Received 16/04/17

RECENT DEVELOPMENTS OF DISCRETE INEQUALITIES FOR CONVEX FUNCTIONS DEFINED ON LINEAR SPACES WITH APPLICATIONS

SILVESTRU SEVER DRAGOMIR[1,2]

Abstract. In this paper we survey some recent discrete inequalities for functions defined on convex subsets of general linear spaces. Various refinements and reverses of Jensen's discrete inequality are presented. The Slater inequality version for these functions is outlined. As applications, we establish several bounds for the *mean f-deviation* of an n-tuple of vectors as well as for the *f-divergence* of an n-tuple of vectors given a discrete probability distribution. Examples for the K. Pearson χ^2-*divergence,* the *Kullback-Leibler divergence,* the *Jeffreys divergence,* the *total variation distance* and other divergence measures are also provided.

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 arithmetic mean-geometric mean inequality, Hölder and Minkowski inequalities, Ky Fan's inequality etc. can be obtained as particular cases of it. In order to state some recent reverses of Jensen's discrete inequality for functions of a real variable we need the following facts.

If $x_i,\ y_i \in \mathbb{R}$ and $w_i \geq 0$ $(i = 1, \dots, n)$ with $W_n := \sum_{i=1}^{n} w_i = 1$ then we may consider the *Čebyšev functional*

$$T_w(x, y) := \sum_{i=1}^{n} w_i x_i y_i - \sum_{i=1}^{n} w_i x_i \sum_{i=1}^{n} w_i y_i. \tag{1.1}$$

The following result is known in the literature as the *Grüss inequality*

$$|T_w(x, y)| \leq \frac{1}{4} (\Gamma - \gamma)(\Delta - \delta), \tag{1.2}$$

provided

$$-\infty < \gamma \leq x_i \leq \Gamma < \infty, \ \ -\infty < \delta \leq y_i \leq \Delta < \infty \tag{1.3}$$

for $i = 1, \dots, n$.

The constant $\frac{1}{4}$ is sharp in the sense that it cannot be replaced by a smaller constant.

1991 *Mathematics Subject Classification.* Primary 26D15, 26D20; Secondary 94A05, 47A63.

Key words and phrases. Convex functions on linear spaces, Discrete Jensen's inequality, Reverse of Jensen's inequality, Discrete divergence measures, f-Divergence measures,

If we assume that $-\infty < \gamma \leq x_i \leq \Gamma < \infty$ for $i = 1, \ldots, n$, then by the Grüss inequality for $y_i = x_i$ and by the Schwarz's discrete inequality, we have

$$\sum_{i=1}^{n} w_i \left| x_i - \sum_{j=1}^{n} w_j x_j \right| \leq \left[\sum_{i=1}^{n} w_i x_i^2 - \left(\sum_{j=1}^{n} w_j x_j \right)^2 \right]^{\frac{1}{2}} \leq \frac{1}{2} (\Gamma - \gamma). \tag{1.4}$$

In order to provide a reverse of the celebrated Jensen's inequality for convex functions of a real variable, S. S. Dragomir obtained in 2002 [14] the following result:

Theorem 1. *Let $f : [m, M] \to \mathbb{R}$ be a differentiable convex function on (m, M). If $x_i \in [m, M]$ and $w_i \geq 0 \quad (i = 1, \ldots, n)$ with $W_n := \sum_{i=1}^{n} w_i = 1$, then one has the counterpart of Jensen's weighted discrete inequality:*

$$\begin{aligned} 0 &\leq \sum_{i=1}^{n} w_i f(x_i) - f\left(\sum_{i=1}^{n} w_i x_i \right) \\ &\leq \sum_{i=1}^{n} w_i f'(x_i) x_i - \sum_{i=1}^{n} w_i f'(x_i) \sum_{i=1}^{n} w_i x_i \\ &\leq \frac{1}{2} [f'(M) - f'(m)] \sum_{i=1}^{n} w_i \left| x_i - \sum_{j=1}^{n} w_j x_j \right|. \end{aligned} \tag{1.5}$$

Remark 1. *We notice that the inequality between the first and the second term in (1.5) was proved in 1994 by Dragomir & Ionescu, see* [28].

On making use of (1.4), we can state the following string of reverse inequalities

$$\begin{aligned} 0 &\leq \sum_{i=1}^{n} w_i f(x_i) - f\left(\sum_{i=1}^{n} w_i x_i \right) \\ &\leq \sum_{i=1}^{n} w_i f'(x_i) x_i - \sum_{i=1}^{n} w_i f'(x_i) \sum_{i=1}^{n} w_i x_i \\ &\leq \frac{1}{2} [f'(M) - f'(m)] \sum_{i=1}^{n} w_i \left| x_i - \sum_{j=1}^{n} w_j x_j \right| \\ &\leq \frac{1}{2} [f'(M) - f'(m)] \left[\sum_{i=1}^{n} w_i x_i^2 - \left(\sum_{j=1}^{n} w_j x_j \right)^2 \right]^{\frac{1}{2}} \\ &\leq \frac{1}{4} [f'(M) - f'(m)] (M - m), \end{aligned} \tag{1.6}$$

provided that $f : [m, M] \subset \mathbb{R} \to \mathbb{R}$ is a differentiable convex function on (m, M), $x_i \in [m, M]$ and $w_i \geq 0 \quad (i = 1, \ldots, n)$ with $W_n := \sum_{i=1}^{n} w_i = 1$.

Remark 2. *We notice that the inequality between the first, second and last term from (1.6) was proved in the general case of positive linear functionals in 2001 by S. S. Dragomir in* [13].

The following reverse Jensen's inequality for convex functions of a real variable also holds:

Theorem 2 (Dragomir, 2013 [23]). *Let $f : I \to \mathbb{R}$ be a continuous convex function on the interval of real numbers I and $m, M \in \mathbb{R}$, $m < M$ with $[m, M] \subset \mathring{I}$, $\mathring{I}$ is the interior of I. If $x_i \in [m, M]$ and $w_i \geq 0$ $(i = 1, \ldots, n)$ with $W_n := \sum_{i=1}^n w_i = 1$, then*

$$
\begin{aligned}
(1.7) \qquad 0 &\leq \sum_{i=1}^{n} w_i f(x_i) - f\left(\sum_{i=1}^{n} w_i x_i\right) \\
&\leq \frac{\left(M - \sum_{i=1}^n w_i x_i\right)\left(\sum_{i=1}^n w_i x_i - m\right)}{M - m} \Psi_f\left(\sum_{i=1}^{n} w_i x_i; m, M\right) \\
&\leq \frac{\left(M - \sum_{i=1}^n w_i x_i\right)\left(\sum_{i=1}^n w_i x_i - m\right)}{M - m} \sup_{t \in (m,M)} \Psi_f(t; m, M) \\
&\leq \left(M - \sum_{i=1}^{n} w_i x_i\right)\left(\sum_{i=1}^{n} w_i x_i - m\right) \frac{f'_-(M) - f'_+(m)}{M - m} \\
&\leq \frac{1}{4}(M - m)\left[f'_-(M) - f'_+(m)\right],
\end{aligned}
$$

where $\Psi_f(\cdot; m, M) : (m, M) \to \mathbb{R}$ is defined by

$$
\Psi_f(t; m, M) = \frac{f(M) - f(t)}{M - t} - \frac{f(t) - f(m)}{t - m}.
$$

We also have the inequality

$$
\begin{aligned}
(1.8) \qquad 0 &\leq \sum_{i=1}^{n} w_i f(x_i) - f\left(\sum_{i=1}^{n} w_i x_i\right) \\
&\leq \frac{\left(M - \sum_{i=1}^n w_i x_i\right)\left(\sum_{i=1}^n w_i x_i - m\right)}{M - m} \Psi_f\left(\sum_{i=1}^{n} w_i x_i; m, M\right) \\
&\leq \frac{1}{4}(M - m)\Psi_f\left(\sum_{i=1}^{n} w_i x_i; m, M\right) \\
&\leq \frac{1}{4}(M - m)\sup_{t \in (m,M)} \Psi_f(t; m, M) \\
&\leq \frac{1}{4}(M - m)\left[f'_-(M) - f'_+(m)\right],
\end{aligned}
$$

provided that $\sum_{i=1}^n w_i x_i \in (m, M)$.

The following result also holds:

Theorem 3 (Dragomir, 2013 [23])**.** *With the assumptions of Theorem 2, we have the inequalities*

$$0\le\sum_{i=1}^{n}w_if(x_i)-f\left(\sum_{i=1}^{n}w_ix_i\right)\le 2\max\left\{\frac{M-\sum_{i=1}^{n}w_ix_i}{M-m},\frac{\sum_{i=1}^{n}w_ix_i-m}{M-m}\right\}\times\left[\frac{f(m)+f(M)}{2}-f\left(\frac{m+M}{2}\right)\right]\le\frac{1}{2}\max\left\{M-\sum_{i=1}^{n}w_ix_i,\sum_{i=1}^{n}w_ix_i-m\right\}\left[f'_-(M)-f'_+(m)\right]. \tag{1.9}$$

Remark 3. *Since, obviously,*

$$\frac{M-\sum_{i=1}^{n}w_ix_i}{M-m},\frac{\sum_{i=1}^{n}w_ix_i-m}{M-m}\le 1,$$

then we obtain from the first inequality in (1.9) the simpler, however coarser inequality, namely

$$0\le\sum_{i=1}^{n}w_if(x_i)-f\left(\sum_{i=1}^{n}w_ix_i\right)\le 2\left[\frac{f(m)+f(M)}{2}-f\left(\frac{m+M}{2}\right)\right]. \tag{1.10}$$

This inequality was obtained in 2008 by S. Simić in [31].

The following result also holds:

Theorem 4 (Dragomir, 2013 [24])**.** *Let $\Phi:I\to\mathbb{R}$ be a continuous convex function on the interval of real numbers I and m, $M\in\mathbb{R}$, $m<M$ with $[m,M]\subset\mathring{I}$, $\mathring{I}$ is the interior of I. If $x_i\in I$ and $w_i\ge 0$ for $i\in\{1,...,n\}$ with $\sum_{i=1}^{n}w_i=1$, denote $\bar{x}_w:=\sum_{i=1}^{n}w_ix_i\in I$, then we have the inequality*

$$0\le\sum_{i=1}^{n}w_i\Phi(x_i)-\Phi(\bar{x}_w)\le\frac{(M-\bar{x}_w)\int_m^{\bar{x}_w}|\Phi'(t)|\,dt+(\bar{x}_w-m)\int_{\bar{x}_w}^{M}|\Phi'(t)|\,dt}{M-m}:=\Theta_\Phi(\bar{x}_w;m,M), \tag{1.11}$$

where $\Theta_\Phi(\bar{x}_w;m,M)$ satisfies the bounds

$$\Theta_\Phi(\bar{x}_w;m,M)\le\begin{cases}\left[\frac{1}{2}+\frac{\left|\bar{x}_w-\frac{m+M}{2}\right|}{M-m}\right]\int_m^M|\Phi'(t)|\,dt\\[2ex]\left[\frac{1}{2}\int_m^M|\Phi'(t)|\,dt+\frac{1}{2}\left|\int_{\bar{x}_w}^M|\Phi'(t)|\,dt-\int_m^{\bar{x}_w}|\Phi'(t)|\,dt\right|\right],\end{cases} \tag{1.12}$$

$$\Theta_\Phi(\bar{x}_w;m,M)\le\frac{(\bar{x}_w-m)(M-\bar{x}_w)}{M-m}\left[\|\Phi'\|_{[\bar{x}_w,M],\infty}+\|\Phi'\|_{[m,\bar{x}_w],\infty}\right]\le\frac{1}{2}(M-m)\frac{\|\Phi'\|_{[\bar{w}_p,M],\infty}+\|\Phi'\|_{[m,\bar{w}_p],\infty}}{2}\le\frac{1}{2}(M-m)\|\Phi'\|_{[m,M],\infty} \tag{1.13}$$

and

$$\begin{aligned}(1.14)\quad &\Theta_{\Phi}\left(\bar{x}_w; m, M\right) \\ &\leq \frac{1}{M-m}\Big[\left(\bar{x}_w - m\right)\left(M - \bar{x}_w\right)^{1/q}\left\|\Phi'\right\|_{[\bar{x}_w, M], p} \\ &+ \left(M - \bar{x}_w\right)\left(\bar{x}_w - m\right)^{1/q}\left\|\Phi'\right\|_{[m, \bar{x}_w], p}\Big] \\ &\leq \frac{1}{M-m}\left[\left(\bar{x}_w - m\right)^q\left(M - \bar{x}_w\right) + \left(M - \bar{x}_w\right)^q\left(\bar{x}_w - m\right)\right]^{1/q}\left\|\Phi'\right\|_{[m, M], p}\end{aligned}$$

where $p > 1$, $\frac{1}{p} + \frac{1}{q} = 1$.

For a real function $g : [m, M] \to \mathbb{R}$ and two distinct points α, $\beta \in [m, M]$ we recall that the *divided difference* of g in these points is defined by

$$[\alpha, \beta; g] := \frac{g(\beta) - g(\alpha)}{\beta - \alpha}.$$

Theorem 5 (Dragomir, 2011 [22]). *Let $f : I \to \mathbb{R}$ be a continuous convex function on the interval of real numbers I and m, $M \in \mathbb{R}$, $m < M$ with $[m, M] \subset \mathring{I}$, $\mathring{I}$ the interior of I. Let $\bar{\mathbf{a}} = (a_1, \ldots, a_n)$, $\bar{\mathbf{p}} = (p_1, \ldots, p_n)$ be n-tuples of real numbers with $p_i \geq 0$ $(i \in \{1, \ldots, n\})$ and $\sum_{i=1}^{n} p_i = 1$. If $m \leq a_i \leq M$, $i \in \{1, \ldots, n\}$, with $\sum_{i=1}^{n} p_i a_i \neq m$, M, then*

$$\begin{aligned}(1.15)\quad &\left|\sum_{i=1}^{n} p_i \left|f(a_i) - f\left(\sum_{j=1}^{n} p_j a_j\right)\right| sgn\left(a_i - \sum_{j=1}^{n} p_j a_j\right)\right| \\ &\leq \sum_{i=1}^{n} p_i f(a_i) - f\left(\sum_{i=1}^{n} p_i a_i\right) \\ &\leq \frac{1}{2}\left(\left[\sum_{i=1}^{n} p_i a_i, M; f\right] - \left[m, \sum_{i=1}^{n} p_i a_i; f\right]\right)\sum_{i=1}^{n} p_i \left|a_i - \sum_{j=1}^{n} p_j a_j\right| \\ &\leq \frac{1}{2}\left(\left[\sum_{i=1}^{n} p_i a_i, M; f\right] - \left[m, \sum_{i=1}^{n} p_i a_i; f\right]\right) \\ &\times \left[\sum_{i=1}^{n} p_i a_i^2 - \left(\sum_{j=1}^{n} p_j a_j\right)^2\right]^{1/2}.\end{aligned}$$

If the lateral derivatives $f'_{+}(m)$ and $f'_{-}(M)$ are finite, then we also have the inequalities

$$0 \le \sum_{i=1}^{n} p_i f(a_i) - f\left(\sum_{i=1}^{n} p_i a_i\right) \tag{1.16}$$
$$\le \frac{1}{2}\left(\left[\sum_{i=1}^{n} p_i a_i, M; f\right] - \left[m, \sum_{i=1}^{n} p_i a_i; f\right]\right) \sum_{i=1}^{n} p_i \left| a_i - \sum_{j=1}^{n} p_j a_j \right|$$
$$\le \frac{1}{2}\left[f'_{-}(M) - f'_{+}(m)\right] \sum_{i=1}^{n} p_i \left| a_i - \sum_{j=1}^{n} p_j a_j \right|$$
$$\le \frac{1}{2}\left[f'_{-}(M) - f'_{+}(m)\right] \left[\sum_{i=1}^{n} p_i a_i^2 - \left(\sum_{j=1}^{n} p_j a_j\right)^2\right]^{1/2}.$$

In this paper we survey some recent discrete inequalities for functions defined on convex subsets of general linear spaces. Various refinements and reverses of Jensen's discrete inequality are presented. The Slater inequality version for these functions is outlined. As applications, we establish several bounds for the *mean f-deviation* of an n-tuple of vectors as well as for the *f-divergence* of an n-tuple of vectors given a discrete probability distribution. Examples for the K. Pearson *χ^2-divergence,* the *Kullback-Leibler divergence,* the *Jeffreys divergence,* the *total variation distance* and other divergence measures are also provided.

