HOW THE TYPE OF CONVEXITY OF THE CORE FUNCTION
AFFECTS THE CSISZÁR f–DIVERGENCE FUNCTIONAL

MOHSEN KIAN

(*Communicated by C. P. Niculescu*)

Abstract. We investigate how the type of convexity of the core function affects the Csiszár f-divergence functional. A general treatment for the type of convexity has been considered and the associated perspective functions have been studied. In particular, it has been shown that when the core function is MN-convex, then the associated perspective function is jointly MN-convex if the two scalar means M and N are the same. In the case where M $\neq$ N, we study the type of convexity of the perspective function. As an application, we prove that the *Hellinger distance* is jointly GG-convex.

1. Introduction

In the probability theory, the notion of Csiszár f-divergence is well-known in relation with measures between probability distributions. Those kinds of measures have many applications in many directions, like economics, genetics, signal processing and so on. In fact, Csiszár [2, 3] introduced f-divergence functional of a function $f : [0,\infty) \to \mathbb{R}$ by

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

for n-tuples of positive real numbers $\mathbf{p} = (p_1,\dots,p_n)$ and $\mathbf{q} = (q_1,\dots,q_n)$. In above definition, the undefined expressions are interpreted as

$$f(0) = \lim_{t\to 0^+} f(t), \qquad 0f\left(\frac{0}{0}\right) = 0, \qquad 0f\left(\frac{p}{0}\right) = \lim_{\varepsilon\to 0^+} f\left(\frac{p}{\varepsilon}\right) = p\lim_{t\to\infty}\frac{f(t)}{t}.$$

A useful fact concerning the f-divergence functional was proved by Csiszár and Körner [4] as follows. In fact, they showed that the perspective function of a convex function is sub-additive.

THEOREM A. *If $f : [0, \infty) \to \mathbb{R}$ is convex, then $I_f(\mathbf{p}, \mathbf{q})$ is jointly convex in $\mathbf{p}$ and $\mathbf{q}$ and*

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

for all positive n-tuples $\mathbf{p} = (p_1, \ldots, p_n)$ and $\mathbf{q} = (q_1, \ldots, q_n)$, where the perspective function g associated to f is defined by

$$g(x, y) := y f\left(\frac{x}{y}\right).$$

We remark that Theorem A was independently proved by E. K. Godunova in [9]. The reader is referred to [14] for more on perspective functions.

When f varies through convex functions, the Csiszár f-divergence produces different known measures. Among others, we mention the following notable measures:

— *Kullback–Leibler distance* is defined by $KL(\mathbf{p}, \mathbf{q}) := \sum_{j=1}^{n} p_j \log\left(\frac{p_j}{q_j}\right)$ and $KL = I_f$, when $f(t) = t \ln t \ (t > 0)$.

— *Total variation distance* is defined by $V(\mathbf{p}, \mathbf{q}) := \sum_{j=1}^{n} |p_j - q_j|$ and $V = I_f$, when $f(t) = |t - 1| \ (t \geqslant 0)$.

— *Hellinger distance* is defined by $H^2(\mathbf{p}, \mathbf{q}) := 2\sum_{j=1}^{n} \left(\sqrt{p_j} - \sqrt{q_j}\right)^2$ and $H^2 = I_f$, when $f(t) = 2(\sqrt{t} - 1)^2 \ (t \geqslant 0)$.

— χ^2*-distance* is defined by $D_{\chi^2}(\mathbf{p}, \mathbf{q}) := \sum_{j=1}^{n} \frac{(p_j - q_j)^2}{q_j}$ and $D_{\chi^2} = I_f$, when $f(t) = (t-1)^2 \ (t \geqslant 0)$.

— *Rényi's divergences* are defined by $R_\alpha(\mathbf{p}, \mathbf{q}) := \frac{1}{\alpha(\alpha - 1)} \ln \rho_\alpha(\mathbf{p}, \mathbf{q})$ for every $\alpha \in \mathbb{R} \setminus \{0, 1\}$, where $\rho_\alpha(\mathbf{p}, \mathbf{q}) = \sum_{j=1}^{n} p_j^\alpha q_j^{1-\alpha}$ and $\rho_\alpha = I_f$, when $f(t) = t^\alpha \ (t > 0)$.

For more information about f-divergence functional and its properties, the reader is referred to [8, 10, 11, 12, 15] and references therein.

For every two positive real numbers x, y and every $\alpha \in [0, 1]$, the most well-known scalar means read as follows:

$$\begin{aligned} \mathbf{A}_\alpha(x, y) &= \alpha x + (1 - \alpha)y & &\text{Arithmetic mean} \\ \mathbf{G}_\alpha(x, y) &= x^\alpha y^{1-\alpha} & &\text{Geometric mean} \\ \mathbf{H}_\alpha(x, y) &= \left(\alpha x^{-1} + (1 - \alpha)y^{-1}\right)^{-1} & &\text{Harmonic mean.} \end{aligned}$$

The Arithmetic-Geometric-Harmonic means inequality is well-known:

$$\mathbf{H}_\alpha(x, y) \leqslant \mathbf{G}_\alpha(x, y) \leqslant \mathbf{A}_\alpha(x, y), \qquad (x, y \geqslant 0, \ \alpha \in [0, 1]). \tag{2}$$

