# Perspective Functions: Properties, Constructions, and Examples*


Patrick L. Combettes

North Carolina State University
Department of Mathematics
Raleigh, NC 27695-8205, USA
plc@math.ncsu.edu



**Abstract**

Many functions encountered in applied mathematics and in statistical data analysis can be expressed in terms of perspective functions. One of the earliest examples is the Fisher information, which appeared in statistics in the 1920s. We analyze various algebraic and convex-analytical properties of perspective functions and provide general schemes to construct lower semicontinuous convex functions from them. Several new examples are presented and existing instances are featured as special cases.


## 1 Introduction

Let $\mathcal{G}$ be a real Hilbert space and let $\varphi\colon \mathcal{G} \to \left]-\infty,+\infty\right]$ be a convex function. The perspective function of $\varphi$ is (see Figure 1)

$$\mathscr{P}_\varphi\colon \mathbb{R}\times\mathcal{G} \to \left]-\infty,+\infty\right] : (\eta,y) \mapsto \begin{cases} \eta\varphi(y/\eta), & \text{if } \eta>0;\\ +\infty, & \text{otherwise.}\end{cases} \tag{1.1}$$

The properties of $\mathscr{P}_\varphi$ were first investigated in [57], where it was shown in particular that $\mathscr{P}_\varphi$ is convex if and only if $\varphi$ is convex (see also [6, 27, 34]). The term "perspective function" was coined by Claude Lemaréchal ca. 1987-1988 [43] and first appeared in print in [34, Section IV.2.2]. Special cases of the construction (1.1) arise in various areas of applied mathematics and data analysis. One of the oldest instances involving perspective functions is the Fisher information of a differentiable probability density $x\colon \mathbb{R}^N \to \left]0,+\infty\right[$, that is,

$$\int_{\mathbb{R}^N} \frac{\|\nabla x(t)\|_2^2}{x(t)}dt, \tag{1.2}$$


*Contact author: P. L. Combettes, plc@math.ncsu.edu, phone: +1 (919) 515 2671.

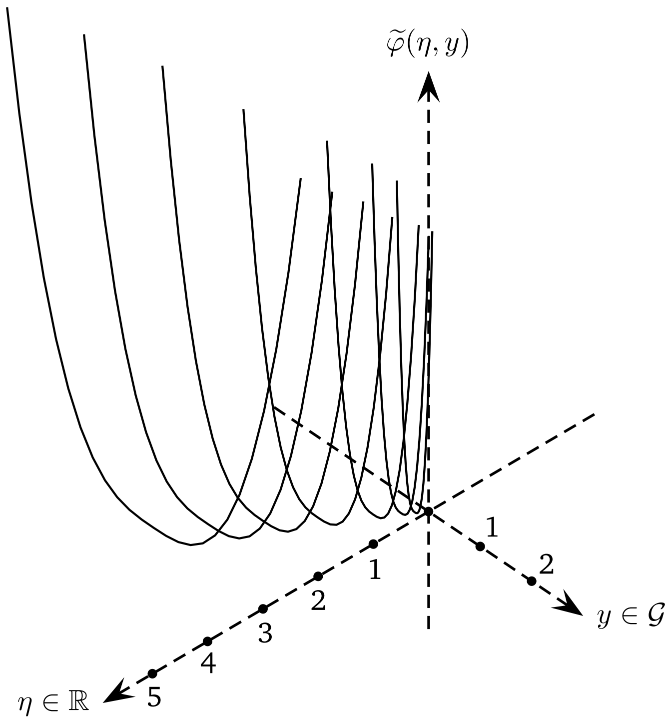


Figure 1: Slices of the graph of the perspective function of $\varphi\colon y \mapsto 1/2 + 8\|y\|^3$ for fixed values of $\eta \in \{1/4, 1/2, 1, 2, 3, 4, 5\}$; the value $\eta = 1$ provides the graph of $\varphi$.

where $\|\cdot\|_2$ is the standard Euclidean norm on $\mathbb{R}^N$. This notion, which dates back to the work of Fisher in statistics [29], has found applications in many contexts, e.g., [9, 13, 14, 31, 52, 58, 60]. More generally, (1.1) can be used to construct convex integrands of integral functionals such as

$$\int_{\mathbb{R}^N} \mathscr{P}_\varphi\big(x(t), y(t)\big)\, dt = \int_{\mathbb{R}^N} x(t)\varphi\left(\frac{y(t)}{x(t)}\right) dt, \tag{1.3}$$

where $x\colon \mathbb{R}^N \to \,]0,+\infty[$ and $y\colon \mathbb{R}^N \to \mathcal{G}$. In the case when $N = 1$ and $\mathcal{G} = \mathbb{R}$, it corresponds to a notion of $\varphi$-divergence which originates in [2, 26] and that has been used extensively in information theory, statistics, signal processing, and pattern recognition [4, 10, 44, 55]; see also [7, 35] for a discussion of discrete counterparts. In the case when $\mathcal{G} = \mathbb{R}^{N\times N}$, $y = \nabla x$, and $\varphi = \|\cdot\|_2^2$, one recovers (1.2). Furthermore, choosing $\varphi = \|\cdot\|_2^p$ with $p \in \,]1,+\infty[$ provides the extension of the Fisher information (1.2) found in [12] in the case when $N = 1$. Instances of perspective functions can also be identified in robust estimation [37, Section 7.7] (see also [49, 53] for recent developments), transportation theory [8, 18, 30, 54], sparse regression [11, 25, 42], control theory [38, 47], mixed-integer programming [33], computer vision [61], disjunctive programming [20], game theory [1], machine learning [46], and mean-field games [19].

Although perspective functions appear explicitly or implicitly in an increasing number of diverse research areas, little effort has been dedicated to the systematic study of their properties, especially in general Hilbert spaces. It is the goal of the present paper to propose such an investigation, with a special focus on the construction of lower semicontinuous convex functions around perspective functions. As is well known, these two properties are of paramount importance in the modeling, analysis, and numerical solution of variational problems. Section 2 focuses on algebraic and convex-analytical properties. On the basis of these results, several examples of lower semicontinuous convex perspective functions are provided in Section 3. Finally, integral functions with perspective function-based integrands are studied in Section 4. Many of the functions we propose are new and suggest new problem formulations in various applications areas. In particular, our results are exploited in the companion paper [25], which investigates the proximity operator of perspective functions and explores new models and algorithms in high-dimensional statistics.

**Notation.** Throughout, $\mathcal{H}$ and $\mathcal{G}$ are real Hilbert spaces and $\mathcal{H} \oplus \mathcal{G}$ denotes their Hilbert direct sum. The closed ball with center $x \in \mathcal{H}$ and radius $\rho \in \left]0,+\infty\right[$ in $\mathcal{H}$ is denoted by $B(x;\rho)$. $\Gamma_0(\mathcal{H})$ is the class of lower semicontinuous convex functions $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ such that $\operatorname{dom} f = \big\{x \in \mathcal{H} \;\big|\; f(x) < +\infty\big\} \neq \varnothing$. Let $f \in \Gamma_0(\mathcal{H})$. Then $f^*$ denotes the conjugate of $f$, $\operatorname{epi} f$ the epigraph of $f$, $\operatorname{rec} f$ the recession function of $f$, and $\partial f$ the subdifferential of $f$. Let $C$ be a subset of $\mathcal{H}$. Then $\iota_C$ is the indicator function of $C$, $d_C$ the distance function to $C$, $\operatorname{rec} C$ the recession cone of $C$, and $\sigma_C$ the support function of $C$. See [6, 41] for background on hilbertian convex analysis and [34, 57] for the Euclidean setting.

## 2 Properties of perspective functions

In this section we study various properties of perspective functions. We start our discussion by noting that, if $\varphi \in \Gamma_0(\mathcal{G})$, the construction (1.1) does not necessarily produce a lower semicontinuous function. For this reason, we shall use the following variant, first proposed in [57] for $\mathcal{G} = \mathbb{R}^N$.

**Definition 2.1** Let $\varphi \in \Gamma_0(\mathcal{G})$ and let $\operatorname{rec}\varphi$ be its recession function, i.e., given any $z \in \operatorname{dom}\varphi$,

$$(\forall y \in \mathcal{G}) \quad (\operatorname{rec}\varphi)(y) = \sup_{x \in \operatorname{dom}\varphi} \big(\varphi(x+y) - \varphi(x)\big) = \lim_{\alpha \to +\infty} \frac{\varphi(z+\alpha y)}{\alpha}. \tag{2.1}$$

The lower semicontinuous envelope of the perspective of $\varphi$ is

$$\widetilde{\varphi}\colon \mathbb{R} \times \mathcal{G} \to \left]-\infty,+\infty\right] : (\eta, y) \mapsto \begin{cases} \eta\varphi(y/\eta), & \text{if } \eta > 0;\\ (\operatorname{rec}\varphi)(y), & \text{if } \eta = 0;\\ +\infty, & \text{otherwise.} \end{cases} \tag{2.2}$$

For simplicity, $\widetilde{\varphi}$ is called the *perspective* of $\varphi$.

**Lemma 2.2** *Let $\varphi \in \Gamma_0(\mathcal{G})$. Then the following hold:*

(i) $\operatorname{rec}\operatorname{epi}\varphi = \operatorname{epi}\operatorname{rec}\varphi$ [41, Proposition 6.8.3].

(ii) $\operatorname{rec}\varphi = \sigma_{\operatorname{dom}\varphi^*}$ [41, Théorème 6.8.5].

The following result records basic topological and convex analytical properties of the perspective function (2.2).

**Proposition 2.3** *Let $\varphi \in \Gamma_0(\mathcal{G})$. Then the following hold:*

(i) $\widetilde{\varphi}$ *is positively homogeneous.*

(ii) $\widetilde{\varphi} \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$.

(iii) $\widetilde{\varphi}$ *is sublinear.*

(iv) *Let* $C = \big\{(\mu, u) \in \mathbb{R} \times \mathcal{G} \;\big|\; \mu + \varphi^*(u) \leqslant 0\big\}$. *Then* $(\widetilde{\varphi})^* = \iota_C$ *and* $\widetilde{\varphi} = \sigma_C$.

