Variation of Grassman Powers and Spectra

Rajendra Bhatia
School of Mathematics
Tata Institute of Fundamental Research
Homi Bhabha Road
Bombay 400 005, India

and

Shmuel Friedland
Institute of Mathematics
The Hebrew University
Jerusalem, Israel

Submitted by Chandler Davis

ABSTRACT

The operator norm of the derivative of the map which takes a finite-dimensional linear operator to its kth Grassman power (the kth compound) is evaluated. This leads to a bound for the distance between the Grassman powers of two operators. As an important application, a bound for the distance between the eigenvalues of two operators is obtained.

1. INTRODUCTION

Our principal aim in this article is to obtain an explicit quantitative estimate of the variation of one of the basic objects of multilinear algebra, known variously as the *k th Grassman power*, the *k th exterior power*, and the *k th compound* of a linear operator.

Let $A \to \Lambda^k A$ be the map which takes a linear operator A on an n-dimensional vector space V to its kth exterior power $\Lambda^k A$, which is a linear operator on the $\binom{n}{k}$-dimensional vector space $\Lambda^k V$. We will evaluate the operator norm of the derivative of this map. Using this, we can then estimate the distance $\|\Lambda^k A - \Lambda^k B\|$ in terms of $\|A - B\|$.

52 Vanderbilt Ave., New York, NY 10017 0024-3795/81/0600001+18$02.50

As an important application of this, we derive an estimate for the distance between the eigenvalues of A and those of B in terms of $\|A-B\|$, where the norm $\|\cdot\|$ is the operator norm.

We shall use some standard facts from multilinear algebra and calculus in Banach spaces. A good reference for the former is Marcus [6]; for the latter, Dieudonné [3]. For the convenience of the reader and to establish notation, we enumerate some of these facts in Section 2.

2. PRELIMINARIES

Let V be a complex n-dimensional Hilbert space with the usual Euclidean norm $\|\cdot\|$. The set of all endomorphisms of V will be denoted by $\mathfrak{S}(V)$, while the set of all $n\times n$ (complex) matrices will be denoted by $M(n)$. Whenever we find it convenient (and we almost always do in this paper), we identify $\mathfrak{S}(V)$ with $M(n)$. The set of all unitary operators on V is denoted by $U(n)$. For an element A of $\mathfrak{S}(V)$, define its *Banach norm*, *operator norm*, or *spectral norm* as

$$\|A\|=\sup\{\|Ax\|:\|x\|=1\}. \tag{1}$$

If A^* denotes the adjoint of the operator A, then A^*A is a positive operator. The positive square roots of the eigenvalues of A^*A are called the *singular values* of A. We arrange the singular values of A as

$$\nu_1(A)\geqslant\nu_2(A)\geqslant\cdots\geqslant\nu_n(A)\geqslant 0. \tag{2}$$

When no confusion is likely to arise from the omission, we shall denote $\nu_i(A)$ simply by ν_i. It is easy to see that

$$\|A\|=\nu_1(A). \tag{3}$$

Let

$$A^+=\operatorname{diag}(\nu_1,\ldots,\nu_n) \tag{4}$$

be the diagonal matrix whose entries are the singular values of A. Then

$$A=UA^+V \qquad \text{for some} \quad U,V\in U(n), \tag{5}$$

$$(PAQ)^+=A^+ \qquad \text{for all} \quad P,Q\in U(n). \tag{6}$$

In particular, we have

$$\|PAQ\|=\|A\| \qquad \text{for all} \quad P,Q\in U(n). \tag{7}$$

We call this last property the *biunitary invariance* of $\|\cdot\|$.

For $1\leqslant k\leqslant n$, let $\Lambda^k V$ denote the kth *exterior power* of V. This is a Hilbert space of dimension $\binom{n}{k}$. The map Λ^k of vector spaces induces a natural map Λ^k from the space of operators $\mathfrak{S}(V)$ to the space of operators $\mathfrak{S}(\Lambda^k V)$. If A is looked upon as a matrix, then $\Lambda^k A$ is its *k th compound*. We regard $\Lambda^k:\mathfrak{S}(V)\to\mathfrak{S}(\Lambda^k V)$ as a map between Banach spaces, where both spaces are equipped with the respective operator norms defined by (1).

Following Marcus and Minc [7], we adopt a compact notation. For $1\leqslant k\leqslant n$, let $Q_{k,n}$ denote the collection of strictly increasing sequences of k integers chosen from $1,2,\dots,n$. In other words,

$$Q_{k,n}=\{\alpha:\alpha=(\alpha_1,\dots,\alpha_k),\ 1\leqslant\alpha_1<\alpha_2<\cdots<\alpha_k\leqslant n\}.$$

Then, $Q_{k,n}$ is a set of cardinality $\binom{n}{k}$. We order $Q_{k,n}$ by the usual *lexicographic ordering*. We shall call elements of $Q_{k,n}$ *k-indices*.