2. Refinements of Jensen's Inequality

2.1. **Preliminary Facts.** Let C be a convex subset of the linear space X and f a convex function on C. If $\mathbf{p} = (p_1, \dots, p_n)$ is a probability sequence and $\mathbf{x} = (x_1, \dots, x_n) \in C^n$, then

$$f\left(\sum_{i=1}^{n} p_i x_i\right) \le \sum_{i=1}^{n} p_i f(x_i), \tag{2.1}$$

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

In 1989, J. Pečarić and the author [30] obtained the following refinement of (2.1):

$$f\left(\sum_{i=1}^{n} p_i x_i\right) \le \sum_{i_1,\dots,i_{k+1}=1}^{n} p_{i_1} \dots p_{i_{k+1}} f\left(\frac{x_{i_1} + \dots + x_{i_{k+1}}}{k+1}\right) \tag{2.2}$$
$$\le \sum_{i_1,\dots,i_k=1}^{n} p_{i_1} \dots p_{i_k} f\left(\frac{x_{i_1} + \dots + x_{i_k}}{k}\right)$$
$$\le \dots \le \sum_{i=1}^{n} p_i f(x_i),$$

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

If $q_1, \dots, q_k \geq 0$ with $\sum_{j=1}^{k} q_j = 1$, then the following refinement obtained in 1994 by the author [2] also holds:

$$f\left(\sum_{i=1}^{n} p_i x_i\right) \leq \sum_{i_1,\dots,i_k=1}^{n} p_{i_1} \dots p_{i_k} f\left(\frac{x_{i_1} + \dots + x_{i_k}}{k}\right) \leq \sum_{i_1,\dots,i_k=1}^{n} p_{i_1} \dots p_{i_k} f\left(q_1 x_{i_1} + \dots + q_k x_{i_k}\right) \leq \sum_{i=1}^{n} p_i f(x_i), \tag{2.3}$$

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

For other refinements and applications related to Ky Fan's inequality, the arithmetic mean-geometric mean inequality, the generalized triangle inequality etc., see [3]-[29].

2.2. General Results.

The following result may be stated.

Theorem 6 (Dragomir, 2010 [19]). *Let $f : C \to \mathbb{R}$ be a convex function on the convex subset C of the linear space X, $x_i \in C$, $p_i > 0$, $i \in \{1, \dots, n\}$ with $\sum_{i=1}^{n} p_i = 1$. Then*

$$\begin{aligned} f\left(\sum_{j=1}^{n} p_j x_j\right) &\leq \min_{k\in\{1,\dots,n\}} \left[(1-p_k) f\left(\frac{\sum_{j=1}^{n} p_j x_j - p_k x_k}{1-p_k}\right) + p_k f(x_k)\right] \\ &\leq \frac{1}{n}\left[\sum_{k=1}^{n} (1-p_k) f\left(\frac{\sum_{j=1}^{n} p_j x_j - p_k x_k}{1-p_k}\right) + \sum_{k=1}^{n} p_k f(x_k)\right] \\ &\leq \max_{k\in\{1,\dots,n\}} \left[(1-p_k) f\left(\frac{\sum_{j=1}^{n} p_j x_j - p_k x_k}{1-p_k}\right) + p_k f(x_k)\right] \\ &\leq \sum_{j=1}^{n} p_j f(x_j). \end{aligned} \tag{2.4}$$

In particular,

$$\begin{aligned} f\left(\frac{1}{n}\sum_{j=1}^{n} x_j\right) &\leq \frac{1}{n} \min_{k\in\{1,\dots,n\}} \left[(n-1) f\left(\frac{\sum_{j=1}^{n} x_j - x_k}{n-1}\right) + f(x_k)\right] \\ &\leq \frac{1}{n^2}\left[(n-1)\sum_{k=1}^{n} f\left(\frac{\sum_{j=1}^{n} x_j - x_k}{n-1}\right) + \sum_{k=1}^{n} f(x_k)\right] \end{aligned} \tag{2.5}$$

$$\begin{aligned} &\leq \frac{1}{n} \max_{k\in\{1,\dots,n\}} \left[(n-1) f\left(\frac{\sum_{j=1}^{n} x_j - x_k}{n-1}\right) + f(x_k)\right] \\ &\leq \frac{1}{n}\sum_{j=1}^{n} f(x_j). \end{aligned}$$

Proof. For any $k \in \{1, \dots, n\}$, we have

$$\sum_{j=1}^{n} p_j x_j - p_k x_k = \sum_{\substack{j=1 \\ j \neq k}}^{n} p_j x_j = \frac{\sum_{\substack{j=1 \\ j \neq k}}^{n} p_j}{\sum_{\substack{j=1 \\ j \neq k}}^{n} p_j} \sum_{\substack{j=1 \\ j \neq k}}^{n} p_j x_j = (1 - p_k) \cdot \frac{1}{\sum_{\substack{j=1 \\ j \neq k}}^{n} p_j} \sum_{\substack{j=1 \\ j \neq k}}^{n} p_j x_j,$$

which implies that

$$\frac{\sum_{j=1}^{n} p_j x_j - p_k x_k}{1 - p_k} = \frac{1}{\sum_{\substack{j=1 \\ j \neq k}}^{n} p_j} \sum_{\substack{j=1 \\ j \neq k}}^{n} p_j x_j \in C \tag{2.6}$$

for each $k \in \{1, \dots, n\}$, since the right side of (2.6) is a convex combination of the elements $x_j \in C$, $j \in \{1, \dots, n\} \setminus \{k\}$.

Taking the function f on (2.6) and applying the Jensen inequality, we get successively

$$f\left(\frac{\sum_{j=1}^{n} p_j x_j - p_k x_k}{1 - p_k}\right) = f\left(\frac{1}{\sum_{\substack{j=1 \\ j \neq k}}^{n} p_j} \sum_{\substack{j=1 \\ j \neq k}}^{n} p_j x_j\right) \leq \frac{1}{\sum_{\substack{j=1 \\ j \neq k}}^{n} p_j} \sum_{\substack{j=1 \\ j \neq k}}^{n} p_j f(x_j)$$

$$= \frac{1}{1 - p_k} \left[\sum_{j=1}^{n} p_j f(x_j) - p_k f(x_k)\right]$$

for any $k \in \{1, \dots, n\}$, which implies

$$(1 - p_k) f\left(\frac{\sum_{j=1}^{n} p_j x_j - p_k x_k}{1 - p_k}\right) + p_k f(x_k) \leq \sum_{j=1}^{n} p_j f(x_j) \tag{2.7}$$

for each $k \in \{1, \dots, n\}$.

Utilising the convexity of f, we also have

$$(1 - p_k) f\left(\frac{\sum_{j=1}^{n} p_j x_j - p_k x_k}{1 - p_k}\right) + p_k f(x_k) \tag{2.8}$$

$$\geq f\left[(1 - p_k) \cdot \frac{\sum_{j=1}^{n} p_j x_j - p_k x_k}{1 - p_k} + p_k x_k\right] = f\left(\sum_{j=1}^{n} p_j x_j\right)$$

for each $k \in \{1, \dots, n\}$.

Taking the minimum over k in (2.8), utilising the fact that

$$\min_{k \in \{1, \dots, n\}} \alpha_k \leq \frac{1}{n} \sum_{k=1}^{n} \alpha_k \leq \max_{k \in \{1, \dots, n\}} \alpha_k$$

and then taking the maximum in (2.7), we deduce the desired inequality (2.4). □

After setting $x_j = y_j - \sum_{l=1}^{n} q_l y_l$ and $p_j = q_j$, $j \in \{1, \dots, n\}$, Theorem 6 becomes the following corollary:

Corollary 1 (Dragomir, 2010 [19]). *Let $f : C \to \mathbb{R}$ be a convex function on the convex subset C, $0 \in C$, $y_j \in X$ and $q_j > 0$, $j \in \{1, \ldots, n\}$ with $\sum_{j=1}^{n} q_j = 1$. If $y_j - \sum_{l=1}^{n} q_l y_l \in C$ for any $j \in \{1, \ldots, n\}$, then*

$$
\begin{aligned}
(2.9)\quad & f(0) \\
& \leq \min_{k\in\{1,\ldots,n\}} \left\{ (1-q_k) f\left[\frac{q_k}{1-q_k} \left(\sum_{l=1}^{n} q_l y_l - y_k \right) \right] + q_k f\left(y_k - \sum_{l=1}^{n} q_l y_l \right) \right\} \\
& \leq \frac{1}{n} \left\{ \sum_{l=1}^{n} (1-q_k) f\left[\frac{q_k}{1-q_k} \left(\sum_{l=1}^{n} q_l y_l - y_k \right) \right] + \sum_{l=1}^{n} q_k f\left(y_k - \sum_{l=1}^{n} q_l y_l \right) \right\} \\
& \leq \max_{k\in\{1,\ldots,n\}} \left\{ (1-q_k) f\left[\frac{q_k}{1-q_k} \left(\sum_{l=1}^{n} q_l y_l - y_k \right) \right] + q_k f\left(y_k - \sum_{l=1}^{n} q_l y_l \right) \right\} \\
& \leq \sum_{j=1}^{n} q_j f\left(y_j - \sum_{l=1}^{n} q_l y_l \right).
\end{aligned}
$$

In particular, if $y_j - \frac{1}{n} \sum_{l=1}^{n} y_l \in C$ for any $j \in \{1, \ldots, n\}$, then

$$
\begin{aligned}
(2.10)\quad & f(0) \\
& \leq \frac{1}{n} \min_{k\in\{1,\ldots,n\}} \left\{ (n-1) f\left[\frac{1}{n-1} \left(\frac{1}{n} \sum_{l=1}^{n} y_l - y_k \right) \right] + f\left(y_k - \frac{1}{n} \sum_{l=1}^{n} y_l \right) \right\} \\
& \leq \frac{1}{n^2} \left\{ (n-1) \sum_{k=1}^{n} f\left[\frac{1}{n-1} \left(\frac{1}{n} \sum_{l=1}^{n} y_l - y_k \right) \right] + \sum_{k=1}^{n} f\left(y_k - \frac{1}{n} \sum_{l=1}^{n} y_l \right) \right\} \\
& \leq \frac{1}{n} \max_{k\in\{1,\ldots,n\}} \left\{ (n-1) f\left[\frac{1}{n-1} \left(\frac{1}{n} \sum_{l=1}^{n} y_l - y_k \right) \right] + f\left(y_k - \frac{1}{n} \sum_{l=1}^{n} y_l \right) \right\} \\
& \leq \frac{1}{n} \sum_{j=1}^{n} f\left(y_j - \frac{1}{n} \sum_{l=1}^{n} y_l \right).
\end{aligned}
$$

The above results can be applied for various convex functions related to celebrated inequalities as mentioned in the introduction.

Application 1. If $(X, \|\cdot\|)$ is a normed linear space and $p \geq 1$, then the function $f : X \to \mathbb{R}$, $f(x) = \|x\|^p$ is convex on X. Now, on applying Theorem 6 and Corollary

1 for $x_i \in X$, $p_i > 0$, $i \in \{1, \dots, n\}$ with $\sum_{i=1}^n p_i = 1$, we get:

$$\begin{aligned}\left\|\sum_{j=1}^n p_j x_j\right\|^p &\leq \min_{k\in\{1,\dots,n\}}\left[(1-p_k)^{1-p}\left\|\sum_{j=1}^n p_j x_j - p_k x_k\right\|^p + p_k\|x_k\|^p\right] \\ &\leq \frac{1}{n}\left[\sum_{k=1}^n (1-p_k)^{1-p}\left\|\sum_{j=1}^n p_j x_j - p_k x_k\right\|^p + \sum_{k=1}^n p_k\|x_k\|^p\right] \\ &\leq \max_{k\in\{1,\dots,n\}}\left[(1-p_k)^{1-p}\left\|\sum_{j=1}^n p_j x_j - p_k x_k\right\|^p + p_k\|x_k\|^p\right] \\ &\leq \sum_{j=1}^n p_j\|x_j\|^p \end{aligned} \tag{2.11}$$

and

$$\max_{k\in\{1,\dots,n\}}\left\{\left[(1-p_k)^{1-p}p_k^p + p_k\right]\left\|x_k - \sum_{l=1}^n p_l x_l\right\|^p\right\} \leq \sum_{j=1}^n p_j\left\|x_k - \sum_{l=1}^n p_l x_l\right\|^p. \tag{2.12}$$

In particular, we have the inequality:

$$\begin{aligned}\left\|\frac{1}{n}\sum_{j=1}^n x_j\right\|^p &\leq \frac{1}{n}\min_{k\in\{1,\dots,n\}}\left[(n-1)^{1-p}\left\|\sum_{j=1}^n x_j - x_k\right\|^p + \|x_k\|^p\right] \\ &\leq \frac{1}{n^2}\left[(n-1)^{1-p}\sum_{k=1}^n\left\|\sum_{j=1}^n x_j - x_k\right\|^p + \sum_{k=1}^n\|x_k\|^p\right] \\ &\leq \frac{1}{n}\max_{k\in\{1,\dots,n\}}\left[(n-1)^{1-p}\left\|\sum_{j=1}^n x_j - x_k\right\|^p + \|x_k\|^p\right] \\ &\leq \frac{1}{n}\sum_{j=1}^n\|x_j\|^p\end{aligned} \tag{2.13}$$

and

$$\left[(n-1)^{1-p}+1\right]\max_{k\in\{1,\dots,n\}}\left\|x_k - \frac{1}{n}\sum_{l=1}^n x_l\right\|^p \leq \sum_{j=1}^n\left\|x_j - \frac{1}{n}\sum_{l=1}^n x_l\right\|^p. \tag{2.14}$$

If we consider the function $h_p(t) := (1-t)^{1-p}t^p + t$, $p \geq 1$, $t \in [0,1)$, then we observe that

$$h_p'(t) = 1 + pt^{p-1}(1-t)^{1-p} + (p-1)t^p(1-t)^{-p},$$

which shows that h_p is strictly increasing on $[0,1)$. Therefore,

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

where $p_m := \min_{k\in\{1,\dots,n\}} p_k$. By (2.12), we then obtain the following inequality:

$$\left[p_m + (1-p_m)^{1-p} \cdot p_m^p\right] \max_{k\in\{1,\dots,n\}} \left\|x_k - \sum_{l=1}^n p_l x_l\right\|^p \leq \sum_{j=1}^n p_j \left\|x_j - \sum_{l=1}^n p_l x_l\right\|^p . \tag{2.15}$$

Application 2. Let $x_i, p_i > 0$, $i \in \{1,\dots,n\}$ with $\sum_{i=1}^n p_i = 1$. The following inequality is well known in the literature as the *arithmetic mean-geometric mean* inequality:

$$\sum_{j=1}^n p_j x_j \geq \prod_{j=1}^n x_j^{p_j}. \tag{2.16}$$

The equality case holds in (2.16) iff $x_1 = \cdots = x_n$.

