Available online at www.sciencedirect.com
ScienceDirect
J. Math. Anal. Appl. 328 (2007) 1309–1320
Journal of MATHEMATICAL ANALYSIS AND APPLICATIONS
www.elsevier.com/locate/jmaa

Convex solutions of a functional equation arising in information theory

J.-B. Hiriart-Urruty [a], J.-E. Martínez-Legaz [b,*,1]

[a] *Laboratoire MIP, UMR CNRS 5640, Institut de Mathématiques, Université Paul Sabatier, 118, route de Narbonne, 31062, Toulouse cedex 09, France*

[b] *Department d'Economia i d'Història Econòmica, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain*

Received 3 March 2006

Available online 14 July 2006

Submitted by Steven G. Krantz

Abstract

Given a convex function f defined for positive real variables, the so-called Csiszár f-divergence is a function I_f defined for two n-dimensional probability vectors $p = (p_1, \dots, p_n)$ and $q = (q_1, \dots, q_n)$ as $I_f(p, q) := \sum_{i=1}^{n} q_i f(\frac{p_i}{q_i})$. For this generalized measure of entropy to have distance-like properties, especially symmetry, it is necessary for f to satisfy the following functional equation: $f(x) = xf(\frac{1}{x})$ for all $x > 0$. In the present paper we determine all the convex solutions of this functional equation by proposing a way of generating all of them. In doing so, existing usual f-divergences are recovered and new ones are proposed.

Keywords: Csiszár divergence; Convex functions; Information theory; Functional equations

* Corresponding author.
E-mail address: juanenrique.martinez@uab.es (J.-E. Martínez-Legaz).

[1] The research of the second author was supported in part by the Ministerio de Ciencia y Tecnología, Project MTM2005-08572-C03-03. He also thanks the support of the Barcelona Economics Program of CREA.

0022-247X/$ – see front matter
doi:10.1016/j.jmaa.2006.06.035

1. Motivation

Given a convex function $f:\mathbb{R}_+ \to \mathbb{R}\cup\{+\infty\}$, the so-called *Csiszár f-divergence* is a function

$$I_f:\mathbb{R}_+^n \times \mathbb{R}_+^n \to \mathbb{R},$$
$$(p,q) \mapsto I_f(p,q) := \sum_{i=1}^n q_i f\left(\frac{p_i}{q_i}\right).$$

(Here $\mathbb{R}_+^n$ stands for the positive orthant of $\mathbb{R}^n$, that is $\{p=(p_1,\dots,p_n) \mid p_i \geqslant 0 \text{ for all } i\}$, and the specific meaning of $q_i f(\frac{p_i}{q_i})$ for $q_i = 0$ will be made precise further.) The function I_f is a generalized measure of entropy whose distance-like properties make it useful in information theory, stochastic optimization and several other applications. This general notion of divergence measures in a certain sense the "distance" between two probability distributions. The function f is sometimes called the *kernel* of the f-divergence I_f. Indeed $I_f(p,q)$ enjoys certain properties of a distance between $p=(p_1,\dots,p_n)$ and $q=(q_1,\dots,q_n)$, but, without further assumptions on the kernel function f, it is not a metric. First of all, to have $I_f(p,q)=0$, it is necessary to impose $f(1)=0$, a *normalization* which is always possible to get at, provided that a specific affine function is added to f (see the first example in Section 2). Secondly, observe that $I_f(q,p) = I_{f^\diamond}(p,q)$, where $f^\diamond$ is another kernel function defined as follows: for all $x>0$, $f^\diamond(x) = xf(\frac{1}{x})$ (with an *ad hoc* limiting value at 0, see below). Moreover, as known in convex analysis (cf. [1, vol. I, p. 5] for example), *this involution $f \rightsquigarrow f^\diamond$ preserves convexity*: $f^\diamond$ is convex if and only if f is convex. So, for symmetry reasons on I_f, it is highly desirable that the kernel function f satisfy $f^\diamond = f$. Thirdly, the convexity of f ensures that of I_f (more will be said on that later on), whence $I_f(p+r,q+s) \leqslant I_f(p,q)+I_f(r,s)$. All these questions motivate the present work the aim of which is to answer the following question: *what are the convex functions f solving the functional equation*

$$f^\diamond = f? \tag{1}$$

The sequel of the paper is organized into two sections. In Section 2, we list some examples of convex functions solving (1): some of them give rise to the best known divergence functions used in mathematical statistics, information theory and signal processing; some are new and are built up from the characterization theorem (of solutions of (1)) derived in Section 3. Section 3 contains the main result: a convex function f satisfies (1) if and only if a certain closed convex set C_f in the plane, associated one-to-one with f, is symmetric with respect to the first bisecting line. Thus, the "functional" equation $f = f^\diamond$ is translated in terms of the simplest "geometrical" involution in the plane: a symmetry with respect to a line. This allows us to generate systematically and easily all the convex solutions of (1).

