KYUNGPOOK Math. J. 52(2012), 495-511
http://dx.doi.org/10.5666/KMJ.2012.52.4.495

Improvement of Jensen's Inequality in terms of Gâteaux Derivatives for Convex Functions in Linear Spaces with Applications

Muhammad Adil Khan*
Department of Mathematics, University of Peshawar, Pakistan
e-mail : adilbandai@yahoo.com

Sadia Khalid
Abdus Salam School of Mathematical Sciences, GC University, 68-B, New Muslim Town, Lahore 54600, Pakistan
e-mail : saadiakhalid176@gmail.com

Josip Pečarić
Faculty of Textile Technology, University of Zagreb, Prilaz baruna Filipovića 28a, 10000 Zagreb, Croatia
e-mail : pecaric@element.hr

Abstract. In this paper, we prove some inequalities in terms of Gâteaux derivatives for convex functions defined on linear spaces and also give improvement of Jensen's inequality. Furthermore, we give applications for norms, mean f-deviations and f-divergence measures.

1. Introduction

Undoubtedly, Jensen's inequality is the most important inequality in analysis, because it implies at once the main part of the other classical inequalities (e.g. Hölder's inequality, Minkowski's inequality, Young's inequality, AGM inequality, generalized triangle inequality, etc.). There is an extensive literature devoted to Jensen's inequality concerning different generalizations, refinements, counterparts and converse results see e.g. [1], [5]-[16] and also the references in them.

Let C be a convex subset of the linear space X and f be a convex function on

* Corresponding Author.
Received September 2, 2011; accepted August 22, 2012.
2010 Mathematics Subject Classification: 26D15.
Key words and phrases: Convex functions, Gâteaux derivatives, Jensen's inequality, Norms, Mean f-deviations, f-Divergence measures.
This research work is partially supported by Higher Education Commission, Pakistan. The research of the third author is supported by the Croatian Ministry of Science, Education and Sports under the Research Grants 117-1170889-0888.

C. If $\mathbf{p} = (p_,...,p_n)$ is a probability sequence and $\mathbf{x} = (x_1,...,x_n) \in C^n$, then the Jensen inequality

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

holds.

Assume that $f : X \to \mathbb{R}$ is a convex function defined on a 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 limit exists

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

and is called the left(right) Gâteaux derivative of the function f at the point x in the direction y.

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

$$\begin{aligned} \frac{f(x+ty)-f(x)}{t} &\geq \nabla_+ f(x)(y) = \inf_{t>0}\left[\frac{f(x+ty)-f(x)}{t}\right] \\ \geq \sup_{s<0}\left[\frac{f(x+sy)-f(x)}{s}\right] &= \nabla_- f(x)(y) \geq \frac{f(x+sy)-f(x)}{s} \end{aligned} \tag{1.2}$$

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

$$\nabla_- f(u)(u-v) \geq f(u) - f(v) \geq \nabla_+ f(v)(u-v) \tag{1.3}$$

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

The following properties are also of great importance:

$$\nabla_+ f(x)(-y) = -\nabla_- f(x)(y) \tag{1.4}$$

and

$$\nabla_{+(-)} f(x)(\alpha y) = \alpha \nabla_{+(-)} f(x)(y) \tag{1.5}$$

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

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

$$\nabla_+ f(x)(y+z) \leq \nabla_+ f(x)(y) + \nabla_+ f(x)(z) \tag{1.6}$$

and

$$\nabla_- f(x)(y+z) \geq \nabla_- f(x)(y) + \nabla_- f(x)(z) \tag{1.7}$$

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

Some natural examples can be provided by using the normed spaces.

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

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

For a comprehensive study of the properties of these mappings in the Geometry of Banach Spaces see the monograph [10].

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 $y \in X$.

This class of functions will be used to illustrate the inequalities obtained in the general case of convex functions defined on entire linear spaces.

In [9], the author proved the following refinement of Jensen's inequality. As applications, inequalities for norms, mean f- deviation and f-divergence measure are also given.

Theorem 1.1. *Let $f : X \to \mathbb{R}$ be a convex function defined on a real linear space X. Then for any n-tuple of vectors $\boldsymbol{x} = (x_1, ..., x_n) \in X^n$ and for any probability distribution $\boldsymbol{p} = (p_1, ..., p_n) \in \mathbb{P}^n$, we have*

$$\sum_{j=1}^{n} p_j f(x_j) - f\left(\sum_{j=1}^{n} p_j x_j\right) \geq \sum_{k=1}^{n} p_k \nabla_+ f\left(\sum_{j=1}^{n} p_j x_j\right)(x_k) - \nabla_+ f\left(\sum_{j=1}^{n} p_j x_j\right)\left(\sum_{j=1}^{n} p_j x_j\right) \geq 0. \tag{1.8}$$

In the same paper author also proved the following reverse of Jensen's inequality:

Theorem 1.2. *Under the assumptions of Theorem 1.1, the following inequality holds*

$$\sum_{j=1}^{n} p_j f(x_j) - f\left(\sum_{j=1}^{n} p_j x_j\right) \leq \sum_{k=1}^{n} p_k \nabla_- f(x_k)(x_k) - \sum_{k=1}^{n} p_k \nabla_- f(x_k)\left(\sum_{j=1}^{n} p_j x_j\right). \tag{1.9}$$