(v) *Let* $\eta \in \mathbb{R}$ *and* $y \in \mathcal{G}$. *Then*

$$\partial\widetilde{\varphi}(\eta, y) = \begin{cases} \big\{\big(\varphi(y/\eta) - \langle y \mid u\rangle/\eta, u\big) \;\big|\; u \in \partial\varphi(y/\eta)\big\}, & \text{if } \eta > 0;\\ \big\{(\mu, u) \in C \;\big|\; \sigma_{\operatorname{dom}\varphi^*}(y) = \langle y \mid u\rangle\big\}, & \text{if } \eta = 0 \text{ and } y \neq 0;\\ C, & \text{if } \eta = 0 \text{ and } y = 0;\\ \varnothing, & \text{if } \eta < 0. \end{cases} \tag{2.3}$$

*Proof*. (i): This follows from (2.1) and (2.2).

(ii): Set $D = \{1\} \times \operatorname{epi}\varphi$ and $g = \mathscr{P}_\varphi$, and let $z \in \operatorname{dom}\varphi$. Then $(1, z) \in \operatorname{dom} g$. On the other hand, since $D$ is convex, $\operatorname{epi} g = \operatorname{cone} D$ is convex and $g$ is therefore a proper convex function. Let us denote by $\breve{g}$ the largest lower semicontinuous convex function majorized by $g$. To show that $\widetilde{\varphi} \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$, it is enough to show that

$$\widetilde{\varphi} = \breve{g}. \tag{2.4}$$

This can be done using the following argument due to H. H. Bauschke. Since $(0, 0) \notin D$, it follows from [6, Theorem 9.9 and Corollary 6.52], Lemma 2.2(i), and [6, Lemma 1.6(ii)] that $\operatorname{epi}\breve{g} = \overline{\operatorname{epi} g} = \overline{\operatorname{cone}} D = (\operatorname{cone} D) \cup (\operatorname{rec} D) = (\operatorname{epi} g) \cup (\{0\} \times \operatorname{rec}\operatorname{epi}\varphi) = (\operatorname{epi} g) \cup (\{0\} \times \operatorname{epi}\operatorname{rec}\varphi) = (\operatorname{epi} g) \cup \operatorname{epi}(\iota_{\{0\}} \oplus \operatorname{rec}\varphi) = \operatorname{epi}\min\{g, \iota_{\{0\}} \oplus \operatorname{rec}\varphi\} = \operatorname{epi}\widetilde{\varphi}$.

(iii): This follows from (i) and (ii).

(iv): Set $g = \mathscr{P}_\varphi$. Then $g^* = \iota_C$ [6, Example 13.8]. Hence, we derive from (2.4) and [6, Proposition 13.14] that $(\widetilde{\varphi})^* = (\breve{g})^* = g^* = \iota_C$. In turn, (ii) and [6, Corollary 13.33] yield $\widetilde{\varphi} = (\widetilde{\varphi})^{**} = \iota_C^* = \sigma_C$.

(v): Let $\mu \in \mathbb{R}$ and $u \in \mathcal{G}$. It follows from the Fenchel-Young identity [6, Proposition 16.13] and (iv) that

$$\begin{aligned} (\mu, u) \in \partial\widetilde{\varphi}(\eta, y) \quad &\Leftrightarrow \quad \widetilde{\varphi}(\eta, y) + (\widetilde{\varphi})^*(\mu, u) = \eta\mu + \langle y \mid u\rangle \\ &\Leftrightarrow \quad \widetilde{\varphi}(\eta, y) = \eta\mu + \langle y \mid u\rangle \;\text{ and }\; \mu + \varphi^*(u) \leqslant 0. \end{aligned} \tag{2.5}$$

We consider three cases.

- $\eta < 0$: Then (2.2) and (2.5) yield $\partial\widetilde{\varphi}(\eta, y) = \varnothing$.
- $\eta = 0$: We deduce from (2.5), (2.2), and Lemma 2.2(ii) that

$$\begin{aligned} (\mu, u) \in \partial\widetilde{\varphi}(\eta, y) \quad &\Leftrightarrow \quad (\operatorname{rec}\varphi)(y) = \langle y \mid u\rangle \;\text{ and }\; \mu + \varphi^*(u) \leqslant 0 \\ &\Leftrightarrow \quad \sigma_{\operatorname{dom}\varphi^*}(y) = \langle y \mid u\rangle \;\text{ and }\; (\mu, u) \in C. \end{aligned} \tag{2.6}$$

Since $\sigma_{\operatorname{dom}\varphi^*}(0) = 0 = \langle 0 \mid u\rangle$, we obtain the desired results.

- $\eta > 0$: Using successively (2.5), (2.2), the Fenchel-Young inequality [6, Proposition 13.13], and the Fenchel-Young identity, we obtain

$$\begin{aligned}(\mu, u) \in \partial\widetilde{\varphi}(\eta, y) \quad &\Leftrightarrow \quad \mu = \varphi(y/\eta) - \langle y \mid u\rangle/\eta \ \text{ and } \ \varphi(y/\eta) + \varphi^*(u) \leqslant \langle y/\eta \mid u\rangle \\ &\Leftrightarrow \quad \mu = \varphi(y/\eta) - \langle y \mid u\rangle/\eta \ \text{ and } \ \varphi(y/\eta) + \varphi^*(u) = \langle y/\eta \mid u\rangle \\ &\Leftrightarrow \quad \mu = \varphi(y/\eta) - \langle y \mid u\rangle/\eta \ \text{ and } \ u \in \partial\varphi(y/\eta). \end{aligned} \tag{2.7}$$

We have thus proved (2.3). ☐

**Remark 2.4** Some of the results of Proposition 2.3 have already been obtained in the case when $\mathcal{G} = \mathbb{R}^N$ with different tools, some of which are specific to the finite-dimensional setting. Thus, items (ii) and (iv) can be found in [57], and the case $\eta > 0$ of (v) appears in [20, Proposition 4].

As shown in [25], (2.3) is instrumental in computing the proximity operator of a perspective function. Here is an important refinement.

**Corollary 2.5** *Let $\varphi \in \Gamma_0(\mathcal{G})$ and denote by $\operatorname{bar}\operatorname{dom}\varphi^*$ the barrier cone of $\operatorname{dom}\varphi^*$. Let $\eta \in \mathbb{R}$, let $y \in \mathcal{G}$, and suppose that one of the following holds:*

(i) $y \notin \operatorname{bar}\operatorname{dom}\varphi^*$.

(ii) $\operatorname{dom}\varphi^*$ *is open.*

(iii) $\operatorname{dom}\varphi^* = \mathcal{G}$.

(iv) $\varphi$ *is supercoercive:* $\lim_{\|y\|\to+\infty} \varphi(y)/\|y\| = +\infty$.

(v) *For every $v \in \mathcal{G}$, $\varphi - \langle \cdot \mid v\rangle$ is coercive.*

*Then*

$$\partial\widetilde{\varphi}(\eta, y) = \begin{cases} \big\{\big(\varphi(y/\eta) - \langle y \mid u\rangle/\eta, u\big) \;\big|\; u \in \partial\varphi(y/\eta)\big\}, & \text{if } \eta > 0; \\ C, & \text{if } \eta = 0 \text{ and } y = 0; \\ \varnothing, & \text{otherwise.} \end{cases} \tag{2.8}$$

*Proof*. In view of Proposition 2.3(v), it suffices to suppose that $y \neq 0$ and to show that

$$D = \big\{(\mu, u) \in \mathbb{R} \times \mathcal{G} \;\big|\; \mu + \varphi^*(u) \leqslant 0 \ \text{ and } \ \sigma_{\operatorname{dom}\varphi^*}(y) = \langle u \mid y\rangle\big\} = \varnothing. \tag{2.9}$$

Now denote by $\operatorname{spts}\operatorname{dom}\varphi^*$ the set of support points of $\operatorname{dom}\varphi^*$. Then

$$D = \big\{(\mu, u) \in \mathbb{R} \times (\operatorname{spts}\operatorname{dom}\varphi^*) \;\big|\; \mu + \varphi^*(u) \leqslant 0 \ \text{ and } \ \sigma_{\operatorname{dom}\varphi^*}(y) = \langle u \mid y\rangle\big\}. \tag{2.10}$$

(i): We have $\sigma_{\operatorname{dom}\varphi^*}(y) = +\infty$ and therefore (2.9) yields $D = \varnothing$.

(ii): We have spts dom $\varphi^* = \varnothing$ and therefore (2.10) yields $D = \varnothing$.

(iii)⇒(ii): Clear.

(iv)⇒(iii): [6, Proposition 14.15].

(v)⇒(iii): Let $v \in \mathcal{G}$. Then by the Moreau-Rockafellar theorem [6, Theorem 14.17], $\varphi - \langle \cdot \mid v \rangle$ is coercive if and only if $v \in \operatorname{int} \operatorname{dom} \varphi^*$. Hence $\mathcal{G} \subset \operatorname{int} \operatorname{dom} \varphi^*$. ☐

Next, we provide an example of a perspective function $g \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$ such that $g|_{\operatorname{dom} g}$ is discontinuous.

**Example 2.6** Suppose that $\mathcal{G} \neq \{0\}$, let $p \in \left]1, +\infty\right[$, and set

$$g\colon \mathbb{R} \oplus \mathcal{G} \to \left]-\infty, +\infty\right] : (\eta, y) \mapsto \begin{cases} \|y\|^p / \eta^{p-1}, & \text{if } \eta > 0; \\ 0, & \text{if } \eta = 0 \text{ and } y = 0; \\ +\infty, & \text{otherwise.} \end{cases} \tag{2.11}$$

Then $g \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$ and $g|_{\operatorname{dom} g}$ is not continuous at $(0,0)$. Indeed, set $\varphi = \|\cdot\|^p$. Then $\varphi$ is a supercoercive function in $\Gamma_0(\mathcal{G})$, and it thus follows from (2.1) that $\operatorname{rec} \varphi = \iota_{\{0\}}$. Hence (2.11) coincides with (2.2) and the first claim is therefore an application of Proposition 2.3(ii) with $g = \widetilde{\varphi}$. Now set $y = (0,0) \in \mathbb{R} \times \mathcal{G}$, let $v \in \mathcal{G}$ be such that $\|v\| = 1$, fix a sequence $(\alpha_n)_{n\in\mathbb{N}}$ in $\left]0, +\infty\right[$ such that $\alpha_n \downarrow 0$, and set $(\forall n \in \mathbb{N})\ y_n = (\alpha_n^{p/(p-1)}, \alpha_n v)$. Then $(y_n)_{n\in\mathbb{N}}$ lies in $\operatorname{dom} g$ and $y_n \to y$, but $\lim g(y_n) = 1 \neq 0 = g(y)$.

