

DERIVATIVES OF TENSOR POWERS AND THEIR NORMS*

RAJENDRA BHATIA[†], PRIYANKA GROVER[†], AND TANVI JAIN[†]

Abstract. The norm of the m th derivative of the map that takes an operator to its k th antisymmetric tensor power is evaluated. The case $m = 1$ has been studied earlier by Bhatia and Friedland [R. Bhatia and S. Friedland. Variation of Grassman powers and spectra. *Linear Algebra and its Applications*, 40:1–18, 1981]. For this purpose a multilinear version of a theorem of Russo and Dye is proved: it is shown that a positive m -linear map between C^* -algebras attains its norm at the m -tuple (I, I, \dots, I) . Expressions for derivatives of the maps that take an operator to its k th tensor power and k th symmetric tensor power are also obtained. The norms of these derivatives are computed. Derivatives of the map taking a matrix to its permanent are also evaluated.

Key words. Determinant, Derivative, Norm, Tensor power, Antisymmetric tensor power, Symmetric tensor power, Permanent, Positive linear map, Russo-Dye theorem.

AMS subject classifications. 15A15, 15A18, 15A60, 15A69, 47A30, 47A80.

1. Introduction. Let $\mathcal{B}(\mathcal{H})$ be the space of linear operators on an n dimensional Hilbert space \mathcal{H} . Let $s_1(A) \geq s_2(A) \geq \dots \geq s_n(A) \geq 0$ be the decreasingly ordered singular values of an operator A . Let $\wedge^k \mathcal{H}$ be the k th antisymmetric tensor power of \mathcal{H} , $1 \leq k \leq n$; and let $\wedge^k : \mathcal{B}(\mathcal{H}) \rightarrow \mathcal{B}(\wedge^k \mathcal{H})$ be the map that takes an element A of \mathcal{H} to its k th antisymmetric tensor power $\wedge^k(A)$. Let $D \wedge^k(A)$ be the derivative of this map. This is a linear map from $\mathcal{B}(\mathcal{H})$ into $\mathcal{B}(\wedge^k \mathcal{H})$ and its norm is defined as

$$(1.1) \quad \|D \wedge^k(A)\| = \sup_{\|X\|=1} \|D \wedge^k(A)(X)\|,$$

where $\|X\|$ is the operator norm of X as a linear operator on \mathcal{H} ; i.e.,

$$\|X\| = \sup_{\|u\|=1} \|Xu\|.$$

An alternative expression for this is $\|X\| = s_1(X)$. Finding $\|X\|$ involves solving a maximisation problem, which is not easy.

Motivated by problems in perturbation theory of eigenvalues, Bhatia and Friedland [7] studied the problem of finding the norm (1.1) and obtained a striking formula

*Received by the editors on January 27, 2013. Accepted for publication on July 8, 2013. Handling Editor: Moshe Goldberg.

[†]Theoretical Statistics and Mathematics Unit, Indian Statistical Institute, 7, S.J.S. Sansanwal Marg, New Delhi-110016, India (rbh@isid.ac.in, pgrover8r@isid.ac.in, tanvi@isid.ac.in).

$$(1.2) \quad \|D^{\wedge^k}(A)\| = p_{k-1}(s_1(A), \dots, s_k(A)),$$

where $p_{k-1}(x_1, \dots, x_k)$ is the $(k-1)$ th elementary symmetric polynomial in k variables x_1, \dots, x_k . Analogues of this formula for other kinds of tensor products have been established; see [3], [6].

This paper is concerned with higher order derivatives of \wedge^k and of other multilinear operators and functions, and is related to some other recent work of two of the authors. The famous Jacobi formula gives the derivative of the determinant function on $n \times n$ matrices as

$$(1.3) \quad D \det(A)(X) = \operatorname{tr}(\operatorname{adj}(A)X),$$

where the symbol $\operatorname{adj}(A)$ stands for the transpose of the matrix whose (i, j) -entry is $(-1)^{i+j} \det A(i|j)$, called the *adjugate* (or the *classical adjoint*) of A . Analogous formulas for higher order derivatives of \det were obtained in [8]. Then the more general problem of evaluating higher order derivatives of the map \wedge^k was studied in [13]. For $1 \leq m \leq k$, $D^m \wedge^k(A)$ is a multilinear map

$$(1.4) \quad D^m \wedge^k(A) : \underbrace{\mathcal{B}(\mathcal{H}) \times \dots \times \mathcal{B}(\mathcal{H})}_{m\text{-fold}} \rightarrow \mathcal{B}(\wedge^k \mathcal{H}).$$

Its norm is defined as

$$(1.5) \quad \|D^m \wedge^k(A)\| = \sup_{\|X^1\|=\dots=\|X^m\|=1} \|D^m \wedge^k(A)(X^1, \dots, X^m)\|.$$

In [13], Jain obtained a formula for (1.4) and used it to evaluate (1.5). This last result can be stated as:

THEOREM 1.1. *With notations as above, we have*

$$(1.6) \quad \|D^m \wedge^k(A)\| = m! p_{k-m}(s_1(A), \dots, s_k(A)),$$

where p_{k-m} is the $(k-m)$ th elementary symmetric polynomial.

Recall that for $1 \leq r \leq k$, the r th elementary symmetric polynomial is defined as

$$p_r(x_1, \dots, x_k) = \sum_{1 \leq i_1 < \dots < i_r \leq k} x_{i_1} x_{i_2} \dots x_{i_r}.$$

The first step in the Bhatia-Friedland proof of (1.2) is the observation that

$$\|D^{\wedge^k}(A)\| = \|D^{\wedge^k}(|A|)\|,$$

where $|A|$ is the absolute value of A , defined as $|A| = (A^*A)^{1/2}$. This was exploited by Sunder [17], who obtained another proof of (1.2) by invoking a well-known theorem from the theory of positive linear maps. A linear map Φ from $\mathcal{B}(\mathcal{H})$ into $\mathcal{B}(\mathcal{K})$ is said to be *positive* if $\Phi(A)$ is a positive semidefinite operator whenever A is positive semidefinite. A famous theorem of Russo and Dye [5, p. 42] says that if Φ is a positive linear map, then $\|\Phi\| = \|\Phi(I)\|$. The proof that we give for (1.6) uses a multilinear version of the Russo-Dye Theorem that we prove in this paper. This is of independent interest and is likely to be useful in other situations.