Applying the inequality (2.4) for the convex function $f : (0,\infty) \to \mathbb{R}$, $f(x) = -\ln x$ and performing the necessary computations, we derive the following refinement of (2.16):

$$\begin{aligned} \sum_{i=1}^n p_i x_i &\geq \max_{k\in\{1,\dots,n\}} \left\{ \left(\frac{\sum_{j=1}^n p_j x_j - p_k x_k}{1-p_k}\right)^{1-p_k} \cdot x_k^{p_k} \right\} \\ &\geq \prod_{k=1}^n \left[\left(\frac{\sum_{j=1}^n p_j x_j - p_k x_k}{1-p_k}\right)^{1-p_k} \cdot x_k^{p_k} \right]^{\frac{1}{n}} \\ &\geq \min_{k\in\{1,\dots,n\}} \left\{ \left(\frac{\sum_{j=1}^n p_j x_j - p_k x_k}{1-p_k}\right)^{1-p_k} \cdot x_k^{p_k} \right\} \geq \prod_{i=1}^n x_i^{p_i}. \end{aligned} \tag{2.17}$$

In particular, we have the inequality:

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^n x_i &\geq \max_{k\in\{1,\dots,n\}} \left\{ \left(\frac{\sum_{j=1}^n x_j - x_k}{n-1}\right)^{\frac{n-1}{n}} \cdot x_k^{\frac{1}{n}} \right\} \\ &\geq \prod_{k=1}^n \left[\left(\frac{\sum_{j=1}^n x_j - x_k}{n-1}\right)^{\frac{n-1}{n}} \cdot x_k^{\frac{1}{n}} \right]^{\frac{1}{n}} \\ &\geq \min_{k\in\{1,\dots,n\}} \left\{ \left(\frac{\sum_{j=1}^n x_j - x_k}{n-1}\right)^{\frac{n-1}{n}} \cdot x_k^{\frac{1}{n}} \right\} \geq \left(\prod_{i=1}^n x_i\right)^{\frac{1}{n}}. \end{aligned}$$

2.3. **Applications for f-Divergences.** The following refinement of the positivity property of f-divergence may be stated.

Theorem 7 (Dragomir, 2010 [19]). *For any* $\mathbf{p}$, $\mathbf{q} \in \mathbb{P}^n$, *namely* $\mathbf{p}$, $\mathbf{q}$ *are probability distributions, we have the inequalities*

$$\begin{aligned} I_f(\mathbf{p},\mathbf{q}) &\geq \max_{k\in\{1,\dots,n\}}\left[(1-q_k) f\left(\frac{1-p_k}{1-q_k}\right)+q_k f\left(\frac{p_k}{q_k}\right)\right] \\ &\geq \frac{1}{n}\left[\sum_{k=1}^{n}(1-q_k) f\left(\frac{1-p_k}{1-q_k}\right)+\sum_{k=1}^{n} q_k f\left(\frac{p_k}{q_k}\right)\right] \\ &\geq \min_{k\in\{1,\dots,n\}}\left[(1-q_k) f\left(\frac{1-p_k}{1-q_k}\right)+q_k f\left(\frac{p_k}{q_k}\right)\right] \geq 0, \end{aligned} \tag{2.18}$$

provided $f:[0,\infty)\to\mathbb{R}$ *is convex and normalized on* $[0,\infty)$.

The proof is obvious by Theorem 6 applied for the convex function $f:[0,\infty)\to\mathbb{R}$ and for the choice $x_i=\frac{p_i}{q_i}$, $i\in\{1,\dots,n\}$ and the probabilities q_i, $i\in\{1,\dots,n\}$.

If we consider a new divergence measure $R_f(\mathbf{p},\mathbf{q})$ defined for $\mathbf{p},\mathbf{q}\in\mathbb{P}^n$ by

$$R_f(\mathbf{p},\mathbf{q}) := \frac{1}{n-1}\sum_{k=1}^{n}(1-q_k) f\left(\frac{1-p_k}{1-q_k}\right) \tag{2.19}$$

and call it the *reverse* f*-divergence*, we observe that

$$R_f(\mathbf{p},\mathbf{q}) = I_f(\mathbf{r},\mathbf{t}) \tag{2.20}$$

with

$$\mathbf{r}=\left(\frac{1-p_1}{n-1},\dots,\frac{1-p_n}{n-1}\right),\quad \mathbf{t}=\left(\frac{1-q_1}{n-1},\dots,\frac{1-q_n}{n-1}\right)\qquad (n\geq 2).$$

With this notation, we can state the following corollary of the above proposition.

Corollary 2. *For any* $\mathbf{p}$, $\mathbf{q}\in\mathbb{P}^n$, *we have*

$$I_f(\mathbf{p},\mathbf{q}) \geq R_f(\mathbf{p},\mathbf{q}) \geq 0. \tag{2.21}$$

The proof is obvious by the second inequality in (2.18) and the details are omitted.

In what follows, we point out some particular inequalities for various instances of divergence measures such as: the *total variation distance*, χ^2*-divergence*, *Kullback-Leibler divergence*, *Jeffreys divergence*.

The *total variation distance* is defined by the convex function $f(t)=|t-1|$, $t\in\mathbb{R}$ and given in:

$$V(p,q) := \sum_{j=1}^{n} q_j\left|\frac{p_j}{q_j}-1\right| = \sum_{j=1}^{n}|p_j-q_j|. \tag{2.22}$$

The following improvement of the positivity inequality for the total variation distance can be stated as follows.

Proposition 1. *For any* $\mathbf{p}$, $\mathbf{q}\in\mathbb{P}^n$, *we have the inequality:*

$$V(p,q) \geq 2\max_{k\in\{1,\dots,n\}}|p_k-q_k| \quad (\geq 0). \tag{2.23}$$

The proof follows by the first inequality in (2.18) for $f(t)=|t-1|$, $t\in\mathbb{R}$.

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

$$\chi^2(p,q) := \sum_{j=1}^{n} q_j \left(\frac{p_j}{q_j} - 1\right)^2 = \sum_{j=1}^{n} \frac{(p_j - q_j)^2}{q_j}. \tag{2.24}$$

Proposition 2. *For any* $\mathbf{p}$, $\mathbf{q} \in \mathbb{P}^n$,

$$\chi^2(p,q) \geq \max_{k\in\{1,\dots,n\}} \left\{\frac{(p_k - q_k)^2}{q_k(1-q_k)}\right\} \geq 4 \max_{k\in\{1,\dots,n\}} (p_k - q_k)^2 \quad (\geq 0). \tag{2.25}$$

Proof. On applying the first inequality in (2.18) for the function $f(t) = (1-t)^2$, $t \in \mathbb{R}$, we get

$$\begin{aligned}\chi^2(p,q) &\geq \max_{k\in\{1,\dots,n\}} \left\{(1-q_k)\left(\frac{1-p_k}{1-q_k} - 1\right)^2 + q_k\left(\frac{p_k}{q_k} - 1\right)^2\right\} \\ &= \max_{k\in\{1,\dots,n\}} \left\{\frac{(p_k - q_k)^2}{q_k(1-q_k)}\right\}.\end{aligned}$$

Since

$$q_k(1-q_k) \leq \frac{1}{4}[q_k + (1-q_k)]^2 = \frac{1}{4},$$

then

$$\frac{(p_k - q_k)^2}{q_k(1-q_k)} \geq 4(p_k - q_k)^2$$

for each $k \in \{1, \dots, n\}$, which proves the last part of (2.25). □

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

$$KL(p,q) := \sum_{j=1}^{n} q_j \cdot \frac{p_j}{q_j} \ln\left(\frac{p_j}{q_j}\right) = \sum_{j=1}^{n} p_j \ln\left(\frac{p_j}{q_j}\right). \tag{2.26}$$

Proposition 3. *For any* $\mathbf{p}$, $\mathbf{q} \in \mathbb{P}^n$, *we have:*

$$KL(p,q) \geq \ln\left[\max_{k\in\{1,\dots,n\}} \left\{\left(\frac{1-p_k}{1-q_k}\right)^{1-p_k} \cdot \left(\frac{p_k}{q_k}\right)^{p_k}\right\}\right] \geq 0. \tag{2.27}$$

Proof. The first inequality is obvious by Theorem 7. Utilising the inequality between the *geometric mean and the harmonic mean*,

$$x^{\alpha} y^{1-\alpha} \geq \frac{1}{\frac{\alpha}{x} + \frac{1-\alpha}{y}}, \qquad x, y > 0,\ \alpha \in [0,1]$$

we have

$$\left(\frac{1-p_k}{1-q_k}\right)^{1-p_k} \cdot \left(\frac{p_k}{q_k}\right)^{p_k} \geq 1,$$

for any $k \in \{1, \dots, n\}$, which implies the second part of (2.27). □

Another divergence measure that is of importance in Information Theory is the *Jeffreys divergence*

$$J(p,q) := \sum_{j=1}^{n} q_j \cdot \left(\frac{p_j}{q_j} - 1\right) \ln\left(\frac{p_j}{q_j}\right) = \sum_{j=1}^{n} (p_j - q_j) \ln\left(\frac{p_j}{q_j}\right), \tag{2.28}$$

which is an f-divergence for $f(t)=(t-1)\ln t$, $t>0$.

Proposition 4. *For any* $\mathbf{p}$, $\mathbf{q}\in\mathbb{P}^n$, *we have:*

$$J(p,q)\geq\max_{k\in\{1,\dots,n\}}\left\{(q_k-p_k)\ln\left[\frac{(1-p_k)\,q_k}{(1-q_k)\,p_k}\right]\right\}\\ \geq\max_{k\in\{1,\dots,n\}}\left[\frac{(q_k-p_k)^2}{p_k+q_k-2p_kq_k}\right]\geq 0. \tag{2.29}$$

Proof. Writing the first inequality in Theorem 7 for $f(t)=(t-1)\ln t$, we have

$$\begin{aligned}J(p,q)&\geq\max_{k\in\{1,\dots,n\}}\left\{(1-q_k)\left[\left(\frac{1-p_k}{1-q_k}-1\right)\ln\left(\frac{1-p_k}{1-q_k}\right)\right]+q_k\left(\frac{p_k}{q_k}-1\right)\ln\left(\frac{p_k}{q_k}\right)\right\}\\&=\max_{k\in\{1,\dots,n\}}\left\{(q_k-p_k)\ln\left(\frac{1-p_k}{1-q_k}\right)-(q_k-p_k)\ln\left(\frac{p_k}{q_k}\right)\right\}\\&=\max_{k\in\{1,\dots,n\}}\left\{(q_k-p_k)\ln\left[\frac{(1-p_k)\,q_k}{(1-q_k)\,p_k}\right]\right\},\end{aligned}$$

proving the first inequality in (2.29).

Utilising the elementary inequality for positive numbers,

$$\frac{\ln b-\ln a}{b-a}\geq\frac{2}{a+b},\qquad a,b>0$$

we have

$$\begin{aligned}&(q_k-p_k)\left[\ln\left(\frac{1-p_k}{1-q_k}\right)-\ln\left(\frac{p_k}{q_k}\right)\right]\\&=(q_k-p_k)\cdot\frac{\ln\left(\frac{1-p_k}{1-q_k}\right)-\ln\left(\frac{p_k}{q_k}\right)}{\frac{1-p_k}{1-q_k}-\frac{p_k}{q_k}}\cdot\left[\frac{1-p_k}{1-q_k}-\frac{p_k}{q_k}\right]\\&=\frac{(q_k-p_k)^2}{q_k(1-q_k)}\cdot\frac{\ln\left(\frac{1-p_k}{1-q_k}\right)-\ln\left(\frac{p_k}{q_k}\right)}{\frac{1-p_k}{1-q_k}-\frac{p_k}{q_k}}\\&\geq\frac{(q_k-p_k)^2}{q_k(1-q_k)}\cdot\frac{2}{\frac{1-p_k}{1-q_k}+\frac{p_k}{q_k}}=\frac{2\,(q_k-p_k)^2}{p_k+q_k-2p_kq_k}\geq 0,\end{aligned}$$

for each $k\in\{1,\dots,n\}$, giving the second inequality in (2.29). □

2.4. **More General Results.** Let C be a convex subset in the real linear space X and assume that $f:C\to\mathbb{R}$ is a convex function on C. If $x_i\in C$ and $p_i>0, i\in\{1,...,n\}$ with $\sum_{i=1}^n p_i=1$, then for any nonempty subset J of $\{1,...,n\}$ we put $\bar{J}:=\{1,...,n\}\setminus J\,(\neq\emptyset)$ and define $P_J:=\sum_{i\in J}p_i$ and $\bar{P}_J:=P_{\bar{J}}=\sum_{j\in\bar{J}}p_j=1-\sum_{i\in J}p_i$. For the convex function f and the n-tuples $\mathbf{x}:=(x_1,...,x_n)$ and $\mathbf{p}:=(p_1,...,p_n)$ as above, we can define the following functional

$$D(f,\mathbf{p},\mathbf{x};J):=P_Jf\left(\frac{1}{P_J}\sum_{i\in J}p_ix_i\right)+\bar{P}_Jf\left(\frac{1}{\bar{P}_J}\sum_{j\in\bar{J}}p_jx_j\right) \tag{2.30}$$

where here and everywhere below $J\subset\{1,...,n\}$ with $J\neq\emptyset$ and $J\neq\{1,...,n\}$.

