Mathematical
Inequalities
& Applications
Volume 24, Number 1 (2021), 103–113 doi:10.7153/mia-2021-24-08

ON A MERCER LIKE INEQUALITY INVOLVING GENERALIZED CSISZÁR f–DIVERGENCES

MAREK NIEZGODA

(*Communicated by M. Praljak*)

Abstract. In this note, an upper bound for values of a convex function f is shown for some specific arguments of the function. Thus a Mercer like inequality involving generalized Csiszár f-divergences is obtained. Special cases of the result are studied.

1. Introduction

We begin with the following result due to A. McD. Mercer [10].

THEOREM A. [10, Theorem 1.2] *Let f be a real convex function on an interval $[a_1,a_2]$, $a_1<a_2$, such that*

$$a_1 \leqslant x_k \leqslant a_2 \quad \text{for } k \in \{1,\dots,N\}. \tag{1}$$

Then

$$f\left(a_1+a_2-\sum_{i=1}^N t_k x_k\right) \leqslant f(a_1)+f(a_2)-\sum_{i=1}^N t_k f(x_k), \tag{2}$$

where $\sum_{k=1}^N t_k = 1$ *with* $t_k > 0$.

Throughout the notation $\mathbb{R}_+=[0,\infty)$ and $\mathbb{R}_{++}=(0,\infty)$ is used. Elements of the Euclidean space $\mathbb{R}^n$ are thought of as row n-vectors.

Given a convex function $f:\mathbb{R}_{++}\to\mathbb{R}$ and two n-tuples $\mathbf{p}=(p_1,p_2,\dots,p_n)\in\mathbb{R}^n_{++}$ and $\mathbf{q}=(q_1,q_2,\dots,q_n)\in\mathbb{R}^n_{++}$, the quantity

$$C_f(\mathbf{p},\mathbf{q}) = \sum_{j=1}^n p_j f\left(\frac{q_j}{p_j}\right) \tag{3}$$

is called *Csiszár f-divergence* (see [1, 2, 3]).

Mathematics subject classification (2010): 94A17, 26D15, 15B48.
Keywords and phrases: Convex function, Mercer inequality, generalized Csiszár f-divergence, nonnegative (entrywise) matrix, column (row) stochastic matrix, $\mathbf{p}$-majorization.

The following *Csiszár-Körner inequality* [2] holds:

$$\sum_{j=1}^{n} p_j f\left(\frac{\sum_{j=1}^{n} q_j}{\sum_{j=1}^{n} p_j}\right) \leqslant C_f(\mathbf{p},\mathbf{q}). \tag{4}$$

For properties of f-divergence, see [3, 4, 7, 13].

For example, we now give definitions of some f-divergences (relative entropies) induced by the convex functions $-\log t$, $t\log t$, $-\frac{t^u-1}{u}$ and $-\frac{[1-v+vt^u]^{1/u}-1}{v}$ for $t>0$, as follows

$$S(\mathbf{p},\mathbf{q}) = -\sum_{i=1}^{n} p_i \log\frac{q_i}{p_i} \quad \text{(relative entropy)}, \tag{5}$$

$$S_1(\mathbf{p},\mathbf{q}) = \sum_{i=1}^{n} p_i \left(\frac{q_i}{p_i}\right)\log\left(\frac{q_i}{p_i}\right) = S(\mathbf{q},\mathbf{p}), \tag{6}$$

$$T_u(\mathbf{p},\mathbf{q}) = -\sum_{i=1}^{n} p_i \frac{\left(\frac{q_i}{p_i}\right)^u - 1}{u}, \quad u\in(0,1], \quad \text{(Tsallis relative entropy)}, \tag{7}$$

and

$$T_{v,u}(\mathbf{p},\mathbf{q}) = -\sum_{i=1}^{n} p_i \frac{\left[1-v+v\left(\frac{q_i}{p_i}\right)^u\right]^{1/u} - 1}{v} \quad \text{(parametrized Tsallis relative entropy)}, \tag{8}$$

where $v,u\in(0,1]$ (see [5, 6, 12, 16]).

For a convex function $f:\mathbb{R}_{++}\to\mathbb{R}$ and for three n-tuples $\mathbf{p}=(p_1,p_2,\ldots,p_n)\in\mathbb{R}^n_{++}$, $\mathbf{q}=(q_1,q_2,\ldots,q_n)\in\mathbb{R}^n_{++}$ and $\mathbf{c}=(c_1,c_2,\ldots,c_n)\in\mathbb{R}^n_{+}$, the *generalized Csiszár f-divergence* of $\mathbf{p}$ and $\mathbf{q}$ with respect to $\mathbf{c}$ is defined by

$$C_f(\mathbf{p},\mathbf{q};\mathbf{c}) = \sum_{j=1}^{n} c_j p_j f\left(\frac{q_j}{p_j}\right) \tag{9}$$

(see [8]).

We say that an $n\times m$ real matrix $\mathbf{R}=(r_{ji})$ is *nonnegative* (entrywise), written as $\mathbf{R}\geqslant 0$, if $r_{ji}\geqslant 0$ for all $j\in\{1,\ldots,n\}$ and $i\in\{1,\ldots,m\}$.

In what follows, we use the symbol $\mathbf{R}^T$ to denote the transpose of a matrix $\mathbf{R}$.