We now turn to some algebraic properties.

**Proposition 2.7** *Let $\varphi \in \Gamma_0(\mathcal{G})$. Then the following hold:*

(i) *Let $\psi \in \Gamma_0(\mathcal{G})$ be such that $\operatorname{dom} \varphi \cap \operatorname{dom} \psi \neq \varnothing$, and let $\lambda \in \left]0, +\infty\right[$. Then $[\lambda\varphi + \psi]^\sim = \lambda\widetilde{\varphi} + \widetilde{\psi} \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$.*

(ii) *Let $\Lambda\colon \mathcal{H} \to \mathcal{G}$ be linear, bounded, and such that $\operatorname{ran} \Lambda \cap \operatorname{dom} \varphi \neq \varnothing$. Set $\widetilde{\Lambda}\colon \mathbb{R} \oplus \mathcal{H} \to \mathbb{R} \oplus \mathcal{G}\colon (\xi, x) \mapsto (\xi, \Lambda x)$. Then $[\varphi \circ \Lambda]^\sim = \widetilde{\varphi} \circ \widetilde{\Lambda} \in \Gamma_0(\mathbb{R} \oplus \mathcal{H})$.*

(iii) *Suppose that $\varphi$ is positively homogeneous with $\operatorname{dom} \varphi = \mathcal{G}$, let $\phi \in \Gamma_0(\mathbb{R})$ be increasing on $\operatorname{ran} \varphi$ and such that $0 \in \operatorname{dom} \phi$, let $\eta \in \mathbb{R}$, and let $y \in \mathcal{G}$. Then $[\phi \circ \varphi]^\sim \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$ and $[\phi \circ \varphi]^\sim(\eta, y) = \widetilde{\phi}(\eta, \varphi(y))$.*

*Proof.* (i): We have $\operatorname{dom}(\varphi + \psi) \neq \varnothing$. Hence $\varphi + \psi \in \Gamma_0(\mathcal{G})$ and (2.1) implies that $\operatorname{rec}(\lambda\varphi + \psi) = \lambda \operatorname{rec} \varphi + \operatorname{rec} \psi$. The claim therefore follows from (2.2) and Proposition 2.3(ii).

(ii): Let $\xi \in \mathbb{R}$ and $x \in \mathcal{H}$. If $\xi > 0$, then $[\varphi \circ \Lambda]^\sim(\xi, x) = \xi(\varphi \circ \Lambda)(x/\xi) = \xi\varphi(\Lambda x/\xi) = (\widetilde{\varphi} \circ \widetilde{\Lambda})(\xi, x)$. Furthermore, we have $\operatorname{dom}(\varphi \circ \Lambda) \neq \varnothing$. Hence, $\varphi \circ \Lambda \in \Gamma_0(\mathcal{H})$ and (2.1) yields $\operatorname{rec}(\varphi \circ \Lambda) = (\operatorname{rec} \varphi) \circ \Lambda$. Hence, we derive from (2.2) that

$$[\varphi \circ \Lambda]^\sim(0, x) = \operatorname{rec}(\varphi \circ \Lambda)(x) = (\operatorname{rec} \varphi)(\Lambda x) = (\widetilde{\varphi} \circ \widetilde{\Lambda})(0, x). \tag{2.12}$$

Finally, if $\xi < 0$, then $[\varphi \circ \Lambda]^{\sim}(\xi, x) = +\infty = (\widetilde{\varphi} \circ \widetilde{\Lambda})(\xi, x)$. Altogether, the conclusion follows from Proposition 2.3(ii).

(iii): The assumptions imply that $\varphi$ is continuous and that $\varphi(0) = 0$. In turn $\phi \circ \varphi$ is lower semicontinuous and $0 \in \operatorname{dom}(\phi \circ \varphi)$. It also follows from the assumptions that $\phi \circ \varphi$ is convex. Altogether, $\phi \circ \varphi \in \Gamma_0(\mathcal{G})$ and we deduce from Proposition 2.3(ii) that $[\phi \circ \varphi]^{\sim} \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$. Now suppose that $\eta > 0$. Then

$$[\phi \circ \varphi]^{\sim}(\eta, y) = \eta\phi\big(\varphi(y/\eta)\big) = \eta\phi\big(\varphi(y)/\eta\big) = \widetilde{\phi}\big(\eta, \varphi(y)\big). \tag{2.13}$$

Next, we observe that, since $0 \in \operatorname{dom}(\phi \circ \varphi)$ and $0 \in \operatorname{dom}\phi$, (2.2) and (2.1) yield

$$\begin{aligned}
[\phi \circ \varphi]^{\sim}(0, y) &= \operatorname{rec}(\phi \circ \varphi)(y) \\
&= \lim_{\alpha \to +\infty} \frac{(\phi \circ \varphi)(0 + \alpha y)}{\alpha} \\
&= \lim_{\alpha \to +\infty} \frac{\phi\big(\varphi(\alpha y)\big)}{\alpha} \\
&= \lim_{\alpha \to +\infty} \frac{\phi\big(0 + \alpha\varphi(y)\big)}{\alpha} \\
&= (\operatorname{rec}\phi)\big(\varphi(y)\big) \\
&= \widetilde{\phi}\big(0, \varphi(y)\big).
\end{aligned} \tag{2.14}$$

Finally, if $\eta < 0$, then $[\phi \circ \varphi]^{\sim}(\eta, y) = +\infty = \widetilde{\phi}(\eta, \varphi(y))$. ☐

**Corollary 2.8** *Let $\psi \in \Gamma_0(\mathcal{G})$ and let $C$ be a closed convex subset of $\mathcal{G}$ such that $C \cap \operatorname{dom}\psi \neq \varnothing$. Set*

$$g\colon \mathbb{R} \times \mathcal{G} \to \left]-\infty, +\infty\right] : (\eta, y) \mapsto \begin{cases} \eta\psi(y/\eta), & \text{if } \eta > 0 \text{ and } y \in \eta(C \cap \operatorname{dom}\psi);\\ (\operatorname{rec}\psi)(y), & \text{if } \eta = 0 \text{ and } y \in \operatorname{rec}C;\\ +\infty, & \text{otherwise.} \end{cases} \tag{2.15}$$

*Then $g \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$.*

*Proof*. This is an application of Proposition 2.7(i) with $\lambda = 1$ and $\varphi = \iota_C$. Indeed, in this setting, $\operatorname{rec}(\varphi + \psi) = \operatorname{rec}\iota_C + \operatorname{rec}\psi = \iota_{\operatorname{rec}C} + \operatorname{rec}\psi$ and (2.15) yields $g = [\iota_C + \psi]^{\sim}$. ☐

**Corollary 2.9** *Let $\varphi \in \Gamma_0(\mathcal{G})$, let $\psi \in \Gamma_0(\mathcal{G})$ be a positively homogeneous function such that $\operatorname{dom}\varphi \cap \operatorname{dom}\psi \neq \varnothing$, and let $\delta \in \mathbb{R}$. Then $[\varphi + \psi + \delta]^{\sim} \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$ and*

$$(\forall \eta \in \mathbb{R})(\forall y \in \mathcal{G}) \quad [\varphi + \psi + \delta]^{\sim}(\eta, y) = \widetilde{\varphi}(\eta, y) + \psi(y) + \delta\eta. \tag{2.16}$$

*Proof*. This follows from (2.2) and Proposition 2.7(i) since $\operatorname{rec}(\varphi + \psi + \delta) = (\operatorname{rec}\varphi) + (\operatorname{rec}\psi) = (\operatorname{rec}\varphi) + \psi$. ☐

**Corollary 2.10** *Let $\varphi \in \Gamma_0(\mathcal{G})$. Then $(\forall (\zeta, \eta) \in \mathbb{R}^2)(\forall y \in \mathcal{G})\ \widetilde{\widetilde{\varphi}}(\zeta, \eta, y) = \widetilde{\varphi}(\eta, y)$.*

*Proof*. By Proposition 2.3(i)–(ii), $\widetilde{\varphi}$ is a positively homogeneous function in $\Gamma_0(\mathbb{R} \oplus \mathcal{G})$. Hence the claim follows from Corollary 2.9. ☐

**Proposition 2.11** *Let $I$ be a finite set and let $\eta \in \mathbb{R}$. For every $i \in I$, let $\mathcal{G}_i$ be a real Hilbert space, let $\varphi_i \in \Gamma_0(\mathcal{G}_i)$, and let $y_i \in \mathcal{G}_i$. Set $\bigoplus_{i\in I} \varphi_i \colon \bigoplus_{i\in I} \mathcal{G}_i \to \left]-\infty,+\infty\right] : (z_i)_{i\in I} \mapsto \sum_{i\in I} \varphi_i(z_i)$. Then*

$$\left(\bigoplus_{i\in I} \varphi_i\right)^{\sim} \Big(\eta, (y_i)_{i\in I}\Big) = \left(\bigoplus_{i\in I} \widetilde{\varphi}_i\right)\Big(\big(\eta, y_i\big)\Big)_{i\in I}. \tag{2.17}$$

*Proof*. Suppose that $\eta > 0$. Then

$$\left(\bigoplus_{i\in I} \varphi_i\right)^{\sim} (\eta, (y_i)_{i\in I}) = \eta\left(\bigoplus_{i\in I} \varphi_i\right)(y_i/\eta)_{i\in I} = \sum_{i\in I} \eta\varphi_i(y_i/\eta) = \left(\bigoplus_{i\in I} \widetilde{\varphi}_i\right)\Big(\big(\eta, y_i\big)\Big)_{i\in I}. \tag{2.18}$$

Now suppose that $\eta = 0$. Then (2.1) implies that $\operatorname{rec} \bigoplus_{i\in I} \varphi_i = \bigoplus_{i\in I} \operatorname{rec} \varphi_i$ and (2.17) follows. Finally, if $\eta < 0$, then both sides of (2.17) are equal to $+\infty$. ☐

Perspective functions can be used to provide examples of nonintuitive behaviors for minimizing sequences in optimization problems.