2. MN-convexity

Convex functions are known to be defined using the Arithmetic mean: A real function f is convex when

$$f(\mathbf{A}_\alpha(x,y)) \leqslant \mathbf{A}_\alpha(f(x),f(y))$$

for all x,y in domain of f and every $\alpha \in [0,1]$. However, when the Arithmetic means are replaced by other means in both sides of the above inequality, different types of convexities for functions can be derived. In the next definition, we limit the domain and the range of our function to the positive half-line, while it will be possible to consider this sets more general subsets of real functions depending on occasions. If $\mathbf{M}_\alpha$ and $\mathbf{N}_\alpha$ are two α-weighted scalar means, a positive real function f on $(0,\infty)$ is said to be MN-convex, when

$$f(\mathbf{M}_\alpha(x,y)) \leqslant \mathbf{N}_\alpha(f(x),f(y)) \tag{3}$$

holds for all $x,y \geqslant 0$ and every $\alpha \in [0,1]$. Note that an AA-convex function is simply called convex. Moreover, some of these functions enjoy well-known titles. For example, HA-convex functions are called Harmonically convex and AG-convex functions are known as log-convex functions.

Some basic facts concerning MN-convex functions are given in the following lemma. The reader is referred to [1, 7, 13, 14, 16] to see the proofs and more information about these functions.

LEMMA 1. *Let f be a positive real function on $(0,\infty)$.*
(i) *If f is AG-convex if and only if $\log f$ is convex;*
(ii) *f is AH-convex if and only if $1/f$ is concave;*
(iii) *f is GA-convex (concave) if and only if $f \circ \exp$ is convex (concave);*
(iv) *If h is convex (concave), then $f(t) = h(\ln t)$ is GA-convex (concave);*
(v) *f is GG-convex if and only if the function $h = \ln \circ f \circ \exp$ is convex;*
(vi) *f is GG-convex if and only if $h = \ln \circ f$ is GA-convex;*
(vii) *f is GH-convex (concave) if and only if $f \circ \exp$ is AH-convex (concave);*
(iix) *f is HG-convex if and only if $h(t) = t \ln f(t)$ is convex;*
(ix) *f is HG-convex if and only if $\ln f$ is HA-convex;*
(x) *f is HH-convex (concave) if and only if $h(t) = t/f(t)$ is concave (convex).*

We note that each class of MN-convex functions we mentioned in Lemma 1 actually contains many examples. So we give several examples before we continue.

EXAMPLE 1. The functions $t \mapsto 1/\sqrt{t}$ and $t \mapsto -t^{-3}$ are AH-convex on $(0,\infty)$.
The functions $t \mapsto \exp t$ and $t \mapsto t^r$ $(r < 0)$ are AG-convex on $\mathbb{R}$ and $(0,\infty)$, respectively.
The function $t \mapsto \log(1+t)$ is GA-convex on $(0,\infty)$. Moreover, recall that the well-known digamma function is defined by $\psi(t) = \frac{d}{dt} \log \Gamma(t) = \frac{\Gamma(t)}{\Gamma'(t)}$ on $(0,\infty)$, where Γ denotes the gamma function, i.e., $\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx$. It is known that [16] the

functions $t \mapsto \psi(t) + \frac{1}{2t}$ and $t \mapsto \psi(t) + \frac{1}{2t} + \frac{1}{12t^2}$ are GA-concave and GA-convex, respectively, on $(0, \infty)$.

It has been shown in [13] that if $f(t) = \sum_{n=0}^{\infty} c_n t^n$ is a real analytic function whose radius of convergence is $r > 0$ and whose coefficients c_n are non-negative, then f is a GG-convex function on $(0, r)$. This implies that the functions exp, sinh and cosh are GG-convex on $\mathbb{R}$ and the functions sec, csc and tan are GG-convex on $(0, \pi/2)$. In addition, the functions $t \mapsto (1 - t)^{-1}$ and $t \mapsto \frac{1+t}{1-t}$ are GG-convex on $(0, 1)$, see [5].

The functions $t \mapsto \frac{1}{\sqrt{\ln t}}$ and $t \mapsto -(\ln t)^{-3}$ are GH-convex on $(0, \infty)$.

For all $r \geqslant 0$ and $r \leqslant -1$, the function $t \mapsto \exp(t^r)$ are HG-convex on $(0, \infty)$.

The functions $t \mapsto \frac{t}{\ln t}$ and $t \mapsto t^r$ $(0 \leqslant r \leqslant 1)$ are HH-convex on $(0, \infty)$.

Regarding the Jensen inequality, Lemma 1 can be used to demonstrate variants of the Jensen inequality for every MN-convex function. For more information on MN-convexity the reader can refer to [1, 5, 13, 14].

3. The effect of type of convexity of the core function on the f-divergence functional

We begin with modifications of the celebrated result of Csiszár, Theorem A. A consequence of Theorem A is that if f is convex, the associated perspective function g_f is convex in both variables. In the next theorem, we investigate the effect of the type of convexity of the generating function f on the convexity of the associated perspective function g_f. When there is no fear of ambiguity, we briefly use g for the associated perspective function of f. Once more, we note that although we restrict the domain and the range of our function to the positive half-line, depending on the situation, it is possible to consider this sets more general subsets of real functions.

