Mathematical
Inequalities
& Applications
Volume 26, Number 1 (2023), 39–45 doi:10.7153/mia-2023-26-04

ON GENERALIZED CSISZÁR (f,g)–DIVERGENCE WITH AN APPLICATION FOR p–MAJORIZATION

MAREK NIEZGODA

(*Communicated by M. Praljak*)

Abstract. In this note, we develop some ideas from [8]. We introduce and investigate generalized Csiszár (f,g)-divergence generated by a convex function f and a concave function g. We derive a Csiszár-Körner type inequality for such (f,g)-divergences. We also study some special cases of the obtained inequality. In particular, we give a result for $\mathbf{p}$-majorization.

1. Introduction

Given a convex function $f:\mathbb{R}_+\to\mathbb{R}$ and vectors $\mathbf{p}=(p_1,\dots,p_n)\in\mathbb{R}_{++}$ and $\mathbf{q}=(q_1,\dots,q_n)\in\mathbb{R}_+$, the *Csiszár f-divergence* is defined by

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

(see [1, 2, 3]).

The *Csiszár-Körner inequality* says that

$$\sum_{i=1}^n p_i f\left(\frac{\sum_{i=1}^n q_i}{\sum_{i=1}^n p_i}\right)\leqslant C_f(\mathbf{p},\mathbf{q}) \tag{2}$$

(see [2, 6]). In the special case $\sum_{i=1}^n q_i=\sum_{i=1}^n p_i$ with $f(1)=0$, inequality (2) implies that

$$0\leqslant C_f(\mathbf{p},\mathbf{q}). \tag{3}$$

As noted in [4], definition (1) can be extended as follows.

Let $f:\mathbb{R}_+\to\mathbb{R}$ be a convex function on $\mathbb{R}_+$, and $\mathbf{p}=(p_1,\dots,p_n)\in\mathbb{R}^n_{++}$, $\mathbf{q}=(q_1,\dots,q_n)\in\mathbb{R}^n_+$, $\mathbf{r}=(r_1,\dots,r_n)\in\mathbb{R}^n_+$. Then the *generalized Csiszár f-divergence* is defined by

$$C_f(\mathbf{p},\mathbf{q};\mathbf{r})=\sum_{i=1}^n r_i p_i f\left(\frac{q_i}{p_i}\right). \tag{4}$$

Evidently, $C_f(\mathbf{p},\mathbf{q};\mathbf{e})=C_f(\mathbf{p},\mathbf{q})$ with $\mathbf{e}=(1,\dots,1)\in\mathbb{R}^n$.

Mathematics subject classification (2020): 94A17, 26D15, 15B48, 15B51.
Keywords and phrases: Convex/concave function, entrywise positive matrix, column stochastic matrix, generalized Csiszár (f,g)-divergence, $\mathbf{p}$-stochastic matrix, $\mathbf{p}$-majorization.

A real matrix $\mathbf{A} = (a_{ij})$ of size $k \times l$ is called *column-stochastic*, if $a_{ij} \geqslant 0$ for $i = 1, \dots, k$ and $j = 1, \dots, l$, and, in addition, $\sum_{i=1}^{k} a_{ij} = 1$ for $j = 1, \dots, l$ (cf. [5, p. 29]).

In this paper, we develop some results of [8] by applying both convex and concave functions. To do so, we give further extension of (1) and (4). Namely, we introduce the notion of generalized Csiszár (f,g)-divergence induced by a convex function f and a concave function g. Furthermore, we show a generalization of the Csiszár-Körner inequality (2) for such (f,g)-divergences. To this end, we employ positive (entrywise) matrices, column stochastic matrices, and vectors related by Sherman type conditions [10]. We also consider some special cases of the obtained inequality. Finally, we give an application for $\mathbf{p}$-stochastic matrices (see [5, p. 585], [9]).

2. Generalized Csiszár (f,g)-divergence

In this section we introduce and study a generalization of Csiszár divergences (1) and (4) induced by two functions.