THEOREM B. [8] *Let $f:\mathbb{R}_{++}\to\mathbb{R}$ be a convex function on $\mathbb{R}_{++}$. Let $\mathbf{p}=(p_1,p_2,\ldots,p_n)\in\mathbb{R}^n_{++}$, $\mathbf{q}=(q_1,q_2,\ldots,q_n)\in\mathbb{R}^n_{++}$ and $\mathbf{d}=(d_1,d_2,\ldots,d_m)\in\mathbb{R}^m_{+}$.*

Let $\mathbf{R}$ be an $n\times m$ nonnegative (entrywise) matrix. Denote

$$\widetilde{\mathbf{p}} = \mathbf{pR}, \quad \widetilde{\mathbf{q}} = \mathbf{qR} \quad \text{and} \quad \mathbf{c} = \mathbf{dR}^T. \tag{10}$$

Assume $\widetilde{\mathbf{p}}\in\mathbb{R}^m_{++}$.

Then

$$C_f(\widetilde{\mathbf{p}},\widetilde{\mathbf{q}};\mathbf{d}) \leqslant C_f(\mathbf{p},\mathbf{q};\mathbf{c}). \tag{11}$$

It is interesting that inequality (11) is a generalization of the Csiszár-Körner inequality (4). Namely, it can be observed from (10) that if $m = n$ and $\mathbf{R}$ is the matrix of ones and $\mathbf{d} = \frac{1}{m}(1,1,\ldots,1) \in \mathbb{R}^m_+$, then $\mathbf{c} = (1,1,\ldots,1) \in \mathbb{R}^n_+$, $\widetilde{\mathbf{p}} = \left(\sum_{j=1}^n p_j,\ldots,\sum_{j=1}^n p_j\right) \in \mathbb{R}^m_{++}$, $\widetilde{\mathbf{q}} = \left(\sum_{j=1}^n q_j,\ldots,\sum_{j=1}^n q_j\right) \in \mathbb{R}^m_{++}$. So, in this situation, (11) reduces to (4) by (3) and (9).

The aim of the present note is to develop the above framework in order to establish an upper bound for some values of a convex function f by using generalized Csiszár f-divergences. In doing so, we apply a transform of a matrix with nonnegative entries to obtain a column stochastic matrix. In result, we are permitted to employ Theorem B, which together with Jensen inequality gives the desired estimate of values of a convex function (see Theorem 1). Next, we consider some specializations of Theorem 1 in Corollaries 1-3. Also, we show an application for $\mathbf{p}$-majorization (see Corollary 4).

2. Mercer type inequality for generalized Csiszár f-divergences

For any l-tuples $\mathbf{a} = (a_1,\ldots,a_l) \in \mathbb{R}^l_{++}$ and $\mathbf{b} = (b_1,\ldots,b_l) \in \mathbb{R}^l_+$, we denote

$$\mathbf{a}\circ\mathbf{b} = (a_1b_1,\ldots,a_lb_l)\ ,\quad \frac{\mathbf{b}}{\mathbf{a}} = \left(\frac{b_1}{a_1},\ldots,\frac{b_l}{a_l}\right) \quad \text{and} \quad \frac{\mathbf{1}}{\mathbf{a}} = \left(\frac{1}{a_1},\ldots,\frac{1}{a_l}\right).$$

An $n\times m$ real matrix $\mathbf{R} = (r_{ji})$ is said to be *column stochastic* if $r_{ji} \geqslant 0$ for $j \in \{1,\ldots,n\}$ and $i \in \{1,\ldots,m\}$, and all column sums of $\mathbf{R}$ are ones, i.e., $\sum_{j=1}^n r_{ji} = 1$ for $i \in \{1,\ldots,m\}$.

An $n\times m$ real matrix $\mathbf{R} = (r_{ji})$ is said to be *row stochastic* if $r_{ji} \geqslant 0$ for $j \in \{1,\ldots,n\}$ and $i \in \{1,\ldots,m\}$, and all row sums of $\mathbf{R}$ are ones, i.e., $\sum_{i=1}^m r_{ji} = 1$ for $j \in \{1,\ldots,n\}$.

An $m\times m$ real matrix $\mathbf{R}$ is called *doubly stochastic* if $\mathbf{R}$ is both column stochastic and row stochastic.

We say that an m-tuple $\mathbf{y} \in \mathbb{R}^m$ is *majorized* by an m-tuple $\mathbf{x} \in \mathbb{R}^m$, written as $\mathbf{y} \prec \mathbf{x}$, if $\mathbf{y} = \mathbf{x}\mathbf{R}$ for some doubly stochastic matrix $\mathbf{R}$ (see [9, p. 33]).

LEMMA 1. *Let* $\mathbf{p} = (p_1,p_2,\ldots,p_n) \in \mathbb{R}^n_{++}$, $\mathbf{q} = (q_1,q_2,\ldots,q_n) \in \mathbb{R}^n_{++}$ $\widetilde{\mathbf{p}} = (\widetilde{p}_1,\widetilde{p}_2,\ldots,\widetilde{p}_m) \in \mathbb{R}^m_{++}$ *and* $\widetilde{\mathbf{q}} = (\widetilde{q}_1,\widetilde{q}_2,\ldots,\widetilde{q}_m) \in \mathbb{R}^m_{++}$.

Let $\mathbf{R} = (r_{ji})$ *be an* $n\times m$ *matrix with nonnegative entries. Let* $\mathbf{S} = (s_{ji})$ *be the* $n\times m$ *matrix such that* $s_{ji} = r_{ji}\frac{p_j}{\widetilde{p}_i}$ *for* $i \in \{1,\ldots,m\}$ *and* $j \in \{1,\ldots,n\}$.

