# Regularizing with Bregman–Moreau envelopes

Heinz H. Bauschke,* Minh N. Dao† and Scott B. Lindstrom‡

May 17, 2017

**Abstract**

Moreau's seminal paper, introducing what is now called the Moreau envelope and the proximity operator (a.k.a. proximal mapping), appeared in 1965. The Moreau envelope of a given convex function provides a regularized version which has additional desirable properties such as differentiability and full domain. Fifty years ago, Attouch proposed to use the Moreau envelope for regularization. Since then, this branch of convex analysis has developed in many fruitful directions. In 1967, Bregman introduced what is nowadays the Bregman distance as a measure of discrepancy between two points generalizing the square of the Euclidean distance. Proximity operators based on the Bregman distance have become a topic of significant research as they are useful in algorithmic solution of optimization problems. More recently, in 2012, Kan and Song studied regularization aspects of the left Bregman–Moreau envelope even for nonconvex functions.

In this paper, we complement previous works by analyzing the left and right Bregman–Moreau envelopes and by providing additional asymptotic results. Several examples are provided.

**2010 Mathematics Subject Classification:** Primary 90C25; Secondary 26A51, 26B25, 47H05, 47H09.

**Keywords:** Bregman distance, convex function, Moreau envelope, proximal mapping, proximity operator, regularization.

## 1 Introduction

We assume throughout that

$$X := \mathbb{R}^J, \tag{1}$$

which we equip with the standard inner product $\langle \cdot, \cdot \rangle$ and the induced Euclidean norm $\|\cdot\|$.

Let $\theta\colon X \to \,]-\infty, +\infty]$ be convex, lower semicontinuous, and proper[1]. The *Moreau envelope* with parameter $\gamma \in \mathbb{R}_{++}$ is the function

$$\operatorname{env}_\theta^\gamma\colon x \mapsto \inf_{y\in X} \theta(y) + \frac{1}{2\gamma}\|x-y\|^2. \tag{2}$$

*Mathematics, University of British Columbia, Kelowna, B.C. V1V 1V7, Canada. E-mail: heinz.bauschke@ubc.ca.

†CARMA, University of Newcastle, Callaghan, NSW 2308, Australia. E-mail: daonminh@gmail.com.

‡CARMA, University of Newcastle, Callaghan, NSW 2308, Australia. E-mail: scott.lindstrom@uon.edu.au.

[1]See [33], [8], [29], and [34] for background material in convex analysis from which we adopt our notation which is standard. We also set $\mathbb{R}_{++} := \{x \in \mathbb{R} \mid x > 0\}$.

Moreau only considered the case when $\gamma = 1$; the systematic study involving the parameter $\gamma$ originated with Attouch (see [2] and [3]). If $\theta = \iota_C$, the indicator function of a nonempty closed convex subset $C$ of $X$, then the corresponding Moreau envelope with parameter $\gamma$ is $\frac{1}{2\gamma}d_C^2$, where $d_C$ is the distance function of the set $C$. While the indicator function has (effective) domain $C$ and is differentiable only on $\operatorname{int} C$, the interior of $C$, the Moreau envelope is much better behaved: for instance, it has full domain and is differentiable everywhere.

Now assume that

$$f\colon X \to \left]-\infty, +\infty\right] \text{ is convex and differentiable on } U := \operatorname{int}\operatorname{dom} f \neq \varnothing. \tag{3}$$

The *Bregman distance*[2] associated with $f$, first explored by Bregman in [12] (see also [19]), is

$$D_f\colon X \times X \to [0, +\infty] : (x, y) \mapsto \begin{cases} f(x) - f(y) - \langle \nabla f(y), x - y \rangle, & \text{if } y \in U; \\ +\infty, & \text{otherwise.} \end{cases} \tag{4}$$

It serves as a measure of discrepancy between two points and thus gives rise to associated projectors (nearest-point mappings) and proximal mappings which have been employed to solve convex feasibility and optimization problems algorithmically; see, e.g., [1], [4], [6], [7], [9], [10], [11], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [25], [26], [27], [31], [32], and [35]. The classical case arises when $f = \frac{1}{2}\|\cdot\|^2$ in which case $D_f(x, y) = \frac{1}{2}\|x - y\|^2 = D_f(y, x)$. This clearly suggests to replace the quadratic term in (2) by the Bregman distance. However, because different assignments of $f$ may allow for cases when $D_f(x, y) \neq D_f(y, x)$, we actually are led to introduce *two* envelopes: The *left* and *right Bregman–Moreau envelopes* are defined by

$$\overleftarrow{\operatorname{env}}_\theta^\gamma\colon X \to [-\infty, +\infty] : y \mapsto \inf_{x \in X} \theta(x) + \frac{1}{\gamma} D_f(x, y) \tag{5}$$

and

$$\overrightarrow{\operatorname{env}}_\theta^\gamma\colon X \to [-\infty, +\infty] : x \mapsto \inf_{y \in X} \theta(y) + \frac{1}{\gamma} D_f(x, y), \tag{6}$$

respectively. It follows from the definition (see also Example 2.3 below) that if $f = \frac{1}{2}\|\cdot\|^2$, then $D_f\colon (x, y) \mapsto \frac{1}{2}\|x - y\|^2$, and $\overleftarrow{\operatorname{env}}_\theta^\gamma = \overrightarrow{\operatorname{env}}_\theta^\gamma = \theta \,\square\, (\frac{1}{2\gamma}\|\cdot\|^2)$ is the classical *Moreau envelope* of $\theta$ of parameter $\gamma$; see [30] and also [8, Section 12.4], [34, Section 1.G]. When $\gamma = 1$, we simply write $\overleftarrow{\operatorname{env}}_\theta$ for $\overleftarrow{\operatorname{env}}_\theta^1$, and $\overrightarrow{\operatorname{env}}_\theta$ for $\overrightarrow{\operatorname{env}}_\theta^1$. Bregman–Moreau envelopes when $\gamma \neq 1$ were previously explored in [21] and [25] for the *left* variant; the authors provided asymptotic results when $\gamma \downarrow 0$.

*The goal of this paper is to present a systematic study of regularization aspects of the Bregman–Moreau envelope.* Our results extend and complement several classical results and provide a novel way to approximate $\theta$. We also obtain new results on the asymptotic behaviour when $\gamma \uparrow +\infty$ and on the *right* Bregman–Moreau envelope.

The remainder of this paper is organized as follows. In Section 2, we collect various useful properties and characterizations of Bregman–Moreau envelopes. In particular, the minimizers of the envelopes are also minimizers of the original function (see Theorem 2.20). Section 3 is devoted to the asymptotic behaviour of the Bregman–Moreau envelopes when $\gamma \downarrow 0$ (Theorem 3.3) and when $\gamma \uparrow +\infty$ (Theorem 3.5). The final Section 4 provides an example where all objects can be computed and nicely visualized.

[2] Note that $D_f$ is *not* a distance in the sense of metric topology; however, this naming convention is now ubiquitous.

## 2 Bregman–Moreau envelopes: basic properties

In this section, we collect various useful properties of the Bregman–Moreau envelopes.

We start by describing the effect of scaling the function.

**Proposition 2.1.** *Let $\theta\colon X \to \left]-\infty,+\infty\right]$, let $\gamma \in \mathbb{R}_{++}$, and let $\mu \in \mathbb{R}_{++}$. Then $\overleftarrow{\operatorname{env}}^{\mu}_{\gamma\theta} = \gamma\,\overleftarrow{\operatorname{env}}^{\gamma\mu}_{\theta}$ and $\overrightarrow{\operatorname{env}}^{\mu}_{\gamma\theta} = \gamma\,\overrightarrow{\operatorname{env}}^{\gamma\mu}_{\theta}$.*

*Proof.* Let $y \in X$. By definition,

$$\overleftarrow{\operatorname{env}}^{\mu}_{\gamma\theta}(y) = \inf_{x\in X}\big((\gamma\theta)(x) + \frac{1}{\mu}D_f(x,y)\big) = \gamma \inf_{x\in X}\big(\theta(x) + \frac{1}{\gamma\mu}D_f(x,y)\big) = \gamma\,\overleftarrow{\operatorname{env}}^{\gamma\mu}_{\theta}(y). \tag{7}$$

We deduce that $\overleftarrow{\operatorname{env}}^{\mu}_{\gamma\theta} = \gamma\,\overleftarrow{\operatorname{env}}^{\gamma\mu}_{\theta}$. The proof that $\overrightarrow{\operatorname{env}}^{\mu}_{\gamma\theta} = \gamma\,\overrightarrow{\operatorname{env}}^{\gamma\mu}_{\theta}$ is analogous. ■

We now turn to regularization properties. (For a variant of Proposition 2.2(i), see [25, Theorem 2.2 and Proposition 2.1(i)].)

**Proposition 2.2.** *Let $\theta\colon X \to \left]-\infty,+\infty\right]$ be such that $U \cap \operatorname{dom}\theta \neq \varnothing$, and let $\gamma \in \mathbb{R}_{++}$. Then the following hold:*

(i) $\operatorname{dom}\overleftarrow{\operatorname{env}}^{\gamma}_{\theta} = U$, *and* $(\forall y \in U)(\forall \mu \in \left]\gamma,+\infty\right[)\ \inf\theta(X) \leq \overleftarrow{\operatorname{env}}^{\mu}_{\theta}(y) \leq \overleftarrow{\operatorname{env}}^{\gamma}_{\theta}(y) \leq \theta(y)$. *Consequently,* $\inf\theta(X) = \inf\overleftarrow{\operatorname{env}}^{\gamma}_{\theta}(X)$, *and* $\overleftarrow{\operatorname{env}}^{\gamma}_{\theta}(y) \downarrow \inf\theta(X)$ *as* $\gamma \uparrow +\infty$.

(ii) $\operatorname{dom}\overrightarrow{\operatorname{env}}^{\gamma}_{\theta} = \operatorname{dom} f$, *and* $(\forall x \in U)(\forall \mu \in \left]\gamma,+\infty\right[)\ \inf\theta(X) \leq \overrightarrow{\operatorname{env}}^{\mu}_{\theta}(x) \leq \overrightarrow{\operatorname{env}}^{\gamma}_{\theta}(x) \leq \theta(x)$. *Consequently,* $\inf\theta(X) = \inf\overrightarrow{\operatorname{env}}^{\gamma}_{\theta}(X)$, *and* $\overrightarrow{\operatorname{env}}^{\gamma}_{\theta}(x) \downarrow \inf\theta(X)$ *as* $\gamma \uparrow +\infty$.

*Proof.* (i): We first show that $\operatorname{dom}\overleftarrow{\operatorname{env}}^{\gamma}_{\theta} = U$. Let $y \in \operatorname{dom}\overleftarrow{\operatorname{env}}^{\gamma}_{\theta}$. Then $\overleftarrow{\operatorname{env}}^{\gamma}_{\theta}(y) = \inf_{x\in X}\theta(x) + \frac{1}{\gamma}D_f(x,y) < +\infty$, and hence there exists $x \in X$ such that $\theta(x) + \frac{1}{\gamma}D_f(x,y) < +\infty$. Since $\theta(x) > -\infty$, this yields $y \in U$.

