

Sherman's inequality and its converse for strongly convex functions with applications to generalized f -divergences

Slavica IVELIĆ BRADANOVIĆ* 

Department of Mathematics, Faculty of Civil Engineering, Architecture And Geodesy,
University of Split, Split, Croatia

Received: 20.05.2019

Accepted/Published Online: 31.08.2019

Final Version: 22.11.2019

Abstract: Considering the weighted concept of majorization, Sherman obtained generalization of majorization inequality for convex functions known as Sherman's inequality. We extend Sherman's result to the class of n -strongly convex functions using extended idea of convexity to the class of strongly convex functions. We also obtain upper bound for Sherman's inequality, called the converse Sherman inequality, and as easy consequences we get Jensen's as well as majorization inequality and their conversions for strongly convex functions. Obtained results are stronger versions for analogous results for convex functions. As applications, we introduced a generalized concept of f -divergence and derived some reverse relations for such concept.

Key words: Jensen inequality, convex function, strongly convex function, majorization, Sherman inequality

1. Introduction

A function $f : [\alpha, \beta] \rightarrow \mathbb{R}$ is called strongly convex with modulus $c > 0$ if

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) - c\lambda(1 - \lambda)(x - y)^2 \quad (1.1)$$

for all $x, y \in [\alpha, \beta]$ and $\lambda \in [0, 1]$.

The concept of strongly convexity has been introduced by Polyak [34]. It has a large number of appearance in many different fields of applications, particular in many branches of mathematics as well as optimization theory, mathematical economics and approximation theory. Strongly convex functions have many nice properties (see [31]).

A function f that satisfies (1.1) with $c = 0$, i.e.

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) \quad (1.2)$$

is convex in usual sense. Specially, if the inequality in (1.2) is strict, then f is called strictly convex.

It is well known that the following implications hold:

$$\text{strongly convex} \Rightarrow \text{strictly convex} \Rightarrow \text{convex.}$$

However, the reverse implications are not true, in general.

*Correspondence: sivelic@gradst.hr

2010 AMS Mathematics Subject Classification: 26D15

Example 1.1 The function $f(x) = x^2$ is strongly convex and also strictly convex and convex. The function $g(x) = e^x$ is strictly convex and convex but not strongly convex. The function $h(x) = x$ is convex but neither strictly nor strongly convex.

In the theory of convex functions, natural generalization are convex functions of higher order, i.e. n -convex functions. The notion of n -convexity was defined in terms of divided differences by T. Popoviciu [35] which we introduce in the sequel.

A function $f : [\alpha, \beta] \rightarrow \mathbb{R}$ is said to be n -convex if for every choice of $n + 1$ distinct points $z_0, \dots, z_n \in [\alpha, \beta]$, the n th order divided difference is nonnegative, i.e.

$$[z_0, z_1, \dots, z_n; f] \geq 0, \tag{1.3}$$

where divided difference may be formally defined by

$$[z_i; f] = f(z_i), \quad i = 0, \dots, n$$

$$[z_0, \dots, z_n; f] = \frac{[z_1, \dots, z_n; f] - [z_0, \dots, z_{n-1}; f]}{z_n - z_0}.$$

The value $[z_0, \dots, z_n; f]$ is independent of the order of the points z_0, \dots, z_n . This definition may be extended to include the case in which some or all the points coincide. Assuming that $f^{(j-1)}(z)$ exists, we define

$$\underbrace{[z, \dots, z; f]}_{j\text{-times}} = \frac{f^{(j-1)}(z)}{(j-1)!}. \tag{1.4}$$

Remark 1.2 It is known that 1-convex function is increasing function and 2-convex function is just ordinary convex function, i.e. convex in usual sense.

If $f^{(n)}$ exists, then f is n -convex iff $f^{(n)} \geq 0$.

Also, if f is n -convex for $n \geq 2$, then $f^{(k)}$ exists and f is $(n - k)$ -convex for $1 \leq k \leq n - 2$. For more information see [33].

Following Gera and Nikodem [11], we say that a function $f : [\alpha, \beta] \rightarrow \mathbb{R}$ is strongly convex of order n with modulus $c > 0$ (or n -strongly convex with modulus $c > 0$) if

$$[z_0, \dots, z_n; f] \geq c \tag{1.5}$$

for all $z_0, \dots, z_n \in [\alpha, \beta]$.

Remark 1.3 Note that 2-strongly convex function with modulus c is just strongly convex function with modulus c as given by (1.1).

For $n = 2$, the condition (1.5) is equivalent to

$$\frac{f(z_0)}{(z_0 - z_1)(z_0 - z_2)} + \frac{f(z_1)}{(z_1 - z_2)(z_1 - z_0)} + \frac{f(z_2)}{(z_2 - z_1)(z_2 - z_0)} \geq c$$

or

$$f(z_1) \leq \frac{z_2 - z_1}{z_2 - z_0} f(z_0) + \frac{z_1 - z_0}{z_2 - z_0} f(z_2) - c(z_2 - z_1)(z_1 - z_0).$$

A function $f : [\alpha, \beta] \rightarrow \mathbb{R}$ is a strongly n -convex with modulus c iff the function $g(x) = f(x) - cx^n$ is n -convex. A function $f : [\alpha, \beta] \rightarrow \mathbb{R}$ is a strongly n -convex with modulus c iff $f^{(n)} \geq cn!$. For more information see [11, 31, 32].

The concept of strongly convexity is a strengthening of the concept of convexity and some properties of strongly convex functions are just stronger versions of analogous properties of convex functions.

For $f : [\alpha, \beta] \rightarrow \mathbb{R}$ strongly convex function with modulus $c > 0$, Jensen’s inequality

$$f\left(\sum_{i=1}^m a_i x_i\right) \leq \sum_{i=1}^m a_i f(x_i) - c \sum_{i=1}^m a_i (x_i - \bar{x})^2 \tag{1.6}$$

holds, where $\mathbf{x} = (x_1, \dots, x_m) \in [\alpha, \beta]^m$, $\mathbf{a} = (a_1, \dots, a_m) \in [0, \infty)^m$ with $\sum_{i=1}^m a_i = 1$ and $\bar{x} = \sum_{i=1}^m a_i x_i$ (see [22]). On the other side, Jensen’s inequality for a classical convex function f has the form

$$f\left(\sum_{i=1}^m a_i x_i\right) \leq \sum_{i=1}^m a_i f(x_i). \tag{1.7}$$

If we compare (1.6) with (1.7), note that the inequality (1.6) includes a better upper bound for $f(\sum_{i=1}^m a_i x_i)$ since $c \sum_{i=1}^m a_i (x_i - \bar{x})^2 \geq 0$. Since specially for $c = 0$ the strongly convexity reduces to the ordinary convexity, then (1.6) becomes (1.7).