It is worth to observe that for $J = \{k\}, k \in \{1, ..., n\}$ we have the functional

$$D_k(f, \mathbf{p}, \mathbf{x}) := D(f, \mathbf{p}, \mathbf{x}; \{k\}) = p_k f(x_k) + (1 - p_k) f\left(\frac{\sum_{i=1}^n p_i x_i - p_k x_k}{1 - p_k}\right) \tag{2.31}$$

that has been investigated in the paper [19].

Theorem 8 (Dragomir, 2010 [20])**.** *Let C be a convex subset in the real linear space X and assume that $f : C \to \mathbb{R}$ is a convex function on C. If $x_i \in C$ and $p_i > 0, i \in \{1, ..., n\}$ with $\sum_{i=1}^n p_i = 1$ then for any nonempty subset J of $\{1, ..., n\}$ we have*

$$\sum_{k=1}^n p_k f(x_k) \geq D(f, \mathbf{p}, \mathbf{x}; J) \geq f\left(\sum_{k=1}^n p_k x_k\right). \tag{2.32}$$

Proof. By the convexity of the function f we have

$$\begin{aligned} D(f, \mathbf{p}, \mathbf{x}; J) &= P_J f\left(\frac{1}{P_J}\sum_{i\in J} p_i x_i\right) + \bar{P}_J f\left(\frac{1}{\bar{P}_J}\sum_{j\in \bar{J}} p_j x_j\right) \\ &\geq f\left[P_J\left(\frac{1}{P_J}\sum_{i\in J} p_i x_i\right) + \bar{P}_J\left(\frac{1}{\bar{P}_J}\sum_{j\in \bar{J}} p_j x_j\right)\right] \\ &= f\left(\sum_{k=1}^n p_k x_k\right) \end{aligned}$$

for any J, which proves the second inequality in (2.32).

By the Jensen inequality we also have

$$\begin{aligned} \sum_{k=1}^n p_k f(x_k) &= \sum_{i\in J} p_i f(x_i) + \sum_{j\in \bar{J}} p_j f(x_j) \\ &\geq P_J f\left(\frac{1}{P_J}\sum_{i\in J} p_i x_i\right) + \bar{P}_J f\left(\frac{1}{\bar{P}_J}\sum_{j\in \bar{J}} p_j x_j\right) \\ &= D(f, \mathbf{p}, \mathbf{x}; J) \end{aligned}$$

for any J, which proves the first inequality in (2.32). □

Remark 4. *We observe that the inequality (2.32) can be written in an equivalent form as*

$$\sum_{k=1}^n p_k f(x_k) \geq \max_{\emptyset \neq J \subset \{1,...,n\}} D(f, \mathbf{p}, \mathbf{x}; J) \tag{2.33}$$

and

$$\min_{\emptyset \neq J \subset \{1,...,n\}} D(f, \mathbf{p}, \mathbf{x}; J) \geq f\left(\sum_{k=1}^n p_k x_k\right). \tag{2.34}$$

These inequalities imply the following results that have been obtained earlier by the author in [19] *utilising a different method of proof slightly more complicated:*

$$\sum_{k=1}^{n} p_k f(x_k) \geq \max_{k\in\{1,...,n\}} D_k(f,\mathbf{p},\mathbf{x}) \tag{2.35}$$

and

$$\min_{k\in\{1,...,n\}} D_k(f,\mathbf{p},\mathbf{x}) \geq f\left(\sum_{k=1}^{n} p_k x_k\right). \tag{2.36}$$

Moreover, since

$$\max_{\emptyset\neq J\subset\{1,...,n\}} D(f,\mathbf{p},\mathbf{x};J) \geq \max_{k\in\{1,...,n\}} D_k(f,\mathbf{p},\mathbf{x})$$

and

$$\min_{k\in\{1,...,n\}} D_k(f,\mathbf{p},\mathbf{x}) \geq \min_{\emptyset\neq J\subset\{1,...,n\}} D(f,\mathbf{p},\mathbf{x};J),$$

then the new inequalities (2.33) and (2.33) are better than the earlier results developed in [19]*.*

The case of uniform distribution, namely, when $p_i = \frac{1}{n}$ for all $\{1,...,n\}$ is of interest as well. If we consider a natural number m with $1 \leq m \leq n-1$ and if we define

$$D_m(f,\mathbf{x}) := \frac{m}{n} f\left(\frac{1}{m}\sum_{i=1}^{m} x_i\right) + \frac{n-m}{n} f\left(\frac{1}{n-m}\sum_{j=m+1}^{n} x_j\right) \tag{2.37}$$

then we can state the following result:

Corollary 3 (Dragomir, 2010 [20])**.** *Let C be a convex subset in the real linear space X and assume that $f : C \to \mathbb{R}$ is a convex function on C. If $x_i \in C$, then for any $m \in \{1,...,n-1\}$ we have*

$$\frac{1}{n}\sum_{k=1}^{n} f(x_k) \geq D_m(f,\mathbf{x}) \geq f\left(\frac{1}{n}\sum_{k=1}^{n} x_k\right). \tag{2.38}$$

In particular, we have the bounds

$$\frac{1}{n}\sum_{k=1}^{n} f(x_k) \geq \max_{m\in\{1,...,n-1\}} \left[\frac{m}{n} f\left(\frac{1}{m}\sum_{i=1}^{m} x_i\right) + \frac{n-m}{n} f\left(\frac{1}{n-m}\sum_{j=m+1}^{n} x_j\right)\right] \tag{2.39}$$

and

$$\min_{m\in\{1,...,n-1\}} \left[\frac{m}{n} f\left(\frac{1}{m}\sum_{i=1}^{m} x_i\right) + \frac{n-m}{n} f\left(\frac{1}{n-m}\sum_{j=m+1}^{n} x_j\right)\right] \geq f\left(\frac{1}{n}\sum_{k=1}^{n} x_k\right). \tag{2.40}$$

The following version of the inequality (2.32) may be useful for symmetric convex functions:

Corollary 4 (Dragomir, 2010 [20])**.** *Let C be a convex function with the property that $0 \in C$. If $y_j \in X$ such that for $p_i > 0, i \in \{1, ..., n\}$ with $\sum_{i=1}^{n} p_i = 1$ we have $y_j - \sum_{i=1}^{n} p_i y_i \in C$ for any $j \in \{1, ..., n\}$, then for any nonempty subset J of $\{1, ..., n\}$ we have*

$$\sum_{k=1}^{n} p_k f\left(y_k - \sum_{i=1}^{n} p_i y_i\right) \geq P_J f\left[\bar{P}_J\left(\frac{1}{P_J}\sum_{i\in J} p_i y_i - \frac{1}{\bar{P}_J}\sum_{j\in \bar{J}} p_j y_j\right)\right] + \bar{P}_J f\left[P_J\left(\frac{1}{\bar{P}_J}\sum_{j\in \bar{J}} p_j y_j - \frac{1}{P_J}\sum_{i\in J} p_i y_i\right)\right] \geq f(0). \tag{2.41}$$

Remark 5. *If C is as in Corollary 4 and $y_j \in X$ such that $y_j - \frac{1}{n}\sum_{i=1}^{n} y_i \in C$ for any $j \in \{1, ..., n\}$ then for any $m \in \{1, ..., n-1\}$ we have*

$$\frac{1}{n}\sum_{k=1}^{n} f\left(y_k - \frac{1}{n}\sum_{i=1}^{n} y_i\right) \geq \frac{m}{n} f\left[\frac{n-m}{n}\left(\frac{1}{m}\sum_{i=1}^{m} y_i - \frac{1}{n-m}\sum_{j=m+1}^{n} y_j\right)\right] + \frac{n-m}{n} f\left[\frac{m}{n}\left(\frac{1}{n-m}\sum_{j=m+1}^{n} y_j - \frac{1}{m}\sum_{i=1}^{m} y_i\right)\right] \geq f(0). \tag{2.42}$$

Remark 6. *It is also useful to remark that if $J = \{k\}$ where $k \in \{1, ..., n\}$ then the particular form we can derive from (2.41) can be written as*

$$\sum_{\ell=1}^{n} p_\ell f\left(y_\ell - \sum_{i=1}^{n} p_i y_i\right) \geq p_k f\left[(1-p_k)\left(y_k - \frac{1}{1-p_k}\left(\sum_{j=1}^{n} p_j y_j - p_k y_k\right)\right)\right] + (1-p_k) f\left[p_k\left(\frac{1}{1-p_k}\left(\sum_{j=1}^{n} p_j y_j - p_k y_k\right) - y_k\right)\right] \geq f(0), \tag{2.43}$$

which is equivalent with

$$\sum_{\ell=1}^{n} p_\ell f\left(y_\ell - \sum_{i=1}^{n} p_i y_i\right) \geq p_k f\left(y_k - \sum_{j=1}^{n} p_j y_j\right) + (1-p_k) f\left[\frac{p_k}{1-p_k}\left(\sum_{j=1}^{n} p_j y_j - y_k\right)\right] \geq f(0) \tag{2.44}$$

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

2.5. **A Lower Bound for Mean f-Deviation.** Let X be a real linear space. For a convex function $f : X \to \mathbb{R}$ with the properties that $f(0) = 0$, define the *mean f-deviation* of an n-tuple of vectors $\mathbf{x} = (x_1, ..., x_n) \in X^n$ with the probability distribution $\mathbf{p}=(p_1,...,p_n)$ by the non-negative quantity

$$K_f(\mathbf{p},\mathbf{x}):=\sum_{i=1}^{n}p_i f\left(x_i-\sum_{k=1}^{n}p_k x_k\right). \tag{2.45}$$

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

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

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

$$K_r(\mathbf{p},\mathbf{x}):=\sum_{i=1}^{n}p_i\left\|x_i-\sum_{k=1}^{n}p_k x_k\right\|^r \tag{2.46}$$

and call it the *mean r-absolute deviation* of the n-tuple of vectors $\mathbf{x}=(x_1,...,x_n)\in X^n$ with the probability distribution $\mathbf{p}=(p_1,...,p_n)$.

The following result that provides a lower bound for the mean f-deviation holds:

Theorem 9 (Dragomir, 2010 [20]). *Let $f:X\to[0,\infty)$ be a convex function with $f(0)=0$. If $\mathbf{x}=(x_1,...,x_n)\in X^n$ and $\mathbf{p}=(p_1,...,p_n)$ is a probability distribution with all p_i nonzero, then*

$$\begin{aligned}K_f(\mathbf{p},\mathbf{x})\geq \max_{\emptyset\neq J\subset\{1,...,n\}}&\left\{P_J f\left[\bar{P}_J\left(\frac{1}{P_J}\sum_{i\in J}p_i x_i-\frac{1}{\bar{P}_J}\sum_{j\in\bar{J}}p_j x_j\right)\right]\right.\\&\left.+P_J f\left(\frac{1}{\bar{P}_J}\sum_{j\in\bar{J}}p_j y_j-\frac{1}{P_J}\sum_{i\in J}p_i y_i\right)\right\}(\geq 0).\end{aligned} \tag{2.47}$$

In particular, we have

$$\begin{aligned}&K_f(\mathbf{p},\mathbf{x})\\&\geq\max_{k\in\{1,...,n\}}\left\{(1-p_k)f\left[\frac{p_k}{1-p_k}\left(\sum_{l=1}^{n}p_l x_l-x_k\right)\right]+p_k f\left(x_k-\sum_{l=1}^{n}p_l x_l\right)\right\}(\geq 0).\end{aligned} \tag{2.48}$$

The proof follows from Corollary 4 and Remark 6.

As a particular case of interest, we have the following:

Corollary 5 (Dragomir, 2010 [20]). *Let $(X,\|\cdot\|)$ be a normed linear space. If $\mathbf{x}=(x_1,...,x_n)\in X^n$ and $\mathbf{p}=(p_1,...,p_n)$ is a probability distribution with all p_i nonzero, then for $r\geq 1$ we have*

$$\begin{aligned}&K_r(\mathbf{p},\mathbf{x})\\&\geq\max_{\emptyset\neq J\subset\{1,...,n\}}\left\{P_J\bar{P}_J\left(\bar{P}_J^{r-1}+P_J^{r-1}\right)\left\|\frac{1}{P_J}\sum_{i\in J}p_i x_i-\frac{1}{\bar{P}_J}\sum_{j\in\bar{J}}p_j x_j\right\|^r\right\}(\geq 0).\end{aligned} \tag{2.49}$$

Remark 7. *By the convexity of the power function $f(t) = t^r, r \geq 1$ we have*

$$\begin{aligned} P_J \bar{P}_J \left(\bar{P}_J^{r-1} + P_J^{r-1} \right) &= P_J \bar{P}_J^r + \bar{P}_J P_J^r \\ &\geq \left(P_J \bar{P}_J + \bar{P}_J P_J \right)^r = 2^r P_J^r \bar{P}_J^r \end{aligned}$$

therefore

$$\begin{aligned} (2.50) \quad & P_J \bar{P}_J \left(\bar{P}_J^{r-1} + P_J^{r-1} \right) \left\| \frac{1}{P_J} \sum_{i \in J} p_i x_i - \frac{1}{\bar{P}_J} \sum_{j \in \bar{J}} p_j x_j \right\|^r \\ & \geq 2^r P_J^r \bar{P}_J^r \left\| \frac{1}{P_J} \sum_{i \in J} p_i x_i - \frac{1}{\bar{P}_J} \sum_{j \in \bar{J}} p_j x_j \right\|^r = 2^r \left\| \bar{P}_J \sum_{i \in J} p_i x_i - P_J \sum_{j \in \bar{J}} p_j x_j \right\|^r . \end{aligned}$$

Since

$$\begin{aligned} (2.51) \quad \bar{P}_J \sum_{i \in J} p_i x_i - P_J \sum_{j \in \bar{J}} p_j x_j &= (1 - P_J) \sum_{i \in J} p_i x_i - P_J \left(\sum_{k=1}^n p_k x_k - \sum_{i \in J} p_i x_i \right) \\ &= \sum_{i \in J} p_i x_i - P_J \sum_{k=1}^n p_k x_k, \end{aligned}$$

then by (2.49)-(2.51) we deduce the coarser but perhaps more useful lower bound

$$(2.52) \qquad K_r(\mathbf{p}, \mathbf{x}) \geq 2^r \max_{\emptyset \neq J \subset \{1,\dots,n\}} \left\{ \left\| \sum_{i \in J} p_i x_i - P_J \sum_{k=1}^n p_k x_k \right\|^r \right\} (\geq 0) .$$

The case for mean r-absolute deviation is incorporated in:

Corollary 6 (Dragomir, 2010 [20])**.** *Let $(X, \|\cdot\|)$ be a normed linear space. If $\mathbf{x} = (x_1, ..., x_n) \in X^n$ and $\mathbf{p} = (p_1, ..., p_n)$ is a probability distribution with all p_i nonzero, then for $r \geq 1$ we have*

$$(2.53) \qquad K_r(\mathbf{p}, \mathbf{x}) \geq \max_{k \in \{1,\dots,n\}} \left\{ \left[(1 - p_k)^{1-r} p_k^r + p_k \right] \left\| x_k - \sum_{l=1}^n p_l x_l \right\|^r \right\} .$$

Remark 8. *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*

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

where $p_m := \min_{k \in \{1,\dots,n\}} p_k$.

