Bull. Aust. Math. Soc. **83** (2011), 500–517
doi:10.1017/S0004972710002054

INEQUALITIES IN TERMS OF THE GÂTEAUX DERIVATIVES FOR CONVEX FUNCTIONS ON LINEAR SPACES WITH APPLICATIONS

S. S. DRAGOMIR

(Received 6 October 2010)

Abstract

Some inequalities in terms of the Gâteaux derivatives related to Jensen's inequality for convex functions defined on linear spaces are given. Applications for norms, mean f-deviations and f-divergence measures are provided as well.

2010 *Mathematics subject classification*: primary 26D15; secondary 94A17.

Keywords and phrases: convex functions, Gâteaux derivatives, Jensen's inequality, norms, mean f-deviations, f-divergence measures.

1. Introduction

The Jensen inequality for convex functions plays a crucial role in the theory of inequalities due to the fact that other inequalities, such as that the arithmetic–geometric mean inequality, Hölder and Minkowski inequalities and Ky Fan's inequality, can be obtained as particular cases of it.

Let C be a convex subset of the linear space X and f a convex function on C. If $\mathbf{p} = (p_1, \ldots, p_n) \in \mathbb{P}^n$, that is, $p_i \geq 0$ for all $i \in \{1, \ldots, n\}$ with $\sum_{i=1}^n p_i = 1$, is a probability sequence and $\mathbf{x} = (x_1, \ldots, x_n) \in C^n$, then

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

is well known in the literature as Jensen's inequality.

Recently the author obtained the following refinement of Jensen's inequality (see [9]):

$$\begin{aligned} f\left(\sum_{j=1}^{n} p_j x_j\right) &\le \min_{k\in\{1,\dots,n\}}\left[(1-p_k) f\left(\frac{\sum_{j=1}^{n} p_j x_j - p_k x_k}{1-p_k}\right) + p_k f(x_k)\right] \\ &\le \frac{1}{n}\left[\sum_{k=1}^{n}(1-p_k) f\left(\frac{\sum_{j=1}^{n} p_j x_j - p_k x_k}{1-p_k}\right) + \sum_{k=1}^{n} p_k f(x_k)\right] \\ &\le \max_{k\in\{1,\dots,n\}}\left[(1-p_k) f\left(\frac{\sum_{j=1}^{n} p_j x_j - p_k x_k}{1-p_k}\right) + p_k f(x_k)\right] \\ &\le \sum_{j=1}^{n} p_j f(x_j), \end{aligned} \tag{1.2}$$

where f, x_k and p_k are as above.

The above result provides a different approach than the earlier one due to Pečarić and the author, namely (see [14]):

$$\begin{aligned} f\left(\sum_{i=1}^{n} p_i x_i\right) &\le \sum_{i_1,\dots,i_{k+1}=1}^{n} p_{i_1}\cdots p_{i_{k+1}} f\left(\frac{x_{i_1}+\cdots+x_{i_{k+1}}}{k+1}\right) \\ &\le \sum_{i_1,\dots,i_k=1}^{n} p_{i_1}\cdots p_{i_k} f\left(\frac{x_{i_1}+\cdots+x_{i_k}}{k}\right) \\ &\le \cdots \le \sum_{i=1}^{n} p_i f(x_i), \end{aligned} \tag{1.3}$$

for $k \ge 1$ and $\mathbf{p}$, $\mathbf{x}$ as above.

If $q_1, \dots, q_k \ge 0$ with $\sum_{j=1}^{k} q_j = 1$, then the following refinement obtained in 1994 by the author [6] also holds:

$$\begin{aligned} f\left(\sum_{i=1}^{n} p_i x_i\right) &\le \sum_{i_1,\dots,i_k=1}^{n} p_{i_1}\cdots p_{i_k} f\left(\frac{x_{i_1}+\cdots+x_{i_k}}{k}\right) \\ &\le \sum_{i_1,\dots,i_k=1}^{n} p_{i_1}\cdots p_{i_k} f(q_1 x_{i_1}+\cdots+q_k x_{i_k}) \\ &\le \sum_{i=1}^{n} p_i f(x_i), \end{aligned} \tag{1.4}$$

where $1 \le k \le n$ and $\mathbf{p}$, $\mathbf{x}$ are as above.

For other refinements and applications related to Ky Fan's inequality, the arithmetic–geometric mean inequality, the generalized triangle inequality, the f-divergence measures and so on, see [3–9, 13].

https://doi.org/10.1017/S0004972710002054 Published online by Cambridge University Press

In this paper, motivated by the above results, some new inequalities in terms of the Gâteaux derivatives related to Jensen's inequality for convex functions defined on linear spaces are given. Applications for norms, mean f-deviations and f-divergence measures are provided as well.

2. The Gâteau derivatives of convex functions

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 limits

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

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

It is obvious that, for any $t > 0 > s$,

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

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 following in order to obtain various results related to Jensen'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*, that is,

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

and

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

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

https://doi.org/10.1017/S0004972710002054 Published online by Cambridge University Press

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 [8].

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

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

for any $y \in X$. If $p = 1$, then

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

for any $y \in X$. This class of functions will be used to illustrate the inequalities obtained in the general case of convex functions defined on an entire linear space.

The following result holds.

THEOREM 2.1. *Let $f : X \to \mathbb{R}$ be a convex function. Then, for any $x, y \in X$ and $t \in [0, 1]$,*

$$\begin{aligned} t(1-t)[\nabla_- f(y)(y-x) - \nabla_+ f(x)(y-x)] &\\ \geq tf(x) + (1-t)f(y) - f(tx+(1-t)y) &\\ \geq t(1-t)[\nabla_+ f(tx+(1-t)y)(y-x) &\\ - \nabla_- f(tx+(1-t)y)(y-x)] \geq 0. & \end{aligned} \tag{2.7}$$

PROOF. Utilizing the gradient inequality (2.2), we have

$$f(tx+(1-t)y) - f(x) \geq (1-t)\nabla_+ f(x)(y-x) \tag{2.8}$$

and

$$f(tx+(1-t)y) - f(y) \geq -t\nabla_- f(y)(y-x). \tag{2.9}$$

If we multiply (2.8) by t and (2.9) by $1-t$ and add the resultant inequalities, we obtain

$$\begin{aligned} f(tx+(1-t)y) - tf(x) - (1-t)f(y) &\\ \geq (1-t)t\nabla_+ f(x)(y-x) - t(1-t)\nabla_- f(y)(y-x), & \end{aligned}$$

which is clearly equivalent to the first part of (2.7).

