All convex invariant functions of hermitian matrices[1])

By CHANDLER DAVIS in New York

Let $\mathfrak{H}_n$ denote the real linear space of all $n \times n$ hermitian matrices $A, B, C, \ldots$. Consider functions f defined on $\mathfrak{H}_n$ with values in a partly-ordered real vector space $\mathfrak{V}$ ($\mathfrak{V}$ might be the real numbers, for example, or $\mathfrak{H}_m$). Such f is called *convex* provided it satisfies identically

(1) $$f((1-t)A + tB) \leqq (1-t) f(A) + t f(B) \quad \text{for} \quad t \in [0,1].$$

Also f is called invariant provided $f(U^{-1}AU) = f(A)$ for any $A \in \mathfrak{H}_n$ and any $n \times n$ unitary U. It has proved useful in the past to know that a real-valued function is convex invariant — for example, the f defined by letting $f(A)$ be the largest eigenvalue of A. This paper finds all convex invariant f.

The problem has not been studied before even in the real-valued case. In that case, however, it is related to a recent theorem of M. D. MARCUS, as explained below.

If f is invariant, $f(A)$ depends only on the set of eigenvalues of A (counted according to their multiplicity), an unordered n-tuple of reals; because this set is a complete set of unitary invariants for A. Therefore f corresponds to a function of n real variables, with values in $\mathfrak{V}$; this function will again be denoted by f; it is symmetric, in the sense that $f(\lambda_1, \ldots, \lambda_n) = f(\mu_1, \ldots, \mu_n)$ whenever $(\mu_1, \ldots, \mu_n)$ is just $(\lambda_1, \ldots, \lambda_n)$ rearranged. Suppose in addition f is convex as a function from $\mathfrak{H}_n$ to $\mathfrak{V}$, and consider diagonal matrices $A = \operatorname{diag}(\alpha_1, \ldots, \alpha_n)$, $B = \operatorname{diag}(\beta_1, \ldots, \beta_n)$, $C = \operatorname{diag}(\gamma_1, \ldots, \gamma_n)$ with $\gamma_i = (1-t)\alpha_i + t\beta_i$, $0 \leqq t \leqq 1$, $i = 1, \ldots, n$. By (1),

$$f(\gamma_1, \ldots, \gamma_n) \leqq (1-t) f(\alpha_1, \ldots, \alpha_n) + t f(\beta_1, \ldots, \beta_n).$$

This proves half of the following theorem.

Theorem. *A unitary-invariant function from $n \times n$ hermitian matrices to a partly-ordered real vector space is convex if and only if the corresponding symmetric function of n real variables is convex.*

Here is the non-trivial half of the proof.

Suppose f symmetric and convex as a function of n real variables. Let again $C = (1-t)A + tB$, $0 \leqq t \leqq 1$, but now let A, B be arbitrary in $\mathfrak{H}_n$. Because f is invariant, there is no loss in generality in assuming C diagonalized. If A_C represents the matrix whose diagonal elements are the same as those of A but whose off-diagonal elements are zero, and B_C similarly, then $C = (1-t)A_C + tB_C$ for the same t. (But of course $A_C \neq A$ and $B_C \neq B$ in general — necessarily whenever $AB \neq BA$.)

[1]) Presented to the American Mathematical Society, February 23, 1957.

The hypothesis on f implies that $f(C) \leqq (1-t)\, f(A_C) + t f(B_C)$, for C, A_C, and B_C are simultaneously diagonalized. If it was known that $f(A_C) \leqq f(A)$ and $f(B_C) \leqq f(B)$ substitution would yield (1).

