Journal of Convex Analysis
Volume 14 (2007), No. 1, 103–108

An Effective Characterization of Schur-Convex Functions with Applications

Czesław Stępniak
Institute of Mathematics, University of Rzeszów,
Al. Rejtana 16 A, 35-310 Rzeszów, Poland
cees@univ.rzeszow.pl

Received: June 29, 2004
Revised manuscript received: December 20, 2005

An effective necessary and sufficient condition for S-convexity is presented and applications are given.

Keywords: S-convex function, majorization

1991 Mathematics Subject Classification: Primary 15A51; Secondary 26B25, 26D15

1. Introduction and summary

In his paper concerning Hadamard´s inequality Schur[7] introduced and characterized a class of functions which were later called Schur-convex or simply S-convex. This characterization was completed by Ostrowski [6]. Hardy, Littlewood and Pólya attached importance to these functions in presenting and proving many results in their book [3], while Marshall and Olkin [5] used the S-functions as a key in creating an unified theory of inequalities.

Both Schur and Ostrowski characterized the S-functions in terms of their differential coefficients. Some further results described relations between S-convexity and convexity (in the sense of Jensen). Among others, it is now well known, that any symmetric convex function is S-convex, but not vice versa.

The aim of this paper is to present a simple and effective characterization of S-convexity. As we will see, this characterization leads directly to some known and unknown results in this subject. Applications of the characterization are also provided.

2. Definitions and main result

Let us consider a subset $\mathcal{A}$ of the real euclidean space R^n, where $n > 1$. Any element of $\mathcal{A}$ will be presented as a row-vector $x = (x_1, ...x_n)$. The set $\mathcal{A}$ is said to be *symmetric* if $(x_1, ..., x_n) \in \mathcal{A}$ implies that $(x_{i_1}, ..., x_{i_n}) \in \mathcal{A}$ for any permutation $x_{i_1}, ..., x_{i_n}$ of the numbers $x_1, ..., x_n$. Moreover, a real function φ on a symmetric set $\mathcal{A}$ is said to be *symmetric*, if $\varphi(x_1, ..., x_n) = \varphi(x_{i_1}, ..., x_{i_n})$. We say that a vector y in a symmetric set

ISSN 0944-6532 / \$ 2.50

$\mathcal{A} \subseteq R^n$ *majorizes* $x \in \mathcal{A}$, and write $x \prec y$, if

$$\max\{\sum_{j=1}^{k} x_{i_j} : 1 \leq i_1 < i_2 < ... < i_k \leq n\}$$
$$\leq \max\{\sum_{j=1}^{k} y_{i_j} : 1 \leq i_1 < i_2 < ... < i_k \leq n\} \text{ for } k = 1, ..., n-1$$

$$\text{and } \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i.$$

The relation $\prec$ possesses the usual properties of preordering, i.e.

(∗) $x \prec x$ for any $x \in \mathcal{A}$,
(∗∗) if $x \prec y$ and $y \prec z$ then $x \prec z$ for all $x, y, z \in \mathcal{A}$.

A real function φ on a convex set $\mathcal{A} \subseteq R^n$ is said to be *convex* if

$$\varphi(\lambda x + (1-\lambda)y) \leq \lambda\varphi(x) + (1-\lambda)\varphi(y) \text{ for any } x, y \in \mathcal{A}. \text{ and } \lambda \in (0,1). \tag{1}$$

Let us mention (cf. [3]) that any continuous function φ on a convex set $\mathcal{A}$ is convex, if and only if,

$$\varphi(\frac{x+y}{2}) \leq \frac{1}{2}[\varphi(x) + \varphi(y)] \text{ for any } x, y \in \mathcal{A}. \tag{2}$$

The following lemma will be useful in Section 3.

Lemma 2.1 (cf. [5], Proposition 16.B.7). *Let f be a convex function on a convex set $\mathcal{A}$ and g be a nondecreasing convex function on R. Then the composition $\phi(x) = g(f(x))$ is convex on $\mathcal{A}$. Moreover, if f is strictly convex and g is increasing strictly convex then ϕ is strictly convex.*

Proof. By definition of ϕ, $\phi(\lambda x + (1-\lambda)y) = g(f(\lambda x + (1-\lambda)y))$. Since f is convex and g is nondecreasing, we get $g(f(\lambda x + (1-\lambda)y)) \leq g(\lambda f(x) + (1-\lambda)f(y))$. Now the desired result follows by the convexity of g. □

Definition 2.2. A real function φ on a symmetric set $\mathcal{A}$ is said to be Schur-convex (or simply S-convex) if it is isotonic relative to the preordering $\prec$, that is, if $x \prec y$ implies $\varphi(x) \leq \varphi(y)$ for any $x, y \in \mathcal{A}$. The function φ is said to be strictly S-convex if the strict inequality $\varphi(x) < \varphi(y)$ does hold for any pair $x, y \in \mathcal{A}$, such that $x \prec y$ but y can not be obtained from x by permutation of its components.

