

Structured Sparsity Promoting Functions

Lixin Shen*

Bruce W. Suter[†]

Erin E. Tripp*

September 19, 2018

Abstract

Motivated by the minimax concave penalty based variable selection in high-dimensional linear regression, we introduce a simple scheme to construct structured semiconvex sparsity promoting functions from convex sparsity promoting functions and their Moreau envelopes. Properties of these functions are developed by leveraging their structure. In particular, we provide sparsity guarantees for the general family of functions. We further study the behavior of the proximity operators of several special functions including indicator functions of closed convex sets, piecewise quadratic functions, and the linear combinations of them. To demonstrate these properties, several concrete examples are presented and existing instances are featured as special cases.

Keywords— Moreau envelope, proximity operator, variable selection, sparsity, thresholding operator

1 Introduction

Natural signals and data streams are often inherently sparse in certain bases or dictionaries where they can be approximately represented by only a few significant components carrying the most relevant information [4, 15, 22]. Regularization methods are a powerful tool for sparse modeling and have been widely used to analyze these data sets. A particular method depends on the choice of penalty used to enforce constraints on the objective. The natural penalty function to promote sparsity is the so-called ℓ_0 -norm, which counts the nonzero components of a vector. However minimizing the ℓ_0 -norm is a combinatorial optimization problem which is known to be NP-hard.

To overcome these computational difficulties, regularization methods with the ℓ_1 -norm as its penalty function like LASSO [24] and Dantzig selectors [4] have been proposed. The convexity of the ℓ_1 -norm makes the implementation of the corresponding methods numerically tractable. However, despite its appealing properties, convex regularization methods can suffer from the bias issue that is inherited from the convexity of the penalty function. To address this, non-convex penalties including the ℓ_q -penalty with $0 < q < 1$ [11], the smoothly clipped absolute deviation penalty (SCAD) [10] and the minimax concave penalty (MCP) [25] have been proposed to replace the ℓ_1 -norm penalty.

In this paper, we introduce a family of semiconvex sparsity promoting functions of which each is the difference of a convex sparsity promoting function with its Moreau envelope. Roughly speaking, a sparsity promoting function is one that admits its global minimum at the origin but is nondifferentiable there; a function is semiconvex if it becomes convex after adding a convex quadratic

*Department of Mathematics, Syracuse University, Syracuse, NY 13244, USA. Email: lshen03@syr.edu and eetripp@syr.edu

[†]Air Force Research Laboratory, Rome, NY. Email: bruce.suter@us.af.mil.

function to it. Semiconvex functions possess useful structure and obey generalizations of many classical results from convex analysis (see, e.g. [3]).

We show that as long as a convex function is a sparsity promoting function, so is the difference between it and its Moreau envelope. This result makes the construction of nonconvex sparsity promoting functions effortless. Some interesting properties of such functions are: (i) they are always nonnegative and semiconvex and (ii) they are a special type of difference of convex (DC) functions with one having a Lipschitz continuous gradient. Due to these properties, we will refer to these functions as structured semiconvex sparsity promoting functions. These properties enable us to make use of the fruitful results, for example, in DC programming [23], to develop efficient algorithms for the associated regularized optimization problems. What's more, these functions provide a bridge between convex and nonconvex sparsity promoting penalties. As a specific example, we recover the MCP from the difference of the ℓ_1 -norm and its envelope. It has been shown (e.g. in [21]) that this closely approximates the ℓ_0 -norm while preserving the continuity and subdifferentiability of ℓ_1 .