Let $f : \mathbb{R}_+ \to \mathbb{R}$ be a convex function and $g : \mathbb{R}_+ \to \mathbb{R}$ be a concave function. Let $\mathbf{p} = (p_1, \dots, p_n) \in \mathbb{R}^n_{++}$, $\mathbf{q} = (q_1, \dots, q_n) \in \mathbb{R}^n_+$, $\mathbf{r} = (r_1, \dots, r_n) \in \mathbb{R}^n_+$, $\mathbf{s} = (s_1, \dots, s_n) \in \mathbb{R}^n_{++}$. We define the *generalized Csiszár* (f,g)*-divergence* by

$$C_{f,g}(\mathbf{p},\mathbf{q};\mathbf{s},\mathbf{r}) = \sum_{i=1}^{n} s_i g\left(\frac{r_i}{s_i}\right) p_i f\left(\frac{q_i}{p_i}\right). \tag{5}$$

If, in particular, $\mathbf{s} = (1, \dots, 1) \in \mathbb{R}^n$ and $g = \mathrm{id}$ is the identity function on $[0,\infty)$, then $C_{f,\mathrm{id}}(\mathbf{p},\mathbf{q};\mathbf{s},\mathbf{r}) = C_f(\mathbf{p},\mathbf{q};\mathbf{r})$.

In the sequel, for two vectors $\mathbf{x} = (x_1, \dots, x_k) \in \mathbb{R}^k$ and $\mathbf{y} = (y_1, \dots, y_k) \in \mathbb{R}^k$, we define

$$\mathbf{x} \circ \mathbf{y} = (x_1 y_1, \dots, x_k y_k)$$

$$\frac{\mathbf{x}}{\mathbf{y}} = \left(\frac{x_1}{y_1}, \dots, \frac{x_k}{y_k}\right), \quad \text{with } \mathbf{y} \in \mathbb{R}^k_{++}.$$

For vectors $\mathbf{x} = (x_1, \dots, x_k) \in \mathbb{R}^k_+$ and $\mathbf{y} = (y_1, \dots, y_l) \in \mathbb{R}^l_{++}$, we introduce a transformation $T_{\mathbf{x},\mathbf{y}}$ of a real matrix $\mathbf{A} = (a_{ij})$ of size $k \times l$, as follows

$$T_{\mathbf{x},\mathbf{y}}\mathbf{A} = \begin{pmatrix} \frac{x_1 a_{11}}{y_1} & \dots & \frac{x_1 a_{1l}}{y_l} \\ \vdots & \dots & \vdots \\ \frac{x_k a_{k1}}{y_1} & \dots & \frac{x_k a_{kl}}{y_l} \end{pmatrix}.$$

It is important that if $\mathbf{y} = \mathbf{x}\mathbf{A}$ then the matrix $T_{\mathbf{x},\mathbf{y}}\mathbf{A}$ is column-stochastic.

It is not hard to verify by a routine algebra that for any entrywise nonnegative (or positive, if necessary) vectors $\mathbf{x} = (x_1, \dots, x_k)$, $\mathbf{y} = (y_1, \dots, y_l)$, $\mathbf{z} = (z_1, \dots, z_k)$, $\mathbf{v} = (v_1, \dots, v_l)$ of appropriate sizes and any (entrywise) nonnegative matrix $\mathbf{A}$, the following identities hold:

$$(\mathbf{y} \circ \mathbf{v})(T_{\mathbf{x},\mathbf{y}}\mathbf{A})^T = \mathbf{x} \circ (\mathbf{v}\mathbf{A}^T), \tag{6}$$

$$\frac{\mathbf{zA}}{\mathbf{y}} = \frac{\mathbf{z}}{\mathbf{x}} T_{\mathbf{x},\mathbf{y}}\mathbf{A}, \tag{7}$$

$$(\mathbf{x}\circ\mathbf{z})\mathbf{A} = \mathbf{y}\circ(\mathbf{z}T_{\mathbf{x},\mathbf{y}}\mathbf{A}). \tag{8}$$

Hereafter real n-tuples (i.e, row vectors of $\mathbb{R}^n$) are thought of as $1\times n$ matrices. Also, juxtaposition of matrices means their standard matrix product. In particular, juxtaposition of an n-tuple and an $n\times l$ matrix means their standard matrix product.

