was really trying to do, why he wanted to do it for operator norms, and what some others had done before him and have done since.

Wielandt's mathematical works [Wiel] straddle two different fields: group theory and matrix analysis. He began with the first, was pulled into the second, and then happily continued with both. The circumstances are best described in his own words:

> *The group-theoretic work was interrupted for several years while, during the second half of the war, at the Göttingen Aerodynamics Research Institute, I had to work on vibration problems. I am indebted to that time for valuable discoveries: on the one hand the applicability of abstract tools to the solution of concrete problems, on the other hand, the—for a pure mathematician—unexpected difficulty and unaccustomed responsibility of numerical evaluation. It was a matter of estimating eigenvalues of non-selfadjoint differential equations and matrices. I attacked the more general problem of developing a metric spectral theory, to begin with for finite complex matrices.*

The links between all parts of our story are contained in the two paragraphs I have quoted from Hoffman and from Wielandt.

By the time Wielandt came to Göttingen in 1942, Hermann Weyl had left. Thirty years earlier Weyl had published a fundamental paper [We] on asymptotics of eigenvalues of partial differential operators. Among the several things Weyl accomplished in that paper are many interesting inequalities relating the eigenvalues of Hermitian matrices A, B, and $A + B$. One of them can be translated into the following perturbation theorem: *If A and B are $n \times n$ Hermitian matrices, and their eigenvalues are enumerated as $\alpha_1 \geq \alpha_2 \geq \cdots \alpha_n$, and $\beta_1 \geq \beta_2 \geq \cdots \geq \beta_n$, respectively, then*

$$\max_{1 \leq j \leq n} |\alpha_j - \beta_j| \leq \|A - B\|. \tag{1}$$

Here $\|A\|$ stands for the norm of A as a linear operator on the Euclidean space $\mathbb{C}^n$; i.e.,

$$\|A\| = \max \{\|Ax\| : x \in \mathbb{C}^n, \|x\| = 1\}. \tag{2}$$

Apart from the intrinsic mathematical interest that Weyl's inequality (1) has, it soothes the analyst's anxiety about "the unaccustomed responsibility of numerical evaluation." If one replaces a Hermitian matrix A by a nearby Hermitian matrix B, then the eigenvalues are changed by no more than the change in the matrix.

Almost the first question that arises now is whether the inequality remains true for a wider class of matrices, and for a mathematician interested in "estimating eigenvalues of non-selfadjoint differential equations and matrices" this would be more than mere curiosity. The first wider class to be considered is that of normal matrices. (An operator A is normal if $AA^* = A^*A$. This is equivalent to the condition that in some orthonormal basis the matrix of A is diagonal. The diagonal entries are the eigenvalues of A, and A is Hermitian if and only if these are all real.)

The eigenvalues of a normal matrix, now being complex, cannot be ordered in any natural way, and we have to define an appropriate distance to replace the left-hand side of (1). If Eig $A = \{\alpha_1, \ldots, \alpha_n\}$ and Eig $B = \{\beta_1, \ldots, \beta_n\}$ are the unordered n-tuples whose elements are the eigenvalues of A and B, respectively, then we define the *optimal matching distance*

$$d(\text{Eig } A, \text{Eig } B) = \min_{\sigma} \max_{1 \leq j \leq n} |\alpha_j - \beta_{\sigma(j)}|, \tag{3}$$

where σ varies over all permutations of the indices $\{1, 2, \ldots, n\}$. The question raised by Weyl's inequality is: if A and B are any two normal matrices, then do we have

$$d(\text{Eig } A, \text{Eig } B) \leq \|A - B\|? \tag{4}$$

This is what Wielandt, and several others over nearly four decades, attempted to prove. We will return to that story later.

The operator norm (2) is the one that every student of functional analysis first learns about. Its definition carries over to all bounded linear operators on an infinite-dimensional Hilbert space. That explains why this norm would have been Wielandt's first choice. There are other possible choices.