The proximity operator, which was first introduced by Moreau in [17] as a generalization of the notion of projection onto a convex set, has been used extensively in nonlinear optimization (see, e.g., [1, 2, 7]). The desired features of the aforementioned regularization methods can be explained in terms of the proximity operators of the corresponding penalties. Therefore to determine the effectiveness of our proposed functions, we must examine the behavior of their proximity operators. The proximity operator of the ℓ_0 -norm is the hard thresholding operator, which annihilates all entries below a certain threshold and keeps all entries above the threshold. In fact, we see that hard thresholding rules are characteristic of penalties which are concave near the origin and constant elsewhere. More generally, we provide sparsity guarantees in terms of thresholding behavior for the entire family of structured semiconvex sparsity promoting functions, with further details for certain special functions.

The rest of the paper is organized as follows. Section 2 provides motivation for the suggested scheme. Section 3 recalls some necessary background in optimization and introduces the concept of the sparsity promoting function. In Section 4, we construct a family of semiconvex sparsity promoting functions which are the difference of convex sparsity promoting functions and their Moreau envelopes. Many interesting properties of this family of functions are presented. In Section 5, several special sparsity promoting functions are presented and discussed thoroughly. Some examples of practical interest are provided in Section 6. We conclude by discussing applications and plans for future work in Section 7.

2 Motivation

Our work on semiconvex sparsity promoting functions was motivated mainly by the minimax concave penalty (MCP) based variable selection in high-dimensional linear regression [25]. Variable selection is fundamental in statistical analysis of high-dimensional data. It is also easily interpretable in terms of sparse signal recovery. We consider a linear regression model with n -dimensional response vector y , $n \times p$ model matrix X , p -dimensional regression vector γ , and n -dimensional error vector ϵ :

$$y = X\gamma + \epsilon.$$

The goal of variable selection is to recover the true underlying sparse model of the pattern $\{j : \gamma_j \neq 0\}$ and to estimate the non-zero regression coefficients γ_j , where γ_j is the j -th component of γ . For small p , subset selection methods can be used to find a good guess of the pattern (see, e.g., [19]). However, subset selection becomes computationally infeasible for large p .

To overcome the computational difficulties of subset selection method, the method of penalized least squares is widely used in variable selection to produce meaningful interpretable models:

$$\min_{\gamma \in \mathbb{R}^p} \frac{1}{2n} \|y - X\gamma\|^2 + \sum_{j=1}^p \rho(|\gamma_j|, \lambda), \quad (1)$$

where $\rho(\cdot, \lambda)$ is a penalty function indexed by $\lambda \geq 0$. The penalty function $\rho(t, \lambda)$, defined on $[0, \infty)$, is assumed to be nondecreasing in t with $\rho(0, \lambda) = 0$ and continuously differentiable for $t \in (0, \infty)$. The formulation in (1) includes many popular variable selection methods. For example, the best subset selection amounts to using the ℓ_0 penalty $\rho(|t|, \lambda) = \frac{\lambda^2}{2} \mathbb{1}_{\{|t| \neq 0\}}$ while LASSO [24] and basis pursuit [5] use the ℓ_1 penalty $\rho(|t|, \lambda) = \lambda|t|$. Here $\mathbb{1}_{\{u \in E\}}$ denotes the characteristic function and $\mathbb{1}_{\{u \in E\}}$ equals 1 if $u \in E$ and 0 otherwise. The estimator (i.e, the hard thresholding operator) with the ℓ_0 penalty suffers from instability in model prediction while the estimator (the soft thresholding operator) with the ℓ_1 penalty suffers from the bias issue, severely interfering with variable selection for large p [10]. To remedy this issue, the SCAD penalty was introduced in [10]. The estimator with the SCAD penalty is continuous and leaves large components not excessively penalized. In [25], the MCP penalty was introduced and is defined as follows

$$\rho(|t|, \lambda) = \lambda \int_0^{|t|} \max \left\{ 0, 1 - \frac{x}{a\lambda} \right\} dx, \quad (2)$$

where the parameter $a > 0$. This penalty function (see [25]) minimizes the maximum concavity

$$\kappa(\rho, \lambda) := \sup_{0 < t_1 < t_2 < \infty} -\frac{\rho(t_2, \lambda) - \rho(t_1, \lambda)}{t_2 - t_1}$$

