The majorisation principle for convex functions

G. J. O. JAMESON

Introduction

Given positive numbers x_j, y_j such that $\sum_{j=1}^{n} x_j = \sum_{j=1}^{n} y_j$, it can happen that $\sum_{j=1}^{n} x_j^2 = \sum_{j=1}^{n} y_j^2$: for example, $(x_j) = (7, 3, 2)$, $(y_j) = (6, 5, 1)$. However, such cases are exceptional. Can we identify conditions that ensure that $\sum_{j=1}^{n} y_j^2 \leqslant \sum_{j=1}^{n} x_j^2$, or more generally $\sum_{j=1}^{n} y_j^p \leqslant \sum_{j=1}^{n} x_j^p$ for all $p > 1$? More generally still, conditions ensuring that $\sum_{j=1}^{n} f(y_j) \leqslant \sum_{j=1}^{n} f(x_j)$ for a suitable class of functions f?

The 'suitable' class is, in fact, the class of convex functions. Recall that a function f is *convex* (informally, curving upwards) if it lies below the straight-line chord between any two points on its graph. Formally, the definition is: f is convex on the interval I if for x_1, x_2 in I and $0 \leqslant \lambda \leqslant 1$, we have $f[(1-\lambda)x_1 + \lambda x_2] \leqslant (1-\lambda)f(x_1) + \lambda f(x_2)$. It is *strictly convex* if strict inequality holds for $0 < \lambda < 1$. Also, f is *concave* if $-f$ is convex.

For differentiable functions, convexity is equivalent to $f'(x)$ increasing with x. Clearly, it is sufficient if $f''(x) \geqslant 0$ (and if $f''(x) > 0$, then f is strictly convex). In particular, x^p is strictly convex for $x > 0$ if $p > 1$ or $p < 0$, and strictly concave if $0 < p < 1$. Also, $\log x$ is strictly concave for $x > 0$.

The property that matters to us here is that a differentiable convex function lies above its tangents. In other words,

$$f(x) - f(x_0) \geqslant (x - x_0)f'(x_0) \qquad (1)$$

for x and x_0 in I (both for $x > x_0$ and for $x < x_0$). This is geometrically compelling; a formal proof is by a simple application of the mean-value theorem. Strict inequality holds if f is strictly convex and $x \neq x_0$.

If f is convex but not differentiable, then it still has right and left derivatives at interior points of I, and (1) still holds with $f'(x_0)$ replaced by either of these one-sided derivatives. This is not hard to prove, but we will not elaborate on it here, because all our applications will involve differentiable functions.

Majorisation and the main result

We shall denote an element $(x_1, x_2, \ldots, x_n)$ of $\mathbb{R}^n$ simply by x. We say that x is positive if each x_j is positive, and that x is decreasing if x_j decreases with j. The sum $\sum_{j=1}^{n} f(x_j)$ does not depend on the order of the x_j, and it will suit our purposes to assume that x is decreasing (after reordering if necessary).

https://doi.org/10.1017/mag.2022.14 Published online by Cambridge University Press

The essential idea driving the results that follow is comparison of partial sums. Let x and y be decreasing. Write $X_k = x_1 + \dots + x_k$ (similarly Y_k for a second element y). If $Y_k \leqslant X_k$ for each k, we write $y \leqslant_S x$. If also $Y_n = X_n$, we write $y \leqslant_M x$, and say that y is *majorised* by x. (We warn the reader that this is not standard notation; in fact, various different symbols have been used in the literature for these relations.)

So, for example, $(5, 4, 2) \leqslant_M (7, 3, 1)$ and $(6, 5, 5, 3) \leqslant_M (9, 4, 4, 2)$. For decreasing elements x, y of $\mathbb{R}^3$ with $X_3 = Y_3$, the statement $y \leqslant_M x$ is equivalent to $y_1 \leqslant x_1$ and $y_3 \geqslant x_3$.

Example 1: Let (x_j) be decreasing and $\bar{x} = X_n / n$. Let $y_j = \bar{x}$ for $1 \leqslant j \leqslant n$. It is elementary that X_k / k (the sequence of averages) decreases with k, hence for $k \leqslant n$, we have $X_k / k \geqslant \bar{x}$. So $Y_k = k\bar{x} \leqslant X_k$, hence $y \leqslant_M x$.