From now on, let $y \in U$, and pick $u \in \operatorname{dom}\theta \cap U$. Then $-f(y) < +\infty$, $\|\nabla f(y)\| < +\infty$, $f(u) < +\infty$, $\theta(u) < +\infty$, and

$$\overleftarrow{\operatorname{env}}^{\gamma}_{\theta}(y) = \inf_{x\in X}\theta(x) + \frac{1}{\gamma}\big(f(x) - f(y) - \langle \nabla f(y), x - y\rangle\,\big), \tag{8a}$$

$$\leq \theta(u) + \frac{1}{\gamma}\big(f(u) - f(y) - \langle \nabla f(y), u - y\rangle\,\big) < +\infty, \tag{8b}$$

which gives $y \in \operatorname{dom}\overleftarrow{\operatorname{env}}^{\gamma}_{\theta}$. Hence, $\operatorname{dom}\overleftarrow{\operatorname{env}}^{\gamma}_{\theta} = U$.

Next, let $\mu \in \left]\gamma,+\infty\right[$. Then $\frac{1}{\mu} < \frac{1}{\gamma}$, $\theta \leq \theta + \frac{1}{\mu}D_f(\cdot,y) \leq \theta + \frac{1}{\gamma}D_f(\cdot,y)$, and so

$$\inf_{x\in X}\theta(x) \leq \overleftarrow{\operatorname{env}}^{\mu}_{\theta}(y) \leq \overleftarrow{\operatorname{env}}^{\gamma}_{\theta}(y) = \inf_{x\in X}\theta(x) + \frac{1}{\gamma}D_f(x,y) \leq \theta(y) + \frac{1}{\gamma}D_f(y,y) = \theta(y). \tag{9}$$

Therefore,

$$\inf\theta(X) \leq \overleftarrow{\operatorname{env}}^{\mu}_{\theta}(y) \leq \overleftarrow{\operatorname{env}}^{\gamma}_{\theta}(y) \leq \theta(y). \tag{10}$$

Taking now the infimum over $y \in U$ yields $\inf\theta(X) = \inf\overleftarrow{\operatorname{env}}^{\gamma}_{\theta}(X)$. Consequently,

$$\inf\theta(X) \leq \varliminf_{\gamma\to+\infty}\overleftarrow{\operatorname{env}}^{\gamma}_{\theta}(y). \tag{11}$$

On the other hand, $(\forall x \in X)$ $\overleftarrow{\operatorname{env}}_\theta^\gamma(y) \le \theta(x) + \frac{1}{\gamma}D_f(x,y)$, which implies that $(\forall x \in X)$ $\overline{\lim}_{\gamma\to+\infty} \overleftarrow{\operatorname{env}}_\theta^\gamma(y) \le \theta(x)$ and thus $\overline{\lim}_{\gamma\to+\infty} \overleftarrow{\operatorname{env}}_\theta^\gamma(y) \le \inf\theta(X)$. Altogether, $\lim_{\gamma\to+\infty} \overleftarrow{\operatorname{env}}_\theta^\gamma(y) = \inf\theta(X)$ and the conclusion follows from (10).

(ii): Similar to (i). ■

Denote by $\Gamma_0(X)$ the set of all proper lower semicontinuous convex functions from $X$ to $]-\infty,+\infty]$. From now on, we strengthen our assumptions by requiring that

$$f \in \Gamma_0(X) \text{ is a convex function of Legendre type and } U := \operatorname{int}\operatorname{dom} f. \tag{12}$$

This will allow us to obtain a quite satisfying theory in which the envelopes are convex functions. Note that $f$ is essentially smooth and essentially strictly convex in the sense of [33, Section 26]. It is well known that

$$\nabla f\colon U \to U^* := \operatorname{int}\operatorname{dom} f^* \text{ is a homeomorphism with } \big(\nabla f\big)^{-1} = \nabla f^*. \tag{13}$$

We will also work with the following standard assumptions:

**A1** $\nabla^2 f$ exists and is continuous on $U$;

**A2** $D_f$ is *jointly convex*, i.e., convex on $X \times X$;

**A3** $(\forall x \in U)$ $D_f(x,\cdot)$ is strictly convex on $U$;

**A4** $(\forall x \in U)$ $D_f(x,\cdot)$ is coercive, i.e., $D_f(x,y) \to +\infty$ as $\|y\| \to +\infty$.

We also henceforth assume that

$$\text{all assumptions } \mathbf{A1}\text{–}\mathbf{A4} \text{ hold.} \tag{14}$$

**Example 2.3 ([11, Example 2.16]).** Assumptions (12) and **A1–A4** hold in the following cases, where $x = (\xi_j)_{1\le j\le J}$ and $y = (\eta_j)_{1\le j\le J}$ are two generic points in $X = \mathbb{R}^J$.

(i) *Energy:* If $f\colon x \mapsto \frac{1}{2}\|x\|^2$, then $U = X$ and

$$D_f(x,y) = \tfrac{1}{2}\|x-y\|^2. \tag{15}$$

(ii) *Boltzmann–Shannon*[3] *entropy:* If $f\colon x \mapsto \sum_{j=1}^{J} \xi_j \ln(\xi_j) - \xi_j$, then $U = \{x \in X \mid x > 0\}$ and one obtains the *Kullback–Leibler divergence*

$$D_f(x,y) = \begin{cases} \sum_{j=1}^{J} \xi_j \ln(\xi_j/\eta_j) - \xi_j + \eta_j, & \text{if } x \ge 0 \text{ and } y > 0;\\ +\infty, & \text{otherwise.}\end{cases} \tag{16}$$

(iii) *Fermi–Dirac entropy:* If $f\colon x \mapsto \sum_{j=1}^{J} \xi_j \ln(\xi_j) + (1-\xi_j)\ln(1-\xi_j)$, then $U = \{x \in X \mid 0 < x < 1\}$ and

$$D_f(x,y) = \begin{cases} \sum_{j=1}^{J} \xi_j \ln(\xi_j/\eta_j) + (1-\xi_j)\log\big((1-\xi_j)/(1-\eta_j)\big), & \text{if } 0 \le x \le 1 \text{ and } 0 < y < 1;\\ +\infty, & \text{otherwise.}\end{cases} \tag{17}$$

[3]When dealing with the Boltzmann–Shannon entropy and Fermi–Dirac entropy, it is understood that $0\cdot\ln(0) := 0$. For two vectors $x$ and $y$ in $X$, expressions such as $x \le y$, $x\cdot y$, and $x/y$ are interpreted coordinate-wise.

The following result relates the Bregman–Moreau envelopes to Fenchel conjugates.

**Proposition 2.4.** *Let* $\theta\colon X \to \left]-\infty,+\infty\right]$ *be such that* $U \cap \operatorname{dom}\theta \neq \varnothing$, *and let* $\gamma \in \mathbb{R}_{++}$. *Then the following hold*[4]*:*

(i) $\gamma\, \overleftarrow{\operatorname{env}}^{\gamma}_{\theta} \circ \nabla f^* = f^* - (\gamma\theta + f)^*$.

(ii) $\gamma\, \overrightarrow{\operatorname{env}}^{\gamma}_{\theta} = f - (f^* + (\gamma\theta \circ \nabla f))^*$.

*Proof.* (i): This follows from [25, Theorem 2.4] and (13).

(ii): Let $x \in X$. Using the fact that $f^*(\nabla f(y)) = \langle \nabla f(y), y\rangle - f(y)$ (see, e.g., [33, Theorem 23.5]) and that $\left(\nabla f\right)^{-1} = \nabla f^*$ (see (13)), we obtain

$$
\begin{aligned}
\overrightarrow{\operatorname{env}}^{\gamma}_{\theta}(x) &= \inf_{y\in X} \left(\theta(y) + \frac{1}{\gamma}\big(f(x) - f(y) - \langle \nabla f(y), x-y\rangle\,\big)\right) && \text{(18a)}\\
&= \frac{f(x)}{\gamma} + \frac{1}{\gamma}\inf_{y\in U}\left(\gamma\theta(y) + f^*(\nabla f(y)) - \langle \nabla f(y), x\rangle\,\right) && \text{(18b)}\\
&= \frac{f(x)}{\gamma} + \frac{1}{\gamma}\inf_{y^*\in U^*}\left(\gamma\theta(\nabla f^*(y^*)) + f^*(y^*) - \langle y^*, x\rangle\,\right) && \text{(18c)}\\
&= \frac{f(x)}{\gamma} - \frac{1}{\gamma}\sup_{y^*\in X}\left(\,\langle x, y^*\rangle - \big((\gamma\theta\circ\nabla f^*) + f^*\big)(y^*)\right) && \text{(18d)}\\
&= \frac{f(x)}{\gamma} - \frac{1}{\gamma}\big((\gamma\theta\circ\nabla f^*) + f^*\big)^*(x). && \text{(18e)}
\end{aligned}
$$

This completes the proof. ■

In the sequel, we shall require the following two facts.

**Fact 2.5.** *The following hold:*

(i) $(\forall x \in X)(\forall y \in U)\;\; D_f(x,y) = 0 \;\;\Leftrightarrow\;\; x = y$.

(ii) $(\forall y \in U)$ $D_f(\cdot,y)$ *is coercive, i.e.,* $D_f(x,y) \to +\infty$ *as* $\|x\| \to +\infty$.

*Proof.* (i): [5, Theorem 3.7.(iv)]. (ii): [5, Theorem 3.7.(iii)]. ■

**Fact 2.6.** *Let* $\theta \in \Gamma_0(X)$ *be such that* $\operatorname{dom}\theta \cap U \neq \varnothing$, *and let* $\gamma \in \mathbb{R}_{++}$. *Consider the following properties:*

(a) $U \cap \operatorname{dom}\theta$ *is bounded.*

(b) $\inf\theta(U) > -\infty$.

(c) $f$ *is supercoercive, i.e.,* $f(x)/\|x\| \to +\infty$ *as* $\|x\| \to +\infty$.

(d) $(\forall x \in U)$ $D_f(x,\cdot)$ *is supercoercive.*

*Then the following hold:*

[4] Indeed, the proof does not require any of **A1**–**A4**.

(i) *If any of the conditions (a), (b), or (c) holds, then*

$$(\forall y \in U) \quad \theta(\cdot) + \frac{1}{\gamma} D_f(\cdot, y) \text{ is coercive} \tag{19}$$

*or, equivalently,*

$$\frac{1}{\gamma} \operatorname{ran} \nabla f \subseteq \operatorname{int} \operatorname{dom} \left( \frac{1}{\gamma} f + \theta \right)^* . \tag{20}$$

(ii) *If any of the conditions (a), (b), or (d) holds, then*

$$(\forall x \in U) \quad \theta(\cdot) + \frac{1}{\gamma} D_f(x, \cdot) \text{ is coercive.} \tag{21}$$

*Proof.* Since $\frac{1}{\gamma} D_f = D_{\frac{1}{\gamma} f}$, the result follows from [9, Lemma 2.12] applied to $\frac{1}{\gamma} f$. ■

The definition of proximal mappings relies on the following result. (For variants of Proposition 2.7(i), see [25, Theorem 2.2 and Theorem 4.3].)