The norm of a multilinear map Φ from $\mathcal{B}(\mathcal{H})^m$ into $\mathcal{B}(\mathcal{K})$ is defined as

$$(1.7) \quad \|\Phi\| = \sup_{\|X^1\|=\dots=\|X^m\|=1} \|\Phi(X^1, \dots, X^m)\|.$$

We say Φ is *positive* if $\Phi(X^1, \dots, X^m)$ is a positive semidefinite operator whenever X^1, \dots, X^m are positive semidefinite. We prove:

THEOREM 1.2. *Let Φ be a positive multilinear map. Then*

$$(1.8) \quad \|\Phi\| = \|\Phi(I, I, \dots, I)\|.$$

The extensions of higher order analogues of Jacobi's formula (1.3) to derivatives of the antisymmetric tensor powers $\wedge^k(A)$ in [13] were obtained by Jain by following an “upwards from the bottom” approach, thinking of $\wedge^k(A)$ as an $\binom{n}{k} \times \binom{n}{k}$ matrix whose entries are $k \times k$ minors of A . Grover [11] followed a similar approach in obtaining an expression for the derivatives of symmetric tensor powers $\vee^k(A)$, $k \geq 1$. Here we look at these problems following a “downwards from the top” approach, thinking of $\wedge^k(A)$ and $\vee^k(A)$ as the restrictions of the tensor power $\otimes^k(A)$ to invariant subspaces. This has several advantages: the proofs become easier and more transparent, the formulas are seen to be valid for infinite dimensional operators as well, the path to studying the same problem for other symmetry classes of tensors becomes clearer.

For the most part, we concentrate on finite dimensional Hilbert spaces. Extensions to infinite dimensional spaces are briefly indicated. In Section 3 we provide the derivatives of the maps that take an operator to its k th tensor power, k th anti-symmetric tensor power and k th symmetric tensor power and compute their norms. Closely related to the determinant is the permanent function. In Section 4 we give expressions for derivatives of all orders for the permanent. These supplement the results in [8]. Results of this section have been reported in a survey article [11]. We provide the details here.

2. A Russo-Dye Theorem for multilinear maps. We first provide a proof of Theorem 1.2, when \mathcal{H} is finite dimensional. We imitate the proof for positive linear maps given in [5, p. 41].

Let U^1, U^2, \dots, U^m be unitary matrices and let

$$U^i = \sum_{j=1}^{r_i} \lambda_j^i P_j^i, \quad 1 \leq i \leq m,$$

be their spectral resolutions. Here, λ_j^i are the eigenvalues of U^i , and P_j^i the corresponding eigenprojections. In particular, $|\lambda_j^i| = 1$, P_j^i are positive semidefinite, and

$$\sum_{j=1}^{r_i} P_j^i = I, \quad 1 \leq i \leq m.$$

Since Φ is multilinear we have

$$\begin{aligned} & \Phi(U^1, \dots, U^m) \\ &= \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} \cdots \sum_{j_m=1}^{r_m} \lambda_{j_1}^1 \lambda_{j_2}^2 \cdots \lambda_{j_m}^m \Phi(P_{j_1}^1, P_{j_2}^2, \dots, P_{j_m}^m), \end{aligned}$$

and

$$\Phi(I, \dots, I) = \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} \cdots \sum_{j_m=1}^{r_m} \Phi(P_{j_1}^1, P_{j_2}^2, \dots, P_{j_m}^m).$$

Since Φ is positive, the operators $\Phi(P_{j_1}^1, P_{j_2}^2, \dots, P_{j_m}^m)$ are positive semidefinite.

Using these two relations, we see that

$$\begin{aligned} & \begin{bmatrix} \Phi(I, \dots, I) & \Phi(U^1, \dots, U^m) \\ \Phi(U^1, \dots, U^m)^* & \Phi(I, \dots, I) \end{bmatrix} \\ &= \sum_{j_1=1}^{r_1} \cdots \sum_{j_m=1}^{r_m} \begin{bmatrix} 1 & \lambda_{j_1}^1 \cdots \lambda_{j_m}^m \\ \overline{\lambda_{j_1}^1 \cdots \lambda_{j_m}^m} & 1 \end{bmatrix} \otimes \Phi(P_{j_1}^1, \dots, P_{j_m}^m). \end{aligned}$$

This is a sum of tensor products of positive semidefinite matrices, and is therefore, positive semidefinite. It follows from Proposition 1.3.2 in [5] that

$$(2.1) \quad \|\Phi(U^1, \dots, U^m)\| \leq \|\Phi(I, \dots, I)\|.$$

Now let X^i , $1 \leq i \leq m$, be matrices with $\|X^i\| = 1$. Then there exist unitary matrices U^i and V^i such that $X^i = \frac{1}{2}(U^i + V^i)$ (see [5, p. 42]).

By the multilinearity of Φ

$$\Phi(X^1, \dots, X^m) = \frac{1}{2^m} \sum \Phi(W^1, \dots, W^m),$$

where the summation is over 2^m terms obtained by choosing each of the W^i to be either U^i or V^i , $1 \leq i \leq m$. It follows from (2.1) that

$$\|\Phi(X^1, \dots, X^m)\| \leq \|\Phi(I, \dots, I)\|.$$

Hence, $\|\Phi\| = \|\Phi(I, \dots, I)\|$. This establishes Theorem 1.2 when \mathcal{H} is finite dimensional.

This theorem is also valid when \mathcal{H} and \mathcal{K} are infinite dimensional Hilbert spaces. It is likely to be useful, and we provide a proof for the infinite dimensional case.

Our proof invokes the well-known fact that if A and B are positive operators on a Hilbert space, then $\begin{bmatrix} A & X \\ X^* & B \end{bmatrix}$ is positive if and only if there exists a contraction K such that $X = A^{1/2} K B^{1/2}$. (See Theorem I.1 in [1]. This is Proposition 1.3.2 in [5] but the proofs given there are only for finite dimensional spaces.) To prove (1.8) we have to show that if X^1, \dots, X^m are operators with $\|X^i\| \leq 1$, then

$$(2.2) \quad \|\Phi(X^1, \dots, X^m)\| \leq \|\Phi(I, \dots, I)\|.$$

Consider first the case when

$$(2.3) \quad X^i = \sum_{j=1}^{r_i} \lambda_j^i P_j^i, \quad 1 \leq i \leq m,$$

where P_j^i are mutually orthogonal projection operators with $\sum_{j=1}^{r_i} P_j^i = I$, and $|\lambda_j^i| = 1$.