By the gradient inequality we also have

$$(1-t)\nabla_- f(tx+(1-t)y)(y-x) \geq f(tx+(1-t)y) - f(x)$$

and

$$-t\nabla_+ f(tx+(1-t)y)(y-x) \geq f(tx+(1-t)y) - f(y),$$

which by the same procedure as above yields the second part of (2.7). □

https://doi.org/10.1017/S0004972710002054 Published online by Cambridge University Press

The following particular case for norms may be stated.

COROLLARY 2.2. *If x and y are two vectors in the normed linear space $(X, \|\cdot\|)$ such that*

$$0 \notin [x, y] := \{(1-s)x + sy,\ s \in [0, 1]\},$$

then for any $p \geq 1$ we have the inequalities

$$\begin{aligned} &pt(1-t)[\|y\|^{p-2}\langle y-x, y\rangle_i - \|x\|^{p-2}\langle y-x, x\rangle_s] \\ &\quad \geq t\|x\|^p + (1-t)\|y\|^p - \|tx + (1-t)y\|^p \\ &\quad \geq pt(1-t)\|tx+(1-t)y\|^{p-2} \\ &\qquad \times [\langle y-x, tx+(1-t)y\rangle_s - \langle y-x, tx+(1-t)y\rangle_i] \geq 0 \end{aligned} \tag{2.10}$$

for any $t \in [0, 1]$. If $p \geq 2$ the inequality holds for any x and y.

REMARK 2.3. For $p = 1$ in (2.10) we derive the result

$$\begin{aligned} &t(1-t)\bigg[\bigg\langle y-x, \frac{y}{\|y\|}\bigg\rangle_i - \bigg\langle y-x, \frac{x}{\|x\|}\bigg\rangle_s\bigg] \\ &\quad \geq t\|x\| + (1-t)\|y\| - \|tx+(1-t)y\| \\ &\quad \geq t(1-t)\bigg[\bigg\langle y-x, \frac{tx+(1-t)y}{\|tx+(1-t)y\|}\bigg\rangle_s - \bigg\langle y-x, \frac{tx+(1-t)y}{\|tx+(1-t)y\|}\bigg\rangle_i\bigg] \\ &\quad \geq 0, \end{aligned} \tag{2.11}$$

while for $p = 2$ we have

$$\begin{aligned} &2t(1-t)[\langle y-x, y\rangle_i - \langle y-x, x\rangle_s] \\ &\quad \geq t\|x\|^2 + (1-t)\|y\|^2 - \|tx+(1-t)y\|^2 \\ &\quad \geq 2t(1-t)[\langle y-x, tx+(1-t)y\rangle_s - \langle y-x, tx+(1-t)y\rangle_i] \geq 0. \end{aligned} \tag{2.12}$$

We notice that inequality (2.12) holds for any $x, y \in X$, while in inequality (2.11) we must assume that x, y and $tx + (1-t)y$ are not zero.

REMARK 2.4. If the normed space is smooth, that is, the norm is Gâteaux differentiable at 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, [8]). In this situation inequality (2.10) becomes

$$\begin{aligned} &pt(1-t)(\|y\|^{p-2}[y-x, y] - \|x\|^{p-2}[y-x, x]) \\ &\quad \geq t\|x\|^p + (1-t)\|y\|^p - \|tx+(1-t)y\|^p \geq 0 \end{aligned} \tag{2.13}$$

and holds for any nonzero x and y. Moreover, if $(X, \langle\cdot, \cdot\rangle)$ is an inner product space, then (2.13) becomes

$$\begin{aligned} &pt(1-t)\langle y-x, \|y\|^{p-2}y - \|x\|^{p-2}x\rangle \\ &\quad \geq t\|x\|^p + (1-t)\|y\|^p - \|tx+(1-t)y\|^p \geq 0. \end{aligned} \tag{2.14}$$

https://doi.org/10.1017/S0004972710002054 Published online by Cambridge University Press

From (2.14) we deduce the particular inequalities of interest:

$$t(1-t)\left\langle y-x, \frac{y}{\|y\|}-\frac{x}{\|x\|}\right\rangle \geq t\|x\|+(1-t)\|y\|-\|tx+(1-t)y\| \geq 0 \quad (2.15)$$

and

$$2t(1-t)\|y-x\|^2 \geq t\|x\|^2+(1-t)\|y\|^2-\|tx+(1-t)y\|^2 \geq 0. \quad (2.16)$$

Obviously, inequality (2.16) can be proved directly on utilizing the properties of the inner products.

PROBLEM 2.5. It is an open question for the author whether or not inequality (2.16) characterizes the class of inner product spaces within the class of normed spaces.

3. A refinement of Jensen's inequality

For a convex function $f: X \to \mathbb{R}$ defined on a linear space X, perhaps one of the most important result is the well-known Jensen's inequality

$$f\left(\sum_{i=1}^{n} p_i x_i\right) \leq \sum_{i=1}^{n} p_i f(x_i), \quad (3.1)$$

which holds for any n-tuple of vectors $\mathbf{x}=(x_1, \ldots, x_n) \in X^n$ and any probability distribution $\mathbf{p}=(p_1, \ldots, p_n) \in \mathbb{P}^n$. The following refinement of Jensen's inequality holds.