For two k-indices, $\alpha=(\alpha_1,\dots,\alpha_k)$ and $\beta=(\beta_1,\dots,\beta_k)$, we shall denote by $A[\alpha|\beta]$ or by $A[(\alpha_1,\dots,\alpha_k)|(\beta_1,\dots,\beta_k)]$ the submatrix of A constructed from the rows $\alpha_1,\dots,\alpha_k$ and columns $\beta_1,\dots,\beta_k$, i.e., this is the submatrix whose (ij)th entry is $a_{\alpha_i\beta_j}$. The entries of the matrix $\Lambda^k A$ of order $\binom{n}{k}$ can be indexed by pairs α,β chosen from $Q_{k,n}$. The (α,β) entry of $\Lambda^k A$ then is $\det A[\alpha|\beta]$, where det denotes the determinant. We write this as

$$\Lambda^k A=(\det A[\alpha|\beta])_{\alpha,\beta\in Q_{k,n}}. \tag{8}$$

It follows that

$$\Lambda^k(AB)=\Lambda^k(A)\Lambda^k(B), \qquad \Lambda^k(I)=I, \qquad (\Lambda^k(A))^*=\Lambda^k(A^*). \tag{9}$$

Here, I denotes the identity matrix of order n, as well as that of order $\binom{n}{k}$. No confusion should arise from this notational abuse. In particular, this implies that

$$\Lambda^k(U)\in U\left(\binom{n}{k}\right) \qquad \text{if} \quad U\in U(n). \tag{10}$$

If the singular values of A are given by (2), then the singular values of $\Lambda^k A$ are

$$\{\nu_{\alpha_1}\nu_{\alpha_2}\cdots\nu_{\alpha_k}:(\alpha_1,\dots,\alpha_k)\in Q_{k,n}\}.$$

Among other things, this implies that

$$(\Lambda^k A)^+ = \Lambda^k(A^+) \tag{11}$$

and

$$\|\Lambda^k A\| = \prod_{j=1}^{k} \nu_j(A). \tag{12}$$

Finally, the expression (8) shows that the map

$$\Lambda^k : M(n) \to M\left(\binom{n}{k}\right)$$

is holomorphic. The *derivative* of this map at a point A of $M(n)$, denoted as $D\Lambda^k(A)$, is a linear map from $M(n)$ to $M\binom{n}{k}$. It is defined as

$$\big(D\Lambda^k(A)\big)(B) = \frac{d}{dt}\bigg|_{t=0} \Lambda^k(A+tB) \tag{13}$$

for all $B\in M(n)$. This can be thought of as the directional derivative of Λ^k at A in the direction of B. We shall find a bound for the operator norm of $D\Lambda^k(A)$ which is defined as

$$\|D\Lambda^k(A)\| = \sup\{\big(D\Lambda^k(A)\big)(B) : B\in M(N),\ \|B\|=1\}. \tag{14}$$

3. VARIATION OF Λ^k

We shall prove the following theorem:

THEOREM 1. *Let A be an element of $\mathfrak{S}(V)$ with singular values $\nu_1(A)\geqslant \nu_2(A)\geqslant\cdots\geqslant\nu_n(A)\geqslant 0$. Then, for $1\leqslant k\leqslant n$,*

$$\|D\Lambda^k(A)\| = \sum_{p=1}^{k}\prod_{\substack{j=1\\ j\neq p}}^{k} \nu_j(A). \tag{15}$$

In particular, this implies

$$\|D\Lambda^k(A)\| \leqslant k\|A\|^{k-1}, \tag{16}$$

for $1 \leqslant k \leqslant n$.

To localize the confusion caused by the multitude of indices that enter the picture, we split the proof of this theorem into several small lemmas. Also, to avoid terminological excesses, we do not name objects that we construct, use, and discard. So statements of lemmas will include definitions of objects they deal with.

LEMMA 1. *Let* A^+ *be as defined by* (4). *Then*

$$\|D\Lambda^k(A)\| = \|D\Lambda^k(A^+)\|.$$

Proof. Let U,V be unitary matrices such that $A = UA^+V$. Let $P = \Lambda^k U$, $Q = \Lambda^k V$. Then P, Q are unitary matrices too. So for all $B \in M(n)$, we have

$$\begin{aligned}
\|(D\Lambda^k(A))(B)\| &= \|P((D\Lambda^k(A))(B))Q\| \\
&= \left\|\Lambda^k U\left(\frac{d}{dt}\bigg|_{t=0}\Lambda^k(A+tB)\right)\Lambda^k V\right\| \\
&= \left\|\frac{d}{dt}\bigg|_{t=0}\Lambda^k U\cdot\Lambda^k(A+tB)\cdot\Lambda^k V\right\| \\
&= \left\|\frac{d}{dt}\bigg|_{t=0}\Lambda^k(A^+ + tUBV)\right\| \\
&= \|(D\Lambda^k(A^+))(UBV)\|.
\end{aligned}$$

Now, $\|UBV\| = \|B\|$, and the set $\{UBV: \|B\| = 1\}$ includes all matrices of norm 1. So we have

$$\begin{aligned}
\|D\Lambda^k(A)\| &= \sup_{\|B\|=1}\|(D\Lambda^k(A))(B)\| \\
&= \sup_{\|B\|=1}\|(D\Lambda^k(A^+))(UBV)\| \\
&= \|D\Lambda^k(A^+)\|.
\end{aligned}$$