It can be seen that $\|X^i\| \leq 1$. (See [16, p. 11].) Arguing as in the finite dimensional case we see that the inequality (2.2) holds in this case. Now if U^i , $1 \leq i \leq m$, are unitary operators, then each U^i is a limit of a sequence of operators of the form (2.3). This shows that the inequality (2.2) holds when X^i are unitary. From here one can see that the inequality continues to hold if each X^i is a convex combination of unitary operators. Finally, since the closed unit ball in $\mathcal{B}(\mathcal{H})$ is the closed convex hull of unitary operators (see [12, p. 75]) the inequality is valid when X^i are any operators with $\|X^i\| \leq 1$.

3. Formulas for $D^m \otimes^k(A)$, $D^m \wedge^k(A)$, $D^m \vee^k(A)$ and their norms. Let \mathcal{H} be a Hilbert space and let $\otimes^k \mathcal{H}$ be its k -fold tensor power $\mathcal{H} \otimes \mathcal{H} \otimes \dots \otimes \mathcal{H}$. Let $\wedge^k \mathcal{H}$ and $\vee^k \mathcal{H}$ be the subspaces of $\otimes^k \mathcal{H}$ consisting of antisymmetric tensors and symmetric tensors, respectively. (See [4, Chap. I] for definitions, notations and basic facts.) If $\dim \mathcal{H} = n$, then $\dim \wedge^k \mathcal{H} = \binom{n}{k}$ for $1 \leq k \leq n$, and $\dim \vee^k \mathcal{H} = \binom{n+k-1}{k}$ for $k \geq 1$. For $k > n$, the space $\wedge^k \mathcal{H}$ is taken to be zero. For every A in $\mathcal{B}(\mathcal{H})$, we denote by $\otimes^k A$ its k -fold tensor power $A \otimes A \otimes \dots \otimes A$. This is an operator on $\otimes^k \mathcal{H}$ that leaves invariant the subspaces $\wedge^k \mathcal{H}$ and $\vee^k \mathcal{H}$. The restriction of $\otimes^k A$ to these subspaces

are denoted by $\wedge^k A$, the k th antisymmetric tensor power of A and $\vee^k A$, the k th symmetric tensor power of A , respectively. We wish to describe the m th derivatives of the maps $A \mapsto \otimes^k A$, $A \mapsto \wedge^k A$, and $A \mapsto \vee^k A$, from $\mathcal{B}(\mathcal{H})$ into $\mathcal{B}(\otimes^k \mathcal{H})$, $\mathcal{B}(\wedge^k \mathcal{H})$, and $\mathcal{B}(\vee^k \mathcal{H})$, respectively.

If $f : X \rightarrow Y$ is a map between normed spaces, then its m th derivative at a point a (if it exists) is a map $D^m f(a)$ from the m -fold product $X \times \cdots \times X$ into Y . As a function of its m variables $D^m f(a)(x^1, \dots, x^m)$ is symmetric and linear in each variable. One way of computing it is by using the relation

$$(3.1) \quad \begin{aligned} D^m f(a)(x^1, \dots, x^m) \\ = \frac{\partial^m}{\partial t_1 \cdots \partial t_m} \Big|_{t_1 = \cdots = t_m = 0} f(a + t_1 x^1 + \cdots + t_m x^m). \end{aligned}$$

We refer the reader to Chapter X of [4] for basic facts about differential calculus on matrix spaces. Given A^1, \dots, A^k in $\mathcal{B}(\mathcal{H})$ we define their *symmetrised tensor product* as

$$(3.2) \quad A^1 \widetilde{\otimes} A^2 \widetilde{\otimes} \cdots \widetilde{\otimes} A^k = \frac{1}{k!} \sum_{\sigma \in S_k} A^{\sigma(1)} \otimes A^{\sigma(2)} \otimes \cdots \otimes A^{\sigma(k)},$$

where S_k is the set of all permutations on k symbols. One can check that this operator on $\otimes^k \mathcal{H}$ leaves invariant the subspaces $\wedge^k \mathcal{H}$ and $\vee^k \mathcal{H}$. The restriction of the symmetrised tensor product to $\wedge^k \mathcal{H}$ and $\vee^k \mathcal{H}$ will be denoted by

$$A^1 \wedge A^2 \wedge \cdots \wedge A^k \text{ and } A^1 \vee A^2 \vee \cdots \vee A^k,$$

respectively and called the *symmetrised antisymmetric tensor product* and the *symmetrised symmetric tensor product* of A^1, A^2, \dots, A^k . The operator $A^1 \wedge A^2 \wedge \cdots \wedge A^k$ acts on product vectors $u_1 \wedge u_2 \wedge \cdots \wedge u_k$ as

$$\begin{aligned} (A^1 \wedge A^2 \wedge \cdots \wedge A^k)(u_1 \wedge u_2 \wedge \cdots \wedge u_k) \\ = \frac{1}{k!} \sum_{\sigma \in S_k} (A^{\sigma(1)} u_1) \wedge (A^{\sigma(2)} u_2) \wedge \cdots \wedge (A^{\sigma(k)} u_k). \end{aligned}$$

Similarly the operator $A^1 \vee A^2 \vee \cdots \vee A^k$ acts on $u_1 \vee u_2 \vee \cdots \vee u_k$ as

$$\begin{aligned} (A^1 \vee A^2 \vee \cdots \vee A^k)(u_1 \vee u_2 \vee \cdots \vee u_k) \\ = \frac{1}{k!} \sum_{\sigma \in S_k} (A^{\sigma(1)} u_1) \vee (A^{\sigma(2)} u_2) \vee \cdots \vee (A^{\sigma(k)} u_k). \end{aligned}$$

Let f be the real function $f(t) = t^k$. Then $f^{(m)}(t) = k(k-1) \cdots (k-m+1)t^{k-m} = \frac{k!}{(k-m)!} t^{k-m}$ for $1 \leq m \leq k$, and $f^{(m)}(t) = 0$ for $m > k$. The following theorem is an

operator analogue of this. With the above notations we have:

THEOREM 3.1. *Let $1 \leq m \leq k$. The m th derivatives of the maps \otimes^k, \wedge^k and \vee^k are given by the formulas*

$$(3.3) \quad D^m \otimes^k (A)(X^1, \dots, X^m) = \frac{k!}{(k-m)!} \underbrace{A \tilde{\otimes} \dots \tilde{\otimes} A}_{k-m \text{ copies}} \tilde{\otimes} X^1 \tilde{\otimes} X^2 \tilde{\otimes} \dots \tilde{\otimes} X^m,$$

$$(3.4) \quad D^m \wedge^k (A)(X^1, \dots, X^m) = \frac{k!}{(k-m)!} \underbrace{A \wedge \dots \wedge A}_{k-m \text{ copies}} \wedge X^1 \wedge X^2 \wedge \dots \wedge X^m,$$

and

$$(3.5) \quad D^m \vee^k (A)(X^1, \dots, X^m) = \frac{k!}{(k-m)!} \underbrace{A \vee \dots \vee A}_{k-m \text{ copies}} \vee X^1 \vee X^2 \vee \dots \vee X^m.$$

If $m > k$, then all the derivatives are zero.

Proof. By the formula (3.1),

$$(3.6) \quad \begin{aligned} & D^m \otimes^k (A)(X^1, \dots, X^m) \\ &= \frac{\partial^m}{\partial t_1 \dots \partial t_m} \Big|_{t_1=\dots=t_m=0} \otimes^k (A + t_1 X^1 + \dots + t_m X^m). \end{aligned}$$

To evaluate this we expand the k -fold tensor product on the right hand side. The resulting expansion is a polynomial in the variables t_1, \dots, t_m . The derivative in (3.6) is evidently the coefficient of the term $t_1 t_2 \dots t_m$ in this polynomial. One can check that this is given by the expression (3.3).

Next we prove (3.4) using (3.3). The proof for (3.5) is similar. The chain rule of differentiation for a composite function $g \circ f$ says that

$$D(g \circ f)(a)(x) = Dg(f(a))(Df(a)(x)).$$

If L is a linear map, then its derivative is equal to L , and in this case

$$D(L \circ f)(a)(x) = L(Df(a)(x)).$$

Repeating this argument one sees that if f is m times differentiable, then

$$(3.7) \quad D^m (L \circ f)(a)(x^1, \dots, x^m) = L(D^m f(a)(x^1, \dots, x^m)).$$

Now let $Q_k : \wedge^k \mathcal{H} \rightarrow \otimes^k \mathcal{H}$ be the inclusion map. Then $Q_k^* : \otimes^k \mathcal{H} \rightarrow \wedge^k \mathcal{H}$ is the projection given by

$$Q_k^*(x_1 \otimes \dots \otimes x_k) = \frac{1}{k!} \sum_{\sigma \in S_k} \varepsilon_\sigma x_{\sigma(1)} \otimes \dots \otimes x_{\sigma(k)},$$

where $\varepsilon_\sigma = \pm 1$, depending on whether σ is an even or an odd permutation. Define $\tilde{Q}_k : \mathcal{B}(\otimes^k \mathcal{H}) \rightarrow \mathcal{B}(\wedge^k \mathcal{H})$ by

$$\tilde{Q}_k(T) = Q_k^* T Q_k.$$

Also $\wedge^k : \mathcal{B}(\mathcal{H}) \rightarrow \mathcal{B}(\wedge^k \mathcal{H})$ factors through $\otimes^k : \mathcal{B}(\mathcal{H}) \rightarrow \mathcal{B}(\otimes^k \mathcal{H})$ via \tilde{Q}_k as

$$\wedge^k A = \tilde{Q}_k(\otimes^k A) \quad \forall A \in \mathcal{B}(\mathcal{H}).$$

Since \tilde{Q}_k is linear, by (3.7) we have

$$(3.8) \quad D^m \wedge^k (A) = \tilde{Q}_k \circ D^m \otimes^k (A).$$

Using this we obtain the expression (3.4).

From (3.3), we see that

$$D^k \otimes^k (A)(X^1, \dots, X^k) = k! X^1 \tilde{\otimes} X^2 \tilde{\otimes} \dots \tilde{\otimes} X^k.$$

This expression does not involve A . Hence $D^m \otimes^k (A) = 0$ if $m > k$. Similarly $D^m \wedge^k (A) = 0$ and $D^m \vee^k (A) = 0$ if $m > k$. \square

REMARK 3.2. By putting $k = n$ in (3.4), we get

$$D^m \det(A)(X^1, \dots, X^m) = \frac{n!}{(n-m)!} \underbrace{A \wedge \dots \wedge A}_{n-m \text{ copies}} \wedge X^1 \wedge X^2 \wedge \dots \wedge X^m.$$

This can also be written as

$$D^m \det(A)(X^1, \dots, X^m) = \frac{n!}{(n-m)!} \Delta(A, \dots, A, X^1, \dots, X^m),$$

where $\Delta(A, \dots, A, X^1, \dots, X^m)$ denotes the mixed discriminant of the matrices $A, \dots, A, X^1, \dots, X^m$. This is the same as Theorem 1 in [8].

From these formulas we obtain the values of the norms of these derivatives. We separate the cases of $\|D^m \otimes^k (A)\|$ and $\|D^m \vee^k (A)\|$. The evaluation of these norms is independent of Theorem 1.2, whereas we make essential use of this theorem in calculating $\|D^m \wedge^k (A)\|$.