THEOREM 3.1. *Let $f: X \to \mathbb{R}$ be a convex function defined on a linear space X. Then for any n-tuple of vectors $\mathbf{x}=(x_1, \ldots, x_n) \in X^n$ and any probability distribution $\mathbf{p}=(p_1, \ldots, p_n) \in \mathbb{P}^n$ we have the inequality*

$$\begin{aligned}\sum_{i=1}^{n} p_i f(x_i) - f\left(\sum_{i=1}^{n} p_i x_i\right) &\geq \sum_{k=1}^{n} p_k \nabla_+ f\left(\sum_{i=1}^{n} p_i x_i\right)(x_k) \\ &\quad - \nabla_+ f\left(\sum_{i=1}^{n} p_i x_i\right)\left(\sum_{i=1}^{n} p_i x_i\right) \geq 0.\end{aligned} \quad (3.2)$$

In particular, for the uniform distribution,

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n} f(x_i) - f\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) \\ &\quad \geq \frac{1}{n}\left[\sum_{k=1}^{n} \nabla_+ f\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)(x_k) - \nabla_+ f\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} x_i\right)\right] \geq 0.\end{aligned} \quad (3.3)$$

https://doi.org/10.1017/S0004972710002054 Published online by Cambridge University Press

PROOF. Utilizing the gradient inequality (2.2), we have

$$f(x_k) - f\left(\sum_{i=1}^n p_i x_i\right) \geq \nabla_+ f\left(\sum_{i=1}^n p_i x_i\right)\left(x_k - \sum_{i=1}^n p_i x_i\right) \tag{3.4}$$

for any $k \in \{1, \ldots, n\}$. By the subadditivity of the functional $\nabla_+ f(\cdot)(\cdot)$ in the second variable, we also have

$$\begin{aligned} &\nabla_+ f\left(\sum_{i=1}^n p_i x_i\right)\left(x_k - \sum_{i=1}^n p_i x_i\right) \\ &\quad \geq \nabla_+ f\left(\sum_{i=1}^n p_i x_i\right)(x_k) - \nabla_+ f\left(\sum_{i=1}^n p_i x_i\right)\left(\sum_{i=1}^n p_i x_i\right) \end{aligned} \tag{3.5}$$

for any $k \in \{1, \ldots, n\}$.

Utilizing inequalities (3.4) and (3.5) gives

$$\begin{aligned} &f(x_k) - f\left(\sum_{i=1}^n p_i x_i\right) \\ &\quad \geq \nabla_+ f\left(\sum_{i=1}^n p_i x_i\right)(x_k) - \nabla_+ f\left(\sum_{i=1}^n p_i x_i\right)\left(\sum_{i=1}^n p_i x_i\right) \end{aligned} \tag{3.6}$$

for any $k \in \{1, \ldots, n\}$. Now, if we multiply (3.6) by $p_k \geq 0$ and sum over k from 1 to n, then we deduce the first inequality in (3.2). The second inequality is obvious by the subadditivity property of the functional $\nabla_+ f(\cdot)(\cdot)$ in the second variable. □

The following particular case which provides a refinement for the generalized triangle inequality in normed linear spaces is of interest.

COROLLARY 3.2. *Let $(X, \|\cdot\|)$ be a normed linear space. Then for any $p \geq 1$, for any n-tuple of vectors $\mathbf{x} = (x_1, \ldots, x_n) \in X^n$ and any probability distribution $\mathbf{p} = (p_1, \ldots, p_n) \in \mathbb{P}^n$ with $\sum_{i=1}^n p_i x_i \neq 0$ we have the inequality*

$$\begin{aligned} &\sum_{i=1}^n p_i \|x_i\|^p - \left\|\sum_{i=1}^n p_i x_i\right\|^p \\ &\quad \geq p\left\|\sum_{i=1}^n p_i x_i\right\|^{p-2}\left[\sum_{k=1}^n p_k\left\langle x_k, \sum_{j=1}^n p_j x_j\right\rangle_s - \left\|\sum_{i=1}^n p_i x_i\right\|^2\right] \geq 0. \end{aligned} \tag{3.7}$$

If $p \geq 2$ the inequality holds for any n-tuple of vectors and probability distribution.

In particular, we have the norm inequalities

$$\begin{aligned} &\sum_{i=1}^n p_i \|x_i\| - \left\|\sum_{i=1}^n p_i x_i\right\| \\ &\quad \geq \left[\sum_{k=1}^n p_k\left\langle x_k, \frac{\sum_{i=1}^n p_i x_i}{\|\sum_{i=1}^n p_i x_i\|}\right\rangle_s - \left\|\sum_{i=1}^n p_i x_i\right\|\right] \geq 0 \end{aligned} \tag{3.8}$$

https://doi.org/10.1017/S0004972710002054 Published online by Cambridge University Press

and

$$\begin{aligned}\sum_{i=1}^{n} p_i \|x_i\|^2 - \Big\|\sum_{i=1}^{n} p_i x_i\Big\|^2 \\ \geq 2\Big[\sum_{k=1}^{n} p_k \Big\langle x_k, \sum_{i=1}^{n} p_i x_i\Big\rangle_s - \Big\|\sum_{i=1}^{n} p_i x_i\Big\|^2\Big] \geq 0.\end{aligned} \tag{3.9}$$

We notice that the first inequality in (3.9) is equivalent to

$$\sum_{i=1}^{n} p_i \|x_i\|^2 + \Big\|\sum_{i=1}^{n} p_i x_i\Big\|^2 \geq 2\sum_{k=1}^{n} p_k \Big\langle x_k, \sum_{i=1}^{n} p_i x_i\Big\rangle_s,$$

which provides the result

$$\frac{1}{2}\Big[\sum_{i=1}^{n} p_i \|x_i\|^2 + \Big\|\sum_{i=1}^{n} p_i x_i\Big\|^2\Big] \geq \sum_{k=1}^{n} p_k \Big\langle x_k, \sum_{i=1}^{n} p_i x_i\Big\rangle_s \quad \Big(\geq \Big\|\sum_{i=1}^{n} p_i x_i\Big\|^2\Big) \tag{3.10}$$

for any n-tuple of vectors and probability distribution.

REMARK 3.3. If in inequality (3.7) we consider the uniform distribution, then we get

$$\begin{aligned}\sum_{i=1}^{n} \|x_i\|^p - n^{1-p}\Big\|\sum_{i=1}^{n} x_i\Big\|^p \\ \geq pn^{1-p}\Big\|\sum_{i=1}^{n} x_i\Big\|^{p-2}\Big[\sum_{k=1}^{n}\Big\langle x_k, \sum_{i=1}^{n} x_i\Big\rangle_s - \Big\|\sum_{i=1}^{n} x_i\Big\|^2\Big] \geq 0.\end{aligned} \tag{3.11}$$

4. A reverse of Jensen's inequality

The following result is of interest as well.