The *Frobenius norm* of an $n \times n$ matrix A is defined as

$$\|A\|_F = (\text{tr } A^* A)^{1/2} = \Big(\sum_{i,j} |a_{ij}|^2\Big)^{1/2}. \tag{5}$$

This norm arises from the inner product $\langle A, B\rangle = \text{tr } A^* B$, and, for this reason, it has pleasant geometric features. It can be easily computed from the entries of A. If we replace the norm (2) with (5), then we must make a similar change in the distance (3) and define

$$d_F(\text{Eig } A, \text{Eig } B) = \min_{\sigma} \Big[\sum_{j=1}^{n} |\alpha_j - \beta_{\sigma(j)}|^2\Big]^{1/2}. \tag{6}$$

Instead of (4), Hoffman and Wielandt proved the following.

THEOREM 1 *Let A and B be any two normal matrices. Then*

$$d_F(\text{Eig } A, \text{Eig } B) \leq \|A - B\|_F. \tag{7}$$

Hoffman credits J. H. Wilkinson [Wil] with the publicity responsible for the theorem's popularity. Wilkinson writes

The Wielandt-Hoffman theorem does not seem to have attracted as much attention as those arising from the direct application of norms. In my experience it is the most useful result for the error analysis of techniques based on orthogonal transformations in floating-point arithmetic.

He also gives an elementary proof for the (most interesting) special case when A and B are Hermitian. In spite of Wilkinson's reversal of the order of names of its authors, the theorem is known as the Hoffman-Wielandt theorem.

Unknown, it would seem, to Hoffman and Wielandt, and

RAJENDRA BHATIA has been associated with ISI Delhi most of the time since his graduate student days. According to the Mathematics Genealogy Project, he is scientifically a direct descendant of Arthur Cayley, via A. Forsyth, E. Whittaker, James Jeans, R.A. Fisher, C.R. Rao, K.R. Parthasarathy. The photograph of him here (taken by George Bergman, whom we thank) shows him on his arrival at University of California Berkeley on a post-doctoral fellowship, 1979.

Indian Statistical Institute Delhi
New Delhi, 110016
India
e-mail: rbh@isid.ac.in

to Wilkinson, the Hermitian special case of (7) had been announced several years earlier, by Karl Löwner in 1934 [Lo]. This paper is very well-known for its deep analysis of operator monotone functions. Somewhat surprisingly, there is no reference to it in most of the papers and books where the inequality (7) is discussed. (Incidentally, Löwner was at the University of Berlin between 1922 and 1928. Wielandt came to study there in 1929 and obtained a Ph.D. in 1935. Löwner's original Czech name was Karel but, because his education was in German, he was known as Karl. Later, when he had to move to the United States, he adopted the name Charles Loewner.) Löwner does not offer a proof and says that the inequality can be established via a simple variational consideration.

One such consideration might go as follows. When $x = (x_1, \ldots, x_n)$ is any vector with real coordinates, let $x^{\downarrow} = (x_1^{\downarrow}, \ldots, x_n^{\downarrow})$ and $x^{\uparrow} = (x_1^{\uparrow}, \ldots, x_n^{\uparrow})$ be the decreasing and increasing rearrangements of x. This means that the numbers $x_1, \ldots, x_n$ are rearranged as $x_1^{\downarrow} \geq \cdots \geq x_n^{\downarrow}$ and as $x_1^{\uparrow} \leq \cdots \leq x_n^{\uparrow}$. Then for any two vectors x and y, we have

$$\text{(8)} \qquad \sum_{j=1}^{n} x_j^{\downarrow} y_j^{\uparrow} \leq \sum_{j=1}^{n} x_j y_j \leq \sum_{j=1}^{n} x_j^{\downarrow} y_j^{\downarrow}.$$

To see this, first note that the general case can be reduced to the special case $n = 2$. This amounts to showing that whenever $x_1 \geq x_2$ and $y_1 \geq y_2$, then $x_1y_1 + x_2y_2 \geq x_1y_2 + x_2y_1$. The latter inequality can be written as $(x_1 - x_2)(y_1 - y_2) \geq 0$ and is obviously true.