2. Mathematical setting and first examples

We begin by setting the mathematical context of our presentation and the notations used. The reader is supposed to be familiar with basic concepts of convex analysis, for which we will adopt the standard terminology of [1,2].

2.1.

We denote by $\Gamma(\mathbb{R}_+)$ the set of functions $\varphi : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ satisfying the following properties:

$$\left.\begin{array}{l} \bullet\ \varphi(x) = +\infty \text{ for all } x < 0, \\ \quad \varphi \text{ is finite at some point } x_0 > 0; \\ \bullet\ \varphi \text{ is convex and lower-semicontinuous on } \mathbb{R} \\ \quad \text{(one also says that it is closed convex on } \mathbb{R}). \end{array}\right\}$$

The properties of $\varphi \in \Gamma(\mathbb{R}_+)$ are classical and can be found in any book on convex analysis, see [1, Chapter I, vol. I] for example. We cast some of them:

- If $\varphi \in \Gamma(\mathbb{R}_+)$, the value $\varphi(0)$ is determined by the values of φ on $(0, \epsilon)$ for $\epsilon > 0$ as small as desired;
- If $\varphi \in \Gamma(\mathbb{R}_+)$, so is $\varphi^\diamond$. In particular, $\varphi^\diamond(0)$ is the "slope at infinity" of φ, that is
$$\varphi'_\infty = \lim_{t \to +\infty} \frac{\varphi(x_0 + t) - \varphi(x_0)}{t},$$
where $x_0 > 0$ is any point at which φ is finite.

2.2.

Let us denote by $\mathfrak{G}(\mathbb{R}_+)$ the set of $f \in \Gamma(\mathbb{R}_+)$ such that $f = f^\diamond$. The next properties clearly come from the definitions themselves: if $f \in \mathfrak{G}(\mathbb{R}_+)$,

- $f(0) = f'_\infty$ (equality in $\mathbb{R} \cup \{+\infty\}$);
- Either f is finite on some $(0, \epsilon)$, $\epsilon > 0$, in which case dom f (the set on which f is finite) is $(0, +\infty)$ or f takes the value $+\infty$ on some $(0, \epsilon)$, in which case dom f is a line-segment with end-points r and $1/r$ for some $0 < r < 1$;
- f is always finite-valued at $x_0 = 1$. The "extreme" case is when f is the indicator function of the singleton $\{1\}$ plus a constant (i.e., dom $f = \{1\}$, that is: $f(1) = c \in \mathbb{R}$, $f(x) = +\infty$ otherwise). In the other cases, dom f contains a line-segment $[r, 1/r]$ for some $0 < r < 1$ (so that $1 \in \operatorname{int}(\operatorname{dom} f)$).

If we start with φ in $\Gamma(\mathbb{R}_+)$, $T(\varphi) := \frac{1}{2}(\varphi + \varphi^\diamond)$ clearly lies in $\mathfrak{G}(\mathbb{R}_+)$, if dom f contains a line-segment $[r, 1/r]$ for some $0 < r \leqslant 1$. Evidently, $T(f) = f$ if and only if $f \in \mathfrak{G}(\mathbb{R}_+)$. Note also that $\varphi^\diamond$ is normalized (i.e., $\varphi^\diamond(1) = 0$) whenever φ is; so is $T(\varphi)$. We call $T(\varphi)$ the *symmetrized form* of the kernel function φ.

The transform $T(\cdot)$ may serve to build up functions in $\mathfrak{G}(\mathbb{R}_+)$; it however does not help much for parameterizing $\mathfrak{G}(\mathbb{R}_+)$, for two reasons: firstly it is not one-to-one, secondly (and more important), it is difficult to devise the appropriate φ such that $T(\varphi) = f$ when specific properties on f are desired (like the behavior around a point). $\mathfrak{G}(\mathbb{R}_+)$ is a *convex cone* (of $\Gamma(\mathbb{R}_+)$), closed for the topology of pointwise convergence. Note also the following properties:

- The only functions $f \in \mathfrak{G}(\mathbb{R}_+)$ which are constant on their domain (with $1 \in \operatorname{int}(\operatorname{dom} f)$) are indicator functions of line-segments $[r, 1/r]$ for some $0 < r < 1$, plus the other "extreme" case, the indicator function of $\mathbb{R}_+$. Thus, if $f \in \mathfrak{G}(\mathbb{R}_+)$ and $1 \in \operatorname{int}(\operatorname{dom} f)$, $f + c \notin \mathfrak{G}(\mathbb{R}_+)$ whenever $c \neq 0$;
- $\mathfrak{G}(\mathbb{R}_+)$ is stable by the max operation: if f_1 and f_2 are in $\mathfrak{G}(\mathbb{R}_+)$, so is $\max(f_1, f_2)$.