**Proposition 2.7.** *Let $\theta\colon X \to \left]-\infty, +\infty\right]$ be convex and such that $U \cap \operatorname{dom} \theta \neq \varnothing$, and let $\gamma \in \mathbb{R}_{++}$. Then the following hold:*

(i) $\overleftarrow{\operatorname{env}}_\theta^\gamma$ *is convex and continuous on $U$, and*
  (a) *If* (19) *holds, i.e., $(\forall y \in U)$ $\theta(\cdot) + \frac{1}{\gamma} D_f(\cdot, y)$ is coercive, then $\overleftarrow{\operatorname{env}}_\theta^\gamma$ is proper;*
  (b) *If $\theta \in \Gamma_0(X)$ and $\theta(\cdot) + \frac{1}{\gamma} D_f(\cdot, y)$ is coercive for a given $y \in U$, then there exists a unique point $z \in U$ such that $\overleftarrow{\operatorname{env}}_\theta^\gamma(y) = \theta(z) + \frac{1}{\gamma} D_f(z, y)$.*

(ii) $\overrightarrow{\operatorname{env}}_\theta^\gamma$ *is convex and continuous on $U$, and*
  (a) *If* (21) *holds, i.e., $(\forall x \in U)$ $\theta(\cdot) + \frac{1}{\gamma} D_f(x, \cdot)$ is coercive, then $\overrightarrow{\operatorname{env}}_\theta^\gamma$ is proper;*
  (b) *If $\theta \in \Gamma_0(X)$ and $\theta(\cdot) + \frac{1}{\gamma} D_f(x, \cdot)$ is coercive for a given $x \in U$, then there exists a unique point $z \in U$ such that $\overrightarrow{\operatorname{env}}_\theta^\gamma(z) = \theta(z) + \frac{1}{\gamma} D_f(x, z)$.*

*Proof.* Since $\frac{1}{\gamma} D_f = D_{\frac{1}{\gamma} f}$, the result follows from [9, Propositions 3.4&3.5] applied to $\frac{1}{\gamma} f$. ■

In view of Proposition 2.7, we define the following operators on $U$; see also [9, Definition 3.7].

**Definition 2.8 (Bregman proximity operators).** Let $\theta \in \Gamma_0(X)$ be such that $U \cap \operatorname{dom} \theta \neq \varnothing$. If (19) holds for $\gamma = 1$, then the *left proximity operator* associated with $\theta$ is

$$\overleftarrow{\mathrm{P}}_\theta\colon U \to U\colon y \mapsto \underset{x \in X}{\operatorname{argmin}}\ \theta(x) + D_f(x, y). \tag{22}$$

If (21) holds for $\gamma = 1$, then the *right proximity operator* associated with $\theta$ is

$$\overrightarrow{\mathrm{P}}_\theta\colon U \to U\colon x \mapsto \underset{y \in X}{\operatorname{argmin}}\ \theta(y) + D_f(x, y). \tag{23}$$

**Remark 2.9.** Suppose that $f = \frac{1}{2}\|\cdot\|^2$ and let $\theta \in \Gamma_0(X)$. Then $U = \operatorname{int} \operatorname{dom} f = X$ and hence $U \cap \operatorname{dom} \theta = \operatorname{dom} \theta \neq \varnothing$. Since $f(x)/\|x\| = \frac{1}{2}\|x\| \to +\infty$ as $\|x\| \to +\infty$, Fact 2.6 implies that (19) and (21) hold for all $\gamma \in \mathbb{R}_{++}$. In this case, $D_f\colon (x, y) \mapsto \frac{1}{2}\|x - y\|^2$ and $\overleftarrow{\mathrm{P}}_\theta = \overrightarrow{\mathrm{P}}_\theta = \operatorname{Prox}_\theta$ is the classical *Moreau proximity operator* of $\theta$ [30].

Given a closed convex subset $C$ of $X$ with $C \cap U \neq \varnothing$, we have that $\iota_C \in \Gamma_0(X)$, $\operatorname{dom} \iota_C = C$, hence $U \cap \operatorname{dom} \iota_C = U \cap C \neq \varnothing$ and also $\inf \iota_C(U) = 0 > -\infty$, which together with Fact 2.6 imply that (19) and (21) hold for all $\gamma \in \mathbb{R}_{++}$. This leads to the following definition.

**Definition 2.10 (Bregman projectors).** Let $C$ be a closed convex subset of $X$ such that $U \cap C \neq \varnothing$. Then $\overleftarrow{\mathrm{P}}_C := \overleftarrow{\mathrm{P}}_{\iota_C}$ is the *left Bregman projector* onto $C$ and $\overrightarrow{\mathrm{P}}_C := \overrightarrow{\mathrm{P}}_{\iota_C}$ is the *right Bregman projector* onto $C$.

**Remark 2.11.** In view of Remark 2.9, if $f = \frac{1}{2}\|\cdot\|^2$, then $\overleftarrow{\mathrm{P}}_C = \overrightarrow{\mathrm{P}}_C = \mathrm{P}_C$ is the *orthogonal projector* onto $C$. Note that $\overleftarrow{\mathrm{P}}_C$, $\overrightarrow{\mathrm{P}}_C$, and $\mathrm{P}_C$ are not, in general, the same when $f \neq \frac{1}{2}\|\cdot\|^2$. Before we give a corresponding example, let us show that these projectors are the same when $X = \mathbb{R}$.

**Proposition 2.12.** *Suppose that $X = \mathbb{R}$ and let $C$ be a closed convex subset of $\mathbb{R}$ such that $U \cap C \neq \varnothing$. Then $\overleftarrow{\mathrm{P}}_C = \overrightarrow{\mathrm{P}}_C = \mathrm{P}_C$ on $U$.*

*Proof.* Let $y \in U$. Because $X = \mathbb{R}$, $(\forall z \in C)$ $(\exists\, \lambda_z \in [0,1])$ $\mathrm{P}_C y = \lambda_z z + (1-\lambda_z)y$. Since $D_f(\cdot, y)$ is convex, nonnegative, and $D_f(y,y) = 0$, it follows that

$$(\forall z \in C) \quad D_f(\mathrm{P}_C y, y) \leq \lambda_z D_f(z,y) + (1-\lambda_z) D_f(y,y) = \lambda_z D_f(z,y) \leq D_f(z,y). \tag{24}$$

This combined with Definition 2.10 yields $\overleftarrow{\mathrm{P}}_C(y) = \mathrm{P}_C(y)$. The proof that $\overrightarrow{\mathrm{P}}_C = \mathrm{P}_C$ is similar. ■

**Example 2.13.** We illustrate how Bregman projectors may differ from the orthogonal projector. We adapt [5, Example 6.15] which illustrates in the setting where $f$ is an entropy function on $\mathbb{R}^J$ and $C$ is the "probabilistic hyperplane" $\{x \in \mathbb{R}^J \mid \sum_j \xi_j = 1\}$. For simplicity, we work in $X = \mathbb{R}^2$. Suppose that $f_1$ is the energy from Example 2.3(i) while $f_2$ is the negative Boltzmann–Shannon entropy from Example 2.3(ii). Since we work in $\mathbb{R}^2$, the probabilistic hyperplane is described by $\xi_2 = 1 - \xi_1$. We compute $\overleftarrow{\mathrm{P}}_C(1,0)$ by substituting $\eta_1 = 1, \eta_2 = 0, \xi_2 = 1 - \xi_1$ and minimizing the resulting Bregman distance over $\xi_1$. We obtain

(i) $\overleftarrow{\mathrm{P}}_C(1,2) = (0,1)$ for $D_{f_1}$;
(ii) $\overleftarrow{\mathrm{P}}_C(1,2) = (1/3, 2/3)$ for $D_{f_2}$.

We illustrate in Figure 1. For $i \in \{1,2\}$, we sketch the contour plot of $D_{f_i}(\cdot, (1,2))$ for the level given by $D_{f_i}(\overleftarrow{\mathrm{P}}_C(1,2), (1,2))$ together with the set $C$.

**Remark 2.14.** Let $\theta \in \Gamma_0(X)$ be such that $U \cap \operatorname{dom} \theta \neq \varnothing$ and let $\gamma \in \mathbb{R}_{++}$. Proposition 2.1 implies that $\overleftarrow{\mathrm{env}}_{\gamma\theta} = \gamma\, \overleftarrow{\mathrm{env}}_\theta^\gamma$ and $\overrightarrow{\mathrm{env}}_{\gamma\theta} = \gamma\, \overrightarrow{\mathrm{env}}_\theta^\gamma$. We thus derive from the definition that if (19) holds, then

$$\overleftarrow{\mathrm{env}}_\theta^\gamma(y) = \theta\big(\overleftarrow{\mathrm{P}}_{\gamma\theta}(y)\big) + \frac{1}{\gamma} D_f\big(\overleftarrow{\mathrm{P}}_{\gamma\theta}(y), y\big) \tag{25a}$$

and, by combining with Proposition 2.2(i),

$$\theta\big(\overleftarrow{\mathrm{P}}_{\gamma\theta}(y)\big) \leq \overleftarrow{\mathrm{env}}_\theta^\gamma(y) \leq \theta(y). \tag{25b}$$

Similarly, if (21) holds, then

$$\overrightarrow{\mathrm{env}}_\theta^\gamma(x) = \theta\big(\overrightarrow{\mathrm{P}}_{\gamma\theta}(x)\big) + \frac{1}{\gamma} D_f\big(x, \overrightarrow{\mathrm{P}}_{\gamma\theta}(x)\big) \tag{26a}$$

and

$$\theta\big(\overrightarrow{\mathrm{P}}_{\gamma\theta}(x)\big) \leq \overrightarrow{\mathrm{env}}_\theta^\gamma(x) \leq \theta(x). \tag{26b}$$

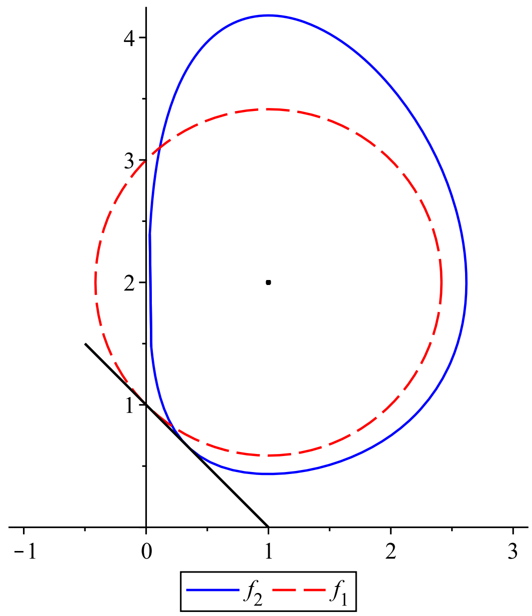


Figure 1: Example 2.13 is illustrated

The next result provides information on the proximal mapping when the parameter is varied. For a variant of the last inequality in (27), see [25, Proposition 2.1(ii)].