Remark 2.3. By definition of the preordering $\prec$, any function ϕ is (strictly) S-convex, if and only if, for arbitrary real functions g and h, such that g is positive, the function $g(\sum_{i=1}^{n} x_i)\phi(x) + h(\sum_{i=1}^{n} x_i)$ is (strictly) S-convex. Hence, S-convex function may not be continuous, or even measurable.

It is well known (cf. [6, 7]) that a continuously differentiable function φ defined on the cartesian product I^n of an open interval $I \subseteq R$ is S-convex, if and only if,

(a) $\varphi(x) = \varphi(x_1, ..., x_n)$ is a symmetric, and,

(b) $(x_i - x_j)(\varphi_{(i)}(x) - \varphi_{(j)}(x)) \geq 0$ for all $x \in I^n$, where $\varphi_{(k)}(x) = \frac{\partial \varphi(x)}{\partial x_k}$.

It is also known, that any symmetric convex function φ defined on a symmetric convex set $\mathcal{A}$ in R^n is S-convex (see [7], or [5]).

We shall prove the following

Theorem 2.4. *Let φ be any function defined on a symmetric convex set $\mathcal{A}$ in R^n. Then*

(A) *The function φ is S-convex if and only if*

- (i) $\forall (x_1, ...x_n) \in \mathcal{A}, \forall 1 \leq i < j \leq n$, $\varphi(x_1, ..., x_{i-1}, x_i, x_{i+1}, ..., x_{j-1}, x_j, x_{j+1}, ..., x_n) = \varphi(x_1, ..., x_{i-1}, x_j, x_{i+1}, ..., x_{j-1}, x_i, x_{j+1}, ..., x_n)$, *and*
- (ii) $\forall (x_1, ...x_n) \in \mathcal{A}, \forall\ 0 < \lambda < 1$, $\varphi(\lambda x_1 + (1-\lambda)x_2, \lambda x_2 + (1-\lambda)x_1, x_3, ..., x_n) \leq \varphi(x_1, .., x_n)$.

(B) *The function φ is strictly S-convex if and only if it satisfies the conditions (i) and (ii) with the strict inequality $<$ instead of $\leq$ for all $(x_1, ..., x_n) \in \mathcal{A}$ such that $x_1 \neq x_2$.*

Remark 2.5. The condition (i) is equivalent to symmetry of φ, while (ii) is weaker than either convexity or quasiconvexity (see [5], Sec. 3.C, for definition).

Proof. Part (A). Necessity of the conditions (i) and (ii).

Necessity of (i) is evident. To show the necessity of (ii) we note that $(\lambda x_1 + (1-\lambda)x_2, \lambda x_2 + (1-\lambda)x_1, x_3, ..., x_n) = xT$, where

$$T = \begin{bmatrix} \lambda & 1-\lambda & 0 & 0 & ... & 0 \\ 1-\lambda & \lambda & 0 & 0 & ... & 0 \\ 0 & 0 & 1 & 0 & ... & 0 \\ . & . & . & . & . & . \\ 0 & 0 & 0 & 0 & ... & 1 \end{bmatrix}.$$

Since T is a doubly stochastic matrix, we get $xT \prec x$ (cf. [5], Th. 2.A.4). Now the necessity of (ii) follows by definition of S-convexity.

Sufficiency of (i) and (ii). Let x and y be arbitrary vectors in $\mathcal{A}$ such that $x \prec y$. Then, by Theorem of Muirhead (see [3], or [5], Lemma 2.B.1) there exist matrices $T_k, k = 1, ...n-1$, such that $x = yT_1T_2...T_{n-1}$, while each T_k may be rewritten in the form (t_{ij}) with

$$t_{ij} = \begin{cases} \lambda & \text{if either } i = j = r \text{ or } i = j = s, \\ 1 & \text{if } \ i = j \neq r, s, \\ 1-\lambda & \text{if either } i = r \text{ and } j = s \text{ or } i = s \text{ and } j = r, \\ 0 & \text{otherwise,} \end{cases}$$

where $r = r(k)$ and $s = s(k)$ are positive integers not greater than n and $\lambda = \lambda(k)$ belongs to the interval $(0, 1)$. Let us define vectors $y^{(k)}, k = 0, 1, ..., n-1$, by setting $y^{(0)} = y$ and $y^{(k)} = y^{(k-1)}T_k$ for $k = 1, .., n-1$. To see that $\varphi(x) \leq \varphi(y)$ we only need to notice that the conditions (i) and (ii) imply $\varphi(y^{(k)}) \leq \varphi(y^{(k-1)})$ for $k = 1, ..., n-1$.