Closely connected to Jensen’s inequality (1.7) is the Lah-Ribarič inequality

$$\sum_{i=1}^m a_i f(x_i) \leq \frac{\beta - \bar{x}}{\beta - \alpha} f(\alpha) + \frac{\bar{x} - \alpha}{\beta - \alpha} f(\beta) \tag{1.8}$$

which holds for every convex function $f : [\alpha, \beta] \rightarrow \mathbb{R}$ and $\mathbf{x} = (x_1, \dots, x_m) \in [\alpha, \beta]^m$, $\mathbf{a} = (a_1, \dots, a_m) \in [0, \infty)^m$ with $\sum_{i=1}^m a_i = 1$ and $\bar{x} = \sum_{i=1}^m a_i x_i$ (see [19]). The Lah-Ribarič inequality gives the upper bound for the term $\sum_{i=1}^m a_i f(x_i)$ and often called the converse Jensen inequality.

2. Preliminaries

For two vectors $\mathbf{x} = (x_1, \dots, x_m), \mathbf{y} = (y_1, \dots, y_m) \in [\alpha, \beta]^m$, let $x_{[i]}, y_{[i]}$ denote their increasing order. We say that \mathbf{x} majorizes \mathbf{y} or \mathbf{y} is majorized by \mathbf{x} and write

$$\mathbf{y} \prec \mathbf{x}$$

if

$$\sum_{i=1}^k y_{[i]} \leq \sum_{i=1}^k x_{[i]}, \quad k = 1, \dots, m, \tag{2.1}$$

with equality in (2.1) for $k = m$.

The term majorization is introduced in the space \mathbb{R}^m , in which the order is not defined, to compare and detect potential links between vectors. The majorization relation is reflexive and transitive but it is not antisymmetric (see [21, p. 79]) and hence is a preordering not a partial ordering. The majorization preorder on

vectors is known as vector majorization or classical majorization. This classical concept was initially studied by Hardy et al. [20]. A superb reference on the subject is [21].

It is well known that

$$\mathbf{y} \prec \mathbf{x} \quad \text{iff} \quad \mathbf{y} = \mathbf{x}\mathbf{A}$$

for some doubly stochastic matrix $\mathbf{A} = (a_{ij}) \in \mathcal{M}_{mm}(\mathbb{R})$, i.e. a matrix with nonnegative entries and rows and columns sums equal to 1.

Moreover, $\mathbf{y} \prec \mathbf{x}$ implies

$$\sum_{i=1}^m f(y_i) \leq \sum_{i=1}^m f(x_i)$$

for every continuous convex function $f : [\alpha, \beta] \rightarrow \mathbb{R}$. This result, obtained by Hardy et al. [20], is well known as majorization inequality and plays an important role in the study of majorization theory.

Sherman [37] considered the weighted concept of majorization between two vectors $\mathbf{x} = (x_1, \dots, x_m) \in [\alpha, \beta]^m$ and $\mathbf{y} = (y_1, \dots, y_l) \in [\alpha, \beta]^l$ with nonnegative weights $\mathbf{a} = (a_1, \dots, a_m)$ and $\mathbf{b} = (b_1, \dots, b_l)$. The concept of weighted majorization is defined by assumption of existence of row stochastic matrix $\mathbf{A} = (a_{ij}) \in \mathcal{M}_{lm}(\mathbb{R})$, i.e. matrix with nonnegative entries and rows sums equal to 1, such that

$$a_j = \sum_{i=1}^m b_j a_{ij}, \quad j = 1, \dots, l, \tag{2.2}$$

$$y_i = \sum_{j=1}^l x_j a_{ij}, \quad i = 1, \dots, m.$$

Sherman proved that under conditions (2.2), the inequality

$$\sum_{i=1}^m b_i f(y_i) \leq \sum_{j=1}^l a_j f(x_j) \tag{2.3}$$

holds for every convex function $f : [\alpha, \beta] \rightarrow \mathbb{R}$.

We can write the conditions (2.2) in the matrix form

$$\mathbf{a} = \mathbf{b}\mathbf{A} \quad \text{and} \quad \mathbf{y} = \mathbf{x}\mathbf{A}^T, \tag{2.4}$$

where \mathbf{A}^T denotes transpose matrix.

In the sequel, we write

$$(\mathbf{y}, \mathbf{b}) \prec (\mathbf{x}, \mathbf{a})$$

and say that a pair (\mathbf{y}, \mathbf{b}) is weighted majorized by (\mathbf{x}, \mathbf{a}) if vectors \mathbf{x}, \mathbf{y} and corresponding weights \mathbf{a}, \mathbf{b} satisfy conditions (2.2) for some row stochastic matrix \mathbf{A} .

Sherman's generalization contains Jensen's as well as majorization inequality as special cases as we pointed in the next remark.

Remark 2.1 a) For $m = 1$ and $\mathbf{b} = [1]$, Sherman's inequality (2.3) reduces to Jensen's inequality (2682).

b) For $m = l$ and $\mathbf{b} = \mathbf{e} = (1, \dots, 1)$, because $\mathbf{y} \prec \mathbf{x}$ gives $\mathbf{y} = \mathbf{x}\mathbf{A}^T$ with some doubly stochastic matrix \mathbf{A} and

$\mathbf{a} = \mathbf{bA} = \mathbf{e}$, from Sherman's inequality (2.3) we get majorization inequality

$$\sum_{i=1}^m f(y_i) \leq \sum_{i=1}^m f(x_i). \tag{2.5}$$

c) When $m = l$, and all weights b_i and a_j are equal, the condition $\mathbf{a} = \mathbf{bA}$ assures the stochasticity on columns, so in that case we deal with doubly stochastic matrices. Moreover, Sherman's inequality (2.3) reduces to

$$\sum_{i=1}^m a_i f(y_i) \leq \sum_{i=1}^m a_i f(x_i), \tag{2.6}$$

known as Fuchs' inequality (see [10]).

Recently, Sherman's result has attracted the interest of several mathematicians (see [1–5], [12–15], [23–30]).

This paper is organized as follows. In Section 3 we obtain the Lah-Ribarich inequality for strongly convex functions. We deal with Sherman's inequality and its converse for strongly convex function. As easy consequences, we get Jensen's and majorization inequalities and their conversions for strongly convex functions. In Section 4, we obtain some inequalities for generalized concept of f -divergence. In the last section, we extend Sherman's result to the class of strongly convex functions of higher order.

3. Sherman's type inequalities and conversions

We start with the Lah-Ribarich inequality for strongly convex functions.