A particular case of interest is for $f(x) = ||x||^p$, where $(X, ||.||)$ is a normed linear space. Then for any $p \geq 1$, for any n-tuple of vectors $\mathbf{x} = (x_1, ..., x_n) \in X^n$ and any probability distribution $\mathbf{p} = (p_1, ..., p_n) \in \mathbb{P}$ with $\sum_{i=1}^{n} p_i x_i \neq 0$, we have

$$\sum_{j=1}^{n} p_j ||x_j||^p - \Big|\Big|\sum_{j=1}^{n} p_j x_j\Big|\Big|^p \geq p\Big|\Big|\sum_{j=1}^{n} p_j x_j\Big|\Big|^{p-2}\left[\sum_{j=1}^{n} p_j \Big\langle x_j, \sum_{k=1}^{n} p_k x_k\Big\rangle_s - \Big|\Big|\sum_{j=1}^{n} p_j x_j\Big|\Big|^2\right]. \tag{1.10}$$

Also, for any $p \geq 1$, for any n-tuple of vectors $\mathbf{x} = (x_1, ..., x_n) \in X^n \setminus \{(0, ..., 0)\}$ and any probability distribution $\mathbf{p} = (p_1, ..., p_n) \in \mathbb{P}$, we have

$$\sum_{j=1}^{n} p_j ||x_j||^p - \Big|\Big|\sum_{j=1}^{n} p_j x_j\Big|\Big|^p \leq p\left[\sum_{j=1}^{n} p_j ||x_j||^p - \sum_{j=1}^{n} p_j ||x_j||^{p-2}\Big\langle \sum_{k=1}^{n} p_k x_k, x_j\Big\rangle_i\right]. \tag{1.11}$$

This paper is organized in the following manner: in Section 2, we prove some inequalities in terms of Gâteaux derivatives for convex functions defined on linear spaces, which implies inequalities (1.8) and (1.9). We also discuss a particular case for norm. In Section 3, we give improvement of Jensen's inequality. Particularly, we provide an improvement for the generalized triangle inequality. In the remaining parts of this paper, we give applications for mean f-deviations and f-divergence measures.

2. Inequalities for convex functions

Theorem 2.1. *Let $f : X \to \mathbb{R}$ be a convex function defined on a linear space X, $\boldsymbol{x} = (x_1, ..., x_n) \in X^n$ be any n-tuple of vectors and $\boldsymbol{p} = (p_1, ..., p_n) \in \mathbb{P}^n$ be any* *probability distribution. If $c, d \in X$ are arbitrary chosen vectors, then we have*

$$(2.1) \quad f(c) + \sum_{i=1}^{n} p_i \nabla_+ f(c)(x_i) - \nabla_+ f(c)(c) \leq \sum_{i=1}^{n} p_i f(x_i) \leq f(d) + \sum_{i=1}^{n} p_i \nabla_- f(x_i)(x_i) - \sum_{i=1}^{n} p_i \nabla_- f(x_i)(d).$$

Proof. For fix index i, we can take $u = x_i$ and $v = c$ in the second inequality of (1.3) to obtain

$$(2.2) \quad f(x_i) - f(c) \geq \nabla_+ f(c)(x_i - c).$$

By using the subadditivity of $\nabla_+ f(.)(.)$ in the second variable, we have

$$(2.3) \quad \nabla_+ f(c)(x_i - c) \geq \nabla_+ f(c)(x_i) - \nabla_+ f(c)(c).$$

Combining (2.3) and (2.2), we get

$$(2.4) \quad f(x_i) - f(c) \geq \nabla_+ f(c)(x_i) - \nabla_+ f(c)(c).$$

Now, if we multiply (2.4) by p_i and summing over $i = 1, 2, ..., n$, we deduce the first inequality in (2.1).

To obtain the second inequality in (2.1), we first put $u = x_i$ and $v = d$ in the first inequality of (1.3) and rewrite it in the form

$$(2.5) \quad f(x_i) - f(d) \leq \nabla_- f(x_i)(x_i - d).$$

By using the superadditivity of $\nabla_- f(.)(.)$ in the second variable, we have

$$(2.6) \quad \nabla_- f(x_i)(x_i - d) \leq \nabla_- f(x_i)(x_i) - \nabla_- f(x_i)(d).$$

Combining (2.6) and (2.5), we get

$$(2.7) \quad f(x_i) - f(d) \leq \nabla_- f(x_i)(x_i) - \nabla_- f(x_i)(d).$$

Multiplying by p_i and summing over $i = 1, 2, ..., n$, we get second inequality in (2.1). □

Remark 2.2. If we set $c = d = \sum_{k=1}^{n} p_k x_k$ in (2.1), then we have (1.8) and (1.9).

Remark 2.3. Related inequalities in terms of subdifferential of a convex function defined on linear space, have been proved by Matić and Pečarić in [15].

The following particular case for norms may be stated:

Corollary 2.4. *Let $(X, ||.||)$ be a normed linear space, $p \geq 1$, $\boldsymbol{x} = (x_1, ..., x_n) \in$*

$X^n\backslash\{(0,...,0)\}$ be any n-tuple of vectors and $\boldsymbol{p}=(p_1,...,p_n)\in\mathbb{P}^n$ be any probability distribution. If $c,d\in X$, $c\neq 0$, are arbitrary chosen vectors, then we have

$$\begin{aligned}||c||^p+p\sum_{j=1}^n p_j||c||^{p-2}\langle x_j,c\rangle_s-p||c||^p &\le \sum_{j=1}^n p_j||x_j||^p\\ &\le ||d||^p+p\sum_{j=1}^n p_j||x_j||^p-p\sum_{j=1}^n p_j||x_j||^{p-2}\langle d,x_j\rangle_i. \qquad (2.8)\end{aligned}$$

If $p\ge 2$, then (2.8) holds for any $c,d,x_j\in X(j=1,..,n)$ and any probability distribution.