subject to the unbiasedness $\frac{\partial}{\partial t} \rho(t, \lambda) = 0$ for all $t > a\lambda$ and selection features $\frac{\partial}{\partial t} \rho(0+, \lambda) = \lambda$. The number $\kappa(\rho, \lambda)$ is related to the computational complexity of regularization method for solving (1). The simulations in [10, 25] gave a strong statistical evidence that the estimators from the non-convex penalty functions SCAD and MCP are useful in variable selection. Recently, an application of MCP to signal processing was reported in [20].

Due to its success in applications, we take a closer look at MCP. The MCP function in (2) can be rewritten as

$$\rho(|t|, \lambda) = \lambda(|t| - \text{env}_{a\lambda|\cdot|}(t)),$$

where $\text{env}_{a\lambda|\cdot|}$ is the Moreau envelope of $|\cdot|$ with index $a\lambda$ (see next section). Clearly, the MCP penalty can be considered as a variation of the ℓ_1 penalty function, that is, the absolute function $|\cdot|$ is replaced by $|\cdot| - \text{env}_{a\lambda|\cdot|}$. From this simple observation, we are drawn to consider a family of penalty functions defined by

$$f - \text{env}_{\alpha} f$$

with f satisfying some proper properties and $\alpha > 0$.

The goal of this paper is to have a comprehensive study on mathematical properties of this family of functions, particularly their proximity operators that are closely related to selection features when adopted in (1).

3 Sparsity Promoting Functions: Definition

In this section, we provide a formal definition of sparsity promoting and characterize convex sparsity promoting functions. We begin by collecting the necessary definitions and facts from convex analysis.

All functions in this work are defined on \mathbb{R}^n , Euclidean space equipped with the standard inner product $\langle \cdot, \cdot \rangle$ and the induced Euclidean norm $\|\cdot\|$. We use $\Gamma(\mathbb{R}^n)$ (respectively $\Gamma_0(\mathbb{R}^n)$) to represent the set of proper lower semicontinuous (respectively convex) functions on \mathbb{R}^n . The domain of an operator A (respectively a function g) is denoted $\text{dom}(A)$ (respectively $\text{dom}(g)$). The boundary of a set S denoted by $\text{bd}(S)$ is the set of points in the closure \bar{S} which are not in the interior $\text{int}(S)$. The relative interior of a set S denoted by $\text{ri}(S)$ is the interior of S when it is viewed as a subset of the affine space it spans. For any $x \in \mathbb{R}^n$ and any $\delta > 0$, we use $B_\delta(x)$ to denote the open ball centered at x with radius δ . In particular, we are interested in $B_{\|x\|}(x) = \{u : \|u - x\| < \|x\|\}$. For a real number a , the signum function $\text{sgn}(a)$ is defined as

$$\text{sgn}(a) = \begin{cases} -1, & \text{if } a < 0; \\ 0, & \text{if } a = 0; \\ 1, & \text{if } a > 0. \end{cases}$$

For any $g \in \Gamma(\mathbb{R}^n)$, the Fréchet subdifferential of g at $x \in \text{dom}(g)$ is the set

$$\partial g(x) := \left\{ d \in \mathbb{R}^n : \liminf_{u \rightarrow x} \frac{g(u) - g(x) - \langle d, u - x \rangle}{\|u - x\|} \geq 0 \right\}.$$

For any $x \notin \text{dom}(g)$, $\partial g(x) = \emptyset$. If $\partial g(x)$ is single-valued, $\partial g(x) = \{\nabla g(x)\}$. We leave off the brackets when there is no risk of confusion. If $g \in \Gamma_0(\mathbb{R}^n)$, the above subdifferential reduces to the usual one

$$\partial g(x) = \{d \in \mathbb{R}^n : g(y) \geq g(x) + \langle d, y - x \rangle, \forall y \in \mathbb{R}^n\}.$$

If $g \in \Gamma_0(\mathbb{R}^n)$, then ∂g is a monotone operator; that is, for any $x, y \in \text{dom}(g)$, $d \in \partial g(x)$, and $\tilde{d} \in \partial g(y)$, $\langle \tilde{d} - d, y - x \rangle \geq 0$.