Theorem 3.1 Let $\mathbf{x} = (x_1, \dots, x_l) \in [\alpha, \beta]^l$ and $\mathbf{a} = (a_1, \dots, a_l) \in [0, \infty)^l$ with $\sum_{j=1}^l a_j = 1$ and $\bar{x} = \sum_{j=1}^l a_j x_j$. If $f : [\alpha, \beta] \rightarrow \mathbb{R}$ is strongly convex with modulus $c > 0$, then

$$\sum_{j=1}^l a_j f(x_j) \leq \frac{\beta - \bar{x}}{\beta - \alpha} f(\alpha) + \frac{\bar{x} - \alpha}{\beta - \alpha} f(\beta) - c \sum_{j=1}^l a_j (\beta - x_j)(x_j - \alpha). \tag{3.1}$$

Proof Since for strongly convex function we have

$$f(z_1) \leq \frac{z_2 - z_1}{z_2 - z_0} f(z_0) + \frac{z_1 - z_0}{z_2 - z_0} f(z_2) - c(z_2 - z_1)(z_1 - z_0),$$

by substituting $z_1 = x_j$, $z_2 = \beta$ and $z_0 = \alpha$, we get

$$f(x_j) \leq \frac{\beta - x_j}{\beta - \alpha} f(\alpha) + \frac{x_j - \alpha}{\beta - \alpha} f(\beta) - c(\beta - x_j)(x_j - \alpha).$$

Now, multiplying with a_j and summing over j we have

$$\sum_{j=1}^l a_j f(x_j) \leq \frac{\beta - \sum_{j=1}^l a_j x_j}{\beta - \alpha} f(\alpha) + \frac{\sum_{j=1}^l a_j x_j - \alpha}{\beta - \alpha} f(\beta) - c \sum_{j=1}^l a_j (\beta - x_j)(x_j - \alpha)$$

what we need to prove. □

Now we give Sherman’s inequality for strongly convex functions.

Theorem 3.2 Let $\mathbf{x} = (x_1, \dots, x_l) \in [\alpha, \beta]^l$, $\mathbf{y} = (y_1, \dots, y_m) \in [\alpha, \beta]^m$, $\mathbf{a} = (a_1, \dots, a_l) \in [0, \infty)^l$, and $\mathbf{b} = (b_1, \dots, b_m) \in [0, \infty)^m$ be such that $(\mathbf{y}, \mathbf{b}) \prec (\mathbf{x}, \mathbf{a})$. Then for every $f : [\alpha, \beta] \rightarrow \mathbb{R}$ strongly convex with modulus $c > 0$, we have

$$\sum_{i=1}^m b_i f(y_i) \leq \sum_{j=1}^l a_j f(x_j) - c \left(\sum_{j=1}^l a_j x_j^2 - \sum_{i=1}^m b_i y_i^2 \right). \tag{3.2}$$

Proof Using (2.2) and applying (1.6), we have

$$\begin{aligned} \sum_{i=1}^m b_i f(y_i) &= \sum_{i=1}^m b_i f \left(\sum_{j=1}^l x_j a_{ij} \right) \\ &\leq \sum_{i=1}^m b_i \left(\sum_{j=1}^l a_{ij} f(x_j) - c \sum_{j=1}^l a_{ij} (x_j - y_i)^2 \right) \\ &= \sum_{j=1}^l a_j f(x_j) - c \sum_{i=1}^m b_i \sum_{j=1}^l a_{ij} (x_j - y_i)^2. \end{aligned} \tag{3.3}$$

By an easy calculation, we get

$$\begin{aligned} &\sum_{j=1}^l a_j f(x_j) - c \sum_{i=1}^m b_i \sum_{j=1}^l a_{ij} (x_j - y_i)^2 \\ &= \sum_{j=1}^l a_j f(x_j) - c \sum_{i=1}^m b_i \sum_{j=1}^l a_{ij} (x_j^2 - 2x_j y_i + y_i^2) \\ &= \sum_{j=1}^l a_j f(x_j) - c \left(\sum_{j=1}^l a_j x_j^2 - \sum_{i=1}^m b_i y_i^2 \right). \end{aligned} \tag{3.4}$$

Now, combining (3.3) and (3.4), we get (3.2). □

Remark 3.3 If we compare (3.2) with (2.3), note that the inequality (3.2) includes a better upper bound for $\sum_{i=1}^m b_i f(y_i)$ since $c \left(\sum_{j=1}^l a_j x_j^2 - \sum_{i=1}^m b_i y_i^2 \right) \geq 0$ because $t \mapsto t^2$ is convex function and then by Sherman’s inequality we have $\sum_{j=1}^l a_j x_j^2 - \sum_{i=1}^m b_i y_i^2 \geq 0$. Moreover, we get the double inequality

$$\begin{aligned} \sum_{i=1}^m b_i f(y_i) &\leq \sum_{j=1}^l a_j f(x_j) - c \left(\sum_{j=1}^l a_j x_j^2 - \sum_{i=1}^m b_i y_i^2 \right) \\ &\leq \sum_{j=1}^l a_j f(x_j). \end{aligned} \tag{3.5}$$

a) Specially, for $m = 1$ and $\mathbf{b} = (1)$, (3.5) becomes

$$\begin{aligned} f\left(\sum_{j=1}^l a_j x_j\right) &\leq \sum_{j=1}^l a_j f(x_j) - c\left(\sum_{j=1}^l a_j x_j^2 - \bar{x}^2\right) \\ &= \sum_{j=1}^l a_j f(x_j) - c\sum_{j=1}^l a_j (x_j - \bar{x})^2 \\ &\leq \sum_{j=1}^l a_j f(x_j), \end{aligned}$$

where $\bar{x} = \sum_{j=1}^l a_j x_j$, i.e. we get Jensen's inequality (1.6) for strongly convex function.

b) For $m = l$ and $\mathbf{b} = \mathbf{e} = (1, \dots, 1)$, (3.5) becomes

$$\begin{aligned} \sum_{i=1}^m f(y_i) &\leq \sum_{i=1}^m f(x_i) - c\left(\sum_{i=1}^m x_i^2 - \sum_{i=1}^m y_i^2\right) \\ &\leq \sum_{i=1}^m f(x_i), \end{aligned}$$

i.e. we get majorization inequality for strongly convex function.

c) When $m = l$, and all weights b_i and a_j are equal, then (3.2) becomes

$$\begin{aligned} \sum_{i=1}^m a_i f(y_i) &\leq \sum_{i=1}^m a_i f(x_i) - c\left(\sum_{i=1}^m a_i x_i^2 - \sum_{i=1}^m a_i y_i^2\right) \\ &\leq \sum_{i=1}^m a_i f(x_i), \end{aligned}$$

i.e. we get Fuchs' inequality for strongly convex function.

Next we give conversion to Sherman's inequality for strongly convex functions.

Theorem 3.4 Let $\mathbf{x} = (x_1, \dots, x_l) \in [\alpha, \beta]^l$, $\mathbf{y} = (y_1, \dots, y_m) \in [\alpha, \beta]^m$, $\mathbf{a} = (a_1, \dots, a_l) \in [0, \infty)^l$ and $\mathbf{b} = (b_1, \dots, b_m) \in [0, \infty)^m$ be such that $(\mathbf{y}, \mathbf{b}) \prec (\mathbf{x}, \mathbf{a})$. Let $B_m = \sum_{i=1}^m b_i$. If $f : [\alpha, \beta] \rightarrow \mathbb{R}$ is strongly convex with modulus $c > 0$, then

$$\begin{aligned} \sum_{j=1}^l a_j f(x_j) &\leq \frac{B_m \beta - \sum_{j=1}^l a_j x_j}{\beta - \alpha} f(\alpha) + \frac{\sum_{j=1}^l a_j x_j - B_m \alpha}{\beta - \alpha} f(\beta) \\ &\quad - c \sum_{j=1}^l a_j (\beta - x_j)(x_j - \alpha). \end{aligned} \tag{3.6}$$

