INEQUALITIES FOR CERTAIN CLASSES OF CONVEX FUNCTIONS

by L. MIRSKY
(Received 27th June 1959)

MAKING use of properties of doubly-stochastic matrices, I recently gave a simple proof (4) of a theorem of Ky Fan (Theorem 2b below) on symmetric gauge functions. I now propose to show that the same idea can be employed to derive a whole series of results on convex functions; in particular, certain well-known inequalities of Hardy-Littlewood-Pólya and of Pólya will emerge as special cases.

Our notation is as follows. All numbers are understood to be real; n denotes a positive integer and $\mathbf{x}=(x_1, \ldots, x_n)$ and $\mathbf{y}=(y_1, \ldots, y_n)$ denote two given vectors. If $\bar{x}_1, \ldots, \bar{x}_n$ are the numbers $x_1, \ldots, x_n$ arranged in non-ascending order of magnitude and if $\bar{y}_1, \ldots, \bar{y}_n$ are defined similarly, then $\mathbf{x} \ll \mathbf{y}$ means that

$$\sum_{i=1}^{k} \bar{x}_i \leqslant \sum_{i=1}^{k} \bar{y}_i \quad (1 \leqslant k \leqslant n). \qquad \text{(1)}$$

Further, we write $\mathbf{x} \prec \mathbf{y}$ if (1) holds and if, in addition, there is equality for $k=n$.

We put $a^+=\max(a, 0)$ and $\mathbf{x}^+=(x_1^+, \ldots, x_n^+)$. A typical permutation of $1, \ldots, n$ is denoted by π, and $\mathbf{x}_\pi$ is the vector obtained when π is applied to the components of $\mathbf{x}$. The set of symmetric functions, i.e. of functions F such that $F(\mathbf{u}_\pi)=F(\mathbf{u})$ for all vectors $\mathbf{u}$ and all permutations π, is denoted by $\mathscr{S}$. The set of increasing functions, i.e. of functions F such that $F(\mathbf{u}) \leqslant F(\mathbf{v})$ whenever † $\mathbf{u} \leqslant \mathbf{v}$, is denoted by $\mathscr{I}$. Again, $\mathscr{Z}$ denotes the set of functions F such that $F(\mathbf{u}\Lambda) \leqslant F(\mathbf{u})$ for every vector $\mathbf{u}$ and every diagonal matrix Λ whose diagonal elements are equal to 0 or 1. Thus, for $n=1, f \in \mathscr{Z}$ means that $\min_u f(u)=f(0)$.

The function F is called *convex* if, for any vectors $\mathbf{u}$, $\mathbf{v}$ and any numbers λ, μ such that $\lambda \geqslant 0$, $\mu \geqslant 0$, $\lambda+\mu=1$, we have

$$F(\lambda\mathbf{u}+\mu\mathbf{v}) \leqslant \lambda F(\mathbf{u})+\mu F(\mathbf{v}).$$

The set of convex functions is denoted by $\mathscr{C}$.

Following von Neumann (6), we call F a *symmetric gauge function* if it satisfies the following conditions. (i) $F(\mathbf{u})>0$ $(\mathbf{u} \neq \mathbf{0})$; (ii) $F(\rho\mathbf{u})=|\rho| F(\mathbf{u})$; (iii) $F(\mathbf{u}+\mathbf{v}) \leqslant F(\mathbf{u})+F(\mathbf{v})$; (iv) $F(\mathbf{u}_\pi\Delta)=F(\mathbf{u})$. Here ρ is any real number; $\mathbf{u}$, $\mathbf{v}$ are any vectors; π any permutation; and Δ any diagonal matrix with diagonal elements ± 1. The set of symmetric gauge functions will be denoted by $\mathscr{G}$.