In particular, we have the norm inequalities

$$\sum_{j=1}^n p_j\langle x_j,\frac{c}{||c||}\rangle_s\le\sum_{j=1}^n p_j||x_j||\le ||d||+\sum_{j=1}^n p_j||x_j||-\sum_{j=1}^n p_j\big\langle d,\frac{x_j}{||x_j||}\big\rangle_i. \qquad (2.9)$$

for $x_j,c\neq 0,\ j\in\{1,...,n\}$ and

$$\begin{aligned}2\sum_{j=1}^n p_j\langle x_j,c\rangle_s-||c||^2 &\le \sum_{j=1}^n p_j||x_j||^2 \qquad (2.10)\\ &\le ||d||^2+2\sum_{j=1}^n p_j||x_j||^2-2\sum_{j=1}^n p_j\langle d,x_j\rangle_i.\end{aligned}$$

Remark 2.5. If we set $c=d=\sum_{k=1}^n p_kx_k$ and apply Corollary 2.1, then we have (1.10) and (1.11).

3. Improvement of Jensen's inequality

Theorem 3.1. *Let $f:X\to\mathbb{R}$ be a convex function defined on a linear space X, $\boldsymbol{x}=(x_1,...,x_n)\in X^n$ be any n-tuple of vectors and $\boldsymbol{p}=(p_1,...,p_n)\in\mathbb{P}^n$ be any probability distribution. If $c\in X$ is arbitrary chosen vector, then we have*

$$\begin{aligned}&\sum_{i=1}^n p_if(x_i)-f(c)-\nabla_+f(c)\left(\sum_{i=1}^n p_ix_i-c\right)\\ &\qquad\ge\left|\sum_{i=1}^n p_i\Big|f(x_i)-f(c)\Big|-\sum_{i=1}^n p_i\Big|\nabla_+f(c)(x_i-c)\Big|\right|. \qquad (3.1)\end{aligned}$$

Proof. From (2.2), we have

$$f(x_i)-f(c)-\nabla_+f(c)(x_i-c)\ge 0. \qquad (3.2)$$

Therefore

$$(3.3)\quad f(x_i) - f(c) - \nabla_+ f(c)(x_i - c) = \left| f(x_i) - f(c) - \nabla_+ f(c)(x_i - c) \right| \geq \left| \left| f(x_i) - f(c) \right| - \left| \nabla_+ f(c)(x_i - c) \right| \right|.$$

Multiplying (3.3) by p_i and summing over $i = 1, 2, ..., n$, we get

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

$$(3.4)\quad \geq \left| \sum_{i=1}^{n} p_i \left| f(x_i) - f(c) \right| - \sum_{i=1}^{n} p_i \left| \nabla_+ f(c)(x_i - c) \right| \right|.$$

Using (1.5) and (1.6), we have

$$(3.5)\quad \nabla_+ f(c) \left(\sum_{i=1}^{n} p_i x_i - c \right) \leq \sum_{i=1}^{n} p_i \nabla_+ f(c)(x_i - c).$$

Now, by using (3.5) in (3.4), we have (3.1). □

The following improvement of Jensen's inequality is valid:

Corollary 3.2. *Let $f : X \to \mathbb{R}$ be a convex function defined on a linear space X. Then for any n-tuple of vectors $\boldsymbol{x} = (x_1, ..., x_n) \in X^n$ and any probability distribution $\boldsymbol{p} = (p_1, ..., p_n) \in \mathbb{P}^n$, we have*

$$(3.6)\quad \sum_{i=1}^{n} p_i f(x_i) - f\left(\sum_{i=1}^{n} p_i x_i \right) \geq \left| \sum_{i=1}^{n} p_i \left| f(x_i) - f\left(\sum_{k=1}^{n} p_k x_k \right) \right| - \sum_{i=1}^{n} p_i \left| \nabla_+ f\left(\sum_{k=1}^{n} p_k x_k \right) \left(x_i - \sum_{k=1}^{n} p_k x_k \right) \right| \right|.$$

In particular, for the uniform distribution, we have

$$(3.7)\quad \frac{1}{n} \sum_{i=1}^{n} f(x_i) - f\left(\frac{1}{n} \sum_{i=1}^{n} x_i \right) \geq \left| \frac{1}{n} \sum_{i=1}^{n} \left| f(x_i) - f\left(\frac{1}{n} \sum_{k=1}^{n} x_k \right) \right| - \frac{1}{n} \sum_{i=1}^{n} \left| \nabla_+ f\left(\frac{1}{n} \sum_{k=1}^{n} x_k \right) \left(x_i - \frac{1}{n} \sum_{k=1}^{n} x_k \right) \right| \right|.$$

Proof. By setting $c = \sum_{k=1}^{n} p_k x_k$ in (3.1), we get (3.6). □

Remark 3.3. If the function f is defined on the Euclidian space $\mathbb{R}^n$ and is differentiable and convex, then from (3.6), we have

$$\sum_{i=1}^{n} p_i f(x_i) - f\left(\sum_{i=1}^{n} p_i x_i\right) \tag{3.8}$$
$$\geq \left|\sum_{i=1}^{n} p_i \left|f(x_i) - f\left(\sum_{k=1}^{n} p_k x_k\right)\right| - \sum_{i=1}^{n} p_i \left|\left\langle \nabla f\left(\sum_{k=1}^{n} p_k x_k\right), x_i - \sum_{k=1}^{n} p_k x_k \right\rangle\right|\right|,$$

where, as usual, $x_i = (x_i^1, ..., x_i^n)$ and $\nabla f(x_i) = \left(\frac{\partial f(x_i)}{\partial x_i^1}, ..., \frac{\partial f(x_i)}{\partial x_i^n}\right)$.