Proof Using (2.2) we have

$$\sum_{j=1}^l a_j f(x_j) = \sum_{j=1}^l \left(\sum_{i=1}^m b_j a_{ij}\right) f(x_j) = \sum_{i=1}^m b_j \left(\sum_{j=1}^l a_{ij} f(x_j)\right). \tag{3.7}$$

Applying (3.1) we get

$$\sum_{j=1}^l a_{ij} f(x_j) \leq \frac{\beta - \sum_{j=1}^l a_{ij} x_j}{\beta - \alpha} f(\alpha) + \frac{\sum_{j=1}^l a_{ij} x_j - \alpha}{\beta - \alpha} f(\beta) - c \sum_{j=1}^l a_{ij} (\beta - x_j)(x_j - \alpha). \tag{3.8}$$

Now, combining (3.7) and (3.8), we get (3.6). □

Remark 3.5 a) Specially, if $m = 1$ and $\mathbf{b} = (1)$, then (3.5) and (3.6) gives the following series of inequalities

$$\begin{aligned} f\left(\sum_{j=1}^l a_j x_j\right) &\leq \sum_{j=1}^l a_j f(x_j) - c \sum_{j=1}^l a_j (x_j - \bar{x})^2 \\ &\leq \sum_{j=1}^l a_j f(x_j) \\ &\leq \frac{\beta - \sum_{j=1}^l a_j x_j}{\beta - \alpha} f(\alpha) + \frac{\sum_{j=1}^l a_j x_j - \alpha}{\beta - \alpha} f(\beta) \\ &\quad - c \sum_{j=1}^l a_j (\beta - x_j)(x_j - \alpha), \end{aligned}$$

i.e. we get Jensen's inequality and its conversion for strongly convex functions.

b) If $m = l$ and $\mathbf{b} = \mathbf{e} = (1, \dots, 1)$, then (3.5) and (3.6) gives

$$\begin{aligned} \sum_{i=1}^m f(y_i) &\leq \sum_{i=1}^m f(x_i) - c \left(\sum_{i=1}^m x_i^2 - \sum_{i=1}^m y_i^2 \right) \\ &\leq \sum_{i=1}^m f(x_i) \\ &\leq \frac{\beta - \sum_{j=1}^l x_j}{\beta - \alpha} f(\alpha) + \frac{\sum_{j=1}^l x_j - \alpha}{\beta - \alpha} f(\beta) \\ &\quad - c \sum_{j=1}^l (\beta - x_j)(x_j - \alpha), \end{aligned}$$

i.e. we get majorization inequality and its conversion for strongly convex functions.

c) If $m = l$, and all weights b_i and a_j are equal, then (3.5) and (3.6) gives

$$\begin{aligned} \sum_{i=1}^m a_i f(y_i) &\leq \sum_{i=1}^m a_i f(x_i) - c \left(\sum_{i=1}^m a_i x_i^2 - \sum_{i=1}^m a_i y_i^2 \right) \\ &\leq \sum_{i=1}^m a_i f(x_i) \\ &\leq \frac{A_m \beta - \sum_{i=1}^m a_i x_i}{\beta - \alpha} f(\alpha) + \frac{\sum_{i=1}^m a_i x_i - A_m \alpha}{\beta - \alpha} f(\beta) \\ &\quad - c \sum_{i=1}^m a_i (\beta - x_i)(x_i - \alpha), \end{aligned}$$

where $\sum_{i=1}^m a_i = A_m$, i.e. we get Fuchs' inequality and its conversion for strongly convex functions.

4. Applications to f -divergences

Shannon [36] introduced a statistical concept of entropy in the theory of communication and transmission of information, the measure of information defined by

$$H(\mathbf{p}) = \sum_{i=1}^n p_i \ln \frac{1}{p_i}, \tag{4.1}$$

where $\mathbf{p} = (p_1, \dots, p_n)$ is a positive probability distribution, i.e. $p_i > 0$, $i = 1, \dots, n$, with $\sum_{i=1}^n p_i = 1$, for some discrete random variable X . It satisfied the estimation:

$$0 \leq H(\mathbf{p}) \leq \ln n.$$

Shannon's entropy quantifies the unevenness in the probability distribution \mathbf{p} .

As a slight modification of the previous formula, we get the Kullback–Leibler divergence [18] or relative entropy of \mathbf{q} with respect to \mathbf{p} defined by

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

It is a measure of the difference between two positive probability distributions \mathbf{q} and \mathbf{p} over the same variable. In statistics, it arises as the expected logarithm of difference between the probability \mathbf{q} of data in the original distribution with the approximating distribution \mathbf{p} . It satisfies the following estimates

$$KL(\mathbf{q}, \mathbf{p}) \geq 0.$$

The previous two concepts we can get as special cases of the Csiszár f -divergence functional

$$D_f(\mathbf{q}, \mathbf{p}) = \sum_{i=1}^n p_i f \left(\frac{q_i}{p_i} \right), \tag{4.2}$$

where $f : (0, \infty) \rightarrow \mathbb{R}$ is a convex function and $\mathbf{p} = (p_1, \dots, p_n)$, $\mathbf{q} = (q_1, \dots, q_n)$ with $p_i, q_i > 0$, $i = 1, \dots, n$ (see [6], [7]).

Note that

$$H(\mathbf{p}) = -\sum_{i=1}^n p_i \ln p_i = -D_f(\mathbf{e}, \mathbf{p}), \quad f(t) = -\ln t,$$

$$D_{KL}(\mathbf{q}, \mathbf{p}) = \sum_{i=1}^m q_i \ln \frac{q_i}{p_i} = D_f(\mathbf{q}, \mathbf{p}), \quad f(t) = t \ln t.$$

Csiszár with Körner [7] proved Jensen’s inequality for the f -divergence functional as follows

$$\sum_{i=1}^n q_i f\left(\frac{\sum_{i=1}^n p_i}{\sum_{i=1}^n q_i}\right) \leq D_f(\mathbf{q}, \mathbf{p}). \tag{CK}$$

Specially, if f is normalized, i.e. $f(1) = 0$ and $\sum_{i=1}^n p_i = \sum_{i=1}^n q_i$, then

$$0 \leq D_f(\mathbf{q}, \mathbf{p}). \tag{4.3}$$

Csiszár f -divergence functional (4.2) is widely employed in different scientific fields among which we point out mathematical statistics and specially information theory with deep connections in topics as diverse as artificial intelligence, statistical physics, and biological evolution. For suitable choices of the kernel f , the general aspect of the Csiszár f -divergence functional (4.2) can be interpreted as a series of the well-known divergencies (see [8, 16, 17]). Here we give some examples:

- Hellinger divergence

$$h^2(\mathbf{q}, \mathbf{p}) = \frac{1}{2} \sum_{i=1}^n (\sqrt{p_i} - \sqrt{q_i})^2, \quad f(t) = \frac{1}{2} (\sqrt{t} - 1)^2,$$

- Variational distance

$$V(\mathbf{q}, \mathbf{p}) = \sum_{i=1}^n |p_i - q_i|, \quad f(t) = |t - 1|,$$

- Harmonic divergence

$$D_H(\mathbf{q}, \mathbf{p}) = \sum_{i=1}^n \frac{2p_i q_i}{p_i + q_i}, \quad f(t) = \frac{2t}{1+t},$$

- Bhattacharya distance

$$D_B(\mathbf{q}, \mathbf{p}) = -D_f(\mathbf{q}, \mathbf{p}) = \sum_{i=1}^n \sqrt{p_i q_i}, \quad f(t) = -\sqrt{t},$$

- Triangular discrimination

$$D_T(\mathbf{q}, \mathbf{p}) = \sum_{i=1}^n \frac{(p_i - q_i)^2}{p_i + q_i}, \quad f(t) = \frac{(t-1)^2}{t+1},$$

- Chi square distance

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

- Rényi α -order entropy ($\alpha > 1$)

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

We extend definition of f -divergence functional (4.2) as follows.