THEOREM 4.1. *Let $f : X \to \mathbb{R}$ be a convex function defined on a linear space X. Then for any n-tuple of vectors $\mathbf{x} = (x_1, \ldots, x_n) \in X^n$ and any probability distribution $\mathbf{p} = (p_1, \ldots, p_n) \in \mathbb{P}^n$ we have the inequality*

$$\begin{aligned}\sum_{k=1}^{n} p_k \nabla_- f(x_k)(x_k) - \sum_{k=1}^{n} p_k \nabla_- f(x_k)\Big(\sum_{i=1}^{n} p_i x_i\Big) \\ \geq \sum_{i=1}^{n} p_i f(x_i) - f\Big(\sum_{i=1}^{n} p_i x_i\Big).\end{aligned} \tag{4.1}$$

In particular, for the uniform distribution,

$$\begin{aligned}\frac{1}{n}\Big[\sum_{k=1}^{n} \nabla_- f(x_k)(x_k) - \sum_{k=1}^{n} \nabla_- f(x_k)\Big(\frac{1}{n}\sum_{i=1}^{n} x_i\Big)\Big] \\ \geq \frac{1}{n}\sum_{i=1}^{n} f(x_i) - f\Big(\frac{1}{n}\sum_{i=1}^{n} x_i\Big).\end{aligned} \tag{4.2}$$

https://doi.org/10.1017/S0004972710002054 Published online by Cambridge University Press

PROOF. Utilizing the gradient inequality (2.2), we can state that

$$\nabla_- f(x_k)\left(x_k - \sum_{i=1}^n p_i x_i\right) \geq f(x_k) - f\left(\sum_{i=1}^n p_i x_i\right) \tag{4.3}$$

for any $k \in \{1, \ldots, n\}$. By the superadditivity of the functional $\nabla_- f(\cdot)(\cdot)$ in the second variable we also have

$$\nabla_- f(x_k)(x_k) - \nabla_- f(x_k)\left(\sum_{i=1}^n p_i x_i\right) \geq \nabla_- f(x_k)\left(x_k - \sum_{i=1}^n p_i x_i\right) \tag{4.4}$$

for any $k \in \{1, \ldots, n\}$. Therefore, by (4.3) and (4.4), we get

$$\nabla_- f(x_k)(x_k) - \nabla_- f(x_k)\left(\sum_{i=1}^n p_i x_i\right) \geq f(x_k) - f\left(\sum_{i=1}^n p_i x_i\right) \tag{4.5}$$

for any $k \in \{1, \ldots, n\}$. Finally, by multiplying (4.5) by $p_k \geq 0$ and summing over k from 1 to n, we deduce the desired inequality (4.1). □

REMARK 4.2. If the function f is defined on the Euclidian space $\mathbb{R}^n$ and is differentiable and convex, then from (4.1) we get the inequality

$$\begin{aligned}\sum_{k=1}^n p_k \langle \nabla f(x_k), x_k\rangle &- \left\langle \sum_{k=1}^n p_k \nabla f(x_k), \sum_{i=1}^n p_i x_i \right\rangle \\ &\geq \sum_{i=1}^n p_i f(x_i) - f\left(\sum_{i=1}^n p_i x_i\right)\end{aligned} \tag{4.6}$$

where, as usual, for $x_k = (x_k^1, \ldots, x_k^n)$,

$$\nabla f(x_k) = \left(\frac{\partial f(x_k)}{\partial x^1}, \ldots, \frac{\partial f(x_k)}{\partial x^n}\right).$$

This inequality was first obtained by Dragomir and Goh in 1996; see [11].

In one dimension we get the inequality

$$\sum_{k=1}^n p_k x_k f'(x_k) - \sum_{i=1}^n p_i x_i \sum_{k=1}^n p_k f'(x_k) \geq \sum_{i=1}^n p_i f(x_i) - f\left(\sum_{i=1}^n p_i x_i\right), \tag{4.7}$$

discovered in 1994 by Dragomir and Ionescu; see [12].

https://doi.org/10.1017/S0004972710002054 Published online by Cambridge University Press

The following reverse of the generalized triangle inequality holds.

COROLLARY 4.3. *Let $(X, \|\cdot\|)$ be a normed linear space. Then for any $p \geq 1$, for any n-tuple of vectors $\mathbf{x} = (x_1, \ldots, x_n) \in X^n \setminus \{(0, \ldots, 0)\}$ and any probability distribution $\mathbf{p} = (p_1, \ldots, p_n) \in \mathbb{P}^n$ we have the inequality*

$$p\left[\sum_{k=1}^{n} p_k \|x_k\|^p - \sum_{k=1}^{n} p_k \|x_k\|^{p-2} \left\langle \sum_{i=1}^{n} p_i x_i, x_k \right\rangle_i \right] \geq \sum_{i=1}^{n} p_i \|x_i\|^p - \left\| \sum_{i=1}^{n} p_i x_i \right\|^p. \tag{4.8}$$

In particular, we have the norm inequalities

$$\sum_{k=1}^{n} p_k \|x_k\| - \sum_{k=1}^{n} p_k \left\langle \sum_{i=1}^{n} p_i x_i, \frac{x_k}{\|x_k\|} \right\rangle_i \geq \sum_{i=1}^{n} p_i \|x_i\| - \left\| \sum_{i=1}^{n} p_i x_i \right\| \tag{4.9}$$

for $x_k \neq 0$, $k \in \{1, \ldots, n\}$ and

$$2\left[\sum_{k=1}^{n} p_k \|x_k\|^2 - \sum_{k=1}^{n} p_k \left\langle \sum_{j=1}^{n} p_j x_j, x_k \right\rangle_i \right] \geq \sum_{i=1}^{n} p_i \|x_i\|^2 - \left\| \sum_{i=1}^{n} p_i x_i \right\|^2, \tag{4.10}$$

for any x_k. We observe that inequality (4.10) is equivalent to

$$\sum_{i=1}^{n} p_i \|x_i\|^2 + \left\| \sum_{i=1}^{n} p_i x_i \right\|^2 \geq 2 \sum_{k=1}^{n} p_k \left\langle \sum_{j=1}^{n} p_j x_j, x_k \right\rangle_i,$$

which provides the interesting result

$$\frac{1}{2}\left[\sum_{i=1}^{n} p_i \|x_i\|^2 + \left\| \sum_{i=1}^{n} p_i x_i \right\|^2 \right] \geq \sum_{k=1}^{n} p_k \left\langle \sum_{j=1}^{n} p_j x_j, x_k \right\rangle_i \left(\geq \sum_{k=1}^{n} \sum_{j=1}^{n} p_j p_k \langle x_j, x_k \rangle_i \right) \tag{4.11}$$

for any n-tuple of vectors and probability distribution.