■

To state our next lemma concisely, we adopt some special notations: Let $\alpha=(\alpha_1,\dots,\alpha_k)$ and $w=(w_1,\dots,w_{k-1})$ be elements of $Q_{k,n}$ and $Q_{k-1,n}$ respectively. If α can be obtained by adjoining an entry p to w, i.e., if there exists p such that $\{\alpha_1,\dots,\alpha_k\}=\{w_1,\dots,w_{k-1}\}\cup\{p\}$, then we shall write $\alpha=\overline{w,p}$. Note that given an $\alpha\in Q_{k,n}$ we can find k elements w of $Q_{k-1,n}$ for which $\alpha=\overline{w,p}$ for some p; and given a $w\in Q_{k-1,n}$ there are $n-(k-1)$ elements α of $Q_{k,n}$ such that $\alpha=\overline{w,p}$ for some p. For $w\in Q_{k-1,n}$ and $p\in\{1,2,\dots,n\}\backslash\{w_1,\dots,w_{k-1}\}$ we define the signature function $\varepsilon_w(p)$ as

$$\varepsilon_w(p)=1 \qquad \text{if an odd number of entries of } w \text{ are less than } p,$$

$$\varepsilon_w(p)=-1 \qquad \text{otherwise.}$$

With this notation we have:

LEMMA 2. *Let $A=\operatorname{diag}(\nu_1,\dots,\nu_n)$ be a diagonal matrix. Then for any $B\in M(n)$,*

$$\left.\frac{d}{dt}\right|_{t=0}\det(A+tB)[\alpha|\beta]=0 \tag{17}$$

for all pairs of indices $\alpha,\beta\in Q_{k,n}$ except those which have the form

$$\alpha=\overline{w,p}\,,\qquad \beta=\overline{w,q} \tag{18}$$

for some $w\in Q_{k-1,n}$.

If α,β are of the form (18), *and if $\alpha\neq\beta$ (i.e. $p\neq q$), then*

$$\left.\frac{d}{dt}\right|_{t=0}\det(A+tB)[\alpha|\beta]=\varepsilon_w(p)\varepsilon_w(q)b_{pq}\left(\prod_{i=1}^{k-1}\nu_{w_i}\right). \tag{19}$$

If $\alpha=\beta$, i.e., α,β are of the form (18) *with $p=q$, then*

$$\left.\frac{d}{dt}\right|_{t=0}\det(A+tB)[\alpha|\alpha]=\sum_{j=1}^{k}b_{\alpha_j\alpha_j}\left(\prod_{\substack{i=1\\ i\neq j}}^{k}\nu_{\alpha_i}\right). \tag{20}$$

Proof. Let $\alpha=(\alpha_1,\ldots,\alpha_k)$, $\beta=(\beta_1,\ldots,\beta_k)$ be any two elements of $Q_{k,n}$. By the rule for differentiating a determinant we have

$$\frac{d}{dt}\bigg|_{t=0}\det(A+tB)[\alpha|\beta]$$

$$=\sum_{i,j=1}^{k}(-1)^{i+j}b_{\alpha_i\beta_j}\det A\big[(\alpha_1,\ldots,\hat{\alpha}_i,\ldots,\alpha_k)|(\beta_1,\ldots,\hat{\beta}_j,\ldots,\beta_k)\big], \tag{21}$$

where the circumflex indicates that the index under it has been omitted. Now note that if $A=\operatorname{diag}(\nu_1,\ldots,\nu_n)$ then

$$\det A[\alpha|\beta]=0 \qquad \text{for all} \quad \alpha,\beta\in Q_{k,n} \quad \text{such that} \quad \alpha\neq\beta,$$

$$\det A[\alpha|\alpha]=\prod_{i=1}^{k}\nu_{\alpha_i} \qquad \text{for all} \quad \alpha\in Q_{k,n},$$

for all $1\leqslant k\leqslant n$. So if α,β do not satisfy (18), all the determinants on the right-hand side of (21) vanish. Thus, in this case, (17) holds. If α,β satisfy (18) with $p\neq q$, then we can have $(\alpha_1,\ldots,\hat{\alpha}_i,\ldots,\alpha_k)=(\beta_1,\ldots,\hat{\beta}_j,\ldots,\beta_k)$ if and only if $\alpha_i=p$, $\beta_j=q$. So, in this case, (21) leads to (19). If $\alpha=\beta$ then $(\alpha_1,\ldots,\hat{\alpha}_i,\ldots,\alpha_k)=(\beta_1,\ldots,\hat{\beta}_j,\ldots,\beta_k)$ if and only if $i=j$. In this case, (21) leads to (20). ■

LEMMA 3. *Let $A=\operatorname{diag}(\nu_1,\ldots,\nu_n)$ be a diagonal matrix, and let $B\in M(n)$ be arbitrary. For each element w of $Q_{k-1,n}$ define a matrix $B^{(w)}(A)$ of order $\binom{n}{k}$ as follows. The entries $b^{(w)}_{\alpha,\beta}(A)$ of $B^{(w)}(A)$ are indexed by $\alpha,\beta\in Q_{k,n}$ and are defined as*

$$b^{(w)}_{\alpha,\beta}(A)=0 \qquad \text{if} \quad \alpha,\beta \text{ do not satisfy (18)} \tag{22}$$

and

$$b^{(w)}_{\alpha,\beta}(A)=\varepsilon_w(p)\varepsilon_w(q)b_{pq}\left(\prod_{i=1}^{k-1}\nu_{w_i}\right) \qquad \text{if} \quad \alpha,\beta \text{ satisfy (18)}. \tag{23}$$