(i) *If* $\widetilde{\mathbf{q}} = \mathbf{q}\mathbf{R}$ *then*

$$\frac{\widetilde{\mathbf{q}}}{\widetilde{\mathbf{p}}} = \frac{\mathbf{q}}{\mathbf{p}}\mathbf{S}.$$

(ii) *If* $\widetilde{\mathbf{p}} = \mathbf{p}\mathbf{R}$ *then* $\mathbf{S}$ *is column stochastic.*

(iii) *If* $\frac{\mathbf{1}}{\mathbf{p}} = \frac{\mathbf{1}}{\widetilde{\mathbf{p}}}\mathbf{R}^T$ *then* $\mathbf{S}$ *is row stochastic.*

Proof. **(i)**. Since $\widetilde{\mathbf{q}} = \mathbf{q}\mathbf{R}$, the following identity holds

$$\frac{\widetilde{q}_i}{\widetilde{p}_i} = \frac{q_1}{p_1} r_{1i} \frac{p_1}{\widetilde{p}_i} + \ldots + \frac{q_n}{p_n} r_{ni} \frac{p_n}{\widetilde{p}_i} = \frac{q_1}{p_1} s_{1i} + \ldots + \frac{q_n}{p_n} s_{ni} \quad \text{for } i \in \{1,\ldots,m\}. \tag{12}$$

So, we find that

$$\frac{\widetilde{\mathbf{q}}}{\widetilde{\mathbf{p}}} = \left(\frac{\widetilde{q}_1}{\widetilde{p}_1}, \ldots, \frac{\widetilde{q}_m}{\widetilde{p}_m} \right) = \left(\frac{q_1}{p_1}, \ldots, \frac{q_n}{p_n} \right) \begin{pmatrix} r_{11}\frac{p_1}{\widetilde{p}_1} & \ldots & r_{1m}\frac{p_1}{\widetilde{p}_m} \\ \vdots & \ddots & \vdots \\ r_{n1}\frac{p_n}{\widetilde{p}_1} & \ldots & r_{nm}\frac{p_n}{\widetilde{p}_m} \end{pmatrix}$$

$$= \left(\frac{q_1}{p_1}, \ldots, \frac{q_n}{p_n} \right) \begin{pmatrix} s_{11} & \ldots & s_{1m} \\ \vdots & \ddots & \vdots \\ s_{n1} & \ldots & s_{nm} \end{pmatrix} = \frac{\mathbf{q}}{\mathbf{p}} \mathbf{S}.$$

(ii). In light of the equality $\widetilde{\mathbf{p}} = \mathbf{p}\mathbf{R}$, it is not hard to check that

$$s_{1i} + \ldots + s_{ni} = r_{1i} \frac{p_1}{\widetilde{p}_i} + \ldots + r_{ni} \frac{p_n}{\widetilde{p}_i} = \frac{\sum\limits_{j=1}^{n} p_j r_{ji}}{\sum\limits_{j=1}^{n} p_j r_{ji}} = 1 \quad \text{for } i \in \{1,\ldots,m\}. \tag{13}$$

For this reason the matrix $\mathbf{S} = (s_{ji})$ is column stochastic.

(iii). Assume $\frac{\mathbf{1}}{\mathbf{p}} = \frac{\mathbf{1}}{\widetilde{\mathbf{p}}}\mathbf{R}^T$. Hence $\frac{1}{p_j} = \frac{1}{\widetilde{p}_1} r_{j1} + \ldots + \frac{1}{\widetilde{p}_m} r_{jm}$ for $j \in \{1,\ldots,n\}$.

In conseqence, we get

$$s_{j1} + \ldots + s_{jm} = r_{j1} \frac{p_j}{\widetilde{p}_1} + \ldots + r_{jm} \frac{p_j}{\widetilde{p}_m} = p_j \left(\frac{1}{\widetilde{p}_1} r_{j1} + \ldots + \frac{1}{\widetilde{p}_m} r_{jm} \right) = 1 \quad \text{for } j \in \{1,\ldots,n\}. \tag{14}$$

That is, the matrix $\mathbf{S} = (s_{ji})$ is row stochastic. $\square$

THEOREM 1. *Let* $f : \mathbb{R}_{++} \to \mathbb{R}$ *be a convex function on* $\mathbb{R}_{++}$. *Let* $\mathbf{p} = (p_1, p_2, \ldots, p_n) \in \mathbb{R}^n_{++}$, $\mathbf{q} = (q_1, q_2, \ldots, q_n) \in \mathbb{R}^n_{++}$, $\mathbf{d} = (d_1, d_2, \ldots, d_m) \in \mathbb{R}^m_{+}$, $d_m > 0$, $\mathbf{p}_k = (p^{(k)}_1, p^{(k)}_2, \ldots, p^{(k)}_m) \in \mathbb{R}^m_{++}$, $\mathbf{q}_k = (q^{(k)}_1, q^{(k)}_2, \ldots, q^{(k)}_m) \in \mathbb{R}^m_{++}$, $\mathbf{c}_k = (c^{(k)}_1, c^{(k)}_2, \ldots, c^{(k)}_n) \in \mathbb{R}^n_{+}$, $k \in \{1,\ldots,N\}$.

Let $\mathbf{R}_k = \left(r^{(k)}_{ji} \right)$, $k \in \{1,\ldots,N\}$, *be an* $n \times m$ *nonnegative (entrywise) matrix such that*

$$\mathbf{p}_k = \mathbf{p}\mathbf{R}_k\,, \quad \mathbf{q}_k = \mathbf{q}\mathbf{R}_k \quad \textit{and} \quad \mathbf{c}_k = \mathbf{d}\mathbf{R}^T_k \quad \textit{for } k \in \{1,\ldots,N\}. \tag{15}$$