A matrix analogue of this inequality is given in the following proposition. If A is a Hermitian matrix we denote by $\mathrm{Eig}^{\downarrow}(A) = (\lambda_1^{\downarrow}(A), \ldots, \lambda_n^{\downarrow}(A))$ the vector whose coordinates are the eigenvalues of A arranged in decreasing order. Similarly $\mathrm{Eig}^{\uparrow}(A) = (\lambda_1^{\uparrow}(A), \ldots, \lambda_n^{\uparrow}(A))$ is the vector whose coordinates are the same numbers arranged in increasing order. The bracket $\langle x, y\rangle$ stands for the usual scalar product $\sum_{j=1}^{n} x_j y_j$.

PROPOSITION 2 *Let A and B be $n \times n$ Hermitian matrices. Then*

$$\text{(9)} \quad \langle \mathrm{Eig}^{\downarrow}(A), \mathrm{Eig}^{\uparrow}(B)\rangle \leq \mathrm{tr}\, AB \leq \langle \mathrm{Eig}^{\downarrow}(A), \mathrm{Eig}^{\downarrow}(B)\rangle.$$

Proof. If A and B are commuting Hermitian matrices, this reduces to (8). The general case can be reduced to this special one as follows.

Let $U(n)$ be the set of all $n \times n$ unitary matrices, and let

$$\mathcal{U}_B = \{U BU^* : U \in U(n)\},$$

be the unitary orbit of B. If we replace B by any element of $\mathcal{U}_B$, then the eigenvalues of B are not changed, and hence neither are the two inner products in (9). Consider the function $f(X) = \mathrm{tr}\, AX$ defined on the compact set $\mathcal{U}_B$. The two inequalities in (9) are lower and upper bounds for $f(X)$. Both will follow if we show that every extreme point X_0 for f commutes with A.

A point X_0 on $\mathcal{U}_B$ is an extreme point if and only if

$$\left.\frac{d}{dt}\right|_{t=0} \mathrm{tr}\, AU(t)X_0U(t)^* = 0$$

for every differentiable curve $U(t)$ with $U(0) = I$. This is equivalent to saying

$$\left.\frac{d}{dt}\right|_{t=0} \mathrm{tr}\, Ae^{tK}X_0e^{-tK} = 0$$

for every skew-Hermitian matrix K. Expanding the exponentials into series, this condition reduces to

$$\mathrm{tr}\,(AKX_0 - AX_0K) = 0.$$

The trace of a product being invariant under cyclic permutation of the factors, this is the same as saying

$$\mathrm{tr}\, K(X_0A - AX_0) = 0.$$

Since $\langle K, L\rangle = -\mathrm{tr}KL$ is an inner product on the space of skew-Hermitian matrices, this is possible if and only if $X_0A - AX_0 = 0$. □

Using the second inequality in (9) we see that

$$\begin{aligned} \|A - B\|_F^2 &= \|A\|_F^2 + \|B\|_F^2 - 2\mathrm{tr}AB \\ &\geq \|A\|_F^2 + \|B\|_F^2 - 2\langle \mathrm{Eig}^{\downarrow}(A), \mathrm{Eig}^{\downarrow}(B)\rangle \\ \text{(10)} \qquad &= \sum_{j=1}^{n} |\lambda_j^{\downarrow}(A) - \lambda_j^{\downarrow}(B)|^2. \end{aligned}$$

This proves the inequality (7) for Hermitian matrices.

The same argument, using the first inequality in (9), shows that

$$\text{(11)} \qquad \|A - B\|_F^2 \leq \sum_{j=1}^{n} |\lambda_j^{\downarrow}(A) - \lambda_j^{\uparrow}(B)|^2.$$

There is another way of proving Proposition 2 that Löwner would have known. In 1923, Issai Schur, the adviser for Wielandt's Ph.D. thesis at Berlin, proved a very interesting relation between the diagonal of a Hermitian matrix and its eigenvalues. This says that if $d = (d_1, \ldots, d_n)$ and $\lambda = (\lambda_1, \ldots, \lambda_n)$ are, respectively, the diagonal entries and the eigenvalues of a Hermitian matrix A, then d is *majorised* by λ. This, by definition, means that

$$\text{(12)} \qquad \sum_{j=1}^{k} d_j^{\downarrow} \leq \sum_{j=1}^{k} \lambda_j^{\downarrow}, \quad \text{for} \quad 1 \leq k \leq n$$

and

$$\text{(13)} \qquad \sum_{j=1}^{n} d_j^{\downarrow} = \sum_{j=1}^{n} \lambda_j^{\downarrow}.$$

The notation $d \prec \lambda$ is used to express that all of the relations (12) and (13) hold. Schur's theorem has been generalized in various directions (see, e.g., the work of Kostant [K] and Atiyah [A]), and it provided a strong stimulus for the theory of majorisation [MO, p4].