We then obtain the following simpler inequality:

$$(2.54) \qquad K_r(\mathbf{p}, \mathbf{x}) \geq \left[p_m + (1 - p_m)^{1-r} \cdot p_m^r \right] \max_{k \in \{1,\dots,n\}} \left\| x_k - \sum_{l=1}^n p_l x_l \right\|^p ,$$

which is perhaps more useful for applications (see also [19]*).*

2.6. **Applications for f-Divergence Measures.** We endeavour to extend the concept of f-divergence for functions defined on a cone in a linear space as follows.

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 $\mathbf{z} = (z_1, ..., z_n) \in K^n$ 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* $\mathbf{q}$

$$I_f(\mathbf{z}, \mathbf{q}) := \sum_{i=1}^{n} q_i f\left(\frac{z_i}{q_i}\right). \tag{2.55}$$

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

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

and

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

where, as above

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

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:

Theorem 10 (Dragomir, 2010 [20]). *Let $f : K \to \mathbb{R}$ be a convex function on the cone K. Then for any n-tuple of vectors $\mathbf{x} = (x_1, ..., x_n) \in K^n$, a probability distribution $\mathbf{q} \in \mathbb{P}^n$ with all values nonzero and for any nonempty subset J of $\{1, ..., n\}$ we have*

$$\begin{aligned} I_f(\mathbf{x}, \mathbf{q}) &\geq \max_{\emptyset \neq J \subset \{1,...,n\}} I_f(\mathbf{x}_J, \mathbf{q}_J) \geq I_f(\mathbf{x}_J, \mathbf{q}_J) \\ &\geq \min_{\emptyset \neq J \subset \{1,...,n\}} I_f(\mathbf{x}_J, \mathbf{q}_J) \geq f(X_n) \end{aligned} \tag{2.56}$$

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

The proof follows by Theorem 8 and the details are omitted.

We observe that, for a given n-tuple of vectors $\mathbf{x} = (x_1, ..., x_n) \in K^n$, 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:

Corollary 7 (Dragomir, 2010 [20]). *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*

$$I_f(\mathbf{p}, \mathbf{q}) \geq \max_{\emptyset \neq J \subset \{1,\ldots,n\}} \left[Q_J f\left(\frac{P_J}{Q_J}\right) + (1 - Q_J) f\left(\frac{1 - P_J}{1 - Q_J}\right) \right] (\geq 0). \tag{2.57}$$

In what follows we provide some lower bounds for a number of f-divergences that are used in various fields of Information Theory, Probability Theory and Statistics.

The *total variation distance* is defined by the convex function $f(t) = |t - 1|$, $t \in \mathbb{R}$ and given in:

$$V(p, q) := \sum_{j=1}^{n} q_j \left| \frac{p_j}{q_j} - 1 \right| = \sum_{j=1}^{n} |p_j - q_j|. \tag{2.58}$$

The following improvement of the positivity inequality for the total variation distance can be stated as follows.

Proposition 5. *For any $\mathbf{p}$, $\mathbf{q} \in \mathbb{P}^n$, we have the inequality:*

$$V(p, q) \geq 2 \max_{\emptyset \neq J \subset \{1,\ldots,n\}} |P_J - Q_J| \quad (\geq 0). \tag{2.59}$$

The proof follows by the inequality (2.57) for $f(t) = |t - 1|$, $t \in \mathbb{R}$.

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

$$\chi^2(p, q) := \sum_{j=1}^{n} q_j \left(\frac{p_j}{q_j} - 1\right)^2 = \sum_{j=1}^{n} \frac{(p_j - q_j)^2}{q_j}. \tag{2.60}$$

Proposition 6. *For any $\mathbf{p}$, $\mathbf{q} \in \mathbb{P}^n$,*

$$\begin{aligned} \chi^2(p, q) &\geq \max_{\emptyset \neq J \subset \{1,\ldots,n\}} \left\{ \frac{(P_J - Q_J)^2}{Q_J(1 - Q_J)} \right\} \\ &\geq 4 \max_{\emptyset \neq J \subset \{1,\ldots,n\}} (P_J - Q_J)^2 \quad (\geq 0). \end{aligned} \tag{2.61}$$

Proof. On applying the inequality (2.57) for the function $f(t) = (1 - t)^2$, $t \in \mathbb{R}$, we get

$$\begin{aligned} \chi^2(p, q) &\geq \max_{\emptyset \neq J \subset \{1,\ldots,n\}} \left\{ (1 - Q_J)\left(\frac{1 - P_J}{1 - Q_J} - 1\right)^2 + Q_J\left(\frac{P_J}{Q_J} - 1\right)^2 \right\} \\ &= \max_{\emptyset \neq J \subset \{1,\ldots,n\}} \left\{ \frac{(P_J - Q_J)^2}{Q_J(1 - Q_J)} \right\}. \end{aligned}$$

Since

$$Q_J(1 - Q_J) \leq \frac{1}{4}[Q_J + (1 - Q_J)]^2 = \frac{1}{4},$$

then

$$\frac{(P_J - Q_J)^2}{Q_J(1 - Q_J)} \geq 4(P_J - Q_J)^2$$

for each $J \subset \{1, \ldots, n\}$, which proves the last part of (2.61). □

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

$$KL(p,q) := \sum_{j=1}^{n} q_j \cdot \frac{p_j}{q_j} \ln\left(\frac{p_j}{q_j}\right) = \sum_{j=1}^{n} p_j \ln\left(\frac{p_j}{q_j}\right). \tag{2.62}$$

Proposition 7. *For any* $\mathbf{p}$, $\mathbf{q} \in \mathbb{P}^n$, *we have:*

$$KL(p,q) \geq \ln\left[\max_{\emptyset \neq J \subset \{1,\dots,n\}} \left\{ \left(\frac{1-P_J}{1-Q_J}\right)^{1-P_J} \cdot \left(\frac{P_J}{Q_J}\right)^{P_J} \right\}\right] \geq 0. \tag{2.63}$$

Proof. The first inequality is obvious by Corollary 7. Utilising the inequality between the *geometric mean and the harmonic mean*,

$$x^{\alpha} y^{1-\alpha} \geq \frac{1}{\frac{\alpha}{x} + \frac{1-\alpha}{y}}, \qquad x, y > 0,\ \alpha \in [0,1]$$

we have for $x = \frac{P_J}{Q_J}, y = \frac{1-P_J}{1-Q_J}$ and $\alpha = P_J$ that

$$\left(\frac{1-P_J}{1-Q_J}\right)^{1-P_J} \cdot \left(\frac{P_J}{Q_J}\right)^{P_J} \geq 1,$$

for any $J \subset \{1,\dots,n\}$, which implies the second part of (2.63). □

Another divergence measure that is of importance in Information Theory is the *Jeffreys divergence*

$$J(p,q) := \sum_{j=1}^{n} q_j \cdot \left(\frac{p_j}{q_j} - 1\right) \ln\left(\frac{p_j}{q_j}\right) = \sum_{j=1}^{n} (p_j - q_j) \ln\left(\frac{p_j}{q_j}\right), \tag{2.64}$$

which is an f-divergence for $f(t) = (t-1)\ln t$, $t > 0$.

Proposition 8. *For any* $\mathbf{p}$, $\mathbf{q} \in \mathbb{P}^n$, *we have:*

$$\begin{aligned} J(p,q) &\geq \ln\left(\max_{\emptyset \neq J \subset \{1,\dots,n\}} \left\{ \left[\frac{(1-P_J) Q_J}{(1-Q_J) P_J}\right]^{(Q_J - P_J)} \right\}\right) \\ &\geq \max_{\emptyset \neq J \subset \{1,\dots,n\}} \left[\frac{(Q_J - P_J)^2}{P_J + Q_J - 2P_J Q_J}\right] \geq 0. \end{aligned} \tag{2.65}$$

Proof. On making use of the inequality (2.57) for $f(t) = (t-1)\ln t$, we have

$$\begin{aligned} &J(p,q) \\ &\geq \max_{k \in \{1,\dots,n\}} \left\{ (1-Q_J)\left[\left(\frac{1-P_J}{1-Q_J} - 1\right) \ln\left(\frac{1-P_J}{1-Q_J}\right)\right] + Q_J \left(\frac{P_J}{Q_J} - 1\right) \ln\left(\frac{P_J}{Q_J}\right) \right\} \\ &= \max_{k \in \{1,\dots,n\}} \left\{ (Q_J - P_J) \ln\left(\frac{1-P_J}{1-Q_J}\right) - (Q_J - P_J) \ln\left(\frac{P_J}{Q_J}\right) \right\} \\ &= \max_{k \in \{1,\dots,n\}} \left\{ (Q_J - P_J) \ln\left[\frac{(1-P_J) Q_J}{(1-Q_J) P_J}\right] \right\}, \end{aligned}$$

proving the first inequality in (2.65).

Utilising the elementary inequality for positive numbers,

$$\frac{\ln b - \ln a}{b - a} \geq \frac{2}{a+b}, \qquad a, b > 0$$

we have

$$\begin{aligned}
&(Q_J - P_J)\left[\ln\left(\frac{1-P_J}{1-Q_J}\right) - \ln\left(\frac{P_J}{Q_J}\right)\right] \\
&= (Q_J - P_J)\cdot\frac{\ln\left(\frac{1-P_J}{1-Q_J}\right) - \ln\left(\frac{P_J}{Q_J}\right)}{\frac{1-P_J}{1-Q_J} - \frac{P_J}{Q_J}}\cdot\left[\frac{1-P_J}{1-Q_J} - \frac{P_J}{Q_J}\right] \\
&= \frac{(Q_J - P_J)^2}{Q_J(1-Q_J)}\cdot\frac{\ln\left(\frac{1-P_J}{1-Q_J}\right) - \ln\left(\frac{P_J}{Q_J}\right)}{\frac{1-P_J}{1-Q_J} - \frac{P_J}{Q_J}} \\
&\geq \frac{(Q_J - P_J)^2}{Q_J(1-Q_J)}\cdot\frac{2}{\frac{1-P_J}{1-Q_J} + \frac{P_J}{Q_J}} = \frac{2(Q_J - P_J)^2}{P_J + Q_J - 2P_JQ_J} \geq 0,
\end{aligned}$$

for each $J \subset \{1, \dots, n\}$, giving the second inequality in (2.65). □

3. Inequalities in Terms of Gâteaux Derivatives

3.1. **Gâteaux Derivatives.** 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

$$\begin{aligned}
\frac{f(x+ty) - f(x)}{t} &\geq \nabla_+ f(x)(y) = \inf_{t>0}\left[\frac{f(x+ty) - f(x)}{t}\right] \\
&\geq \sup_{s<0}\left[\frac{f(x+sy) - f(x)}{s}\right] = \nabla_- f(x)(y) \geq \frac{f(x+sy) - f(x)}{s}
\end{aligned} \tag{3.1}$$

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

$$\nabla_- f(u)(u - v) \geq f(u) - f(v) \geq \nabla_+ f(v)(u - v) \tag{3.2}$$

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 to Jensen's inequality.

The following properties are also of importance:

$$\nabla_+ f(x)(-y) = -\nabla_- f(x)(y), \tag{3.3}$$

and

$$\nabla_{+(-)}f(x)(\alpha y) = \alpha\nabla_{+(-)}f(x)(y) \tag{3.4}$$

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

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

$$\nabla_+ f(x)(y + z) \leq \nabla_+ f(x)(y) + \nabla_+ f(x)(z) \tag{3.5}$$

and

$$\nabla_- f(x)(y + z) \geq \nabla_- f(x)(y) + \nabla_- f(x)(z) \tag{3.6}$$

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 [16].

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 an entire linear space.

The following result holds:

Theorem 11 (Dragomir, 2011 [21])**.** *Let $f : X \to \mathbb{R}$ be a convex function. Then for any $x, y \in X$ and $t \in [0,1]$ we have*

$$\begin{aligned} &t(1-t)[\nabla_- f(y)(y-x) - \nabla_+ f(x)(y-x)] \\ &\geq tf(x) + (1-t)f(y) - f(tx + (1-t)y) \\ &\geq t(1-t)[\nabla_+ f(tx+(1-t)y)(y-x) - \nabla_- f(tx+(1-t)y)(y-x)] \geq 0. \end{aligned} \tag{3.7}$$

Proof. Utilising the gradient inequality (3.2) we have

$$f(tx + (1-t)y) - f(x) \geq (1-t)\nabla_+ f(x)(y-x) \tag{3.8}$$

and

$$f(tx + (1-t)y) - f(y) \geq -t\nabla_- f(y)(y-x). \tag{3.9}$$

If we multiply (3.8) with t and (3.9) with $1-t$ 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 with the first part of (3.7).

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 (3.7). □

The following particular case for norms may be stated:

Corollary 8 (Dragomir, 2011 [21]). *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*

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

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

Remark 9. *We observe that for $p = 1$ in (3.10) we derive the result*

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

while for $p = 2$ we have

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

We notice that the inequality (3.12) holds for any $x, y \in X$ while in the inequality (3.11) we must assume that x, y and $tx + (1-t)y$ are not zero.

Remark 10. *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 [16]). In this situation the inequality (3.10) becomes*

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

and holds for any nonzero x and y.

Moreover, if $(X, \langle\cdot,\cdot\rangle)$ is an inner product space, then (3.13) becomes

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

From (3.14) we deduce the particular inequalities of interest

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

and

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

Obviously, the inequality (3.16) can be proved directly on utilising the properties of the inner products.