Definition 4.1 Let $f : (0, \infty) \rightarrow \mathbb{R}$ be a strongly convex function with modulus $c > 0$ and $\mathbf{p} = (p_1, \dots, p_n)$, $\mathbf{q} = (q_1, \dots, q_n)$ with $p_i, q_i > 0$, $i = 1, \dots, n$. We define

$$\tilde{D}_f(\mathbf{q}, \mathbf{p}) = \sum_{i=1}^n p_i f\left(\frac{q_i}{p_i}\right). \tag{4.4}$$

In this section our intention is to derive mutual bounds for the generalized f -divergence functional (4.4). We obtain some reverse relations for the generalized f -divergence functional that correspond to the class of strongly convex functions.

Through the rest of the paper we always assume that $\alpha, \beta > 0$.

Corollary 4.2 Let $\mathbf{p} = (p_1, \dots, p_l) \in [\alpha, \beta]^l$, $\mathbf{q} = (q_1, \dots, q_l) \in [\alpha, \beta]^l$, and $\mathbf{R} = (r_{ij}) \in \mathcal{M}_{ml}(\mathbb{R})$ be column stochastic matrix. Let us define $\langle \mathbf{p}, \mathbf{r}_i \rangle = \sum_{j=1}^l p_j r_{ij} > 0$, $\langle \mathbf{q}, \mathbf{r}_i \rangle = \sum_{j=1}^l q_j r_{ij}$, $i = 1, \dots, m$. Then for every $f : [\alpha, \beta] \rightarrow \mathbb{R}$ strongly convex with modulus $c > 0$, we have

$$\begin{aligned} \sum_{i=1}^m \langle \mathbf{p}, \mathbf{r}_i \rangle f\left(\frac{\langle \mathbf{q}, \mathbf{r}_i \rangle}{\langle \mathbf{p}, \mathbf{r}_i \rangle}\right) &\leq \tilde{D}_f(\mathbf{q}, \mathbf{p}) - c \left(\sum_{j=1}^l \frac{q_j^2}{p_j} - \sum_{i=1}^m \frac{\langle \mathbf{q}, \mathbf{r}_i \rangle^2}{\langle \mathbf{p}, \mathbf{r}_i \rangle} \right) \\ &\leq \tilde{D}_f(\mathbf{q}, \mathbf{p}) \\ &\leq \frac{\sum_{i=1}^m \langle \mathbf{p}, \mathbf{r}_i \rangle \beta - \sum_{j=1}^l q_j}{\beta - \alpha} f(\alpha) + \frac{\sum_{j=1}^l q_j - \sum_{i=1}^m \langle \mathbf{p}, \mathbf{r}_i \rangle \alpha}{\beta - \alpha} f(\beta) \\ &\quad - c \sum_{j=1}^l p_j \left(\beta - \frac{q_j}{p_j} \right) \left(\frac{q_j}{p_j} - \alpha \right). \end{aligned} \tag{4.5}$$

Proof Let us consider $\mathbf{x} = (x_1, \dots, x_l)$ and $\mathbf{y} = (y_1, \dots, y_m)$, such that $x_j = \frac{q_j}{p_j}$, $j = 1, \dots, l$ and $y_i = \frac{\langle \mathbf{q}, \mathbf{r}_i \rangle}{\langle \mathbf{p}, \mathbf{r}_i \rangle}$, $i = 1, \dots, m$. Let $a_j = \sum_{i=1}^m b_i \frac{p_j r_{ij}}{\langle \mathbf{p}, \mathbf{r}_i \rangle}$, $j = 1, \dots, m$, where $b_i = \langle \mathbf{p}, \mathbf{r}_i \rangle$, $i = 1, \dots, m$.

We have

$$\frac{\langle \mathbf{q}, \mathbf{r}_i \rangle}{\langle \mathbf{p}, \mathbf{r}_i \rangle} = \frac{\sum_{j=1}^l q_j r_{ij}}{\sum_{j=1}^l p_j r_{ij}} = \frac{p_1 r_{i1}}{\sum_{j=1}^l p_j r_{ij}} \frac{q_1}{p_1} + \dots + \frac{p_l r_{il}}{\sum_{j=1}^l p_j r_{ij}} \frac{q_l}{p_l}, \quad i = 1, \dots, m.$$

Moreover, the following identity

$$\left(\frac{\langle \mathbf{q}, \mathbf{r}_1 \rangle}{\langle \mathbf{p}, \mathbf{r}_1 \rangle}, \dots, \frac{\langle \mathbf{q}, \mathbf{r}_m \rangle}{\langle \mathbf{p}, \mathbf{r}_m \rangle} \right) = \left(\frac{q_1}{p_1}, \dots, \frac{q_l}{p_l} \right) \cdot \begin{pmatrix} \frac{p_1 r_{11}}{\langle \mathbf{p}, \mathbf{r}_1 \rangle} & \dots & \frac{p_1 r_{m1}}{\langle \mathbf{p}, \mathbf{r}_m \rangle} \\ \vdots & \ddots & \vdots \\ \frac{p_l r_{1l}}{\langle \mathbf{p}, \mathbf{r}_1 \rangle} & \dots & \frac{p_l r_{ml}}{\langle \mathbf{p}, \mathbf{r}_m \rangle} \end{pmatrix}$$

holds for some row stochastic matrix $\mathbf{A} = (a_{ij}) \in \mathcal{M}_{ml}(\mathbb{R})$, with $a_{ij} = \frac{p_j r_{ij}}{\langle \mathbf{p}, \mathbf{r}_i \rangle}$, $i = 1, \dots, m$, $j = 1, \dots, l$.

Therefore, $\mathbf{y} = \mathbf{x}\mathbf{A}^T$ holds.