Along with (1), the second ingredient of our reasoning is the well-known Abel summation formula for finite sums: for any numbers a_j, x_j, we have

$$\sum_{j=1}^{n} a_j x_j = a_1 X_1 + \sum_{j=2}^{n} a_j \left(X_j - X_{j-1}\right) = \sum_{j=1}^{n-1} \left(a_j - a_{j-1}\right) X_j + a_n X_n. \quad (2)$$

The following Lemma is an obvious consequence of this identity.

Lemma: Suppose that x, y are elements of $\mathbb{R}^n$ with $y \leqslant_S x$. Suppose also that $a_1 \geqslant a_2 \geqslant \dots \geqslant a_n$ and that either $Y_n = X_n$ or $a_n \geqslant 0$. Then

$$\sum_{j=1}^{n} a_j y_j \leqslant \sum_{j=1}^{n} a_j x_j.$$

Proof: We apply (2). For $1 \leqslant j \leqslant n - 1$, we have

$$\left(a_j - a_{j+1}\right) Y_j \leqslant \left(a_j - a_{j+1}\right) X_j.$$

Also, under either hypothesis, $a_n Y_n \leqslant a_n X_n$.

Our main result now follows very pleasantly by combining the Lemma with inequality (1).

Proposition 1: Let x, y be decreasing elements of $\mathbb{R}^n$ with $y \leqslant_S x$. Suppose that the function f is convex on an open interval I containing all x_j and y_j, and that

either (i) $Y_n = X_n$ (so that $y \leqslant_M x$),
or (ii) f is increasing on I.

Then

$$\sum_{j=1}^{n} f\left(y_j\right) \leqslant \sum_{j=1}^{n} f\left(x_j\right). \quad (3)$$

https://doi.org/10.1017/mag.2022.14 Published online by Cambridge University Press

Proof: Assume first that f is differentiable on I. Then by (1),

$$f(x_j) - f(y_j) \geqslant (x_j - y_j)f'(y_j)$$

for each j. Since f is convex, $f'(t)$ increases with t, so $f'(y_j)$ decreases with j. Apply the Lemma, with $a_j = f'(y_j)$. Under either hypothesis, we conclude that $\sum_{j=1}^{n}(x_j - y_j)f'(y_j) \geqslant 0$: in case (ii), this follows from the fact that $f'(y_n) \geqslant 0$.

If f is not differentiable at some y_j, similar reasoning still applies with $f'(y_j)$ replaced by the right-derivative, as explained earlier.

This result is known as the 'majorisation principle'. It seems to have been first formulated by Hardy, Littlewood and Pólya in 1929 [1]. It was rediscovered by Karamata in 1932 [2], and it has also been called 'Karamata's inequality'.

Of course, the reverse of (3) holds if f is concave and either $Y_n = X_n$ or f is decreasing. Some further comments are helpful. First, *strict* inequality holds in (3) if f is strictly convex and $y_j \neq x_j$ for some j. Second, under condition (ii), it follows further that $\sum_{j=1}^{k} f(y_j) \leqslant \sum_{j=1}^{k} f(x_j)$ for each $k \leqslant n$, more closely reflecting the hypothesis $Y_k \leqslant X_k$ for each k. In this form the result extends to infinite sequences.

In our applications the numbers x_j, y_j will be positive and the interval I will be $x > 0$.

General sequences

For sequences that are not decreasing, Proposition 1 can be restated as follows (and often is in the literature). Let x^* be the vector consisting of the terms x_j arranged in decreasing order (the 'decreasing rearrangement' of x). Of course, the sum $\sum_{j=1}^{n} f(x_j)$ is unchanged by rearrangement. So Proposition 1 says that (3) holds if $y^* \leqslant_M x^*$, or if $y^* \leqslant_S x^*$ and f is increasing. The established terminology is that y is 'majorised' by x if $y^* \leqslant_M x^*$.