For one dimensional case, we have

$$\sum_{i=1}^{n} p_i f(x_i) - f\left(\sum_{i=1}^{n} p_i x_i\right) \geq \left|\sum_{i=1}^{n} p_i \left|f(x_i) - f\left(\sum_{k=1}^{n} p_k x_k\right)\right|\right. \tag{3.9}$$
$$\left. - \left|f'\left(\sum_{k=1}^{n} p_k x_k\right)\right| \sum_{i=1}^{n} p_i \left|x_i - \sum_{k=1}^{n} p_k x_k\right|\right|,$$

that was proved in 2008 by Pečarić et al., see [12] (also see [1]).

The following particular case for norms may be stated:

Corollary 3.4. *Let $(X, ||.||)$ be a normed linear space, $p \geq 1$, $\boldsymbol{x} = (x_1, ..., x_n) \in X^n$ be any n-tuple of vectors and $\boldsymbol{p} = (p_1, ..., p_n) \in \mathbb{P}^n$ be any probability distribution. If $c \in X$ is non zero arbitrary chosen vector, then we have*

$$\sum_{i=1}^{n} p_i ||x_i||^p - ||c||^p - p||c||^{p-2} \left\langle \sum_{i=1}^{n} p_i x_i - c, c \right\rangle_s \tag{3.10}$$
$$\geq \left|\sum_{i=1}^{n} p_i \left| ||x_i||^p - ||c||^p \right| - p||c||^{p-2} \sum_{i=1}^{n} p_i \left|\langle x_i - c, c\rangle_s\right|\right|.$$

If $p \geq 2$, then the inequality holds for any vector c.

In particular, we have the norm inequalities

$$\sum_{i=1}^{n} p_i ||x_i|| - ||c|| - \left\langle \sum_{i=1}^{n} p_i x_i - c, \frac{c}{||c||} \right\rangle_s \tag{3.11}$$
$$\geq \left|\sum_{i=1}^{n} p_i \left| ||x_i|| - ||c|| \right| - \sum_{i=1}^{n} p_i \left|\langle x_i - c, \frac{c}{||c||}\rangle_s\right|\right|$$

for $c \neq 0$ and

$$\sum_{i=1}^{n} p_i ||x_i||^2 - ||c||^2 - 2\left\langle \sum_{i=1}^{n} p_i x_i - c, c \right\rangle_s \tag{3.12}$$
$$\geq \left|\sum_{i=1}^{n} p_i \left| ||x_i||^2 - ||c||^2 \right| - 2\sum_{i=1}^{n} p_i \left|\langle x_i - c, c\rangle_s\right|\right|.$$

The following particular case that provides an improvement for the generalized triangle inequality in the normed linear spaces is of interest:

Corollary 3.5. *Let $(X, ||.||)$ be a normed linear space. Then for any $p \geq 1$, $\boldsymbol{x} = (x_1, ..., x_n) \in X^n$ and for any probability distribution $\boldsymbol{p} = (p_1, ..., p_n) \in \mathbb{P}^n$ with $\sum_{i=1}^n p_i x_i \neq 0$, we have*

$$\sum_{i=1}^n p_i ||x_i||^p - \Big|\Big| \sum_{i=1}^n p_i x_i \Big|\Big|^p \geq \Bigg| \sum_{i=1}^n p_i \Big| ||x_i||^p - \Big|\Big| \sum_{k=1}^n p_k x_k \Big|\Big|^p \Big| - p \Big|\Big| \sum_{i=1}^n p_i x_i \Big|\Big|^{p-2} \sum_{i=1}^n p_i \Big| \Big\langle x_i - \sum_{k=1}^n p_k x_k, \sum_{k=1}^n p_k x_k \Big\rangle_s \Big| \Bigg|. \tag{3.13}$$

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

In particular, we have the norm inequalities

$$\sum_{i=1}^n p_i ||x_i|| - \Big|\Big| \sum_{i=1}^n p_i x_i \Big|\Big| \geq \Bigg| \sum_{i=1}^n p_i \Big| ||x_i|| - \Big|\Big| \sum_{k=1}^n p_k x_k \Big|\Big| \Big| - \sum_{i=1}^n p_i \Big| \Big\langle x_i - \sum_{k=1}^n p_k x_k, \frac{\sum_{k=1}^n p_k x_k}{\big|\big| \sum_{k=1}^n p_k x_k \big|\big|} \Big\rangle_s \Big| \Bigg| \tag{3.14}$$

and

$$\sum_{i=1}^n p_i ||x_i||^2 - \Big|\Big| \sum_{i=1}^n p_i x_i \Big|\Big|^2 \geq \Bigg| \sum_{i=1}^n p_i \Big| ||x_i||^2 - \Big|\Big| \sum_{k=1}^n p_k x_k \Big|\Big|^2 \Big| - 2 \sum_{i=1}^n p_i \Big| \Big\langle x_i - \sum_{k=1}^n p_k x_k, \sum_{k=1}^n p_k x_k \Big\rangle_s \Big| \Bigg|. \tag{3.15}$$

Remark 3.6. If in inequality (3.13), we consider the uniform distribution, then we have

$$\sum_{i=1}^n ||x_i||^p - n^{1-p} \Big|\Big| \sum_{i=1}^n x_i \Big|\Big|^p \geq \Bigg| \sum_{i=1}^n \Big| ||x_i||^p - n^{-p} \Big|\Big| \sum_{k=1}^n x_k \Big|\Big|^p \Big| - p n^{2-p} \Big|\Big| \sum_{i=1}^n x_i \Big|\Big|^{p-2} \sum_{i=1}^n \Big| \Big\langle x_i - \frac{1}{n} \sum_{k=1}^n x_k, \frac{1}{n} \sum_{k=1}^n x_k \Big\rangle_s \Big| \Bigg|. \tag{3.16}$$