2.3. Divergence functions

In dealing with the φ-divergence function attached with the kernel function $\varphi \in \Gamma(\mathbb{R}_+)$, we have to make precise the meaning of $q_i\varphi(\frac{p_i}{q_i})$ when q_i equals 0. This is done via the following key-result from convex analysis: the so-called *closed perspective function* of φ defined on $\mathbb{R}^2$ as

$$(x, y) \in \mathbb{R} \times \mathbb{R} \mapsto \widetilde{\varphi}(x, y) = \begin{cases} y\varphi(\frac{x}{y}) & \text{if } y > 0, \\ \varphi'_\infty x & \text{if } x > 0 \text{ and } y = 0, \\ 0 & \text{if } x = 0 \text{ and } y = 0, \\ +\infty & \text{if } x < 0 \text{ and } y = 0, \text{ or } y < 0, \end{cases} \tag{2}$$

is a *closed convex function* [1, vol. II, pp. 40–41], [2, p. 67]. So, $0\varphi(\frac{0}{0})$ is interpreted as 0, while $0\varphi(\frac{x}{0})$ is $\varphi'_\infty x$ for $x > 0$. The function $\widetilde{\varphi}$ turns out to be convex as a function of both x and y, it is actually the support function of a closed convex set in $\mathbb{R}^2$, a property which will be instrumental in deriving the main result in Section 3.

To have the f-divergence function I_f meaningful, it is necessary to have $I_f(p, q) \geqslant 0$ for all $(p, q) \in \mathbb{R}^n_+ \times \mathbb{R}^n_+$; this is indeed achieved when $f \in \mathfrak{G}(\mathbb{R}_+)$ is normalized.

Proposition 1. *Let $f \in \mathfrak{G}(\mathbb{R}_+)$ be normalized; then $f \geqslant 0$ on $\mathbb{R}_+$, so that $I_f \geqslant 0$ on $\mathbb{R}^n_+ \times \mathbb{R}^n_+$.*

Proof. If dom $f = \{1\}$, the only possibility for f is to be the indicator function of $\{1\}$, which is positive.

The other possibility is to have 1 lying in int(dom f) (cf. Section 2.2). Then consider a subderivative s of f at 1 [1, vol. I, p. 22]; we have

$$f(x) \geqslant f(1) + s(x-1) = s(x-1) \quad \text{for all } x \geqslant 0.$$

So,

$$xf\left(\frac{1}{x}\right) \geqslant xs\left(\frac{1}{x} - 1\right) = s(1-x) \quad \text{for all } x > 0.$$

In sum, since $f(x) = xf(\frac{1}{x})$ for all $x > 0$,

$$f(x) \geqslant \left|s(x-1)\right| \quad \text{for all } x > 0, \tag{3}$$

an inequality which extends to $\mathbb{R}_+$ (because $f \in \Gamma(\mathbb{R}_+)$). Thus, $f \geqslant 0$ on $\mathbb{R}_+$ (and therefore on $\mathbb{R}$).

As an immediate consequence,

$$\widetilde{f}(x, y) = x\frac{y}{x}f\left(\frac{1}{y/x}\right) \geqslant 0 \quad \text{whenever } x > 0 \text{ and } y > 0,$$

an inequality which extends to the whole of $\mathbb{R}^2$ (cf. the definitions in (2)). Whence the positivity of I_f on $\mathbb{R}^n_+ \times \mathbb{R}^n_+$ is proved. □

There are shorter proofs of the positivity of f under the normalization assumption; note however that (3) provides an *estimate* of “how positive f is” in terms of the length of subderivatives of f at the point 1.

Note that the positivity of $I_\varphi(p, q)$ was proved in [3, Corollary 3.1] under the sole assumption that the normalized φ is in $\Gamma(\mathbb{R}_+)$, but only for those $p, q \in \mathbb{R}^n_+$ satisfying $\sum_{i=1}^n p_i = \sum_{i=1}^n q_i$.

2.4. Examples

Here we list some examples of $f \in \mathfrak{G}(\mathbb{R}_+)$; in each case we note if the considered f is defined as the symmetrized form of some $\varphi \in \Gamma(\mathbb{R}_+)$, how to normalize it, and what divergence function it gives rise to (when the kernel function is named after someone, we recall it, borrowing from [3] for that).

Example 2.1. For $a \in \mathbb{R}$, let $f : x \geqslant 0 \mapsto f(x) := a(x+1)$. This is one of the simplest functions in $\mathfrak{G}(\mathbb{R}_+)$. Indeed if $f \in \mathfrak{G}(\mathbb{R}_+)$, then

$$x > 0 \mapsto f(x) - \frac{f(1)}{2}(x+1) \tag{4}$$

is again in $\mathfrak{G}(\mathbb{R}_+)$, but normalized now.