Applications

Without further proof, we can write down what Proposition 1 says when applied to x^p for different p:

Proposition 2: Let x, y be decreasing, positive elements of $\mathbb{R}^n$ with $y \leqslant_S x$. Then:

(i) if $p \geqslant 1$, then $\sum_{j=1}^{k} y_j^p \leqslant \sum_{j=1}^{k} x_j^p$ for each $k \leqslant n$.

(ii) if $0 < p < 1$ and also $Y_n = X_n$, then $\sum_{j=1}^{n} y_j^p \geqslant \sum_{j=1}^{n} x_j^p$.

(iii) if $p < 0$ and also $Y_n = X_n$, then $\sum_{j=1}^{n} y_j^p \leqslant \sum_{j=1}^{n} x_j^p$.

There is no question of (ii) or (iii) applying without the condition $Y_n = X_n$, since this would allow each y_j to be arbitrarily small.

We mention that the case $n = 3$ was stated in [3], with a different proof that is implied but not given in detail.

Using the logarithmic and exponential functions, we can relate sums to products.

Proposition 3: Let x, y be distinct, decreasing, positive elements of $\mathbb{R}^n$. If $y \leqslant_M x$, then $y_1 y_2 \ldots y_n > x_1 x_2 \ldots x_n$.

Proof: Apply Proposition 1 with $f(t) = \log t$: since f is concave, (3) is reversed.

By Example 1, this applies, in particular, when $y_j = \bar{x}$ for each j, then stating that $\bar{x}^n \geqslant x_1 x_2 \ldots x_n$, in other words, the arithmetic mean $\bar{x}$ is not less than the geometric mean.

Note: It is quite possible to have $\sum_{j=1}^{n} x_j = \sum_{j=1}^{n} y_j$ and $x_1 x_2 \ldots x_n = y_1 y_2 \ldots y_n$, for example (12, 5, 4) and (10, 8, 3). In fact, if $n = 3$ and x_1, x_2, y_1, y_2 are given, then x_3 and y_3 are determined by a simple pair of linear equations. What Proposition 3 tells us is that this cannot happen with either $y \leqslant_S x$ or $x \leqslant_S y$.

In the opposite direction, we can prove the following result.

Proposition 4: Let x, y be decreasing, positive elements of $\mathbb{R}^n$. If

$$y_1 y_2 \ldots y_k \leqslant x_1 x_2 \ldots x_k$$

for each $k \leqslant n$, then for any $p > 0$, we have $\sum_{j=1}^{k} y_j^p \leqslant \sum_{j=1}^{k} x_j^p$ for each $k \leqslant n$.

Proof: Then $\sum_{j=1}^{k} \log y_j \leqslant \sum_{j=1}^{n} \log x_j$ for each $k \leqslant n$. Let $f(t) = e^{pt}$. Then f is convex and increasing, and $f(\log y_j) = y_j^p$. The conclusion follows, by Proposition 1.

In particular, the hypothesis in Proposition 4 implies that $y \leqslant_S x$. This, with Proposition 2, implies the stated inequality for $p \geqslant 1$, but not for $0 < p < 1$.

https://doi.org/10.1017/mag.2022.14 Published online by Cambridge University Press

Example 2: By Proposition 4, we have $4^p + 3^p \leqslant 6^p + 2^p$ for all $p > 0$. (Readers are invited to satisfy themselves that this is not at all easy to prove by other methods.)

We describe a more general example illustrating the results. It is taken from the article [4], which contains a number of examples presented as problems: this one is Problem 3.

Example 3: Let

$$y = (a, b, c) \qquad x = (a + b - c, c + a - b, b + c - a),$$

where $a \geqslant b \geqslant c > 0$ and $b + c > a$. Then $x_1 \geqslant x_2 \geqslant x_3 > 0$ and $y \leqslant_M x$, since $x_1 \geqslant a$, also $X_2 = 2a \geqslant a + b$ and $X_3 = a + b + c$. So by Propositions 2 and 3, we have

$$a^p + b^p + c^p \leqslant (b + c - a)^p + (c + a - b)^p + (a + b - c)^p$$

for $p > 1$, together with the opposite inequality for $0 < p < 1$, also

$$\frac{1}{a} + \frac{1}{b} + \frac{1}{c} \leqslant \frac{1}{b + c - a} + \frac{1}{c + a - b} + \frac{1}{a + b - c}.$$

$$(b + c - a)(c + a - b)(a + b - c) \leqslant abc.$$

The pair (5, 4, 2) and (7, 3, 1), seen earlier, is a particular case. One can easily verify some actual values, for example $5^2 + 4^2 + 2^2 = 45$, while $7^2 + 3^2 + 1^2 = 59$.