3.2. **A Refinement of Jensen's Inequality.** The following refinement of Jensen's inequality holds:

Theorem 12 (Dragomir, 2011 [21]). *Let $f : X \to \mathbb{R}$ be a convex function defined on a linear space X. Then for any n-tuple of vectors $\mathbf{x} = (x_1, ..., x_n) \in X^n$ and any probability distribution $\mathbf{p} = (p_1, ..., p_n) \in \mathbb{P}^n$ we have the inequality*

$$\begin{aligned}(3.17)\quad & \sum_{i=1}^{n} p_i f(x_i) - f\left(\sum_{i=1}^{n} p_i x_i\right) \\ & \geq \sum_{k=1}^{n} p_k \nabla_+ f\left(\sum_{i=1}^{n} p_i x_i\right)(x_k) - \nabla_+ f\left(\sum_{i=1}^{n} p_i x_i\right)\left(\sum_{i=1}^{n} p_i x_i\right) \geq 0.\end{aligned}$$

In particular, for the uniform distribution, we have

$$\begin{aligned}(3.18)\quad & \frac{1}{n}\sum_{i=1}^{n} f(x_i) - f\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) \\ & \geq \frac{1}{n}\left[\sum_{k=1}^{n} \nabla_+ f\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)(x_k) - \nabla_+ f\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} x_i\right)\right] \geq 0.\end{aligned}$$

Proof. Utilising the gradient inequality (3.2) we have

$$(3.19)\qquad f(x_k) - f\left(\sum_{i=1}^{n} p_i x_i\right) \geq \nabla_+ f\left(\sum_{i=1}^{n} p_i x_i\right)\left(x_k - \sum_{i=1}^{n} p_i x_i\right)$$

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

By the subadditivity of the functional $\nabla_+ f(\cdot)(\cdot)$ in the second variable we also have

$$\begin{aligned}(3.20)\quad & \nabla_+ f\left(\sum_{i=1}^{n} p_i x_i\right)\left(x_k - \sum_{i=1}^{n} p_i x_i\right) \\ & \geq \nabla_+ f\left(\sum_{i=1}^{n} p_i x_i\right)(x_k) - \nabla_+ f\left(\sum_{i=1}^{n} p_i x_i\right)\left(\sum_{i=1}^{n} p_i x_i\right)\end{aligned}$$

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

Utilising the inequalities (3.19) and (3.20) we get

$$\begin{aligned}(3.21)\quad & f(x_k) - f\left(\sum_{i=1}^{n} p_i x_i\right) \\ & \geq \nabla_+ f\left(\sum_{i=1}^{n} p_i x_i\right)(x_k) - \nabla_+ f\left(\sum_{i=1}^{n} p_i x_i\right)\left(\sum_{i=1}^{n} p_i x_i\right)\end{aligned}$$

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

Now, if we multiply (3.21) with $p_k \geq 0$ and sum over k from 1 to n, then we deduce the first inequality in (3.17). The second inequality is obvious by the subadditivity property of the functional $\nabla_+ f(\cdot)(\cdot)$ in the second variable. □

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

Corollary 9 (Dragomir, 2011 [21]). *Let $(X, \|\cdot\|)$ be 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 $\mathbf{p} = (p_1, ..., p_n) \in \mathbb{P}^n$ with $\sum_{i=1}^n p_i x_i \neq 0$ we have the inequality*

$$\sum_{i=1}^{n} p_i \|x_i\|^p - \left\| \sum_{i=1}^{n} p_i x_i \right\|^p \geq p \left\| \sum_{i=1}^{n} p_i x_i \right\|^{p-2} \left[\sum_{k=1}^{n} p_k \left\langle x_k, \sum_{j=1}^{n} p_j x_j \right\rangle_s - \left\| \sum_{i=1}^{n} p_i x_i \right\|^2 \right] \geq 0. \tag{3.22}$$

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

In particular, we have the norm inequalities

$$\sum_{i=1}^{n} p_i \|x_i\| - \left\| \sum_{i=1}^{n} p_i x_i \right\| \geq \left[\sum_{k=1}^{n} p_k \left\langle x_k, \frac{\sum_{i=1}^{n} p_i x_i}{\left\| \sum_{i=1}^{n} p_i x_i \right\|} \right\rangle_s - \left\| \sum_{i=1}^{n} p_i x_i \right\| \right] \geq 0. \tag{3.23}$$

and

$$\sum_{i=1}^{n} p_i \|x_i\|^2 - \left\| \sum_{i=1}^{n} p_i x_i \right\|^2 \geq 2 \left[\sum_{k=1}^{n} p_k \left\langle x_k, \sum_{i=1}^{n} p_i x_i \right\rangle_s - \left\| \sum_{i=1}^{n} p_i x_i \right\|^2 \right] \geq 0. \tag{3.24}$$

We notice that the first inequality in (3.24) is equivalent with

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

which provides the result

$$\frac{1}{2} \left[\sum_{i=1}^{n} p_i \|x_i\|^2 + \left\| \sum_{i=1}^{n} p_i x_i \right\|^2 \right] \geq \sum_{k=1}^{n} p_k \left\langle x_k, \sum_{i=1}^{n} p_i x_i \right\rangle_s \left(\geq \left\| \sum_{i=1}^{n} p_i x_i \right\|^2 \right) \tag{3.25}$$

for any n-tuple of vectors and probability distribution.

Remark 11. *If in the inequality (3.22) we consider the uniform distribution, then we get*

$$\sum_{i=1}^{n}\|x_i\|^p - n^{1-p}\left\|\sum_{i=1}^{n}x_i\right\|^p \geq pn^{1-p}\left\|\sum_{i=1}^{n}x_i\right\|^{p-2}\left[\sum_{k=1}^{n}\left\langle x_k,\sum_{i=1}^{n}x_i\right\rangle_s - \left\|\sum_{i=1}^{n}x_i\right\|^2\right] \geq 0. \tag{3.26}$$

3.3. **A Reverse of Jensen's Inequality.** The following result is of interest as well:

Theorem 13 (Dragomir, 2011 [21])**.** *Let $f : X \to \mathbb{R}$ be a convex function defined on a linear space X. Then for any n-tuple of vectors $\mathbf{x} = (x_1, ..., x_n) \in X^n$ and any probability distribution $\mathbf{p} = (p_1, ..., p_n) \in \mathbb{P}^n$ we have the inequality*

$$\sum_{k=1}^{n}p_k\nabla_- f(x_k)(x_k) - \sum_{k=1}^{n}p_k\nabla_- f(x_k)\left(\sum_{i=1}^{n}p_ix_i\right) \geq \sum_{i=1}^{n}p_if(x_i) - f\left(\sum_{i=1}^{n}p_ix_i\right). \tag{3.27}$$

In particular, for the uniform distribution, we have

$$\frac{1}{n}\left[\sum_{k=1}^{n}\nabla_- f(x_k)(x_k) - \sum_{k=1}^{n}\nabla_- f(x_k)\left(\frac{1}{n}\sum_{i=1}^{n}x_i\right)\right] \geq \frac{1}{n}\sum_{i=1}^{n}f(x_i) - f\left(\frac{1}{n}\sum_{i=1}^{n}x_i\right). \tag{3.28}$$

Proof. Utilising the gradient inequality (3.2) we can state that

$$\nabla_- f(x_k)\left(x_k - \sum_{i=1}^{n}p_ix_i\right) \geq f(x_k) - f\left(\sum_{i=1}^{n}p_ix_i\right) \tag{3.29}$$

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

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

$$\nabla_- f(x_k)(x_k) - \nabla_- f(x_k)\left(\sum_{i=1}^{n}p_ix_i\right) \geq \nabla_- f(x_k)\left(x_k - \sum_{i=1}^{n}p_ix_i\right) \tag{3.30}$$

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

Therefore, by (3.29) and (3.30) we get

$$\nabla_- f(x_k)(x_k) - \nabla_- f(x_k)\left(\sum_{i=1}^{n}p_ix_i\right) \geq f(x_k) - f\left(\sum_{i=1}^{n}p_ix_i\right) \tag{3.31}$$

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

Finally, by multiplying (3.31) with $p_k \geq 0$ and summing over k from 1 to n we deduce the desired inequality (3.27). □

Remark 12. *If the function f is defined on the Euclidian space $\mathbb{R}^n$ and is differentiable and convex, then from (3.27) we get the inequality*

$$\sum_{k=1}^{n} p_k \left\langle \nabla f(x_k), x_k \right\rangle - \left\langle \sum_{k=1}^{n} p_k \nabla f(x_k), \sum_{i=1}^{n} p_i x_i \right\rangle \geq \sum_{i=1}^{n} p_i f(x_i) - f\left(\sum_{i=1}^{n} p_i x_i\right) \tag{3.32}$$

where, as usual, for $x_k = \left(x_k^1, ..., x_k^n\right)$, $\nabla f(x_k) = \left(\frac{\partial f(x_k)}{\partial x^1}, ..., \frac{\partial f(x_k)}{\partial x^n}\right)$. This inequality was obtained firstly by Dragomir & Goh in 1996, see [26].

For one dimension we get the inequality

$$\sum_{k=1}^{n} p_k x_k f'(x_k) - \sum_{i=1}^{n} p_i x_i \sum_{k=1}^{n} p_k f'(x_k) \geq \sum_{i=1}^{n} p_i f(x_i) - f\left(\sum_{i=1}^{n} p_i x_i\right) \tag{3.33}$$

that was discovered in 1994 by Dragomir and Ionescu, see [28].

The following reverse of the generalized triangle inequality holds:

Corollary 10 (Dragomir, 2011 [21])**.** *Let $(X, \|\cdot\|)$ be a normed linear space. Then for any $p \geq 1$, for any n-tuple of vectors $\mathbf{x} = (x_1, ..., x_n) \in X^n \setminus \{(0, ..., 0)\}$ and any probability distribution $\mathbf{p} = (p_1, ..., p_n) \in \mathbb{P}^n$ we have the inequality*

$$p\left[\sum_{k=1}^{n} p_k \|x_k\|^p - \sum_{k=1}^{n} p_k \|x_k\|^{p-2} \left\langle \sum_{i=1}^{n} p_i x_i, x_k \right\rangle_i\right] \geq \sum_{i=1}^{n} p_i \|x_i\|^p - \left\|\sum_{i=1}^{n} p_i x_i\right\|^p. \tag{3.34}$$

In particular, we have the norm inequalities

$$\sum_{k=1}^{n} p_k \|x_k\| - \sum_{k=1}^{n} p_k \left\langle \sum_{i=1}^{n} p_i x_i, \frac{x_k}{\|x_k\|} \right\rangle_i \geq \sum_{i=1}^{n} p_i \|x_i\| - \left\|\sum_{i=1}^{n} p_i x_i\right\| \tag{3.35}$$

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

$$2\left[\sum_{k=1}^{n} p_k \|x_k\|^2 - \sum_{k=1}^{n} p_k \left\langle \sum_{j=1}^{n} p_j x_j, x_k \right\rangle_i\right] \geq \sum_{i=1}^{n} p_i \|x_i\|^2 - \left\|\sum_{i=1}^{n} p_i x_i\right\|^2, \tag{3.36}$$

for any x_k.

We observe that the inequality (3.36) is equivalent with

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

which provides the interesting result

$$\frac{1}{2} \left[\sum_{i=1}^{n} p_i \|x_i\|^2 + \left\| \sum_{i=1}^{n} p_i x_i \right\|^2 \right] \geq \sum_{k=1}^{n} p_k \left\langle \sum_{j=1}^{n} p_j x_j, x_k \right\rangle_i \left(\geq \sum_{k=1}^{n} \sum_{j=1}^{n} p_j p_k \langle x_j, x_k \rangle_i \right) \tag{3.37}$$

holding for any n-tuple of vectors and probability distribution.

Remark 13. *If in the inequality (3.34) we consider the uniform distribution, then we get*

$$p \left[\sum_{k=1}^{n} \|x_k\|^p - \frac{1}{n} \sum_{k=1}^{n} \|x_k\|^{p-2} \left\langle \sum_{j=1}^{n} x_j, x_k \right\rangle_i \right] \geq \sum_{i=1}^{n} \|x_i\|^p - n^{1-p} \left\| \sum_{i=1}^{n} x_i \right\|^p . \tag{3.38}$$

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

3.4. **Bounds for the Mean f-Deviation.** Utilising the result from [19] we can state then the following result providing a non-trivial lower bound for the mean f-deviation:

Theorem 14. *Let* $f : X \to [0, \infty)$ *be a convex function with* $f(0) = 0$. *If* $y= (y_1, ..., y_n) \in X^n$ *and* $\mathbf{p} = (p_1, ..., p_n)$ *is a probability distribution with all* p_i *nonzero, then*

$$K_f(\mathbf{p}, \mathbf{y}) \geq \max_{k \in \{1,\dots,n\}} \left\{ (1 - p_k) f \left[\frac{p_k}{1 - p_k} \left(y_k - \sum_{l=1}^{n} p_l y_l \right) \right] + p_k f \left(y_k - \sum_{l=1}^{n} p_l y_l \right) \right\} (\geq 0) . \tag{3.39}$$

The case for mean r-absolute deviation is incorporated in

Corollary 11. *Let* $(X, \|\cdot\|)$ *be a normed linear space. If* $y= (y_1, ..., y_n) \in X^n$ *and* $\mathbf{p} = (p_1, ..., p_n)$ *is a probability distribution with all* p_i *nonzero, then for* $r \geq 1$ *we have*

$$K_r(\mathbf{p}, \mathbf{y}) \geq \max_{k \in \{1,\dots,n\}} \left\{ \left[(1 - p_k)^{1-r} p_k^r + p_k \right] \left\| y_k - \sum_{l=1}^{n} p_l y_l \right\|^r \right\} . \tag{3.40}$$

Remark 14. *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*

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

where $p_m := \min_{k\in\{1,\dots,n\}} p_k$. *We then obtain the following simpler inequality:*

$$K_r(\mathbf{p},\mathbf{y}) \geq \left[p_m + (1-p_m)^{1-r}\cdot p_m^r\right] \max_{k\in\{1,\dots,n\}} \left\|y_k - \sum_{l=1}^n p_l y_l\right\|^p, \tag{3.41}$$

which is perhaps more useful for applications.

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

Theorem 15 (Dragomir, 2011 [21])**.** *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*

$$K_{\nabla_- f(\cdot)(\cdot)}(\mathbf{p},\mathbf{y}) \geq K_{f(\cdot)}(\mathbf{p},\mathbf{y}) \geq K_{\nabla_+ f(0)(\cdot)}(\mathbf{p},\mathbf{y}) \geq 0. \tag{3.42}$$

Proof. If we use the inequality (3.17) for $x_i = y_i - \sum_{k=1}^n p_k y_k$ we get

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

which is equivalent with the second part of (3.42).

Now, by utilising the inequality (3.27) for the same choice of x_i we get

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

which in its turn is equivalent with the first inequality in (3.42). □

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

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

where $g : [0,\infty) \to [0,\infty)$ is a monotonic nondecreasing convex function with $g(0) = 0$.