This will now be shown. The proof is almost exactly the same as one by M. D. MARCUS [3]; see also [4], [2]. The idea is to regard $f(A_C)$, for fixed A, as a function of C — more exactly, as a function of S defined below. Let $x_1, \ldots, x_n$ be orthonormal eigenvectors of A: for all u, $A u = \sum_{j=1}^{n} \alpha_j (u\, x_j)\, x_j$. Let $z_1, \ldots, z_n$ be orthonormal eigenvectors of C and of A_C: for all u, $A_C u = \sum_{j=1}^{n} \alpha'_j\, (u\, z_j)\; z_j$. By invariance, $f(A)$ depends only on $\alpha_1, \ldots, \alpha_n$, and $f(A_C)$ only on $\alpha'_1, \ldots, \alpha'_n$. Now

$$\alpha'_i = (A_C\, z_i, z_i) = (A z_i, z_i) = \left(\sum_{j=1}^{n} \alpha_j (z_i, x_j)\, x_j, z_i\right) =$$

$$= \sum_{j=1}^{n} \alpha_j \,|(x_j, z_i)|^2 = \sum_j S_{ij}\, \alpha_j\,.$$

The matrix S here, defined by $S_{ij} = |(x_j, z_i)|^2$, is *doubly stochastic*; that is, $S_{ij} \geqq 0$, $\sum_{j=1}^{n} S_{ij} = \sum_{i=1}^{n} S_{ij} = 1$. A permutation matrix is a doubly stochastic matrix each element of which equals either 0 or 1. It is a theorem of G. BIRKHOFF [1] that every $n \times n$ doubly stochastic matrix is a convex combination of the $n \times n$ permutation matrices P^k, so write $S = \sum_{k=1}^{n!} \mu_k\, P^k$, $\mu_k \geqq 0$, $\sum_{k=1}^{n!} \mu_k = 1$. Using these facts, and the convexity hypothesis on f,

$$f(A_C) = f(\alpha'_1, \ldots, \alpha'_n) = f\left(\sum_k \mu_k \sum_j P^k_{1j}\, \alpha_j, \ldots, \sum_k \mu_k \sum_j P^k_{nj}\, \alpha_j\right) \leqq$$

$$\leqq \sum_k \mu_k\, f\left(\sum_j P^k_{1j}\, \alpha_j, \ldots, \sum_j P^k_{nj}\, \alpha_j\right) = \sum_k \mu_k\, f(\beta^k_1, \ldots, \beta^k_n)\,.$$

But for each k, since P^k is a permutation matrix, $(\beta^k_1, \ldots, \beta^k_n)$ is just $(\alpha_1, \ldots, \alpha_n)$ rearranged; so by the symmetry of f, $f(\beta^k_1, \ldots, \beta^k_n) = f(\alpha_1, \ldots, \alpha_n) = f(A)$. Making this substitution, $f(A_C) \leqq \sum_k \mu_k\, f(A) = f(A)$. The theorem is proved.

If $\mathfrak{B}$, the range of f, is the real numbers, $f(A_C) \leqq f(A)$ is just a specialization of MARCUS's theorem [3] in a different notation. But his proof does not apply as it is to non-simply-ordered $\mathfrak{B}$ because a convex function defined on doubly stochastic matrices with values in $\mathfrak{B}$ need not assume any maximum. Here is a trivial example in which the function g is even linear: let $\mathfrak{B}$ be real diagonal 2×2 matrices with the usual ordering, and $g\begin{pmatrix} a & 1-a \\ 1-a & a \end{pmatrix} = \begin{pmatrix} a & 0 \\ 0 & 1-a \end{pmatrix}$ for $0 \leqq a \leqq 1$.

Corollary 1. *The theorem is also true for orthogonal-invariant functions of real symmetric matrices.*

Corollary 2. *The theorem and Corollary* 1 *remain true if the functions are defined only for matrices whose spectra are restricted to a given finite or infinite interval.*

These modifications can be made without changing the proof.

References

[1] G. BIRKHOFF, Three observations on linear algebra. Univ. nac. Tucumán, Revista, Ser. A, **5**, 147–151 (1946). (In Spanish.)

[2] A. HORN, Doubly stochastic matrices and the diagonal of a rotation matrix. Amer. J. Math. **76**, 620–630 (1954).

[3] M. D. MARCUS, Convex functions of quadratic forms. To appear.

[4] I. SCHUR, Über eine Klasse von Mittelbildungen mit Anwendungen auf die Determinantentheorie. Sitzungsberichte der Berliner Mathematischen Gesellschaft **22**, 9–20 (1923).

Eingegangen am 22. 6. 1957