**Proposition 2.15.** *Let $\theta \in \Gamma_0(X)$ be such that $U \cap \operatorname{dom}\theta \neq \varnothing$ and let $\gamma \in \mathbb{R}_{++}$.*

(i) *If* (19) *holds, then* $(\forall y \in U)(\forall \mu \in \left]\gamma, +\infty\right[)$

$$\theta\big(\overleftarrow{\mathrm{P}}_{\mu\theta}(y)\big) \leq \theta\big(\overleftarrow{\mathrm{P}}_{\gamma\theta}(y)\big) \quad \textit{and} \quad D_f\big(\overleftarrow{\mathrm{P}}_{\mu\theta}(y), y\big) \geq D_f\big(\overleftarrow{\mathrm{P}}_{\gamma\theta}(y), y\big). \tag{27}$$

(ii) *If* (21) *holds, then* $(\forall x \in U)(\forall \mu \in \left]\gamma, +\infty\right[)$

$$\theta\big(\overrightarrow{\mathrm{P}}_{\mu\theta}(x)\big) \leq \theta\big(\overleftarrow{\mathrm{P}}_{\gamma\theta}(x)\big) \quad \textit{and} \quad D_f\big(\overleftarrow{\mathrm{P}}_{\mu\theta}(x), x\big) \geq D_f\big(\overleftarrow{\mathrm{P}}_{\gamma\theta}(x), x\big). \tag{28}$$

*Proof.* This follows from Remark 2.14 and [28, Proposition 7.6.1]. ■

The left and right proximal mappings can be characterized in various ways:

**Proposition 2.16.** *Let $\theta \in \Gamma_0(X)$ be such that $\operatorname{dom}\theta \cap U \neq \varnothing$ and let $\gamma \in \mathbb{R}_{++}$.*

(i) *Suppose that* (19) *holds. Then for every $(x, y) \in U \times U$, the following conditions are equivalent:*
   (a) $x = \overleftarrow{\mathrm{P}}_{\gamma\theta}(y)$;
   (b) $0 \in \gamma\partial\theta(x) + \nabla f(x) - \nabla f(y)$;
   (c) $(\forall z \in X) \quad \langle \nabla f(y) - \nabla f(x), z - x\rangle + \gamma\theta(x) \leq \gamma\theta(z)$.

   *Moreover,*

$$\overleftarrow{\mathrm{P}}_{\gamma\theta} = (\nabla f + \gamma\partial\theta)^{-1} \circ \nabla f \tag{29}$$

   *is continuous on $U$.*

(ii) *Suppose that* (21) *holds. Then for every $(x, y) \in U \times U$, the following conditions are equivalent:*

(a) $y = \overrightarrow{\mathrm{P}}_{\gamma\theta}(x)$;
(b) $0 \in \gamma\partial\theta(y) + \nabla^2 f(y)(y-x)$;
(c) $(\forall z \in X) \quad \big\langle \nabla^2 f(y)(x-y), z-y \big\rangle + \gamma\theta(y) \leq \gamma\theta(z)$.

*Moreover,* $\overrightarrow{\mathrm{P}}_{\gamma\theta}$ *is continuous on* $U$.

*Proof.* Apply [9, Proposition 3.10] to $\gamma\theta$. ■

**Remark 2.17.** Consider Proposition 2.16 and its notation.

(i) In the case of item (i) and when $U^* = X$, we note that, by (13),

$$\overleftarrow{\mathrm{P}}_{\gamma\theta} \circ \nabla f^* = \big(\nabla f + \gamma\partial\theta\big)^{-1} = \partial\big(f + \gamma\theta\big)^* \quad \text{is maximally (cyclically) monotone;} \tag{30}$$

see also [25, Theorem 4.2] for a more general result.

(ii) In the case of item (ii), let us prove the variant of [25, Theorem 4.1] stating that

$$\nabla f \circ \overrightarrow{\mathrm{P}}_{\gamma\theta} \quad \text{is monotone.} \tag{31}$$

Indeed, let $x_1$ and $x_2$ be in $U$, and set $y_i = \overrightarrow{\mathrm{P}}_{\gamma\theta}(x_i)$ for $i \in \{1,2\}$. Then $\theta(y_1) + \frac{1}{\gamma}D_f(x_1,y_1) \leq \theta(y_2) + \frac{1}{\gamma}D_f(x_1,y_2)$ and $\theta(y_2) + \frac{1}{\gamma}D_f(x_2,y_2) \leq \theta(y_1) + \frac{1}{\gamma}D_f(x_2,y_1)$ Adding and simplifying yields

$$0 \leq D_f(x_1,y_2) + D_f(x_2,y_1) - D_f(x_1,y_1) - D_f(x_2,y_2). \tag{32}$$

A direct expansion (or the *four-point identity* from [10, Remark 2.5]) shows that (32) is the same as

$$0 \leq \langle \nabla f(y_1) - \nabla f(y_2), x_1 - x_2 \rangle\,; \tag{33}$$

therefore, (31) follows. We do not know whether or not in general the operator in (31) is the gradient of a convex function.

**Corollary 2.18.** *Let* $C$ *be a closed convex subset of* $X$ *such that* $U \cap C \neq \varnothing$, *let* $(x,y) \in U \times U$, *and let* $p \in U \cap C$. *Then the following hold:*

(i) $p = \overleftarrow{\mathrm{P}}_C y \;\Leftrightarrow\; (\forall z \in C)\ \langle \nabla f(y) - \nabla f(p), z-p \rangle \leq 0$.
(ii) $p = \overrightarrow{\mathrm{P}}_C x \;\Leftrightarrow\; (\forall z \in C)\ \big\langle \nabla^2 f(p)(x-p), z-p \big\rangle \leq 0$.

*Proof.* In light of Definition 2.10, we apply Proposition 2.16 (see also [5, Proposition 3.16]). ■

The derivatives of the left and right Bregman–Moreau envelopes feature the corresponding proximal mappings as follows:

**Proposition 2.19.** *Let* $\theta \in \Gamma_0(X)$ *be such that* $U \cap \operatorname{dom}\theta \neq \varnothing$ *and let* $\gamma \in \mathbb{R}_{++}$. *Then the following hold:*

(i) *If* (19) *holds, then* $\overleftarrow{\mathrm{env}}^{\gamma}_{\theta}$ *is differentiable on* $U$ *and*

$$(\forall y \in U) \quad \nabla\, \overleftarrow{\mathrm{env}}^{\gamma}_{\theta}(y) = \frac{1}{\gamma}\nabla^2 f(y)(y - \overleftarrow{\mathrm{P}}_{\gamma\theta}(y)). \tag{34}$$

(ii) *If* (21) *holds, then* $\overrightarrow{\operatorname{env}}_\theta^\gamma$ *is differentiable on* $U$ *and*

$$(\forall x \in U) \quad \nabla \overrightarrow{\operatorname{env}}_\theta^\gamma(x) = \frac{1}{\gamma}\nabla f(x) - \frac{1}{\gamma}\nabla f(\overrightarrow{\mathrm{P}}_{\gamma\theta}(x)). \tag{35}$$

*Proof.* Combine Remark 2.14 with [9, Proposition 3.12]. ■

The following result, which is a variant of [24, Theorem XV.4.1.7], highlights the connection to convex optimization:

**Theorem 2.20.** *Let* $\theta \in \Gamma_0(X)$ *be such that* $U \cap \operatorname{dom}\theta \neq \varnothing$, *let* $\gamma \in \mathbb{R}_{++}$, *let* $x \in U$, *and let* $y \in U$.

(i) *Suppose that* (19) *holds. Then the following are equivalent:*
- (a) $y \in \operatorname{argmin}\theta$;
- (b) $y \in \operatorname{Fix} \overleftarrow{\mathrm{P}}_{\gamma\theta}$;
- (c) $y \in \operatorname{argmin} \overleftarrow{\operatorname{env}}_\theta^\gamma$;
- (d) $\theta(\overleftarrow{\mathrm{P}}_{\gamma\theta}(y)) = \theta(y)$;
- (e) $\overleftarrow{\operatorname{env}}_\theta^\gamma(y) = \theta(y)$.

*Consequently,*

$$U \cap \operatorname{argmin}\theta = \operatorname{Fix} \overleftarrow{\mathrm{P}}_{\gamma\theta} = \operatorname{argmin} \overleftarrow{\operatorname{env}}_\theta^\gamma. \tag{36}$$

(ii) *Suppose that* (21) *holds. Then the following are equivalent:*
- (a) $x \in \operatorname{argmin}\theta$;
- (b) $x \in \operatorname{Fix} \overrightarrow{\mathrm{P}}_{\gamma\theta}$;
- (c) $x \in \operatorname{argmin} \overrightarrow{\operatorname{env}}_\theta^\gamma$;
- (d) $\theta(\overrightarrow{\mathrm{P}}_{\gamma\theta}(x)) = \theta(x)$;
- (e) $\overrightarrow{\operatorname{env}}_\theta^\gamma(x) = \theta(x)$.

*Consequently,*

$$U \cap \operatorname{argmin}\theta = \operatorname{Fix} \overrightarrow{\mathrm{P}}_{\gamma\theta} = U \cap \operatorname{argmin} \overrightarrow{\operatorname{env}}_\theta^\gamma. \tag{37}$$

*Proof.* (i): Using Proposition 2.16(i), we have

$$y \in \operatorname{argmin}\theta \Leftrightarrow\ 0 \in \partial\theta(y) \Leftrightarrow\ 0 \in \gamma\partial\theta(y) + \nabla f(y) - \nabla f(y) \tag{38a}$$

$$\Leftrightarrow\ y = \overleftarrow{\mathrm{P}}_{\gamma\theta}(y) \Leftrightarrow\ y \in \operatorname{Fix} \overleftarrow{\mathrm{P}}_{\gamma\theta}. \tag{38b}$$

This proves that (i)(a) ⇔ (i)(b).

Assume that (i)(b) holds, i.e., $y = \overleftarrow{\mathrm{P}}_{\gamma\theta}(y)$. Then $\nabla \overleftarrow{\operatorname{env}}_\theta^\gamma(y) = 0$ by Proposition 2.19(i), and thus (i)(c) holds by the convexity of $\overleftarrow{\operatorname{env}}_\theta^\gamma$ shown in Proposition 2.7(i). Next, (i)(d) is obvious and (i)(e) holds due to (25a).

Now recall from (25) that

$$\theta(\overleftarrow{\mathrm{P}}_{\gamma\theta}(y)) \leq \theta(\overleftarrow{\mathrm{P}}_{\gamma\theta}(y)) + \frac{1}{\gamma}D_f(\overleftarrow{\mathrm{P}}_{\gamma\theta}(y), y) = \overleftarrow{\operatorname{env}}_\theta^\gamma(y) \leq \theta(y). \tag{39}$$

If (i)(c) holds, then since $\inf\theta(X) = \inf \overleftarrow{\operatorname{env}}_\theta^\gamma(X)$ (see Proposition 2.2(i)), combining with (39) yields

$$\inf\theta(X) \leq \theta(\overleftarrow{\mathrm{P}}_{\gamma\theta}(y)) \leq \overleftarrow{\operatorname{env}}_\theta^\gamma(y) = \min \overleftarrow{\operatorname{env}}_\theta^\gamma(X) = \inf\theta(X), \tag{40}$$

which implies that $D_f(\overleftarrow{\mathrm{P}}_{\gamma\theta}(y), y) = 0$, so $y = \overleftarrow{\mathrm{P}}_{\gamma\theta}(y)$ due to Fact 2.5(i), and we get (i)(b).

If (i)(d) holds, then by (39), $D_f(\overleftarrow{\mathrm{P}}_{\gamma\theta}(y), y) = 0$, and (i)(b) thus holds.

Finally, if (i)(e) holds, then $\overleftarrow{\mathrm{env}}_\theta^\gamma(y) = \theta(y) + \frac{1}{\gamma} D_f(y, y)$, and using Proposition 2.7(i) and (25a), we must have $y = \overleftarrow{\mathrm{P}}_{\gamma\theta}(y)$ and therefore get (i)(b).