**Example 2.12** Suppose that $\mathcal{G} = \mathbb{R}$. Then Proposition 2.3(ii) asserts that the function

$$g = \big[\,|\cdot|^2\big]^{\sim} \colon \mathbb{R}^2 \to \left]-\infty,+\infty\right] : (\xi_1,\xi_2) \mapsto \begin{cases} \xi_2^2/\xi_1, & \text{if } \xi_1 > 0;\\ 0, & \text{if } \xi_1 = \xi_2 = 0;\\ +\infty, & \text{otherwise} \end{cases} \tag{2.19}$$

belongs to $\Gamma_0(\mathbb{R}^2)$. Moreover, $\operatorname{Argmin} g = \left[0,+\infty\right[ \times \{0\}$. Now let $p \in \left[1,+\infty\right[$ and set $(\forall n \in \mathbb{N})$ $x_n = ((n+1)^{p+2}, n+1)$. Then $(x_n)_{n\in\mathbb{N}}$ is a minimizing sequence of $g$ since $g(x_n) - \min g(\mathbb{R}^2) = 1/(n+1)^p \downarrow 0$. However, $d_{\operatorname{Argmin} g}(x_n) = n+1 \uparrow +\infty$. To sum up,

$$g(x_n) - \min g(\mathbb{R}^2) = O(1/n^p), \quad \text{while} \quad (\forall x \in \operatorname{Argmin} g) \quad \|x_n - x\| \uparrow +\infty. \tag{2.20}$$

This illustrates the fact that, even if it induces a very good convergence rate of the objective values $(g(x_n))_{n\in\mathbb{N}}$, a minimizing sequence $(x_n)_{n\in\mathbb{N}}$ may have extremely poor properties in terms of actually approaching a solution to the underlying minimization problem.

We now describe constructions of lower semicontinuous convex functions based on perspective functions. The first result is based on the composition of the perspective of a convex function with an affine operator.

**Proposition 2.13** *Let $L \colon \mathcal{H} \to \mathcal{G}$ be linear and bounded, let $\varphi \in \Gamma_0(\mathcal{G})$, let $r \in \mathcal{G}$, let $u \in \mathcal{H}$, let $\rho \in \mathbb{R}$, and set*

$$f \colon \mathcal{H} \to \left]-\infty,+\infty\right] : x \mapsto \begin{cases} (\langle x \mid u\rangle - \rho)\varphi\left(\dfrac{Lx - r}{\langle x \mid u\rangle - \rho}\right), & \text{if } \langle x \mid u\rangle > \rho;\\ (\operatorname{rec}\varphi)\big(Lx - r\big), & \text{if } \langle x \mid u\rangle = \rho;\\ +\infty, & \text{if } \langle x \mid u\rangle < \rho. \end{cases} \tag{2.21}$$

*Suppose that there exists $z \in \mathcal{H}$ such that $Lz \in r + (\langle z \mid u\rangle - \rho)\operatorname{dom}\varphi$ and $\langle z \mid u\rangle \geqslant \rho$, and set $A \colon \mathcal{H} \to \mathbb{R} \oplus \mathcal{G} \colon x \mapsto (\langle x \mid u\rangle - \rho, Lx - r)$. Then $f = \widetilde{\varphi} \circ A \in \Gamma_0(\mathcal{H})$.*

*Proof*. By construction, $A$ is a continuous affine operator, while $\widetilde{\varphi} \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$ by Proposition 2.3(ii). Therefore $f = \widetilde{\varphi} \circ A$ is lower semicontinuous and convex. Finally, to show that $f$ is proper, suppose first that $\langle z \mid u \rangle > \rho$. Then $(Lz - r)/(\langle z \mid u \rangle - \rho) \in \operatorname{dom} \varphi$ and hence $z \in \operatorname{dom} f$. On the other hand, if $\langle z \mid u \rangle = \rho$, then $Lz - r \in \{0\}$. In turn, $f(z) = (\operatorname{rec} \varphi)(0) = 0$ and therefore $z \in \operatorname{dom} f$. ☐

The next result involves the marginal of a perspective function (see [1] for a special case in the context of game theory).

**Proposition 2.14** *Let $\varphi \in \Gamma_0(\mathcal{G})$ and let $K$ be a nonempty closed bounded interval in $[0, +\infty[$. Define*

$$g\colon \mathcal{G} \to \mathbb{R}\colon y \mapsto \inf_{\eta \in K} \widetilde{\varphi}(\eta, y). \tag{2.22}$$

*Then $g \in \Gamma_0(\mathcal{G})$.*

*Proof*. Proposition 2.3(ii) asserts that $\widetilde{\varphi} \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$. In turn, it follows from [6, Proposition 8.26] that $g$ is convex and from [6, Lemma 1.29] that it is lower semicontinuous and proper. ☐

# 3 Examples of perspective functions

Our first construction involves a difference of convex functions.

**Corollary 3.1** *Let $\psi \in \Gamma_0(\mathcal{G})$ and let $\operatorname{env}(\psi^*)\colon u \mapsto \inf_{v \in \mathcal{G}}(\psi^*(v) + \|u - v\|^2/2)$ be the Moreau envelope of $\psi^*$. Set*

$$g\colon \mathbb{R} \times \mathcal{G} \to \left]-\infty, +\infty\right] : (\eta, y) \mapsto \begin{cases} \dfrac{\|y\|^2}{2\eta} - \eta(\operatorname{env} \psi)(y/\eta), & \text{if } \eta > 0; \\ \sigma_{\operatorname{dom} \psi}(y), & \text{if } \eta = 0; \\ +\infty, & \text{if } \eta < 0. \end{cases} \tag{3.1}$$

*Then $g = [\operatorname{env}(\psi^*)]^\sim \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$.*

*Proof*. Set $q = \|\cdot\|^2/2$ and $\varphi = q - \operatorname{env} \psi$, and let $\square$ denote the infimal convolution operation. It follows from Moreau's decomposition [48] (see also [6, Theorem 14.3(i)]) that $\varphi = \operatorname{env}(\psi^*) \in \Gamma_0(\mathcal{G})$. In addition, from basic convex analysis,

$$\varphi^* = (\psi^* \,\square\, q)^* = \psi^{**} + q = \psi + q \tag{3.2}$$

and therefore Lemma 2.2(ii) yields

$$\operatorname{rec} \varphi = \sigma_{\operatorname{dom} \varphi^*} = \sigma_{\operatorname{dom} \psi}. \tag{3.3}$$

In view of (2.2) and Proposition 2.3(ii), we conclude that $g = \widetilde{\varphi} \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$. ☐

**Example 3.2 (generalized Huber function)** Let $C$ be a nonempty closed convex subset of $\mathcal{G}$ and let $P_C$ denote its projector. Upon setting $\psi = \iota_C$ in Corollary 3.1, we deduce that the function

$$g\colon \mathbb{R}\times\mathcal{G} \to \left]-\infty,+\infty\right] : (\eta,y) \mapsto \begin{cases} \langle y \mid P_C(y/\eta)\rangle - \dfrac{\eta\|P_C(y/\eta)\|^2}{2}, & \text{if } y \notin \eta C \text{ and } \eta > 0;\\ \dfrac{\|y\|^2}{2\eta}, & \text{if } y \in \eta C \text{ and } \eta > 0;\\ \sigma_C(y), & \text{if } \eta = 0;\\ +\infty, & \text{if } \eta < 0 \end{cases} \tag{3.4}$$

is in $\Gamma_0(\mathbb{R}\oplus\mathcal{G})$. More precisely, $g = \widetilde{\varphi}$, where $\varphi = \operatorname{env}(\psi^*) = \operatorname{env}\sigma_C$. Let us further specialize by taking $C = B(0;\rho)$ for some $\rho \in \left]0,+\infty\right[$. Then (3.4) reduces to

$$g\colon \mathbb{R}\times\mathcal{G} \to \left]-\infty,+\infty\right] : (\eta,y) \mapsto \begin{cases} \rho\|y\| - \dfrac{\eta\rho^2}{2}, & \text{if } \|y\| > \eta\rho \text{ and } \eta > 0;\\ \dfrac{\|y\|^2}{2\eta}, & \text{if } \|y\| \leqslant \eta\rho \text{ and } \eta > 0;\\ \rho\|y\|, & \text{if } \eta = 0;\\ +\infty, & \text{if } \eta < 0. \end{cases} \tag{3.5}$$

We infer from Corollary 3.1 that $g = \widetilde{\varphi}$, where $\varphi = \operatorname{env}(\rho\|\cdot\|) = q - d_C^2/2$, that is,

$$\varphi\colon \mathcal{G} \to \left]-\infty,+\infty\right] : y \mapsto \begin{cases} \rho\|y\| - \dfrac{\rho^2}{2}, & \text{if } \|y\| > \rho;\\ \dfrac{\|y\|^2}{2}, & \text{if } \|y\| \leqslant \rho. \end{cases} \tag{3.6}$$

In particular, if $\mathcal{G} = \mathbb{R}$, then $\varphi$ is known as the Huber function. This function was introduced in [36] and it plays an important role in robust statistics and signal processing [37, 51], while its perspective function appears implicitly in robust regression problems [37, 39, 53]. The fact that the Huber function is the Moreau envelope of the absolute value function can already be found in [16]; see also [17]. On the other hand, if we specialize the perspective function (3.5) to the case when $\mathcal{G} = \mathbb{R}$ and $\rho = 1$, we obtain the function

$$g\colon \mathbb{R}^2 \to \left]-\infty,+\infty\right] : (\eta,y) \mapsto \begin{cases} |y| - \dfrac{\eta}{2}, & \text{if } |y| > \eta \text{ and } \eta > 0;\\ \dfrac{|y|^2}{2\eta}, & \text{if } |y| \leqslant \eta \text{ and } \eta > 0;\\ |y|, & \text{if } \eta = 0;\\ +\infty, & \text{if } \eta < 0, \end{cases} \tag{3.7}$$

which is used in computer vision [61], where it is called the bivariate Huber function.