For functions $f:\mathbb{R}_+\to\mathbb{R}$, $g:\mathbb{R}_+\to\mathbb{R}$ and for a vector $\mathbf{z}=(z_1,\dots,z_l)\in\mathbb{R}^l_+$ we define

$$f(\mathbf{z}) = (f(z_1),\dots,f(z_l)) \quad \text{and} \quad g(\mathbf{z}) = (g(z_1),\dots,g(z_l)).$$

In the forthcoming theorem we present a comparison of two generalized Csiszár (f,g)-divergences with a convex function f and a concave function g (cf. [8]).

THEOREM 1. *Let $f:\mathbb{R}_+\to\mathbb{R}$ be a convex function on $\mathbb{R}_+$ and $g:\mathbb{R}_+\to\mathbb{R}$ be a concave function on $\mathbb{R}_+$. Let $\mathbf{p}=(p_1,\dots,p_n)\in\mathbb{R}^n_{++}$, $\mathbf{q}=(q_1,\dots,q_n)\in\mathbb{R}^n_+$, $\mathbf{s}=(s_1,\dots,s_m)\in\mathbb{R}^m_{++}$, $\mathbf{r}=(r_1,\dots,r_m)\in\mathbb{R}^m_+$, $\widetilde{\mathbf{p}}=(\widetilde{p}_1,\dots,\widetilde{p}_m)\in\mathbb{R}^m_{++}$, $\widetilde{\mathbf{q}}=(\widetilde{q}_1,\dots,\widetilde{q}_m)\in\mathbb{R}^m_+$, $\widetilde{\mathbf{s}}=(\widetilde{s}_1,\dots,\widetilde{s}_n)\in\mathbb{R}^n_{++}$, $\widetilde{\mathbf{r}}=(\widetilde{r}_1,\dots,\widetilde{r}_n)\in\mathbb{R}^n_+$. Assume $f\left(\frac{\mathbf{q}}{\mathbf{p}}\right)\geqslant 0$ and $g\left(\frac{\mathbf{r}}{\mathbf{s}}\right)\geqslant 0$.*

Let $\mathbf{S}=(s_{ij})$ and $\mathbf{R}=(r_{ji})$ be matrices with nonnegative entries of sizes $n\times m$ and $m\times n$, respectively, such that $\mathbf{S}^T\leqslant\mathbf{R}$ (in the entrywise order $\leqslant$).

If

$$\widetilde{\mathbf{p}} = \mathbf{pS}, \quad \widetilde{\mathbf{q}} = \mathbf{qS}, \tag{9}$$

$$\widetilde{\mathbf{s}} = \mathbf{sR}, \quad \widetilde{\mathbf{r}} = \mathbf{rR}, \tag{10}$$

then

$$C_{f,g}(\widetilde{\mathbf{p}},\widetilde{\mathbf{q}};\mathbf{s},\mathbf{r}) \leqslant C_{f,g}(\mathbf{p},\mathbf{q};\widetilde{\mathbf{s}},\widetilde{\mathbf{r}}). \tag{11}$$