THEOREM 1. *Let $f : (0, \infty) \to (0, \infty)$ be a real function.*

(i) *f is AH-convex if and only if g is AH-convex on the first coordinate and convex on the second coordinate. In particular, the inequality*

$$g\left(\mathbf{A}_\alpha(a,b), \mathbf{A}_\alpha(x,y)\right) \leqslant \mathbf{H}_\alpha\left\{[\mathbf{A}_\alpha(g(a,x), g(a,y))], [\mathbf{A}_\alpha(g(b,x), g(b,y))]\right\} \quad (4)$$

holds for all $a, b, x, y \geqslant 0$ and every $\alpha \in [0, 1]$.

(ii) *f is AG-convex if and only if g is AG-convex on the first coordinate and convex on the second coordinate. In particular, the inequality*

$$g\left(\mathbf{A}_\alpha(a,b), \mathbf{A}_\alpha(x,y)\right) \leqslant \mathbf{G}_\alpha\left\{\mathbf{A}_\alpha(g(a,x), g(a,y)), \mathbf{A}_\alpha(g(b,x), g(b,y))\right\} \quad (5)$$

holds, for all $a, b, x, y \geqslant 0$ and every $\alpha \in [0, 1]$.

(iii) *f is GG-convex if and only if g is jointly GG-convex. In particular,*

$$g\left(\mathbf{G}_\alpha(a,b), \mathbf{G}_\alpha(x,y)\right) \leqslant \mathbf{G}_\alpha\left\{g(a,x), g(b,y)\right\} \quad (6)$$

for all $a, b, x, y \geqslant 0$ and every $\alpha \in [0, 1]$.

(iv) f is HH-*convex if and only if* g *is jointly* HH-*convex. In particular,*

$$g\big(\mathbf{H}_\alpha(a,b),\mathbf{H}_\alpha(x,y)\big) \leqslant \mathbf{H}_\alpha\{g(a,x),g(b,y))\} \tag{7}$$

for all $a,b,x,y \geqslant 0$ *and every* $\alpha \in [0,1]$.

(v) f is GH-*convex if and only if* g *is* GH-*convex in its first variable and* GG-*convex in its second variable. In particular, the inequality*

$$g\big(\mathbf{G}_\alpha(a,b),\mathbf{G}_\alpha(x,y)\big) \leqslant \mathbf{H}_\alpha\{\mathbf{G}_\alpha(g(a,x),g(a,y)),\mathbf{G}_\alpha(g(b,x),g(b,y))\} \tag{8}$$

holds for all $a,b,x,y \geqslant 0$ *and every* $\alpha \in [0,1]$. *Moreover, in this case* g *is jointly* GG-*convex, i.e.,* (6) *holds.*

Before proving Theorem 1, we would like to note that if f is MN-convex, then g is not necessarily MN-convex in both variables, unless $M = N$. For example, if f is AH-convex, then part (i) of Theorem 1 shows that g is AH-convex on the first coordinate and convex on the second coordinate. However, g is not AH-convex in both variables. To see this, consider the AH-convex function $f(t) = 1/\sqrt{t}$ and put $\alpha = 1/2$, $a = 1$, $x = 2$ and $y = 4$. Then

$$3\sqrt{3} = g(a,\mathbf{A}_{1/2}(x,y)) \nleqslant \mathbf{H}_{1/2}(g(a,x),g(a,y)) = \frac{16\sqrt{2}}{4+\sqrt{2}}.$$

Note in addition that when f is AH-convex, g is AA-convex in both variables. However, the the reverse direction does not hold, i.e., if g is AA-convex in both variables, then f is not necessarily AH-convex.

Proof of Theorem 1. First assume that f is AH-convex. For all $x,y,a \geqslant 0$ and every $\alpha \in [0,1]$ we have

$$\begin{aligned}
g\big(\alpha a + (1-\alpha)b,x\big) &= xf\Big(\frac{1}{x}\big(\alpha a + (1-\alpha)b\big)\Big) \\
&\leqslant x\Big[\alpha f\big(\tfrac{a}{x}\big)^{-1} + (1-\alpha)f\big(\tfrac{b}{x}\big)^{-1}\Big]^{-1} \\
&= \Big[\alpha x^{-1}f\big(\tfrac{a}{x}\big)^{-1} + (1-\alpha)x^{-1}f\big(\tfrac{b}{x}\big)^{-1}\Big]^{-1} \\
&= \Big[\alpha g(a,x)^{-1} + (1-\alpha)g(b,x)^{-1}\Big]^{-1}.
\end{aligned}$$

This ensures that g is AH-convex on the first coordinate. Therefore

$$g\big(\alpha(a,x) + (1-\alpha)(b,y)\big) \leqslant \big(\alpha g(a,z)^{-1} + (1-\alpha)g(b,z)^{-1}\big)^{-1},$$

where $z = \alpha x + (1-\alpha)y$. This means that

$$g\big(\mathbf{A}_\alpha(a,b),\mathbf{A}_\alpha(x,y)\big) \leqslant \mathbf{H}_\alpha\{g(a,\mathbf{A}_\alpha(x,y)),g(b,\mathbf{A}_\alpha(x,y))\}. \tag{9}$$