Then, for any $t_k \geqslant 0$, $k \in \{1,\ldots,N\}$, *with* $\sum\limits_{k=1}^{N} t_k = 1$, *the following inequality holds:*

$$f\left(\sum_{k=1}^{N} t_k \sum_{j=1}^{n} \lambda^{(k)}_j \frac{q_j}{p_j} - \sum_{k=1}^{N} t_k \sum_{i=1}^{m-1} \frac{q^{(k)}_i}{p^{(k)}_i} \right) \leqslant \sum_{k=1}^{N} \frac{t_k}{d_m p^{(k)}_m} \left(C_f(\mathbf{p},\mathbf{q};\mathbf{c}_k) - C_f(\widehat{\mathbf{p}}_k, \widehat{\mathbf{q}}_k; \widehat{\mathbf{d}}) \right), \tag{16}$$

where $\lambda_j^{(k)}$ *is the* j*th row sum of the matrix* $\mathbf{S}_k = \left(s_{ji}^{(k)}\right)$ *with*

$$s_{ji}^{(k)} = r_{ji}^{(k)} \frac{p_j}{p_i^{(k)}} \quad \textit{for } j \in \{1,\dots,n\} \textit{ and } i \in \{1,\dots,m\}, \tag{17}$$

and $\widehat{\mathbf{p}}_k = (p_1^{(k)}, p_2^{(k)}, \dots, p_{m-1}^{(k)}) \in \mathbb{R}_{++}^{m-1}$, $\widehat{\mathbf{q}}_k = (q_1^{(k)}, q_2^{(k)}, \dots, q_{m-1}^{(k)}) \in \mathbb{R}_{++}^{m-1}$, $\widehat{\mathbf{d}} = (d_1, d_2, \dots, d_{m-1}) \in \mathbb{R}_{+}^{m-1}$.

Proof. It follows from Jensen's inequality that

$$f\left(\sum_{k=1}^{N} t_k \frac{q_m^{(k)}}{p_m^{(k)}}\right) \leqslant \sum_{k=1}^{N} t_k f\left(\frac{q_m^{(k)}}{p_m^{(k)}}\right). \tag{18}$$

It is easily seen from Lemma 1, part **(i)**, and from (15) that

$$\frac{\mathbf{q}_k}{\mathbf{p}_k} = \frac{\mathbf{q}}{\mathbf{p}} \mathbf{S}_k \quad \text{for } k \in \{1,\dots,N\}. \tag{19}$$

On account of Lemma 1, part **(ii)**, and (15), the matrix $\mathbf{S}_k$ is column stochastic.

Applying (15)–(16) leads to

$$\sum_{i=1}^{m} \frac{q_i^{(k)}}{p_i^{(k)}} = \sum_{j=1}^{n} \lambda_j^{(k)} \frac{q_j}{p_j} \quad \text{for } k \in \{1,\dots,N\}. \tag{20}$$

In fact, because of the equalities $\lambda_j^{(k)} = \sum_{i=1}^{m} s_{ji}^{(k)} = \sum_{i=1}^{m} r_{ji}^{(k)} \frac{p_j}{p_i^{(k)}}$ for $j \in \{1,\dots,n\}$, $\mathbf{q}_k = \mathbf{q}\mathbf{R}_k$ and $q_i^{(k)} = \sum_{j=1}^{n} r_{ji}^{(k)} q_j$ for $i \in \{1,\dots,m\}$, we can write

$$\sum_{j=1}^{n} \lambda_j^{(k)} \frac{q_j}{p_j} = \sum_{j=1}^{n} \left(\sum_{i=1}^{m} r_{ji}^{(k)} \frac{p_j}{p_i^{(k)}}\right) \frac{q_j}{p_j} = \sum_{i=1}^{m} \sum_{j=1}^{n} r_{ji}^{(k)} \frac{q_j}{p_i^{(k)}} = \sum_{i=1}^{m} \frac{1}{p_i^{(k)}} \sum_{j=1}^{n} r_{ji}^{(k)} q_j = \sum_{i=1}^{m} \frac{q_i^{(k)}}{p_i^{(k)}}.$$

Therefore (20) yields

$$0 < \frac{q_m^{(k)}}{p_m^{(k)}} = \sum_{j=1}^{n} \lambda_j^{(k)} \frac{q_j}{p_j} - \sum_{i=1}^{m-1} \frac{q_i^{(k)}}{p_i^{(k)}} \quad \text{for } k \in \{1,\dots,N\}, \tag{21}$$

whence

$$f\left(\sum_{k=1}^{N} t_k \frac{q_m^{(k)}}{p_m^{(k)}}\right) = f\left(\sum_{k=1}^{N} t_k \sum_{j=1}^{n} \lambda_j^{(k)} \frac{q_j}{p_j} - \sum_{k=1}^{N} t_k \sum_{i=1}^{m-1} \frac{q_i^{(k)}}{p_i^{(k)}}\right). \tag{22}$$

By (18), (21) and (22) we get

$$f\left(\sum_{k=1}^{N} t_k \sum_{j=1}^{n} \lambda_j^{(k)} \frac{q_j}{p_j} - \sum_{k=1}^{N} t_k \sum_{i=1}^{m-1} \frac{q_i^{(k)}}{p_i^{(k)}}\right) \leqslant \sum_{k=1}^{N} t_k f\left(\sum_{j=1}^{n} \lambda_j^{(k)} \frac{q_j}{p_j} - \sum_{i=1}^{m-1} \frac{q_i^{(k)}}{p_i^{(k)}}\right). \tag{23}$$

On the other hand, by virtue of Theorem B, for any $k \in \{1,\dots,N\}$, we have

$$C_f\left(\mathbf{p}_k,\mathbf{q}_k;\mathbf{d}\right) \leqslant C_f\left(\mathbf{p},\mathbf{q};\mathbf{c}_k\right), \tag{24}$$

that is,

$$\sum_{i=1}^{m} d_i p_i^{(k)} f\left(\frac{q_i^{(k)}}{p_i^{(k)}}\right) \leqslant \sum_{j=1}^{n} c_j^{(k)} p_j f\left(\frac{q_j}{p_j}\right). \tag{25}$$

