Some Slater's Type Inequalities for Convex Functions Defined on Linear Spaces and Applications

SOME SLATER'S TYPE INEQUALITIES FOR CONVEX FUNCTIONS DEFINED ON LINEAR SPACES AND APPLICATIONS

S.S. DRAGOMIR

Abstract. Some inequalities of the Slater type for convex functions defined on general linear spaces are given. Applications for norm inequalities and f-divergence measures are also provided.

1. Introduction

Suppose that I is an interval of real numbers with interior $\mathring{I}$ and $f : I \to \mathbb{R}$ is a convex function on I. Then f is continuous on $\mathring{I}$ and has finite left and right derivatives at each point of $\mathring{I}$. Moreover, if $x, y \in \mathring{I}$ and $x < y$, then $f'_{-}(x) \leq f'_{+}(x) \leq f'_{-}(y) \leq f'_{+}(y)$ which shows that both f'_{-} and f'_{+} are nondecreasing function on $\mathring{I}$. It is also known that a convex function must be differentiable except for at most countably many points.

For a convex function $f : I \to \mathbb{R}$, the subdifferential of f denoted by ∂f is the set of all functions $\varphi : I \to [-\infty, \infty]$ such that $\varphi\left(\mathring{I}\right) \subset \mathbb{R}$ and

$$f(x) \geq f(a) + (x - a)\varphi(a) \text{ for any } x, a \in I.$$

It is also well known that if f is convex on I, then ∂f is nonempty, f'_{-}, $f'_{+} \in \partial f$ and if $\varphi \in \partial f$, then

$$f'_{-}(x) \leq \varphi(x) \leq f'_{+}(x) \text{ for any } x \in \mathring{I}.$$

In particular, φ is a nondecreasing function.

If f is differentiable and convex on $\mathring{I}$, then $\partial f = \{f'\}$.

The following result is well known in the literature as *the Slater inequality:*

Theorem 1 (Slater, 1981, [5]). *If $f : I \to \mathbb{R}$ is a nonincreasing (nondecreasing) convex function, $x_i \in I, p_i \geq 0$ with $P_n := \sum_{i=1}^n p_i > 0$ and $\sum_{i=1}^n p_i\varphi(x_i) \neq 0$, where $\varphi \in \partial f$, then*

$$\frac{1}{P_n}\sum_{i=1}^n p_i f(x_i) \leq f\left(\frac{\sum_{i=1}^n p_i x_i \varphi(x_i)}{\sum_{i=1}^n p_i \varphi(x_i)}\right). \tag{1.1}$$

As pointed out in [4, p. 208], the monotonicity assumption for the derivative φ can be replaced with the condition

$$\frac{\sum_{i=1}^n p_i x_i \varphi(x_i)}{\sum_{i=1}^n p_i \varphi(x_i)} \in I, \tag{1.2}$$

Date: February 14, 2009.

1991 *Mathematics Subject Classification.* 26D15; 94.

Key words and phrases. Slater's inequality, Convex functions, Norm inequalities, Semi-inner products, f-divergence measures.

which is more general and can hold for suitable points in I and for not necessarily monotonic functions.

The main aim of the present paper is to extend Slater's inequality for convex functions defined on general linear spaces. A reverse of the Slater's inequality is also obtained. Natural applications for norm inequalities and f-divergence measures are provided as well.

2. Slater's Inequality for Functions Defined on Linear Spaces

Assume that $f : X \to \mathbb{R}$ is a *convex function* on the real linear space X. Since for any vectors $x, y \in X$ the function $g_{x,y} : \mathbb{R} \to \mathbb{R}$, $g_{x,y}(t) := f(x + ty)$ is convex it follows that the following limits exist

$$\nabla_{+(-)} f(x)(y) := \lim_{t \to 0+(-)} \frac{f(x+ty) - f(x)}{t}$$

and they are called the *right(left) Gâteaux derivatives* of the function f in the point x over the direction y.

It is obvious that for any $t > 0 > s$ we have

$$\frac{f(x+ty) - f(x)}{t} \geq \nabla_+ f(x)(y) = \inf_{t>0} \left[\frac{f(x+ty) - f(x)}{t} \right] \geq \sup_{s<0} \left[\frac{f(x+sy) - f(x)}{s} \right] = \nabla_- f(x)(y) \geq \frac{f(x+sy) - f(x)}{s} \tag{2.1}$$

for any $x, y \in X$ and, in particular,

$$\nabla_- f(u)(u - v) \geq f(u) - f(v) \geq \nabla_+ f(v)(u - v) \tag{2.2}$$

for any $u, v \in X$. We call this *the gradient inequality* for the convex function f. It will be used frequently in the sequel in order to obtain various results related toSlater's inequality.

The following properties are also of importance:

$$\nabla_+ f(x)(-y) = -\nabla_- f(x)(y), \tag{2.3}$$

and

$$\nabla_{+(-)} f(x)(\alpha y) = \alpha \nabla_{+(-)} f(x)(y) \tag{2.4}$$

for any $x, y \in X$ and $\alpha \geq 0$.

The right Gâteaux derivative is *subadditive* while the left one is *superadditive*, i.e.,

$$\nabla_+ f(x)(y + z) \leq \nabla_+ f(x)(y) + \nabla_+ f(x)(z) \tag{2.5}$$

and

$$\nabla_- f(x)(y + z) \geq \nabla_- f(x)(y) + \nabla_- f(x)(z) \tag{2.6}$$

for any $x, y, z \in X$.

Some natural examples can be provided by the use of normed spaces.

Assume that $(X, \|\cdot\|)$ is a real normed linear space. The function $f : X \to \mathbb{R}$, $f(x) := \frac{1}{2}\|x\|^2$ is a convex function which generates *the superior* and *the inferior semi-inner products*

$$\langle y, x \rangle_{s(i)} := \lim_{t \to 0+(-)} \frac{\|x+ty\|^2 - \|x\|^2}{t}.$$

For a comprehensive study of the properties of these mappings in the Geometry of Banach Spaces see the monograph [3].