For this kind of functions we have by direct computation that

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

and

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

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

Corollary 12 (Dragomir, 2011 [21])**.** *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* $y = (y_1, ..., y_n) \in X^n$ *and* $\mathbf{p} = (p_1, ..., p_n)$ *a probability distribution, we have*

$$\sum_{i=1}^{n} p_i g'_- \left(\left\| y_i - \sum_{k=1}^{n} p_k y_k \right\| \right) \left\| y_i - \sum_{k=1}^{n} p_k y_k \right\| \geq \sum_{i=1}^{n} p_i g \left(\left\| y_i - \sum_{k=1}^{n} p_k y_k \right\| \right) \geq g'_+(0) \sum_{i=1}^{n} p_i \left\| y_i - \sum_{k=1}^{n} p_k y_k \right\|. \tag{3.43}$$

3.5. **Bounds for f-Divergence Measures.** The following inequality for the f-divergence of an n-tuple of vectors in a linear space holds [20]:

Theorem 16. *Let* $f : K \to \mathbb{R}$ *be a convex function on the cone* K. *Then for any* n*-tuple of vectors* $\mathbf{x} = (x_1, ..., x_n) \in K^n$, *a probability distribution* $\mathbf{q} \in \mathbb{P}^n$ *with all values nonzero and for any nonempty subset* J *of* $\{1, ..., n\}$ *we have*

$$I_f(\mathbf{x}, \mathbf{q}) \geq \max_{\emptyset \neq J \subset \{1,...,n\}} I_f(\mathbf{x}_J, \mathbf{q}_J) \geq I_f(\mathbf{x}_J, \mathbf{q}_J) \geq \min_{\emptyset \neq J \subset \{1,...,n\}} I_f(\mathbf{x}_J, \mathbf{q}_J) \geq f(X_n) \tag{3.44}$$

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

We observe that, for a given n-tuple of vectors $\mathbf{x} = (x_1, ..., x_n) \in K^n$, 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 [20]:

Corollary 13. *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*

$$I_f(\mathbf{p}, \mathbf{q}) \geq \max_{\emptyset \neq J \subset \{1,...,n\}} \left[Q_J f\left(\frac{P_J}{Q_J}\right) + (1 - Q_J) f\left(\frac{1 - P_J}{1 - Q_J}\right) \right] (\geq 0). \tag{3.45}$$

In what follows, by the use of the results in Theorem 12 and Theorem 13 we can provide an upper and a lower bound for the positive difference $I_f(\mathbf{x}, \mathbf{q}) - f(X_n)$.

Theorem 17 (Dragomir, 2011 [21])**.** *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* $\mathbf{q} \in \mathbb{P}^n$ *with all values nonzero we have*

$$I_{\nabla_- f(\cdot)(\cdot)}(\mathbf{x}, \mathbf{q}) - I_{\nabla_- f(\cdot)(X_n)}(\mathbf{x}, \mathbf{q}) \geq I_f(\mathbf{x}, \mathbf{q}) - f(X_n) \geq I_{\nabla_+ f(X_n)(\cdot)}(\mathbf{x}, \mathbf{q}) - \nabla_+ f(X_n)(X_n) \geq 0. \tag{3.46}$$

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

Corollary 14. *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*

$$I_{f'_-(\cdot)(\cdot)}(\mathbf{p}, \mathbf{q}) - I_{f'_-(\cdot)}(\mathbf{p}, \mathbf{q}) \geq I_f(\mathbf{p}, \mathbf{q}) \geq 0, \tag{3.47}$$

or, equivalently,

$$I_{f'_-(\cdot)[(\cdot)-1]}(\mathbf{p}, \mathbf{q}) \geq I_f(\mathbf{p}, \mathbf{q}) \geq 0. \tag{3.48}$$

The above corollary is useful to provide 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 9. *Let $f:[0,\infty)\to\mathbb{R}$ be a normalized convex function and $\mathbf{p}$, $\mathbf{q}\in\mathbb{P}^n$. If there exists the constants γ and Γ with*

$$-\infty<\gamma\leq f'_-\left(\frac{p_k}{q_k}\right)\leq\Gamma<\infty \text{ for all } k\in\{1,...,n\},$$

then we have the inequality

$$0\leq I_f(\mathbf{p},\mathbf{q})\leq\frac{1}{2}(\Gamma-\gamma)V(\mathbf{p},\mathbf{q}), \tag{3.49}$$

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 the inequality (3.48) we have successively that

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

which proves the desired result (3.49). □

Corollary 15. *Let $f:[0,\infty)\to\mathbb{R}$ be a normalized convex function and $\mathbf{p}$, $\mathbf{q}\in\mathbb{P}^n$. If there exist the constants r and R with*

$$0<r\leq\frac{p_k}{q_k}\leq R<\infty \text{ for all } k\in\{1,...,n\},$$

then we have the inequality

$$0\leq I_f(\mathbf{p},\mathbf{q})\leq\frac{1}{2}\left[f'_-(R)-f'_-(r)\right]V(\mathbf{p},\mathbf{q}). \tag{3.50}$$

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

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

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

Proposition 10. *Let $f:[0,\infty)\to\mathbb{R}$ be a normalized convex function and $\mathbf{p}$, $\mathbf{q}\in\mathbb{P}^n$. If there exists the constant $0<\Delta<\infty$ with*

$$\left|\frac{f'_-\left(\frac{p_i}{q_i}\right)-f'_-(1)}{\frac{p_i}{q_i}-1}\right|\leq\Delta \text{ for all } k\in\{1,...,n\}, \tag{3.51}$$

then we have the inequality

$$0 \leq I_f(\mathbf{p}, \mathbf{q}) \leq \Delta\chi^2(p, q). \tag{3.52}$$

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

$$\left|f'_-(x) - f'_-(1)\right| \leq \Delta|x - 1| \text{ for any } x \in (0, \infty) \tag{3.53}$$

then (3.52) holds true for any $\mathbf{p}$, $\mathbf{q} \in \mathbb{P}^n$.

Proof. We have from (3.48) that

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

and the inequality (3.52) is obtained. □

Remark 15. *It is obvious that if one chooses in the above inequalities particular normalized convex functions that generates the Kullback-Leibler, Jeffreys, Hellinger or other divergence measures or discrepancies, that one can obtain some results of interest. However the details are not provided here.*

4. Inequalities of Slater's Type

4.1. **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 18 (Slater, 1981, [32])**.** *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*

$$\frac{1}{P_n}\sum_{i=1}^n p_i f(x_i)\leq f\left(\frac{\sum_{i=1}^n p_i x_i\varphi(x_i)}{\sum_{i=1}^n p_i\varphi(x_i)}\right). \tag{4.1}$$

As pointed out in [17, p. 208], the monotonicity assumption for the derivative φ can be replaced with the condition

$$\frac{\sum_{i=1}^n p_i x_i\varphi(x_i)}{\sum_{i=1}^n p_i\varphi(x_i)}\in I, \tag{4.2}$$

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

The main aim of the next section 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.

4.2. **Slater's Inequality for Functions Defined on Linear Spaces.** 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

$$Sla_{+(-)}(f,\mathbf{x}):=\left\{v\in X\mid \nabla_{+(-)}f(x_i)(v-x_i)\geq 0\text{ for all } i\in\{1,...,n\}\right\} \tag{4.3}$$

and

$$Sla_{+(-)}(f,\mathbf{x},\mathbf{p}):=\left\{v\in X\mid \sum_{i=1}^n p_i\nabla_{+(-)}f(x_i)(v-x_i)\geq 0\right\} \tag{4.4}$$

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 (Dragomir, 2012 [25])**.** *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. □

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 [16]). 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 19 (Dragomir, 2012 [25])**.** *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*

$$\nabla_- f(v)(v) - \sum_{i=1}^{n} p_i \nabla_- f(v)(x_i) \geq f(v) - \sum_{i=1}^{n} p_i f(x_i) \geq 0. \tag{4.5}$$

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

$$\nabla_- f(v)(v - x_i) \geq f(v) - f(x_i) \geq \nabla_+ f(x_i)(v - x_i) \tag{4.6}$$

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

By multiplying (4.6) with $p_i \geq 0$ and summing over i from 1 to n we get

$$\sum_{i=1}^{n} p_i \nabla_- f(v)(v - x_i) \geq f(v) - \sum_{i=1}^{n} p_i f(x_i) \geq \sum_{i=1}^{n} p_i \nabla_+ f(x_i)(v - x_i). \tag{4.7}$$

Now, since $v \in Sla_+(f, \mathbf{x}, \mathbf{p})$, then the right hand side of (4.7) is nonnegative, which proves the second inequality in (4.5).

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

$$\nabla_- f(v)(v) - \sum_{i=1}^{n} p_i \nabla_- f(v)(x_i) \geq \sum_{i=1}^{n} p_i \nabla_- f(v)(v - x_i). \tag{4.8}$$

Utilising (4.7) and (4.8) we deduce the desired result (4.5). □

Remark 16. *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*

$$\sum_{j=1}^{n} p_j \|x_j\|^{p-2} \langle v, x_j \rangle_s \geq \sum_{j=1}^{n} p_j \|x_j\|^p, \ p \geq 1, \tag{4.9}$$

we have the inequalities

$$p \left[\|v\|^p - \sum_{j=1}^{n} p_j \|x_j\|^{p-2} \langle v, x_j \rangle_i \right] \geq \|v\|^p - \sum_{j=1}^{n} p_j \|x_j\|^p \geq 0. \tag{4.10}$$

Rearranging the first inequality in (4.10) we also have that

$$(p-1)\|v\|^p + \sum_{j=1}^{n} p_j \|x_j\|^p \geq p \sum_{j=1}^{n} p_j \|x_j\|^{p-2} \langle v, x_j \rangle_i. \tag{4.11}$$

If the space is smooth, then the condition (4.9) becomes

$$\sum_{j=1}^{n} p_j \|x_j\|^{p-2} [v, x_j] \geq \sum_{j=1}^{n} p_j \|x_j\|^p, \ p \geq 1, \tag{4.12}$$

implying the inequality

$$p\left[\|v\|^{p}-\sum_{j=1}^{n}p_{j}\|x_{j}\|^{p-2}[v,x_{j}]\right]\geq\|v\|^{p}-\sum_{j=1}^{n}p_{j}\|x_{j}\|^{p}\geq 0. \tag{4.13}$$

Notice also that the first inequality in (4.13) is equivalent with

$$(p-1)\|v\|^{p}+\sum_{j=1}^{n}p_{j}\|x_{j}\|^{p}\geq p\sum_{j=1}^{n}p_{j}\|x_{j}\|^{p-2}[v,x_{j}] \left(\geq p\sum_{j=1}^{n}p_{j}\|x_{j}\|^{p}\geq 0\right). \tag{4.14}$$

The following corollary is of interest:

Corollary 16 (Dragomir, 2012 [25])**.** *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*

$$\sum_{i=1}^{n}p_{i}\nabla_{+}f(x_{i})(x_{i})\geq(<)\,0 \tag{4.15}$$

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

$$\sum_{i=1}^{n}p_{i}\nabla_{+(-)}f(x_{i})(s)\geq(\leq)\,1 \tag{4.16}$$

then

$$\begin{aligned}\nabla_{-}f\left(\sum_{j=1}^{n}p_{j}\nabla_{+}f(x_{j})(x_{j})\,s\right)\left(\sum_{j=1}^{n}p_{j}\nabla_{+}f(x_{j})(x_{j})\,s\right)\\-\sum_{i=1}^{n}p_{i}\nabla_{-}f\left(\sum_{j=1}^{n}p_{j}\nabla_{+}f(x_{j})(x_{j})\,s\right)(x_{i})\\\geq f\left(\sum_{j=1}^{n}p_{j}\nabla_{+}f(x_{j})(x_{j})\,s\right)-\sum_{i=1}^{n}p_{i}f(x_{i})\geq 0.\end{aligned} \tag{4.17}$$

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 19 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

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

Therefore $w \in Sla_+(f, \mathbf{x}, \mathbf{p})$ and by Theorem 19 we get the desired result. □

It is natural to consider the case of normed spaces.

Remark 17. *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*

$$p\sum_{i=1}^{n} p_i \|x_i\|^{p-2} \langle s, x_i\rangle_s \geq 1, \tag{4.18}$$

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.

4.3. **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 $x = (x^1, ..., x^m) \in C$ then for any $y = (y^1, ..., y^m) \in \mathbb{R}^m$ we have

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

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 $x_i = (x_i^1, ..., x_i^m)$ with $i \in \{1, ..., n\}$, we consider the sets

$$\begin{aligned} Sla(f, \mathbf{x}, C) := \Bigg\{ v \in C \mid & \sum_{k=1}^{m} \frac{\partial f(x_i^1, ..., x_i^m)}{\partial x^k} \cdot v^k \\ & \geq \sum_{k=1}^{m} \frac{\partial f(x_i^1, ..., x_i^m)}{\partial x^k} \cdot x_i^k \text{ for all } i \in \{1, ..., n\} \Bigg\} \end{aligned} \tag{4.19}$$

and

$$\begin{aligned} Sla(f, \mathbf{x}, \mathbf{p}, C) := \Bigg\{ v \in C \mid & \sum_{i=1}^{n}\sum_{k=1}^{m} p_i \frac{\partial f(x_i^1, ..., x_i^m)}{\partial x^k} \cdot v^k \\ & \geq \sum_{i=1}^{n}\sum_{k=1}^{m} p_i \frac{\partial f(x_i^1, ..., x_i^m)}{\partial x^k} \cdot x_i^k \Bigg\} \end{aligned} \tag{4.20}$$

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 $\text{clo}(C)$, 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 11. *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 $v = (v^1, ..., v^n) \in$*

$Sla\,(f, \mathbf{x}, \mathbf{p}, C)$ *we have the inequalities*

$$\sum_{k=1}^{m} \frac{\partial f\left(v^1, ..., v^m\right)}{\partial x^k} \cdot v^k - \sum_{i=1}^{n} \sum_{k=1}^{m} p_i \; \frac{\partial f\left(x_i^1, ..., x_i^m\right)}{\partial x^k} \cdot v^k \geq f\left(v^1, ..., v^n\right) - \sum_{i=1}^{n} p_i f\left(x_i^1, ..., x_i^m\right) \geq 0. \tag{4.21}$$

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

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

and

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

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 $\sum_{i=1}^{n} p_i f'_+\,(x_i) \neq 0$ then for $\sum_{i=1}^{n} p_i f'_+\,(x_i) > 0$ we have

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

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

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

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

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

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

Proposition 12. *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*

$$f'_-\,(v) \left(v - \sum_{i=1}^{n} p_i x_i \right) \geq f\,(v) - \sum_{i=1}^{n} p_i f\,(x_i) \geq 0 \tag{4.22}$$