Hence, for any $k \in \{1,\dots,N\}$,

$$f\left(\frac{q_m^{(k)}}{p_m^{(k)}}\right) \leqslant \frac{1}{d_m p_m^{(k)}}\left(\sum_{j=1}^{n} c_j^{(k)} p_j f\left(\frac{q_j}{p_j}\right) - \sum_{i=1}^{m-1} d_i p_i^{(k)} f\left(\frac{q_i^{(k)}}{p_i^{(k)}}\right)\right), \tag{26}$$

and further,

$$\sum_{k=1}^{N} t_k f\left(\frac{q_m^{(k)}}{p_m^{(k)}}\right) \leqslant \sum_{k=1}^{N} \frac{t_k}{d_m p_m^{(k)}}\left(\sum_{j=1}^{n} c_j^{(k)} p_j f\left(\frac{q_j}{p_j}\right) - \sum_{i=1}^{m-1} d_i p_i^{(k)} f\left(\frac{q_i^{(k)}}{p_i^{(k)}}\right)\right). \tag{27}$$

By combining (21) and (27) we obtain

$$\sum_{k=1}^{N} t_k f\left(\sum_{j=1}^{n} \lambda_j^{(k)} \frac{q_j}{p_j} - \sum_{i=1}^{m-1} \frac{q_i^{(k)}}{p_i^{(k)}}\right) \tag{28}$$

$$\leqslant \sum_{k=1}^{N} \frac{t_k}{d_m p_m^{(k)}}\left(\sum_{j=1}^{n} c_j^{(k)} p_j f\left(\frac{q_j}{p_j}\right) - \sum_{i=1}^{m-1} d_i p_i^{(k)} f\left(\frac{q_i^{(k)}}{p_i^{(k)}}\right)\right).$$

Simultaneously, by (9) we have

$$C_f\left(\mathbf{p},\mathbf{q};\mathbf{c}_k\right) = \sum_{j=1}^{n} c_j^{(k)} p_j f\left(\frac{q_j}{p_j}\right) \quad \text{and} \quad C_f\left(\widehat{\mathbf{p}}_k,\widehat{\mathbf{q}}_k;\widehat{\mathbf{d}}\right) = \sum_{i=1}^{m-1} d_i p_i^{(k)} f\left(\frac{q_i^{(k)}}{p_i^{(k)}}\right).$$

Now, we deduce from (23) and (28) that

$$f\left(\sum_{k=1}^{N} t_k \sum_{j=1}^{n} \lambda_j^{(k)} \frac{q_j}{p_j} - \sum_{k=1}^{N} t_k \sum_{i=1}^{m-1} \frac{q_i^{(k)}}{p_i^{(k)}}\right) \leqslant \sum_{k=1}^{N} \frac{t_k}{d_m p_m^{(k)}}\left(C_f\left(\mathbf{p},\mathbf{q};\mathbf{c}_k\right) - C_f\left(\widehat{\mathbf{p}}_k,\widehat{\mathbf{q}}_k;\widehat{\mathbf{d}}\right)\right), \tag{29}$$

completing the proof. □

REMARK 1. It is easy to verify by (21) that in the case $N = 1$, $k = 1$, $t_k = t_1 = 1$, inequality (16) takes the form

$$C_f\left(\mathbf{p}_1,\mathbf{q}_1;\mathbf{d}\right) \leqslant C_f\left(\mathbf{p},\mathbf{q};\mathbf{c}_1\right). \tag{30}$$

In other words, inequality (16) includes (30), as a special case.

We now investigate some special cases of Theorem 1.

COROLLARY 1. *Let $f:\mathbb{R}_{++}\to\mathbb{R}$ be a convex function on $\mathbb{R}_{++}$. Let $\mathbf{q}=(q_1,q_2,\dots,q_n)\in\mathbb{R}^n_{++}$, $\mathbf{d}=(d_1,d_2,\dots,d_m)\in\mathbb{R}^m_{+}$, $d_m>0$, $\mathbf{q}_k=(q_1^{(k)},q_2^{(k)},\dots,q_m^{(k)})\in\mathbb{R}^m_{++}$, $\mathbf{c}_k=(c_1^{(k)},c_2^{(k)},\dots,c_n^{(k)})\in\mathbb{R}^n_{+}$, $k\in\{1,\dots,N\}$.*

Let $\mathbf{R}_k=\left(r_{ji}^{(k)}\right)$, $k\in\{1,\dots,N\}$, be an $n\times m$ column stochastic matrix such that

$$\mathbf{q}_k=\mathbf{q}\mathbf{R}_k \quad \textit{and} \quad \mathbf{c}_k=\mathbf{d}\mathbf{R}_k^T \quad \textit{for } k\in\{1,\dots,N\}. \tag{31}$$

Then, for any $t_k\geqslant 0$, $k\in\{1,\dots,N\}$, with $\sum\limits_{k=1}^{N}t_k=1$, the following inequality holds:

$$f\left(\sum_{k=1}^{N}t_k\left(\sum_{j=1}^{n}\lambda_j^{(k)}q_j-\sum_{i=1}^{m-1}q_i^{(k)}\right)\right)\leqslant\sum_{k=1}^{N}\frac{t_k}{d_m}\left(\sum_{j=1}^{n}c_j^{(k)}f(q_j)-\sum_{i=1}^{m-1}d_if\left(q_i^{(k)}\right)\right), \tag{32}$$

where $\lambda_j^{(k)}$ is the jth row sum of the matrix $\mathbf{R}_k$.