For the convex function $f_p : X \to \mathbb{R}$, $f_p(x) := \|x\|^p$ with $p > 1$, we have

$$\nabla_{+(-)} f_p(x)(y) = \begin{cases} p\|x\|^{p-2}\langle y, x\rangle_{s(i)} & \text{if } x \neq 0, \\ \\ 0 & \text{if } x = 0 \end{cases}$$

for any $y \in X$.

If $p = 1$, then we have

$$\nabla_{+(-)} f_1(x)(y) = \begin{cases} \|x\|^{-1}\langle y, x\rangle_{s(i)} & \text{if } x \neq 0, \\ \\ +(-)\|y\| & \text{if } x = 0 \end{cases}$$

for any $y \in X$.

For a given convex function $f : X \to \mathbb{R}$ and a given n-tuple of vectors $\mathbf{x} = (x_1, ..., x_n) \in X^n$ we consider the sets

$$Sla_{+(-)}(f, \mathbf{x}) := \left\{ v \in X \mid \nabla_{+(-)} f(x_i)(v - x_i) \geq 0 \text{ for all } i \in \{1, ..., n\} \right\} \tag{2.7}$$

and

$$Sla_{+(-)}(f, \mathbf{x}, \mathbf{p}) := \left\{ v \in X \mid \sum_{i=1}^{n} p_i \nabla_{+(-)} f(x_i)(v - x_i) \geq 0 \right\} \tag{2.8}$$

where $\mathbf{p} = (p_1, ..., p_n) \in \mathbb{P}^n$ is a given probability distribution.

Since $\nabla_{+(-)} f(x)(0) = 0$ for any $x \in X$, then we observe that $\{x_1, ..., x_n\} \subset Sla_{+(-)}(f, \mathbf{x}, \mathbf{p})$, therefore the sets $Sla_{+(-)}(f, \mathbf{x}, \mathbf{p})$ are not empty for each $f, \mathbf{x}$ and $\mathbf{p}$ as above.

The following properties of these sets hold:

Lemma 1. *For a given convex function $f : X \to \mathbb{R}$, a given n-tuple of vectors $\mathbf{x} = (x_1, ..., x_n) \in X^n$ and a given probability distribution $\mathbf{p} = (p_1, ..., p_n) \in \mathbb{P}^n$ we have*

(i) $Sla_-(f, \mathbf{x}) \subset Sla_+(f, \mathbf{x})$ *and* $Sla_-(f, \mathbf{x}, \mathbf{p}) \subset Sla_+(f, \mathbf{x}, \mathbf{p})$;

(ii) $Sla_-(f, \mathbf{x}) \subset Sla_-(f, \mathbf{x}, \mathbf{p})$ *and* $Sla_+(f, \mathbf{x}) \subset Sla_+(f, \mathbf{x}, \mathbf{p})$ *for all* $\mathbf{p} = (p_1, ..., p_n) \in \mathbb{P}^n$;

(iii) The sets $Sla_-(f, \mathbf{x})$ *and* $Sla_-(f, \mathbf{x}, \mathbf{p})$ *are convex.*

Proof. The properties (i) and (ii) follow from the definition and the fact that $\nabla_+ f(x)(y) \geq \nabla_- f(x)(y)$ for any x, y.

(iii) Let us only prove that $Sla_-(f, \mathbf{x})$ is convex.

If we assume that $y_1, y_2 \in Sla_-(f, \mathbf{x})$ and $\alpha, \beta \in [0, 1]$ with $\alpha + \beta = 1$, then by the superadditivity and positive homogeneity of the Gâteaux derivative $\nabla_- f(\cdot)(\cdot)$ in the second variable we have

$$\begin{aligned} \nabla_- f(x_i)(\alpha y_1 + \beta y_2 - x_i) &= \nabla_- f(x_i)[\alpha(y_1 - x_i) + \beta(y_2 - x_i)] \\ &\geq \alpha \nabla_- f(x_i)(y_1 - x_i) + \beta \nabla_- f(x_i)(y_2 - x_i) \geq 0 \end{aligned}$$

for all $i \in \{1, ..., n\}$, which shows that $\alpha y_1 + \beta y_2 \in Sla_-(f, \mathbf{x})$.

The proof for the convexity of $Sla_-(f, \mathbf{x}, \mathbf{p})$ is similar and the details are omitted. $\square$

For the convex function $f_p : X \to \mathbb{R}$, $f_p(x) := \|x\|^p$ with $p \geq 1$, defined on the normed linear space $(X, \|\cdot\|)$ and for the n-tuple of vectors $\mathbf{x} = (x_1, ..., x_n) \in X^n \setminus \{(0, ..., 0)\}$ we have, by the well known property of the semi inner products

$$\langle y + \alpha x, x \rangle_{s(i)} = \langle y, x \rangle_{s(i)} + \alpha \|x\|^2 \text{ for any } x, y \in X \text{ and } \alpha \in \mathbb{R},$$

that

$$\begin{aligned} Sla_{+(-)}(\|\cdot\|^p, \mathbf{x}) &= Sla_{+(-)}(\|\cdot\|, \mathbf{x}) \\ &:= \left\{ v \in X \mid \langle v, x_j \rangle_{s(i)} \geq \|x_j\|^2 \text{ for all } j \in \{1, ..., n\} \right\} \end{aligned}$$

which, as can be seen, does not depend of p. We observe that, by the continuity of the semi-inner products in the first variable that $Sla_{+(-)}(\|\cdot\|, \mathbf{x})$ is closed in $(X, \|\cdot\|)$. Also, we should remarks that if $v \in Sla_{+(-)}(\|\cdot\|, \mathbf{x})$ then for any $\gamma \geq 1$ we also have that $\gamma v \in Sla_{+(-)}(\|\cdot\|, \mathbf{x})$.