THEOREM 3.3. For $1 \leq m \leq k$, we have

$$(3.9) \quad \|D^m \otimes^k (A)\| = \frac{k!}{(k-m)!} \|A\|^{k-m}$$

and

$$(3.10) \quad \|D^m \vee^k (A)\| = \frac{k!}{(k-m)!} \|A\|^{k-m}.$$

Proof. To compute the norm $\|D^m \otimes^k (A)\|$, we first see that by definition of the symmetrised tensor product (3.2) and by the triangle inequality we get

$$\|\underbrace{A \otimes \cdots \otimes A}_{k-m \text{ copies}} \otimes X^1 \otimes X^2 \otimes \cdots \otimes X^m\| \leq \frac{1}{k!} \sum_{\sigma \in S_k} \|Y^{\sigma(1)} \otimes Y^{\sigma(2)} \otimes \cdots \otimes Y^{\sigma(k)}\|,$$

where $k-m$ of the Y 's are equal to A and the rest are X^1, X^2, \dots, X^m . Each of the terms in the summation is equal to $\|A\|^{k-m} \|X^1\| \|X^2\| \cdots \|X^m\|$. By the definition of the norm of a multilinear map (1.7) we obtain

$$\|D^m \otimes^k (A)\| \leq \frac{k!}{(k-m)!} \|A\|^{k-m}.$$

Also note that

$$\left\| D^m \otimes^k (A) \left(\frac{A}{\|A\|}, \frac{A}{\|A\|}, \dots, \frac{A}{\|A\|} \right) \right\| = \frac{k!}{(k-m)!} \|A\|^{k-m}.$$

This shows that

$$\|D^m \otimes^k (A)\| \geq \frac{k!}{(k-m)!} \|A\|^{k-m}.$$

Hence we obtain (3.9). This argument works equally well in infinite dimensions.

Let $R_k : \vee^k \mathcal{H} \rightarrow \otimes^k \mathcal{H}$ be the inclusion map. Define $\tilde{R}_k : \mathcal{B}(\otimes^k \mathcal{H}) \rightarrow \mathcal{B}(\vee^k \mathcal{H})$ by

$$\tilde{R}_k(T) = R_k^* T R_k.$$

Then $\|\tilde{R}_k\| \leq 1$. Arguments similar to those in the proof of Theorem 3.1 lead to an expression similar to (3.8):

$$D^m \vee^k (A) = \tilde{R}_k \circ D^m \otimes^k (A).$$

It follows that

$$(3.11) \quad \|D^m \vee^k (A)\| \leq \|\tilde{R}_k\| \|D^m \otimes^k (A)\| \leq \frac{k!}{(k-m)!} \|A\|^{k-m}.$$

Let us now consider the case when \mathcal{H} is an n dimensional space. The polar decomposition theorem tells us that

$$A = U |A|,$$

where U is unitary and $|A| = (A^* A)^{1/2}$ is positive semidefinite. Since $UU^* = I$, we have

$$\begin{aligned} & \vee^k (A + t_1 X^1 + t_2 X^2 + \cdots + t_m X^m) \\ &= \vee^k (U(|A| + t_1 U^* X^1 + t_2 U^* X^2 + \cdots + t_m U^* X^m)) \\ &= (\vee^k U) \vee^k (|A| + t_1 U^* X^1 + t_2 U^* X^2 + \cdots + t_m U^* X^m). \end{aligned}$$

So from (3.1) we obtain

$$D^m \vee^k (A)(X^1, \dots, X^m) = (\vee^k U) D^m \vee^k (|A|)(U^* X^1, \dots, U^* X^m).$$

Now $\vee^k U$ is unitary and the norm is unitarily invariant. So we have

$$\|D^m \vee^k (A)(X^1, \dots, X^m)\| = \|D^m \vee^k (|A|)(U^* X^1, \dots, U^* X^m)\|.$$

The condition $\|X^j\| = 1$ is equivalent to $\|U^* X^j\| = 1$ for $1 \leq j \leq m$. So we have proved that

$$(3.12) \quad \|D^m \vee^k (A)\| = \|D^m \vee^k (|A|)\|.$$

Now assume A is positive semidefinite and let u be an eigenvector corresponding to its maximal eigenvalue $\|A\|$. Consider the vector $w = u \vee u \vee \dots \vee u$ in $\vee^k \mathcal{H}$. If $T = Y^1 \vee Y^2 \vee \dots \vee Y^k$ is an operator in which $k - m$ of the Y 's are equal to A and the rest of them are equal to I , then $Tw = \|A\|^{k-m} w$. It then follows from (3.5) that

$$(D^m \vee^k (A)(I, \dots, I))w = \frac{k!}{(k-m)!} \|A\|^{k-m} w.$$

This shows that

$$\|D^m \vee^k (A)\| \geq \frac{k!}{(k-m)!} \|A\|^{k-m}.$$

We have already noted the reverse inequality in (3.11). So we have (3.10) in the case when A is positive semidefinite. The relation (3.12) then shows that (3.10) is valid for all A .

We now indicate the modifications needed in this proof to handle the infinite dimensional case. In this case A has a *maximal polar representation* $A = U|A|$ in which U is either an isometry ($U^*U = I$) or a coisometry ($UU^* = I$) ([12, p. 75]). When \mathcal{H} is finite dimensional these two conditions are equivalent and U is unitary. Our argument using the polar decomposition for proving (3.12) can be modified. A very similar idea is used in [17] and we refer the reader to that paper for details.

To prove (3.10) in the infinite dimensional case we may, therefore, again assume that A is a positive operator. If A has pure point spectrum, then the arguments given for the finite dimensional case serve equally well here. In particular, (3.10) is valid for compact operators. Every positive operator is a limit of a sequence of positive operators with pure point spectrum. Using this fact one can see that (3.10) is valid for all operators. \square

Note that for the above proof no use of Theorem 1.2 has been made. The formula (1.6) for $\|D^m \wedge^k (A)\|$ is more interesting, and to prove it we do need to invoke Theorem

1.2. To compute $\|D^m \wedge^k(A)\|$ we first note that the symmetrised antisymmetric product of positive semidefinite operators is positive semidefinite. It follows from (3.4) that if A is positive semidefinite, then the map $D^m \wedge^k(A)$ from $\mathcal{B}(\mathcal{H})^m$ into $\mathcal{B}(\wedge^k \mathcal{H})$ is a positive multilinear map. So, we have from Theorem 1.2

$$(3.13) \quad \|D^m \wedge^k(A)\| = \|D^m \wedge^k(A)(I, \dots, I)\|.$$

Arguments similar to the ones used in the proof of Theorem 3.3 show that

$$(3.14) \quad \|D^m \wedge^k(A)\| = \|D^m \wedge^k(|A|)\|.$$

So we assume A to be positive semidefinite. By (3.13), we have

$$\begin{aligned} \|D^m \wedge^k(A)\| &= \|D^m \wedge^k(A)(I, \dots, I)\| \\ &= \left\| \frac{\partial^m}{\partial t_1 \cdots \partial t_m} \Big|_{t_1=\dots=t_m=0} \wedge^k(A + t_1 I + \cdots + t_m I) \right\|. \end{aligned}$$