Proof. We consider the vectors $\mathbf{p}=(1,1,\dots,1)\in\mathbb{R}^n$ and $\mathbf{p}_k=(1,1,\dots,1)\in\mathbb{R}^m$. Since $\mathbf{R}_k$ is column stochastic, it follows that

$$\mathbf{p}_k=\mathbf{p}\mathbf{R}_k \quad \text{for } k\in\{1,\dots,N\}.$$

We introduce the matrix $\mathbf{S}_k=\left(s_{ji}^{(k)}\right)$ with $s_{ji}^{(k)}=r_{ji}^{(k)}\frac{p_j}{p_i^{(k)}}$ for $j\in\{1,\dots,n\}$ and $i\in\{1,\dots,m\}$. Therefore we have $\mathbf{S}_k=\mathbf{R}_k$ and $\mathbf{S}_k$ is column stochastic for $k\in\{1,\dots,N\}$.

So, we are allowed to apply Theorem 1. By inequality (16) we obtain

$$f\left(\sum_{k=1}^{N}t_k\left(\sum_{j=1}^{n}\lambda_j^{(k)}\frac{q_j}{p_j}-\sum_{i=1}^{m-1}\frac{q_i^{(k)}}{p_i^{(k)}}\right)\right)$$
$$\leqslant\sum_{k=1}^{N}\frac{t_k}{d_mp_m^{(k)}}\left(\sum_{j=1}^{n}c_j^{(k)}p_jf\left(\frac{q_j}{p_j}\right)-\sum_{i=1}^{m-1}d_ip_i^{(k)}f\left(\frac{q_i^{(k)}}{p_i^{(k)}}\right)\right). \tag{33}$$

Since $p_j=1$ and $p_i^{(k)}=1$ for $j\in\{1,\dots,n\}$ and $i\in\{1,\dots m\}$, (33) reduces to inequality (32), as claimed. □

COROLLARY 2. *Let $f:\mathbb{R}_{++}\to\mathbb{R}$ be a convex function on $\mathbb{R}_{++}$. Let $\mathbf{q}=(q_1,q_2,\dots,q_n)\in\mathbb{R}^n_{++}$, $\mathbf{q}_k=(q_1^{(k)},q_2^{(k)},\dots,q_m^{(k)})\in\mathbb{R}^m_{++}$, $k\in\{1,\dots,N\}$.*

Let $\mathbf{R}_k=\left(r_{ji}^{(k)}\right)$, $k\in\{1,\dots,N\}$, be an $n\times m$ column stochastic matrix such that

$$\mathbf{q}_k=\mathbf{q}\mathbf{R}_k \quad \textit{for } k\in\{1,\dots,N\}. \tag{34}$$

Then, for any $t_k\geqslant 0$, $k\in\{1,\dots,N\}$, with $\sum\limits_{k=1}^{N}t_k=1$, the following inequality holds:

$$f\left(\sum_{k=1}^{N}t_k\left(\sum_{j=1}^{n}\lambda_j^{(k)}q_j-\sum_{i=1}^{m-1}q_i^{(k)}\right)\right)\leqslant\sum_{k=1}^{N}t_k\left(\sum_{j=1}^{n}\lambda_j^{(k)}f(q_j)-\sum_{i=1}^{m-1}f\left(q_i^{(k)}\right)\right), \tag{35}$$

where $\lambda_j^{(k)}$ *is the* j*th row sum of the matrix* $\mathbf{R}_k$*.*

Proof. We take $\mathbf{d} = (1,1,\dots,1) \in \mathbb{R}^m$ and $\mathbf{c}_k = \mathbf{d}\mathbf{R}_k^T$ for $k \in \{1,\dots,N\}$. Hence $\mathbf{c}_k = \lambda_k$, where $\lambda_k = (\lambda_1^{(k)}, \lambda_2^{(k)}, \dots, \lambda_n^{(k)})$ is the vector of row sums of the matrix $\mathbf{R}_k$. So, we have $c_j^{(k)} = \lambda_j^{(k)}$ for $j \in \{1,\dots,n\}$. Thus all assumptions of Corollary 1 are fulfilled.

In this situation inequality (32) takes the form (35), as wanted. □

REMARK 2. Similar results to (35) can be found in [14].

In the rest of this section we assume that $m = n$.

COROLLARY 3. *Let* $f : \mathbb{R}_{++} \to \mathbb{R}$ *be a convex function on* $\mathbb{R}_{++}$*. Let* $\mathbf{q} = (q_1, q_2, \dots, q_m) \in \mathbb{R}^m_{++}$*,* $\mathbf{q}_k = (q_1^{(k)}, q_2^{(k)}, \dots, q_m^{(k)}) \in \mathbb{R}^m_{++}$*,* $k \in \{1,\dots,N\}$*.*

Assume that

$$\mathbf{q}_k \prec \mathbf{q} \quad \textit{for } k \in \{1,\dots,N\}. \tag{36}$$

Then, for any $t_k \geqslant 0$*,* $k \in \{1,\dots,N\}$*, with* $\sum_{k=1}^{N} t_k = 1$*, the following inequality holds:*

$$f\left(\sum_{j=1}^{m} q_j - \sum_{k=1}^{N} t_k \sum_{i=1}^{m-1} q_i^{(k)}\right) \leqslant \sum_{j=1}^{m} f(q_j) - \sum_{k=1}^{N} t_k \sum_{i=1}^{m-1} f\left(q_i^{(k)}\right). \tag{37}$$

Proof. Due to (36) there exists an $m \times m$ doubly stochastic matrix $\mathbf{R}_k$ such that

$$\mathbf{q}_k = \mathbf{q}\mathbf{R}_k \quad \text{for } k \in \{1,\dots,N\}. \tag{38}$$

It follows from the double stochasticity of the matrix $\mathbf{R}_k$ that

$$\lambda_k = (1,1,\dots,1) \in \mathbb{R}^m \quad \text{for } k \in \{1,\dots,N\},$$

i.e., $\lambda_j^{(k)} = 1$ is the jth row sum of the matrix $\mathbf{R}_k$ for $j \in \{1,\dots,m\}$.