A good part of this theory had been developed by the time Hardy, Littlewood, and Pólya wrote their famous book [HLP] in 1934, the same year as that of Löwner's paper. The condition $d \prec \lambda$ is equivalent to the condition that the vector d is in the convex hull of the vectors λ_σ whose coordinates are permutations of the coordinates of λ.

Schur's theorem leads to an easy proof of (9). We can apply a unitary similarity and assume that A is diagonal, and its diagonal entries are $\lambda_j^{\downarrow}(A)$, $1 \leq j \leq n$. Then

$$\mathrm{tr}\, AB = \sum_{j=1}^{n} \lambda_j^{\downarrow}(A) d_j(B) = \langle \mathrm{Eig}^{\downarrow}(A), d(B)\rangle$$

where $d(B) = (d_1(B), \ldots, d_n(B))$ is the diagonal of B. By Schur's theorem, this vector is in the convex set Ω whose vertices are $\lambda_\sigma(B)$. On this set the function $f(\omega) = \sum_{j=1}^{n} \lambda_j^{\downarrow}(A)\omega_j$ is affine, and hence attains its maximum and minimum on vertices of Ω. Now the inequalities (9) follow from (8).

The ideas occurring in this proof are extremely close to those in the paper of Hoffman and Wielandt. I now give their argument in a simpler version due to Ludwig Elsner.

A matrix S is said to be *doubly stochastic* if its entries s_{ij} are nonnegative, $\sum_{j=1}^{n} s_{ij} = 1$, and $\sum_{i=1}^{n} s_{ij} = 1$. The set Ω consisting of $n \times n$ doubly stochastic matrices is convex. A famous theorem, attributed to Garrett Birkhoff [B], says that the vertices of Ω are the permutation matrices.

Now let A and B be normal matrices, and choose unitary matrices U and V such that $UAU^* = D_1$, and $VBV^* = D_2$, where D_1 and D_2 are diagonal matrices whose diagonal entries are $\alpha_1, \ldots, \alpha_n$, and $\beta_1, \ldots, \beta_n$, respectively. Then

$$(14) \quad \|A - B\|_F^2 = \|U^*D_1U - V^*D_2V\|_F^2 = \|D_1W - WD_2\|_F^2,$$

where $W = UV^*$ is another unitary matrix. The second equality in (14) is a consequence of the fact that the Frobenius norm is unitarily invariant; i.e., that $\|XTY\|_F = \|T\|_F$, for all T, and all unitary X, Y. If the matrix W has entries w_{ij}, then the equality (14) can be expressed as

$$\|A - B\|_F^2 = \sum_{i,j=1}^{n} |\alpha_i - \beta_j|^2 \, |w_{ij}|^2.$$

The matrix $(|w_{ij}|^2)$ is doubly stochastic, and the function $f(S) = \sum_{i,j} |\alpha_i - \beta_j|^2 \, s_{ij}$ on the set Ω consisting of doubly stochastic matrices is an affine function. So, the minimum of f is attained at one of the vertices of Ω, and by Birkhoff's theorem this vertex is a permutation matrix $P = (p_{ij})$. Thus

$$\|A - B\|_F^2 \geq \sum_{i,j=1}^{n} |\alpha_i - \beta_j|^2 \, p_{ij}.$$

If the matrix P corresponds to the permutation σ, then this inequality says that

$$\|A - B\|_F^2 \geq \sum_{i=1}^{n} |\alpha_i - \beta_{\sigma(i)}|^2.$$

This is exactly the Hoffman-Wielandt inequality (7).