Proof. According to (5) we have to prove that

$$\sum_{j=1}^{m} s_j g\left(\frac{r_j}{s_j}\right)\widetilde{p}_j f\left(\frac{\widetilde{q}_j}{\widetilde{p}_j}\right) \leqslant \sum_{i=1}^{n}\widetilde{s}_i g\left(\frac{\widetilde{r}_i}{\widetilde{s}_i}\right) p_i f\left(\frac{q_i}{p_i}\right),$$

that is,

$$\left\langle \mathbf{s}\circ g\left(\frac{\mathbf{r}}{\mathbf{s}}\right), \widetilde{\mathbf{p}}\circ f\left(\frac{\widetilde{\mathbf{q}}}{\widetilde{\mathbf{p}}}\right)\right\rangle \leqslant \left\langle \widetilde{\mathbf{s}}\circ g\left(\frac{\widetilde{\mathbf{r}}}{\widetilde{\mathbf{s}}}\right), \mathbf{p}\circ f\left(\frac{\mathbf{q}}{\mathbf{p}}\right)\right\rangle, \tag{12}$$

where $\langle\cdot,\cdot\rangle$ denotes the standard inner product (on $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively).

On account of (7) we have

$$\frac{\widetilde{\mathbf{q}}}{\widetilde{\mathbf{p}}} = \frac{\mathbf{q}}{\mathbf{p}} T_{\mathbf{p},\widetilde{\mathbf{p}}}\mathbf{S}. \tag{13}$$

Therefore,

$$\left\langle \mathbf{s}\circ g\left(\frac{\mathbf{r}}{\mathbf{s}}\right), \widetilde{\mathbf{p}}\circ f\left(\frac{\widetilde{\mathbf{q}}}{\widetilde{\mathbf{p}}}\right)\right\rangle = \left\langle \widetilde{\mathbf{p}}\circ\mathbf{s}\circ g\left(\frac{\mathbf{r}}{\mathbf{s}}\right), f\left(\frac{\widetilde{\mathbf{q}}}{\widetilde{\mathbf{p}}}\right)\right\rangle = \left\langle \widetilde{\mathbf{p}}\circ\mathbf{s}\circ g\left(\frac{\mathbf{r}}{\mathbf{s}}\right), f\left(\frac{\mathbf{q}}{\mathbf{p}} T_{\mathbf{p},\widetilde{\mathbf{p}}}\mathbf{S}\right)\right\rangle$$

$$\leqslant \left\langle \widetilde{\mathbf{p}}\circ\mathbf{s}\circ g\left(\frac{\mathbf{r}}{\mathbf{s}}\right), \left(f\left(\frac{\mathbf{q}}{\mathbf{p}}\right)\right) T_{\mathbf{p},\widetilde{\mathbf{p}}}\mathbf{S}\right\rangle = \left\langle \left(\widetilde{\mathbf{p}}\circ\mathbf{s}\circ g\left(\frac{\mathbf{r}}{\mathbf{s}}\right)\right)(T_{\mathbf{p},\widetilde{\mathbf{p}}}\mathbf{S})^T, f\left(\frac{\mathbf{q}}{\mathbf{p}}\right)\right\rangle. \tag{14}$$

The last inequality follows from the statement

$$f\left(\frac{\mathbf{q}}{\mathbf{p}} T_{\mathbf{p},\widetilde{\mathbf{p}}}\mathbf{S}\right) \leqslant \left(f\left(\frac{\mathbf{q}}{\mathbf{p}}\right)\right) T_{\mathbf{p},\widetilde{\mathbf{p}}}\mathbf{S} \quad \text{(entrywise)},$$

because f is convex, the matrix $T_{\mathbf{p},\widetilde{\mathbf{p}}}\mathbf{S}$ is column stochastic, and, in addition, $\widetilde{\mathbf{p}} \circ \mathbf{s} \circ g\left(\frac{\mathbf{r}}{\mathbf{s}}\right) \geqslant 0$ (entrywise).