The larger classes, which are dependent on the probability distribution $\mathbf{p} \in \mathbb{P}^n$ are described by

$$Sla_{+(-)}(\|\cdot\|^p, \mathbf{x}, \mathbf{p}) := \left\{ v \in X \mid \sum_{j=1}^n p_j \|x_j\|^{p-2} \langle v, x_j \rangle_{s(i)} \geq \sum_{j=1}^n p_j \|x_j\|^p \right\}.$$

If the normed space is smooth, i.e., the norm is Gâteaux differentiable in any nonzero point, then the superior and inferior semi-inner products coincide with the Lumer-Giles semi-inner product $[\cdot, \cdot]$ that generates the norm and is linear in the first variable (see for instance [3]). In this situation

$$Sla(\|\cdot\|, \mathbf{x}) = \left\{ v \in X \mid [v, x_j] \geq \|x_j\|^2 \text{ for all } j \in \{1, ..., n\} \right\}$$

and

$$Sla(\|\cdot\|^p, \mathbf{x}, \mathbf{p}) = \left\{ v \in X \mid \sum_{j=1}^n p_j \|x_j\|^{p-2} [v, x_j] \geq \sum_{j=1}^n p_j \|x_j\|^p \right\}.$$

If $(X, \langle \cdot, \cdot \rangle)$ is an inner product space then $Sla(\|\cdot\|^p, \mathbf{x}, \mathbf{p})$ can be described by

$$Sla(\|\cdot\|^p, \mathbf{x}, \mathbf{p}) = \left\{ v \in X \mid \left\langle v, \sum_{j=1}^n p_j \|x_j\|^{p-2} x_j \right\rangle \geq \sum_{j=1}^n p_j \|x_j\|^p \right\}$$

and if the family $\{x_j\}_{j=1,...,n}$ is orthogonal, then obviously, by the Pythagoras theorem, we have that the sum $\sum_{j=1}^n x_j$ belongs to $Sla(\|\cdot\|, \mathbf{x})$ and therefore to $Sla(\|\cdot\|^p, \mathbf{x}, \mathbf{p})$ for any $p \geq 1$ and any probability distribution $\mathbf{p} = (p_1, ..., p_n) \in \mathbb{P}^n$.

We can state now the following results that provides a generalization of Slater's inequality as well as a counterpart for it.

Theorem 2. *Let $f : X \to \mathbb{R}$ be a convex function on the real linear space X, $\mathbf{x} = (x_1, ..., x_n) \in X^n$ an n-tuple of vectors and $\mathbf{p} = (p_1, ..., p_n) \in \mathbb{P}^n$ a probability distribution. Then for any $v \in Sla_+(f, \mathbf{x}, \mathbf{p})$ we have the inequalities*

$$\nabla_- f(v)(v) - \sum_{i=1}^n p_i \nabla_- f(v)(x_i) \geq f(v) - \sum_{i=1}^n p_i f(x_i) \geq 0. \tag{2.9}$$

Proof. If we write the gradient inequality for $v \in Sla_+(f, \mathbf{x}, \mathbf{p})$ and x_i, then we have that

$$\nabla_- f(v)(v - x_i) \geq f(v) - f(x_i) \geq \nabla_+ f(x_i)(v - x_i) \tag{2.10}$$

for any $i \in \{1, ..., n\}$.

By multiplying (2.10) with $p_i \geq 0$ and summing over i from 1 to n we get

$$\sum_{i=1}^{n} p_i \nabla_- f(v)(v - x_i) \geq f(v) - \sum_{i=1}^{n} p_i f(x_i) \geq \sum_{i=1}^{n} p_i \nabla_+ f(x_i)(v - x_i). \tag{2.11}$$

Now, since $v \in Sla_+(f, \mathbf{x}, \mathbf{p})$, then the right hand side of (2.11) is nonnegative, which proves the second inequality in (2.9).

By the superadditivity of the Gâteaux derivative $\nabla_- f(\cdot)(\cdot)$ in the second variable we have

$$\nabla_- f(v)(v) - \nabla_- f(v)(x_i) \geq \nabla_- f(v)(v - x_i),$$

which, by multiplying with $p_i \geq 0$ and summing over i from 1 to n, produces the inequality

$$\nabla_- f(v)(v) - \sum_{i=1}^{n} p_i \nabla_- f(v)(x_i) \geq \sum_{i=1}^{n} p_i \nabla_- f(v)(v - x_i). \tag{2.12}$$

Utilising (2.11) and (2.12) we deduce the desired result (2.9). □

Remark 1. *The above result has the following form for normed linear spaces. Let $(X, \|\cdot\|)$ be a normed linear space, $\mathbf{x} = (x_1, ..., x_n) \in X^n$ an n-tuple of vectors from X and $\mathbf{p} = (p_1, ..., p_n) \in \mathbb{P}^n$ a probability distribution. Then for any vector $v \in X$ with the property*

$$\sum_{j=1}^{n} p_j \|x_j\|^{p-2} \langle v, x_j \rangle_s \geq \sum_{j=1}^{n} p_j \|x_j\|^p, \ p \geq 1, \tag{2.13}$$

we have the inequalities

$$p \left[\|v\|^p - \sum_{j=1}^{n} p_j \|x_j\|^{p-2} \langle v, x_j \rangle_i \right] \geq \|v\|^p - \sum_{j=1}^{n} p_j \|x_j\|^p \geq 0. \tag{2.14}$$

Rearranging the first inequality in (2.14) we also have that

$$(p-1)\|v\|^p + \sum_{j=1}^{n} p_j \|x_j\|^p \geq p \sum_{j=1}^{n} p_j \|x_j\|^{p-2} \langle v, x_j \rangle_i. \tag{2.15}$$

If the space is smooth, then the condition (2.13) becomes

$$\sum_{j=1}^{n} p_j \|x_j\|^{p-2} [v, x_j] \geq \sum_{j=1}^{n} p_j \|x_j\|^p, \ p \geq 1, \tag{2.16}$$

implying the inequality

$$p \left[\|v\|^p - \sum_{j=1}^{n} p_j \|x_j\|^{p-2} [v, x_j] \right] \geq \|v\|^p - \sum_{j=1}^{n} p_j \|x_j\|^p \geq 0. \tag{2.17}$$