4. 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 n-tuple of vectors $\mathbf{y} = (y_1, ..., y_n) \in X^n$ with the probability distribution $\mathbf{p} = (p_1, ..., p_n) \in \mathbb{P}^n$ by the non-negative quantity

$$K_{f(.)}(\mathbf{p}, \mathbf{y}) = K_f(\mathbf{p}, \mathbf{y}) := \sum_{i=1}^{n} p_i f\left(y_i - \sum_{k=1}^{n} p_k y_k\right). \tag{4.1}$$

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

$$K_f(\mathbf{p}, \mathbf{y}) \geq f\left(\sum_{i=1}^{n} p_i \left(y_i - \sum_{k=1}^{n} p_k y_k\right)\right) = f(0) \geq 0. \tag{4.2}$$

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 $(X; ||.||)$. We denote this by

$$K_f(\mathbf{p}, \mathbf{y}) := \sum_{i=1}^{n} p_i \left|\left| y_i - \sum_{k=1}^{n} p_k y_k \right|\right|^r \tag{4.3}$$

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

Utilizing (1.8) and (1.9), we can state the following result providing a non-trivial lower and upper bound for the mean f-deviation (see [9]).

Theorem 4.1. *Let $f : X \to [0, \infty)$ be a convex function with $f(0) = 0$. If $\boldsymbol{y} = (y_1, ..., y_n) \in X^n$ and $\boldsymbol{p} = (p_1, ..., p_n) \in \mathbb{P}^n$ is the probability distribution with all $p_{i's}$ $(i = 1, ..., n)$ non zero, then*

$$K_{\nabla_+ f(0)(.)}(\boldsymbol{p}, \boldsymbol{y}) \leq K_{f(.)}(\boldsymbol{p}, \boldsymbol{y}) \leq K_{\nabla_- f(.)(.)}(\boldsymbol{p}, \boldsymbol{y}). \tag{4.4}$$

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

Theorem 4.2. *Let $f : X \to [0, \infty)$ be a convex function with $f(0) = 0$, $\boldsymbol{y} = (y_1, ..., y_n) \in X^n$ and $\boldsymbol{p} = (p_1, ..., p_n) \in \mathbb{P}^n$ be any probability distribution. If $c, d \in X$ are arbitrary chosen vectors, then we have*

$$\begin{aligned} f(c) + K_{\nabla_+ f(c)(.)}(\boldsymbol{p}, \boldsymbol{y}) - \nabla_+ f(c)(c) &\leq K_{f(.)}(\boldsymbol{p}, \boldsymbol{y}) \\ &\leq f(d) + K_{\nabla_- f(.)(.)}(\boldsymbol{p}, \boldsymbol{y}) - K_{\nabla_- f(.)(d)}(\boldsymbol{p}, \boldsymbol{y}). \end{aligned} \tag{4.5}$$

Proof. If we use the second inequality of (2.1) for $x_i = y_i - \sum_{k=1}^{n} p_k y_k$, we have

$$\begin{aligned}(4.6)\qquad &\sum_{i=1}^{n} p_i f\left(y_i - \sum_{k=1}^{n} p_k y_k\right)\\ &\leq f(d) + \sum_{i=1}^{n} p_i \nabla_- f\left(y_i - \sum_{k=1}^{n} p_k y_k\right)\left(y_i - \sum_{k=1}^{n} p_k y_k\right)\\ &- \sum_{i=1}^{n} p_i \nabla_- f\left(y_i - \sum_{k=1}^{n} p_k y_k\right)(d),\end{aligned}$$

which is equivalent to the second part of (4.5).

Now, by utilizing the first inequality of (2.1) for the same choice of x_i, we have

$$\begin{aligned}(4.7)\qquad f(c) + \sum_{i=1}^{n} p_i \nabla_+ f(c)\left(y_i - \sum_{k=1}^{n} p_k y_k\right) - \nabla_+ f(c)(c)\\ \leq \sum_{i=1}^{n} p_i f\left(y_i - \sum_{k=1}^{n} p_k y_k\right),\end{aligned}$$

which in turns is equivalent to the first inequality of (4.5). □

Remark 4.3. If all the assumptions of Theorem 4.1 are satisfied and if we set $c = d = 0$ in (4.5), then we have (4.4).

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

Theorem 4.4. *Let $f : X \to [0, \infty)$ be a convex function, $\boldsymbol{y} = (y_1, ..., y_n) \in X^n$ and $\boldsymbol{p} = (p_1, ..., p_n) \in \mathbb{P}^n$ be any probability distribution. If $c \in X$ is arbitrary chosen vector, then we have*

$$\begin{aligned}(4.8)\ K_f(\boldsymbol{p}, \boldsymbol{y}) - f(c) - \nabla_+ f(c)(-c) \geq \Bigg|\sum_{i=1}^{n} p_i \Bigg| f\left(y_i - \sum_{k=1}^{n} p_k y_k\right) - f(c)\Bigg|\\ - \sum_{i=1}^{n} p_i \Bigg|\nabla_+ f(c)\left(y_i - \sum_{k=1}^{n} p_k y_k - c\right)\Bigg|\Bigg|.\end{aligned}$$

Proof. By using the inequality (3.1) for $x_i = y_i - \sum_{k=1}^{n} p_k y_k$, we have (4.8). □

By using Theorem 4.4, we can give the following result providing a non-trivial lower bound for the mean f-deviation.