We find from (6) that

$$(\widetilde{\mathbf{p}} \circ \mathbf{v})\,(T_{\mathbf{p},\widetilde{\mathbf{p}}}\mathbf{S})^T = \mathbf{p} \circ (\mathbf{v}\mathbf{S}^T) \quad \text{for } \mathbf{v} \in \mathbb{R}^m_+. \tag{15}$$

Hence, by putting $\mathbf{v} = \mathbf{s} \circ g\left(\frac{\mathbf{r}}{\mathbf{s}}\right)$, we deduce that

$$\begin{aligned} \left\langle \left(\widetilde{\mathbf{p}} \circ \mathbf{s} \circ g\left(\frac{\mathbf{r}}{\mathbf{s}}\right)\right)(T_{\mathbf{p},\widetilde{\mathbf{p}}}\mathbf{S})^T, f\left(\frac{\mathbf{q}}{\mathbf{p}}\right)\right\rangle &= \left\langle \mathbf{p} \circ \left((\mathbf{s} \circ g\left(\frac{\mathbf{r}}{\mathbf{s}}\right))\mathbf{S}^T\right), f\left(\frac{\mathbf{q}}{\mathbf{p}}\right)\right\rangle \\ &= \left\langle \left(\mathbf{s} \circ g\left(\frac{\mathbf{r}}{\mathbf{s}}\right)\right)\mathbf{S}^T, \mathbf{p} \circ f\left(\frac{\mathbf{q}}{\mathbf{p}}\right)\right\rangle. \end{aligned} \tag{16}$$

But $\mathbf{S}^T \leqslant \mathbf{R}$, so we obtain

$$\left\langle \left(\mathbf{s} \circ g\left(\frac{\mathbf{r}}{\mathbf{s}}\right)\right)\mathbf{S}^T, \mathbf{p} \circ f\left(\frac{\mathbf{q}}{\mathbf{p}}\right)\right\rangle \leqslant \left\langle \left(\mathbf{s} \circ g\left(\frac{\mathbf{r}}{\mathbf{s}}\right)\right)\mathbf{R}, \mathbf{p} \circ f\left(\frac{\mathbf{q}}{\mathbf{p}}\right)\right\rangle, \tag{17}$$

because $\mathbf{p} \circ f\left(\frac{\mathbf{q}}{\mathbf{p}}\right) \geqslant 0$ and $\left(\mathbf{s} \circ g\left(\frac{\mathbf{r}}{\mathbf{s}}\right)\right)\mathbf{S}^T \leqslant \left(\mathbf{s} \circ g\left(\frac{\mathbf{r}}{\mathbf{s}}\right)\right)\mathbf{R}$ with $\mathbf{s} \circ g\left(\frac{\mathbf{r}}{\mathbf{s}}\right) \geqslant 0$.

By virtue of (8) we can write

$$\left(\mathbf{s} \circ g\left(\frac{\mathbf{r}}{\mathbf{s}}\right)\right)\mathbf{R} = \widetilde{\mathbf{s}} \circ \left(g\left(\frac{\mathbf{r}}{\mathbf{s}}\right) T_{\mathbf{s},\widetilde{\mathbf{s}}}\mathbf{R}\right). \tag{18}$$

Therefore we get

$$\left\langle \left(\mathbf{s} \circ g\left(\frac{\mathbf{r}}{\mathbf{s}}\right)\right)\mathbf{R}, \mathbf{p} \circ f\left(\frac{\mathbf{q}}{\mathbf{p}}\right)\right\rangle = \left\langle \widetilde{\mathbf{s}} \circ \left(g\left(\frac{\mathbf{r}}{\mathbf{s}}\right) T_{\mathbf{s},\widetilde{\mathbf{s}}}\mathbf{R}\right), \mathbf{p} \circ f\left(\frac{\mathbf{q}}{\mathbf{p}}\right)\right\rangle. \tag{19}$$

The function g is concave and the matrix $T_{\mathbf{s},\widetilde{\mathbf{s}}}\mathbf{R}$ is column-stochastic, so we derive the inequality

$$\left(g\left(\frac{\mathbf{r}}{\mathbf{s}}\right)\right) T_{\mathbf{s},\widetilde{\mathbf{s}}}\mathbf{R} \leqslant g\left(\frac{\mathbf{r}}{\mathbf{s}} T_{\mathbf{s},\widetilde{\mathbf{s}}}\mathbf{R}\right). \tag{20}$$