(ii): This is proved similarly to (i) by using Proposition 2.2(ii), Proposition 2.7(ii), Proposition 2.16(ii), Proposition 2.19(ii), and (26a). The difference between (36) and (37) is because $\operatorname{dom} \overleftarrow{\mathrm{env}}_\theta^\gamma = U$ while $\operatorname{dom} \overrightarrow{\mathrm{env}}_\theta^\gamma = \operatorname{dom} f$. ■

# 3 Bregman–Moreau envelopes: behaviour when $\gamma \downarrow 0$ or $\gamma \uparrow +\infty$

The results in this section, almost all of which are new, extend or complement results for the classical energy case and for left variants studied in [21] and [25]. We will require the following lemma.

**Lemma 3.1.** *Let $C$ be a compact subset of a Hausdorff space $\mathcal{X}$, let $\phi\colon \mathcal{X} \to [-\infty, +\infty]$ be lower semicontinuous, let $(x_a)_{a\in A}$ be a net in $C$, and suppose that $\phi(x_a) \to \inf \phi(\mathcal{X})$. Then $\operatorname{argmin} \phi \neq \varnothing$ and all cluster points of $(x_a)_{a\in A}$ lie in $\operatorname{argmin} \phi$. Consequently, if $\phi$ attains its minimum at a unique point $u$, then $x_a \to u$.*

*Proof.* Let $x$ be an arbitrary cluster point of $(x_a)_{a\in A}$. Then there exists a subnet $(x_{k(a)})_{a\in A}$ of $(x_a)_{a\in A}$ such that $\lim x_{k(a)} = x$. By assumption, $\phi(x_{k(a)}) \to \inf \phi(\mathcal{X})$, which combined with the lower semicontinuity of $\phi$ implies

$$\inf \phi(\mathcal{X}) \le \phi(x) \le \lim \phi(x_{k(a)}) = \inf \phi(\mathcal{X}), \tag{41}$$

and therefore, $\phi(x) = \inf \phi(\mathcal{X})$. We deduce that $x \in \operatorname{argmin} \phi$ and $\operatorname{argmin} \phi \neq \varnothing$. Now if $\phi$ attains its minimum at a unique point $u$, then $(x_a)_{a\in A}$ admits a unique cluster point $u$ and the conclusion follows from [8, Lemma 1.14]. ■

What happens when $\gamma \downarrow 0$? The next two results provide answers.

**Proposition 3.2.** *Let $\theta \in \Gamma_0(X)$ be such that $U \cap \operatorname{dom} \theta \neq \varnothing$, let $x \in U$, and let $y \in U$. Then the following hold:*

(i) *If* (19) *holds for some $\mu \in \mathbb{R}_{++}$, then $\overleftarrow{\mathrm{P}}_{\gamma\theta}(y) \to y$ as $\gamma \downarrow 0$.*
(ii) *If* (21) *holds for some $\mu \in \mathbb{R}_{++}$, then $\overrightarrow{\mathrm{P}}_{\gamma\theta}(x) \to x$ as $\gamma \downarrow 0$.*

*Proof.* (i): Noting that $(\forall \gamma \in \left]0, \mu\right])\ \theta + \frac{1}{\gamma} D_f(\cdot, y) \ge \theta + \frac{1}{\mu} D_f(\cdot, y)$, we have that (19) holds for all $\gamma \in \left]0, \mu\right]$. In particular, $g := \theta + \frac{1}{\mu} D_f(\cdot, y)$ is coercive. By Proposition 2.2(i) and (25a),

$$(\forall \gamma \in \left]0, \mu\right])\quad \theta(y) \ge \overleftarrow{\mathrm{env}}_\theta^\gamma(y) = \theta\big(\overleftarrow{\mathrm{P}}_{\gamma\theta}(y)\big) + \frac{1}{\gamma} D_f\big(\overleftarrow{\mathrm{P}}_{\gamma\theta}(y), y\big) \ge g\big(\overleftarrow{\mathrm{P}}_{\gamma\theta}(y)\big) \tag{42}$$

and so $\overleftarrow{\mathrm{P}}_{\gamma\theta}(y) \in \operatorname{lev}_{\le \theta(y)} g$. The coercivity of $g$ and [8, Proposition 11.12] imply that $\nu := \sup_{\gamma\in]0,\mu]} \|\overleftarrow{\mathrm{P}}_{\gamma\theta}(y)\| < +\infty$. Now by [8, Theorem 9.20], there exist $u \in X$ and $\eta \in \mathbb{R}$ such that $\theta \ge \langle \cdot, u\rangle + \eta$. Using (42) and Cauchy–Schwarz yields

$$(\forall \gamma \in \left]0, \mu\right])\quad \theta(y) \ge \theta\big(\overleftarrow{\mathrm{P}}_{\gamma\theta}(y)\big) + \frac{1}{\gamma} D_f\big(\overleftarrow{\mathrm{P}}_{\gamma\theta}(y), y\big) \tag{43a}$$

$$\geq \left\langle \overleftarrow{\mathrm{P}}_{\gamma\theta}(y), u \right\rangle + \eta + \frac{1}{\gamma} D_f\big(\overleftarrow{\mathrm{P}}_{\gamma\theta}(y), y\big) \tag{43b}$$

$$\geq -\nu\|u\| + \eta + \frac{1}{\gamma} D_f\big(\overleftarrow{\mathrm{P}}_{\gamma\theta}(y), y\big), \tag{43c}$$

which gives

$$0 \leq D_f\big(\overleftarrow{\mathrm{P}}_{\gamma\theta}(y), y\big) \leq \gamma\big(\theta(y) + \nu\|u\| - \eta\big) \to 0 \quad \text{as} \quad \gamma \downarrow 0, \tag{44}$$

and thus $D_f(\overleftarrow{\mathrm{P}}_{\gamma\theta}(y), y) \to 0$ as $\gamma \downarrow 0$. Observe that $D_f(\cdot, y) = f(\cdot) - f(y) - \langle \nabla f(y), \cdot - y\rangle$ is lower semicontinuous, that $\operatorname{argmin} D_f(\cdot, y) = \{y\}$ by Fact 2.5(i), and that $\sup_{\gamma\in]0,1[} \|\overleftarrow{\mathrm{P}}_{\gamma\theta}(y)\| < +\infty$, it follows from Lemma 3.1 that $\overleftarrow{\mathrm{P}}_{\gamma\theta}(y) \to y$ as $\gamma \downarrow 0$.

(ii): Similar to (i). ■

**Theorem 3.3.** *Let $\theta \in \Gamma_0(X)$ be such that $U \cap \operatorname{dom}\theta \neq \varnothing$, let $x \in U$, and let $y \in U$. Then the following hold:*

(i) *If* (19) *holds for some $\mu \in \mathbb{R}_{++}$ and $\gamma \downarrow 0$, then $\overleftarrow{\operatorname{env}}_\theta^\gamma(y) \uparrow \theta(y)$, $\theta(\overleftarrow{\mathrm{P}}_{\gamma\theta}(y)) \uparrow \theta(y)$, and $\frac{1}{\gamma} D_f(\overleftarrow{\mathrm{P}}_{\gamma\theta}(y), y) \to 0$.*

(ii) *If* (21) *holds for some $\mu \in \mathbb{R}_{++}$, and $\gamma \downarrow 0$, then $\overrightarrow{\operatorname{env}}_\theta^\gamma(x) \uparrow \theta(x)$, $\theta(\overrightarrow{\mathrm{P}}_{\gamma\theta}(x)) \uparrow \theta(x)$, and $\frac{1}{\gamma} D_f(x, \overrightarrow{\mathrm{P}}_{\gamma\theta}(x)) \to 0$.*

*Proof.* (i): According to Proposition 2.2(i), there exists $\beta \in \mathbb{R}$ such that $\overleftarrow{\operatorname{env}}_\theta^\gamma(y) \uparrow \beta \leq \theta(y)$ as $\gamma \downarrow 0$. Combining with (25a), we have

$$(\forall \gamma \in ]0, \mu]) \quad \theta(y) \geq \beta \geq \overleftarrow{\operatorname{env}}_\theta^\gamma(y) = \theta\big(\overleftarrow{\mathrm{P}}_{\gamma\theta}(y)\big) + \frac{1}{\gamma} D_f\big(\overleftarrow{\mathrm{P}}_{\gamma\theta}(y), y\big) \geq \theta\big(\overleftarrow{\mathrm{P}}_{\gamma\theta}(y)\big). \tag{45}$$

This together with the fact that $\lim_{\gamma\downarrow 0} \overleftarrow{\mathrm{P}}_{\gamma\theta}(y) = y$ by Proposition 3.2(i), and the lower semicontinuity of $\theta$ implies

$$\theta(y) \geq \beta \geq \varliminf_{\gamma\downarrow 0} \theta\big(\overleftarrow{\mathrm{P}}_{\gamma\theta}(y)\big) \geq \theta(y) \geq \varlimsup_{\gamma\downarrow 0} \theta\big(\overleftarrow{\mathrm{P}}_{\gamma\theta}(y)\big), \tag{46}$$

and then $\beta = \theta(y) = \lim_{\gamma\downarrow 0} \theta(\overleftarrow{\mathrm{P}}_{\gamma\theta}(y))$. Now recall (45) and Proposition 2.15(i).

(ii): Similar to (i). ■

For a variant of the result from Theorem 3.3(i) that $\overleftarrow{\operatorname{env}}_\theta^\gamma(y) \uparrow \theta(y)$ as $\gamma \downarrow 0$, see [25, Theorem 2.5]. Note that $D_f(\overleftarrow{\mathrm{P}}_{\gamma\theta}(y), y)$ is monotone with respect to $\gamma$, as shown in Proposition 2.15(i), but the same is not necessarily true for $\frac{1}{\gamma} D_f(\overleftarrow{\mathrm{P}}_{\gamma\theta}(y), y)$ (see Figure 2).

The two following results describe the behaviour when $\gamma \uparrow +\infty$.