The part (B) of this Theorem may be proved in the same way. $\square$

From Theorem 2.4 we get the following corollaries.

Corollary 2.6 (Schur [7], Marshall and Olkin [5], Th. 3.C.2). *Any symmetric convex function defined on a symmetric convex set $\mathcal{A}$ is S-convex.*

Corollary 2.7 (Marshall and Olkin [5], Th. 3.C.3). *If φ is a symmetric function defined on a symmetric convex set $\mathcal{A}$ in R^n satisfying the condition $\varphi(\lambda x + (1-\lambda)y) \leq \max\{\varphi(x), \varphi(y)\}$ for all $\lambda \in (0,1)$ and for all $x, y \in \mathcal{A}$ then φ is S-convex.*

3. Further results and applications

We shall start from some auxiliary results.

Lemma 3.1. *Any real measurable function f defined on an open interval I in R is convex, if and only if,*

$$f(\lambda x + (1-\lambda)y) + f(\lambda y + (1-\lambda)x) \leq f(x) + f(y) \text{ for any } x, y \in I \text{ and } \lambda \in (0,1). \quad (3)$$

Moreover, f is strictly convex, if and only if, the inequality in (3) is strict.

Proof. For necessity we shall use the fact that any measurable convex function φ on an open interval $I \subseteq R$ is continuous (cf. [4]). Let us recall that, in this case, the conditions (1) and (2) are equivalent. Thus, if f is convex, then by (1), $f(\lambda x+(1-\lambda)y) \leq \lambda f(x)+(1-\lambda)f(y)$ and $f(\lambda y+(1-\lambda)x) \leq \lambda f(y)+(1-\lambda)f(x)$ for $x, y \in I$ and $\lambda \in (0,1)$. This implies (3). Conversely, by setting in (3) $\lambda = \frac{1}{2}$ we get (2). □

Remark 3.2. By Bernstein-Doetsch Theorem (see [1]) the word "measurable" in Lemma 3.1 and throughout this section may be replaced by "locally bounded".

The following lemma is a direct consequence of Theorem 2.4 and Lemma 3.1.

Lemma 3.3 (Propositions 3.C.1 and 3.C.1.c in [5]). *For a given measurable function f on an open interval $I \subseteq R$ the following are equivalent:*

(*a*) *The function f is (strictly) convex on I.*
(*b*) *The function $\varphi(x) = \sum_{i=1}^n f(x_i)$ is (strictly) S-convex on I^n.*

Taking into account Remark 2.3, one can state the following theorem.

Theorem 3.4. *For arbitrary real functions f, g and h on an open interval $I \subseteq R$, such that g is positive and f is measurable, the following are equivalent:*

(*a*) *The function f is (strictly) convex,*
(*b*) *The function $\phi_n(x_1, ..., x_n) = g(\sum_{i=1}^n x_i) \sum_{i=1}^n f(x_i) + h(\sum_{i=1}^n x_i)$ is (strictly) S-convex.*

Now, for a given f, let us consider the function $\psi_n(x) = \prod_{i=1}^n f(x_i)$. To state a necessary and sufficient condition for the S-convexity of ψ_n we need some notion.

A positive function f on an interval $I \subseteq R$ is said to be *logconvex* if $\ln f$ is convex.

Remark 3.5. We note that f is logconvex, if and only if, $\log_a f$ is convex for some $a > 1$. Moreover, by Lemma 2.1, any logconvex function is convex. However, a convex function may not be logconvex. For instance, the function $f(x) = x^2$ is convex, while $\ln f(x) = 2 \ln x$ is concave.

The following theorem is an extension of Proposition 3.E.1 in [5].

Theorem 3.6. *For arbitrary real functions f, g and h on an open interval $I \subseteq R$, such that g is positive and f is measurable and positive, the following are equivalent:*

(a) *The function f is (strictly) logconvex,*

(b) *The function $\psi_n(x_1, ..., x_n) = g(\sum_{i=1}^n x_i) \prod_{i=1}^n f(x_i) + h(\sum_{i=1}^n x_i)$ is (strictly) S-convex.*

Proof. By Remark 2.3, we only need to consider the case $\psi_n(x) = \prod_{i=1}^n f(x_i)$. Since g is positive, the condition (ii) in Theorem 2.4 may be equivalently replaced by $\psi_2(\lambda x + (1-\lambda)y), \lambda y + (1-\lambda)x) \leq \psi_2(x) + \psi_2(y)$, and, by definition of ψ_n, the last one reduces to $\ln f(\lambda x + (1-\lambda)y) + \ln f(\lambda y + (1-\lambda)x) \leq \ln f(x) + \ln f(y)$. Now the desired result follows from Theorem 3.4. □

Let us present some applications of Theorems 3.4 and 3.6.