Then

$$\big(D\Lambda^k(A)\big)(B)=\sum_{w\in Q_{k-1,n}}B^{(w)}(A). \tag{24}$$

Proof. Both sides of (24) are matrices of order $\binom{n}{k}$. The (α, β)th entry of $(D\Lambda^k(A))\,(B)$ is given by

$$\frac{d}{dt}\bigg|_{t=0} \det(A+tB)[\alpha|\beta], \qquad \alpha, \beta \in Q_{k,n}.$$

If α, β do not satisfy (18) for any w, then this is zero by Lemma 2, while the (α, β)th entry of the right-hand side is zero by definition. Let α, β satisfy (18) and let $\alpha \neq \beta$. Note that the w occurring in (18) is unique in this case. The (α, β)th entries of the two sides of (24) agree because of (19) and (23). Finally, let $\alpha = \beta$. In this case α and β satisfy (18) for exactly k different choices of p, q, and w, viz.

$$p = q = \alpha_j,$$

$$w = (\alpha_1, \ldots, \hat{\alpha}_j, \ldots, \alpha_k),$$

$j = 1, 2, \ldots, k$. Once again, the (α, β)th entries of the two sides of (24) agree in view of (20) and (23). ■

LEMMA 4. *Let $B \in M(n)$, $B = ((b_{ij}))$. For $w \in Q_{k-1,n}$ define a matrix $\tilde{B}$ of order $n-(k-1)$ as follows: The entries $\tilde{b}_{pq}$ of $\tilde{B}$ are indexed by $p, q \in \{1, 2, \ldots, n\} \setminus \{w_1, \ldots, w_{k-1}\}$ and are defined as*

$$\tilde{b}_{pq} = \varepsilon_w(p)\varepsilon_w(q) b_{pq}.$$

Then

$$\|\tilde{B}\| \leqslant \|B\|.$$

Proof. Let $\{e_1, \ldots, e_n\}$ be an orthonormal basis for V, with respect to which B is identified as an operator on V. Let P be the projection operator in V corresponding to the subspace spanned by the basis vectors e_p, $p \notin \{w_1, \ldots, w_{k-1}\}$. Let D be the diagonal operator with matrix $(d_1, \ldots, d_n)$ such that

$$d_p = \begin{cases} \varepsilon_w(p) & \text{if } \; p \notin \{w_1, \ldots, w_{k-1}\}, \\ 0 & \text{if } \; p \in \{w_1, \ldots, w_{k-1}\} \end{cases}.$$

Then there exists a unitary operator U on V such that

$$\tilde{B}\oplus 0 = UDPBPDU^{-1}.$$

Hence

$$\|\tilde{B}\| \leqslant \|B\|. \qquad \blacksquare$$

Let $w\in Q_{k-1,n}$. For a vector x in $\Lambda^k V$ with components x_α, $\alpha\in Q_{k,n}$, define

$$\|x\|_w = \left(\sum_{\{\alpha:\alpha=\overline{w,q}\}} |x_\alpha|^2\right)^{1/2}.$$

Here, the summation is over all α which can be obtained from the given $(k-1)$-index w by adjoining an entry q to it. We shall write this briefly as

$$\|x\|_w = \left(\sum_q |x_{\overline{w,q}}|^2\right)^{1/2} \tag{25}$$

with the understanding that the summation is over all

$$q\in\{1,2,\ldots,n\}\backslash\{w_1,\ldots,w_{k-1}\}.$$

LEMMA 5. *Let $w\in Q_{k-1,n}$, and let A, B, $B^{(w)}(A)$ be as in Lemma 3. Then for any $x,y\in\Lambda^k V$ we have*

$$|\langle B^{(w)}(A)x,y\rangle| \leqslant \prod_{i=1}^{k-1}|\nu_{w_i}|\,\|B\|\,\|x\|_w\|y\|_w, \tag{26}$$

where $\langle\cdot,\cdot\rangle$ denotes the inner product in $\Lambda^k V$ and $\|\cdot\|_w$ is defined by (25).

Proof. On using the definition of $B^{(w)}(A)$ given in Lemma 3 we obtain

$$\begin{aligned}\langle B^{(w)}(A)x,y\rangle &= \sum_{\alpha,\beta\in Q_{k,n}} b_{\alpha,\beta}^{(w)}(A)x_\beta\bar{y}_\alpha\\ &= \left(\prod_{i=1}^{k-1}\nu_{w_i}\right)\sum_{p,q}\varepsilon_w(p)\varepsilon_w(q)b_{pq}x_{\overline{w,q}}\bar{y}_{\overline{w,p}},\end{aligned}$$

where in the last summation p, q vary over $\{1,2,\dots,n\}\setminus\{w_1,\dots,w_{k-1}\}$. If $\tilde{B}$ is the matrix defined in Lemma 4, this gives

$$\left|\langle B^{(w)}(A)x, y\rangle\right| \leqslant \prod_{i=1}^{k-1} |\nu_{w_i}| \, \|\tilde{B}\| \, \|x\|_w \|y\|_w.$$