We now describe a further application involving x^p, in which there is no longer any assumption about partial sums.

Proposition 5: Let x, y be non-negative elements of $\mathbb{R}^n$, both decreasing or both increasing. If $p > 1$, then

$$\left(\sum_{j=1}^{n} x_j^p y_j^p\right)\left(\sum_{j=1}^{n} x_j\right)^p \geqslant \left(\sum_{j=1}^{n} x_j y_j\right)^p \left(\sum_{j=1}^{n} x_j^p\right). \tag{4}$$

The reverse inequality holds if $0 < p < 1$.

Proof: We prove the statement for the case where (x_j) and (y_j) are decreasing. The case where they are increasing then follows by considering x_j and y_j in reversed order.

Write $z_j = x_j y_j$ and $Z_n = cX_n$. We show that $cX_k \leqslant Z_k$ for $k \leqslant n$. By Proposition 2, it then follows, for $p > 1$, that $c^p \sum_{j=1}^{n} x_j^p \leqslant \sum_{j=1}^{n} z_j^p$, which equates to (4).

Let $Z_k = c_k X_k$, so $c_n = c$. We have to show that $c_k \geqslant c$. This will follow if we show that $c_{k+1} \leqslant c_k$ for each $k < n$. Now $Z_k = \sum_{j=1}^{k} x_j y_j \geqslant y_k X_k$,

https://doi.org/10.1017/mag.2022.14 Published online by Cambridge University Press

so $c_k \geqslant y_k$, hence $c_k \geqslant y_{k+1}$. Now

$$Z_{k+1} = c_k X_k + x_{k+1} y_{k+1} \leqslant c_k \left(X_k + x_{k+1}\right) = c_k X_{k+1},$$

hence $c_{k+1} \leqslant c_k$, as required.

This result can be restated neatly in terms of ℓ_p-norms. Define $\|x\|_p$ to be $\left(\sum_{j=1}^{n} |x_j|^p\right)^{1/p}$ (note that $\|x\|_2$ is the ordinary Euclidean norm). Then (4) equates to

$$\frac{\|xy\|_p}{\|xy\|_1} \geqslant \frac{\|x\|_p}{\|x\|_1}.$$

Some further results

There is a converse to Proposition 1, essentially showing that the property stated there characterises majorised pairs. For this we will use non-differentiable convex functions, in fact functions of the following form: for fixed x, let

$$f(t) = (t - x)^+ = \begin{cases} t - x & \text{if } t \geqslant x, \\ 0 & \text{if } t < x. \end{cases}$$

Such functions are obviously convex and increasing.

Proposition 6: Let x, y be decreasing elements of $\mathbb{R}^n$. If (3) holds for all increasing, convex f, then $y \leqslant_S x$. If (3) holds for all convex f, then $y \leqslant_M x$.

Proof: Suppose first that (3) holds for all convex f. Then it holds, in particular, for $f(t) = \pm t$. This implies at once that $Y_n = X_n$.

Now suppose that (3) holds for increasing, convex f. Choose $k \leqslant n$, and let $f(t) = (t - x_k)^+$. Then f is convex and increasing, and

$$\sum_{j=1}^{n} f\left(x_j\right) = \sum_{j=1}^{k} \left(x_j - x_k\right) = X_k - kx_k.$$

Also, since $f(t)$ is not less than both $t - x_k$ and 0 for all t, we have

$$\sum_{j=1}^{n} f\left(y_j\right) \geqslant \sum_{j=1}^{k} \left(y_j - x_k\right) = Y_k - kx_k,$$

hence $Y_k \leqslant X_k$.