On the other hand we can write

$$f\left(\frac{a}{\mathbf{A}_\alpha(x,y)}\right) = f\left(\frac{\frac{\alpha a}{yx} + \frac{(1-\alpha)a}{yx}}{\frac{\alpha}{y} + \frac{(1-\alpha)}{x}}\right) = f\left(\beta\left(\frac{a}{x}\right) + (1-\beta)\left(\frac{a}{y}\right)\right), \quad (10)$$

where $\beta = \frac{\frac{\alpha}{y}}{\frac{\alpha}{y} + \frac{(1-\alpha)}{x}} = \frac{\alpha x}{\mathbf{A}_\alpha(x,y)}$. Since f is convex, (10) implies that

$$f\left(\frac{a}{\mathbf{A}_\alpha(x,y)}\right) \leqslant \frac{\alpha x}{\mathbf{A}_\alpha(x,y)} f\left(\frac{a}{x}\right) + \frac{(1-\alpha)y}{\mathbf{A}_\alpha(x,y)} f\left(\frac{a}{y}\right).$$

Multiplying both sides by $\mathbf{A}_\alpha(x,y)$ we get

$$g\big(a, \mathbf{A}_\alpha(x,y)\big) \leqslant \mathbf{A}_\alpha(g(a,x), g(a,y)). \quad (11)$$

Similarly

$$g\big(b, \mathbf{A}_\alpha(x,y)\big) \leqslant \mathbf{A}_\alpha(g(b,x), g(b,y)). \quad (12)$$

Since the Harmonic mean is monotone, it follows from (11) and (12) that

$$\begin{aligned} g\big(\mathbf{A}_\alpha(a,b), \mathbf{A}_\alpha(x,y)\big) &\leqslant \mathbf{H}_\alpha\{g(a, \mathbf{A}_\alpha(x,y)), g(b, \mathbf{A}_\alpha(x,y))\} \quad \text{(by (9))} \\ &\leqslant \mathbf{H}_\alpha\{\mathbf{A}_\alpha(g(a,x), g(a,y)), \mathbf{A}_\alpha(g(b,x), g(b,y))\} \end{aligned}$$

which is the desired inequality (4). With $x = y$, this gives the AH-convexity of g in the first coordinate and with $a = b$ this implies the convexity of g in the second coordinate.

Conversely, if g is AH-convexity in the first coordinate, then $f(t) = g(t,1)$ is an AH-convex function, too. This completes the proof of (i).

Next suppose that f is AG-convex. For all $x, y, a \geqslant 0$ and every $\alpha \in [0,1]$ we have

$$\begin{aligned} g\big(\alpha a + (1-\alpha)b, x\big) &= xf\left(\frac{1}{x}(\alpha a + (1-\alpha)b)\right) \\ &\leqslant x\left[f\left(\frac{a}{x}\right)^\alpha f\left(\frac{b}{x}\right)^{1-\alpha}\right] \\ &= \left[xf\left(\frac{a}{x}\right)\right]^\alpha \left[xf\left(\frac{b}{x}\right)\right]^{1-\alpha} \\ &= g(a,x)^\alpha g(b,x)^{1-\alpha}, \end{aligned}$$

whence g is AG-convex in its first variable. Hence

$$g\big(\mathbf{A}_\alpha(a,b), \mathbf{A}_\alpha(x,y)\big) \leqslant \mathbf{G}_\alpha\{g(a, \mathbf{A}_\alpha(x,y)), g(b, \mathbf{A}_\alpha(x,y))\}. \quad (13)$$

Furthermore, taking into account the Arithmetic-Geometric means inequality, we know that f is a convex function so that (11) and (12) hold. Regarding the monotonicity of the Geometric mean in its both variables we conclude from (11) and (12) that

$$\mathbf{G}_\alpha\{g(a, \mathbf{A}_\alpha(x,y)), g(b, \mathbf{A}_\alpha(x,y))\} \leqslant \mathbf{G}_\alpha\{\mathbf{A}_\alpha(g(a,x), g(a,y)), \mathbf{A}_\alpha(g(b,x), g(b,y))\}. \quad (14)$$

Combining (13) and (14) we infer (5). Putting $x = y$, the AG-convexity of g in first coordinate follows from (5) and with $a = b$, (5) gives the convexity of g in the second coordinate. Conversely, if g is AG-convex in its first coordinate, then $f(t) = g(t,1)$ is AG-convex as well.