Corollary 4.5. *Under the assumptions of Theorem 4.1, we have*

$$(4.9)\qquad K_f(\boldsymbol{p}, \boldsymbol{y}) \geq \Bigg| K_f(\boldsymbol{p}, \boldsymbol{y}) - \sum_{i=1}^{n} p_i \Bigg|\nabla_+ f(0)\left(y_i - \sum_{k=1}^{n} p_k y_k\right)\Bigg|\Bigg|.$$

We can consider the function

$$f(x) := g(||x||), \ x \in X \tag{4.10}$$

as an example of convex function defined on the normed linear space $(X, ||.||)$ and vanishes at 0, where $g : [0, \infty) \to [0, \infty)$ is monotonic nondecreasing convex function with $g(0) = 0$. For this kind of functions, by direct computation, we have

$$\nabla_+ f(0)(u) = g'_+(0)||u|| \text{ for any } u \in X. \tag{4.11}$$

We then have the following norm inequality that is of interest:

Corollary 4.6. Let $(X, ||.||)$ 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 n-tuple of vectors $\mathbf{y} = (y_1, ..., y_n) \in X^n$ and for any probability distribution $\mathbf{p} = (p_1, ..., p_n) \in \mathbb{P}^n$, we have

$$\begin{aligned} \sum_{i=1}^{n} p_i g\left(\left|\left|y_i - \sum_{k=1}^{n} p_k y_k\right|\right|\right) & \\ \geq \left| \sum_{i=1}^{n} p_i g\left(\left|\left|y_i - \sum_{k=1}^{n} p_k y_k\right|\right|\right) \right. & \left. - g'_+(0) \sum_{i=1}^{n} p_i \left|\left|y_i - \sum_{k=1}^{n} p_k y_k\right|\right| \right|. \end{aligned} \tag{4.12}$$

5. Bounds for f-divergence measure

Given a convex function $f : \mathbb{R}_+ \to \mathbb{R}_+$, the f-divergence functional

$$I_f(\mathbf{p}, \mathbf{q}) := \sum_{i=1}^{n} q_i f\left(\frac{p_i}{q_i}\right), \tag{5.1}$$

where $\mathbf{p} = (p_1, ..., p_n)$, $\mathbf{q} = (q_1, ..., q_n)$ are positive sequences, was introduced by Csiszár in [3], as a generalized measure of information, a distance function on the set of probability distributions $\mathbb{P}^n$. As in [3], we interpret undefined expressions by

$$f(0) = \lim_{t \to 0^+} f(t), \quad 0f\left(\frac{0}{0}\right) = 0,$$

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

The following results were essentially given by Csiszár and Körner [4]:

(i) If f is convex, then $I_f(\mathbf{p}, \mathbf{q})$ is jointly convex in $\mathbf{p}$ and $\mathbf{q}$;

(ii) For every $\mathbf{p}, \mathbf{q} \in \mathbb{R}_+^n$, we have

$$I_f(\mathbf{p}, \mathbf{q}) \geq \sum_{j=1}^{n} q_j f\left(\frac{\sum_{j=1}^{n} p_j}{\sum_{j=1}^{n} q_j}\right). \tag{5.2}$$

If f is strictly convex, equality holds in (5.2) if and only if

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

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

$$I_f(\mathbf{p}, \mathbf{q}) \geq 0. \tag{5.3}$$

In particular, if $\mathbf{p}, \mathbf{q} \in \mathbb{P}^n$, then (5.3) holds. This is the well-known positivity property of the f-divergence.

We give some examples of divergence measures in Information Theory which are particular cases of Csiszár f-divergences such as Kullback-Leibler divergence, χ^2-divergence, α-order entropy distance and Bhattacharyya distance etc..

The Kullback-Leibler divergence (see[14]) can be obtained for the convex function $f : (0, \infty) \to \mathbb{R}$ defined by $f(x) = x \log x$ and is given by

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

The K. Pearson χ^2-divergence can be obtained for the convex function $f : (0, \infty) \to \mathbb{R}$ defined by $f(x) = (1 - x)^2,\ x \in \mathbb{R}$ and is given by

$$\chi^2(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^{n} \frac{(p_i - q_i)^2}{q_i}.$$

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

$$I_f(\mathbf{p}, \mathbf{q}) := \sum_{i=1}^{n} q_i f\left(\frac{p_i}{q_i}\right) = -\sum_{i=1}^{n} q_i \log\left(\frac{p_i}{q_i}\right) = \sum_{i=1}^{n} q_i \log\left(\frac{q_i}{p_i}\right) = KL(\mathbf{q}, \mathbf{p}). \tag{5.4}$$

For $\alpha > 1$, let $f(x) = x^\alpha$, where $x > 0$. Then α-order entropy (see[17]) is

$$I_\alpha(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^{n} p_i^\alpha q_i^{1-\alpha}.$$

For the convex function $f(x) = -\sqrt{x},\ x > 0$, we have

$$I_f(\mathbf{p}, \mathbf{q}) = -\sum_{i=1}^{n} \sqrt{p_i q_i} = -B(\mathbf{p}, \mathbf{q}),$$

where $B(\mathbf{p},\mathbf{q}) = \sum_{i=1}^{n} \sqrt{p_i q_i}$ is Bhattacharyya distance (see for example [13]).

We endeavour to extend this concept for functions defined on a cone in a linear space as follows (see [11]).

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 the convex function $f : K \to \mathbb{R}$, we can define the following f-divergence of $\mathbf{z}$ with the distribution $\mathbf{q}$

$$I_f(\mathbf{z},\mathbf{q}) := \sum_{i=1}^{n} q_i f\left(\frac{z_i}{q_i}\right), \tag{5.5}$$

where $\mathbf{z} = (z_1, ..., z_n) \in K^n$ is the n-tuple of vectors and $\mathbf{q} \in \mathbb{P}^n$ is the probability distribution with all values non zero.
It is obvious that if $X = \mathbb{R}$, $K = [0,\infty)$ and $\mathbf{z} = \mathbf{p} \in \mathbb{P}^n$, then we obtain the usual concept of the f-divergence associated with a function $f : [0,\infty) \to \mathbb{R}$.