For this reason inequality (35) becomes the following

$$f\left(\sum_{k=1}^{N} t_k \left(\sum_{j=1}^{m} q_j - \sum_{i=1}^{m-1} q_i^{(k)}\right)\right) \leqslant \sum_{k=1}^{N} t_k \left(\sum_{j=1}^{m} f(q_j) - \sum_{i=1}^{m-1} f\left(q_i^{(k)}\right)\right), \tag{39}$$

which easily implies (37). □

By setting $m = 2$, we conclude from (37) that

$$f\left(q_1 + q_2 - \sum_{k=1}^{N} t_k q_1^{(k)}\right) \leqslant f(q_1) + f(q_2) - \sum_{k=1}^{N} t_k f(q_1^{(k)}).$$

This is the classical Mercer inequality (see Theorem A).

REMARK 3. Corollary 3 is equivalent to Theorem 2.1 in [11]. Because Corollary 3 follows from Corollary 2, the latter extends this theorem from doubly stochastic matrices in [11] to column stochastic matrices in the present paper.

Let $\mathbf{p} = (p_1, p_2, \dots, p_m) \in \mathbb{R}^m_{++}$ be a given m-tuple. Following [9, Definition B.1., p. 585], we say that an $m \times m$ matrix $\mathbf{R} = (r_{ji})$ is $\mathbf{p}$-*stochastic*, if

(i) $r_{ji} \geqslant 0$ for $j, i \in \{1, \dots, m\}$,

(ii) $\mathbf{p} = \mathbf{pR}$,

(iii) $\mathbf{e} = \mathbf{eR}^T$, where $\mathbf{e} = (1, 1, \dots, 1) \in \mathbb{R}^m$.

We say that an m-tuple $\widetilde{\mathbf{q}} \in \mathbb{R}^m$ is $\mathbf{p}$-*majorized* by an m-tuple $\mathbf{q} \in \mathbb{R}^m$, written as $\widetilde{\mathbf{q}} \prec_{\mathbf{p}} \mathbf{q}$, if $\widetilde{\mathbf{q}} = \mathbf{qR}$ for some $\mathbf{p}$-stochastic matrix $\mathbf{R}$ (see [9, Definition B.2., p. 585]).

In the special case when $\mathbf{p} = \mathbf{e}$, the $\mathbf{p}$-stochastic matrices are exactly doubly stochastic matrices, and, in consequence, the relation of $\mathbf{e}$-majorization $\prec_{\mathbf{e}}$ becomes the standard majorization $\prec$ on $\mathbb{R}^m$ [9].

It is interesting that the relation $\widetilde{\mathbf{q}} \prec_{\mathbf{p}} \mathbf{q}$ holds if and only if the following inequality

$$\sum_{i=1}^{m} p_i \psi\left(\frac{\widetilde{q}_i}{p_i}\right) \leqslant \sum_{j=1}^{m} p_j \psi\left(\frac{q_j}{p_j}\right) \tag{40}$$

is satisfied for all real convex (continuous) functions ψ on $\mathbb{R}_+$, where $\mathbf{q} = (q_1, q_2, \dots, q_m) \in \mathbb{R}^m_{++}$ and $\widetilde{\mathbf{q}} = (\widetilde{q}_1, \widetilde{q}_2, \dots, \widetilde{q}_m) \in \mathbb{R}^m_{++}$ (see [15, Proposition 4.2] and [9, Proposition B.4., pp. 586–587]).

With the aid of the above conditions **(i)**, **(ii)** and **(iii)**, observe that the statement (40) is of the form (11) (see Theorem B for details).

We finish this section by providing an application of Theorem 1 for $\mathbf{p}$-majorization.

COROLLARY 4. *Let $f : \mathbb{R}_{++} \to \mathbb{R}$ be a convex function on $\mathbb{R}_{++}$. Let $\mathbf{p} = (p_1, p_2, \dots, p_m) \in \mathbb{R}^m_{++}$ be fixed. Let $\mathbf{q} = (q_1, q_2, \dots, q_m) \in \mathbb{R}^m_{++}$, $\mathbf{q}_k = (q_1^{(k)}, q_2^{(k)}, \dots, q_m^{(k)}) \in \mathbb{R}^m_{++}$ for $k \in \{1, \dots, N\}$.*

Assume that

$$\mathbf{q}_k \prec_{\mathbf{p}} \mathbf{q} \quad \textit{for } k \in \{1, \dots, N\}. \tag{41}$$

Then, for any $t_k \geqslant 0$, $k \in \{1, \dots, N\}$, with $\sum_{k=1}^{N} t_k = 1$, the following inequality holds

$$f\left(\sum_{k=1}^{N} t_k \sum_{j=1}^{m} \lambda_j^{(k)} \frac{q_j}{p_j} - \sum_{k=1}^{N} t_k \sum_{i=1}^{m-1} \frac{q_i^{(k)}}{p_i}\right) \leqslant \frac{1}{p_m} C_f(\mathbf{p}, \mathbf{q}) - \sum_{k=1}^{N} \frac{t_k}{p_m} C_f(\widehat{\mathbf{p}}, \widehat{\mathbf{q}}_k), \tag{42}$$

where $\lambda_j^{(k)}$ is the jth row sum of the matrix $\mathbf{S}_k = \left(s_{ji}^{(k)}\right)$ with $s_{ji}^{(k)} = r_{ji}^{(k)} \frac{p_j}{p_i}$ for $i, j \in \{1, \dots, m\}$, and $\widehat{\mathbf{p}} = (p_1, p_2, \dots, p_{m-1}) \in \mathbb{R}^{m-1}_{++}$ and $\widehat{\mathbf{q}}_k = (q_1^{(k)}, q_2^{(k)}, \dots, q_{m-1}^{(k)}) \in \mathbb{R}^{m-1}_{++}$ for $k \in \{1, \dots, N\}$.