Furthermore, we have

$$a_j = \sum_{i=1}^m \langle \mathbf{p}, \mathbf{r}_i \rangle \frac{p_j r_{ij}}{\langle \mathbf{p}, \mathbf{r}_i \rangle} = p_j \sum_{i=1}^m r_{ij} = p_j, j = 1, \dots, l,$$

i.e. $\mathbf{a} = \mathbf{b}\mathbf{A}$. Therefore, the assumptions of Theorem 3.2 and Theorem 3.4 are fulfilled. Now applying (3.2) and (3.6), we get

$$\begin{aligned} \sum_{i=1}^m \langle \mathbf{p}, \mathbf{r}_i \rangle f \left(\frac{\langle \mathbf{q}, \mathbf{r}_i \rangle}{\langle \mathbf{p}, \mathbf{r}_i \rangle} \right) &\leq \sum_{j=1}^l p_j f \left(\frac{q_j}{p_j} \right) - c \left(\sum_{j=1}^l \frac{q_j^2}{p_j} - \sum_{i=1}^m \frac{\langle \mathbf{q}, \mathbf{r}_i \rangle^2}{\langle \mathbf{p}, \mathbf{r}_i \rangle} \right) \\ &\leq \sum_{j=1}^l p_j f \left(\frac{q_j}{p_j} \right) \\ &\leq \frac{\sum_{i=1}^m \langle \mathbf{p}, \mathbf{r}_i \rangle \beta - \sum_{j=1}^l q_j}{\beta - \alpha} f(\alpha) + \frac{\sum_{j=1}^l q_j - \sum_{i=1}^m \langle \mathbf{p}, \mathbf{r}_i \rangle \alpha}{\beta - \alpha} f(\beta) \\ &\quad - c \sum_{j=1}^l p_j \left(\beta - \frac{q_j}{p_j} \right) \left(\frac{q_j}{p_j} - \alpha \right), \end{aligned}$$

which is equivalent to (4.5). □

Specially, for $m = 1$, the previous result reduces to the next corollary.

Corollary 4.3 *Let $\mathbf{p} = (p_1, \dots, p_l) \in [\alpha, \beta]^l$, $\mathbf{q} = (q_1, \dots, q_l) \in [\alpha, \beta]^l$ and $\mathbf{r} = (r_1, \dots, r_l) \in [\alpha, \beta]^l$. Let us define $\langle \mathbf{p}, \mathbf{r} \rangle = \sum_{j=1}^l p_j r_j > 0$, $\langle \mathbf{q}, \mathbf{r} \rangle = \sum_{j=1}^l q_j r_j$. Then for every $f : [\alpha, \beta] \rightarrow \mathbb{R}$ strongly convex with modulus $c > 0$, we have*

$$\begin{aligned} \langle \mathbf{p}, \mathbf{r} \rangle f \left(\frac{\langle \mathbf{q}, \mathbf{r} \rangle}{\langle \mathbf{p}, \mathbf{r} \rangle} \right) &\leq \tilde{D}_f(\mathbf{q}, \mathbf{p}) - c \left(\sum_{j=1}^l \frac{q_j^2}{p_j} - \frac{\langle \mathbf{q}, \mathbf{r} \rangle^2}{\langle \mathbf{p}, \mathbf{r} \rangle} \right) \\ &\leq \tilde{D}_f(\mathbf{q}, \mathbf{p}) \\ &\leq \frac{\langle \mathbf{p}, \mathbf{r} \rangle \beta - \sum_{j=1}^l q_j}{\beta - \alpha} f(\alpha) + \frac{\sum_{j=1}^l q_j - \langle \mathbf{p}, \mathbf{r} \rangle \alpha}{\beta - \alpha} f(\beta) \\ &\quad - c \sum_{j=1}^l p_j \left(\beta - \frac{q_j}{p_j} \right) \left(\frac{q_j}{p_j} - \alpha \right). \end{aligned}$$

If in addition $\mathbf{r} = \mathbf{e} = (1, \dots, 1)$, then

$$\begin{aligned} \sum_{j=1}^l p_j f\left(\frac{\sum_{j=1}^l q_j}{\sum_{j=1}^l p_j}\right) &\leq \tilde{D}_f(\mathbf{q}, \mathbf{p}) - c \left(\sum_{j=1}^l \frac{q_j^2}{p_j} - \frac{\left(\sum_{j=1}^l q_j\right)^2}{\sum_{j=1}^l p_j} \right) \\ &\leq \tilde{D}_f(\mathbf{q}, \mathbf{p}) \\ &\leq \frac{\sum_{j=1}^l p_j \beta - \sum_{j=1}^l q_j}{\beta - \alpha} f(\alpha) + \frac{\sum_{j=1}^l q_j - \sum_{j=1}^l p_j \alpha}{\beta - \alpha} f(\beta) \\ &\quad - c \sum_{j=1}^l p_j \left(\beta - \frac{q_j}{p_j}\right) \left(\frac{q_j}{p_j} - \alpha\right). \end{aligned}$$

Moreover, if f is normalized, i.e. $f(1) = 0$ and $\sum_{j=1}^l p_j = \sum_{j=1}^l q_j$, we get

$$\begin{aligned} 0 &\leq \tilde{D}_f(\mathbf{q}, \mathbf{p}) - c \left(\sum_{j=1}^l \frac{q_j^2}{p_j} - \sum_{j=1}^l p_j \right) \\ &\leq \tilde{D}_f(\mathbf{q}, \mathbf{p}) \\ &\leq \frac{\sum_{j=1}^l p_j (\beta - 1)}{\beta - \alpha} f(\alpha) + \frac{\sum_{j=1}^l p_j (1 - \alpha)}{\beta - \alpha} f(\beta) - c \sum_{j=1}^l p_j \left(\beta - \frac{q_j}{p_j}\right) \left(\frac{q_j}{p_j} - \alpha\right). \end{aligned}$$

5. Generalization of Sherman’s inequality for strongly n -convex function

The technique that we use in this section is based on an application of Fink’s identity [9]

$$\begin{aligned} f(x) &= \frac{n}{\beta - \alpha} \int_{\alpha}^{\beta} f(t) dt - \sum_{w=1}^{n-1} \frac{n-w}{w!} \cdot \frac{f^{(w-1)}(\alpha)(x-\alpha)^w - f^{(w-1)}(\beta)(x-\beta)^w}{\beta - \alpha} \\ &\quad + \frac{1}{(n-1)!(\beta - \alpha)} \int_{\alpha}^{\beta} (x-t)^{n-1} k(t, x) f^{(n)}(t) dt, \end{aligned} \tag{5.1}$$

where

$$k(t, x) = \begin{cases} t - \alpha, & \alpha \leq t \leq x \leq \beta \\ t - \beta, & \alpha \leq x < t \leq \beta \end{cases}, \tag{5.2}$$

which holds for every $f : [\alpha, \beta] \rightarrow \mathbb{R}$ such that $f^{(n-1)}$ is absolutely continuous for some $n \geq 1$. The sum in (5.1) is zero when $n = 1$.

We start with an identity which is very useful for us to obtain generalizations.