Notice also that the first inequality in (2.17) is equivalent with

$$(p-1)\|v\|^{p}+\sum_{j=1}^{n}p_{j}\|x_{j}\|^{p}\geq p\sum_{j=1}^{n}p_{j}\|x_{j}\|^{p-2}[v,x_{j}] \left(\geq p\sum_{j=1}^{n}p_{j}\|x_{j}\|^{p}\geq 0\right). \tag{2.18}$$

The following corollary is of interest:

Corollary 1. *Let $f:X\to\mathbb{R}$ be a convex function on the real linear space X, $\mathbf{x}=(x_1,...,x_n)\in X^n$ an n-tuple of vectors and $\mathbf{p}=(p_1,...,p_n)\in\mathbb{P}^n$ a probability distribution. If*

$$\sum_{i=1}^{n}p_{i}\nabla_{+}f(x_{i})(x_{i})\geq(<)\,0 \tag{2.19}$$

and there exists a vector $s\in X$ with

$$\sum_{i=1}^{n}p_{i}\nabla_{+(-)}f(x_{i})(s)\geq(\leq)\,1 \tag{2.20}$$

then

$$\begin{aligned}\nabla_{-}f\left(\sum_{j=1}^{n}p_{j}\nabla_{+}f(x_{j})(x_{j})\,s\right)&\left(\sum_{j=1}^{n}p_{j}\nabla_{+}f(x_{j})(x_{j})\,s\right)\\ &-\sum_{i=1}^{n}p_{i}\nabla_{-}f\left(\sum_{j=1}^{n}p_{j}\nabla_{+}f(x_{j})(x_{j})\,s\right)(x_{i})\\ &\geq f\left(\sum_{j=1}^{n}p_{j}\nabla_{+}f(x_{j})(x_{j})\,s\right)-\sum_{i=1}^{n}p_{i}f(x_{i})\geq 0.\end{aligned} \tag{2.21}$$

Proof. Assume that $\sum_{i=1}^{n}p_{i}\nabla_{+}f(x_{i})(x_{i})\geq 0$ and $\sum_{i=1}^{n}p_{i}\nabla_{+}f(x_{i})(s)\geq 1$ and define $v:=\sum_{j=1}^{n}p_{j}\nabla_{+}f(x_{j})(x_{j})\,s$. We claim that $v\in Sla_{+}(f,\mathbf{x},\mathbf{p})$.

By the subadditivity and positive homogeneity of the mapping $\nabla_{+}f(\cdot)(\cdot)$ in the second variable we have

$$\begin{aligned}
& \sum_{i=1}^{n} p_i \nabla_+ f(x_i)(v - x_i) \\
\geq\ & \sum_{i=1}^{n} p_i \nabla_+ f(x_i)(v) - \sum_{i=1}^{n} p_i \nabla_+ f(x_i)(x_i) \\
=\ & \sum_{i=1}^{n} p_i \nabla_+ f(x_i)\left(\sum_{j=1}^{n} p_j \nabla_+ f(x_j)(x_j)\, s\right) - \sum_{i=1}^{n} p_i \nabla_+ f(x_i)(x_i) \\
=\ & \sum_{j=1}^{n} p_j \nabla_+ f(x_j)(x_j) \sum_{i=1}^{n} p_i \nabla_+ f(x_i)(s) - \sum_{i=1}^{n} p_i \nabla_+ f(x_i)(x_i) \\
=\ & \sum_{j=1}^{n} p_j \nabla_+ f(x_j)(x_j)\left[\sum_{i=1}^{n} p_i \nabla_+ f(x_i)(s) - 1\right] \geq 0,
\end{aligned}$$

as claimed. Applying Theorem 2 for this v we get the desired result.

If $\sum_{i=1}^{n} p_i \nabla_+ f(x_i)(x_i) < 0$ and $\sum_{i=1}^{n} p_i \nabla_- f(x_i)(s) \leq 1$ then for

$$w := \sum_{j=1}^{n} p_j \nabla_+ f(x_j)(x_j)\, s$$

we also have that

$$\begin{aligned}
& \sum_{i=1}^{n} p_i \nabla_+ f(x_i)(w - x_i) \\
\geq\ & \sum_{i=1}^{n} p_i \nabla_+ f(x_i)\left(\sum_{j=1}^{n} p_j \nabla_+ f(x_j)(x_j)\, s\right) - \sum_{i=1}^{n} p_i \nabla_+ f(x_i)(x_i) \\
=\ & \sum_{i=1}^{n} p_i \nabla_+ f(x_i)\left(\left(-\sum_{j=1}^{n} p_j \nabla_+ f(x_j)(x_j)\right)(-s)\right) - \sum_{i=1}^{n} p_i \nabla_+ f(x_i)(x_i) \\
=\ & \left(-\sum_{j=1}^{n} p_j \nabla_+ f(x_j)(x_j)\right) \sum_{i=1}^{n} p_i \nabla_+ f(x_i)(-s) - \sum_{i=1}^{n} p_i \nabla_+ f(x_i)(x_i) \\
=\ & \left(-\sum_{j=1}^{n} p_j \nabla_+ f(x_j)(x_j)\right)\left(1 + \sum_{i=1}^{n} p_i \nabla_+ f(x_i)(-s)\right) \\
=\ & \left(-\sum_{j=1}^{n} p_j \nabla_+ f(x_j)(x_j)\right)\left(1 - \sum_{i=1}^{n} p_i \nabla_- f(x_i)(s)\right) \geq 0
\end{aligned}$$

where, for the last equality we have used the property (2.3). Therefore $w \in Sla_+(f, \mathbf{x}, \mathbf{p})$ and by Theorem 2 we get the desired result. □

It is natural to consider the case of normed spaces.