To prove (iii), let f be a GG-convex function. Then

$$f\left(\frac{\mathbf{G}_\alpha(a,b)}{\mathbf{G}_\alpha(x,y)}\right) = f\left(a^\alpha b^{1-\alpha}x^{-\alpha}y^{\alpha-1}\right) = f\left(\mathbf{G}_\alpha\left(\frac{a}{x},\frac{b}{y}\right)\right) \leqslant \mathbf{G}_\alpha\left(f\left(\frac{a}{x}\right),f\left(\frac{b}{y}\right)\right).$$

Hence

$$g\left(\mathbf{G}_\alpha(a,b),\mathbf{G}_\alpha(x,y)\right) = \mathbf{G}_\alpha(x,y)f\left(\frac{\mathbf{G}_\alpha(a,b)}{\mathbf{G}_\alpha(x,y)}\right)$$
$$\leqslant \mathbf{G}_\alpha(x,y)\mathbf{G}_\alpha\left(f\left(\frac{a}{x}\right),f\left(\frac{b}{y}\right)\right)$$
$$= \mathbf{G}_\alpha\left\{g(a,x),g(b,y)\right\}$$

as required. This proves (iii).

Next suppose that f is a HH-convex function. We write

$$f\left(\frac{\mathbf{H}_\alpha(a,b)}{\mathbf{H}_\alpha(x,y)}\right) = f\left(\left(\frac{\alpha a^{-1}+(1-\alpha)b^{-1}}{\alpha x^{-1}+(1-\alpha)y^{-1}}\right)^{-1}\right)$$
$$= f\left(\left(\frac{\alpha\frac{xy}{a}+(1-\alpha)\frac{xy}{b}}{\alpha y+(1-\alpha)x}\right)^{-1}\right)$$
$$= f\left(\left(\beta\frac{x}{a}+(1-\beta)\frac{y}{b}\right)^{-1}\right)$$

in which we set $\beta = \frac{\alpha y}{\alpha y+(1-\alpha)x}$. Since f is HH-convex, we obtain

$$f\left(\frac{\mathbf{H}_\alpha(a,b)}{\mathbf{H}_\alpha(x,y)}\right) = f\left(\mathbf{H}_\beta\left(\frac{a}{x},\frac{b}{y}\right)\right) \leqslant \mathbf{H}_\beta\left(f\left(\frac{a}{x}\right),f\left(\frac{b}{y}\right)\right)$$

so that

$$g\left(\mathbf{H}_\alpha(a,b),\mathbf{H}_\alpha(x,y)\right) = \mathbf{H}_\alpha(x,y)f\left(\frac{\mathbf{H}_\alpha(a,b)}{\mathbf{H}_\alpha(x,y)}\right) \leqslant \mathbf{H}_\alpha(x,y)\mathbf{H}_\beta\left(f\left(\frac{a}{x}\right),f\left(\frac{b}{y}\right)\right). \tag{15}$$

A simple calculation shows that

$$\mathbf{H}_\alpha(x,y)\mathbf{H}_\beta\left(f\left(\frac{a}{x}\right),f\left(\frac{b}{y}\right)\right) = \mathbf{H}_\alpha\left\{g(a,x),g(b,y)\right\}.$$

Consequently, (7) follows from (15). Hence g is jointly HH-convex. Conversely, if g is jointly HH-convex, then $f(t) = g(t,1)$ is HH-convex. This proves (iv).

Let f be a GH-convex function. For all $a,b,x \geqslant 0$ and $\alpha \in [0,1]$ we have

$$g(\mathbf{G}_\alpha(a,b),x) = xf\left(\frac{\mathbf{G}_\alpha(a,b)}{x}\right) = xf\left(\mathbf{G}_\alpha\left(\frac{a}{x},\frac{b}{x}\right)\right)$$
$$\leqslant x\mathbf{H}_\alpha\left(f\left(\frac{a}{x}\right),f\left(\frac{b}{x}\right)\right)$$
$$= \mathbf{H}_\alpha\left(g(a,x),g(b,x)\right), \qquad (16)$$

whence g is GH-convex function in its first coordinate. Furthermore, we can write

$$g(a,\mathbf{G}_\alpha(x,y)) = \mathbf{G}_\alpha(x,y)f\left(\frac{a}{\mathbf{G}_\alpha(x,y)}\right)$$
$$= \mathbf{G}_\alpha(x,y)f\left(\mathbf{G}_\alpha\left(\frac{a}{x},\frac{a}{y}\right)\right)$$
$$\leqslant \mathbf{G}_\alpha(x,y)\mathbf{H}_\alpha\left(f\left(\frac{a}{x}\right),f\left(\frac{a}{y}\right)\right)$$
$$\leqslant \mathbf{G}_\alpha(x,y)\mathbf{G}_\alpha\left(f\left(\frac{a}{x}\right),f\left(\frac{a}{y}\right)\right)$$
$$= \mathbf{G}_\alpha\left(g(a,x),g(a,y)\right), \qquad (17)$$

where the last inequality follows from the Harmonic-Geometric mean inequality. This ensures that g is GG-convex in the second coordinate. Furthermore, combining (16) and (17) and using the monotonicity of the Harmonic mean, we reach (8). In addition, a similar argument as in (16) shows that g is jointly GG-convex. Indeed,

$$g(\mathbf{G}_\alpha(a,b),\mathbf{G}_\alpha(x,y)) = \mathbf{G}_\alpha(x,y)f\left(\frac{\mathbf{G}_\alpha(a,b)}{\mathbf{G}_\alpha(x,y)}\right)$$
$$= \mathbf{G}_\alpha(x,y)f\left(\mathbf{G}_\alpha\left(\frac{a}{x},\frac{b}{y}\right)\right)$$
$$\leqslant \mathbf{G}_\alpha(x,y)\mathbf{H}_\alpha\left(f\left(\frac{a}{x}\right),f\left(\frac{b}{y}\right)\right)$$
$$\leqslant \mathbf{G}_\alpha(x,y)\mathbf{G}_\alpha\left(f\left(\frac{a}{x}\right),f\left(\frac{b}{y}\right)\right)$$
$$= \mathbf{G}_\alpha\left(g(a,x),g(b,y)\right),$$

and so g is jointly GG-convex as we claimed. The converse follows similarly as previous parts. $\square$