Let me interject here that ideas very similar to these lead to a quick proof of Schur's theorem about the diagonal. Let A be a Hermitian matrix and let $A = U\Lambda U^*$ be its spectral representation, where Λ is a diagonal matrix. If d and λ are the vectors corresponding to the diagonals of A and Λ, respectively, then we have $d = S\lambda$, where S is the matrix with entries $s_{ij} = |u_{ij}|^2$. This matrix is doubly stochastic. Hence, we have $d \prec \lambda$.

Now let us return to inequality (4) involving operator norms, the thing Wielandt and Hoffman wanted. Apart from Hermitians, dealt with by Weyl's result (1), there is another equally important subclass of normal matrices: the unitary matrices. Thirty years after [HW], R. Bhatia and C. Davis [BD] proved that the inequality (4) is true when A and B are unitary. There were other papers a little earlier proving the inequality in special cases. One by this author [B1] showed that (4) is true when not only A and B but also $A - B$ are normal. The case of Hermitian A, B is included in this. V. S. Sunder [S] proved the inequality when A is Hermitian and B skew-Hermitian. In 1983 R. Bhatia, C. Davis, and A. McIntosh [BDM] proved that there exists a number c such that for all normal matrices A and B (of any size n) we have

$$(15) \quad d(\text{Eig } A, \text{Eig } B) \leq c\, \|A - B\|.$$

A few years later R. Bhatia, C. Davis, and P. Koosis [BDK] showed that this number c is no bigger than 3. Thus it came to be believed, more strongly than before, that the inequality (4) is very likely true, in general, for normal A and B.

Surprise: in 1992 J. Holbrook [H] published an example of two 3×3 normal matrices A and B for which $d(\text{Eig } A, \text{Eig } B) > \|A - B\|$. (When $n = 2$, this is not possible.) Holbrook found his example by a directed computer search.

As a sidelight, I should mention that a namesake of Wielandt, Helmut Wittmeyer [Wit], claimed that he had proved (4) for all normal A, B. For a proof he referred the reader to his Ph.D. thesis at the Technical University, Darmstadt, written in 1935, the same year as Wielandt's. There is no mention of this in Wielandt's papers, and so he must have been unaware of Wittmeyer's claim.

Hoffman mentions, without any detail, that Wielandt had something "quite lovely, involving a path in matrix space." An argument using paths in the space of normal matrices was discovered by this author [B1]. This led to some new results and some new proofs. It also raises an intriguing problem in differential geometry. Let me explain these ideas.

Though the inequality (4) fails to hold "globally," it is true "locally" in a small neighbourhood of a normal matrix A, even when B is not normal. More precisely, we have the following.

THEOREM 3 *Let A be a normal matrix, and B any matrix such that $\|A - B\|$ is smaller than half the distance between each pair of distinct eigenvalues of A. Then $d(\text{Eig } A, \text{Eig } B) \leq \|A - B\|$.*

Proof. Let $\varepsilon = \|A - B\|$. First I show that any eigenvalue β of B is within distance ε of some eigenvalue α_j of A. By applying a translation, we may assume that $\beta = 0$. If no eigenvalue of A is within a distance ε of this, then A is invertible. Since A is normal, we have $\|A^{-1}\| = 1/\min|\alpha_j| < 1/\varepsilon$. Hence

$$\|A^{-1}(B - A)\| \leq \|A^{-1}\| \, \|B - A\| < 1.$$

This means that $I + A^{-1}(B - A)$ is invertible, and so is $A(I + A^{-1}(B - A)) = B$. But then $\beta = 0$ could not have been an eigenvalue of B, and we have a contradiction.

Now let $\alpha_1, \ldots, \alpha_k$ be all the distinct eigenvalues of A, and let D_j be the closed disk with centre α_j and radius $\varepsilon = \|A - B\|$. By the hypothesis of the theorem, the disks D_j, $1 \leq j \leq k$, are disjoint. By what we have seen above, all the eigenvalues of B lie in the union of these k disks. The rest of the proof consists of showing that if the eigenvalue α_j has multiplicity m_j, then the disk D_j contains exactly m_j eigenvalues of B counted with their respective multiplicities. (It is clear that this implies the theorem.)