REMARK 4.4. If in inequality (4.8) we consider the uniform distribution, then we get

$$p\left[\sum_{k=1}^{n} \|x_k\|^p - \frac{1}{n} \sum_{k=1}^{n} \|x_k\|^{p-2} \left\langle \sum_{j=1}^{n} x_j, x_k \right\rangle_i \right] \geq \sum_{i=1}^{n} \|x_i\|^p - n^{1-p} \left\| \sum_{i=1}^{n} x_i \right\|^p. \tag{4.12}$$

For $p \in [1, 2)$, all vectors x_k should not be zero.

https://doi.org/10.1017/S0004972710002054 Published online by Cambridge University Press

5. Bounds for the mean f-deviation

Let X be a real linear space. For a convex function $f : X \to \mathbb{R}$ with the property that $f(0) \geq 0$, we define the *mean f-deviation* of an n-tuple of vectors $y = (y_1, \ldots, y_n) \in X^n$ with the probability distribution $\mathbf{p} = (p_1, \ldots, p_n) \in \mathbb{P}^n$ by the nonnegative quantity

$$K_{f(\cdot)}(\mathbf{p}, \mathbf{y}) = K_f(\mathbf{p}, \mathbf{y}) := \sum_{i=1}^{n} p_i f\left(y_i - \sum_{k=1}^{n} p_k y_k\right). \tag{5.1}$$

The fact that $K_f(\mathbf{p}, \mathbf{y})$ is nonnegative follows by Jensen's inequality, namely

$$K_f(\mathbf{p}, \mathbf{y}) \geq f\left(\sum_{i=1}^{n} p_i \left(y_i - \sum_{k=1}^{n} p_k y_k\right)\right) = f(0) \geq 0.$$

Of course the concept can be extended for any function defined on X. However, if the function is not convex, or if it is convex but $f(0) < 0$, then we are not sure about the positivity of the quantity $K_f(\mathbf{p}, \mathbf{y})$.

A natural example of such deviations can be provided by the convex function $f(y) := \|y\|^r$ with $r \geq 1$ defined on a normed linear space $(X, \|\cdot\|)$. We denote this by

$$K_r(\mathbf{p}, \mathbf{y}) := \sum_{i=1}^{n} p_i \left\| y_i - \sum_{k=1}^{n} p_k y_k \right\|^r \tag{5.2}$$

and call it the *mean r-absolute deviation* of the n-tuple of vectors $y = (y_1, \ldots, y_n) \in X^n$ with the probability distribution $\mathbf{p} = (p_1, \ldots, p_n)$.

Utilizing the result from [9], we can state then the following result providing a nontrivial lower bound for the mean f-deviation.

THEOREM 5.1. *Let $f : X \to [0, \infty)$ be a convex function with $f(0) = 0$. If $y = (y_1, \ldots, y_n) \in X^n$ and $\mathbf{p} = (p_1, \ldots, p_n)$ is a probability distribution with all p_i nonzero, then*

$$\begin{aligned} K_f(\mathbf{p}, \mathbf{y}) \geq \max_{k \in \{1,\ldots,n\}} \bigg\{ &(1 - p_k) f\left[\frac{p_k}{1 - p_k}\left(y_k - \sum_{l=1}^{n} p_l y_l\right)\right] \\ &+ p_k f\left(y_k - \sum_{l=1}^{n} p_l y_l\right)\bigg\} \quad (\geq 0). \end{aligned} \tag{5.3}$$

The case for mean r-absolute deviation is incorporated in the following corollary.

COROLLARY 5.2. *Let $(X, \|\cdot\|)$ be a normed linear space. If $y = (y_1, \ldots, y_n) \in X^n$ and $\mathbf{p} = (p_1, \ldots, p_n)$ is a probability distribution with all p_i nonzero, then for $r \geq 1$ we have*

$$K_r(\mathbf{p}, \mathbf{y}) \geq \max_{k \in \{1,\ldots,n\}} \left\{ [(1 - p_k)^{1-r} p_k^r + p_k] \left\| y_k - \sum_{l=1}^{n} p_l y_l \right\|^r \right\}. \tag{5.4}$$

REMARK 5.3. Since the function $h_r(t) := (1-t)^{1-r} t^r + t$, $r \geq 1$, $t \in [0, 1)$, is strictly increasing on [0, 1), then

$$\min_{k\in\{1,\dots,n\}} \{(1-p_k)^{1-r} p_k^r + p_k\} = p_m + (1-p_m)^{1-r} p_m^r,$$

where $p_m := \min_{k\in\{1,\dots,n\}} p_k$. By (5.4), we then obtain the simpler inequality

$$K_r(\mathbf{p}, \mathbf{y}) \geq [p_m + (1-p_m)^{1-r} \cdot p_m^r] \max_{k\in\{1,\dots,n\}} \left\| y_k - \sum_{l=1}^{n} p_l y_l \right\|^p, \tag{5.5}$$

which is perhaps more useful for applications.

We have the following double inequality for the mean f-mean deviation.

THEOREM 5.4. *Let $f : X \to [0, \infty)$ be a convex function with $f(0) = 0$. If $y = (y_1, \dots, y_n) \in X^n$ and $\mathbf{p} = (p_1, \dots, p_n)$ is a probability distribution with all p_i nonzero, then*

$$K_{\nabla_- f(\cdot)(\cdot)}(\mathbf{p}, \mathbf{y}) \geq K_{f(\cdot)}(\mathbf{p}, \mathbf{y}) \geq K_{\nabla_+ f(0)(\cdot)}(\mathbf{p}, \mathbf{y}) \geq 0. \tag{5.6}$$

PROOF. If we use inequality (3.2) for $x_i = y_i - \sum_{k=1}^n p_k y_k$, we get

$$\begin{aligned}
&\sum_{i=1}^{n} p_i f\left(y_i - \sum_{k=1}^{n} p_k y_k\right) - f\left(\sum_{i=1}^{n} p_i \left(y_i - \sum_{k=1}^{n} p_k y_k\right)\right) \\
&\quad \geq \sum_{j=1}^{n} p_j \nabla_+ f\left(\sum_{i=1}^{n} p_i \left(y_i - \sum_{k=1}^{n} p_k y_k\right)\right)\left(y_j - \sum_{k=1}^{n} p_k y_k\right) \\
&\qquad - \nabla_+ f\left(\sum_{i=1}^{n} p_i \left(y_i - \sum_{k=1}^{n} p_k y_k\right)\right)\left(\sum_{i=1}^{n} p_i \left(y_i - \sum_{k=1}^{n} p_k y_k\right)\right) \geq 0,
\end{aligned}$$

which is equivalent to the second part of (5.6).