Finally, there is a continuous version of majorisation, in which integrals replace discrete sums. The proof is analogous, but Abel summation is replaced by integration by parts.

Proposition 7: Let x, y be decreasing, differentiable functions on $[a, b]$, with values in an interval I. Write $X(t) = \int_a^t x(s)\, ds$, similarly $Y(t)$. Suppose that $Y(t) \leqslant X(t)$ for $a \leqslant t \leqslant b$. Let f be a function that is convex and twice differentiable on I. Suppose further that either $Y(b) = X(b)$ or f is increasing. Then

$$\int_a^b f[y(t)]\, dt \leqslant \int_a^b f[x(t)]\, dt.$$

Proof: By inequality (1),

$$f[x(t)] - f[y(t)] \geqslant [x(t) - y(t)] f'[y(t)].$$

Integrating by parts, we find

$$\int_a^b [x(t) - y(t)] f'[y(t)]\, dt$$

$$= [[X(t) - Y(t)] f'[y(t)]]_a^b - \int_a^b [X(t) - Y(t)] f''[y(t)]\, y'(t)\, dt$$

$$= [X(b) - Y(b)] f'[y(b)] - \int_a^b [X(t) - Y(t)] f''[y(t)]\, y'(t)\, dt.$$

Under either of the alternative hypotheses, the first term is non-negative. Since $f''[y(t)] \geqslant 0$ and $y'(t) \leqslant 0$, the second term is non-negative.

Applications analogous to Propositions 2 and 5 can be derived as before.

There is a substantial further body of theory concerning majorisation and its applications. An account of it can be seen in chapters 2 and 3 of [5], and it forms the topic of the whole book [6]. More recent developments are given in [7]. Here we just mention without proof a purely algebraic characterisation [5, pp. 46-49]. An $n \times n$ matrix $P = (p_{j.k})$ is *doubly stochastic* if the entries are non-negative and all row sums and all column sums equal 1:

$$\sum_{j=1}^{n} p_{j,k} = 1 \text{ for each } k, \qquad \sum_{k=1}^{n} p_{j,k} = 1 \text{ for each } j.$$

The statement $y^* \leqslant_M x^*$ is equivalent to the existence of a doubly stochastic matrix P (not necessarily unique) such that $y = Px$. One way round is quite easy: if $y = Px$, so that $y_j = \sum_{k=1}^{n} p_{j,k} x_k$, then Jensen's inequality gives $f(y_j) \leqslant \sum_{k=1}^{n} p_{j,k} f(x_k)$ for convex f. Summation over j then gives (3). By Proposition 6, this implies that $y^* \leqslant_M x^*$.

https://doi.org/10.1017/mag.2022.14 Published online by Cambridge University Press

Acknowledgement

The author is grateful to the referee for several useful suggestions and for drawing attention to reference [4].

References

1. G. H. Hardy, J. Littlewood and G. Pólya, Some simple inequalities satisfied by convex functions, *Messenger Math*. **58** (1929) pp. 145-152.
2. J. Karamata, Sur une inégalité relative aux fonctions convexes, *Publ. Math. Univ. Belgrade* **1** (1932) pp. 145-148.
3. G. J. O. Jameson, Monotonicity of the midpoint and trapezium estimates for integrals, *Math. Gaz*. **105** (2021) pp. 433-441.
4. Z. Kadelburg, D. Dukić, M. Lukić and I. Matić, Inequalities of Karamata, Schur and Muirhead, and some applications, *Teaching of Math*. **8** (2005) pp. 31-45.
5. G. H. Hardy, J. Littlewood and G. Pólya, *Inequalities* (2nd edn.), Cambridge University Press (1967).
6. A. W. Marshall and I. Olkin, *Inequalities*: *Theory of majorization and its applications*, Academic Press (1979).
7. Grahame Bennett, Some forms of majorization, *Houston J. Math*. **36** (2010) pp. 1037-1066.

G. J. O. JAMESON
13 Sandown Road,
Lancaster LA1 4LN
e-mail: *pgjameson@talktalk.net*