For a function g in $\Gamma(\mathbb{R}^n)$, the Moreau envelope of f with parameter α , denoted by $\text{env}_{\alpha g}$, is

$$\text{env}_{\alpha g}(x) = \inf_{u \in \mathbb{R}^n} \left\{ g(u) + \frac{1}{2\alpha} \|u - x\|^2 \right\}.$$

The associated proximity operator of g with parameter α at x is the set of all points at which the above infimum is attained, denoted by $\text{prox}_{\alpha g}(x)$:

$$\text{prox}_{\alpha g}(x) = \text{argmin}_{u \in \mathbb{R}^n} \left\{ g(u) + \frac{1}{2\alpha} \|u - x\|^2 \right\}.$$

When $\text{prox}_{\alpha g}(x) \neq \emptyset$, $\text{env}_{\alpha g}(x) = g(p) + \frac{1}{2\alpha} \|p - x\|^2$ for all $p \in \text{prox}_{\alpha g}(x)$.

Recall that for a proper function g on \mathbb{R}^n , the Fenchel conjugate g^* is defined as

$$g^*(x) = \sup_{u \in \mathbb{R}^n} \{ \langle u, x \rangle - g(u) \}.$$

The Fenchel conjugate is closely related to the Moreau envelope. Indeed, it is shown in [2] that for any $x \in \mathbb{R}^n$ and $\alpha > 0$,

$$\left(g + \frac{1}{2\alpha} \|\cdot\|^2 \right)^* (\alpha^{-1}x) = \left(-\text{env}_{\alpha g} + \frac{1}{2\alpha} \|\cdot\|^2 \right) (x). \quad (3)$$

We now rigorously define what is meant by sparsity promoting and discuss how this captures the behavior described in the previous section.

Definition 1. Let $f \in \Gamma(\mathbb{R}^n)$. Then f is said to be a sparsity promoting function provided that (i) $f(0) = 0$ and f achieves its global minimum at the origin; and (ii) the set $\partial f(0)$ contains at least one nonzero element.

From the above definition, if $f \in \Gamma(\mathbb{R}^n)$ is a sparsity promoting function, then by Fermat's rule $0 \in \partial f(0)$ and f must be nondifferentiable at the origin. As pointed out in [10], the non-differentiability of f at the origin is necessary for f to be a suitable penalty in (1) for variable selection.

One typical sparsity promoting function is the absolute value function on \mathbb{R} . The global minimum is $|0| = 0$, and $\partial|\cdot|(0) = [-1, 1]$. We will return to this example throughout to illustrate various properties and connect them to our motivating example MCP. Another example of a sparsity promoting function is the indicator function ι_C that is defined by

$$\iota_C(x) := \begin{cases} 0, & \text{if } x \in C; \\ +\infty, & \text{otherwise,} \end{cases}$$

where C is a closed, convex set such that $0 \in \text{bd } C$ and $\{0\} \subsetneq C$. For further discussion of this example, we refer to Section 5. Beyond these examples, the following result shows that every norm $\|\cdot\|$ on \mathbb{R}^n is a sparsity promoting function.

Proposition 1. Let $\|\cdot\|$ be a norm on \mathbb{R}^n . Then the norm $\|\cdot\|$ is a sparsity promoting function.