Lemma 4 then leads to (26). ∎

Proof of Theorem 1. By Lemma 1, we can assume, without loss of generality, that $A = \operatorname{diag}(\nu_1,\dots,\nu_n)$, $\nu_1 \geqslant \nu_2 \geqslant \cdots \geqslant \nu_n \geqslant 0$. Let $B \in M(n)$, and let x, y be any two vectors in $\Lambda^k V$. Let $B^{(w)}(A)$ be defined as in Lemma 3. Then by Lemma 5 and the Cauchy-Schwarz inequality we have

$$\begin{aligned}\left|\left\langle \sum_w B^{(w)}(A)x, y\right\rangle\right| &\leqslant \|B\| \sum_w \left(\prod_{i=1}^{k-1} \nu_{w_i}\right) \|x\|_w \|y\|_w \\ &\leqslant \|B\| \left[\sum_w \left(\prod_{i=1}^{k-1} \nu_{w_i}\right) \|x\|_w^2\right]^{1/2} \left[\sum_w \left(\prod_{i=1}^{k-1} \nu_{w_i}\right) \|y\|_w^2\right]^{1/2},\end{aligned} \tag{27}$$

where in all the sums w varies over $Q_{k-1,n}$. We now claim that

$$\sum_{w \in Q_{k-1,n}} \left(\prod_{i=1}^{k-1} \nu_{w_i}\right) \|x\|_w^2 = \sum_{\alpha \in Q_{k,n}} \left(\sum_{p=1}^{k} \prod_{\substack{i=1\\ i\neq p}}^{k} \nu_{\alpha_i}\right) |x_\alpha|^2. \tag{28}$$

Recall that each element α of $Q_{k,n}$ can be written in exactly k different ways as $\alpha = w, p$, namely, by choosing $p = \alpha_i$ and $w = (\alpha_1,\dots,\hat{\alpha}_i,\dots,\alpha_k)$, $i = 1,2,\dots,k$. This fact, together with the definition of $\|x\|_w$ given by (25), leads to (28). Since $\nu_1 \geqslant \nu_2 \geqslant \cdots \geqslant \nu_n \geqslant 0$, we have, for any $\alpha \in Q_{k,n}$, the inequality

$$\sum_{p=1}^{k} \prod_{\substack{i=1\\ i\neq p}}^{k} \nu_{\alpha_i} \leqslant \sum_{p=1}^{k} \prod_{\substack{i=1\\ i\neq p}}^{k} \nu_i. \tag{29}$$

If $\|x\|$ is defined, as usual, as $\|x\|^2 = \sum_\alpha |x_\alpha|^2$, then (28) and (29) give

$$\sum_{w \in Q_{k-1,n}} \left(\prod_{i=1}^{k-1} \nu_{w_i}\right) \|x\|_w^2 \leqslant \left(\sum_{p=1}^{k} \prod_{\substack{i=1\\ i\neq p}}^{k} \nu_i\right) \|x\|^2.$$

Using this estimate and (27) we have

$$\left|\left\langle \sum_w B^{(w)}(A)x, y \right\rangle\right| \leqslant \|B\| \left(\sum_{p=1}^{k} \prod_{\substack{i=1 \\ i\neq p}}^{k} \nu_i \right) \|x\|\,\|y\|.$$

Hence,

$$\left\| \sum_w B^{(w)}(A) \right\| \leqslant \|B\| \left(\sum_{p=1}^{k} \prod_{\substack{i=1 \\ i\neq p}}^{k} \nu_i \right).$$

But by Lemma 3, $\Sigma_w B^{(w)}(A) = D\Lambda^k(A)(B)$. Hence,

$$\|D\Lambda^k(A)\| = \sup_{\|B\|=1} \|D\Lambda^k(A)(B)\| \leqslant \sum_{p=1}^{k} \prod_{\substack{i=1 \\ i\neq p}}^{k} \nu_i. \tag{30}$$

We will show that the last inequality is actually an equality by showing that the supremum in question is attained at $B=I$. Since A is diagonal, $\Lambda^k(A+tI)$ is also diagonal, and its entries are

$$\det(A+tI)[\alpha|\alpha] = \prod_{i=1}^{k} (\nu_{\alpha_i} + t), \qquad \alpha \in Q_{k,n}.$$

Thus, $(D\Lambda^k(A))(I)$ is also a diagonal matrix, with entries

$$d_{\alpha,\alpha} = \left.\frac{d}{dt}\right|_{t=0} \det(A+tI)[\alpha|\alpha] = \sum_{p=1}^{k} \prod_{\substack{i=1 \\ i\neq p}}^{k} \nu_{\alpha_i}.$$

Since $\nu_1 \geqslant \nu_2 \geqslant \cdots \geqslant \nu_n \geqslant 0$, we have

$$\left\|\left(D\Lambda^k(A)\right)(I)\right\| = \max_{\alpha \in Q_{k,n}} d_{\alpha,\alpha} = \sum_{p=1}^{k} \prod_{\substack{i=1 \\ i\neq p}}^{k} \nu_i.$$

This proves the theorem completely. ■

REMARK. The bound (16) is best possible. It is attained when $\nu_1(A) = \nu_2(A) = \cdots = \nu_k(A)$.