By the spectral theorem there exists a unitary W such that $A = WDW^*$, where D is the diagonal matrix whose diagonal entries are $\alpha_1 \geq \cdots \geq \alpha_n (\geq 0)$, the eigenvalues of A . The matrix $\wedge^k W$ is again unitary, and our norm is unitarily invariant. So in the right hand side of the equation above we can replace A by D . Now

$$\wedge^k(D + t_1 I + \cdots + t_m I)$$

is a diagonal matrix of order $\binom{n}{k}$. Its norm is equal to its top diagonal entry, which is

$$\frac{\partial^m}{\partial t_1 \cdots \partial t_m} \Big|_{t_1=\dots=t_m=0} \prod_{j=1}^k (\alpha_j + t_1 + \cdots + t_m).$$

A calculation shows that this is equal to

$$m! p_{k-m}(\alpha_1, \dots, \alpha_k).$$

This establishes (1.6) in the case when A is positive semidefinite. The general case follows from (3.14). Theorem 1.1 can be modified for infinite dimensional operators. The statement of this theorem involves the sequence $s_1(A) \geq s_2(A) \geq \cdots$. If we stretch the definitions and interpret a point of the essential spectrum of $|A|$ as an eigenvalue of infinite multiplicity, then Theorem 1.1 is valid for infinite dimensional operators too. The proof is similar to the proof for symmetric tensor powers.

4. Formulas for $D^m \text{per}(A)$. The *permanent* of $A = (a_{ij}) \in \mathbb{M}(n)$, written as $\text{per } A$, is defined by

$$\text{per } A = \sum_{\sigma \in S_n} a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}.$$

Since the definitions of per and \det are similar, it is natural to expect a formula for $D \text{per}(A)$ similar to the Jacobi formula (1.3). Applying the special case $m = 1$ of (3.1) to the per function, we see that $D \text{per}(A)(X)$ is the coefficient of t in the polynomial $\text{per}(A + tX)$. For $1 \leq j \leq n$, let $A(j; X)$ be the matrix obtained from A by replacing the j^{th} column of A by the j^{th} column of X and keeping the rest of the columns unchanged. Since per is a linear function in each of the columns, we get

$$(4.1) \quad D \text{per}(A)(X) = \sum_{j=1}^n \text{per} A(j; X).$$

To give a formula analogous to the Jacobi formula, we define the *permanental adjoint* of A as the $n \times n$ matrix whose (i, j) -entry is $\text{per} A(i|j)$, where $A(i|j)$ denotes the $(n-1) \times (n-1)$ submatrix obtained from A by deleting its i th row and j th column (see [14, p. 237]). Note that the adjugate of A , $\text{adj} A$, is defined as the transpose of the matrix whose (i, j) -entry is $(-1)^{i+j} \det A(i|j)$, whereas in the definition of padj , the transpose is not taken. This is just a matter of convention. The expression (4.1) can be rewritten as follows.

THEOREM 4.1. For each $X \in \mathbb{M}(n)$,

$$(4.2) \quad D \text{per}(A)(X) = \text{tr}(\text{padj}(A)^t X).$$

Our next aim is to obtain higher order derivatives of the permanent function. The expressions obtained are analogous to the ones for the \det function given in [8]. Applying (3.1) to the per function, we see that $D^m \text{per} A(X^1, \dots, X^m)$ is the coefficient of $t_1 \cdots t_m$ in the expansion of $\text{per}(A + t_1 X^1 + \cdots + t_m X^m)$. To write an explicit expression for this, we require some notations.

Let $Q_{m,n} = \{(i_1, \dots, i_m) \mid i_1, \dots, i_m \in \mathbb{N}, 1 \leq i_1 < \cdots < i_m \leq n\}$. For $m > n$, $Q_{m,n} = \emptyset$ by convention. Let $G_{m,n} = \{(i_1, \dots, i_m) \mid i_1, \dots, i_m \in \mathbb{N}, 1 \leq i_1 \leq \cdots \leq i_m \leq n\}$. Note that for $m \leq n$, $Q_{m,n}$ is a subset of $G_{m,n}$. For $\mathcal{J} = (j_1, \dots, j_m) \in Q_{m,n}$, we denote by $A(\mathcal{J}; X^1, \dots, X^m)$, the matrix obtained from A by replacing the j_p^{th} column of A by the j_p^{th} column of X^p for $1 \leq p \leq m$, and keeping the rest of the columns unchanged. Expanding $\text{per}(A + t_1 X^1 + \cdots + t_m X^m)$ by using the fact that per is a linear function in each of the columns, we obtain an expression for $D^m \text{per} A(X^1, \dots, X^m)$ as follows. This is a generalisation of (4.1).

THEOREM 4.2. For $1 \leq m \leq n$,

$$(4.3) \quad D^m \text{per}(A)(X^1, \dots, X^m) = \sum_{\sigma \in S_m} \sum_{\mathcal{J} \in Q_{m,n}} \text{per} A(\mathcal{J}; X^{\sigma(1)}, X^{\sigma(2)}, \dots, X^{\sigma(m)}).$$

In particular,

$$D^m \text{per}(A)(X, \dots, X) = m! \sum_{\mathcal{J} \in Q_{m,n}} \text{per} A(\mathcal{J}; X, \dots, X).$$

The *Laplace expansion theorem* for permanents [15, p. 16] says that for any $1 \leq m \leq n$, and for any $\mathcal{I} \in Q_{m,n}$,

$$(4.4) \quad \text{per } A = \sum_{\mathcal{J} \in Q_{m,n}} \text{per } A[\mathcal{I}|\mathcal{J}] \text{per } A(\mathcal{I}|\mathcal{J}),$$

where $A[\mathcal{I}|\mathcal{J}]$ denotes the $m \times m$ submatrix obtained from A by picking rows \mathcal{I} and columns \mathcal{J} and $A(\mathcal{I}|\mathcal{J})$ denotes the $(n-m) \times (n-m)$ submatrix obtained from A by deleting rows \mathcal{I} and columns \mathcal{J} . In particular, for any $i, 1 \leq i \leq n$,

$$\text{per } A = \sum_{j=1}^n a_{ij} \text{per } A(i|j).$$

Using this, equation (4.1) can be rewritten as

$$(4.5) \quad \text{D per } (A)(X) = \sum_{i=1}^n \sum_{j=1}^n x_{ij} \text{per } A(i|j).$$

We obtain a generalisation of this expression for higher order derivatives. Let $Y_{[j]}$ denote the j th column of the matrix Y . Let σ be a permutation on m symbols, then by $Y_{[\mathcal{J}]}^\sigma$, we mean the matrix in which $Y_{[j_p]}^\sigma = X_{[j_p]}^{\sigma(p)}$ for $1 \leq p \leq m$ and $Y_{[\ell]}^\sigma = 0$ if ℓ does not occur in \mathcal{J} . By using the Laplace expansion (4.4) for each term in the summation of (4.3), we obtain the following expression for $\text{D}^m \text{per } (A)(X^1, \dots, X^m)$.