**Proposition 3.4.** *Let $\theta \in \Gamma_0(X)$ be such that $U \cap \operatorname{dom}\theta \neq \varnothing$, let $x \in U$, and let $y \in U$. Then the following hold:*

(i) *If* (19) *holds for all $\gamma \in \mathbb{R}_{++}$, then $\theta(\overleftarrow{\mathrm{P}}_{\gamma\theta}(y)) \to \inf \theta(X)$ as $\gamma \uparrow +\infty$.*

(ii) *If* (21) *holds for all $\gamma \in \mathbb{R}_{++}$, then $\theta(\overrightarrow{\mathrm{P}}_{\gamma\theta}(x)) \to \inf \theta(X)$ as $\gamma \uparrow +\infty$.*

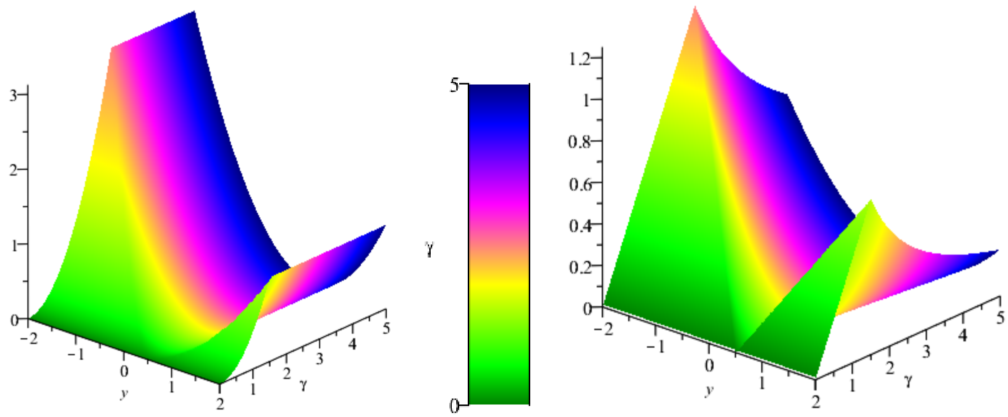


Figure 2: $D_f(\overleftarrow{\mathrm{P}}_{\gamma\theta}(y), y)$ (left) and $\frac{1}{\gamma}D_f(\overleftarrow{\mathrm{P}}_{\gamma\theta}(y), y)$ (right) when $X = \mathbb{R}$, $f$ is the energy, and $\theta$ is the function $x \mapsto |x - \frac{1}{2}|$

*Proof.* We shall just prove (i) because the proof of (ii) is similar. Assume that (19) holds for all $\gamma \in \mathbb{R}_{++}$. Combining (25b) with Proposition 2.2(i) yields

$$\inf \theta(X) \leq \theta(\overleftarrow{\mathrm{P}}_{\gamma\theta}(y)) \leq \overleftarrow{\mathrm{env}}_\theta^\gamma(y) \to \inf \theta(X) \quad \text{as} \quad \gamma \uparrow +\infty, \tag{47}$$

which implies that $\theta(\overleftarrow{\mathrm{P}}_{\gamma\theta}(y)) \to \inf \theta(X)$ as $\gamma \uparrow +\infty$. ■

**Theorem 3.5.** *Let $\theta \in \Gamma_0(X)$ be coercive such that $U \cap \operatorname{dom} \theta \neq \varnothing$, let $x \in U$, and let $y \in U$. Then the following hold:*

(i) *The net $(\overleftarrow{\mathrm{P}}_{\gamma\theta}(y))_{\gamma\in\mathbb{R}_{++}}$ is bounded with all cluster points as $\gamma \uparrow +\infty$ lying in* $\operatorname{argmin} \theta$. *Moreover:*
   - (a) *If* $\operatorname{argmin}\theta$ *is a singleton, then* $\overleftarrow{\mathrm{P}}_{\gamma\theta}(y) \to \operatorname{argmin}\theta$ *as* $\gamma \uparrow +\infty$;
   - (b) *If* $\operatorname{argmin}\theta \subseteq U$, *then* $\overleftarrow{\mathrm{P}}_{\gamma\theta}(y) \to \overleftarrow{\mathrm{P}}_{\operatorname{argmin}\theta} y$ *as* $\gamma \uparrow +\infty$.

(ii) *The net $(\overrightarrow{\mathrm{P}}_{\gamma\theta}(x))_{\gamma\in\mathbb{R}_{++}}$ is bounded with all cluster points as $\gamma \uparrow +\infty$ lying in* $\operatorname{argmin} \theta$. *Moreover:*
   - (a) *If* $\operatorname{argmin}\theta$ *is a singleton, then* $\overrightarrow{\mathrm{P}}_{\gamma\theta}(x) \to \operatorname{argmin}\theta$ *as* $\gamma \uparrow +\infty$;
   - (b) *If* $\operatorname{argmin}\theta \subseteq U$, *then* $\overrightarrow{\mathrm{P}}_{\gamma\theta}(x) \to \overrightarrow{\mathrm{P}}_{\operatorname{argmin}\theta} x$ *as* $\gamma \uparrow +\infty$.

*Proof.* First, by assumption, [8, Proposition 11.15(i)] gives $\operatorname{argmin}\theta \neq \varnothing$. This combined with [8, Lemma 1.24 and Corollary 8.5] implies that $\operatorname{argmin}\theta = \operatorname{lev}_{\leq \inf\theta(X)} \theta$ is a nonempty closed convex subset of $X$. Now since $\theta$ is coercive and since $D_f \geq 0$, we immediately get that (19) and (21) hold for all $\gamma \in \mathbb{R}_{++}$.

(i): It follows from (25b) that

$$(\forall \gamma \in \mathbb{R}_{++}) \quad \overleftarrow{\mathrm{P}}_{\gamma\theta}(y) \in \operatorname{lev}_{\leq \theta(y)} \theta, \tag{48}$$

and then from the coercivity of $\theta$ and [8, Proposition 11.12] that $(\overleftarrow{\mathrm{P}}_{\gamma\theta}(y))_{\gamma\in\mathbb{R}_{++}}$ is bounded. In turn, Proposition 3.4(i) and Lemma 3.1 imply that all cluster points of $(\overleftarrow{\mathrm{P}}_{\gamma\theta}(y))_{\gamma\in\mathbb{R}_{++}}$ as $\gamma \uparrow +\infty$ lie in $\operatorname{argmin}\theta$, and we get (i)(a).

Now assume that $\operatorname{argmin}\theta \subseteq U$. Let $y'$ be a cluster point of $(\overleftarrow{\mathrm{P}}_{\gamma\theta}(y))_{\gamma\in\mathbb{R}_{++}}$ as $\gamma \uparrow +\infty$. Then $y' \in \operatorname{argmin}\theta \subseteq U$ and there exists a sequence $(\gamma_n)_{n\in\mathbb{N}}$ in $\mathbb{R}_{++}$ such that $\gamma_n \uparrow +\infty$ and $\overleftarrow{\mathrm{P}}_{\gamma_n\theta}(y) \to y'$ as $n \to +\infty$. Let $z \in \operatorname{argmin}\theta$. We have $(\forall n \in \mathbb{N})\ \theta(z) \leq \theta(\overleftarrow{\mathrm{P}}_{\gamma_n\theta}(y))$, and by Proposition 2.16(i),

$$\Big\langle \nabla f(y) - \nabla f\big(\overleftarrow{\mathrm{P}}_{\gamma_n\theta}(y)\big), z - \overleftarrow{\mathrm{P}}_{\gamma_n\theta}(y)\Big\rangle \leq \gamma_n\Big(\theta(z) - \theta\big(\overleftarrow{\mathrm{P}}_{\gamma_n\theta}(y)\big)\Big) \leq 0. \tag{49}$$

Taking the limit as $n \to +\infty$ and using the continuity of $\nabla f$ yield

$$\big\langle \nabla f(y) - \nabla f(y'), z - y'\big\rangle \leq 0. \tag{50}$$

Since $z \in \operatorname{argmin}\theta$ was chosen arbitrarily and since $\operatorname{argmin}\theta$ is a closed convex subset of $X$ with $U \cap \operatorname{argmin}\theta \neq \varnothing$, in view of Corollary 2.18(i), $y' = \overleftarrow{\mathrm{P}}_{\operatorname{argmin}\theta}y$, and $\overleftarrow{\mathrm{P}}_{\operatorname{argmin}\theta}y$ is thus the only cluster point of $(\overleftarrow{\mathrm{P}}_{\gamma\theta}(y))_{\gamma\in\mathbb{R}_{++}}$ as $\gamma \uparrow +\infty$. Hence, (i)(b) holds.

(ii): The proof is similar to the one of (i). ■

**Remark 3.6.** Suppose that $f = \frac{1}{2}\|\cdot\|^2$ and let $\theta \in \Gamma_0(X)$ be coercive. By Remark 2.11 and Theorem 3.5,

$$(\forall x \in X) \quad \operatorname{Prox}_{\gamma\theta}(x) \to P_{\operatorname{argmin}\theta}x \quad \text{and} \quad \gamma \uparrow +\infty. \tag{51}$$

**Corollary 3.7.** *Let $\theta \in \Gamma_0(\mathbb{R})$ be coercive such that* $\operatorname{argmin}\theta \subseteq U$ *and let $z \in U$. Then $\overleftarrow{\mathrm{P}}_{\gamma\theta}(z) \to P_{\operatorname{argmin}\theta}z$ and $\overrightarrow{\mathrm{P}}_{\gamma\theta}(z) \to P_{\operatorname{argmin}\theta}z$ as $\gamma \uparrow +\infty$.*

*Proof.* As shown in the proof of Theorem 3.5, $\operatorname{argmin}\theta$ is a nonempty closed convex subset of $X$ and hence $U \cap \operatorname{argmin}\theta \neq \varnothing$. It now suffices to apply Theorem 3.5(i)(b)&(ii)(b) and to use Proposition 2.12. ■

## 4 Examples

In this final section, we illustrate our theory by considering the case when $\theta$ is the *nonsmooth* function $x \mapsto |x - \frac{1}{2}|$.

**Example 4.1.** Suppose that $X = \mathbb{R}$, and let $\theta\colon \mathbb{R} \to \mathbb{R}\colon x \mapsto |x - \frac{1}{2}|$. Then $\theta \in \Gamma_0(X)$, $\operatorname{dom}\theta = X$, and $\theta$ is coercive with $\operatorname{argmin}\theta = \{\frac{1}{2}\}$. It follows that $U \cap \operatorname{dom}\theta = U \neq \varnothing$ and, by Fact 2.6, the assumptions (19) and (21) hold for all $\gamma \in \mathbb{R}_{++}$. We revisit Example 2.3 (with $J = 1$) to illustrate Theorem 2.20, Proposition 3.2, and Theorem 3.5. Let $\gamma \in \mathbb{R}_{++}$. We recall from Proposition 2.16 that

$$\overleftarrow{\mathrm{P}}_{\gamma\theta} = (\nabla f + \gamma\partial\theta)^{-1} \circ \nabla f \tag{52}$$

and that

$$(\forall (x,y) \in U \times U) \quad y = \overrightarrow{\mathrm{P}}_{\gamma\theta}(x) \quad \Leftrightarrow \quad 0 \in \gamma\partial\theta(y) + \nabla^2 f(y)(y - x). \tag{53}$$

Note that

$$\partial\theta(x) = \begin{cases} 1, & \text{if } x > \frac{1}{2}; \\ -1, & \text{if } x < \frac{1}{2}; \\ [-1,1], & \text{if } x = \frac{1}{2}. \end{cases} \tag{54}$$

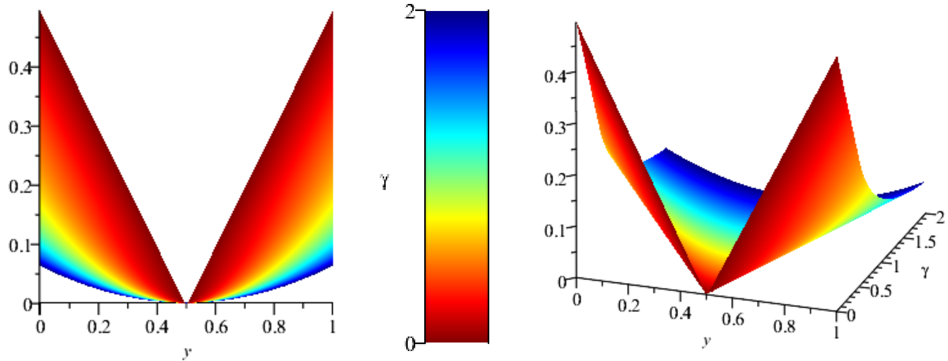


Figure 3: Bregman envelope from Example 4.1(i)