Clearly, $\widetilde{\mathbf{s}} \geqslant 0$ and $\mathbf{p} \circ f\left(\frac{\mathbf{q}}{\mathbf{p}}\right) \geqslant 0$. In consequence, by (20), we have

$$\begin{aligned} \left\langle \widetilde{\mathbf{s}} \circ \left(\left(g\left(\frac{\mathbf{r}}{\mathbf{s}}\right)\right) T_{\mathbf{s},\widetilde{\mathbf{s}}}\mathbf{R}\right), \mathbf{p} \circ f\left(\frac{\mathbf{q}}{\mathbf{p}}\right)\right\rangle &\leqslant \left\langle \widetilde{\mathbf{s}} \circ g\left(\frac{\mathbf{r}}{\mathbf{s}} T_{\mathbf{s},\widetilde{\mathbf{s}}}\mathbf{R}\right), \mathbf{p} \circ f\left(\frac{\mathbf{q}}{\mathbf{p}}\right)\right\rangle \\ &= \left\langle \widetilde{\mathbf{s}} \circ g\left(\frac{\widetilde{\mathbf{r}}}{\widetilde{\mathbf{s}}}\right), \mathbf{p} \circ f\left(\frac{\mathbf{q}}{\mathbf{p}}\right)\right\rangle. \end{aligned} \tag{21}$$

The last equality holds by the equality

$$\frac{\mathbf{r}}{\mathbf{s}} T_{\mathbf{s},\widetilde{\mathbf{s}}} \mathbf{R} = \frac{\widetilde{\mathbf{r}}}{\widetilde{\mathbf{s}}} \tag{22}$$

(see (7)).

By combining (14), (16), (17), (19) and (21), we get the required assertions (12) and (11). This completes the proof. □

COROLLARY 1. *With the assumptions of Theorem 1, let* $\mathbf{s} = \mathbf{r}$ *and* $g(1) > 0$. *Then*

$$\sum_{j=1}^{m} s_j \widetilde{p}_j f\left(\frac{\widetilde{q}_j}{\widetilde{p}_j}\right) \leqslant \sum_{i=1}^{n} \widetilde{s}_i p_i f\left(\frac{q_i}{p_i}\right). \tag{23}$$

Proof. Since $\mathbf{s} = \mathbf{r}$, we also have $\widetilde{\mathbf{s}} = \widetilde{\mathbf{r}}$ by (10). Hence we get $\frac{r_j}{s_j} = 1$ and $\frac{\widetilde{r}_i}{\widetilde{s}_i} = 1$ for $j = 1, \ldots, m$ and $i = 1, \ldots, n$. Therefore (11) holds and reduces to

$$\sum_{j=1}^{m} s_j g(1) \widetilde{p}_j f\left(\frac{\widetilde{q}_j}{\widetilde{p}_j}\right) \leqslant \sum_{i=1}^{n} \widetilde{s}_i g(1) p_i f\left(\frac{q_i}{p_i}\right). \tag{24}$$

However, $g(1) > 0$, so (24) leads to (23). □

COROLLARY 2. *With the assumptions of Theorem 1, let* $\mathbf{S}$ *and* $\mathbf{R}$ *be matrices of ones of sizes* $n \times m$ *and* $m \times n$*, respectively.*

Then

$$\sum_{i=1}^{n} p_i f\left(\frac{\sum_{i=1}^{n} q_i}{\sum_{i=1}^{n} p_i}\right) \sum_{j=1}^{m} s_j g\left(\frac{r_j}{s_j}\right) \leqslant \sum_{j=1}^{m} s_j g\left(\frac{\sum_{j=1}^{m} r_j}{\sum_{j=1}^{m} s_j}\right) \sum_{i=1}^{n} p_i f\left(\frac{q_i}{p_i}\right). \tag{25}$$