Remark 2. *Let $(X, \|\cdot\|)$ be a normed linear space, $\mathbf{x} = (x_1, ..., x_n) \in X^n$ an n-tuple of vectors from X and $\mathbf{p} = (p_1, ..., p_n) \in \mathbb{P}^n$ a probability distribution. Then for any* *vector $s \in X$ with the property that*

$$p\sum_{i=1}^{n}p_i\|x_i\|^{p-2}\langle s,x_i\rangle_s\geq 1, \tag{2.22}$$

we have the inequalities

$$\begin{aligned}p^p\|s\|^{p-1}\left(\sum_{j=1}^{n}p_j\|x_j\|^p\right)^{p-1}&\left(p\|s\|\sum_{j=1}^{n}p_j\|x_j\|^p-\sum_{j=1}^{n}p_j\langle x_j,s\rangle_i\right)\\&\geq p^p\|s\|^p\left(\sum_{j=1}^{n}p_j\|x_j\|^p\right)^p-\sum_{j=1}^{n}p_j\|x_j\|^p\geq 0.\end{aligned}$$

The case of smooth spaces can be easily derived from the above, however the details are left to the interested reader.

3. The Case of Finite Dimensional Linear Spaces

Consider now the finite dimensional linear space $X=\mathbb{R}^m$ and assume that C is an open convex subset of $\mathbb{R}^m$. Assume also that the function $f:C\to\mathbb{R}$ is differentiable and convex on C. Obviously, if $x=\left(x^1,...,x^m\right)\in C$ then for any $y=\left(y^1,...,y^m\right)\in\mathbb{R}^m$ we have

$$\nabla f(x)(y)=\sum_{k=1}^{m}\frac{\partial f\left(x^1,...,x^m\right)}{\partial x^k}\cdot y^k$$

For the convex function $f:C\to\mathbb{R}$ and a given n-tuple of vectors $\mathbf{x}=(x_1,...,x_n)\in C^n$ with $x_i=\left(x_i^1,...,x_i^m\right)$ with $i\in\{1,...,n\}$, we consider the sets

$$\begin{aligned}Sla(f,\mathbf{x},C):=\Bigg\{v\in C\mid &\sum_{k=1}^{m}\frac{\partial f\left(x_i^1,...,x_i^m\right)}{\partial x^k}\cdot v^k\\&\geq\sum_{k=1}^{m}\frac{\partial f\left(x_i^1,...,x_i^m\right)}{\partial x^k}\cdot x_i^k\text{ for all }i\in\{1,...,n\}\Bigg\}\end{aligned} \tag{3.1}$$

and

$$\begin{aligned}Sla(f,\mathbf{x},\mathbf{p},C):=\Bigg\{v\in C\mid &\sum_{i=1}^{n}\sum_{k=1}^{m}p_i\frac{\partial f\left(x_i^1,...,x_i^m\right)}{\partial x^k}\cdot v^k\\&\geq\sum_{i=1}^{n}\sum_{k=1}^{m}p_i\frac{\partial f\left(x_i^1,...,x_i^m\right)}{\partial x^k}\cdot x_i^k\Bigg\}\end{aligned} \tag{3.2}$$

where $\mathbf{p}=(p_1,...,p_n)\in\mathbb{P}^n$ is a given probability distribution.

As in the previous section the sets $Sla(f,\mathbf{x},C)$ and $Sla(f,\mathbf{x},\mathbf{p},C)$ are convex and closed subsets of $\mathrm{clo}(C)$, the closure of C. Also $\{x_1,...,x_n\}\subset Sla(f,\mathbf{x},C)\subset Sla(f,\mathbf{x},\mathbf{p},C)$ for any $\mathbf{p}=(p_1,...,p_n)\in\mathbb{P}^n$ a probability distribution.

Proposition 1. *Let $f:C\to\mathbb{R}$ be a convex function on the open convex set C in the finite dimensional linear space $\mathbb{R}^m$, $(x_1,...,x_n)\in C^n$ an n-tuple of vectors* *and* $(p_1, ..., p_n) \in \mathbb{P}^n$ *a probability distribution. Then for any* $v = \left(v^1, ..., v^n\right) \in Sla\left(f, \mathbf{x}, \mathbf{p}, C\right)$ *we have the inequalities*

$$\sum_{k=1}^{m} \frac{\partial f\left(v^1, ..., v^m\right)}{\partial x^k} \cdot v^k - \sum_{i=1}^{n} \sum_{k=1}^{m} p_i \ \frac{\partial f\left(x_i^1, ..., x_i^m\right)}{\partial x^k} \cdot v^k \geq f\left(v^1, ..., v^n\right) - \sum_{i=1}^{n} p_i f\left(x_i^1, ..., x_i^m\right) \geq 0. \tag{3.3}$$

The unidimensional case, i.e., $m = 1$ is of interest for applications. We will state this case with the general assumption that $f : I \to \mathbb{R}$ is a convex function on an *open* interval I. For a given n-tuple of vectors $\mathbf{x} = (x_1, ..., x_n) \in I^n$ we have

$$Sla_{+(-)}\left(f, \mathbf{x}, I\right) := \left\{v \in I \mid f'_{+(-)}\left(x_i\right) \cdot \left(v - x_i\right) \geq 0 \text{ for all } i \in \{1, ..., n\}\right\}$$

and

$$Sla_{+(-)}\left(f, \mathbf{x}, \mathbf{p}, \mathbf{I}\right) := \left\{v \in I | \ \sum_{i=1}^{n} p_i f'_{+(-)}\left(x_i\right) \cdot \left(v - x_i\right) \geq 0 \right\},$$

where $(p_1, ..., p_n) \in \mathbb{P}^n$ is a probability distribution. These sets inherit the general properties pointed out in Lemma 1. Moreover, if we make the assumption that $\sum_{i=1}^{n} p_i f'_{+}\left(x_i\right) \neq 0$ then for $\sum_{i=1}^{n} p_i f'_{+}\left(x_i\right) > 0$ we have

$$Sla_{+}\left(f, \mathbf{x}, \mathbf{p}, \mathbf{I}\right) = \left\{v \in I | \ v \geq \frac{\sum_{i=1}^{n} p_i f'_{+}\left(x_i\right) x_i}{\sum_{i=1}^{n} p_i f'_{+}\left(x_i\right)} \right\}$$

while for $\sum_{i=1}^{n} p_i f'_{+}\left(x_i\right) < 0$ we have

$$v = \left\{v \in I | \ v \leq \frac{\sum_{i=1}^{n} p_i f'_{+}\left(x_i\right) x_i}{\sum_{i=1}^{n} p_i f'_{+}\left(x_i\right)} \right\}.$$

Also, if we assume that $f'_{+}\left(x_i\right) \geq 0$ for all $i \in \{1, ..., n\}$ and $\sum_{i=1}^{n} p_i f'_{+}\left(x_i\right) > 0$ then

$$v_s := \frac{\sum_{i=1}^{n} p_i f'_{+}\left(x_i\right) x_i}{\sum_{i=1}^{n} p_i f'_{+}\left(x_i\right)} \in I$$

due to the fact that $x_i \in I$ and I is a convex set.