This bound can be obtained in another way as follows. Let

$$\otimes^k V = V \otimes \cdots \otimes V$$

be the k-fold tensor product of V. The space $\Lambda^k V$ can be identified as a subspace of $\otimes^k V$. Indeed, the map P_k defined on product vectors as

$$P_k(x_1 \otimes \cdots \otimes x_k) = \frac{1}{k!} \sum_{\sigma} \operatorname{sgn}(\sigma) x_{\sigma(1)} \otimes \cdots \otimes x_{\sigma(k)},$$

where $\operatorname{sgn}(\sigma)$ is the signature of the permutation σ, extends linearly to all of $\otimes^k V$ as an orthoprojector with range $\Lambda^k V$. We have

$$x_1 \wedge \cdots \wedge x_k = P_k(x_1 \otimes \cdots \otimes x_k).$$

Let $Q_k: \Lambda^k V \to \otimes^k V$ be the inclusion map. Consider the induced map $\tilde{P}_k: \mathfrak{S}(\otimes^k V) \to \mathfrak{S}(\Lambda^k V)$ defined as

$$\tilde{P}_k(T) = P_k T Q_k, \qquad T \in \mathfrak{S}(\otimes^k V).$$

Then P_k is a projection and $\|\tilde{P}_k\| \leqslant 1$. The map $\Lambda^k: \mathfrak{S}(V) \to \mathfrak{S}(\Lambda^k V)$ factors through the map $\otimes^k: \mathfrak{S}(V) \to \mathfrak{S}(\otimes^k V)$ via the projection $\tilde{P}_k$, i.e.,

$$\Lambda^k(A) = P_k(\otimes^k A) \qquad \text{for all} \quad A \in \mathfrak{S}(V).$$

By the chain rule of differentiation, and the fact that the derivative of the linear map $\tilde{P}_k$ is $\tilde{P}_k$ itself, we have

$$D\Lambda^k(A) = \tilde{P}_k \circ D \otimes^k(A).$$

(Note that $D\otimes^k(A)$ is a linear map from $\mathfrak{S}(V)$ to $\mathfrak{S}(\otimes^k V)$, and $D\Lambda^k(A)$ is a linear map from $\mathfrak{S}(V)$ to $\mathfrak{S}(\Lambda^k V)$). Now,

$$\begin{aligned}\big(D\otimes^k(A)\big)(B) &= \left.\frac{d}{dt}\right|_{t=0} \otimes^k(A+tB) \\ &= B \otimes A \otimes \cdots \otimes A + A \otimes B \otimes \cdots \otimes A + \cdots + A \otimes \cdots \otimes A \otimes B.\end{aligned}$$

Since $\|A \otimes B\| = \|A\| \, \|B\|$, this gives

$$\left\|D\otimes^k(A)(B)\right\| \leqslant k \|A\|^{k-1} \|B\|.$$

Taking supremum over B such that $\|B\|=1$, and considering the special case $B=A/\|A\|$, we obtain

$$\left\|D\otimes^k(A)\right\|=k\|A\|^{k-1}.$$

So

$$\left\|D\Lambda^k(A)\right\|\leqslant\left\|\tilde{P}_k\right\|\left\|D\otimes^k(A)\right\|\leqslant k\|A\|^{k-1},$$

which is the estimate (16).

This result has been generalized in [4] in two directions. There it is shown that a bound analogous to (16) is applicable when Λ^k is replaced by any irreducible representation τ of $\mathfrak{S}(V)$ in $\mathfrak{S}(\otimes^k V)$ given by the Young diagram, and that this result applies to infinite-dimensional Banach spaces as well.

Between the writing of the two versions of this article, Sunder has given a different proof of Theorem 1, which is also valid for infinite-dimensional Hilbert spaces. In his proof he uses what has been called Lemma 1 above and the fact that a completely positive map of C^* algebras attains its norm at the identity. Note that in the course of the proof of Theorem 1 we also use the fact that the supremum in (30) is attained at I.

As Corollaries to Theorem 1, we have:

COROLLARY 1. *For any two elements A, B of $\mathfrak{S}(V)$ we have*

$$\left\|\Lambda^kB-\Lambda^kA\right\|\leqslant k\|A\|^{k-1}\|B-A\|+O\left(\|B-A\|^2\right).$$

Proof. This is an application of Taylor's formula. (See 8.14.3 in [3]). ■

More precisely, we have:

COROLLARY 2. *For any two elements A, B of $\mathfrak{S}(V)$,*

$$\left\|\Lambda^kB-\Lambda^kA\right\|\leqslant kM^{k-1}\|B-A\|,$$

where $M=\max(\|A\|,\|B\|)$.

Proof. Consider the map $f:[0,1]\to\mathfrak{S}(V)$ defined as $f(t)=(1-t)A+tB$. This is a linear path joining A and B. By the mean-value theorem (8.5.4 in [3]) applied to the composite map $\Lambda^k \circ f:[0,1] \to \mathfrak{S}(\Lambda^k V)$, we get

$$\begin{aligned}\|\Lambda^k B - \Lambda^k A\| &\leqslant \sup_{0 \leqslant t \leqslant 1} \|D\Lambda^k(f(t))\| \|Df(t)\| \\ &\leqslant \sup_{0 \leqslant t \leqslant 1} k \|(1-t)A + tB\|^{k-1} \|B - A\| \\ &\leqslant kM^{k-1} \|B - A\|. \qquad \blacksquare\end{aligned}$$

4. APPLICATIONS TO THE VARIATION OF SPECTRA

Results of the preceding section can be applied to the study of spectral variation following the ideas introduced by Bhatia and Mukherjea in [2]. Results in this section supplement those in [1] and [2].

