References

[1] R.B. Burckel, Iterating analytic self-maps of discs, Amer. Math. Monthly, 88(1981), 396-406.

[2] A. Carlsson, Om Iterader Funktioner, Ph.D., Uppsala (1907).

[3] L. Euler, De formulis exponentialibus replicatus, Opera Omnia, Series Primus XV, 268-297; Acta Academiae Petropolitanae, 1 (1777), 38-60.

[4] R.A. Knoebel, Exponentials reiterated, Amer. Math. Monthly, 88(1981), 235-252.

[5] A.J. Macintyre, Convergence of $i^{i^{\cdots}}$, Proc. Amer. Math. Soc., 17(1966), 67.

[6] D.L. Shell, On the convergence of infinite exponentials, Proc. Amer. Math. Soc., 13(1962), 678-681.

[7] W.J. Thron, Convergence of infinite exponentials with complex elements, Proc. Amer. Math. Soc., 8(1957), 1040-1043.

University College,
Cork.

Majorization and Schur Functions

Philip J. Boland
University College Dublin

The concepts of majorization and Schur functions lay the basis for a rich and elegant theory in which many classical and applicable inequalities may be viewed. In this expository paper the basic definitions and properties of majorization and Schur functions are presented, together with a variety of applications emphasizing in particular some in reliability theory. For a thorough and recent account of majorization and Schur functions, the interested reader should consult the excellent Inequalities: Theory of Majorization and its Applications by Marshall and Olkin (1979).

1. Majorization

Given a vector $\underline{x} = (x_1,\ldots,x_n) \in R^n$, let $x_{[1]} \geq \ldots \geq x_{[n]}$ denote a decreasing rearrangement of $x_1,\ldots,x_n$.

Definition 1.1 If $\underline{x},\underline{y} \in R^n$, then $\underline{x} \prec \underline{y}$ if

$$\sum_1^j x_{[i]} \leq \sum_1^j y_{[i]} \qquad \text{for } j = 1,\ldots,n-1$$

$$\text{and} \qquad \sum_1^n x_{[i]} = \sum_1^n y_{[i]}.$$

If $\underline{x} \prec \underline{y}$, we say that $\underline{x}$ is majorized by $\underline{y}$. Note that if $\underline{x} \prec \underline{y}$, then the components of $\underline{y}$ are more "spread out" than those of $\underline{x}$. For example $(\frac{1}{n},\ldots,\frac{1}{n},\frac{1}{n}) \prec (\frac{1}{n-1},\ldots,\frac{1}{n-1},0) \prec (1,0,\ldots,0)$, and $(\bar{x},\bar{x},\ldots,\bar{x}) \prec (x_1,\ldots,x_n)$ where $\bar{x} = \Sigma x_i/n$.

One of the origins of majorization can be found in the work of Schur (1923) on Hadamard's determinant inequality (which states that for any $n \times n$ positive semi-definite Hermitian matrix $M = (m_{ij})$, $\det M \leq \prod_i^n m_{ii}$). Preliminary to proving this result, Schur showed that the diagonal elements $m_{11}, \ldots, m_{nn}$ of a positive semi-definite Hermitian matrix M are majorized by the characteristic roots $(\lambda_1, \ldots, \lambda_n)$. Horn (1954) later showed that this relationship actually characterizes those vectors $\underline{m} = (m_{11}, \ldots, m_{nn})$ and $\underline{\lambda} = (\lambda_1, \ldots, \lambda_n)$ that can arise together as respectively the diagonal and characteristic root vectors of the same Hermitian matrix.

Many basic inequalities reduce to an inequality of the form $f(\bar{y}, \ldots, \bar{y}) \leq f(y_1, \ldots, y_n)$ for some appropriate f. This suggests perhaps considering comparisons of the type $f(x_1, \ldots, x_n) \leq f(y_1, \ldots, y_n)$ where $\underline{x} \prec \underline{y}$. Hardy, Littlewood and Polya (1923) asked the following question: what conditions on $\underline{x} = (x_1, \ldots, x_n)$ and $\underline{y} = (y_1, \ldots, y_n)$ ensure that

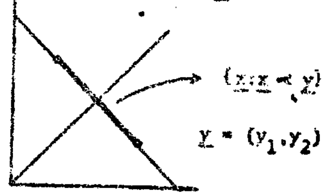
$$\Sigma g(x_i) \leq \Sigma g(y_i)$$

for all convex functions $g: R \to R$? They proved in fact that * is true for all convex g if and only if $\underline{x} \prec \underline{y}$.