Proof. It is obvious that the norm $\|\cdot\|$ is convex and $0 = \|0\| = \min_{x \in \mathbb{R}^n} \|x\|$. We further know that

$$\partial\|\cdot\|(0) = \{s \in \mathbb{R}^n : \max_{\|u\| \leq 1} \langle s, u \rangle \leq 1\},$$

which is the unit ball associated with the dual norm of $\|\cdot\|$ (see, e.g. [13]). The result of this proposition follows. \square

It is well known that the relationship between the subdifferential and proximity operator of a function $f \in \Gamma_0(\mathbb{R}^n)$ is characterized as follows (see, e.g., [2, 16]): for any $\alpha > 0$

$$x \in \alpha \partial f(y) \Leftrightarrow y = \text{prox}_{\alpha f}(x + y). \quad (4)$$

From this relationship, we get the following characterization of convex sparsity promoting functions.

Lemma 1. Let $f \in \Gamma_0(\mathbb{R}^n)$ be a sparsity promoting function and let $\alpha > 0$. Then the following statements hold.

- (i) If $x \in \alpha \partial f(0)$, then $\text{prox}_{\alpha f}(x) = 0$.
- (ii) For all $x \in \text{dom}(f)$, $\|\text{prox}_{\alpha f}(x)\| \leq \|x\|$.

Proof. (i): This is a direct consequence of equation (4).

(ii): Note that $\text{prox}_{\alpha f}(0) = 0$ due to $0 \in \alpha \partial f(0)$ and Item (i). Since $\text{prox}_{\alpha f}$ is a nonexpansive operator, then for all $x \in \text{dom}(f)$, $\|\text{prox}_{\alpha f}(x)\| = \|\text{prox}_{\alpha f}(x) - \text{prox}_{\alpha f}(0)\| \leq \|x - 0\|$. \square

It follows from Lemma 1 that the proximity operator of a convex sparsity promoting function shrinks all input towards the origin, and all input below a certain threshold are sent to zero. As an example, the proximity operator of $|\cdot|$ is $\text{prox}_{\alpha|\cdot|}(x) = \text{sign}(x) \max\{|x| - \alpha, 0\}$, which is the well-known soft thresholding operator in wavelet literature [9]. This very behavior for the ℓ_1 -penalty is described by Tibshirani in the name LASSO: least absolute shrinkage and selection operator [24].

4 Semiconvex Sparsity Promoting Functions

In this section, we introduce the titular family of semiconvex sparsity promoting functions. For any $f \in \Gamma_0(\mathbb{R}^n)$ and any positive number $\alpha > 0$, we define

$$f_\alpha(x) := f(x) - \text{env}_{\alpha f}(x). \quad (\mathcal{F}_\alpha)$$

Clearly $f_\alpha \in \Gamma(\mathbb{R}^n)$ is the difference of two convex functions. Returning to the example $f(x) = |x|$, we see that $f_\alpha(x) = \min\{|x| - \frac{1}{2\alpha}x^2, \frac{\alpha}{2}\}$. As discussed in the previous section, this is the scaled the minimax concave penalty (MCP) given in [25].

Sparsity promotion depends entirely on the behavior of a function and its subdifferential at the origin. Since the Moreau envelope of any function f in $\Gamma_0(\mathbb{R}^n)$ is differentiable (see, e.g. [2]), the subdifferentials of f_α and f are related as follows (see [6]):

$$\partial f_\alpha(x) = \partial f(x) - \nabla \text{env}_{\alpha f}(x). \quad (5)$$

Due to this inherent relationship between ∂f_α and ∂f , we see immediately that f_α must be sparsity promoting if f is.