Let $\mathfrak{G}_0(\mathbb{R}_+)$ denote the set of $f \in \mathfrak{G}(\mathbb{R}_+)$ which are normalized. Then, as for $\mathfrak{G}(\mathbb{R}_+)$, $\mathfrak{G}_0(\mathbb{R}_+)$ is a closed convex cone and, according to what has just been explained,

$$\mathfrak{G}(\mathbb{R}_+) = \mathfrak{G}_0(\mathbb{R}_+) + \mathbb{R} f_1, \tag{5}$$

where f_1 stands for the basic function $x > 0 \mapsto x + 1$.

Example 2.2 *(Kullback–Leibler).* Let $\varphi \in \Gamma(\mathbb{R}_+)$ be defined as $x > 0 \mapsto \varphi(x) := -\log(x)$. The symmetrized form of φ is

$$x > 0 \mapsto \frac{1}{2}\big[x \log(x) - \log(x)\big],$$

which turns out to be normalized.

The corresponding f-divergence is

$$(p, q) \in \mathbb{R}_+^n \times \mathbb{R}_+^n \mapsto I_f(p, q) = \frac{1}{2}\sum_{i=1}^{n}\left[p_i \log\left(\frac{p_i}{q_i}\right) + q_i \log\left(\frac{q_i}{p_i}\right)\right].$$

Example 2.3 *(Hellinger).* Let $f \in \mathfrak{G}(\mathbb{R}_+)$ be defined as: $f : x \geqslant 0 \mapsto f(x) := (1 - \sqrt{x})^2$. This function is normalized, and the corresponding f-divergence is

$$I_f(p, q) = \sum_{i=1}^{n}(\sqrt{p_i} - \sqrt{q_i})^2.$$

Example 2.4 *(Renyi).* For $\alpha > 1$, let $\varphi \in \Gamma(\mathbb{R}_+)$ be defined as $\varphi(x) := x^\alpha$. Then the symmetrized and normalized form of φ is

$$f : x > 0 \mapsto \frac{1}{2}\big[x^\alpha + x^{1-\alpha} - (x+1)\big].$$

The corresponding f-divergence is

$$I_f(p, q) = \frac{1}{2}\sum_{i=1}^{n}\big[p_i^\alpha q_i^{1-\alpha} + p_i^{1-\alpha} q_i^\alpha - (p_i + q_i)\big].$$

Example 2.5 *(The χ^2-kernel).* Let $\varphi \in \Gamma(\mathbb{R}_+)$ be defined by $\varphi(x) := (x-1)^2$. The symmetrized and normalized form of φ is

$$f : x > 0 \mapsto f(x) = \frac{1}{2}\left(x^2 - x + \frac{1}{x} - 1\right).$$

The associated f-divergence is

$$\begin{aligned} I_f(p,q) &= \frac{1}{2}\sum_{i=1}^{n}\left(\frac{p_i^2}{q_i} + \frac{q_i^2}{p_i} - p_i - q_i\right) \\ &= \frac{1}{2}\sum_{i=1}^{n}\left(\frac{(p_i - q_i)^2}{p_i} + \frac{(q_i - p_i)^2}{q_i}\right) \end{aligned}$$

(the so-called *symmetrized χ^2-distance* between p and q).

Example 2.6. For $a > 0$, let $f : x \geqslant 0 \mapsto f(x) := a|x-1|$. This is one of the most used kernel functions. Clearly, $f \in \mathfrak{G}_0(\mathbb{R}_+)$, and the attached f-divergence is

$$I_f(p,q) = a\sum_{i=1}^{n}|p_i - q_i|$$

(called the *variation distance* between p and q).

Example 2.7. Let $f \in \mathfrak{G}(\mathbb{R}_+)$ be defined by $f(x) := \sqrt{x^2+1}$. Its normalized version is

$$f_0 : x > 0 \mapsto f_0(x) = \sqrt{x^2+1} - \frac{\sqrt{2}}{2}(x+1).$$

Whence we get the associated f_0-divergence:

$$I_{f_0}(p,q) = \sum_{i=1}^{n}\left[\sqrt{p_i^2 + q_i^2} - \frac{\sqrt{2}}{2}(p_i + q_i)\right].$$

Example 2.8. Let $f \in \mathfrak{G}(\mathbb{R}_+)$ be defined by $f(x) := -\sqrt{x}$. When normalized, this kernel function gives rise to

$$f_0 : x > 0 \mapsto f_0(x) = -\sqrt{x} + \frac{1}{2}(x+1).$$

The corresponding f_0-divergence has the following expression:

$$I_{f_0}(p,q) = \sum_{i=1}^{n}\left[\frac{1}{2}(p_i + q_i) - \sqrt{p_i q_i}\right]$$

(summation of the gaps between the arithmetical means and the geometrical means of the (p_i, q_i)).