Let $A(t) = (1 - t)A + tB$, $0 \leq t \leq 1$, be the straight line segment joining A and B. Then $\|A - A(t)\| = t\varepsilon$, and so all eigenvalues of $A(t)$ also lie in the disks D_j. By a well-known continuity principle, as t moves from 0 to 1 the eigenvalues of $A(t)$ trace continuous curves starting at the eigenvalues of A and ending at those of B. None of these curves can jump from one of the disks D_j to another. So if we start with m_j such curves in D_j, then we must end up with exactly as many. This proves the theorem. □

You may recognize the reasoning in the second part of the proof above as an idea much used in complex analysis around the Argument Principle.

Can the local estimate of Theorem 3 be extended to a global one? Let **N** be the set of all normal matrices of a fixed size n. If A is in **N**, then so is tA for any real t. Thus **N** is a path-connected set. Let $\gamma(t)$, $0 \leq t \leq 1$, be a continuous curve in **N**, and let $\gamma(0) = A$, $\gamma(1) = B$. We say γ is a *normal path* joining A and B. The *length* of γ with respect to the norm $\|\cdot\|$ is defined, as usual, by

$$\ell_{\|\cdot\|}(\gamma) = \sup \sum_{k=0}^{m-1} \|\gamma(t_{k+1}) - \gamma(t_k)\|,$$

where the supremum is taken over all partitions of [0, 1] as $0 = t_0 < t_1 < \cdots < t_m = 1$. If this length is finite, γ is said to be *rectifiable*. If $\gamma(t)$ is a piecewise C^1 function, then

$$\ell_{\|\cdot\|}(\gamma) = \int_0^1 \|\gamma'(t)\| \, dt.$$

From Theorem 3 it is not difficult to obtain, using familiar ideas in differential geometry, the following.

THEOREM 4 *Let A and B be normal matrices, and let γ be a rectifiable normal path joining A and B. Then*

$$d(\text{Eig } A, \text{Eig } B) \leq \ell_{\|\cdot\|}(\gamma). \tag{16}$$

If we could find the length of the shortest normal path joining A and B, then (16) would give a good estimate for $d(\text{Eig A, Eig } B)$. The set **N** does not have an easily tractable geometric structure, and the norm $\|\cdot\|$ is not Euclidean. So we are dealing here with non-Riemannian geometry (Finsler geometry) of a complicated set. Nevertheless, interesting information can be extracted from (16).

In a variety of special cases Theorem 4 leads to the inequality (4). For example, this works when A and B lie in a "flat" part of **N**. By this I mean that the entire line segment $\gamma(t) = (1 - t)A + tB$ is in **N**. A small calculation shows that this is the case if and only if A, B, and $A - B$ are normal—in particular, when A and B are Hermitian.

Much more interesting is the fact that there are sets in **N** that are not affine but are "metrically flat." We say that a subset **S** of **N** is *metrically flat* if any two points A and B of **S** can be joined by a path γ that lies entirely within **S** and has length $\|A - B\|$. An interesting example is given by the following theorem.

THEOREM 5 *Let **S** consist of all $n \times n$ matrices of the form zU where z is a complex number and U is a unitary matrix. Then **S** is a metrically flat subset of **N**.*

Proof. Any two elements of **S** can be represented as $A_0 = r_0 U_0$ and $A_1 = r_1 U_1$, where r_0 and r_1 are nonnegative real numbers. Choose an orthonormal basis in which the unitary matrix $U_1 U_0^{-1}$ is diagonal:

$$U_1 U_0^{-1} = \text{diag}(e^{i\theta_1}, \ldots, e^{i\theta_n}),$$

where

$$|\theta_n| \leq \cdots \leq |\theta_1| \leq \pi.$$

Let $K = \text{diag}(i\theta_1, \ldots, i\theta_n)$. Then K is a skew-Hermitian matrix whose eigenvalues are in $(-i\pi, i\pi]$. We have

$$\begin{aligned}\|A_0 - A_1\| = \|r_0 U_0 - r_1 U_1\| &= \|r_0 I - r_1 U_1 U_0^{-1}\| \\ &= \max_j |r_0 - r_1 \exp(i\theta_j)| = |r_0 - r_1 \exp(i\theta_1)|.\end{aligned}$$