The following inequalities for the f-divergence of n-tuple of vectors in the linear spaces hold (see [9]):

Theorem 5.1. *Let $f : K \to \mathbb{R}$ be a convex function on the cone K. Then for any n-tuple of vectors $\mathbf{x} = (x_1, ..., x_n) \in K^n$ and a probability distribution $\boldsymbol{q} = (q_1, ..., q_n) \in \mathbb{P}^n$ with all values non zero, we have*

$$\begin{aligned} &I_{\nabla_+ f(\sum_{i=1}^{n} x_i)(.)}(\boldsymbol{x},\boldsymbol{q}) - \nabla_+ f(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} x_i) \\ &\quad \leq I_f(\boldsymbol{x},\boldsymbol{q}) - f\Big(\sum_{i=1}^{n} x_i\Big) \leq I_{\nabla_- f(.)(.)}(\boldsymbol{x},\boldsymbol{q}) - I_{\nabla_- f(.)(\sum_{i=1}^{n} x_i)}(\boldsymbol{x},\boldsymbol{q}). \end{aligned} \tag{5.6}$$

By using the results of Theorem 2.1, we can provide a lower and upper bound of $I_f(\mathbf{x},\mathbf{q})$.

Theorem 5.2. *Let $f : K \to \mathbb{R}$ be a convex function on the cone K, $\mathbf{x} = (x_1, ..., x_n) \in K^n$ be n-tuple of vectors and $\boldsymbol{q} = (q_1, ..., q_n) \in \mathbb{P}^n$ be the probability distribution with all values non zero. If $c, d \in X$ are arbitrary chosen vectors, then we have*

$$\begin{aligned} &f(c) + I_{\nabla_+ f(c)(.)}(\boldsymbol{x},\boldsymbol{q}) - \nabla_+ f(c)(c) \\ &\quad \leq I_f(\boldsymbol{x},\boldsymbol{q}) \leq f(d) + I_{\nabla_- f(.)(.)}(\boldsymbol{x},\boldsymbol{q}) - I_{\nabla_- f(.)(d)}(\boldsymbol{x},\boldsymbol{q}). \end{aligned} \tag{5.7}$$

Remark 5.3. If all the assumptions of Theorem 5.2 are satisfied and if we set $c = d = \sum_{i=1}^{n} x_i$ in (5.7), then we have (5.6).

Remark 5.4. Theorem 5.2 for the case of real variable normalized convex function is useful for applications (see [9]).

By using the results of Theorem 3.1, we can provide a lower bound for $I_f(\mathbf{x}, \mathbf{q})$.

Theorem 5.5. *Under the assumptions of Theorem 5.2, we have*

$$(5.8) \qquad I_f(\boldsymbol{x}, \boldsymbol{q}) - f(c) - \nabla_+ f(c)\left(\sum_{i=1}^{n} x_i - c\right) \geq \left|\sum_{i=1}^{n} q_i \left| f\left(\frac{x_i}{q_i}\right) - f(c)\right| - \sum_{i=1}^{n} q_i \left|\nabla_+ f(c)\left(\frac{x_i}{q_i} - c\right)\right|\right|.$$

The special case of Theorem 5.5 provides a lower bound for the positive difference $I_f(\mathbf{x}, \mathbf{q}) - f\left(\sum_{i=1}^{n} x_i\right)$.

Corollary 5.6. *Under the assumptions of Theorem 5.2, we have*

$$(5.9) \quad I_f(\boldsymbol{x}, \boldsymbol{q}) - f\left(\sum_{i=1}^{n} x_i\right) \geq \left|\sum_{i=1}^{n} q_i \left| f\left(\frac{x_i}{q_i}\right) - f\left(\sum_{i=1}^{n} x_i\right)\right| - \sum_{i=1}^{n} q_i \left|\nabla_+ f\left(\sum_{i=1}^{n} x_i\right)\left(\frac{x_i}{q_i} - \sum_{i=1}^{n} x_i\right)\right|\right|.$$

If the function f is differentiable and convex and K is the subset of Euclidean space $\mathbb{R}^n$, then from (5.9), we have

$$(5.10) \quad I_f(\boldsymbol{x}, \boldsymbol{q}) - f\left(\sum_{i=1}^{n} x_i\right) \geq \left|\sum_{i=1}^{n} q_i \left| f\left(\frac{x_i}{q_i}\right) - f\left(\sum_{i=1}^{n} x_i\right)\right| - \sum_{i=1}^{n} q_i \left|\left\langle \nabla f\left(\sum_{i=1}^{n} x_i\right), \frac{x_i}{q_i} - \sum_{i=1}^{n} x_i\right\rangle\right|\right|.$$

The special case of Theorem 5.5 for functions of real variable that is of interest for applications:

Theorem 5.7. *Let $f : [0, \infty) \to \mathbb{R}$ be a differentiable convex function, $\boldsymbol{p}, \boldsymbol{q} \in \mathbb{P}^n$ be any two probability distributions with all values nonzero. If $c \in [0, \infty]$, then we have*

$$(5.11) \qquad I_f(\boldsymbol{p}, \boldsymbol{q}) - f(c) - f'(c)(1 - c) \geq \left|\sum_{i=1}^{n} q_i \left| f\left(\frac{p_i}{q_i}\right) - f(c)\right| - \left|f'(c)\right| \sum_{i=1}^{n} q_i \left|\frac{p_i}{q_i} - c\right|\right|.$$