† Inequalities between vectors are interpreted component-wise.

https://doi.org/10.1017/S0013091500021969 Published online by Cambridge University Press

Theorem 1. *The inequality*

$$F(\mathbf{x}) \leqslant F(\mathbf{y}) \quad \text{.....................................(2)}$$

holds for all $F \in \mathscr{C} \cap \mathscr{S}$ *if and only if* †

$$\mathbf{x} \prec \mathbf{y}. \quad \text{...(3)}$$

Let (3) be given. Then, by (**8**), $\mathbf{x}$ lies in the convex hull of the $n\,!$ vectors $\mathbf{y}_\pi$, i.e.

$$\mathbf{x} = \sum_\pi t_\pi \mathbf{y}_\pi,$$

where the t's are non-negative numbers with sum 1. Hence, for $F \in \mathscr{C} \cap \mathscr{S}$,

$$F(\mathbf{x}) \leqslant \sum_\pi t_\pi F(\mathbf{y}_\pi) = \sum_\pi t_\pi F(\mathbf{y}) = F(\mathbf{y}).$$

Suppose, on the other hand, that (2) holds for all $F \in \mathscr{C} \cap \mathscr{S}$. The functions $(u - y_k)^+$ $(1 \leqslant k \leqslant n)$ of the single variable u are all convex. Hence

$$F_k(\mathbf{u}) = \sum_{i=1}^n (u_i - y_k)^+ \in \mathscr{C} \cap \mathscr{S} \quad (1 \leqslant k \leqslant n),$$

and so $F_k(\mathbf{x}) \leqslant F_k(\mathbf{y})$ $(1 \leqslant k \leqslant n)$. By the reasoning of Hardy, Littlewood, and Pólya (**2**, pp. 89-90), we now infer (3).

Theorem 1a. (Hardy-Littlewood-Pólya (**2**, Theorem 108); Karamata (**3**)). *The inequality*

$$\sum_{i=1}^n f(x_i) \leqslant \sum_{i=1}^n f(y_i) \quad \text{....................................(4)}$$

holds for all ‡ $f \in \mathscr{C}$ *if and only if* (3) *is satisfied.*

Let (3) be given. If $f \in \mathscr{C}$, then

$$F(\mathbf{u}) = \sum_{i=1}^n f(u_i) \in \mathscr{C} \cap \mathscr{S}\,; \quad \text{................................(5)}$$

and (4) follows by Theorem 1. If, on the other hand, (4) holds for all $f \in \mathscr{C}$, then we infer (3) exactly as in (**2**), viz. by considering the functions u, $-u$, $(u - y_k)^+$ $(1 \leqslant k \leqslant n)$.

Theorem 2. *The inequality* (2) *holds for all* $F \in \mathscr{C} \cap \mathscr{S} \cap \mathscr{I}$ *if and only if*

$$\mathbf{x} \ll \mathbf{y} \quad \text{...(6)}$$

Let (6) be given. Then, by (**5**), there exists a doubly-stochastic matrix § $\mathbf{D}$ such that $\mathbf{x} \leqslant \mathbf{yD}$. Hence, for $F \in \mathscr{I}$, $F(\mathbf{x}) \leqslant F(\mathbf{yD})$. But, by (**8**), $\mathbf{yD}$ lies in the convex hull of the vectors $\mathbf{y}_\pi$ and so, for $F \in \mathscr{C} \cap \mathscr{S}$, $F(\mathbf{yD}) \leqslant F(\mathbf{y})$. Hence (2) holds for all $F \in \mathscr{C} \cap \mathscr{S} \cap \mathscr{I}$.

† To avoid trivial complications, we consider throughout those functions which possess the specified properties *everywhere*. Naturally, our theorems could be made a little sharper if we restricted ourselves to appropriate regions or intervals.

‡ In (**2**) the continuity of f was postulated explicitly, but our definition of convexity implies that f is continuous.

§ A (square) matrix is said to be doubly-stochastic if its elements are non-negative and if the sum of the elements in each row and in each column is equal to 1.

https://doi.org/10.1017/S0013091500021969 Published online by Cambridge University Press

Again, the functions $(u-y_k)^+$ $(1 \leqslant k \leqslant n)$ of u all belong to $\mathscr{C} \cap \mathscr{I}$. Hence the functions $F_k(\mathbf{u})(1 \leqslant k \leqslant n)$, defined in the proof of Theorem 1, belong to $\mathscr{C} \cap \mathscr{S} \cap \mathscr{I}$. If follows, by precisely the same argument as above, that, if (2) holds for all $F \in \mathscr{C} \cap \mathscr{S} \cap \mathscr{I}$, then (6) is satisfied.

Theorem 2a. (Pólya (7)).