This last quantity is the length of the straight line joining the points r_0 and $r_1 \exp(i\theta_1)$. *If* $|\theta_1| < \pi$, this line segment can be parametrised as $r(t)\exp(it\theta_1)$, $0 \leq t \leq 1$. The equation above can then be expressed as

$$\begin{aligned}\|A_0 - A_1\| &= \int_0^1 |[r(t)\exp(it\theta_1)]'| \, dt \\ &= \int_0^1 |r'(t) + r(t) i\theta_1| \, dt.\end{aligned}$$

Let $A(t) = r(t)e^{tK}U_0$, $0 \leq t \leq 1$. This is a smooth curve in **S** joining A_0 and A_1, and its length is

$$\begin{aligned}\int_0^1 \|A'(t)\| dt &= \int_0^1 \|r'(t)e^{tK}U_0 + r(t)Ke^{tK}U_0\| \, dt \\ &= \int_0^1 \|r'(t)I + r(t)K\| \, dt,\end{aligned}$$

since $e^{tK}U_0$ is unitary. But

$$\|r'(t)I + r(t)K\| = \max_j |r'(t) + ir(t)\theta_j| = |r'(t) + ir(t)\theta_1|.$$

The last three equations show that the path $A(t)$ joining A_0 and A_1 has length $\|A_0 - A_1\|$.

If $|\theta_1| = \pi$, the argument above is not needed. In this case $\|A_0 - A_1\| = |r_0 - r_1 \exp(i\theta_1)| = r_0 + r_1$. This is the length of the piecewise linear path joining A_0 *to* 0 and then to A_1. □

Theorems 4 and 5 together show that the inequality (4) is true when A and B are scalar multiples of unitaries. Theorem 4, in a more general form and with a different proof, was given in [B1]. Theorem 5 was first proved in [BH].

When $n = 2$, the entire set **N** is metrically flat. This can be seen as follows. Let A and B be 2×2 normal matrices. The eigenvalues of A and those of B lie either on two parallel lines, or on two concentric circles. In the first case, we may assume that the lines are parallel to the real axis. Then the skew-Hermitian part of $A - B$ is a scalar, and hence $A - B$ is normal. We have seen that in this case the line segment joining A and B lies in **N**. In the second case, if α is the common centre of the two circles, then A and B are in the set $\alpha I + \mathbf{S}$, which is metrically flat.

Since the inequality (4) is not always true for 3×3 normal matrices, the set **N** must not be metrically flat when $n \geq 3$. I have pointed out some metrically flat subsets of **N**. There may well be others.

An intriguing problem, that seems hard, is that of finding a "curvature constant" for the set **N**. For each n, let $k(n)$ be the smallest number with the following property. Given any two $n \times n$ normal matrices A and B there exists a normal path γ joining them such that

$$\ell_{\|\cdot\|}(\gamma) \leq k(n) \, \|A - B\|.$$

We know that $k(2) = 1$, and $k(3) > 1$. Is the sequence $k(n)$ bounded? If so, is the supremum of $k(n)$ some familiar number like, say, $\pi/2$?

It will be appropriate to end with a related story in which Wielandt played an important role. In 1950, V. B. Lidskii [Li] published a short note in which he gave a matrix-theoretic proof of a theorem that arose in the work of F. Berezin and I. M. Gel'fand on Lie groups. Lidskii's theorem says that if A and B are Hermitian matrices, then the vector $\mathrm{Eig}^{\downarrow}(A) - \mathrm{Eig}^{\downarrow}(B)$ lies in the convex hull of the vectors obtained by permuting the coordinates of $\mathrm{Eig}(A - B)$. In another formulation, it says that for all $0 \leq k \leq n$, and indices $1 \leq i_1 < i_2 < \cdots < i_k \leq n$, we have

$$\sum_{j=1}^{k} \lambda_{i_j}^{\downarrow}(A + B) \leq \sum_{j=1}^{k} \lambda_{i_j}^{\downarrow}(A) + \sum_{j=1}^{k} \lambda_{j}^{\downarrow}(B). \tag{17}$$

Wielandt [Wie2] discovered a remarkable maximum principle from which he derived these inequalities as he "did not succeed in completing the interesting sketch of a proof given by Lidskii."