Majorization is inherent in the work of economists studying income inequality in the early part of this century. Lorenz (1905) did so in introducing what is now known as a "Lorenz curve". Let $\underline{x} = (x_1, \ldots, x_n)$ be the wealth vector for a population of size n, i.e. x_i is the wealth of individual i. We let $S_0 = 0$, and $S_k = \sum_{i=n-k+1}^{n} x_{[i]}$ be the total wealth of the k poorest individuals. If we plot the points $(k/n, S_k/S_n)$ for $k = 0, \ldots, n$ and connect them in a linear fashion we obtain the Lorenz curve for the wealth vector $\underline{x}$. The Lorenz curve is always convex and is a straight line if and only if the wealth is uniformly distributed. Suppose now $\underline{y}$ is another wealth vector from a population of size n. One would say that the wealth distribution of $\underline{x}$ is more equal than that of $\underline{y}$ if the Lorenz curve

of $\underline{x}$ lies above that of $\underline{y}$. When the total wealth of the two populations is the same this is equivalent to saying that $\underline{x} \prec \underline{y}$.

Hardy, Littlewood and Polya (1929) showed that $\underline{x} \prec \underline{y}$ if and only if there is a doubly stochastic matrix P (a matrix with nonnegative elements whose rows and columns both sum to one) such that $\underline{x} = \underline{y}P$. "Hitting" a vector $\underline{y}$ with a doubly stochastic matrix P has the effect of averaging or smoothing out its components. Birkhoff (1946) proved that the set of doubly stochastic matrices is the convex hull of the permutation matrices (and moreover that the permutation matrices are the extreme points of this set). Birkhoff's result together with the above characterization of Hardy, Littlewood and Polya enable one to show that for a given $\underline{y}$, $\{\underline{x}: \underline{x} \prec \underline{y}\}$ is the convex hull of the orbit of $\underline{y}$ under permutations (the set of points obtained by permuting the components of $\underline{y}$).



2. Schur Functions

Majorization represents a partial ordering on R^n. A Schur function is a real valued function which is monotone with respect to this ordering.

Definition 2.1 A function f satisfying the property that $f(\underline{x}) \leq (\geq) f(\underline{y})$ whenever $\underline{x} \prec \underline{y}$ is Schur convex (concave). Functions which are either Schur convex or Schur concave are called Schur functions.

Note that a Schur function is necessarily symmetric or permutation invariant, that is $f(\underline{x}) = f(\underline{x}^{\Pi})$ where $\underline{x}^{\Pi}$ is an arbitrary rearrangement of the coordinates of $\underline{x}$.

The terminology "Schur convex (concave)" is rather misleading. Although a symmetric convex function on R^n is Schur convex, a Schur convex function may be far from convex in any usual sense of the word. "Schur increasing" would perhaps be more appropriate than "Schur convex", although (unfortunately) for historical reasons the later term is now conventional.

A useful characterization of Schur functions is given by the following result of Ostrowski (1952), sometimes referred to as the Schur Ostrowski condition.

Theorem 2.2 Let $A \subset R^n$ be convex and permutation invariant with non empty interior. If $f: A \to R$ is continuously differentiable on the interior of A and continuous and symmetric on A, then

$$f \text{ is Schur convex (concave)}$$
$$\Longleftrightarrow$$
$$(x_i - x_j)\left(\frac{\partial f}{\partial x_i}(\underline{x}) - \frac{\partial f}{\partial x_j}(\underline{x})\right) \geq (\leq)\ 0 \quad \text{for all } i \neq j.$$

Example 2.3 Let $S_k(\underline{x})$ be the k^{th} elementary symmetric function of $\underline{x}$, for $k = 0,1,\ldots,n$. That is $S_0(\underline{x}) \equiv 1$, $S_1(\underline{x}) = \Sigma x_i$, $S_2(\underline{x}) = \sum_{i<j} x_i x_j$, $S_3(\underline{x}) = \sum_{i<j<k} x_i x_j x_k, \ldots,$ $S_n(\underline{x}) = x_1 \cdots x_n$. Verifying the Schur Ostrowski condition, one can see that $S_k(\underline{x})$ is Schur concave on $(0,\infty)^n$ for $k = 0,1,\ldots,n$.