For $A, B \in \mathfrak{S}(V)$, let $\operatorname{Eig} A = \{\alpha_1, \dots, \alpha_n\}$ and $\operatorname{Eig} B = \{\beta_1, \dots, \beta_n\}$ denote their respective eigenvalues counted with multiplicity. A distance between these n-tuples can be defined as

$$d(\operatorname{Eig} A, \operatorname{Eig} B) = \min_{\sigma \in \Pi_n} \max_{1 \leqslant i \leqslant n} |\alpha_i - \beta_{\sigma(i)}|, \tag{31}$$

where Π_n is the group of permutations on n symbols. A question of obvious interest and importance is: What is $d(\operatorname{Eig} A, \operatorname{Eig} B)$ in terms of $\|A - B\|$? One answer to this question is provided by estimates derived below.

Let the characteristic polynomial of A be written as

$$\chi_A(t) = t^n - \varphi_1(A) t^{n-1} + \cdots + (-1)^n \varphi_n(A).$$

It is well known that

$$\varphi_k(A) = \operatorname{tr} \Lambda^k(A), \qquad 1 \leqslant k \leqslant n, \tag{32}$$

where tr stands for the trace of an operator. With this notation, we have:

Proposition 1. *For* $1 \leqslant k \leqslant n$,

$$\|D\varphi_k(A)\| \leqslant k \binom{n}{k} \|A\|^{k-1}. \tag{33}$$

Proof. Let $\dim V = n$, and let $(\mathrm{tr})_n : \mathfrak{S}(V) \to \mathbb{C}$ be the trace map. Since this map is linear, we have

$$\|D(\mathrm{tr})_n(A)\| = \|(\mathrm{tr})_n\| = n \qquad \text{for all} \quad A \in \mathfrak{S}(V).$$

Since

$$\dim \Lambda^k(V) = \binom{n}{k},$$

the proposition follows from (32), Theorem 1, and the chain rule of differentiation. ■

REMARK. Using a different method, it was shown in [2] that

$$\|D\varphi_k(A)\| \leqslant k^{1-k/2}\binom{n}{k}\|A\|_F^{k-1}, \tag{34}$$

where the *Frobenius norm* $\|\cdot\|_F$ on $\mathfrak{S}(V)$ is defined as:

$$\|A\|_F = (\mathrm{tr}\, A^*A)^{1/2} = \left(\sum_{i,j} |a_{ij}|^2\right)^{1/2}. \tag{35}$$

We have for all $A \in \mathfrak{S}(V)$

$$\|A\| \leqslant \|A\|_F \leqslant n^{1/2}\|A\|. \tag{36}$$

It is not possible to derive either (33) or (34) from the other using the relation (36).

Applying the mean-value theorem as before, we have:

COROLLARY. *For all* $A, B \in \mathfrak{S}(V)$ *and* $1 \leqslant k \leqslant n$,

$$|\varphi_k(A) - \varphi_k(B)| \leqslant k\binom{n}{k}M^{k-1}\|B - A\|, \tag{37}$$

where $M = \max(\|A\|, \|B\|)$.

This gives us an estimate of the distance between the coefficients of the characteristic polynomials of A and B. From this we can estimate $d(\mathrm{Eig}\, A, \mathrm{Eig}\, B)$, using the celebrated theorem of Ostrowski [8] stated below.

THEOREM (Ostrowski). *Let $f(z)=z^n+a_1z^{n-1}+\cdots+a_n$ and $g(z)=z^n+b_1z^{n-1}+\cdots+b_n$ be two polynomials of degree n. Let the roots of f and g be $\alpha_1,\dots,\alpha_n$ and $\beta_1,\dots,\beta_n$, respectively. Let $\mu_1=\max_k|\alpha_k|$, $\mu_2=\max_k|\beta_k|$, $\mu=\max(\mu_1,\mu_2)$. Let $\theta=\{\sum_{k=1}^n|b_k-a_k|\mu^{n-k}\}^{1/n}$. Then the roots can be arranged in such a way that*

$$|\alpha_k-\beta_k|<(2n-1)\theta \qquad \textit{for all} \quad 1\leqslant k\leqslant n.$$

Using this we obtain

THEOREM 2. *For all $A, B\in\mathfrak{S}(V)$*

$$d(\operatorname{Eig} A,\operatorname{Eig} B)\leqslant 2^{1-1/n}(2n-1)n^{1/n}M^{1-1/n}\|B-A\|^{1/n}, \tag{38}$$

where $M=\max(\|A\|,\|B\|)$.