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

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

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

then

$$f'_{-}\left(\frac{\sum_{i=1}^{n}p_if'_{+}(x_i)x_i}{\sum_{i=1}^{n}p_if'_{+}(x_i)}\right)\left[\frac{\sum_{i=1}^{n}p_if'_{+}(x_i)x_i}{\sum_{i=1}^{n}p_if'_{+}(x_i)}-\sum_{i=1}^{n}p_ix_i\right] \geq f\left(\frac{\sum_{i=1}^{n}p_if'_{+}(x_i)x_i}{\sum_{i=1}^{n}p_if'_{+}(x_i)}\right)-\sum_{i=1}^{n}p_if(x_i)\geq 0 \tag{4.23}$$

Moreover, if $f'_{+}(x_i)\geq 0$ *for all* $i\in\{1,...,n\}$ *and* $\sum_{i=1}^{n}p_if'_{+}(x_i)>0$ *then (4.23) holds true as well.*

Remark 18. *We remark that the first inequality in (4.23) provides a reverse inequality for the classical result due to Slater.*

4.4. **Some Applications for f-divergences.** 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

$$Sla_{+}(f,\mathbf{p},\mathbf{q}):=\left\{v\in[0,\infty)|\ \sum_{i=1}^{n}q_if'_{+}\left(\frac{p_i}{q_i}\right)\cdot\left(v-\frac{p_i}{q_i}\right)\geq 0\right\}. \tag{4.24}$$

Now, observe that

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

is equivalent with

$$v\sum_{i=1}^{n}q_if'_{+}\left(\frac{p_i}{q_i}\right)\geq\sum_{i=1}^{n}p_if'_{+}\left(\frac{p_i}{q_i}\right). \tag{4.25}$$

If $\sum_{i=1}^{n}q_if'_{+}\left(\frac{p_i}{q_i}\right)>0$, then (4.25) is equivalent with

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

therefore in this case

$$Sla_{+}(f,\mathbf{p},\mathbf{q})=\begin{cases}[0,\infty) & \text{if } \sum_{i=1}^{n}p_if'_{+}\left(\frac{p_i}{q_i}\right)<0,\\[2ex] \left[\frac{\sum_{i=1}^{n}p_if'_{+}\left(\frac{p_i}{q_i}\right)}{\sum_{i=1}^{n}q_if'_{+}\left(\frac{p_i}{q_i}\right)},\infty\right) & \text{if } \sum_{i=1}^{n}p_if'_{+}\left(\frac{p_i}{q_i}\right)\geq 0.\end{cases} \tag{4.26}$$

If $\sum_{i=1}^{n}q_if'_{+}\left(\frac{p_i}{q_i}\right)<0$, then (4.25) is equivalent with

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

therefore

$$Sla_{+}(f,\mathbf{p},\mathbf{q})=\begin{cases}\left[0,\frac{\sum_{i=1}^{n}p_{i}f'_{+}\left(\frac{p_{i}}{q_{i}}\right)}{\sum_{i=1}^{n}q_{i}f'_{+}\left(\frac{p_{i}}{q_{i}}\right)}\right] & \text{if } \sum_{i=1}^{n}p_{i}f'_{+}\left(\frac{p_{i}}{q_{i}}\right)\leq 0,\\ \\ \emptyset & \text{if } \sum_{i=1}^{n}p_{i}f'_{+}\left(\frac{p_{i}}{q_{i}}\right)>0.\end{cases} \tag{4.27}$$

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

Theorem 20 (Dragomir, 2012 [25])**.** *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*

$$f'_{-}(v)(v-1)\geq f(v)-I_{f}(\mathbf{p},\mathbf{q})\geq 0. \tag{4.28}$$

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

$$\begin{aligned} f'_{-}\left(\frac{I_{f'_{+}(\cdot)(\cdot)}(\mathbf{p},\mathbf{q})}{I_{f'_{+}}(\mathbf{p},\mathbf{q})}\right)&\left[\frac{I_{f'_{+}(\cdot)(\cdot)}(\mathbf{p},\mathbf{q})}{I_{f'_{+}}(\mathbf{p},\mathbf{q})}-1\right]\\ &\geq f\left(\frac{I_{f'_{+}(\cdot)(\cdot)}(\mathbf{p},\mathbf{q})}{I_{f'_{+}}(\mathbf{p},\mathbf{q})}\right)-I_{f}(\mathbf{p},\mathbf{q})\geq 0. \end{aligned} \tag{4.29}$$

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 (4.29) holds true as well.

The proof follows immediately from Proposition 12 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

$$\chi^{2}(\mathbf{p},\mathbf{q}):=\sum_{j=1}^{n}q_{j}\left(\frac{p_{j}}{q_{j}}-1\right)^{2}=\sum_{j=1}^{n}\frac{(p_{j}-q_{j})^{2}}{q_{j}}=\sum_{j=1}^{n}\frac{p_{i}^{2}}{q_{i}}-1. \tag{4.30}$$

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

$$KL(\mathbf{p},\mathbf{q}):=\sum_{j=1}^{n}q_{j}\cdot\frac{p_{j}}{q_{j}}\ln\left(\frac{p_{j}}{q_{j}}\right)=\sum_{j=1}^{n}p_{j}\ln\left(\frac{p_{j}}{q_{j}}\right). \tag{4.31}$$

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

$$\begin{aligned} I_{f}(\mathbf{p},\mathbf{q}) \quad :&=\sum_{i=1}^{n}q_{i}f\left(\frac{p_{i}}{q_{i}}\right)=-\sum_{i=1}^{n}q_{i}\ln\left(\frac{p_{i}}{q_{i}}\right)\\ &=\sum_{i=1}^{n}q_{i}\ln\left(\frac{q_{i}}{p_{i}}\right)=KL(\mathbf{q},\mathbf{p}). \end{aligned} \tag{4.32}$$

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

$$Sla(-\ln, \mathbf{p}, \mathbf{q}) := \left\{ v \in [0,\infty) | \ -\sum_{i=1}^{n} q_i \left(\frac{p_i}{q_i}\right)^{-1} \cdot \left(v - \frac{p_i}{q_i}\right) \geq 0 \right\} = \left\{ v \in [0,\infty) | \ v\sum_{i=1}^{n} \frac{q_i^2}{p_i} - 1 \leq 0 \right\} = \left[0, \frac{1}{\chi^2(\mathbf{q},\mathbf{p})+1}\right]. \tag{4.33}$$

Utilising the first part of the Theorem 20 we can state the following

Proposition 13. *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*

$$\frac{1-v}{v} \geq -\ln(v) - KL(\mathbf{q},\mathbf{p}) \geq 0. \tag{4.34}$$

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

$$\chi^2(\mathbf{q},\mathbf{p}) \geq \ln\left[\chi^2(\mathbf{q},\mathbf{p}) + 1\right] - KL(\mathbf{q},\mathbf{p}) \geq 0. \tag{4.35}$$

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

$$\begin{aligned} &Sla((\cdot)\ln(\cdot), \mathbf{p}, \mathbf{q}) \\ &:= \left\{ v \in [0,\infty) | \ \sum_{i=1}^{n} q_i \left(\ln\left(\frac{p_i}{q_i}\right) + 1\right) \cdot \left(v - \frac{p_i}{q_i}\right) \geq 0 \right\} \\ &= \left\{ v \in [0,\infty) | \ v\sum_{i=1}^{n} q_i \left(\ln\left(\frac{p_i}{q_i}\right) + 1\right) - \sum_{i=1}^{n} p_i \cdot \left(\ln\left(\frac{p_i}{q_i}\right) + 1\right) \geq 0 \right\} \\ &= \{ v \in [0,\infty) | \ v(1 - KL(\mathbf{q},\mathbf{p})) \geq 1 + KL(\mathbf{p},\mathbf{q}) \}. \end{aligned} \tag{4.36}$$

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

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

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

By the use of Theorem 20 we can state now the following

Proposition 14. *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*

$$(\ln v + 1)(v - 1) \geq v\ln(v) - KL(\mathbf{p},\mathbf{q}) \geq 0. \tag{4.37}$$

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

$$\begin{aligned} &\left(\ln\left[\frac{1 + KL(\mathbf{p},\mathbf{q})}{1 - KL(\mathbf{q},\mathbf{p})}\right] + 1\right)\left(\frac{1 + KL(\mathbf{p},\mathbf{q})}{1 - KL(\mathbf{q},\mathbf{p})} - 1\right) \\ &\qquad \geq \frac{1 + KL(\mathbf{p},\mathbf{q})}{1 - KL(\mathbf{q},\mathbf{p})} \ln\left[\frac{1 + KL(\mathbf{p},\mathbf{q})}{1 - KL(\mathbf{q},\mathbf{p})}\right] - KL(\mathbf{p},\mathbf{q}) \geq 0. \end{aligned} \tag{4.38}$$

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] P. CERONE and S. S. DRAGOMIR, A refinement of the Grüss inequality and applications. *Tamkang J. Math.* **38** (2007), no. 1, 37–49. Preprint RGMIA *Res. Rep. Coll.* **5** (2002), No. 2, Art. 14. [Online http://rgmia.org/papers/v5n2/RGIApp.pdf].

[2] S. S. DRAGOMIR, An improvement of Jensen's inequality, *Bull. Math. Soc. Sci. Math. Roumanie*, **34** (**82**) (1990), No. 4, 291-296.

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

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

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

[6] S. S. DRAGOMIR, A new improvement of Jensen's inequality, *Indian J. Pure and Appl. Math.*, **26** (10) (1995), 959-968.

[7] S. S. DRAGOMIR, New estimation of the remainder in Taylor's formula using Grüss' type inequalities and applications, *Math. Ineq. and Appl.*, **2** (2) (1999), 183-193.

[8] S. S. DRAGOMIR, On the Ostrowski's integral inequality for mappings of bounded variation and applications, *RGMIA Res. Rep. Coll.* **2** (1999) No. 1, 63–70. [Online http://rgmia.org/papers/v2n1/v2n1-7.pdf].

[9] S. S. DRAGOMIR, An improvement of the remainder estimate in the generalised Taylor's formula, *RGMIA Res. Rep. Coll.,* **3** (2000), No. 1, Article 1. [Online http://rgmia.org/papers/v3n1/IREGTF.pdf].

[10] S. S. DRAGOMIR, Other inequalities for Csiszár divergence and applications, Preprint, RGMIA Monographs, Victoria University, 2000. [Online http://rgmia.org/papers/Csiszar/ICDApp.pdf].

[11] S. S. DRAGOMIR, Some inequalities for (m, M)-convex mappings and applications for the Csiszár f-divergence in Information Theory, *Math. J. Ibaraki Univ.,* 33 (2001), 35-50. [Online http://rgmia.org/papers/Csiszar/ImMCMACFDIT.pdf].

[12] S. S. DRAGOMIR, Some inequalities for two Csiszár divergences and applications, *Mat. Bilten* (Macedonia), **25** (2001), 73-90. [Online http://rgmia.org/papers/Csiszar/In2CsisDivApp.pdf].

[13] S. S. DRAGOMIR, On a reverse of Jessen's inequality for isotonic linear functionals, *J. Ineq. Pure & Appl. Math.*, **2**(2001), No. 3, Article 36.

[14] S. S. DRAGOMIR, A Grüss type inequality for isotonic linear functionals and applications. *Demonstratio Math.* **36** (2003), no. 3, 551–562. Preprint *RGMIA Res. Rep. Coll.* **5** (2002), Suplement, Art. 12. [Online http://rgmia.org/papers/v5e/GTIILFApp.pdf].

[15] S. S. DRAGOMIR, A converse inequality for the Csiszár Φ-divergence. *Tamsui Oxf. J. Math. Sci.* **20** (2004), no. 1, 35–53. Preprint in S. S. Dragomir (Ed.): *Inequalities for Csiszár f-Divergence in Information Theory,* RGMIA Monographs, Victoria University, 2000, [Online http://rgmia.org/papers/Csiszar/Csiszar.pdf].

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

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

[18] S. S. DRAGOMIR, Bounds for the normalized Jensen functional, *Bull. Austral. Math. Soc.* **74**(3)(2006), 471-476.

[19] S. S. DRAGOMIR, A refinement of Jensen's inequality with applications for f-divergence measures, *Taiwanese J. Math.* **14** (2010), no. 1, 153–164.

[20] S. S. DRAGOMIR, A new refinement of Jensen's inequality in linear spaces with applications. *Math. Comput. Modelling* **52** (2010), no. 9-10, 1497–1505.

[21] S. S. DRAGOMIR, Inequalities in terms of the Gâteaux derivatives for convex functions on linear spaces with applications. *Bull. Aust. Math. Soc.* **83** (2011), no. 3, 500–517.

[22] S. S. DRAGOMIR, A refinement and a divided difference reverse of Jensen's inequality with applications, *Rev. Colombiana Mat.* **50** (2016), No. 1, 17–39. Preprint *RGMIA Res. Rep. Coll.* **14** (2011), Art. 74. [Online http://rgmia.org/papers/v14/v14a74.pdf].

[23] S. S. DRAGOMIR, Some reverses of the Jensen inequality with applications. *Bull. Aust. Math. Soc.* **87** (2013), no. 2, 177–194.

[24] S. S. DRAGOMIR, Reverses of the Jensen inequality in terms of first derivative and applications. *Acta Math. Vietnam.* **38** (2013), no. 3, 429–446

[25] S. S. DRAGOMIR, Some Slater's type inequalities for convex functions defined on linear spaces and applications. *Abstr. Appl. Anal.* **2012**, Art. ID 168405, 16 pp.

[26] 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.

[27] S. S. DRAGOMIR and C. J. GOH, Some bounds on entropy measures in information theory, *Appl. Math. Lett.,* **10** (3) (1997) 23–28.

[28] S. S. DRAGOMIR and N. M. IONESCU, Some converse of Jensen's inequality and applications. *Rev. Anal. Numér. Théor. Approx.* **23** (1994), no. 1, 71–78.

[29] S. S. DRAGOMIR, J. PEČARIĆ and L. E. PERSSON, Properties of some functionals related to Jensen's inequality, *Acta Math. Hung.*, **70**(1-2) (1996), 129-143.

[30] J. PEČARIĆ and S. S. DRAGOMIR, A refinements of Jensen inequality and applications, *Studia Univ. Babeş-Bolyai, Mathematica,* **24** (1) (1989), 15-19.

[31] S. SIMIĆ, On a global upper bound for Jensen's inequality, *J. Math. Anal. Appl.* **343**(2008), 414-419.

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

[1]Mathematics, College of Engineering & Science, Victoria University, PO Box 14428, Melbourne City, MC 8001, Australia.

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

URL: http://rgmia.org/dragomir

[2]DST-NRF Centre of Excellence in the Mathematical and Statistical Sciences, School of Computer Science & Applied Mathematics, University of the Witwatersrand, Private Bag 3, Johannesburg 2050, South Africa