Example 2.4 If g is a convex (concave) function of one real variable, then $f(\underline{x}) = \Sigma g(x_i)$ is Schur convex (concave). This enables one to construct many Schur functions.

a) $f(\underline{x}) = \Sigma \frac{1}{x_i}$ is Schur convex on $(0,\infty)^n$. One may use this result to prove an inequality due to Schweitzer (1914):

Let $0 < m \leq a_i \leq M$ for $i = 1,\ldots,n$. Then

$$\left(\frac{1}{n}\Sigma a_i\right)\left(\frac{1}{n}\Sigma\frac{1}{a_i}\right) \leq \frac{(M+m)^2}{4mM}$$

b) $H(\underline{p}) = -\Sigma p_i \log p_i$ is Schur concave on $(0,1)^n$. $H(\underline{p})$ is called the entropy of $\underline{p}$ when $\Sigma p_i = 1$. Hence the entropy of $\underline{p}$ increases as the p_i's become "more equal".

c) $s(\underline{x}) = \left(\frac{1}{n}\Sigma(x_i - \bar{x})^2\right)^{1/2}$ is Schur convex on R^n. $s(\underline{x})$ is the sample "standard deviation" for the sample vector $\underline{x}$.

One may show that a convex symmetric real valued function f is Schur convex. If f is convex, there are various methods of symmetrizing f while preserving its convexity (and hence generating a Schur convex function). Techniques of this sort enable one to prove a famous inequality due to Muirhead (1903) and Hardy, Littlewood and Polya (1934).

Theorem 2.5 Let $\underline{y} = (y_1,\ldots,y_n)$ where $y_i > 0$ for $i = 1,\ldots,n$. If $\underline{a} \prec \underline{b}$, then

$$\sum_{\Pi} y_{\Pi(1)}^{a_1} \cdots y_{\Pi(n)}^{a_n} \leq \sum_{\Pi} y_{\Pi(1)}^{b_1} \cdots y_{\Pi(n)}^{b_n}.$$

Note in particular that taking $\underline{a} = (\frac{1}{n},\ldots,\frac{1}{n})$ and $\underline{b} = (1,0,\ldots,0)$, one obtains the arithmetic-geometric mean inequality:

$$(y_1 \cdots y_n)^{1/n} \leq \Sigma y_i / n.$$

3. Applications in Reliability Theory

Definition 3.1 A system with n independent components which functions if and only if at least k of the components function is a k out of n system.

A parallel system is a 1 out of n system, an n-1 out of n system is called a fail-safe system, and an n out of n system is a series system.

If $\underline{p} = (p_1,\ldots,p_n)$ is the vector of components reliabilities (that is p_i = probability that component i functions), then

$$h_k(\underline{p}) = \sum_{\epsilon_1+\cdots+\epsilon_n \geq k} p_1^{\epsilon_1} \cdots p_n^{\epsilon_n} (1-p_1)^{1-\epsilon_1} \cdots (1-p_n)^{1-\epsilon_n}$$

(where ϵ_i is either 1 or 0)

is the probability that k or more of the components function. $h_k(\underline{p})$ is called the reliability function for a k out of n system. Note also that $h_k(\underline{p})$ may be interpreted as the probability of k or more successes in n independent Bernoulli trials with respective success probabilities $p_1,\ldots,p_n$.

Using the Schur Ostrowski characterization of Schur functions and a monotonicity result, one obtains the following Theorem (Boland-Proschan, 1982):

Theorem 3.2 The reliability function $h_k(\underline{p})$ of a k out of n system is Schur convex in $\left[\frac{k-1}{n-1}, 1\right]^n$ and Schur concave in $\left[0, \frac{k-1}{n-1}\right]^n$.

If k=1, that is we are considering a parallel system, the above result says that $h_k(\underline{p})$ is Schur convex on $[0,1]^n$. This means that subject to the constraint that Σp_i is constant, the more spread out the component reliabilities are the more reliable the system is. When considering a series system (k=n), the opposite is true - subject to the constraint that Σp_i is constant, the more equal the component reliabilities are the more reliable the system is.

Example 3.3 Let us consider a 3 out of 4 system. If $\underline{p} = (p_1,p_2,p_3,p_4)$ is the vector of component reliabilities, then

a) (1.0,.9,.8,.7) yields higher reliability than (.95,.95,.75,.75) which in turn is superior to (.85,.85,.85,.85).

b) (.6,.5,.3,.2) is inferior to (.6,.4,.4,.2) which in turn is inferior to (.4,.4,.4,.4).

If $h(\underline{p})$ is the reliability function of a system, one can measure the importance of a component in contributing to system reliability by the rate at which system reliability improves as the reliability of the component improves. More specifically one can define the reliability importance $I_h(j)$ of component j as $I_h(j) = \frac{\partial h}{\partial p_j}(\underline{p})$. (See Barlow-Proschan, 1975).