Example 2.9. Let $f \in \mathfrak{G}(\mathbb{R}_+)$ be defined as the following:

$$f(x) := \frac{1}{x} \quad \text{if } x \leqslant 1, \qquad x^2 \quad \text{if } x > 1.$$

After normalization we get at

$$f_0(x) := \frac{1}{x} - \frac{1}{2}(x+1) \quad \text{if } x \leqslant 1, \quad x^2 - \frac{1}{2}(x+1) \quad \text{if } x > 1.$$

This gives rise to the rather original f_0-divergence below:

$$I_{f_0}(p,q) = \sum_{i=1}^{n} \left\{ \frac{[\max(p_i, q_i)]^2}{\min(p_i, q_i)} - \frac{1}{2}(p_i + q_i) \right\}.$$

Example 2.10. We end our list with an example of kernel function taking the value $+\infty$ out of some interval of $(0, +\infty)$. For $0 < r \leqslant 1$, let $f_r \in \mathfrak{G}_0(\mathbb{R}_+)$ be defined as the indicator of the line-segment $[r, 1/r]$, that is:

$$f_r(x) := 0 \quad \text{if } r \leqslant x \leqslant \frac{1}{r}, \quad +\infty \quad \text{otherwise.}$$

Then,

$$I_{f_r}(p,q) = \begin{cases} 0 & \text{if all the } p_i \text{ and } q_i \text{ are strictly positive and } r \leqslant \frac{p_i}{q_i} \leqslant \frac{1}{r} \text{ for all } i; \\ +\infty & \text{otherwise.} \end{cases}$$

3. The main result

As we began explaining in Section 2.3, with any $\varphi \in \Gamma(\mathbb{R}_+)$, it is possible to associate its closed perspective function $\widetilde{\varphi}$ (cf. (2)). This function is closed convex and positively homogeneous (jointly in the real variables x and y), it is therefore the support function of a closed convex set of $\mathbb{R}^2$: we call such a set the *generator* of φ, and denote it by C_φ. Since $\widetilde{\varphi}(x,y) = +\infty$ for $x < 0$ or $y < 0$, the asymptotic (or recession) cone of C_φ contains the negative orthant $\mathbb{R}^2_-$. An asymptotic cone of C_φ strictly larger than $\mathbb{R}^2_-$ corresponds exactly to the case where φ takes the value $+\infty$ on some $x > 0$ (this asymptotic cone cannot be the whole of $\mathbb{R}^2$ since we have assumed by definition that φ is finite at some $x_0 > 0$ (cf. beginning of Section 2.1). A little bit simpler but equivalent way of defining C_φ is:

$$C_\varphi = \left\{ (p,q) \in \mathbb{R}^2 \;\middle|\; px + qy \leqslant y\varphi\left(\frac{x}{y}\right) \text{ for all } x > 0 \text{ and } y > 0 \right\} \tag{6}$$

(the support function $\widetilde{\varphi}$ of C_φ is just the closure of the function which takes the value $y\varphi(\frac{x}{y})$ whenever $x > 0$ and $y > 0$, and $+\infty$ elsewhere).

The next result characterizes those f which are in $\mathfrak{G}(\mathbb{R}_+)$ in terms of a very simple geometrical involution: the symmetry with respect to a line. We alluded to this result in our earlier work on the Legendre–Fenchel conjugate of the reciprocal (or inverse) of a function [4, p. 548].

Theorem 2. *$f \in \mathfrak{G}(\mathbb{R}_+)$ if and only if its generator C_f is symmetric with respect to the first bisecting line in the plane $\mathbb{R}^2$ (of equation $y = x$).*

Proof. We have C_f defined as in (6). The keypoint of the proof is the following: To have $f \in \mathfrak{G}(\mathbb{R}_+)$, that is, satisfying

$$f(u) = uf\left(\frac{1}{u}\right) \quad \text{for all } u > 0,$$

is equivalent to having

$$yf\left(\frac{x}{y}\right) = xf\left(\frac{y}{x}\right) \quad \text{for all } x > 0 \text{ and } y > 0. \tag{7}$$

The geometrical interpretation of (7), via (6), is: C_f is invariant by the transformation $(x, y) \mapsto (y, x)$, i.e., C_f is symmetric with respect to the first bisecting line in the plane $\mathbb{R}^2$. □

Let $\mathcal{C}$ denote the collection of closed convex sets in $\mathbb{R}^2$, whose recession cones contain $\mathbb{R}^2_-$ (but different from $\mathbb{R}^2$), and symmetric with respect to the first bisecting line in $\mathbb{R}^2$. The back and forth relation between the $f \in \mathfrak{G}(\mathbb{R}_+)$ and the C in $\mathcal{C}$ is summarized below:

$$\left.\begin{array}{l} \bullet\ f \in \mathfrak{G}(\mathbb{R}_+) \rightsquigarrow C_f \in \mathcal{C} \text{ whose support function is } \widetilde{f} \text{ (cf. (6))} \\ \bullet\ C \in \mathcal{C} \rightsquigarrow f_C : x > 0 \mapsto f_C(x) := \sigma_C(x, 1); \end{array}\right\} \tag{8}$$

above σ_C designates the support function of C.