We give some particular corollaries of of Theorem 1 for some f-divergence functionals. It is easy to see that there are positive real numbers c_1,c_2 for which the function $f(t) = t\log t$ is AH-convex on (c_1,c_2). Theorem 1 implies that the *Kullback–Leibler distance* $KL(\mathbf{p},\mathbf{q}) = \sum_{j=1}^n p_j \log\left(\frac{p_j}{q_j}\right)$ is AH-convex on the first coordinate and convex on the second coordinate. As another example, the function $f(t) = t^r$ is AG-convex on

$(0,\infty)$ for all $r < 0$. By Theorem 1, the generated divergence functional $\rho_r(\mathbf{p},\mathbf{q}) = \sum_{j=1}^n p_j^r q_j^{1-r}$ is AG-convex on its first coordinate and convex on its second coordinate. As another example, the function $f(t) = 2(\sqrt{t}-1)^2$ is GG-convex. Accordingly, the related divergence functional, which is the *Hellinger distance* $H^2(\mathbf{p},\mathbf{q}) := 2\sum_{j=1}^n \left(\sqrt{p_j} - \sqrt{q_j}\right)^2$ is jointly GG-convex.

The MN-convexity of f produces variants of the Jensen inequality. Here, we study inequality (1) in Theorem A, when the core function f enjoys MN-convexity.

THEOREM 2. *Let* $\mathbf{a} = (a_1,\ldots,a_n)$ *and* $\mathbf{b} = (b_1,\ldots,b_n)$ *be two* n-*tuples of positive real numbers with* $\bar{\mathbf{a}} = \sum_{i=1}^n a_i$ *and* $\bar{\mathbf{b}} = \sum_{i=1}^n b_i$. *Let* f *be a positive real function on* $(0,\infty)$.

(i) *If* f *is an* AH-*convex function, then*

$$g\left(\bar{\mathbf{a}},\bar{\mathbf{b}}\right) \leqslant \bar{\mathbf{a}}^2 I_{\frac{1}{f}}(\mathbf{a},\mathbf{b})^{-1} \leqslant I_f(\mathbf{a},\mathbf{b}). \tag{18}$$

(ii) *If* f *is an* AG-*convex function, then*

$$g\left(\bar{\mathbf{a}},\bar{\mathbf{b}}\right) \leqslant \bar{\mathbf{b}} \, \exp\left[\frac{1}{\bar{\mathbf{b}}} \, I_{\log f}(\mathbf{a},\mathbf{b})\right] \leqslant I_f(\mathbf{a},\mathbf{b}). \tag{19}$$

(iii) *If* f *is a* HA-*convex function, then*

$$g\left(\bar{\mathbf{a}},\bar{\mathbf{b}}\right) = \frac{\bar{\mathbf{b}}}{\bar{\mathbf{a}}}g_\varphi\left(\bar{\mathbf{b}},\bar{\mathbf{a}}\right) = \frac{\bar{\mathbf{b}}}{\bar{\mathbf{a}}}\, g_\phi\left(\bar{\mathbf{a}},\bar{\mathbf{b}}\right) \leqslant \frac{\bar{\mathbf{b}}}{\bar{\mathbf{a}}}I_\varphi(\mathbf{b},\mathbf{a}) = \frac{\bar{\mathbf{b}}}{\bar{\mathbf{a}}}I_\phi(\mathbf{a},\mathbf{b}), \tag{20}$$

where $\varphi(t) = f(1/t)$ *and* $\phi(t) = tf(t)$.

(iv) *If* f *is an increasing* GA-*convex function, then*

$$g\left(\bar{\mathbf{a}},\bar{\mathbf{b}}\right) \leqslant g_{f\circ\exp}\left(\bar{\mathbf{a}},\bar{\mathbf{b}}\right) \leqslant I_{f\circ\exp}(\mathbf{a},\mathbf{b}). \tag{21}$$