We now consider a function that combines distance and support functions.

**Example 3.3 (generalized Berhu function)** Let $C$ and $D$ be nonempty closed convex subsets of $\mathcal{G}$, and let $\rho \in \left]0,+\infty\right[$. Then the function

$$g\colon \mathbb{R}\times\mathcal{G}\to \left]-\infty,+\infty\right] : (\eta,y)\mapsto \begin{cases} \dfrac{\eta d_C^2(y/\eta)}{2\rho}+\sigma_D(y), & \text{if } \eta>0 \text{ and } y\notin \eta C;\\ \sigma_D(y), & \text{if } \eta>0 \text{ and } y\in\eta C;\\ \sigma_D(y), & \text{if } \eta=0 \text{ and } y\in\operatorname{rec} C;\\ +\infty, & \text{otherwise} \end{cases} \tag{3.8}$$

is in $\Gamma_0(\mathbb{R}\oplus\mathcal{G})$. To show this, set $q=\|\cdot\|^2/2$, $\varphi=d_C^2/(2\rho)$, and $\psi=\sigma_D$. Then $\varphi\in\Gamma_0(\mathcal{G})$ and $\psi$ is a positively homogeneous function in $\Gamma_0(\mathcal{G})$ such that $0\in\operatorname{dom}\varphi\cap\operatorname{dom}\psi$. Furthermore, since $\varphi=\iota_C\,\square\,(q/\rho)$, we have $\varphi^*=\iota_C^*+(q/\rho)^*=\iota_C^*+\rho q$ and therefore $\operatorname{dom}\varphi^*=\operatorname{dom}\iota_C^*$. In turn, Lemma 2.2 yields

$$\operatorname{rec}\varphi=\sigma_{\operatorname{dom}\varphi^*}=\sigma_{\operatorname{dom}\iota_C^*}=\operatorname{rec}\iota_C=\iota_{\operatorname{rec}C}. \tag{3.9}$$

Altogether,

$$g=\left[\frac{d_C^2}{2\rho}+\sigma_D\right]^{\sim} \tag{3.10}$$

and the claim follows from Corollary 2.9. An especially interesting case is obtained when $C=B(0;\rho)$ and $D=B(0;1)$. Then $\operatorname{rec}C=\{0\}$, $\sigma_D=\|\cdot\|$, and (3.8) therefore becomes

$$g\colon \mathbb{R}\times\mathcal{G}\to \left]-\infty,+\infty\right] : (\eta,y)\mapsto \begin{cases} \dfrac{\|y\|^2+\rho^2\eta^2}{2\eta\rho}, & \text{if } \eta>0 \text{ and } \|y\|>\eta\rho;\\ \|y\|, & \text{if } \eta>0 \text{ and } \|y\|\leqslant\eta\rho;\\ 0, & \text{if } \eta=0 \text{ and } y=0;\\ +\infty, & \text{otherwise.} \end{cases} \tag{3.11}$$

As seen above, $g$ is the perspective function of

$$\vartheta\colon \mathcal{G}\to \left]-\infty,+\infty\right] : y\mapsto \begin{cases} \dfrac{\|y\|^2+\rho^2}{2\rho}, & \text{if } \|y\|>\rho;\\ \|y\|, & \text{if } \|y\|\leqslant\rho. \end{cases} \tag{3.12}$$

In the special case when $\mathcal{G}=\mathbb{R}$, $\vartheta$ arises in mechanics [3, 15] as well as in statistics [53], where it is called the Berhu (or reverse Huber) function. The reason for this terminology is that (3.6) exhibits a quadratic behavior on $B(0;\rho)$ and a sublinear behavior outside, while (3.12) exhibits a sublinear behavior on $B(0;\rho)$ and a quadratic behavior outside. Applications of the perspective of the Berhu function in robust regression can be found in [40] and in [53].

We now turn to a type of function that is used in support vector machines and in computer vision.

**Example 3.4 (generalized Vapnik loss function)** Let $\varepsilon\in\left]0,+\infty\right[$. By applying Proposition 2.7(iii) to $\phi\colon t\mapsto\max\{|t|-\varepsilon,0\}$ and $\varphi=\|\cdot\|$, we obtain that the function

$$g\colon \mathbb{R}\times\mathcal{G}\to \left]-\infty,+\infty\right] : (\eta,y)\mapsto \begin{cases} d_{B(0;\varepsilon\eta)}(y), & \text{if } \eta>0;\\ \|y\|, & \text{if } \eta=0;\\ +\infty, & \text{if } \eta<0. \end{cases} \tag{3.13}$$

is the perspective function of $\vartheta = \max\{\|\cdot\| - \varepsilon, 0\}$ and that it is in $\Gamma_0(\mathbb{R} \oplus \mathcal{G})$. A special case of this function appears in the context of computer vision in [61]. When $\mathcal{G} = \mathbb{R}$, $\vartheta$ is known as Vapnik's $\varepsilon$-insensitive loss function and it is employed in the area of support vector machines [59].

Our next construction involves a mix of positively homogeneous and norm-like functions.

**Example 3.5** Let $\psi\colon \mathcal{G} \to [0,+\infty[$ be a proper, lower semicontinuous, positively homogeneous convex function, let $\delta \in \mathbb{R}$, let $\rho \in [0,+\infty[$, let $p \in [1,+\infty[$, let $v \in \mathcal{G}$, and set

$$g\colon \mathbb{R} \times \mathcal{G} \to ]-\infty,+\infty] : (\eta, y) \mapsto \begin{cases} \delta\eta + \langle y \mid v\rangle + \big|\rho\eta^p + \psi^p(y)\big|^{1/p}, & \text{if } \eta \geqslant 0;\\ +\infty, & \text{if } \eta < 0. \end{cases} \tag{3.14}$$

Then $g = [\delta + \langle\cdot \mid v\rangle + |\rho + \psi^p|^{1/p}]^{\sim} \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$. Indeed, set $\varphi = \delta + \langle\cdot \mid v\rangle + |\rho + \psi^p|^{1/p}$ and $\phi = \big|\rho + |\cdot|^p\big|^{1/p}$. Then $\operatorname{rec}\phi = |\cdot|$ and $\varphi = \delta + \langle\cdot \mid v\rangle + \phi \circ \psi$. Altogether, we derive from Corollary 2.9 and Proposition 2.7(iii) that $g = \widetilde{\varphi} \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$. Let us now consider some special cases of this perspective function.

(i) Set $\psi = \|\cdot\|$, $v = 0$, and $p = 2$. Then (3.14) leads to the perspective function

$$g\colon \mathbb{R} \times \mathcal{G} \to ]-\infty,+\infty] : (\eta, y) \mapsto \begin{cases} \delta\eta + \sqrt{\rho\eta^2 + \|y\|^2}, & \text{if } \eta \geqslant 0;\\ +\infty, & \text{if } \eta < 0. \end{cases} \tag{3.15}$$

In the case when $\mathcal{G} = \mathbb{R}$, $\rho = 1$, and $\delta = -1$, this function shows up in computer vision [32], where $g(\eta, \cdot)$ is called the pseudo-Huber function.

(ii) Let $D$ be a nonempty closed convex cone in $\mathcal{G}$, let $v = 0$, let $\delta = 0$, let $\rho = 1$, and let $|||\cdot|||$ be a norm on $\mathcal{G}$. Set $\mathcal{H} = \mathbb{R} \oplus \mathcal{G}$, $K = [0,+\infty[ \times D$, and $\psi = |||\cdot||| + \iota_D$. Define a norm on $\mathcal{H}$ by $|||\cdot|||_p\colon (\eta, y) \mapsto (|\eta|^p + |||y|||^p)^{1/p}$. Then (3.14) yields $g = |||\cdot|||_p + \iota_K$, i.e.,

$$g\colon \mathcal{H} \to ]-\infty,+\infty] : z \mapsto \begin{cases} |||z|||_p, & \text{if } z \in K;\\ +\infty, & \text{if } z \notin K. \end{cases} \tag{3.16}$$

(iii) Consider the following setting in (ii): $N \geqslant 2$ is an integer, $\mathcal{G} = \mathbb{R}^{N-1}$, $|||\cdot|||$ is the $\ell^p$ norm on $\mathbb{R}^{N-1}$, $D = [0,+\infty[^{N-1}$, and $K = [0,+\infty[^N$. Then, if $\|\cdot\|_p$ denotes the $\ell^p$ norm on $\mathbb{R}^N$, the corresponding perspective function (3.14) is

$$g\colon \mathbb{R}^N \to ]-\infty,+\infty] : z \mapsto \begin{cases} \|z\|_p, & \text{if } z \in [0,+\infty[^N;\\ +\infty, & \text{if } z \notin [0,+\infty[^N. \end{cases} \tag{3.17}$$

(iv) Set $\mathcal{G} = \mathbb{R}$, $\psi = |\cdot|$, $v = -1$, $\rho = 1$, and $\delta = -1$. Then (3.14) yields the generalized Fischer-Burmeister function

$$g\colon \mathbb{R}^2 \to ]-\infty,+\infty] : (\eta, y) \mapsto \begin{cases} -\eta - y + \big|\eta^p + |y|^p\big|^{1/p}, & \text{if } \eta \geqslant 0;\\ +\infty, & \text{if } \eta < 0, \end{cases} \tag{3.18}$$

which is used is nonlinear complementarity problems [22]. The original Fischer-Burmeister function is obtained for $p = 2$.