(i) *Energy:* Suppose that $f$ is the energy. Then $U = \operatorname{int} \operatorname{dom} f = \mathbb{R}$. Since $\nabla f = \mathrm{Id}$, by Remark 2.9 and (52), $\overleftarrow{\mathrm{P}}_{\gamma\theta} = \overrightarrow{\mathrm{P}}_{\gamma\theta} = (\mathrm{Id} + \gamma\partial\theta)^{-1}$. We have that

$$\big(\mathrm{Id} + \gamma\partial\theta\big)(x) = \begin{cases} x - \gamma, & \text{if } x < \frac{1}{2}; \\ x + \gamma, & \text{if } x > \frac{1}{2}; \\ \left[\frac{1}{2} - \gamma, \frac{1}{2} + \gamma\right], & \text{if } x = \frac{1}{2}. \end{cases} \tag{55}$$

Then $(\nabla f + \gamma\partial\theta)^{-1}(y)$ amounts to solving $(\nabla f + \gamma\partial\theta)(x) = y$ piecewise. For example, solving $x - \gamma = y$ for $x < \frac{1}{2}$ yields $x = y + \gamma$ for $y + \gamma < \frac{1}{2}$, so $(\nabla f + \gamma\theta)^{-1}(y) = y + \gamma$ for $y < \frac{1}{2} - \gamma$. Continuing in this fashion,

$$\overleftarrow{\mathrm{P}}_{\gamma\theta}(y) = \overrightarrow{\mathrm{P}}_{\gamma\theta}(y) = (\mathrm{Id} + \gamma\partial\theta)^{-1}(y) = \begin{cases} y + \gamma, & \text{if } y < \frac{1}{2} - \gamma; \\ y - \gamma, & \text{if } y > \frac{1}{2} + \gamma; \\ \frac{1}{2}, & \text{otherwise} \end{cases} \tag{56}$$

and by (25a),

$$\overleftarrow{\mathrm{env}}_{\theta}^{\gamma}(y) = \overrightarrow{\mathrm{env}}_{\theta}^{\gamma}(y) = \begin{cases} -y + \frac{1-\gamma}{2}, & \text{if } y < \frac{1}{2} - \gamma; \\ y - \frac{1-\gamma}{2}, & \text{if } y > \frac{1}{2} + \gamma; \\ \frac{4y^2 - 4y + 1}{8\gamma}, & \text{otherwise.} \end{cases} \tag{57}$$

It is clear that $\overleftarrow{\mathrm{P}}_{\gamma\theta}(\frac{1}{2}) = \frac{1}{2}$, while $(\forall y \in \mathbb{R} \smallsetminus \{\frac{1}{2}\})$ $\overleftarrow{\mathrm{P}}_{\gamma\theta}(y) \neq y$, and so $\operatorname{Fix} \overleftarrow{\mathrm{P}}_{\gamma\theta} = \{\frac{1}{2}\} = \operatorname{argmin} \theta$. As expected, $\overleftarrow{\mathrm{P}}_{\gamma\theta}(y) \to y$ as $\gamma \downarrow 0$, and $\overleftarrow{\mathrm{P}}_{\gamma\theta}(y) \to \frac{1}{2} = \operatorname{argmin} \theta$ as $\gamma \uparrow +\infty$. Moreover, $\overleftarrow{\mathrm{env}}_{\theta}^{\gamma}(y) \to \theta(y)$ as $\theta \downarrow 0$; this is illustrated in Figure 3.

(ii) *Boltzmann–Shannon entropy:* Suppose that $f$ is the Boltzmann–Shannon entropy. Then $\operatorname{dom} f = \mathbb{R}_+$, $U = \operatorname{int} \operatorname{dom} f = \mathbb{R}_{++}$, $\nabla f(x) = \ln x$, and $\nabla^2 f(x) = 1/x$. Again employ-

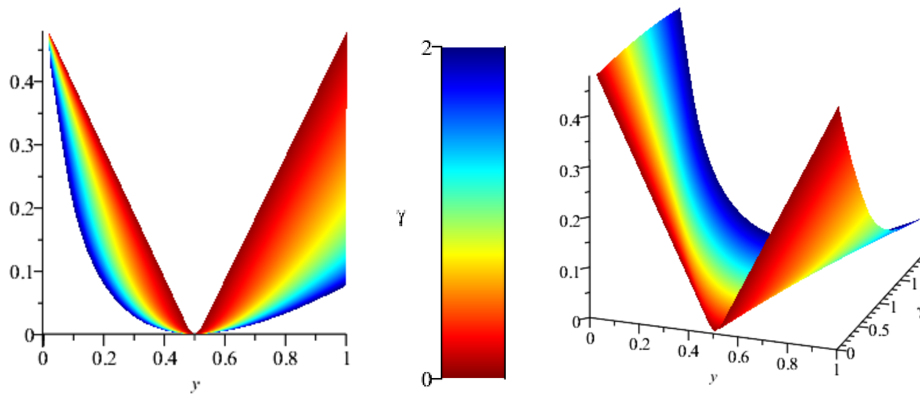


Figure 4: Left Bregman envelope from Example 4.1(ii)

ing (52) and (25a), we have

$$(\nabla f + \gamma\partial\theta)^{-1}(y) = \begin{cases} \exp(y+\gamma), & \text{if } y < -\ln 2 - \gamma; \\ \exp(y-\gamma), & \text{if } y > -\ln 2 + \gamma; \\ \frac{1}{2}, & \text{otherwise,} \end{cases} \tag{58a}$$

$$\overleftarrow{\mathrm{P}}_{\gamma\theta}(y) = \begin{cases} y\exp(\gamma), & \text{if } 0 < y < \frac{1}{2}\exp(-\gamma); \\ y\exp(-\gamma), & \text{if } y > \frac{1}{2}\exp(\gamma); \\ \frac{1}{2}, & \text{otherwise,} \end{cases} \tag{58b}$$

$$\overleftarrow{\mathrm{env}}_{\theta}^{\gamma}(y) = \begin{cases} \frac{y(1-e^{\gamma})}{\gamma} + \frac{1}{2}, & \text{if } 0 < y < \frac{1}{2}\exp(-\gamma); \\ \frac{y(1-e^{-\gamma})}{\gamma} - \frac{1}{2}, & \text{if } \frac{1}{2}\exp(\gamma) < y; \\ \frac{2y-\ln(y)-1-\ln(2)}{2\gamma}, & \text{otherwise.} \end{cases} \tag{58c}$$

Clearly $\operatorname{Fix}\overleftarrow{\mathrm{P}}_{\gamma\theta} = \{\frac{1}{2}\} = \operatorname{argmin}\theta$. It can also be seen that $\overleftarrow{\mathrm{P}}_{\gamma\theta}(y) \to y$ as $\gamma \downarrow 0$, and $\overleftarrow{\mathrm{P}}_{\gamma\theta}(y) \to \frac{1}{2} = \operatorname{argmin}\theta$ as $\gamma \uparrow +\infty$. Moreover, once again $\overleftarrow{\mathrm{env}}_{\theta}^{\gamma}(y) \to \theta(y)$ as $\theta \downarrow 0$. This example is illustrated in Figure 4.

Now (53) implies that for every $(x,y) \in \mathbb{R}_{++} \times \mathbb{R}_{++}$,

$$y = \overrightarrow{\mathrm{P}}_{\gamma\theta}(x) \quad \Leftrightarrow \quad 0 \in \gamma\partial\theta(y) + \frac{1}{y}(y-x) \quad \Leftrightarrow \quad x \in y\big(1+\gamma\partial\theta(y)\big). \tag{59}$$

Solving the induced system of equations yields

$$\overrightarrow{\mathrm{P}}_{\gamma\theta}(x) = \begin{cases} \frac{x}{1-\gamma}, & \text{if } 0 < x < \frac{1-\gamma}{2}; \\ \frac{x}{1+\gamma}, & \text{if } x > \frac{1+\gamma}{2}; \\ \frac{1}{2}, & \text{otherwise.} \end{cases} \tag{60}$$

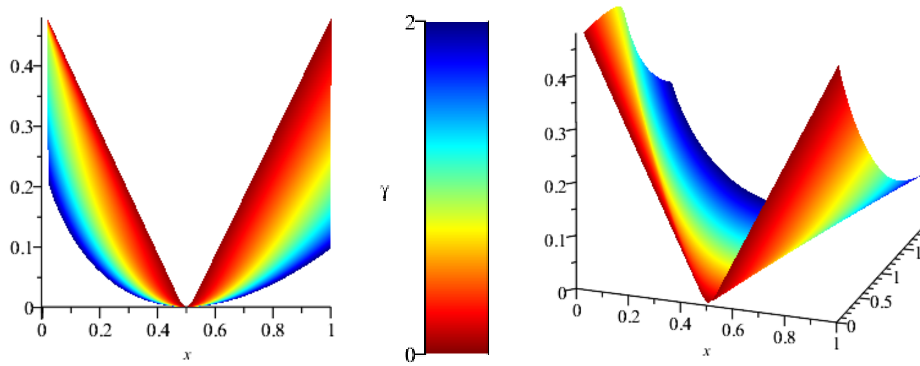


Figure 5: Right Bregman envelope from Example 4.1(ii)

Using (26a) and noting that $\frac{x}{1-\gamma} < \frac{1}{2}$ if $x < \frac{1-\gamma}{2}$ and $\frac{x}{1+\gamma} > \frac{1}{2}$ if $x > \frac{1+\gamma}{2}$, we obtain

$$\overrightarrow{\mathrm{env}}_{\theta}^{\gamma}(x) = \begin{cases} \frac{\ln(1-\gamma)}{\gamma}x + \frac{1}{2}, & \text{if } 0 < x < \frac{1-\gamma}{2}; \\ \frac{\ln(1+\gamma)}{\gamma}x - \frac{1}{2}, & \text{if } x > \frac{1+\gamma}{2}; \\ \frac{1}{\gamma}\left(x\ln(2x) - x + \frac{1}{2}\right), & \text{otherwise.} \end{cases} \tag{61}$$

The right envelope is shown in Figure 5.

(iii) *Fermi–Dirac entropy:* Suppose that $f$ is the Fermi–Dirac entropy. Then $\operatorname{dom} f = [0,1]$, $U = \operatorname{int}\operatorname{dom} f = ]0,1[$, $\nabla f(x) = \ln\left(\frac{x}{1-x}\right)$, and $\nabla^2 f(x) = \frac{1}{x(1-x)}$. Again by (52),

$$(\nabla f + \gamma\partial\theta)^{-1}(y) = \begin{cases} \frac{\exp(y+\gamma)}{\exp(y+\gamma)+1}, & \text{if } y < -\gamma; \\ \frac{\exp(y-\gamma)}{\exp(y-\gamma)+1}, & \text{if } y > \gamma; \\ \frac{1}{2}, & \text{otherwise,} \end{cases} \tag{62a}$$

$$\overleftarrow{\mathrm{P}}_{\gamma\theta}(y) = \begin{cases} \frac{y\exp(\gamma)}{y\exp(\gamma)+1-y} & \text{if } 0 < y < \frac{\exp(-\gamma)}{1+\exp(-\gamma)}; \\ \frac{y\exp(-\gamma)}{y\exp(-\gamma)+1-y}, & \text{if } \frac{\exp(\gamma)}{1+\exp(\gamma)} < y < 1; \\ \frac{1}{2}, & \text{otherwise.} \end{cases} \tag{62b}$$

A formula for $\overleftarrow{\mathrm{env}}_{\theta}^{\gamma}$ may be once again obtained by using (25a):

$$\overleftarrow{\mathrm{env}}_{\gamma\theta} = \begin{cases} -\frac{2\ln(y\exp(\gamma)-y+1)-\gamma}{2\gamma}, & \text{if } 0 < y < \frac{\exp(-\gamma)}{1+\exp(-\gamma)} \\ -\frac{2\ln(y\exp(-\gamma)-y+1)+\gamma}{2\gamma}, & \text{if } 1 > y > \frac{\exp(\gamma)}{1+\exp(\gamma)} \\ -\frac{2\ln(2)+\ln(1-y)+\ln(y)}{2\gamma}, & \text{if } \frac{\exp(-\gamma)}{1+\exp(-\gamma)} \le y \le \frac{\exp(\gamma)}{1+\exp(\gamma)} \end{cases}. \tag{63}$$

We illustrate this envelope in Figure 6.