Proof. Suppose that $\mathbf{a},\mathbf{b}$ are n-tuples of positive real numbers. For every $i = 1,\ldots,n$, we set $\beta_i = \frac{b_i}{\sum_{k=1}^n b_k}$ so that $(\beta_1,\ldots,\beta_n)$ is a probability vector. First assume that f is an AH-convex function for which we can write

$$f\left(\frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n b_i}\right) = f\left(\frac{a_1}{\sum_{k=1}^n b_k} + \cdots + \frac{a_n}{\sum_{k=1}^n b_k}\right)$$

$$= f\left(\frac{a_1}{b_1}\frac{b_1}{\sum_{k=1}^n b_k} + \cdots + \frac{a_n}{b_n}\frac{b_n}{\sum_{k=1}^n b_k}\right)$$

$$\leqslant \left(\sum_{i=1}^n \frac{b_i}{\sum_{k=1}^n b_k}\frac{1}{f(\frac{a_i}{b_i})}\right)^{-1} \tag{22}$$

where we use the AH-convexity of f with $\beta_i = \frac{b_i}{\sum_{k=1}^n b_k}$. Multiplying both sides of (22) by $\sum_{k=1}^n b_k$ we get

$$
\begin{aligned}
\Big(\sum_{k=1}^n b_k\Big) f\Big(\frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n b_i}\Big) &\leqslant \Big(\sum_{k=1}^n b_k\Big)\Big(\sum_{i=1}^n \frac{b_i}{\sum_{k=1}^n b_k}\frac{1}{f(\frac{a_i}{b_i})}\Big)^{-1} \\
&= \Big(\sum_{k=1}^n b_k\Big)^2 \Big(\sum_{i=1}^n b_i \frac{1}{f(\frac{a_i}{b_i})}\Big)^{-1} \\
&= \Big(\sum_{k=1}^n b_k\Big)^2 I_{\frac{1}{f}}(\mathbf{a},\mathbf{b})^{-1},
\end{aligned}
$$

which implies the first inequality in (18). To get the second inequality we use the convexity of the function $t \mapsto t^{-1}$.

$$
\Big(\sum_{k=1}^n b_k\Big)^2 I_{\frac{1}{f}}(\mathbf{a},\mathbf{b})^{-1} = \Big(\sum_{k=1}^n b_k\Big)\Big(\sum_{i=1}^n \frac{b_i}{\sum_{k=1}^n b_k}\frac{1}{f(\frac{a_i}{b_i})}\Big)^{-1} \leqslant \sum_{i=1}^n b_i f\Big(\frac{a_i}{b_i}\Big) = I_f(\mathbf{a},\mathbf{b}).
$$

This completes the proof of (i). Next assume that f is an AG-convex function. Then

$$
f\Big(\frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n b_i}\Big) = f\Big(\sum_{i=1}^n \beta_i \frac{a_i}{b_i}\Big) \leqslant \prod_{i=1}^n f\Big(\frac{a_i}{b_i}\Big)^{\beta_i}, \tag{23}
$$

in which we use the same convex coefficients β_i as in the proof of (i). Moreover,

$$
\begin{aligned}
\prod_{i=1}^n f\Big(\frac{a_i}{b_i}\Big)^{\beta_i} &= \prod_{i=1}^n \exp\Big[\frac{b_i}{\sum_{k=1}^n b_k}\log f\Big(\frac{a_i}{b_i}\Big)\Big] \\
&= \exp\Big[\sum_{i=1}^n \frac{b_i}{\sum_{k=1}^n b_k}\log f\Big(\frac{a_i}{b_i}\Big)\Big] = \exp\Big[\frac{1}{\mathbf{b}} I_{\log f}(\mathbf{a},\mathbf{b})\Big]. \tag{24}
\end{aligned}
$$

The left inequality in (19) follows from (23) and (24). In addition, utilising the Arithmetic-Geometric means inequality we reach

$$
\prod_{i=1}^n f\Big(\frac{a_i}{b_i}\Big)^{\beta_i} \leqslant \sum_{i=1}^n \beta_i f\Big(\frac{a_i}{b_i}\Big) = \frac{1}{\mathbf{b}} I_f(\mathbf{a},\mathbf{b}) \tag{25}
$$

and the right inequality in (19) is derived. This concludes (ii).

Now assume that f is a HA-convex function. It is not hard to see that [5] the functions $\varphi(t) = f(1/t)$ and $\phi(t) = tf(t)$ are convex on proper domains so that Theorem A gives $g_\varphi\big(\bar{\mathbf{a}},\bar{\mathbf{b}}\big) \leqslant I_\varphi(\mathbf{a},\mathbf{b})$ and $g_\phi\big(\bar{\mathbf{a}},\bar{\mathbf{b}}\big) \leqslant I_\phi(\mathbf{a},\mathbf{b})$. We consider the convex coefficients $\alpha_i = \frac{a_i}{\sum_{k=1}^n a_k}$ for $i = 1,\ldots,n$ in such a way that

$$
f\Big(\frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n b_i}\Big) = f\Big(\Big(\frac{\sum_{i=1}^n b_i}{\sum_{i=1}^n a_i}\Big)^{-1}\Big) = f\Big(\Big(\sum_{i=1}^n \frac{b_i}{a_i}\frac{a_i}{\sum_{k=1}^n a_k}\Big)^{-1}\Big) = f\Big(\Big(\sum_{i=1}^n \alpha_i \frac{b_i}{a_i}\Big)^{-1}\Big).
$$

By the HA-convex of f, this concludes that

$$f\left(\frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i}\right) \leqslant \sum_{i=1}^{n} \alpha_i f\left(\frac{a_i}{b_i}\right) = \frac{1}{\sum_{k=1}^{n} a_k} \sum_{i=1}^{n} a_i f\left(\frac{a_i}{b_i}\right).$$