The inequality (4) *holds for all* $f \in \mathscr{C} \cap \mathscr{I}$ *if and only if* (6) *is satisfied.*

Let (6) be given. If $f \in \mathscr{C} \cap \mathscr{I}$, then the function $F(\mathbf{u})$, defined by (5), belongs to $\mathscr{C} \cap \mathscr{S} \cap \mathscr{I}$. Hence (4) follows by Theorem 2. If, on the other hand, (4) holds for all $f \in \mathscr{C} \cap \mathscr{I}$, we consider, as before, the functions $(u-y_k)^+$ $(1 \leqslant k \leqslant n)$ and obtain (6).

Theorem 2b. (Fan (1)).

Suppose that $\mathbf{x} \geqslant \mathbf{0}$, $\mathbf{y} \geqslant \mathbf{0}$. *Then the inequality* (2) *holds for all* $F \in \mathscr{G}$ *if and only if* (6) *is satisfied.*

Denote by $\mathscr{C}^+, \mathscr{S}^+, \mathscr{I}^+$ the sets of functions which are respectively convex, symmetric, and increasing in the positive orthant. Let $F \in \mathscr{G}$; then clearly $F \in \mathscr{C}^+ \cap \mathscr{S}^+$. Moreover, it is well known (9, Lemma 5.16) that $F \in \mathscr{I}^+$. Thus $F \in \mathscr{C}^+ \cap \mathscr{S}^+ \cap \mathscr{I}^+$ and (2) follows by an obvious modification of Theorem 2.

Again, suppose that (2) holds for all $F \in \mathscr{G}$. As was pointed out by Fan (1), relation (6) follows at once if we consider the functions

$$F_k(\mathbf{u}) = \max_{1 \leq i_1 < \ldots < i_k \leq n} (|u_{i_1}| + \ldots + |u_{i_k}|) \; (1 \leqslant k \leqslant n)$$

which obviously belong to $\mathscr{G}$.

Theorem 3. *The inequality* (2) *holds for all* $F \in \mathscr{C} \cap \mathscr{S} \cap \mathscr{Z}$ *if and only if*

$$\mathbf{x} \ll \mathbf{y}^+, \quad -\mathbf{x} \ll (-\mathbf{y})^+. \qquad \ldots\ldots\ldots\ldots(7)$$

Let (7) be satisfied. Then, by (5), $\mathbf{x}$ lies in the convex hull of the vectors $\mathbf{y}_\pi \mathbf{\Lambda}$, where π ranges over all permutations of 1, ..., n and $\mathbf{\Lambda}$ over all diagonal matrices whose diagonal elements are 0 or 1. Thus

$$\mathbf{x} = \sum_{\pi, \Lambda} t_{\pi,\Lambda} \mathbf{y}_\pi \mathbf{\Lambda},$$

where the t's are non-negative and have sum 1. Using successively the relations $F \in \mathscr{C}$, $F \in \mathscr{Z}$, $F \in \mathscr{S}$, we obtain

$$F(\mathbf{x}) \leqslant \sum_{\pi, \Lambda} t_{\pi,\Lambda} F(\mathbf{y}_\pi \mathbf{\Lambda}) \leqslant \sum_{\pi, \Lambda} t_{\pi\Lambda}, F(\mathbf{y}_\pi)$$
$$= \sum_{\pi, \Lambda} t_{\pi,\Lambda} F(\mathbf{y}) = F(\mathbf{y}).$$

Next, let (2) be satisfied for all $F \in \mathscr{C} \cap \mathscr{S} \cap \mathscr{Z}$. Assume, as may be done without loss of generality, that

$$x_1 \geqslant \ldots \geqslant x_n, \quad y_1 \geqslant \ldots \geqslant y_n. \qquad \ldots\ldots\ldots\ldots(8)$$

The functions $(u - y_k^+)^+$ $(1 \leqslant k \leqslant n)$ of u all belong to $\mathscr{C} \cap \mathscr{Z}$. Hence

$$F_k(\mathbf{u}) = \sum_{i=1}^{n} (u_i - y_k^+)^+ \in \mathscr{C} \cap \mathscr{S} \cap \mathscr{Z} \quad (1 \leqslant k \leqslant n),$$

and so

$$F_k(\mathbf{x}) \leqslant F_k(\mathbf{y}) \quad (1 \leqslant k \leqslant n) \qquad \ldots\ldots\ldots\ldots(9)$$

Now we have

$$F_k(\mathbf{x}) = \sum_{i=1}^{n} (x_i - y_k^+)^+ \geqslant \sum_{i=1}^{k} (x_i - y_k^+),$$

i.e.