Proof. We deduce from (41) that there exists a $\mathbf{p}$-stochastic matrix $\mathbf{R}_k$ such that

$$\mathbf{q}_k = \mathbf{q}\mathbf{R}_k\,, \quad \mathbf{p} = \mathbf{p}\mathbf{R}_k \quad \text{and} \quad \mathbf{e} = \mathbf{e}\mathbf{R}_k^T \quad \text{for } k \in \{1,\dots,N\}. \tag{43}$$

For $k \in \{1,\dots,N\}$ with $m = n$, we introduce $\mathbf{p}_k = \mathbf{p}$, $\mathbf{c}_k = \mathbf{d} = \mathbf{e} = (1,1,\dots,1) \in \mathbb{R}^m$, and $\mathbf{p}_k = (p_1^{(k)}, p_2^{(k)}, \dots, p_m^{(k)}) \in \mathbb{R}^m_{++}$, $\mathbf{d} = (d_1, d_2, \dots, d_m) \in \mathbb{R}^m_+$, $\mathbf{c}_k = (c_1^{(k)}, c_2^{(k)}, \dots, c_m^{(k)}) \in \mathbb{R}^m_+$. Hence, $p_i^{(k)} = p_i$ for $i \in \{1,\dots,m\}$, and $c_j^{(k)} = d_j = 1$ for $j \in \{1,\dots,m\}$.

From this we get $\widehat{\mathbf{p}}_k = (p_1^{(k)}, p_2^{(k)}, \dots, p_{m-1}^{(k)}) \in \mathbb{R}^{m-1}_{++}$, $\widehat{\mathbf{d}} = (d_1, d_2, \dots, d_{m-1}) \in \mathbb{R}^{m-1}_+$, $\widehat{\mathbf{d}} = \widehat{\mathbf{e}} = (1,1,\dots,1) \in \mathbb{R}^{m-1}$. Moreover, $C_f(\mathbf{p},\mathbf{q};\mathbf{c}_k)$ and $C_f(\widehat{\mathbf{p}}_k, \widehat{\mathbf{q}}_k; \widehat{\mathbf{d}})$ become $C_f(\mathbf{p},\mathbf{q})$ and $C_f(\widehat{\mathbf{p}}, \widehat{\mathbf{q}}_k)$, respectively. Furthermore, (43) ensures that

$$\mathbf{q}_k = \mathbf{q}\mathbf{R}_k\,, \quad \mathbf{p}_k = \mathbf{p}\mathbf{R}_k \quad \text{and} \quad \mathbf{c}_k = \mathbf{d}\mathbf{R}_k^T \quad \text{for } k \in \{1,\dots,N\}.$$

So, we can utilize inequality (16) in Theorem 1 to obtain (42). This completes the proof. $\square$

REMARK 4. The results of the present paper can be demonstrated for convex functions on $\mathbb{R}_+ = [0,\infty)$ and for $\mathbf{q} \in \mathbb{R}^n_+$. However, this extended approach does not include the standard divergences (entropies) generated by the minus logarithm function, etc., (see (5)–(8)).

Acknowledgement. The author is grateful to an anonymous referee for careful reading the paper and giving useful remarks and suggestions that improved 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, *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*, (Ed. S. S. Dragomir), 2000. `http://rgmia.vu.edu.au/monographs/csiszar.htm`.

[4] S. FURUICHI AND F.-C. MITROI, *Mathematical inequalities for some divergences*, Physica A, 391 (2012) 388–400.

[5] T. FURUTA, *Parametric extensions of Shannon inequality and its reverse one in Hilbert space operators*, Linear Algebra Appl., 381 (2004) 219–235.

[6] H. ISA, M. ITO, E. KAMEI, H. TOHYAMA AND M. WATANABE, *Generalizations of operator Shannon inequality based on Tsallis and Rényi relative entropies*, Linear Algebra Appl., 439 (10) (2013) 3148–3155.

[7] P. KLUZA AND M. NIEZGODA, *Generalizations of Crooks and Lin's results on Jeffreys–Csiszár and Jensen–Csiszár f-divergences*, Physica A, 463 (2016) 383–393.

[8] 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.

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

[10] A. MCD. MERCER, *A variant of Jensen's inequality*, J. Ineq. Pure & Appl. Math., 4 (2003), Article 73.

[11] M. NIEZGODA, *A generalization of Mercer's result on convex functions*, Nonlinear Analysis, 71 (2009) 2771–2779.
[12] M. NIEZGODA, *On Shannon like inequalities for f-connections of positive linear maps and positive operators*, Linear Algebra Appl., 481 (2015) 186–201.
[13] M. NIEZGODA, *Vector joint majorization and generalization of Csiszár–Körner's inequality for f-divergence*, Discrete Applied Math., 198 (2016) 195–205.
[14] Z. PAVIĆ, *Extension of Jensen's inequality to affine combinations*, J. Inequal. Appl., 2014, 2014:298.
[15] F. VOM ENDE AND G. DIRR, *The d-majorization polytope*, arXiv:1911.01061v1, 4 Nov 2019.
[16] K. YANAGI, K. KURIYAMA AND S. FURUICHI, *Generalized Shannon inequalities based on Tsallis relative operator entropy*, Linear Algebra Appl., 394 (2005) 109–118.

(Received January 9, 2020)

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