The example below extends constructions found in robust estimation and in machine learning.

**Example 3.6** Let $\phi \in \Gamma_0(\mathbb{R})$ be an even function, let $v \in \mathcal{G}$, and let $\delta \in \mathbb{R}$. Then $\phi$ in increasing on $[0, +\infty[$ and $0 \in \operatorname{dom} \phi$. In turn, it follows from Corollary 2.9 and Proposition 2.7(iii) that the function

$$g\colon \mathbb{R} \oplus \mathcal{G} \to \left]-\infty, +\infty\right] : (\eta, y) \mapsto \begin{cases} \delta\eta + \langle y \mid v \rangle + \eta\phi(\|y\|/\eta), & \text{if } \eta > 0; \\ \langle y \mid v \rangle + (\operatorname{rec} \phi)(\|y\|), & \text{if } \eta = 0; \\ +\infty, & \text{if } \eta < 0 \end{cases} \tag{3.19}$$

is in $\Gamma_0(\mathbb{R} \oplus \mathcal{G})$. More precisely, $g = [\delta + \langle \cdot \mid v \rangle + \phi \circ \| \cdot \|]^{\sim}$. Now assume further that $\operatorname{dom} \phi^* = \mathbb{R}$. Then [5, Theorem 3.4] implies that $\phi^{**} = \phi$ is supercoercive and, therefore, that $\varphi$ is likewise. In turn, we derive from (2.1) that $\operatorname{rec} \varphi = \iota_{\{0\}}$, which allows us to rewrite (3.19) as

$$g\colon \mathbb{R} \oplus \mathcal{G} \to \left]-\infty, +\infty\right] : (\eta, y) \mapsto \begin{cases} \delta\eta + \langle y \mid v \rangle + \eta\phi(\|y\|/\eta), & \text{if } \eta > 0; \\ 0, & \text{if } \eta = 0 \text{ and } y = 0; \\ +\infty, & \text{otherwise.} \end{cases} \tag{3.20}$$

In particular, when $\mathcal{G} = \mathbb{R}$, $\phi = |\cdot|^2$, and $v = 0$, (3.20) has been used in robust estimation [37] and in machine learning [46].

**Example 3.7** Let $\rho \in \left]0, +\infty\right[$, let $p \in \left[1, +\infty\right[$, and set

$$\begin{aligned} g\colon \mathbb{R} \times \mathcal{G} &\to \left]-\infty, +\infty\right] \\ (\eta, y) &\mapsto \begin{cases} \dfrac{\rho\|y\|^p}{\eta^{p-1}} + p\eta \ln \eta - \eta \ln\left(\eta^p + \rho\|y\|^p\right), & \text{if } \eta > 0; \\ \rho\|y\|, & \text{if } \eta = 0 \text{ and } p = 1; \\ 0, & \text{if } \eta = 0,\ y = 0, \text{ and } p > 1; \\ +\infty, & \text{otherwise.} \end{cases} \end{aligned} \tag{3.21}$$

Upon invoking Proposition 2.3(iii) with $\varphi = \| \cdot \|$ and

$$\phi\colon \mathbb{R} \to \left]-\infty, +\infty\right] : t \mapsto \rho|t|^p - \ln\left(1 + \rho|t|^p\right), \tag{3.22}$$

we see that $g = [\phi \circ \varphi]^{\sim} \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$. For $p = 1$, (3.22) arises in inverse problems [21]. For $\mathcal{G} = \mathbb{R}$ and $\rho = p = 1$, (3.21) is closely related to the so-called "fair" function in robust statistics [56, Section 6.4.5]. For $\mathcal{G} = \mathbb{R}^N$ and $\rho = p = 1$, (3.21) is used in least-squares regularization [28].

**Example 3.8** Let $p \in \left[1, +\infty\right[$ and set

$$\begin{aligned} g\colon \mathbb{R} \times \mathcal{G} &\to \left]-\infty, +\infty\right] \\ (\eta, y) &\mapsto \begin{cases} p\eta \ln \eta - \eta \ln(\eta^p - \|y\|^p), & \text{if } \eta > 0 \text{ and } \|y\| < \eta; \\ 0, & \text{if } \eta = 0 \text{ and } y = 0; \\ +\infty, & \text{otherwise.} \end{cases} \end{aligned} \tag{3.23}$$

It follows from Proposition 2.3(iii) applied to $\varphi = \|\cdot\|$ and

$$\phi\colon \mathbb{R} \to ]-\infty, +\infty] : t \mapsto \begin{cases} -\ln\left(1 - |t|^p\right), & \text{if } |t| < 1; \\ +\infty, & \text{if } |t| \geqslant 1 \end{cases} \tag{3.24}$$

that $g = [\phi \circ \varphi]^{\sim} \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$. For $\mathcal{G} = \mathbb{R}^N$ and $p = 2$, (3.23) is closely related to a standard barrier for the Lorentz cone $\left\{(y, \eta) \in \mathbb{R}^{N+1} \mid \|y\| \leqslant \eta\right\}$ [50, Proposition 5.4.3].

Proposition 2.13 is an effective device for constructing a lower semicontinuous convex function in $\Gamma_0(\mathcal{H})$ by composing a perspective function $\widetilde{\varphi}$, for some $\varphi \in \Gamma_0(\mathcal{G})$, with a continuous affine operator $A\colon \mathcal{H} \to \mathbb{R} \oplus \mathcal{G}$ and, possibly, a suitable convexity preserving operation (see also Proposition 4.2). For instance, the generalized TREX estimator of [25] hinges on a special case of the following example in Euclidean spaces.

**Example 3.9** Let $L\colon \mathcal{H} \to \mathcal{G}$ be linear and bounded, let $|||\cdot|||$ be a norm on $\mathcal{G}$ such that, for some $\chi \in ]0, +\infty[$, $|||\cdot||| \geqslant \chi\|\cdot\|$, let $r \in \mathcal{G}$, let $u \in \mathcal{H}$, let $\rho \in \mathbb{R}$, let $q \in ]1, +\infty[$, and let $s \in [1, +\infty[$. Set

$$h\colon \mathcal{H} \to ]-\infty, +\infty] : x \mapsto \begin{cases} \dfrac{|||Lx - r|||^{qs}}{|\langle x \mid u\rangle - \rho|^{(q-1)s}}, & \text{if } \langle x \mid u\rangle > \rho; \\ 0, & \text{if } Lx = r \text{ and } \langle x \mid u\rangle = \rho; \\ +\infty, & \text{otherwise.} \end{cases} \tag{3.25}$$

Then $h \in \Gamma_0(\mathcal{H})$.

*Proof*. Set $\varphi = |||\cdot|||^q$. Then $\operatorname{dom}\varphi = \mathcal{G}$. In addition, $\varphi(y)/\|y\| \geqslant \chi^q\|y\|^q/\|y\| \to +\infty$ as $\|y\| \to +\infty$ and therefore (2.1) implies that $\operatorname{rec}\varphi = \iota_{\{0\}}$. Thus, (2.21) becomes

$$f\colon \mathcal{H} \to ]-\infty, +\infty] : x \mapsto \begin{cases} \dfrac{|||Lx - r|||^{q}}{|\langle x \mid u\rangle - \rho|^{q-1}}, & \text{if } \langle x \mid u\rangle > \rho; \\ 0, & \text{if } Lx = r \text{ and } \langle x \mid u\rangle = \rho; \\ +\infty, & \text{otherwise,} \end{cases} \tag{3.26}$$

and Proposition 2.13 asserts that $f \in \Gamma_0(\mathcal{H})$. Now let $\phi = |\cdot|^s$ and set $\phi(+\infty) = +\infty$. Then $\phi$ is increasing on $[0, +\infty] = \operatorname{ran} f$, continuous, and convex. Hence it follows from [23, Proposition II.8.4] and [6, Proposition 8.19] that $h = \phi \circ f \in \Gamma_0(\mathcal{H})$. ☐

**Example 3.10** Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space and let $\mathcal{H} = L^2(\Omega, \mathcal{F}, \mathsf{P})$ be the associated Hilbert space of square-integrable random variables. Let $\varphi \in \Gamma_0(\mathcal{H})$ and set

$$f\colon \mathcal{H} \to ]-\infty, +\infty] : X \mapsto \begin{cases} \mathsf{E}X\varphi\left(\dfrac{X}{\mathsf{E}X}\right), & \text{if } \mathsf{E}X > 0; \\ (\operatorname{rec}\varphi)(X), & \text{if } \mathsf{E}X = 0; \\ +\infty, & \text{if } \mathsf{E}X < 0. \end{cases} \tag{3.27}$$

Then $f \in \Gamma_0(\mathcal{H})$.

*Proof*. This is an application of Proposition 2.13 with $\mathcal{G} = \mathcal{H}$, $L = \operatorname{Id}$, $\mu = \mathsf{P}$, $u = 1$ a.s., $r = 0$ a.s., $z = 0$ a.s., and $\rho = 0$. ☐

# 4 Integral functions

In this section we construct lower semicontinuous functions by using as an integrand a perspective function. First, let us extend and formalize the divergence model (1.3).

**Proposition 4.1** *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, let $\mathsf{G}$ be a separable real Hilbert space, and let $\varphi \in \Gamma_0(\mathsf{G})$. Set $\mathcal{H} = L^2((\Omega, \mathcal{F}, \mu); \mathbb{R})$ and $\mathcal{G} = L^2((\Omega, \mathcal{F}, \mu); \mathsf{G})$, and suppose that one of the following holds:*

(i) $\mu(\Omega) < +\infty$.

(ii) $\varphi \geqslant \varphi(0) = 0$.