The inequalities (1) and (10) of Weyl and Löwner are subsumed in (a corollary of) Lidskii's theorem. A norm $|||\cdot|||$ on matrices is said to be *unitarily invariant* if $|||UAV||| = |||A|||$ for all unitary matrices U and V. The operator norm (1) and the Frobenius norm (5) have this property. It follows from Lidskii's theorem that if A and B are Hermitian matrices, then

$$d_{|||\cdot|||}(\mathrm{Eig}\, A, \mathrm{Eig}\, B) \leq |||A - B||| \tag{18}$$

for every unitarily invariant norm.

Fascinated by the inequalities (17), several mathematicians discovered more such relations. This led to a conjecture by Alfred Horn in 1962 specifying *all* possible linear inequalities between eigenvalues of Hermitian matrices A, B, and $A + B$. Horn's conjecture was proved towards the end of the twentieth century by Alexander Klyachko, and Alan Knutson and Terence Tao. In the intervening years it was realised that the problem has ramifications across several major areas of mathematics. The interested reader can find more about this from the expository articles [B3], [F], [KT].

REFERENCES

[A] M. F. Atiyah, Convexity and commuting Hamiltonians, *Bull. London Math. Soc.* 14 (1982), 1–15.

[B1] R. Bhatia, Analysis of spectral variation and some inequalities, *Trans. Amer. Math. Soc.* 272 (1982), 323–332.

[B2] R. Bhatia, *Matrix Analysis*, Springer-Verlag, 1997.

[B3] R. Bhatia, Linear algebra to quantum cohomology: the story of Alfred Horn's inequalities, *Amer. Math. Monthly* 108 (2001), 289–318.

[BD] R. Bhatia and C. Davis, A bound for the spectral variation of a unitary operator, *Lin. Multilin. Algebra* 15 (1984), 71–76.

[BDK] 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.

[BDM] R. Bhatia, C. Davis, and A. McIntosh, Perturbation of spectral subspaces and solution of linear operator equations, *Lin. Alg. Appl.* 52/53 (1983), 45–67.

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

[B] G. Birkhoff, Tres observaciones sobre el algebra lineal, *Univ. Nac. Ucumán Rev.* Ser. A, 5 (1946), 147–151.

[F] W. Fulton, Eigenvalues, invariant factors, highest weights, and Schubert calculus, *Bull. Amer. Math. Soc.* 37 (2000), 209–249.

[HLP] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, Cambridge University Press, 1934.

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

[H] J. Holbrook, Spectral variation of normal matrices, *Lin. Alg. Appl.* 174 (1992), 131–144.

[KT] A. Knutson and T. Tao, Honeycombs and sums of Hermitian matrices, *Notices Amer. Math. Soc.* 48 (2001), 175–186.

[K] B. Kostant, On convexity, the Weyl group and the Iwasawa decomposition, *Ann. Sci. E. N. S.* 6 (1973), 413–455.

[Li] V. B. Lidskii, The proper values of the sum and product of symmetric matrices, *Dokl. Akad. Nauk SSSR* 75 (1950), 769–772.

[Lo] K. Löwner, Über monotone Matrixfunctionen, *Math. Z.* 38 (1934), 177–216.

[MO] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorisation and Its Applications*, Academic Press, 1979.

[S] V. S. Sunder, Distance between normal operators, *Proc. Amer. Math. Soc.* 84 (1982), 483–484.

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

[Wie1] H. Wielandt, *Mathematical Works*, B. Huppert and H. Schneider eds., W. de Gruyter, 1996.

[Wie2] H. W. Wielandt, An extremum property of sums of eigenvalues, *Proc. Amer. Math. Soc.* 6 (1955), 106–110.

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

[Wit] H. Wittmeyer, Einfluss der Änderung einer Matrix . . . , *Zeit. Angew. Math. Mech.* 16 (1936), 287–300.