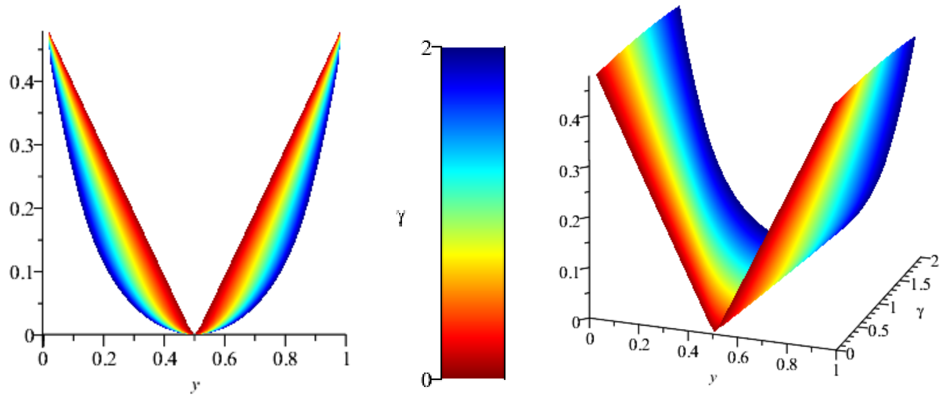


Figure 6: Left Bregman envelope from Example 4.1(iii)

Next we have from (53) that for every $(x,y) \in \,]0,1[\, \times \,]0,1[$,

$$y = \overrightarrow{\mathrm{P}}_{\gamma\theta}(x) \quad \Leftrightarrow \quad 0 \in \gamma\partial\theta(y) + \frac{1}{y(1-y)}(y-x) \quad \Leftrightarrow \quad x \in \gamma y(1-y)\partial\theta(y) + y. \tag{64}$$

Solving the induced system of equations gives

$$\overrightarrow{\mathrm{P}}_{\gamma\theta}(x) = \begin{cases} \frac{\gamma-1+\sqrt{(\gamma-1)^2+4\gamma x}}{2\gamma}, & \text{if } 0 < x < \frac{2-\gamma}{4}; \\ \frac{\gamma+1-\sqrt{(\gamma+1)^2-4\gamma x}}{2\gamma}, & \text{if } \frac{2+\gamma}{4} < x < 1; \\ \frac{1}{2}, & \text{otherwise} \end{cases} \tag{65}$$

and, in turn, (26a) gives

$$\overrightarrow{\mathrm{env}}_{\gamma\theta}(x) = \begin{cases} \frac{2\ln\left(\frac{2\gamma x^x(\gamma+1-\sqrt{\gamma^2+4\gamma x-2\gamma+1})^{x-1}}{(\gamma-1+\sqrt{\gamma^2+4\gamma x-2\gamma+1})^x(1-x)^{x-1}}\right)+1-\sqrt{\gamma^2+4\gamma x-2\gamma+1}}{2\gamma}, & \text{if } 0 < x < \frac{2-\gamma}{4}; \\ \frac{\ln\left(\frac{2\gamma x^x(\gamma-1+\sqrt{\gamma^2-4\gamma x+2\gamma+1})^{x-1}}{(1-x)^{x-1}(\gamma+1-\sqrt{\gamma^2-4\gamma x+2\gamma+1})^x}\right)+1-\sqrt{\gamma^2-4\gamma x+2\gamma+1}}{2\gamma}, & \text{if } \frac{2+\gamma}{4} < x < 1; \\ \frac{x\ln(x)+(1-x)\ln(1-x)+\ln(2)}{\gamma}, & \text{otherwise.} \end{cases} \tag{66}$$

The right envelope is shown in Figure 7.

## Acknowledgments

HHB was partially supported by the Natural Sciences and Engineering Research Council of Canada. MND was partially supported by the Australian Research Council Discovery Project DP160101537.

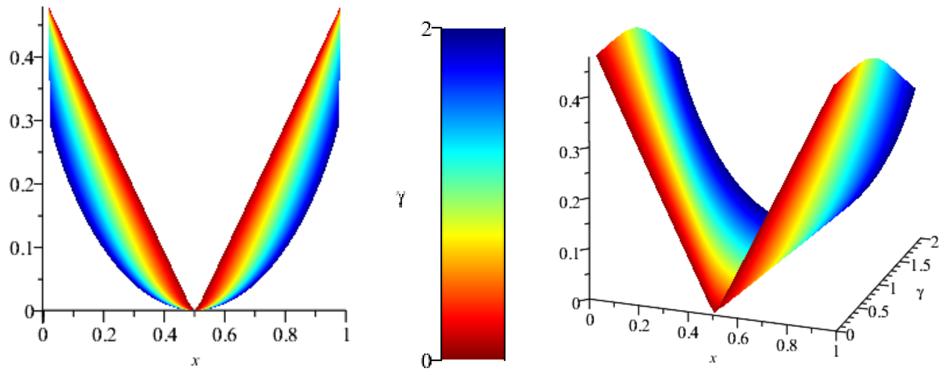


Figure 7: Right Bregman envelope from Example 4.1(iii)

# References


[1] F. Alvarez, R. Correa, and M. Marechal, Regular self-proximal distances are Bregman, *Journal of Convex Analysis* 24 (2017), 135–148.

[2] H. Attouch, Convergence de fonctions convexes, des sous-différentiels et semi-groupes associés, *Comptes Rendus de l'Académie des Sciences de Paris* 284 (1977), 539–542.

[3] H. Attouch, *Variational Convergence for Functions and Operators*, Pitman, 1984.

[4] H.H. Bauschke, J. Bolte, and M. Teboulle, A descent lemma beyond Lipschitz gradient continuity: first-order methods revisited and applications, *Mathematics of Operations Research*, 2016. DOI 10.1287/moor.2016.0817

[5] H.H. Bauschke and J.M. Borwein, Legendre functions and the method of random Bregman projections, *Journal of Convex Analysis* 4 (1997), 27–67.

[6] H.H. Bauschke, J.M. Borwein, and P.L. Combettes, Bregman monotone optimization algorithms, *SIAM Journal on Control and Optimization* 42 (2003), 596–636.

[7] H.H. Bauschke and P.L. Combettes, Iterating Bregman retractions, *SIAM Journal on Optimization* 13 (2003), 1159–1173.

[8] H.H. Bauschke and P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, second edition, Springer, 2017.

[9] H.H. Bauschke, P.L. Combettes, and D. Noll, Joint minimization with alternating Bregman proximity operators, *Pacific Journal of Optimization* 2 (2006), 401–424.

[10] H.H. Bauschke and A.S. Lewis, Dykstra's algorithm with Bregman projectors: a convergence proof, *Optimization* 48 (2000), 409–427.

[11] H.H. Bauschke and D. Noll, The method of forward projections, *Journal of Nonlinear and Convex Analysis* 3 (2002), 191–205.

[12] L.M. Bregman, The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming, *USSR Computational Mathematics and Mathematical Physics* 7 (1967), 200–217.

[13] R. Burachik and G. Kassay, On a generalized proximal point method for solving equilibrium problems in Banach spaces, *Nonlinear Analysis* 75 (2012), 6456–6464.

[14] D. Butnariu and A.N. Iusem, *Totally Convex Functions for Fixed Points Computation and Infinite Dimensional Optimization*, Kluwer, 2000.

[15] C. Byrne and Y. Censor, Proximity function minimization using multiple Bregman projections, with applications to split feasibility and Kullback–Leibler distance minimization, *Annals of Operations Research* 105 (2001), 77–98.

[16] Y. Censor and G. T. Herman, Block-iterative algorithms with underrelaxed Bregman projections, *SIAM Journal on Optimization* 13 (2002), 283–297.

[17] Y. Censor and S. Reich, The Dykstra algorithm with Bregman projections, Communications in Applied Analysis 2 (1998), 407–419.

[18] Y. Censor and S.A. Zenios, Proximal minimization algorithm with D-functions, *Journal of Optimization Theory and Applications* 73 (1992), 451–464.

[19] Y. Censor and S.A. Zenios, *Parallel Optimization: Theory, Algorithms, and Applications*, Oxford University Press, 1997.

[20] G. Chen and M. Teboulle, Convergence analysis of a proximal-like minimization algorithm using Bregman functions, *SIAM Journal on Optimization* 3 (1993), 538–543.

[21] Y.Y. Chen, C. Kan, and W. Song, The Moreau envelope function and proximal mapping with respect to the Bregman distance in Banach spaces, *Vietnam Journal of Mathematics* 40 (2012), 181–199.

[22] P.L. Combettes and Q.V. Nguyen, Solving composite monotone inclusions in reflexive Banach spaces by constructing best Bregman approximations from their Kuhn–Tucker set, *Journal of Convex Analysis* 23 (2016), 481–510.

[23] J. Eckstein, Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming, *Mathematics of Operations Research* 18 (1993), 202–226.

[24] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms II: Advanced Theory and Bundle Methods*, Springer, 1993.

[25] C. Kan and W. Song, The Moreau envelope function and proximal mapping in the sense of the Bregman distance, *Nonlinear Analysis* 75 (2012), 1385–1399.

[26] G. Kassay, S. Reich, and S. Sabach, Iterative methods for solving systems of variational inequalities in reflexive Banach spaces, *SIAM Journal on Optimization* 21 (2011), 1319–1344.

[27] K.C. Kiwiel, Proximal minimization methods with generalized Bregman functions, *SIAM Journal on Control and Optimization* 35 (1997), 1142–1168.

[28] K. Lange, *MM Optimization Algorithms*, SIAM, 2016.

[29] B.S. Mordukhovich and N.M. Nam, *An Easy Path to Convex Analysis and Applications*, Morgan & Claypool Publishers, 2013.

[30] J.-J. Moreau, Proximité et dualité dans un espace hilbertien, *Bulletin de la Société Mathématique de France* 93 (1965), 273–299.

[31] Q.V. Nguyen, Forward-backward splitting with Bregman distances, *Vietnam Journal of Mathematics*, 2017, `DOI 10.1007/s10013-016-0238-3`

[32] Q.V. Nguyen, Variable quasi-Bregman monotone sequences, *Numerical Algorithms* 73 (2016), 1107–1130.

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

[34] R.T. Rockafellar and R.J-B Wets, *Variational Analysis*, Springer-Verlag, 1998.

[35] S. Sabach, Products of finitely many resolvents of maximal monotone mappings in reflexive Banach spaces, *SIAM Journal on Optimization* 21 (2011), 1289–1308.