The second part of this correspondence is actually the process we used for designing the functions f of $\mathfrak{G}(\mathbb{R}_+)$ proposed in Examples 2.7–2.9, etc. Indeed, the geometrical properties of C in $\mathcal{C}$ are translated into functional properties on f_C: for example, a polyhedral C will give rise to a polyhedral f_C, the behavior of the boundary curve of C (when $x \to -\infty$ or $y \to -\infty$ for a boundary point (x, y) of C) yields all the information on the behavior of f_C in the neighborhood of 0 or $+\infty$. In short: given prescribed properties of the aimed function $f \in \mathfrak{G}(\mathbb{R}_+)$, it is possible to devise $C \in \mathcal{C}$ such that $f_C = f$.

Note that, in terms of the generator C_f, normalizing f amounts to translating C_f in an (a, a) vector direction. When f is normalized, the boundary curve of C_f passes through $(0, 0)$.

When considering the problem we have tackled, a natural question in the context of convex analysis we would raise is: what if the Legendre–Fenchel transformation enters into the picture? In fact, for $\varphi \in \Gamma(\mathbb{R}_+)$, the Legendre–Fenchel conjugate φ^* of φ is the support function of $\{(x, -y) \mid (x, y) \in \mathit{epi}\, f^*\}$ ([1, vol. II, Proposition 1.2.1], [2, Corollary 13.5.1]). Thus: $f \in \mathfrak{G}(\mathbb{R}_+)$ if and only if $\mathit{epi}\, f^*$ is symmetric with respect to the second bisecting line in the plane $\mathbb{R}^2$ (of equation $y = -x$). So, to come back to our $C \in \mathcal{C}$, the generator C_f of $f \in \mathfrak{G}(\mathbb{R}_+)$ is nothing else than the copy of $\mathit{epi}\, f^*$ by the symmetry $(x, y) \mapsto (x, -y)$ in $\mathbb{R}^2$.

We end by inviting the reader to put the scheme (8) into practice by drawing the generator C_f of the f in Examples 2.2–2.10, and by devising new $f \in \mathfrak{G}(\mathbb{R}_+)$ from $C \in \mathcal{C}$. Below we present some instances.

From f to C_f. The generator C_f of $f \in \mathfrak{G}(\mathbb{R}_+)$ can also be viewed as the subdifferential of the associated perspective function $\widetilde{f}$ at $(0, 0)$. Consider for the sake of simplicity the case where f is finite on $(0, +\infty)$, so that the closure of $\operatorname{dom} \widetilde{f}$ is $\mathbb{R}^2_+$ and the asymptotic (or recession) cone of $C_f = \partial \widetilde{f}(0, 0)$ is $\mathbb{R}^2_-$. According to [2, Theorem 2.5.6] or [1, vol. I, Section 6.3]:

$$C_f = \overline{co}\,\vec{\nabla} \widetilde{f}(0, 0) + \mathbb{R}^2_-, \tag{9}$$

where $\vec{\nabla} \widetilde{f}(0, 0)$ is the set of all limits of sequences $\{\nabla \widetilde{f}(x_n, y_n)\}$ such that $\widetilde{f}$ is differentiable at (x_n, y_n) and (x_n, y_n) tends to $(0, 0)$ as $n \to +\infty$. Here, the calculation in (9) is made easier due to the fact that $\widetilde{f}$ is positively homogeneous, whence its gradient is constant along the half-rays directed by (x, y), $x > 0$ and $y > 0$. Indeed we have

$$\nabla \widetilde{f}(x, y) = \left(f'\left(\frac{x}{y}\right), f\left(\frac{x}{y}\right) - \frac{x}{y} f'\left(\frac{x}{y}\right)\right)$$

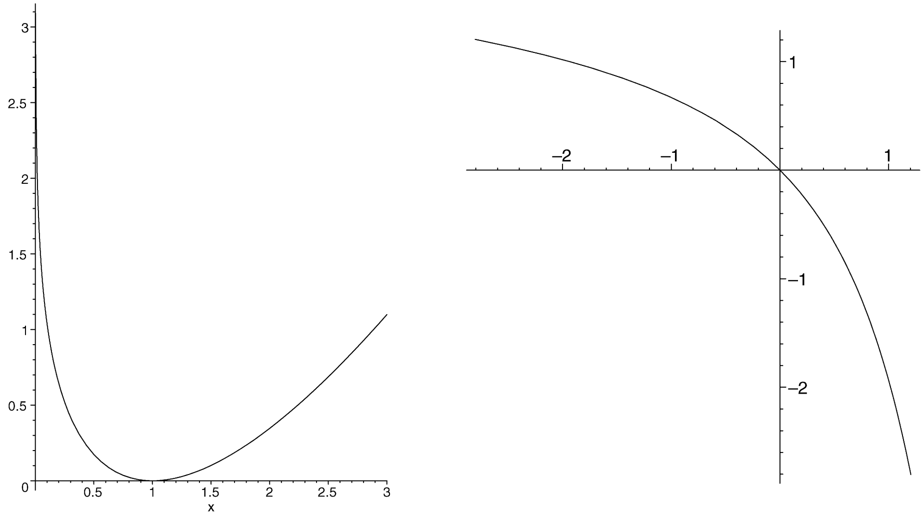

Fig. 1.