Proposition 2. *Let* $f : I \to \mathbb{R}$ *be a convex function on an open interval* I*. For a given* n*-tuple of vectors* $\mathbf{x} = (x_1, ..., x_n) \in I^n$ *and* $(p_1, ..., p_n) \in \mathbb{P}^n$ *a probability distribution we have*

$$f'_{-}\left(v\right) \left(v - \sum_{i=1}^{n} p_i x_i\right) \geq f\left(v\right) - \sum_{i=1}^{n} p_i f\left(x_i\right) \geq 0 \tag{3.4}$$

for any $v \in Sla_{+}\left(f, \mathbf{x}, \mathbf{p}, \mathbf{I}\right)$.

In particular, if we assume that $\sum_{i=1}^{n} p_i f'_{+}\left(x_i\right) \neq 0$ *and*

$$\frac{\sum_{i=1}^{n} p_i f'_{+}\left(x_i\right) x_i}{\sum_{i=1}^{n} p_i f'_{+}\left(x_i\right)} \in I$$

then

$$f'_{-}\left(\frac{\sum_{i=1}^{n}p_{i}f'_{+}(x_{i})x_{i}}{\sum_{i=1}^{n}p_{i}f'_{+}(x_{i})}\right)\left[\frac{\sum_{i=1}^{n}p_{i}f'_{+}(x_{i})x_{i}}{\sum_{i=1}^{n}p_{i}f'_{+}(x_{i})}-\sum_{i=1}^{n}p_{i}x_{i}\right] \geq f\left(\frac{\sum_{i=1}^{n}p_{i}f'_{+}(x_{i})x_{i}}{\sum_{i=1}^{n}p_{i}f'_{+}(x_{i})}\right)-\sum_{i=1}^{n}p_{i}f(x_{i})\geq 0 \tag{3.5}$$

Moreover, if $f'_{+}(x_i)\geq 0$ *for all* $i\in\{1,...,n\}$ *and* $\sum_{i=1}^{n}p_{i}f'_{+}(x_{i})>0$ *then (3.5) holds true as well.*

Remark 3. *We remark that the first inequality in (3.5) provides a reverse inequality for the classical result due to Slater.*

4. Some Applications for f-divergences

Given a convex function $f:[0,\infty)\to\mathbb{R}$, the f-*divergence functional*

$$I_{f}(\mathbf{p},\mathbf{q}):=\sum_{i=1}^{n}q_{i}f\left(\frac{p_{i}}{q_{i}}\right), \tag{4.1}$$

where $\mathbf{p}=(p_1,\dots,p_n)$, $\mathbf{q}=(q_1,\dots,q_n)$ are positive sequences, was introduced by Csiszár in [1], as a generalized measure of information, a "distance function" on the set of probability distributions $\mathbb{P}^n$. As in [1], we interpret undefined expressions by

$$f(0)=\lim_{t\to 0+}f(t),\qquad 0f\left(\frac{0}{0}\right)=0,$$
$$0f\left(\frac{a}{0}\right)=\lim_{q\to 0+}qf\left(\frac{a}{q}\right)=a\lim_{t\to\infty}\frac{f(t)}{t},\quad a>0.$$

The following results were essentially given by Csiszár and Körner [2]:

(i) *If* f *is convex, then* $I_f(\mathbf{p},\mathbf{q})$ *is jointly convex in* $\mathbf{p}$ *and* $\mathbf{q}$;
(ii) *For every* $\mathbf{p},\mathbf{q}\in R_{+}^{n}$, *we have*

$$I_{f}(\mathbf{p},\mathbf{q})\geq\sum_{j=1}^{n}q_{j}f\left(\frac{\sum_{j=1}^{n}p_{j}}{\sum_{j=1}^{n}q_{j}}\right). \tag{4.2}$$

If f *is strictly convex, equality holds in (4.2) iff*

$$\frac{p_1}{q_1}=\frac{p_2}{q_2}=\cdots=\frac{p_n}{q_n}.$$

If f *is normalized, i.e.,* $f(1)=0$, *then for every* $\mathbf{p},\mathbf{q}\in R_{+}^{n}$ *with* $\sum_{i=1}^{n}p_i=\sum_{i=1}^{n}q_i$, *we have the inequality*

$$I_{f}(\mathbf{p},\mathbf{q})\geq 0. \tag{4.3}$$

In particular, if $\mathbf{p},\mathbf{q}\in\mathbb{P}^n$, then (4.3) holds. This is the well-known positivity property of the f-divergence.

It is obvious that the above definition of $I_f(\mathbf{p},\mathbf{q})$ can be extended to any function $f:[0,\infty)\to\mathbb{R}$ however the positivity condition will not generally hold for normalized functions and $\mathbf{p},\mathbf{q}\in R_{+}^{n}$ *with* $\sum_{i=1}^{n}p_i=\sum_{i=1}^{n}q_i$.