Example 3.7. In descriptive statistics the ratio $v = \frac{s}{\overline{x}}$, where $\overline{x} = \frac{1}{n}\sum x_i$ and $s = \sqrt{\frac{1}{n-1}\sum(x_i - \overline{x})^2}$, is said to be the variation coefficient in the sample $x = (x_1, ..., x_n)$, providing $\overline{x} \neq 0$. In general, the set of possible values of v is not bounded, but if $x_1, ..., x_n$ are nonnegative and $\sum x_i > 0$, then it is so. One can suspect that, under this additional condition, the coefficient v is normalized, in the sense that $0 \leq v \leq 1$. We shall show that this conjecture is false.

Given $\overline{x}$ the function $s^2(x)$ may be presented in the form $s^2(x) = \frac{1}{n-1}(\varphi_n(x) - n\overline{x}^2)$, where $\varphi_n(x) = \sum x_i^2$. Thus, by Theorem 3.4, $s^2(x)$ is S-convex. Therefore, its maximal value is attained for the maximal element in the set

$$\mathcal{A} = \left\{(x_1, ..., x_n) : \ x_i \geq 0, \ \sum x_i = n\overline{x}\right\}$$

with respect to the ordering $\prec$. It is well known that this element is $y = (0, ..., 0, n\overline{x})$, and hence $\max_{x \in \mathcal{A}} s^2(x) = s^2(y) = \frac{1}{n-1}(\sum y_i^2 - n\overline{x}^2) = n\ \overline{x}^2$. Therefore, the function v is not normalized but $\frac{v}{\sqrt{n}}$ is so.

Example 3.8. Any quadratic form of $x = (x_1, ..., x_n)$ may be expressed as $\phi(x) = xAx^T$, where A is a symmetric matrix of $n \times n$. It is evident that such a function is symmetric, if and only if, $A = aI + b\mathbf{1}$, $a, b \in R$, where $\mathbf{1}$ is $n \times n$ matrix of ones. Thus, in this case, $xAx^T = a\sum x_i^2 + b(\sum x_i)^2$. Hence, by Theorem 3.4, a quadratic form $\phi(x) = xAx^T$ is S-convex, if and only if, $A = aI + b\mathbf{1}$ with $a \geq 0$, and strictly S-convex, if and only if, $a > 0$.

Example 3.9. Let us consider the function $\psi(x) = g(\sum x_i)e^{xAx^T} + h(\sum x_i)$, where $x = (x_1, ..., x_n)$, while g and h satisfy the assumptions of Theorem 3.6. We are seeking for a necessary and sufficient condition to ψ be S-convex. It follows from Example 3.8, via Remark 2.3, that a necessary one is $A = aI + b\mathbf{1}$ with $a \geq 0$. To show its sufficiency, we only need to set in Theorem 3.6 $f(x) = e^{ax}$. Similarly, ψ is strictly S-convex, if and only if, $a > 0$.

Some interesting applications of Schur-convex functions in statistics can be found in [8] and [9].

Acknowledgements. The author would like to thank a referee for his careful reading and helpful comments on the earlier version of the paper.

References

[1] F. G. Bernstein, G. Doetsch: Zur Theorie der konvexen Funktionen, Math. Ann. 76 (1915) 514–526.

[2] G. H. Hardy, J. E. Littlewood, G. Pólya: Some simple inequalities satisfied by convex functions, Messenger Math. 58 (1929) 145–152.

[3] G. H. Hardy, J. E. Littlewood, G. Pólya: Inequalities, 2nd Ed., Cambridge University Press, Cambridge (1988).

[4] L. D. Kudrjavtsev: Convex function, in: Mathematical Encyclopedia, Vol. 1, I. M. Vinogradov, V. I. Bityutskov, Yu. V. Prokhorov (eds.), Soviet Encyclopedia, Moscow (1977) 794–795 (in Russian).

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

[6] A. M. Ostrowski: Sur quelques applications des fonctions convexes an sens de I. Schur, J. Math. Pures Appl. 31 (1952) 253–282.

[7] I. Schur: Über eine Klasse von Mittelbildungen mit Anwendungen auf die Determinantentheorie, Sitzungsber. Berl. Math. Ges. 22 (1923) 9–20, also in: Issai Schur Collected Works, Vol. II, A. Brauer, A. Rohrbach (eds.), Springer, Berlin (1973) 418–427.

[8] C. Stępniak: Stochastic ordering and Schur-convex functions in comparison of linear experiments, Metrika 36 (1989) 291–298.

[9] C. Stępniak: A note on comparison of genetic experiments, Ann. Stat. 22 (1994) 1630–1632.