Now, by utilizing the inequality (4.1) for the same choice of x_i, we get

$$\begin{aligned}
&\sum_{j=1}^{n} p_j \nabla_- f\left(y_j - \sum_{k=1}^{n} p_k y_k\right)\left(y_j - \sum_{k=1}^{n} p_k y_k\right) \\
&\quad - \sum_{k=1}^{n} p_j \nabla_- f\left(y_j - \sum_{k=1}^{n} p_k y_k\right)\left(\sum_{i=1}^{n} p_i \left(y_i - \sum_{k=1}^{n} p_k y_k\right)\right) \\
&\qquad \geq \sum_{i=1}^{n} p_i f\left(y_i - \sum_{k=1}^{n} p_k y_k\right) - f\left(\sum_{i=1}^{n} p_i \left(y_i - \sum_{k=1}^{n} p_k y_k\right)\right),
\end{aligned}$$

which in its turn is equivalent with the first inequality in (5.6). □

https://doi.org/10.1017/S0004972710002054 Published online by Cambridge University Press

We observe that as examples of convex functions defined on the entire normed linear space $(X, \|\cdot\|)$ that are convex and vanish in 0 we can consider the functions

$$f(x) := g(\|x\|), \quad x \in X,$$

where $g : [0, \infty) \to [0, \infty)$ is a monotonic nondecreasing convex function with $g(0) = 0$. For functions of this kind we have by direct computation that

$$\nabla_+ f(0)(u) = g'_+(0)\|u\| \quad \text{for any } u \in X$$

and

$$\nabla_- f(u)(u) = g'_-(\|u\|)\|u\| \quad \text{for any } u \in X.$$

We then have the following norm inequalities that are of interest.

COROLLARY 5.5. *Let $(X, \|\cdot\|)$ be a normed linear space. If $g : [0, \infty) \to [0, \infty)$ is a monotonic nondecreasing convex function with $g(0) = 0$, then for any $y = (y_1, \ldots, y_n) \in X^n$ and $\mathbf{p} = (p_1, \ldots, p_n)$ a probability distribution,*

$$\begin{aligned} \sum_{i=1}^{n} p_i g'_-\bigg(\bigg\|y_i - \sum_{k=1}^{n} p_k y_k\bigg\|\bigg)\bigg\|y_i - \sum_{k=1}^{n} p_k y_k\bigg\| \\ \geq \sum_{i=1}^{n} p_i g\bigg(\bigg\|y_i - \sum_{k=1}^{n} p_k y_k\bigg\|\bigg) \geq g'_+(0)\sum_{i=1}^{n} p_i \bigg\|y_i - \sum_{k=1}^{n} p_k y_k\bigg\|. \end{aligned} \tag{5.7}$$

6. Bounds for f-divergence measures

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

$$I_f(\mathbf{p}, \mathbf{q}) := \sum_{i=1}^{n} q_i f\left(\frac{p_i}{q_i}\right), \tag{6.1}$$

where $\mathbf{p} = (p_1, \ldots, p_n)$, $\mathbf{q} = (q_1, \ldots, q_n)$ 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

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

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 $\mathbf{q}$.

https://doi.org/10.1017/S0004972710002054 Published online by Cambridge University Press

(ii) For every $\mathbf{p}, \mathbf{q} \in R_+^n$,

$$I_f(\mathbf{p}, \mathbf{q}) \geq \sum_{j=1}^{n} q_j f\left(\frac{\sum_{j=1}^{n} p_j}{\sum_{j=1}^{n} q_j}\right). \tag{6.2}$$

If f is strictly convex, equality holds in (6.2) if and only if

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

If f is normalized, that is, $f(1) = 0$, 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

$$I_f(\mathbf{p}, \mathbf{q}) \geq 0. \tag{6.3}$$

In particular, if $\mathbf{p}, \mathbf{q} \in \mathbb{P}^n$, then (6.3) holds. This is the well-known positivity property of the f-divergence.

We endeavour to extend this concept to functions defined on a cone in a linear space as follows (see also [10]).

Firstly, we recall that the subset K in a linear space X is a *cone* if the following two conditions are satisfied:

(i) for any $x, y \in K$ we have $x + y \in K$;
(ii) for any $x \in K$ and any $\alpha \geq 0$ we have $\alpha x \in K$.

For a given n-tuple of vectors $\mathbf{z} = (z_1, \ldots, z_n) \in K^n$ and a probability distribution $\mathbf{q} \in \mathbb{P}^n$ with all values nonzero, we can define, for the convex function $f : K \to \mathbb{R}$, the following *f-divergence of* $\mathbf{z}$ *with the distribution* $\mathbf{q}$:

$$I_f(\mathbf{z}, \mathbf{q}) := \sum_{i=1}^{n} q_i f\left(\frac{z_i}{q_i}\right). \tag{6.4}$$

It is obvious that if $X = \mathbb{R}$, $K = [0, \infty)$ and $\mathbf{x} = \mathbf{p} \in \mathbb{P}^n$, then we obtain the usual concept of the f-divergence associated with a function $f : [0, \infty) \to \mathbb{R}$.