*For every $x \in \mathcal{H}$, set $\Omega_0(x) = \big\{\omega \in \Omega \;\big|\; x(\omega) = 0\big\}$ and $\Omega_+(x) = \big\{\omega \in \Omega \;\big|\; x(\omega) > 0\big\}$. Define*

$$\Phi \colon \mathcal{H} \oplus \mathcal{G} \to \left]-\infty, +\infty\right] : (x, y) \mapsto \begin{cases} \displaystyle\int_{\Omega_0(x)} (\operatorname{rec}\varphi)\big(y(\omega)\big)\mu(d\omega) + \int_{\Omega_+(x)} x(\omega)\varphi\left(\frac{y(\omega)}{x(\omega)}\right)\mu(d\omega), \\ \qquad\qquad \text{if } \begin{cases} x \geqslant 0 \text{ a.e.} \\ (\operatorname{rec}\varphi)(y)1_{\Omega_0(x)} + x\varphi(y/x)1_{\Omega_+(x)} \in L^1\big((\Omega, \mathcal{F}, \mu); \mathbb{R}\big); \end{cases} \\ +\infty, \qquad \text{otherwise.} \end{cases} \tag{4.1}$$

*Then $\Phi \in \Gamma_0(\mathcal{H} \oplus \mathcal{G})$.*

*Proof*. It follows from Proposition 2.3(ii) that $\widetilde{\varphi} \in \Gamma_0(\mathbb{R} \oplus \mathsf{G})$. Furthermore, we derive from (2.2) and (4.1) that

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{G}) \quad \Phi(x, y) = \int_\Omega \widetilde{\varphi}\big(x(\omega), y(\omega)\big)\mu(d\omega). \tag{4.2}$$

In turn, [6, Proposition 9.32] yields $\Phi \in \Gamma_0(\mathcal{H} \oplus \mathcal{G})$. ☐

**Proposition 4.2** *Let $\Omega$ be a nonempty open subset of $\mathbb{R}^N$ and let $\mathcal{H}$ be the Sobolev space $H^1(\Omega)$, i.e., $\mathcal{H} = \big\{x \in L^2(\Omega) \;\big|\; \nabla x \in (L^2(\Omega))^N\big\}$. For every $x \in \mathcal{H}$, set $\Omega_-(x) = \big\{t \in \Omega \;\big|\; x(t) < 0\big\}$, $\Omega_0(x) = \big\{t \in \Omega \;\big|\; x(t) = 0\big\}$, and $\Omega_+(x) = \big\{t \in \Omega \;\big|\; x(t) > 0\big\}$. Let $\varphi \in \Gamma_0(\mathbb{R}^N)$ be such that $\varphi \geqslant \varphi(0) = 0$, and define*

$$\begin{aligned} f \colon \mathcal{H} &\to \left]-\infty, +\infty\right] \\ x &\mapsto \begin{cases} \displaystyle\int_{\Omega_0(x)} (\operatorname{rec}\varphi)\big(\nabla x(t)\big)dt + \int_{\Omega_+(x)} x(t)\varphi\left(\frac{\nabla x(t)}{x(t)}\right)dt, & \text{if } x \geqslant 0 \text{ a.e.}; \\ +\infty, & \text{otherwise.} \end{cases} \end{aligned} \tag{4.3}$$

*Then $f \in \Gamma_0(\mathcal{H})$.*

*Proof*. Set $\mathcal{G} = (L^2(\Omega))^N$ and $\mathsf{G} = \mathbb{R}^N$, define $\Phi$ as in (4.1), where $(\Omega, \mathcal{F}, \mu)$ is the standard Lebesgue measure space, and let $L\colon \mathcal{H} \to \mathcal{H} \oplus \mathcal{G}\colon x \mapsto (x, \nabla x)$. Then $\Phi \in \Gamma_0(\mathcal{H} \oplus \mathcal{G})$ by Proposition 4.1(ii). On the other hand, since $\nabla\colon \mathcal{H} \to \mathcal{G}$ is bounded, $L$ is linear and continuous. Since $f(0) = 0$, we conclude that $f = \Phi \circ L \in \Gamma_0(\mathcal{H})$. □

The next examples recover two classical functions that have been used extensively in statistics (Fisher information) and in image recovery (total variation).

**Example 4.3** Consider the setting of Proposition 4.2.

(i) By choosing the supercoercive function $\varphi = \|\cdot\|_2^2$, we infer that the Fisher information

$$
\begin{aligned}
f\colon H^1(\Omega) &\to \left]-\infty, +\infty\right] \\
x &\mapsto \begin{cases} \displaystyle\int_{\Omega_+(x)} \frac{\|\nabla x(t)\|_2^2}{x(t)} dt, & \text{if } \begin{cases} x \geqslant 0 \text{ a.e.} \\ [\, x = 0 \;\Rightarrow\; \nabla x = 0 \,] \text{ a.e.;} \end{cases} \\ +\infty, & \text{otherwise} \end{cases}
\end{aligned}
\tag{4.4}
$$

is in $\Gamma_0(H^1(\Omega))$. The convexity properties of (1.2) over the subspace of strictly positive 1-dimensional smooth densities were apparently first discussed in [24]. The convexity and lower semicontinuity properties of extensions of the Fisher information, such as those used in [45] for $N = 1$ and based on $\varphi = |\cdot|^p$, with $p > 1$, or on higher order derivatives, can be obtained analogously.

(ii) By choosing the positively homogeneous function $\varphi = \|\cdot\|_2$, we infer that the total variation function

$$
\begin{aligned}
f\colon H^1(\Omega) &\to \left]-\infty, +\infty\right] \\
x &\mapsto \begin{cases} \displaystyle\int_{\Omega} \|\nabla x(t)\|_2 dt, & \text{if } x \geqslant 0 \text{ a.e.;} \\ +\infty, & \text{otherwise} \end{cases}
\end{aligned}
\tag{4.5}
$$

is in $\Gamma_0(H^1(\Omega))$.

We can also derive from Proposition 4.1 lower semicontinuous versions of a variety of standard divergences in the continuous and discrete cases. In the former, the underlying measure space is the Lebesgue measure space. The latter is illustrated below.

**Example 4.4** Let $N$ be a strictly positive integer, set $I = \{1, \ldots, N\}$, and let $\phi \in \Gamma_0(\mathbb{R})$. For every $x = (\xi_i)_{i \in I} \in \mathbb{R}^N$ and every $y = (\eta_i)_{i \in I} \in \mathbb{R}^N$, set $I_-(x) = \big\{ i \in I \;\big|\; \xi_i < 0 \big\}$, $I_0(x) = \big\{ i \in I \;\big|\; \xi_i = 0 \big\}$, $I_+(x) = \big\{ i \in I \;\big|\; \xi_i > 0 \big\}$, and

$$
\Phi(x, y) = \begin{cases} \displaystyle\sum_{i \in I_0(x)} (\operatorname{rec} \phi)(\eta_i) + \sum_{i \in I_+(x)} \xi_i \phi(\eta_i / \xi_i), & \text{if } I_-(x) = \varnothing; \\ +\infty, & \text{if } I_-(x) \neq \varnothing. \end{cases}
\tag{4.6}
$$

Then $\Phi \in \Gamma_0(\mathbb{R}^{2N})$. Indeed, this is a special case of Proposition 4.1(i), where $\Omega = I$, $\mathcal{F} = 2^I$, $\mu$ is the counting measure (hence $\mathcal{H} = \mathcal{G} = \mathbb{R}^N$), $\varphi = \phi$, and $\mathsf{G} = \mathbb{R}$. For instance, consider

$$\phi\colon \mathbb{R} \to \left]-\infty, +\infty\right] : t \mapsto \begin{cases} t \ln t, & \text{if } \ t > 0; \\ 0, & \text{if } \ t = 0; \\ +\infty, & \text{if } \ t < 0. \end{cases} \tag{4.7}$$

Then $\operatorname{rec} \phi = \iota_{\{0\}}$ and, if we set $J(x,y) = \big\{ i \in I \ \big| \ (\xi_i = 0 \text{ and } \eta_i \neq 0) \text{ or } (\xi_i > 0 \text{ and } \eta_i < 0) \big\}$,

$$\Phi(x,y) = \begin{cases} \displaystyle\sum_{i \in I_+(x) \cap I_+(y)} \eta_i \ln(\eta_i/\xi_i), & \text{if } \ I_-(x) \cup J(x,y) = \varnothing; \\ +\infty, & \text{otherwise} \end{cases} \tag{4.8}$$

is the Kullback-Leibler divergence between $x$ and $y$. This notion is central in statistics and in information theory. Another noteworthy family of discrete divergences is obtained by replacing (4.7) by

$$\phi\colon \mathbb{R} \to \left]-\infty, +\infty\right] : t \mapsto \begin{cases} \big|t^{1/p} - 1\big|^p, & \text{if } \ t \geqslant 0; \\ +\infty, & \text{if } \ t < 0, \end{cases} \qquad \text{where} \quad p \in \left[1, +\infty\right[. \tag{4.9}$$

In this case $\operatorname{rec} \phi = \sigma_{\left]-\infty, 1\right]}$ and, if we set $J(x,y) = \big\{ i \in I \ \big| \ \xi_i \geqslant 0 \text{ and } \eta_i < 0 \big\}$, (4.6) becomes

$$\Phi(x,y) = \begin{cases} \displaystyle\sum_{i \in I_0(x) \cap I_+(y)} \eta_i + \sum_{i \in I_+(x) \smallsetminus I_-(y)} \big|\eta_i^{1/p} - \xi_i^{1/p}\big|^p, & \text{if } \ I_-(x) \cup J(x,y) = \varnothing; \\ +\infty, & \text{otherwise.} \end{cases} \tag{4.10}$$

We recover the Kolmogorov variational divergence for $p = 1$ and the Hellinger divergence for $p = 2$.

**Acknowledgement.** The work of P. L. Combettes was partially supported by the CNRS MASTODONS project under grant 2016TABASCO.

## References


[1] M. Akian, S. Gaubert, and A. Hochart, Minimax representation of nonexpansive functions and application to zero-sum recursive games, *J. Convex Anal.,* to appear.

[2] S. M. Ali and S. D. Silvey, A general class of coefficients of divergence of one distribution from another, *J. Roy. Statist. Soc.,* vol. B28, pp. 131–142, 1966.

[3] J. J. Alibert, G. Bouchitté, I. Fragalà, and I. Lucardesi, A nonstandard free boundary problem arising in the shape optimization of thin torsion rods, *Interfaces Free Bound.*, vol. 15, pp. 95–119, 2013.