THEOREM 4.3. For $1 \leq m \leq n$,

$$\text{D}^m \text{per } (A)(X^1, \dots, X^m) = \sum_{\sigma \in S_m} \sum_{\mathcal{I}, \mathcal{J} \in Q_{m,n}} \text{per } A[\mathcal{I}|\mathcal{J}] \text{per } Y_{[\mathcal{J}]}^\sigma[\mathcal{I}|\mathcal{J}].$$

In particular,

$$\text{D}^m \text{per } (A)(X, \dots, X) = m! \sum_{\mathcal{I}, \mathcal{J} \in Q_{m,n}} \text{per } A[\mathcal{I}|\mathcal{J}] \text{per } X[\mathcal{I}|\mathcal{J}].$$

Note that

$$\text{D}^n \text{per } (A)(X, \dots, X) = n! \text{per } X,$$

and

$$\text{D}^m \text{per } (A)(X^1, \dots, X^m) = 0 \text{ for all } m > n.$$

We now describe a generalisation of (4.2) for higher order derivatives of the per function. Given an orthonormal basis $\{e_1, e_2, \dots, e_n\}$ of an n dimensional Hilbert

space \mathcal{H} , the set $\{m(\alpha)^{-1/2} e_\alpha : \alpha \in G_{m,n}\}$ is an orthonormal basis of $\vee^m \mathcal{H}$. (See [4, p. 17] for details.) Let P_m be the canonical projection of $\vee^m \mathcal{H}$ onto the subspace $\{e_\alpha : \alpha \in Q_{m,n}\}$. Then there is a permutation of the above orthonormal basis of $\vee^m \mathcal{H}$ in which $P_m = \begin{bmatrix} I & O \\ O & O \end{bmatrix}$ and the matrix T_m , defined by $T_m = (\text{per } A[\alpha|\beta])_{\alpha,\beta \in Q_{m,n}}$, is the upper left corner of $\vee^m A$, that is,

$$P_m (\vee^m A) P_m = \begin{bmatrix} T_m & O \\ O & O \end{bmatrix}.$$

Let U be the $\binom{n}{m} \times \binom{n}{m}$ unitary matrix given by $U = \begin{bmatrix} & & & 1 \\ & & 1 & \\ & \ddots & & \\ 1 & & & \end{bmatrix}$. For $\alpha, \beta \in Q_{n-m,n}$, the (α, β) -entry of $U^* T_m U$ is $\text{per } A(\alpha|\beta)$. Let \tilde{U} be the $\binom{n+m-1}{m} \times \binom{n+m-1}{m}$ matrix given by $\tilde{U} = \begin{bmatrix} U & O \\ O & I \end{bmatrix}$. Let $\tilde{\vee}^m A$ denote the matrix $\tilde{U}^* (\vee^m A)^t \tilde{U}$. Then

$$P_m (\tilde{\vee}^m A) P_m = \begin{bmatrix} U^* T_m^t U & O \\ O & O \end{bmatrix}.$$

In particular for $m = n - 1$,

$$P_{n-1} (\tilde{\vee}^{n-1} A) P_{n-1} = \begin{bmatrix} (\text{padj } A)^t & O \\ O & O \end{bmatrix}.$$

Identifying an $n \times n$ matrix X with $\binom{2n-2}{n-1} \times \binom{2n-2}{n-1}$ matrix $\begin{bmatrix} X & O \\ O & O \end{bmatrix}$, equation (4.2) can be rewritten as

$$\text{D per } (A)(X) = \text{tr} \left(P_{n-1} (\tilde{\vee}^{n-1} A) P_{n-1} \right) X.$$

Its generalisation for higher order derivatives is given as follows.

THEOREM 4.4. For $1 \leq m \leq n$,

$$\begin{aligned} & \text{D}^m \text{per } (A)(X^1, \dots, X^m) \\ (4.6) \quad & = m! \text{tr} \left[\left(P_{n-m} (\tilde{\vee}^{n-m} A) P_{n-m} \right) (P_m (X^1 \vee \dots \vee X^m) P_m) \right]. \end{aligned}$$

In particular,

$$\text{D}^m \text{per } (A)(X, \dots, X) = m! \text{tr} \left[\left(P_{n-m} (\tilde{\vee}^{n-m} A) P_{n-m} \right) (P_m (\vee^m X) P_m) \right].$$

To see a proof of Theorem 4.4, we first describe the notion of *mixed permanent* of $m \times m$ matrices T^1, \dots, T^m . (This was first introduced by Bapat in [2].) It is denoted by $\Delta_p(T^1, \dots, T^m)$, and is defined as

$$\Delta_p(T^1, \dots, T^m) = \frac{1}{m!} \sum_{\sigma \in S_m} \text{per} \left[T_{[1]}^{\sigma(1)}, \dots, T_{[m]}^{\sigma(m)} \right].$$

When all $T^j = T$, then $\Delta_p(T, \dots, T) = \text{per } T$. Observe that for $\mathcal{I}, \mathcal{J} \in Q_{m,n}$,

$$\sum_{\sigma \in S_m} \text{per } Y_{[\mathcal{J}]}^{\sigma} [\mathcal{I}|\mathcal{J}] = m! \Delta_p(X^1[\mathcal{I}|\mathcal{J}], \dots, X^m[\mathcal{I}|\mathcal{J}]).$$