Now, for a given n-tuple of vectors $\mathbf{x} = (x_1, \ldots, x_n) \in K^n$, a probability distribution $\mathbf{q} \in \mathbb{P}^n$ with all values nonzero and for any nonempty subset J of $\{1, \ldots, n\}$, we have

$$\mathbf{q}_J := (Q_J, \bar{Q}_J) \in \mathbb{P}^2$$

where $Q_J := \sum_{i \in J} q_j$, $\bar{Q}_J := Q_{\bar{J}}$, $\bar{J} := \{1, \ldots, n\} \setminus J$ and

$$\mathbf{x}_J := (X_J, \bar{X}_J) \in K^2$$

in which, as above,

$$X_J := \sum_{i \in J} x_i \quad \text{and} \quad \bar{X}_J := X_{\bar{J}}.$$

https://doi.org/10.1017/S0004972710002054 Published online by Cambridge University Press

It is obvious that

$$I_f(\mathbf{x}_J, \mathbf{q}_J) = Q_J f\left(\frac{X_J}{Q_J}\right) + \bar{Q}_J f\left(\frac{\bar{X}_J}{\bar{Q}_J}\right).$$

The following inequality for the f-divergence of an n-tuple of vectors in a linear space holds [10].

THEOREM 6.1. *Let $f : K \to \mathbb{R}$ be a convex function on the cone K. Then for any n-tuple of vectors $\mathbf{x} = (x_1, \ldots, x_n) \in K^n$, a probability distribution $\mathbf{q} \in \mathbb{P}^n$ with all values nonzero and for any nonempty subset J of $\{1, \ldots, n\}$ we have*

$$\begin{aligned} I_f(\mathbf{x}, \mathbf{q}) &\geq \max_{\emptyset \neq J \subset \{1,\ldots,n\}} I_f(\mathbf{x}_J, \mathbf{q}_J) \geq I_f(\mathbf{x}_J, \mathbf{q}_J) \\ &\geq \min_{\emptyset \neq J \subset \{1,\ldots,n\}} I_f(\mathbf{x}_J, \mathbf{q}_J) \geq f(X_n) \end{aligned} \tag{6.5}$$

where $X_n := \sum_{i=1}^{n} x_i$.

We observe that, for a given n-tuple of vectors $\mathbf{x} = (x_1, \ldots, x_n) \in K^n$, a sufficient condition for the positivity of $I_f(\mathbf{x}, \mathbf{q})$ for any probability distribution $\mathbf{q} \in \mathbb{P}^n$ with all values nonzero is that $f(X_n) \geq 0$. In the scalar case and if $\mathbf{x} = \mathbf{p} \in \mathbb{P}^n$, then a sufficient condition for the positivity of the f-divergence $I_f(\mathbf{p}, \mathbf{q})$ is that $f(1) \geq 0$.

The case of functions of a real variable that is of interest for applications is incorporated in [10].

COROLLARY 6.2. *Let $f : [0, \infty) \to \mathbb{R}$ be a normalized convex function. Then for any $\mathbf{p}, \mathbf{q} \in \mathbb{P}^n$ we have*

$$I_f(\mathbf{p}, \mathbf{q}) \geq \max_{\emptyset \neq J \subset \{1,\ldots,n\}} \left[Q_J f\left(\frac{P_J}{Q_J}\right) + (1 - Q_J) f\left(\frac{1 - P_J}{1 - Q_J}\right)\right] \quad (\geq 0). \tag{6.6}$$

In what follows, by using the results in Theorems 3.1 and 4.1, we can provide an upper and a lower bound for the positive difference $I_f(\mathbf{x}, \mathbf{q}) - f(X_n)$.

THEOREM 6.3. *Let $f : K \to \mathbb{R}$ be a convex function on the cone K. Then for any n-tuple of vectors $\mathbf{x} = (x_1, \ldots, x_n) \in K^n$ and a probability distribution $\mathbf{q} \in \mathbb{P}^n$ with all values nonzero,*

$$\begin{aligned} I_{\nabla_- f(\cdot)(\cdot)}(\mathbf{x}, \mathbf{q}) - I_{\nabla_- f(\cdot)(X_n)}(\mathbf{x}, \mathbf{q}) &\geq I_f(\mathbf{x}, \mathbf{q}) - f(X_n) \\ &\geq I_{\nabla_+ f(X_n)(\cdot)}(\mathbf{x}, \mathbf{q}) - \nabla_+ f(X_n)(X_n) \geq 0. \end{aligned} \tag{6.7}$$

The case of functions of a real variable that is useful for applications is as follows.

COROLLARY 6.4. *Let $f : [0, \infty) \to \mathbb{R}$ be a normalized convex function. Then for any $\mathbf{p}, \mathbf{q} \in \mathbb{P}^n$ we have*

$$I_{f'_-(\cdot)(\cdot)}(\mathbf{p}, \mathbf{q}) - I_{f'_-(\cdot)}(\mathbf{p}, \mathbf{q}) \geq I_f(\mathbf{p}, \mathbf{q}) \geq 0, \tag{6.8}$$

or, equivalently,

$$I_{f'_-(\cdot)[(\cdot)-1]}(\mathbf{p}, \mathbf{q}) \geq I_f(\mathbf{p}, \mathbf{q}) \geq 0. \tag{6.9}$$

https://doi.org/10.1017/S0004972710002054 Published online by Cambridge University Press

The above corollary is useful for providing an upper bound in terms of the variational distance for the f-divergence $I_f(\mathbf{p}, \mathbf{q})$ of normalized convex functions whose derivatives are bounded above and below.

PROPOSITION 6.5. *Let $f : [0, \infty) \to \mathbb{R}$ be a normalized convex function and $\mathbf{p}, \mathbf{q} \in \mathbb{P}^n$. If there exist constants γ and Γ with*

$$-\infty < \gamma \le f'_-\left(\frac{p_k}{q_k}\right) \le \Gamma < \infty \quad \textit{for all } k \in \{1, \ldots, n\},$$

then we have the inequality

$$0 \le I_f(\mathbf{p}, \mathbf{q}) \le \tfrac{1}{2}(\Gamma - \gamma) V(\mathbf{p}, \mathbf{q}), \tag{6.10}$$

where

$$V(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^{n} q_i \left| \frac{p_i}{q_i} - 1 \right| = \sum_{i=1}^{n} |p_i - q_i|.$$