Theorem 5.1 Let $\mathbf{x} = (x_1, \dots, x_l) \in [\alpha, \beta]^l$, $\mathbf{y} = (y_1, \dots, y_m) \in [\alpha, \beta]^m$, $\mathbf{a} = (a_1, \dots, a_l) \in [0, \infty)^l$, and $\mathbf{b} = (b_1, \dots, b_m) \in [0, \infty)^m$ be such that $(\mathbf{y}, \mathbf{b}) \prec (\mathbf{x}, \mathbf{a})$. Let $k(t, \cdot)$ be defined as in (5.2). Then for every

$f : [\alpha, \beta] \rightarrow \mathbb{R}$, such that $f^{(n-1)}$ is absolutely continuous on $[\alpha, \beta]$, we have

$$\begin{aligned} & \sum_{j=1}^l a_j f(x_j) - \sum_{i=1}^m b_i f(y_i) \\ &= \frac{1}{\beta - \alpha} \sum_{w=2}^{n-1} \frac{n-w}{w!} \cdot f^{(w-1)}(\beta) \left(\sum_{j=1}^l a_j (x_j - \beta)^w - \sum_{i=1}^m b_i (y_i - \beta)^w \right) \\ & - \frac{1}{\beta - \alpha} \sum_{w=2}^{n-1} \frac{n-w}{w!} \cdot f^{(w-1)}(\alpha) \left(\sum_{j=1}^l a_j (x_j - \alpha)^w - \sum_{i=1}^m b_i (y_i - \alpha)^w \right) \\ & + \frac{1}{(n-1)!(\beta - \alpha)} \times \\ & \int_{\alpha}^{\beta} \left[\sum_{j=1}^l a_j (x_j - t)^{n-1} k(t, x_j) - \sum_{i=1}^m b_i (y_i - t)^{n-1} k(t, y_i) \right] f^{(n)}(t) dt. \end{aligned} \tag{5.3}$$

Proof Applying (5.1) to the Sherman difference $\sum_{j=1}^l a_j f(x_j) - \sum_{i=1}^m b_i f(y_i)$, we get (5.3). □

Theorem 5.2 Let all the assumptions of Theorem 5.1 be satisfied. Additionally, let f be n -strongly convex with modulus $c > 0$. If

$$\sum_{j=1}^l a_j (x_j - t)^{n-1} k(t, x_j) - \sum_{i=1}^m b_i (y_i - t)^{n-1} k(t, y_i) \geq 0, \quad \alpha \leq t \leq \beta, \tag{5.4}$$

then

$$\begin{aligned} & \sum_{j=1}^l a_j f(x_j) - \sum_{i=1}^m b_i f(y_i) - c \left(\sum_{j=1}^l a_j x_j^n - \sum_{i=1}^m b_i y_i^n \right) \\ & \geq \frac{1}{\beta - \alpha} \sum_{w=1}^{n-1} \frac{n-w}{w!} \\ & \cdot \left[f^{(w-1)}(\beta) - cn(n-1)\dots(n-w+2)\beta^{n-w+2} \right] \left(\sum_{j=1}^l a_j (x_j - \beta)^w - \sum_{i=1}^m b_i (y_i - \beta)^w \right) \\ & - \frac{1}{\beta - \alpha} \sum_{w=1}^{n-1} \frac{n-w}{w!} \\ & \cdot \left[f^{(w-1)}(\alpha) - cn(n-1)\dots(n-w+2)\alpha^{n-w+2} \right] \left(\sum_{j=1}^l a_j (x_j - \alpha)^w - \sum_{i=1}^m b_i (y_i - \alpha)^w \right). \end{aligned} \tag{5.5}$$

If the reverse inequality in (5.4) holds, then the reverse inequality in (5.5) holds.

Proof Let us consider the function $g(x) = f(x) - cx^n$. Since f is strongly n -convex with modulus c , then g is n -convex. We may assume without loss of generality that f and g are n -times differentiable and $g^{(n)} \geq 0$

on $[\alpha, \beta]$ (see [33, p. 16]).

Applying (5.5) to g , we have

$$\begin{aligned} & \sum_{j=1}^l a_j g(x_j) - \sum_{i=1}^m b_i g(y_i) \tag{5.6} \\ &= \frac{1}{\beta - \alpha} \sum_{w=2}^{n-1} \frac{n-w}{w!} \cdot g^{(w-1)}(\beta) \left(\sum_{j=1}^l a_j (x_j - \beta)^w - \sum_{i=1}^m b_i (y_i - \beta)^w \right) \\ & - \frac{1}{\beta - \alpha} \sum_{w=2}^{n-1} \frac{n-w}{w!} \cdot g^{(w-1)}(\alpha) \left(\sum_{j=1}^l a_j (x_j - \alpha)^w - \sum_{i=1}^m b_i (y_i - \alpha)^w \right) \\ & + \frac{1}{(n-1)!(\beta - \alpha)} \times \\ & \int_{\alpha}^{\beta} \left[\sum_{j=1}^l a_j (x_j - t)^{n-1} k(t, x_j) - \sum_{i=1}^m b_i (y_i - t)^{n-1} k(t, y_i) \right] g^{(n)}(t) dt. \end{aligned}$$

Moreover, if (5.4) holds, then

$$\begin{aligned} & \sum_{j=1}^l a_j g(x_j) - \sum_{i=1}^m b_i g(y_i) \tag{5.7} \\ & \geq \frac{1}{\beta - \alpha} \sum_{w=2}^{n-1} \frac{n-w}{w!} \cdot g^{(w-1)}(\beta) \left(\sum_{j=1}^l a_j (x_j - \beta)^w - \sum_{i=1}^m b_i (y_i - \beta)^w \right) \\ & - \frac{1}{\beta - \alpha} \sum_{w=2}^{n-1} \frac{n-w}{w!} \cdot g^{(w-1)}(\alpha) \left(\sum_{j=1}^l a_j (x_j - \alpha)^w - \sum_{i=1}^m b_i (y_i - \alpha)^w \right) \end{aligned}$$

which is equivalent to (5.5).

If the reverse inequality in (5.4) holds, then the last term in (5.6) is nonpositive and then the reverse inequality in (5.7) holds. This ends the proof. \square

Remark 5.3 Consider the function $s : [\alpha, \beta] \rightarrow \mathbb{R}$ defined by

$$s(x) = (x - t)^{n-1} k(t, x) = \begin{cases} (x - t)^{n-1} (t - \alpha), & \alpha \leq t \leq x \leq \beta \\ (x - t)^{n-1} (t - \beta), & \alpha \leq x < t \leq \beta \end{cases}.$$

We have

$$s''(x) = \begin{cases} (n-1)(n-2)(x-t)^{n-3}(t-\alpha), & \alpha \leq t \leq x \leq \beta \\ (n-1)(n-2)(x-t)^{n-3}(t-\beta), & \alpha \leq x < t \leq \beta \end{cases}.$$

Then for even n , the function s is convex and by Sherman's theorem, we have

$$\sum_{j=1}^l a_j (x_j - t)^{n-1} k(t, x_j) - \sum_{i=1}^m b_i (y_i - t)^{n-1} k(t, y_i) \geq 0,$$

i.e. the assumption (5.4) is immediately satisfied. Therefore, by Theorem 5.2, the inequality (5.5) holds. Specially, for $n = 2$, the inequality (5.5) reduces to (3.2).

Acknowledgment

This research is partially supported through project KK.01.1.1.02.0027, a project co-financed by the Croatian Government and the European Union through the European Regional Development Fund - the Competitiveness and Cohesion Operational Programme.