Using this, Theorem 4.3 can be rewritten as follows:

$$(4.7) \quad \begin{aligned} & D^m \text{per}(A)(X^1, \dots, X^m) \\ &= m! \sum_{\mathcal{I}, \mathcal{J} \in Q_{m,n}} \text{per } A(\mathcal{I}|\mathcal{J}) \Delta_p(X^1[\mathcal{I}|\mathcal{J}], \dots, X^m[\mathcal{I}|\mathcal{J}]). \end{aligned}$$

Next we note that for $\mathcal{I}, \mathcal{J} \in G_{m,n}$, the $(\mathcal{I}, \mathcal{J})$ -entry of $X^1 \vee \dots \vee X^m$ is

$$(4.8) \quad (m(\mathcal{I})m(\mathcal{J}))^{-1/2} \Delta_p(X^1[\mathcal{I}|\mathcal{J}], \dots, X^m[\mathcal{I}|\mathcal{J}]).$$

In particular, if $\mathcal{I}, \mathcal{J} \in Q_{m,n}$, then the $(\mathcal{I}, \mathcal{J})$ -entry of $X^1 \vee \dots \vee X^m$ is $\Delta_p(X^1[\mathcal{I}|\mathcal{J}], \dots, X^m[\mathcal{I}|\mathcal{J}])$. The $(\mathcal{J}, \mathcal{I})$ -entry of $P_{n-m}(\tilde{V}^{n-m} A)P_{n-m}$ is $\text{per } A(\mathcal{I}|\mathcal{J})$. The expression (4.6) can now be easily seen as a reformulation of (4.7).

5. Remarks.

- (1) An alternative proof of Theorem 4.2 can be given using (3.5). We know that $D^m \text{per}(A)(X^1, \dots, X^m)$ is the (α, α) -entry of $D^m \vee^n A(X^1, \dots, X^m)$ for $\alpha = (1, \dots, n)$, which by (3.5) and (4.8) is $\frac{n!}{(n-m)!} \Delta_p(X^1, \dots, X^m, A, \dots, A)$. This is the same as

$$\frac{1}{(n-m)!} \sum_{\tau \in S_n} \text{per} \left[Y_{[1]}^{\tau(1)}, \dots, Y_{[n]}^{\tau(n)} \right],$$

where $k-m$ of the Y 's are equal to A and the rest are X^1, X^2, \dots, X^m . For any given $\mathcal{J} \in Q_{m,n}$ and $\sigma \in S_m$, there are $(n-m)!$ terms in this summation which are equal to $\text{per } A(\mathcal{J}; X^{\sigma(1)}, X^{\sigma(2)}, \dots, X^{\sigma(m)})$. This gives Theorem 4.2.

- (2) An upper bound for norms of the derivatives of the permanent function can be obtained by using (3.10). By using the fact that $D^m \text{per}(A)(X^1, \dots, X^m)$ is one of the entries of the matrix $D^m \vee^n A(X^1, \dots, X^m)$, we obtain

$$(5.1) \quad \|D^m \text{per}(A)\| \leq \frac{n!}{(n-m)!} \|A\|^{n-m}.$$

While we have equality in (3.10), we may have strict inequality here. For example, let $A = \begin{bmatrix} 1 & \\ & 0 \end{bmatrix}$. Then A is a positive semidefinite matrix. So $\text{D per}(A)$ is a positive linear functional. By the Russo-Dye Theorem, we have

$$\|\text{D per}(A)\| = |\text{D per}(A)(I)|,$$

which is equal to 1, by (4.5). But the right hand side of (5.1) is equal to 2.

- (3) In this paper, we have limited ourselves to tensor powers, symmetric tensor powers and antisymmetric tensor powers. There are other symmetry classes of tensors, and the corresponding problems for these classes have been studied by Carvalho and Freitas in [9] and [10]. Norms of first derivatives of the operators induced on the symmetry classes of tensors had been computed earlier by Bhatia and Da Silva [6]. The work in [10] extends this to higher order derivatives.

REFERENCES

- [1] T. Ando. *Topics on Operator Inequalities*. Hokkaido University, Sapporo, 1978.
- [2] R.B. Bapat. Inequalities for mixed Schur functions. *Linear Algebra and its Applications*, 83:143–149, 1986.
- [3] R. Bhatia. Variation of symmetric tensor powers and permanents. *Linear Algebra and its Applications*, 62:269–276, 1984.
- [4] R. Bhatia. *Matrix Analysis*. Springer, New York, 1997.
- [5] R. Bhatia. *Positive Definite Matrices*. Princeton University Press, New Jersey, 2007.
- [6] R. Bhatia and J.A. Dias da Silva. Variation of induced linear operators. *Linear Algebra and its Applications*, 341:391–402, 2002.
- [7] R. Bhatia and S. Friedland. Variation of Grassman powers and spectra. *Linear Algebra and its Applications*, 40:1–18, 1981.
- [8] R. Bhatia and T. Jain. Higher order derivatives and perturbation bounds for determinants. *Linear Algebra and its Applications*, 431:2102–2108, 2009.
- [9] S. Carvalho and P. Freitas. The k -th derivatives of the immanant and the χ -symmetric power of an operator. ArXiv:1305.1143v1 [math.AC].
- [10] S. Carvalho and P. Freitas. The norm of the k -th derivative of the χ -symmetric power of an operator. ArXiv:1305.1144v1 [math.FA].
- [11] P. Grover. Derivatives of multilinear functions of matrices. In: F. Nielsen and R. Bhatia (editors), *Matrix Information Geometry*, Springer, 93–109, 2013.
- [12] P.R. Halmos. *A Hilbert Space Problem Book*. Springer, New York, 1984.
- [13] T. Jain. Derivatives for antisymmetric tensor powers and perturbation bounds. *Linear Algebra and its Applications*, 435:1111–1121, 2011.
- [14] R. Merris. *Multilinear Algebra*. Gordon and Breach Science Publishers, Singapore, 1997.
- [15] H. Minc. *Permanents*. Addison-Wesley Publishing Company, Massachusetts, 1970.
- [16] V. Paulsen. *Completely Bounded Maps and Operator Algebras*. Cambridge University Press, New York, 2003.
- [17] V.S. Sunder. A noncommutative analogue of $|D(X^k)| = |kX^{k-1}|$. *Linear Algebra and its Applications*, 44:87–95, 1982.