PROOF. By inequality (6.9) we have successively that

$$\begin{aligned}
0 &\le I_f(\mathbf{p}, \mathbf{q}) \le I_{f'_-(\cdot)[(\cdot)-1]}(\mathbf{p}, \mathbf{q}) \\
&= \sum_{i=1}^{n} q_i \left(\frac{p_i}{q_i} - 1\right) \left[f'_-\left(\frac{p_i}{q_i}\right) - \frac{\Gamma + \gamma}{2} \right] \\
&\le \sum_{i=1}^{n} q_i \left| \frac{p_i}{q_i} - 1 \right| \left| f'_-\left(\frac{p_i}{q_i}\right) - \frac{\Gamma + \gamma}{2} \right| \\
&\le \frac{1}{2}(\Gamma - \gamma) \sum_{i=1}^{n} q_i \left| \frac{p_i}{q_i} - 1 \right|,
\end{aligned}$$

which proves the desired result (6.10). □

COROLLARY 6.6. *Let $f : [0, \infty) \to \mathbb{R}$ be a normalized convex function and $\mathbf{p}, \mathbf{q} \in \mathbb{P}^n$. If there exist constants r and R with*

$$0 < r \le \frac{p_k}{q_k} \le R < \infty \quad \textit{for all } k \in \{1, \ldots, n\},$$

then we have the inequality

$$0 \le I_f(\mathbf{p}, \mathbf{q}) \le \tfrac{1}{2}[f'_-(R) - f'_-(r)] V(\mathbf{p}, \mathbf{q}). \tag{6.11}$$

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

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

https://doi.org/10.1017/S0004972710002054 Published online by Cambridge University Press

Finally, the following proposition giving another upper bound in terms of the χ^2-divergence can be stated.

PROPOSITION 6.7. *Let $f : [0, \infty) \to \mathbb{R}$ be a normalized convex function and $\mathbf{p}, \mathbf{q} \in \mathbb{P}^n$. If there exists a constant $0 < \Delta < \infty$ with*

$$\left| \frac{f'_-(\frac{p_i}{q_i}) - f'_-(1)}{\frac{p_i}{q_i} - 1} \right| \le \Delta \quad \text{for all } k \in \{1, \ldots, n\}, \tag{6.12}$$

then we have the inequality

$$0 \le I_f(\mathbf{p}, \mathbf{q}) \le \Delta \chi^2(p, q). \tag{6.13}$$

In particular, if $f'_-(\cdot)$ satisfies the local Lipschitz condition

$$|f'_-(x) - f'_-(1)| \le \Delta |x - 1| \quad \text{for any } x \in (0, \infty) \tag{6.14}$$

then (6.13) holds true for any $\mathbf{p}, \mathbf{q} \in \mathbb{P}^n$.

PROOF. We have from (6.9) that

$$\begin{aligned}
0 &\le I_f(\mathbf{p}, \mathbf{q}) \le I_{f'_-(\cdot)[(\cdot)-1]}(\mathbf{p}, \mathbf{q}) \\
&= \sum_{i=1}^n q_i \left(\frac{p_i}{q_i} - 1 \right) \left[f'_- \left(\frac{p_i}{q_i} \right) - f'_-(1) \right] \\
&\le \sum_{i=1}^n q_i \left(\frac{p_i}{q_i} - 1 \right)^2 \left| \frac{f'_-(\frac{p_i}{q_i}) - f'_-(1)}{\frac{p_i}{q_i} - 1} \right| \\
&\le \Delta \sum_{i=1}^n q_i \left(\frac{p_i}{q_i} - 1 \right)^2
\end{aligned}$$

and inequality (6.13) is obtained. □

REMARK 6.8. It is obvious that if one chooses, in the above inequalities, particular normalized convex functions that generate the Kullback–Leibler, Jeffreys, Hellinger or other divergence measures or discrepancies, then one can obtain some results of interest. However, the details are not provided here.

References

[1] I. Csiszár, 'Information-type measures of differences of probability distributions and indirect observations', *Studia Sci. Math. Hungar.* **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, 'An improvement of Jensen's inequality', *Bull. Math. Soc. Sci. Math. Roumanie* **34**(82) (1990), 291–296.

https://doi.org/10.1017/S0004972710002054 Published online by Cambridge University Press

[4] S. S. Dragomir, 'Some refinements of Ky Fan's inequality', *J. Math. Anal. Appl.* **163**(2) (1992), 317–321.
[5] S. S. Dragomir, 'Some refinements of Jensen's inequality', *J. Math. Anal. Appl.* **168**(2) (1992), 518–522.
[6] S. S. Dragomir, 'A further improvement of Jensen's inequality', *Tamkang J. Math.* **25**(1) (1994), 29–36.
[7] S. S. Dragomir, 'A new improvement of Jensen's inequality', *Indian J. Pure Appl. Math.* **26**(10) (1995), 959–968.
[8] S. S. Dragomir, *Semi-inner Products and Applications* (Nova Science Publishers, New York, 2004).
[9] S. S. Dragomir, 'A refinement of Jensen's inequality with applications for f-divergence measures', *Taiwanese J. Math.* **14**(1) (2010), 153–164.
[10] S. S. Dragomir, 'A new refinement of Jensen's inequality in linear spaces with applications', *RGMIA Res. Rep. Coll.* **12** (2009), Supplement, Article 6. http://www.staff.vu.edu.au/RGMIA/v12(E).asp.
[11] S. S. Dragomir and C. J. Goh, 'A counterpart of Jensen's discrete inequality for differentiable convex mappings and applications in information theory', *Math. Comput. Modelling* **24**(2) (1996), 1–11.
[12] S. S. Dragomir and N. M. Ionescu, 'Some converse of Jensen's inequality and applications', *Rev. Anal. Numér. Théor. Approx.* **23**(1) (1994), 71–78.
[13] S. S. Dragomir, J. Pečarić and L. E. Persson, 'Properties of some functionals related to Jensen's inequality', *Acta Math. Hungar.* **70**(1–2) (1996), 129–143.
[14] J. Pečarić and S. S. Dragomir, 'A refinement of Jensen inequality and applications', *Stud. Univ. Babeş-Bolyai Math.* **24**(1) (1989), 15–19.

S. S. DRAGOMIR, Mathematics, School of Engineering and Science,
Victoria University, PO Box 14428 Melbourne City,
MC 8001, Australia
and
School of Computational and Applied Mathematics, University of the Witwatersrand,
Private Bag 3, Wits 2050, Johannesburg, South Africa
e-mail: sever.dragomir@vu.edu.au

https://doi.org/10.1017/S0004972710002054 Published online by Cambridge University Press