whenever f is differentiable at y/x, so that a parameterization of the (essential part of the) boundary curve of C_f is provided by

$$\big(f'(s), f(s) - sf'(s)\big), \quad s > 0. \tag{10}$$

In the smooth case (i.e., when f is differentiable on $(0, +\infty)$), the closed convex hull operation is unnecessary in (9). In the nonsmooth case, we have to connect with line segments the pieces of curves obtained in $\vec{\nabla}\widetilde{f}(0,0)$.

Example 3.1. (*Example 2.2 revisited.*) According to the parameterization process described above, the boundary curve of the generator C_f attached with the Kullback–Leibler function f is

$$\left(\frac{1}{2}\left(1 + \log(s) - \frac{1}{s}\right), \frac{1}{2}\big(-\log(s) - s + 1\big)\right), \quad s > 0. \tag{11}$$

See f and C_f in Fig. 1.

Example 3.2. (*Example 2.9 revisited.*) The essential part of the boundary curve of C_f is provided by two pieces of curves:

$$\begin{aligned}
&\left(-\frac{1}{s^2} - \frac{1}{2}, -\frac{2}{s} - \frac{1}{2}\right), \quad 0 < s \leqslant 1;\\
&\left(2s - \frac{1}{2}, -s^2 - \frac{1}{2}\right), \quad s > 1.
\end{aligned} \tag{12}$$

We then connect them with a line segment, so that to obtain $\vec{\nabla}\widetilde{f}(0,0)$. See f and C_f in Fig. 2.

Example 3.3. (*Example 2.7 revisited.*) We have here:

$$\nabla\widetilde{f}(0,0) = \left\{\left(\frac{s}{\sqrt{s^2+1}} - \frac{\sqrt{2}}{2}, \frac{1}{\sqrt{s^2+1}} - \frac{\sqrt{2}}{2}\right) \;\middle|\; s \geqslant 0\right\}, \tag{13}$$

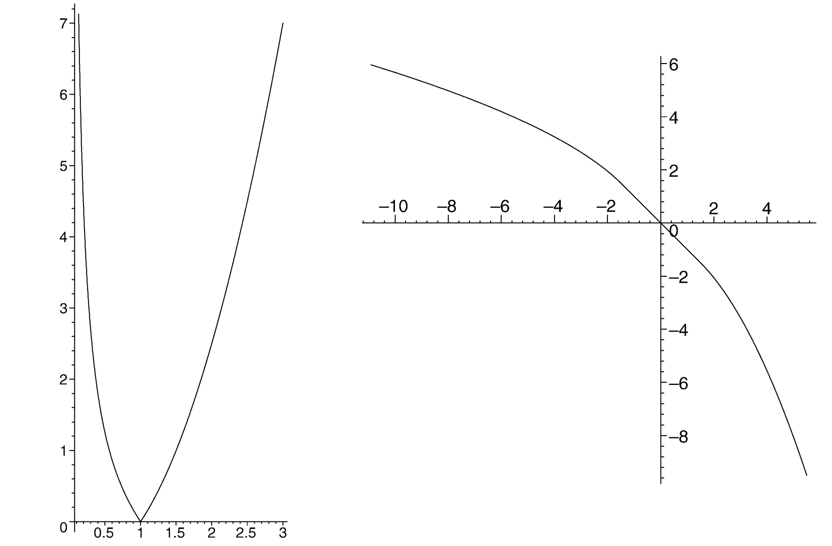

Fig. 2.

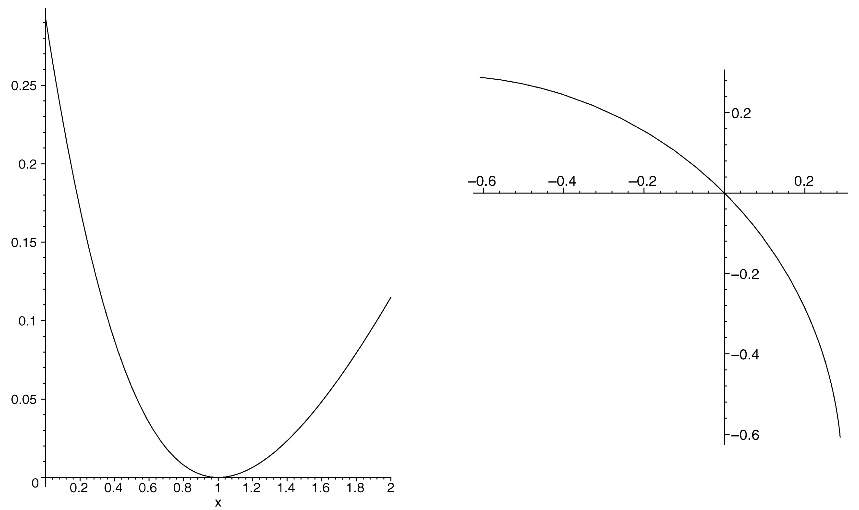

Fig. 3.

a quarter of the unit circle centered at $(-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2})$. We have to add two half-lines (the boundary of $\mathbb{R}^2_-$) to get the whole boundary curve of C_f. See f, $\overrightarrow{\nabla}\widetilde{f}(0,0)$ and C_f in Fig. 3.