[4] M. Basseville, Distance measures for signal processing and pattern recognition, *Signal Processing,* vol. 18, pp. 349–369, 1989.

[5] H. H. Bauschke, J. M. Borwein, and P. L. Combettes, Essential smoothness, essential strict convexity, and Legendre functions in Banach spaces, *Comm. Contemp. Math.*, vol. 3, pp. 615–647, 2001.

[6] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces.* Springer, New York, 2011.

[7] A. Ben-Tal, A. Ben-Israel, and M. Teboulle, Certainty equivalents and information measures: Duality and extremal principles, *J. Math. Anal. Appl.*, vol. 157, pp. 211–236, 1991.

[8] J.-D. Benamou and Y. Brenier, A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem, *Numer. Math.,* vol. 84, pp. 375–393, 2000.

[9] J.-F. Bercher, Some properties of generalized Fisher information in the context of nonextensive thermostatistics, *Physica A,* vol. 392, pp. 3140–3154, 2013.

[10] A. Berlinet and I. Vajda, Selection rules based on divergences, *Statistics*, vol. 45, pp. 479–495, 2011.

[11] J. Bien, I. Gaynanova, J. Lederer, and C. L. Müller, Non-convex global minimization and false discovery rate control for the TREX, *J. Comp. Graph. Stat.,* to appear.

[12] D. E. Boekee, An extension of the Fisher information measure, in: I. Csiszár and P. Elias (eds.), *Topics in Information Theory,* János Bolyai Mathematical Society, vol. 16, pp. 113–123. North-Holland, Keszthely, Hungary, 1977.

[13] J. M. Borwein, A. S. Lewis, M. N. Limber, and D. Noll, Maximum entropy reconstruction using derivative information, part 2: Computational results, *Numer. Math.*, vol. 69, pp. 243–256, 1995.

[14] J. M. Borwein, A. S. Lewis, and D. Noll, Maximum entropy reconstruction using derivative information, part 1: Fisher information and convex duality, *Math. Oper. Res.,* vol. 21, pp. 442–468, 1996.

[15] G. Bouchitté, I. Fragalà, I. Lucardesi, and P. Seppecher, Optimal thin torsion rods and Cheeger sets, *SIAM J. Math. Anal.*, vol. 44, pp. 483–512, 2012.

[16] M. L. Bougeard, Connection between some statistical estimation criteria, lower-C2 functions and Moreau-Yosida approximates, in: *Bulletin International Statistical Institute,* 47th session, contributed papers, vol. 1, 159–160, 1989.

[17] M. L. Bougeard and C. D. Caquineau, Parallel proximal decomposition algorithms for robust estimation, *Ann. Oper. Res.*, vol. 90, pp. 247–270, 1999.

[18] L. Brasco, G. Buttazzo, and F. Santambrogio, A Benamou-Brenier approach to branched transport, *SIAM J. Math. Anal.*, vol. 43, pp. 1023–1040, 2011.

[19] L. M. Briceño-Arias, D. Kalise, and F. J. Silva, Proximal methods for stationary mean field games with local couplings, https://arxiv.org/pdf/1608.07701v1.pdf, 2016.

[20] S. Ceria and J. Soares, Convex programming for disjunctive convex optimization, *Math. Program.,* vol. A86, pp. 595–614, 1999.

[21] C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs, A variational formulation for frame-based inverse problems, *Inverse Problems,* vol. 23, pp. 1495–1518, 2007.

[22] J.-S. Chen, The semismooth-related properties of a merit function and a descent method for the nonlinear complementarity problem, *J. Global Optim.,* vol. 36, pp. 565–580, 2006.

[23] G. Choquet, *Topologie.* Masson, Paris, 1964 (English translation: *Topology.* Academic Press, New York, 1966).

[24] M. L. Cohen, The Fisher information and convexity, *IEEE Trans. Inform. Theory,* vol. 14, pp. 591–592, 1968.

[25] P. L. Combettes and C. L. Müller, Perspective functions: Proximal calculus and applications in high-dimensional statistics, *J. Math. Anal. Appl.*, published online 2016-12-15.

[26] I. Csiszár, Information-type measures of difference of probability distributions and indirect observations, *Studia Sci. Math. Hungar.,* vol. 2, pp. 299–318, 1967.

[27] B. Dacorogna and P. Maréchal, The role of perspective functions in convexity, polyconvexity, rank-one convexity and separate convexity, *J. Convex Anal.,* vol. 15, pp. 271–284, 2008.

[28] M. Elad, B. Matalon, and M. Zibulevsky, Coordinate and subspace optimization methods for linear least squares with non-quadratic regularization, *Appl. Comput. Harmon. Anal.*, vol. 23, pp. 346–367, 2007.

[29] R. A. Fisher, Theory of statistical estimation, *Proc. Cambridge. Philos. Soc.,* vol. 22, pp. 700–725, 1925.

[30] J. H. Fitschen, F. Laus, and G. Steidl, Transport between RGB images motivated by dynamic optimal transport, *J. Math. Imaging Vis.,* vol. 56, pp. 409–429, 2016.

[31] B. R. Frieden and R. A. Gatenby (eds.), *Exploratory Data Analysis Using Fisher Information.* Springer, New York, 2007.

[32] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision,* 2nd ed. Cambridge University Press, 2003.

[33] H. Hijazi, P. Bonami, G. Cornuéjols, and A. Ouorou, Mixed-integer nonlinear programs featuring "on/off" constraints, *Comput. Optim. Appl.,* vol. 52, pp. 537–558, 2012.

[34] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms.* Springer-Verlag, New York, 1993.

[35] J.-B. Hiriart-Urruty and J.-E. Martínez-Legaz, Convex solutions of a functional equation arising in information theory, *J. Math. Anal. Appl.*, vol. 328, pp. 1309–1320, 2007.

[36] P. J. Huber, Robust estimation of a location parameter, *Ann. Stat.,* vol. 35, pp. 73–101, 1964.

[37] P. J. Huber and E. M. Ronchetti, *Robust Statistics,* 2nd ed. Wiley, New York, 2009.

[38] M. N. Jung, C. Kirches, and S. Sager, On perspective functions and vanishing constraints in mixed-integer nonlinear optimal control, in: *Facets of Combinatorial Optimization,* pp. 387–417. Springer, Heidelberg, 2013.

[39] S. Lambert-Lacroix and L. Zwald, Robust regression through the Huber's criterion and adaptive lasso penalty, *Electron. J. Stat.,* vol. 5, pp. 1015–1053, 2011.

[40] S. Lambert-Lacroix and L. Zwald, The adaptive BerHu penalty in robust regression, *J. Nonparametr. Stat.*, vol. 28, pp. 487–514, 2016.

[41] P.-J. Laurent, *Approximation et Optimisation*, Hermann, Paris, 1972.

[42] J. Lederer and C. L. Müller, Don't fall for tuning parameters: Tuning-free variable selection in high dimensions with the TREX, *Proc. Twenty-Ninth AAAI Conf. Artif. Intell.*, pp. 2729–2735. AAAI Press, Austin, 2015.

[43] C. Lemaréchal, personnal communication.

[44] F. Liese and I. Vajda, On divergences and informations in statistics and information theory, *IEEE Trans. Inform. Theory,* vol. 52, pp. 4394–4412, 2006.

[45] P.-L. Lions and G. Toscani, A strengthened central limit theorem for smooth densities, *J. Funct. Anal.,* vol. 129, pp. 148–167, 1995.

[46] C. A. Micchelli, J. M. Morales, and M. Pontil, Regularizers for structured sparsity, *Adv. Comput. Math.,* vol. 38, pp. 455–489, 2013.

[47] N. Moehle and S. Boyd, A perspective-based convex relaxation for switched-affine optimal control, *Systems Control Lett.,* vol. 86, pp. 34–40, 2015.

[48] J. J. Moreau, Fonctions convexes duales et points proximaux dans un espace hilbertien, *C. R. Acad. Sci. Paris Sér. A Math.,* vol. 255, pp. 2897–2899, 1962.

[49] E. Ndiaye, O. Fercoq, A. Gramfort, V. Leclère, and J. Salmon, Efficient smoothed concomitant lasso estimation for high dimensional regression, https://arxiv.org/pdf/1606.02702v1.pdf, 2016.

[50] Yu. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming.* SIAM, Philadelphia, 1994.

[51] M. Nikolova and M. K. Ng, Analysis of half-quadratic minimization methods for signal and image recovery, *SIAM J. Sci. Comput.,* vol. 27, pp. 937–966, 2005.

[52] D. Noll, Reconstruction with noisy data: An approach via eigenvalue optimization, *SIAM J. Optim.,* vol. 8, pp. 82–104, 1998.

[53] A. B. Owen, A robust hybrid of lasso and ridge regression, *Contemp. Math.,* vol. 443, pp. 59–71, 2007.

[54] N. Papadakis, G. Peyré, and E. Oudet, Optimal transport with proximal splitting, *SIAM J. Imaging Sci.,* vol. 7, pp. 212–238, 2014.

[55] L. Pardo, *Statistical Inference Based on Divergence Measures.* Chapman and Hall/CRC, Boca Raton, FL, 2006.

[56] W. J. J. Rey, *Introduction to Robust and Quasi-Robust Statistical Methods.* Springer, Berlin, 1983.

[57] R. T. Rockafellar, *Convex Analysis.* Princeton University Press, Princeton, NJ, 1970.

[58] G. Toscani, A strengthened entropy power inequality for log-concave densities, *IEEE Trans. Inform. Theory,* vol. 61, pp. 6550–6559, 2015.

[59] V. N. Vapnik, *The Nature of Statistical Learning Theory,* 2nd ed. Springer, New York, 2000.

[60] C. Villani, Fisher information estimates for Boltzmann's collision operator, *J. Math. Pures. Appl.,* vol. 77, pp. 821–837, 1998.

[61] C. Zach and M. Pollefeys, Practical methods for convex multi-view reconstruction, *Lecture Notes in Comput. Sci.,* vol. 6314, pp. 354–367, 2010.