Proof. Use Ostrowski's theorem and the estimate (37). Notice that the eigenvalues of A are bounded in modulus by $\|A\|$. So we have

$$\begin{aligned} d(\operatorname{Eig} A,\operatorname{Eig} B)&\leqslant(2n-1)\left\{\sum_{k=1}^{n}|\varphi_k(B)-\varphi_k(A)|M^{n-k}\right\}^{1/n}\\ &\leqslant(2n-1)\left\{\sum_{k=1}^{n}k\binom{n}{k}\right\}^{1/n}M^{1-1/n}\|B-A\|^{1/n}. \end{aligned}$$

Now use the combinatorial identity

$$\sum_{k=0}^{n}k\binom{n}{k}=n2^{n-1}$$

to obtain (38). ∎

REMARKS. In [2], Bhatia and Mukherjea derived an estimate for $d(\operatorname{Eig} A,\operatorname{Eig} B)$ in terms of $\|A-B\|_F$. It is more natural to use the Banach norm $\|\cdot\|$ when the distance d between the eigenvalues is defined as in (31). With the Frobenius norm, it is more natural to use the distance δ between eigenvalues, defined as

$$\delta(\operatorname{Eig} A,\operatorname{Eig} B)=\min_{\sigma\in\Pi_n}\left(\sum_i|\alpha_i-\beta_{\sigma(i)}|^2\right)^{1/2}.$$

The distances d and δ are related by an inequality analogous to (36), viz.

$$d(\cdot,\cdot)\leqslant\delta(\cdot,\cdot)\leqslant n^{1/2}d(\cdot,\cdot). \tag{39}$$

So it is possible to switch from one distance and norm to another at the cost of some precision. The estimate (38) in this paper and the one in [2] cannot be obtained from each other in this way.

Other estimates for $d(\operatorname{Eig} A, \operatorname{Eig} B)$ have been obtained by Ostrowski [8] and by Henrici [5]. The advantages of the estimate (38) over their results are the same as those pointed out in [2] for the estimate derived there.

Finally, the analysis done in [1] for the special case when A, B lie in the classical matrix Lie algebras can be carried through verbatim here to obtain special results for these cases. Using that analysis and the combinatorial identity

$$\sum_{k=0}^{r} 2k\binom{n}{2k}=n2^{n-2},$$

where r is the integral part of $n/2$, we obtain:

THEOREM 3. *Let A, B be two matrices of order $n=2r$ or $n=2r+1$. Suppose A, B are elements either of the Lie algebra $\mathfrak{so}(n,\mathbb{C})$ consisting of complex skew-symmetric matrices or of the symplectic Lie algebra $\mathfrak{sp}(r,\mathbb{C})$. Then*

$$d(\operatorname{Eig} A^2, \operatorname{Eig} B^2)\leqslant 2^{(n-2)/r}(2r-1)n^{1/r}M^{2-1/r}\|B-A\|^{1/r}, \tag{40}$$

where $M=\max(\|A\|,\|B\|)$. Further, if A, B both have 0 as one of their eigenvalues with the same multiplicity, and if the rest of their eigenvalues lie outside a circle of radius c around the origin, then, we have

$$d(\operatorname{Eig} A, \operatorname{Eig} B)\leqslant R/c,$$

where R denotes the right-hand side of (40).

REMARKS. The main inequalities of [1] and [2] can be somewhat strengthened and considerably tidied up. Writing our path as $f(t)=(1-t)A+tB$, as we have done here, Proposition 4.1 in [2] becomes much neater, and this, in turn leads to an improvement of the estimate (4.1) in [2], which now reads as

$$d(\operatorname{Eig} A, \operatorname{Eig} B) \leqslant (2n-1) M_F^{1-1/n} \left\{ \sum_{k=1}^{n} k^{1-k/2} \binom{n}{k} \right\}^{1/n} \| B - A \|_F^{1/n}$$
$$< 2(2n-1) M_F^{1-1/n} \| B - A \|_F^{1/n},$$

where $M_F = \max(\|A\|_F, \|B\|_F)$.

In the same way, the bounds (3.1) in [1] can be improved to

$$d(\operatorname{Eig} A^2, \operatorname{Eig} B^2) \leqslant 2^{(n-1)/r} (2r-1) M_F^{2-1/r} \| B - A \|_F^{1/r}.$$

We would like to thank Professor Chandler Davis, whose comments led to improvements in the inequalities of Section 4, and Professor Miroslav Fiedler, who drew our attention to a grievous mistake in an earlier version of this paper. We would like to thank A. Ramanathan and V. Sunder for several discussions on this problem with the first author.

The earlier version of this article was written in the year 1979–80, when the first author was supported by a Fulbright fellowship at the Department of Mathematics, University of California, Berkeley, and when the second author was at the Mathematics Research Center, University of Wisconsin, Madison. We are indebted to these departments for the facilities they provided us for this work.

REFERENCES

1 R. Bhatia, On the rate of change of spectra of operators. II. *Linear Algebra and Appl.* 36:25–32 (1981).
2 R. Bhatia and K. K. Mukherjea, On the rate of change of spectra of operators, *Linear Algebra and Appl.* 27:147–157 (1979).
3 J. Dieudonné, *Foundations of Modern Analysis*, Academic, New York, 1960.
4 S. Friedland, Variations of tensor powers, to appear.
5 P. Henrici, Bounds for iterates, inverses, spectral variation and fields of values of nonnormal matrices, *Numer. Math.* 4:24–39 (1962).
6 M. Marcus, *Finite-dimensional Multilinear Algebra* (2 vols.), Marcel Dekker, New York, 1973, 1975.
7 M. Marcus and H. Minc, *A Survey of Matrix Theory and Matrix Inequalities*, Prindle, Weber, and Schmidt, Boston, 1964.
8 A. M. Ostrowski, *Solution of Equations and Systems of Equations*, 2nd ed., Academic, New York, 1966.
9 V. S. Sunder, A non-commutative analogue of $D(X^k) = k X^{k-1}$, to appear.

Received 27 May 1980; revised 9 March 1981