$$F_k(\mathbf{x}) \geqslant \sum_{i=1}^{k} x_i - k y_k^+. \qquad \text{.......(10)}$$

Also, as is easily shown,

$$F_k(\mathbf{y}) = \sum_{i=1}^{k} {}_i^+ - k y_k^+. \qquad \text{.......(11)}$$

Hence, by (9), (10) and (11),

$$\sum_{i=1}^{k} x_i \leqslant \sum_{i=1}^{k} y_i^+ \quad (1 \leqslant k \leqslant n);$$

and, in view of (8), this is equivalent to $\mathbf{x} \ll \mathbf{y}^+$. Furthermore,

$$G_k(\mathbf{u}) = \sum_{i=1}^{n} \{-u_i - (-y_k)^+\}^+ \in \mathscr{C} \cap \mathscr{S} \cap \mathscr{Z} \quad (1 \leqslant k \leqslant n),$$

and so $G_k(\mathbf{x}) \leqslant G_k(\mathbf{y})$. But it is easy to verify that

$$G_k(\mathbf{x}) \geqslant -\sum_{i=k}^{n} x_i - (n-k+1)(-y_k)^+,$$

$$G_k(\mathbf{y}) = \sum_{i=k}^{n} (-y_i)^+ - (n-k+1)(-y_k)^+,$$

and therefore

$$-\sum_{i=k}^{n} x_i \leqslant \sum_{i=k}^{n} (-y_i)^+ \quad (1 \leqslant k \leqslant n),$$

i.e. $-\mathbf{x} \ll (-\mathbf{y})^+$. This completes the proof.

Theorem 3a. *The inequality* (4) *holds for all* $f \in \mathscr{C} \cap \mathscr{Z}$ *if and only if* (7) *is satisfied.*

If $f \in \mathscr{C} \cap \mathscr{Z}$, then the function $F(\mathbf{u})$, defined by (5), belongs to $\mathscr{C} \cap \mathscr{S} \cap \mathscr{Z}$. Hence, if (7) is given, then (4) is satisfied by virtue of Theorem 3. On the other hand, suppose that (4) holds for all $f \in \mathscr{C} \cap \mathscr{Z}$. Then, by considering the functions

$$(u - y_k^+)^+, \quad \{-u - (-y_k)^+\}^+ \quad (1 \leqslant k \leqslant n)$$

and arguing exactly as in the proof of Theorem 3, we infer (7).

It may be noted that, if (8) holds, then the set of inequalities

$$\left.\begin{aligned} \sum_{i=1}^{k} x_i &\leqslant \sum_{i=1}^{k} \max(y_i, 0) \\ \sum_{i=k}^{n} x_i &\geqslant \sum_{i=k}^{n} \min(y_i, 0) \end{aligned}\right\} \quad (1 \leqslant k \leqslant n)$$

is equivalent to (7).

In conclusion, I wish to thank Professor R. Rado for a number of helpful comments.

https://doi.org/10.1017/S0013091500021969 Published online by Cambridge University Press

REFERENCES

(1) K. Fan, Maximum properties and inequalities for the eigenvalues of completely continuous operators, *Proc. Nat. Acad. Sci.*, **37** (1951), 760-766.

(2) G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities* (Cambridge, 1934).

(3) J. Karamata, Sur une inégalité relative aux fonctions convexes, *Publ. math. Univ. Belgrade*, **1** (1932), 145-148.

(4) L. Mirsky, Symmetric gauge functions and unitarily invariant norms, *Quart. J. Math.* (Oxford). To appear.

(5) L. Mirsky, On a convex set of matrices, *Archiv der Math.*, **10** (1959), 88-92.

(6) J. von Neumann, Some matrix-inequalities and metrization of matric-space, *Tomsk Univ. Rev.* **1** (1937), 286-299.

(7) G. Pólya, Remark on Weyl's note : Inequalities between the two kinds of eigenvalues of a linear transformation, *Proc. Nat. Acad. Sci.*, **36** (1950), 49-51.

(8) R. Rado, An inequality, *J. London Math. Soc.*, **27** (1952), 1-6.

(9) R. Schatten, *A Theory of Cross-spaces* (Princeton, 1950).

THE UNIVERSITY
SHEFFIELD

https://doi.org/10.1017/S0013091500021969 Published online by Cambridge University Press