Corollary 5.8. *For any two probability distributions* $\boldsymbol{p}, \boldsymbol{q} \in \mathbb{P}^n$ *with all values nonzero and* $c \in (0, \infty)$*, we have*

$$\begin{aligned}(5.12)\quad & KL(\boldsymbol{p}, \boldsymbol{q}) - 1 + c - \log c \\ & \geq \left| \sum_{i=1}^{n} q_i \left| \frac{p_i}{q_i} \log\left(\frac{p_i}{q_i}\right) - c\log c \right| - \left| (1 + \log c) \right| \sum_{i=1}^{n} q_i \left| \frac{p_i}{q_i} - c \right| \right|.\end{aligned}$$

and

$$\begin{aligned}(5.13)\quad & KL(\boldsymbol{q}, \boldsymbol{p}) + \log c + \frac{1}{c} - 1 \\ & \geq \left| \sum_{i=1}^{n} q_i \left| \log\left(\frac{q_i}{p_i}\right) + \log c \right| - \frac{1}{c} \sum_{i=1}^{n} q_i \left| \frac{p_i}{q_i} - c \right| \right|.\end{aligned}$$

Corollary 5.9. *For any two probability distributions* $\boldsymbol{p}, \boldsymbol{q} \in \mathbb{P}^n$ *with all values nonzero and* $c \in \mathbb{R}$*, we have*

$$\begin{aligned}(5.14)\quad & \chi^2(\boldsymbol{p}, \boldsymbol{q}) + (1 - c)^2 \\ & \geq \left| \sum_{i=1}^{n} q_i \left| \left(1 - \frac{p_i}{q_i}\right)^2 - (1 - c)^2 \right| - 2\left|1 - c\right| \sum_{i=1}^{n} q_i \left| \frac{p_i}{q_i} - c \right| \right|.\end{aligned}$$

Corollary 5.10. *For any two probability distributions* $\boldsymbol{p}, \boldsymbol{q} \in \mathbb{P}^n$ *with all values nonzero,* $c \in (0, \infty)$ *and for* $\alpha > 1$*, we have*

$$\begin{aligned}(5.15)\quad & I_\alpha(\boldsymbol{p}, \boldsymbol{q}) + c^\alpha(\alpha - 1) - \alpha c^{\alpha-1} \\ & \geq \left| \sum_{i=1}^{n} q_i \left| p_i^\alpha q_i^{-\alpha} - c^\alpha \right| - \alpha c^{\alpha-1} \sum_{i=1}^{n} q_i \left| \frac{p_i}{q_i} - c \right| \right|.\end{aligned}$$

Corollary 5.11. *For any two probability distributions* $\boldsymbol{p}, \boldsymbol{q} \in \mathbb{P}^n$ *with all values nonzero and* $c \in (0, \infty)$*, we have*

$$(5.16)\quad -B(\boldsymbol{p}, \boldsymbol{q}) + \frac{\sqrt{c}}{2} + \frac{1}{2\sqrt{c}} \geq \left| \sum_{i=1}^{n} q_i \left| \sqrt{\frac{p_i}{q_i}} - \sqrt{c} \right| - \frac{1}{2\sqrt{c}} \sum_{i=1}^{n} q_i \left| \frac{p_i}{q_i} - c \right| \right|.$$

Remark 5.12. It is obvious that if in the above inequalities, one chooses the other particular convex functions that generates Jeffreys, Hellinger or other divergence measures or discrepancies, then one can obtain some results of interest. For some choice of c, the above results are also useful for finding the lower bound of different divergences (see [1, 2, 12]).

References

[1] M. Adil Khan, M. Anwar, J. Jakšetić and J. Pečarić, *On some improvements of the Jensen inequality with some applications*, J. Inequal. and Appl., (2009), Article ID 323615, 15 pages.

[2] M. Anwar, S. Hussain and J. Pečarić, *Some inequalities for Csiszár-divergence measures*, Int. J. Math. Anal., **3**(26)(2009), 1295-1304.

[3] I. Csiszár, *Information-type measures of difference of probability distributions and indirect observations*, Studia Math. Hungarica, **2**(1967), 299-318.

[4] I. Csiszár and J. Korner, Information theory: Coding Theorem for Dicsrete Memoryless systems, Academic Press, New York, 1981.

[5] S. S. Dragomir, *An improvement of Jensen's inequality*, Bull. Math. Soc. Sci. Math. Roumanie, **34**(82)(1990), 291-296.

[6] S. S. Dragomir, *Some refinements of Ky Fan's inequality*, J. Math. Anal. Appl., **163**(2)(1992), 317-321.

[7] S. S. Dragomir, *Some refinements of Jensen's inequality*, J. Math. Anal. Appl., **168**(2)(1992), 518-522.

[8] S. S. Dragomir, *A further improvement of Jensen's inequality*, Tamkang J. Math., **25**(1)(1994), 29-36.

[9] S. S. Dragomir, *Inequalities in terms of the gâteaux derivatives for convex functions on linear spaces with applications*, Bulletin of the Australian Mathematical Society, **83**(2011), 500-517.

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

[11] S. S. Dragomir, *A new refinement of Jensen's inequality in linear spaces with applications*, Mathematical and Computer Modelling, **52**(2010), 1497-1505.

[12] S. Hussain, J. Pečarić, *An improvement of Jensen's inequality with some applications*, Asian European Journal of Mathematics, **2**(1)(2009), 85-94.

[13] J. N. Kapur, *A comparative assessment of various measures of directed divergence*, Advances in Management Studies, **3**(1)(1984), 1-16.

[14] S. Kullback and R. A. Leiber, *On information and sufficency*, Ann. Math. Statist., **22**(1951), 79-86.

[15] M. Matić, J. Pečarić, *Some companion inequalities to Jensen's inequality*, Math. Inequal. Appl., **3**(3)(2000), 355-368.

[16] J. Pečarić and S. S. Dragomir, *A refinements of Jensen's inequality and applications*, Studia Univ. Babes-Bolyai, Mathematica, **24**(1)(1989), 15-19.

[17] A. Rényi, *On measures of entropy and information*, Proc. Fourth Berkeley Symp. Math. Stat. and Prob., University of California Press, **1**(1961), 547-561.