For a normalized convex function $f:[0,\infty)\to\mathbb{R}$ and two probability distributions $\mathbf{p},\mathbf{q}\in\mathbb{P}^n$ we define the set

$$Sla_{+}(f,\mathbf{p},\mathbf{q}) := \left\{ v\in[0,\infty) \middle| \ \sum_{i=1}^{n} q_i f'_{+}\left(\frac{p_i}{q_i}\right)\cdot\left(v-\frac{p_i}{q_i}\right)\geq 0 \right\}. \tag{4.4}$$

Now, observe that

$$\sum_{i=1}^{n} q_i f'_{+}\left(\frac{p_i}{q_i}\right)\cdot\left(v-\frac{p_i}{q_i}\right)\geq 0$$

is equivalent with

$$v\sum_{i=1}^{n} q_i f'_{+}\left(\frac{p_i}{q_i}\right)\geq\sum_{i=1}^{n} p_i f'_{+}\left(\frac{p_i}{q_i}\right). \tag{4.5}$$

If $\sum_{i=1}^{n} q_i f'_{+}\left(\frac{p_i}{q_i}\right)>0$, then (4.5) is equivalent with

$$v\geq\frac{\sum_{i=1}^{n} p_i f'_{+}\left(\frac{p_i}{q_i}\right)}{\sum_{i=1}^{n} q_i f'_{+}\left(\frac{p_i}{q_i}\right)}$$

therefore in this case

$$Sla_{+}(f,\mathbf{p},\mathbf{q}) = \begin{cases} [0,\infty) & \text{if } \sum_{i=1}^{n} p_i f'_{+}\left(\frac{p_i}{q_i}\right)<0 \\ \left[\frac{\sum_{i=1}^{n} p_i f'_{+}\left(\frac{p_i}{q_i}\right)}{\sum_{i=1}^{n} q_i f'_{+}\left(\frac{p_i}{q_i}\right)},\infty\right) & \text{if } \sum_{i=1}^{n} p_i f'_{+}\left(\frac{p_i}{q_i}\right)\geq 0. \end{cases} \tag{4.6}$$

If $\sum_{i=1}^{n} q_i f'_{+}\left(\frac{p_i}{q_i}\right)<0$, then (4.5) is equivalent with

$$v\leq\frac{\sum_{i=1}^{n} p_i f'_{+}\left(\frac{p_i}{q_i}\right)}{\sum_{i=1}^{n} q_i f'_{+}\left(\frac{p_i}{q_i}\right)}$$

therefore

$$Sla_{+}(f,\mathbf{p},\mathbf{q}) = \begin{cases} \left[0,\frac{\sum_{i=1}^{n} p_i f'_{+}\left(\frac{p_i}{q_i}\right)}{\sum_{i=1}^{n} q_i f'_{+}\left(\frac{p_i}{q_i}\right)}\right] & \text{if } \sum_{i=1}^{n} p_i f'_{+}\left(\frac{p_i}{q_i}\right)\leq 0 \\ \emptyset & \text{if } \sum_{i=1}^{n} p_i f'_{+}\left(\frac{p_i}{q_i}\right)>0. \end{cases} \tag{4.7}$$

Utilising the extended f-divergences notation, we can state the following result:

Theorem 3. *Let $f:[0,\infty)\to\mathbb{R}$ be a normalized convex function and $\mathbf{p},\mathbf{q}\in\mathbb{P}^n$ two probability distributions. If $v\in Sla_{+}(f,\mathbf{p},\mathbf{q})$ then we have*

$$f'_{-}(v)(v-1)\geq f(v)-I_f(\mathbf{p},\mathbf{q})\geq 0. \tag{4.8}$$

In particular, if we assume that $I_{f'_{+}}(\mathbf{p},\mathbf{q})\neq 0$ and

$$\frac{I_{f'_{+}(\cdot)(\cdot)}(\mathbf{p},\mathbf{q})}{I_{f'_{+}}(\mathbf{p},\mathbf{q})}\in[0,\infty)$$

then

$$f'_{-}\left(\frac{I_{f'_{+}(\cdot)(\cdot)}(\mathbf{p},\mathbf{q})}{I_{f'_{+}}(\mathbf{p},\mathbf{q})}\right)\left[\frac{I_{f'_{+}(\cdot)(\cdot)}(\mathbf{p},\mathbf{q})}{I_{f'_{+}}(\mathbf{p},\mathbf{q})}-1\right] \geq f\left(\frac{I_{f'_{+}(\cdot)(\cdot)}(\mathbf{p},\mathbf{q})}{I_{f'_{+}}(\mathbf{p},\mathbf{q})}\right)-I_f(\mathbf{p},\mathbf{q})\geq 0. \tag{4.9}$$

Moreover, if $f'_{+}\left(\frac{p_i}{q_i}\right)\geq 0$ *for all* $i\in\{1,...,n\}$ *and* $I_{f'_{+}}(\mathbf{p},\mathbf{q})>0$ *then (4.9) holds true as well.*

The proof follows immediately from Proposition 2 and the details are omitted.

The K. Pearson χ^2*-divergence* is obtained for the convex function $f(t)=(1-t)^2$, $t\in\mathbb{R}$ and given by

$$\chi^2(\mathbf{p},\mathbf{q}):=\sum_{j=1}^{n}q_j\left(\frac{p_j}{q_j}-1\right)^2=\sum_{j=1}^{n}\frac{(p_j-q_j)^2}{q_j}=\sum_{j=1}^{n}\frac{p_i^2}{q_i}-1. \tag{4.10}$$

The *Kullback-Leibler divergence* can be obtained for the convex function $f:(0,\infty)\to\mathbb{R}$, $f(t)=t\ln t$ and is defined by

$$KL(\mathbf{p},\mathbf{q}):=\sum_{j=1}^{n}q_j\cdot\frac{p_j}{q_j}\ln\left(\frac{p_j}{q_j}\right)=\sum_{j=1}^{n}p_j\ln\left(\frac{p_j}{q_j}\right). \tag{4.11}$$

If we consider the convex function $f:(0,\infty)\to\mathbb{R}$, $f(t)=-\ln t$, then we observe that

$$I_f(\mathbf{p},\mathbf{q}):=\sum_{i=1}^{n}q_i f\left(\frac{p_i}{q_i}\right)=-\sum_{i=1}^{n}q_i\ln\left(\frac{p_i}{q_i}\right)=\sum_{i=1}^{n}q_i\ln\left(\frac{q_i}{p_i}\right)=KL(\mathbf{q},\mathbf{p}). \tag{4.12}$$