Now let us consider again a k out of n system. Without loss of generality let us assume that component reliabilities $\underline{p} = (p_1,\ldots,p_n)$ are such that $p_1 \leq p_2 \leq \cdots \leq p_n$. Using the Schur Ostrowski condition and Theorem 3.2, it follows that

a) whenever $\underline{p} \in \left[\frac{k-1}{n-1}, 1\right]^n$ the most reliable component (component n) is the most important to the system, and

b) whenever $\underline{p} \in \left[0, \frac{k-1}{n-1}\right]^n$ the least reliable component is the most important to the system.

Note that this says that for example for parallel systems, the component with highest reliability is the most important to the system. This is intuitively clear as the system functions if only one component functions. On the other hand for a series system, the weakest component is the most important to the system. This reflects the well known adage that a chain is as strong as its weakest link.

References

1. Barlow, R.E. and Proschan, F., (1965). Mathematical Theory of Reliability. John Wiley and Sons, New York.

2. Barlow, R.E. and Proschan, F., (1975). Statistical Theory of Reliability and Life Testing: Probability Models. Holt, Rinehart & Winston, New York.

3. Birkhoff, G., (1946). "Tres observaciones sobre el algebra lineal". Univ. Nac. Tucuman Rev. Ser. A 5, 147-151.

4. Boland, P.J. and Proschan, F., (1982). "The Reliability of k out of n systems". Preprint.

5. Hardy, G.H., Littlewood, J.E., and Polya, G., (1929). "Some simple inequalities satisfied by convex functions". Messenger Math. 58 145-152.

6. Hardy, G.H., Littlewood, J.E., and Polya, G., (1934,1952). Inequalities. 1st ed., 2nd ed. Cambridge Univ. Press, London and New York.

7. Horn, A., (1954). "Doubly stochastic matrices and the diagonal of a rotation matrix". Amer. J. Math. 76, 620-630.

8. Lorenz, M.O., (1905). "Methods of measuring concentration of wealth". J. Amer. Statist. Assoc. 9, 209-219.

9. Marshall, A.W., and Olkin, I., (1979). Inequalities: Theory of Majorization and its Applications, Academic Press, New York.

10. Muirhead, R.F., (1903). "Some methods applicable to identities and inequalities of symmetric algebraic functions of n letters". Proc. Edinburgh Math. Soc. 21, 144-157.

11. Ostrowski, A.M., (1952). "Sur quelques applications des fonctions convexes et concaves au sens de I. Schur". J. Math. Pures Appl. 31, 253-292.

12. Schur, I., (1923). "Über eine Klasse von Mittelbildungen mit Anwendungen die Determinanten" - Theorie Sitzungsber. Berlin. Math. Gesellschaft 22, 9-20.

13. Schweitzer, P., (1914). "An inequality concernign the arithmetic mean" (in Hungarian). Math.-Phys. Lapok 23, 257-261.

NEWS FOR COMPLEX VARIABLES TEACHERS

Aongus Ó Cairbre and Anthony G. O'Farrell

It is good pedagogical practice to mix solid applications (inside or outside mathematics) with the development of general theory [3]. This is routinely done in general courses, but it is just as important in honours courses, because otherwise the students may get the wrong idea of what mathematics is and how it is done. Mathematics is best done with a specific problem in mind.

People who agree with this point of view will be interested to learn that two major applications of complex function theory have recently been simplified to the point where both can now be presented to average honours undergraduates. Hitherto they were, realistically speaking, first year postgraduate material. The results are the big Picard theorem and the prime number theorem.

The original proof of Picard's theorem, using the elliptic modular function and monodromy, remains firmly at postgraduate level. Of course it was undergraduate material long ago, when it was acceptable to be vague about topological problems. Until last year, the proof normally used was basically that in Landau's "Neuere Ergebnisse", via the theorems of Bloch and Schottky. The new proof is a simplification of this latter proof. It gets the result in one page after Schottky's theorem. The entire proof, assuming the maximum principle, Rouché's theorem, and a knowledge of the logarithm and complex powers, may be presented in two lectures. The new idea is due to Bridges, Calder, Julian, Mines, and Richman, and is explained in [2]. Curiously enough, they found this simpler proof, not because they were trying to, but because they sought a constructive proof, i.e. one not using the apparatus of normal families.

The new proof of the prime number theorem is due to Newman [6]. Until it appeared, the simplest proof was that in Heins' "Topics". The latter proof involved the Riemann-Lebesgue lemma and many technical convergence details. Newman actually offered two proofs. He started by giving an ingenious proof of a Tauberian theorem of Ingham. He observed that Landau's equivalent form of the prime number theorem follows at once. He went on to give the details of another proof, based on the fact that the existence of the limit