Proof. Because $\mathbf{S} = (s_{ij})$ with $s_{ij} = 1$ and $\mathbf{R} = (r_{ji})$ with $r_{ji} = 1$ for $j = 1, \ldots, m$, $i = 1, \ldots, n$, conditions (9) and (10) imply that $\widetilde{p}_j = \sum_{i=1}^{n} p_i$, $\widetilde{q}_j = \sum_{i=1}^{n} q_i$ for $j = 1, \ldots, m$, and $\widetilde{s}_i = \sum_{j=1}^{m} s_j$, $\widetilde{r}_i = \sum_{j=1}^{m} r_j$ for $i = 1, \ldots, n$. It is now sufficient to apply inequality (11) in Theorem 1. □

Given an m-tuple $\mathbf{p} = (p_1, \ldots, p_m) \in \mathbb{R}^m_{++}$, an $m \times m$ matrix $\mathbf{S}$ with nonnegative entries is said to be $\mathbf{p}$-*stochastic*, if

(i) $\mathbf{p} = \mathbf{p}\mathbf{S}$,

(ii) $\mathbf{e} = \mathbf{e}\mathbf{S}^T$, where $\mathbf{e} = (1, \ldots, 1) \in \mathbb{R}^m$

(see [5, Definition B.1., p. 585]).

An m-tuple $\widetilde{\mathbf{q}} \in \mathbb{R}^m$ is said to be $\mathbf{p}$-*majorized* by an m-tuple $\mathbf{q} \in \mathbb{R}^m$ (abbreviated as $\widetilde{\mathbf{q}} \prec_{\mathbf{p}} \mathbf{q}$), if $\widetilde{\mathbf{q}} = \mathbf{q}\mathbf{S}$ for some $\mathbf{p}$-stochastic matrix $\mathbf{S}$ (see [5, Definition B.2., p. 585]).

It is not hard to check that if $\mathbf{p} = \mathbf{e}$ is the tuple of ones, then all $\mathbf{p}$-stochastic matrices are usual doubly stochastic matrices [5, p. 29]. For this reason, the relation $\prec_{\mathbf{e}}$ of $\mathbf{e}$-majorization is the standard majorization $\prec$ on $\mathbb{R}^m$ [5, p. 8].

It is known that the relation $\widetilde{\mathbf{q}} \prec_{\mathbf{p}} \mathbf{q}$ is characterized by the inequality

$$\sum_{i=1}^{m} p_i f\left(\frac{\widetilde{q}_i}{p_i}\right) \leqslant \sum_{i=1}^{m} p_i f\left(\frac{q_i}{p_i}\right) \tag{26}$$

for all continuous convex functions $f : \mathbb{R}_+ \to \mathbb{R}$, where $\mathbf{q} = (q_1, \ldots, q_m) \in \mathbb{R}^m_{++}$ and $\widetilde{\mathbf{q}} = (\widetilde{q}_1, \ldots, \widetilde{q}_m) \in \mathbb{R}^m_{++}$ (cf. [5, Proposition B.4., pp. 586–587] and [9, Proposition 4.2]).

In the next corollary we show an extension of the inequality (26).

COROLLARY 3. *Let* $f : \mathbb{R}_+ \to \mathbb{R}$ *be a convex function on* $\mathbb{R}_+$ *and* $g : \mathbb{R}_+ \to \mathbb{R}$ *be a concave function on* $\mathbb{R}_+$. *Let* $\mathbf{p} = (p_1, \ldots, p_m) \in \mathbb{R}^m_{++}$, $\mathbf{q} = (q_1, \ldots, q_m) \in \mathbb{R}^m_{+}$, $\mathbf{r} = (r_1, \ldots, r_m) \in \mathbb{R}^m_{+}$, $\widetilde{\mathbf{q}} = (\widetilde{q}_1, \ldots, \widetilde{q}_m) \in \mathbb{R}^m_{+}$, $\widetilde{\mathbf{r}} = (\widetilde{r}_1, \ldots, \widetilde{r}_m) \in \mathbb{R}^m_{+}$. *Assume* $f\left(\frac{\mathbf{q}}{\mathbf{p}}\right) \geqslant 0$ *and* $g(\mathbf{r}) \geqslant 0$.