From C to f_C. The value at $x > 0$ of the function $f_C \in \mathfrak{G}(\mathbb{R}_+)$ associated with $C \in \mathcal{C}$ is obtained by maximizing a linear form on C, namely:

$$f_C(x) = \sup\{px + q \mid (p,q) \in C\}. \tag{14}$$

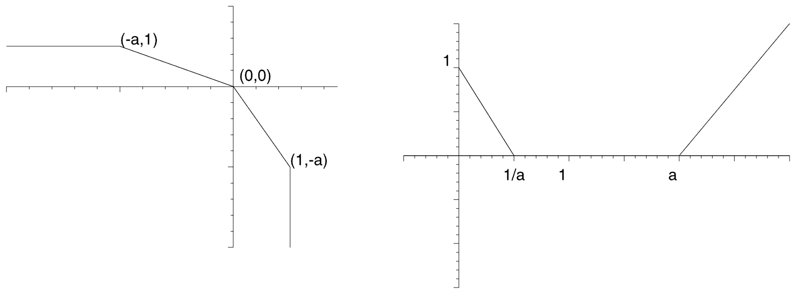

Fig. 4.

Example 3.4. For $a \geqslant 1$, let $C \in \mathcal{C}$ be drawn on the left part of Fig. 4. Then the corresponding $f_C \in \mathfrak{G}(\mathbb{R}_+)$ is defined as following:

$$f_C(x) = \begin{cases} 1 - ax & \text{if } 0 < x \leqslant 1/a, \\ 0 & \text{if } 1/a < x \leqslant a, \\ x - a & \text{if } x > a. \end{cases}$$

We end by giving an analytical characterization of convex functions solving the functional equation (1). The following result deals with the simpler case of nonnegative functions satisfying the normalization condition $f(1) = 0$.

Proposition 3. *Let $f : [1, +\infty) \to \mathbb{R} \cup \{+\infty\}$ be a convex function satisfying $f(1) = 0$. Extend it to $(0, 1]$ by setting $f(x) = xf(\frac{1}{x})$. Then the resulting function is convex (and satisfies the functional equation (1)) if and only if f is nonnegative.*

Proof. The "only if" statement follows from Proposition 1. Conversely, if f is nonnegative then, as its extension is also convex on $(0, 1]$, to deduce convexity on the whole of $(0, +\infty)$ it suffices to observe that 1 is a (global) minimum. □

Theorem 4. *Let $f : [1, +\infty) \to \mathbb{R} \cup \{+\infty\}$ be convex, and continuous at 1 (from the right). Extend it to $(0, 1)$ by setting $f(x) = xf(\frac{1}{x})$. Then the resulting function is convex (and satisfies the functional equation (1)) if and only if*

$$f(1) \leqslant 2f'_+(1).$$

Proof. By Proposition 3, the extension is convex if and only if its normalized form

$$x > 0 \mapsto g(x) = f(x) - \frac{f(1)}{2}(x + 1)$$

is nonnegative, which is in turn equivalent to the nonnegativity of $g'_+(1) = f'_+(1) - \frac{f(1)}{2}$. □

The preceding characterization covers all convex solutions of (1) except the trivial ones (i.e., those consisting of the sum of the indicator function of the singleton $\{1\}$ with a constant).

Proposition 3 and Theorem 4 admit obvious counterparts, which we omit, starting from a convex function $f : (0, 1] \to \mathbb{R} \cup \{+\infty\}$.

References

[1] J.-B. Hiriart-Urruty, C. Lemarechal, Convex Analysis and Minimizations Algorithms I and II, Grundlehren Math. Wiss., vols. 305 and 306, Springer, 1993. Second corrected printing, 1996.
[2] R.T. Rockafellar, Convex Analysis, second ed., Princeton Univ. Press, 1972.
[3] A. Ben-Tal, A. Ben-Israel, M. Teboulle, Certainty equivalents and information measures: Duality and extremal principles, J. Math. Anal. Appl. 157 (1991) 211–236.
[4] J.-B. Hiriart-Urruty, J.-E. Martínez-Legaz, New formulas for the Legendre–Fenchel transform, J. Math. Anal. Appl. 288 (2003) 544–555.