References

- [1] Adil Khan M, Ivelić Bradanović S, Pečarić J. Generalizations of Sherman's inequality by Hermite's interpolating polynomial. *Mathematical Inequalities & Applications* 2016; 19 (4): 1181-1192. doi: 10.7153/mia-19-87
- [2] Adil Khan M, Ivelić Bradanović S, Pečarić J. Generalizations of Sherman's inequality by Hermite's interpolating polynomial and Green function. *Konuralp Journal of Mathematics* 2016; 4 (2): 255-270.
- [3] Agarwal Ravi P, Ivelić Bradanović S, Pečarić J. Generalizations of Sherman's inequality by Lidstone's interpolating polynomial. *Journal of Inequalities and Applications* 2016; 6 (2016): doi: 10.1186/s13660-015-0935-6
- [4] Borcea J. Equilibrium points of logarithmic potentials. *Transactions of the American Mathematical Society* 2007; 359: 3209-3237. doi: 10.1090/S0002-9947-07-04251-1
- [5] Burtea AM. Two examples of weighted majorization. *Annals of the University of Craiova, Mathematics and Computer Science* 2010; 37 (2): 92-99.
- [6] Csiszár I. Information-type measures of difference of probability functions and indirect observations. *Studia Scientiarum Mathematicarum Hungarica* 1967; 2: 299-318.
- [7] Csiszár I, Körner J. *Information Theory: Coding Theorem for Discrete Memoryless Systems*. NY, USA: Academic Press, 1981.
- [8] Dragomir SS. Upper and lower bounds for Csiszár f -divergence in terms of the Kullback-Leibler divergence and applications, *Inequalities for Csiszár f -Divergence in Information Theory*. RGMIA Monographs, Australia: Victoria University, 2000.
- [9] Fink AM. Bounds of the deviation of a function from its averages. *Czechoslovak Mathematical Journal* 1992; 42 (117): 289-310.
- [10] Fuchs L. A new proof of an inequality of Hardy-Littlewood-Pólya. *Mathematisk Tidsskrift B* 1947; (1947): 53-54.
- [11] Gera R, Nikodem K. Strongly convex functions of higher order. *Advances in Nonlinear Analysis* 2011; 74: 661-665.
- [12] Ivelić Bradanović S, Latif N, Pečarić J. On an upper bound for Sherman's inequality. *Journal of Inequalities and Applications* 2016; (2016): doi: 10.1186/s13660-016-1091-3
- [13] Ivelić Bradanović S, Latif N, Pečarić Đ, Pečarić J. Sherman's and related inequalities with applications in information theory. *Journal of Inequalities and Applications* 2018; (2018): doi: 10.1186/s13660-018-1692-0
- [14] Ivelić Bradanović S, Pečarić J. Extensions and improvements of Sherman's and related inequalities for n -convex functions. *Open Mathematics* 2017; 15 (1): doi: 10.1515/math-2017-0077
- [15] Ivelić Bradanović S, Pečarić J. Generalizations of Sherman's inequality. *Periodica Mathematica Hungarica* 2017; 74 (2): doi: 10.1007/s10998-016-0154-z
- [16] Kapur JN. A comparative assessment of various measures of directed divergence. *Advances in Management Studies* 1984; 3 (1): 1-16. doi: 10.1080/02522667.1983.10698762
- [17] Kapur JN. *Maximum-Entropy Models in Science and Engineering*. NY, USA: John Wiley and Sons, 1989. doi: 10.2307/2532770

- [18] Kullback S. Information Theory and Statistics. NY, USA: Wiley, 1959.
- [19] Lah P, Ribarič M. Converse of Jensen's inequality for convex functions. Publikacije Elektrotehničkog fakulteta Univerziteta u Beogradu 1973; (412-460): 201-205.
- [20] Hardy GH, Littlewood JE, Pólya G. Inequalities. 2nd ed. Cambridge, UK: Cambridge University Press, 1952.
- [21] Marshall AW, Olkin I, Barry CA. Inequalities: Theory of Majorization and its Applications. 2nd ed. NY, USA: Springer Series in Statistics, 2011. doi: 10.2307/41305043
- [22] Merentes N, Nikodem K. Remarks on strongly convex functions. Aequationes Mathematicae 2010; 80: 193-199. doi: 10.1007/s00010-010-0043-0
- [23] Niculescu CP. Choquet theory for signed measures. Mathematical Inequalities & Applications 2002; 5: 479-489. doi: 10.7153/mia-05-47
- [24] Niculescu CP, Persson LE. Old and new on the Hermite-Hadamard inequality. Real Analysis Exchange 2004; 29 (2): doi: 10.14321/realanalexch.29.2.0663
- [25] Niculescu CP, Roventă I. An approach of majorization in spaces with a curved geometry. Journal of Mathematical Analysis and Applications 2014; 411: 119-128. doi: 10.1016/j.jmaa.2013.09.038
- [26] Niculescu CP, Roventă I. Relative convexity and its applications. Aequationes Mathematicae 2015; 89: 1389-1400. doi: 10.1007/s00010-014-0319-x
- [27] Niezgoda M. Nonlinear Sherman-type inequalities. Advances in Nonlinear Analysis 2020; 9 (1): doi: 10.1515/anona-2018-0098
- [28] Niezgoda M. Remarks on Sherman like inequalities for (α, β) -convex functions. Mathematical Inequalities & Applications 2014; 17 (4): 1579-1590. doi: 10.7153/mia-17-116
- [29] Niezgoda M. Remarks on convex functions and separable sequences. Discrete Mathematics 2008; 308: 1765-1773. doi: 10.1016/j.disc.2007.04.023
- [30] Niezgoda M. Vector joint majorization and generalization of Csiszar-Korner's inequality for f-divergence. Discrete Applied Mathematics 2016; 198: 195-205. doi: 10.1016/j.dam.2015.06.018
- [31] Nikodem K. On Strongly Convex Functions and Related Classes of Functions. Handbook of Functional Equations. NY, USA: Springer, 2014, 365-405. doi: 10.1007/978-1-4939-1246-9_16
- [32] Nikodem K, Páles Z. Characterizations of inner product spaces by strongly convex functions. Banach Journal of Mathematical Analysis 2011; 5 (1): 83-87. doi: 10.15352/bjma/1313362982
- [33] Pečarić J, Proschan F, Tong YL. Convex functions, Partial Orderings and Statistical Applications. NY, USA: Academic Press, 1992.
- [34] Polyak BT. Existence theorems and convergence of minimizing sequences in extremum problems with restrictions. Soviet mathematics-Doklady 1966; 7: 72-75.
- [35] Popoviciu T. Les Fonctions Convexes. Paris, France: Hermann et Cie, 1944 (in French).
- [36] Shannon CE, Weaver W. The Mathematical Theory of Communication. Urbana, USA: University of Illinois Press, 1949.
- [37] Sherman S. On a theorem of Hardy, Littlewood, Pólya and Blackwell. Proceedings of the National Academy of Sciences of the United States of America 1957; 37 (1): 826-831. doi: 10.1073/pnas.37.12.826