For the function $f(t)=-\ln t$ we have obviously have that

$$\begin{aligned} Sla(-\ln,\mathbf{p},\mathbf{q}) &:= \left\{v\in[0,\infty) \mid -\sum_{i=1}^{n}q_i\left(\frac{p_i}{q_i}\right)^{-1}\cdot\left(v-\frac{p_i}{q_i}\right)\geq 0\right\} \\ &= \left\{v\in[0,\infty) \mid v\sum_{i=1}^{n}\frac{q_i^2}{p_i}-1\leq 0\right\} \\ &= \left[0,\frac{1}{\chi^2(\mathbf{q},\mathbf{p})+1}\right]. \end{aligned} \tag{4.13}$$

Utilising the first part of the Theorem 3 we can state the following

Proposition 3. *Let* $\mathbf{p},\mathbf{q}\in\mathbb{P}^n$ *two probability distributions. If* $v\in\left[0,\frac{1}{\chi^2(\mathbf{q},\mathbf{p})+1}\right]$ *then we have*

$$\frac{1-v}{v}\geq-\ln(v)-KL(\mathbf{q},\mathbf{p})\geq 0. \tag{4.14}$$

In particular, for $v=\frac{1}{\chi^2(\mathbf{q},\mathbf{p})+1}$ *we get*

$$\chi^2(\mathbf{q},\mathbf{p})\geq\ln\left[\chi^2(\mathbf{q},\mathbf{p})+1\right]-KL(\mathbf{q},\mathbf{p})\geq 0. \tag{4.15}$$

If we consider now the function $f:(0,\infty)\to\mathbb{R}$, $f(t)=t\ln t$, then $f'(t)=\ln t+1$ and

$$\begin{aligned}(4.16)\quad & Sla\left((\cdot)\ln(\cdot),\mathbf{p},\mathbf{q}\right)\\ &:=\left\{v\in[0,\infty)\mid\ \sum_{i=1}^{n}q_i\left(\ln\left(\frac{p_i}{q_i}\right)+1\right)\cdot\left(v-\frac{p_i}{q_i}\right)\geq 0\right\}\\ &=\left\{v\in[0,\infty)\mid\ v\sum_{i=1}^{n}q_i\left(\ln\left(\frac{p_i}{q_i}\right)+1\right)-\sum_{i=1}^{n}p_i\cdot\left(\ln\left(\frac{p_i}{q_i}\right)+1\right)\geq 0\right\}\\ &=\left\{v\in[0,\infty)\mid\ v\left(1-KL\left(\mathbf{q},\mathbf{p}\right)\right)\geq 1+KL\left(\mathbf{p},\mathbf{q}\right)\right\}.\end{aligned}$$

We observe that if $\mathbf{p},\mathbf{q}\in\mathbb{P}^n$ two probability distributions such that $0<KL(\mathbf{q},\mathbf{p})<1$, then

$$Sla\left((\cdot)\ln(\cdot),\mathbf{p},\mathbf{q}\right)=\left[\frac{1+KL(\mathbf{p},\mathbf{q})}{1-KL(\mathbf{q},\mathbf{p})},\infty\right).$$

If $KL(\mathbf{q},\mathbf{p})\geq 1$ then $Sla\left((\cdot)\ln(\cdot),\mathbf{p},\mathbf{q}\right)=\emptyset$.

By the use of Theorem 3 we can state now the following

Proposition 4. *Let $\mathbf{p},\mathbf{q}\in\mathbb{P}^n$ two probability distributions such that $0<KL(\mathbf{q},\mathbf{p})<1$. If $v\in\left[\frac{1+KL(\mathbf{p},\mathbf{q})}{1-KL(\mathbf{q},\mathbf{p})},\infty\right)$ then we have*

$$(4.17)\qquad (\ln v+1)(v-1)\geq v\ln(v)-KL(\mathbf{p},\mathbf{q})\geq 0.$$

In particular, for $v=\frac{1+KL(\mathbf{p},\mathbf{q})}{1-KL(\mathbf{q},\mathbf{p})}$ we get

$$\begin{aligned}(4.18)\quad & \left(\ln\left[\frac{1+KL(\mathbf{p},\mathbf{q})}{1-KL(\mathbf{q},\mathbf{p})}\right]+1\right)\left(\frac{1+KL(\mathbf{p},\mathbf{q})}{1-KL(\mathbf{q},\mathbf{p})}-1\right)\\ &\geq\frac{1+KL(\mathbf{p},\mathbf{q})}{1-KL(\mathbf{q},\mathbf{p})}\ln\left[\frac{1+KL(\mathbf{p},\mathbf{q})}{1-KL(\mathbf{q},\mathbf{p})}\right]-KL(\mathbf{p},\mathbf{q})\geq 0.\end{aligned}$$

Similar results can be obtained for other divergence measures of interest such as the *Jeffreys divergence, Hellinger discrimination,* etc...However the details are left to the interested reader.

References

[1] I. Csiszár, Information-type measures of differences of probability distributions and indirect observations, *Studia Sci. Math. Hung.*, **2** (1967), 299-318.

[2] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic Press, New York, 1981.

[3] S.S. Dragomir, *Semi-inner Products and Applications*, Nova Science Publishers Inc., NY, 2004.

[4] S.S. Dragomir, *Discrete Inequalities of the Cauchy-Bunyakovsky-Schwarz Type*, Nova Science Publishers, NY, 2004.

[5] M.S. Slater, A companion inequality to Jensen's inequality, *J. Approx. Theory*, **32**(1981), 160-166.

Research Group in Mathematical Inequalities & Applications, School of Engineering & Science, Victoria University, PO Box 14428, Melbourne City, MC 8001, Australia.

E-mail address: sever.dragomir@vu.edu.au

URL: http://www.staff.vu.edu.au/rgmia/dragomir/