If

$$\widetilde{\mathbf{q}} = \mathbf{q}\mathbf{S}, \tag{27}$$

$$\widetilde{\mathbf{r}} = \mathbf{r}\mathbf{S}^T \tag{28}$$

for some $\mathbf{p}$*-stochastic matrix* $\mathbf{S}$ *of size* $m \times m$, *then*

$$C_{f,g}(\mathbf{p}, \widetilde{\mathbf{q}}; \mathbf{e}, \mathbf{r}) \leqslant C_{f,g}(\mathbf{p}, \mathbf{q}; \mathbf{e}, \widetilde{\mathbf{r}}), \tag{29}$$

i.e.,

$$\sum_{i=1}^{m} g(r_i) p_i f\left(\frac{\widetilde{q}_i}{p_i}\right) \leqslant \sum_{i=1}^{m} g(\widetilde{r}_i) p_i f\left(\frac{q_i}{p_i}\right). \tag{30}$$

Proof. We use Theorem 1 with $m = n$, $\mathbf{R} = \mathbf{S}^T$, $\widetilde{\mathbf{p}} = \mathbf{p}$ and $\widetilde{\mathbf{s}} = \mathbf{s} = \mathbf{e} = (1, \ldots, 1) \in \mathbb{R}^m$. Then $\widetilde{\mathbf{p}} = \mathbf{p}\mathbf{S}$ and $\widetilde{\mathbf{s}} = \mathbf{s}\mathbf{R}$, which together with (27) and (28) gives (9) and (10). So, we are allowed to apply inequality (11). However, in the present situation, (11) takes the form (29). This completes the proof. □

We finish this section with the observation that if $\mathbf{p} = \mathbf{e} = (1, \ldots, 1) \in \mathbb{R}^m$, and $g(t) = t$ is the identity function on $\mathbb{R}_+$, then inequality (30) reduces Sherman's inequality [10, 7].

Acknowledgements. The author would like to thank an anonymous referee for his/her comments improving the previous version of the manuscript.

REFERENCES

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

[2] I. CSISZÁR AND J. KÖRNER, *Information Theory: Coding Theorems for Discrete Memory-less Systems*, Academic Press, New York, 1981.

[3] S. S. DRAGOMIR (Ed.), *Upper and lower bounds for Csiszár f-divergence in terms of the Kullback-Leibler distance and applications*, Inequalities for the Csiszár f-divergence in Information Theory, 2000, http://rgmia.vu.edu.au/monographs/csiszar.htm.

[4] P. KLUZA AND M. NIEZGODA, *On Csiszár and Tsallis type f-divergences induced by superquadratic and convex functions*, Math. Inequal. Appl., **21**, (2018), 455–467.

[5] A. W. MARSHALL, I. OLKIN AND B. C. ARNOLD, *Inequalities: Theory of Majorization and Its Applications*, Springer, 2nd printing, New York, 2011.

[6] M. NIEZGODA, *Vector joint majorization and generalization of Csiszár-Körner's inequality for f-divergence*, Discrete Appl. Math., **198**, (2016), 195–205.

[7] M. NIEZGODA, *Nonlinear Sherman type inequalities*, Adv. Nonlinear Anal., **9**, 1 (2020), 168–175.

[8] M. NIEZGODA, *Inequalities for convex and concave functions and a new concept of majorization intended for two pairs of vectors*, Results in Math., **75**, 1 (2020), article 34.

[9] F. VOM ENDE AND G. DIRR, *The d-majorization polytope*, arXiv:1911.01061v1, 4 Nov 2019.

[10] S. SHERMAN, *On a theorem of Hardy, Littlewood, Pólya, and Blackwell*, Proc. Nat. Acad. Sci. USA, **37**, (1957), 826–831.

(Received August 6, 2021)

Marek Niezgoda
Institute of Mathematics
Pedagogical University of Cracow
Podchorążych 2, 30-084 Kraków, Poland
e-mail: bniezgoda@wp.pl