Multiplying both sides by $\sum_{k=1}^{n} b_k$ we reach

$$g\left(\overline{\mathbf{a}}, \overline{\mathbf{b}}\right) \leqslant \frac{\overline{\mathbf{b}}}{\overline{\mathbf{a}}} I_\varphi(\mathbf{b}, \mathbf{a}).$$

On the other hand,

$$g\left(\overline{\mathbf{a}}, \overline{\mathbf{b}}\right) = \overline{\mathbf{b}} f\left(\frac{\overline{\mathbf{a}}}{\overline{\mathbf{b}}}\right) = \frac{\overline{\mathbf{b}}}{\overline{\mathbf{a}}} \overline{\mathbf{a}} \varphi\left(\frac{\overline{\mathbf{b}}}{\overline{\mathbf{a}}}\right) = \frac{\overline{\mathbf{b}}}{\overline{\mathbf{a}}} g_\varphi\left(\overline{\mathbf{b}}, \overline{\mathbf{a}}\right) = \frac{\overline{\mathbf{b}}}{\overline{\mathbf{a}}} g_\phi\left(\overline{\mathbf{a}}, \overline{\mathbf{b}}\right).$$

Furthermore, we compute

$$I_\varphi\left(\mathbf{b}, \mathbf{a}\right) = \sum_{i=1}^{n} a_i \varphi\left(\frac{b_i}{a_i}\right) = \sum_{i=1}^{n} b_i \frac{a_i}{b_i} f\left(\frac{a_i}{b_i}\right) = I_\phi\left(\mathbf{a}, \mathbf{b}\right),$$

so that we arrive at (iii).

For proving (iv), first note that a function f is GA-convex if and only if the function $t \mapsto f(e^t)$ is convex, indeed, when proper domains are considered. So the Csiszár inequality in Theorem A implies the right inequality of (21):

$$g_{f \circ \exp}\left(\overline{\mathbf{a}}, \overline{\mathbf{b}}\right) \leqslant I_{f \circ \exp}(\mathbf{a}, \mathbf{b}). \tag{26}$$

When f is increasing, we have $f \circ \exp \geqslant f$ on the positive half line. This ensures that the left inequality in (iv) is valid. $\quad\square$

REFERENCES

[1] G. D. ANDERSON, M. K. VAMANAMURTHY AND M. VUORINEN, *Generalized convexity and inequalities*, J. Math. Anal. Appl. **335** (2007), 1294–1308.

[2] I. CSISZÁR, *Information Measures: A Critical Survey*, Trans. 7'th Prague Conf. on Info. Th., Statist. Decis. Funct., Random Processes and 8'th European Meeting of Statist., Volume B, Academia Prague (1978), 73–86.

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

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

[5] S. S. DRAGOMIR, *Hermite-Hadamard type inequalities for MN-convex functions*, Aust. J. Math. Anal. Appl. **18** (2021), No. 1, Art. 1, 127 pp.

[6] S. S. DRAGOMIR, *Inequalities for Csiszár f-divergence in information theory*, RGMIA Monographs, Victoria University, 2000.

[7] S. S. DRAGOMIR, *Inequalities of Jensen type for* AH-*convex functions*, J. Numer. Anal. Approx. Theory, **45** (2016), 128–146.

[8] G. L. GILARDONI, *On Pinsker's and Vajda's type inequalities for Csiszár's f -divergences*, IEEE Trans. Inf. Theory **56** (2010), 5377–5386.

[9] E. K. GODUNOVA, *Convexity of composed functions and its application to the proof of inequalities*, Mat. Zametki **1** (1967), 495–500, (Russian).

[10] J.-B. HIRIART-URRUTY AND J.-E. MARTÍNEZ-LEGAZ, *Convex solutions of a functional equation arising in information theory*, J. Math. Anal. Appl. **328** (2007), 1309–1320.

[11] M. KIAN, *A characterization of mean values for Csiszár's inequality and applications*, Indaga. Math. **25** (2014) 505–515.

[12] M. S. MOSLEHIAN AND M. KIAN, *Jensen type inequalities for Q-class functions*, Bull. Aust. Math. Soc. **85** (2012), 128–142.

[13] C. P. NICULESCU, *Convexity according to the geometric mean*, Math. Inequal. Appl., **3** (2000), 155–167.

[14] C. P. NICULESCU AND L.-E. PERSSON, *Convex Functions and their Applications. A Contemporary Approach, Second Edition*, CMS books in mathematics vol. 23, Springer-Verlag, New York, 2018.

[15] I. SASON, *On f -divergences: integral representations, local behavior, and inequalities*, Entropy **20** (2018), 383.

[16] X.-M. ZHANG, Y.-M. CHU AND X.-H. ZHANG, *The Hermite-Hadamard type inequality of* GA-*convex functions and its application*, Journal of Inequalities and Applications, Volume 2010, Article ID 507560, 11 pages.

Mohsen Kian
Department of Mathematics
University of Bojnord
P. O. Box 1339, Bojnord 94531, Iran
e-mail: kian@ub.ac.ir

Mathematical Inequalities & Applications
www.ele-math.com
mia@ele-math.com