Theorem 1. *Let $f \in \Gamma_0(\mathbb{R}^n)$ be a sparsity promoting function. Then the following statements hold:*

- (i) *For any $\alpha > 0$, the function f_α defined by (\mathcal{F}_α) is a sparsity promoting function. Moreover, $\partial f_\alpha(0) = \partial f(0)$;*
- (ii) *Let $g : x \mapsto f(-x)$. Then both g and g_α are sparsity promoting. Moreover, $g_\alpha = f_\alpha(-\cdot)$ and $\partial g_\alpha(0) = -\partial f(0)$.*

Proof. (i): As a direct consequence of the definition of the Moreau envelope, $\text{env}_{\alpha f}(x) \leq f(x)$ for all $x \in \mathbb{R}^n$, hence $f_\alpha(x) \geq 0$ for all $x \in \text{dom}(f)$. Since f is a sparsity promoting function, we have $f_\alpha(0) = f(0) - \text{env}_{\alpha f}(0) = 0$. Therefore, $\min_{x \in \mathbb{R}^n} f_\alpha(x) = f_\alpha(0) = 0$. On the other hand, from (5) and the relation $\nabla \text{env}_{\alpha f}(x) = \frac{1}{\alpha}(x - \text{prox}_{\alpha f}(x))$, we have $\partial f_\alpha(0) = \partial f(0)$ which contains at least one nonzero element by assumption. Hence, f_α is sparsity promoting.

(ii): Since $g(0) = f(0) = \min_{x \in \mathbb{R}^n} f(x) = \min_{x \in \mathbb{R}^n} g(x)$ and $\partial g(0) = -\partial f(0)$, so g is sparsity promoting. Hence, by Item (i), g_α is sparsity promoting and $\partial g_\alpha(0) = -\partial f(0)$. By the definition of the Moreau envelope, $\text{env}_{\alpha g}(x) = \text{env}_{\alpha f}(-x)$ which leads to $g_\alpha = f_\alpha(-\cdot)$. \square

With Theorem 1, we say f_α is a structured sparsity promoting function if f is a convex sparsity promoting function. We now prove that f_α is semiconvex and show how its semiconvexity depends on the convexity of f . We remind the reader of the definition. A function $g \in \Gamma_0(\mathbb{R}^n)$ is σ -strongly convex if and only if there exists a constant $\sigma > 0$ such that the function $g - \frac{\sigma}{2}\|\cdot\|^2$ is convex. A function $g \in \Gamma(\mathbb{R}^n)$ is ρ -semiconvex if $g + \frac{\rho}{2}\|\cdot\|^2$ is convex.

Proposition 2. *Let f be a function in $\Gamma_0(\mathbb{R}^n)$. Then f_α , defined by (\mathcal{F}_α) , is $\frac{1}{\alpha}$ -semiconvex. If, in addition, f is μ -strongly convex, then f_α is $(\mu - \frac{1}{\alpha})$ -strongly convex if $\mu > \frac{1}{\alpha}$, convex if $\mu = \frac{1}{\alpha}$, and $(\frac{1}{\alpha} - \mu)$ -semiconvex if $\mu < \frac{1}{\alpha}$.*

Proof. Write

$$f_\alpha = f - \text{env}_{\alpha f} = f + \left(-\text{env}_{\alpha f} + \frac{1}{2\alpha}\|\cdot\|^2 \right) - \frac{1}{2\alpha}\|\cdot\|^2.$$

By (3), for all $x \in \mathbb{R}^n$ we have that

$$f_\alpha(x) = f(x) + \left(f + \frac{1}{2\alpha}\|\cdot\|^2 \right)^* (\alpha^{-1}x) - \frac{1}{2\alpha}\|x\|^2, \quad (6)$$

which implies that f_α is $\frac{1}{\alpha}$ -semiconvex.

In addition, if f is μ -strongly convex, then there exists a convex function g such that $f = g + \frac{\mu}{2}\|\cdot\|^2$. Replacing $f(x)$ in (6) by $g(x) + \frac{\mu}{2}\|x\|^2$, we have

$$f_\alpha(x) = g(x) + \left(f + \frac{1}{2\alpha}\|\cdot\|^2\right)^*(\alpha^{-1}x) + \frac{1}{2}\left(\mu - \frac{1}{\alpha}\right)\|x\|^2.$$

It follows from the above equation that f_α is $(\mu - \frac{1}{\alpha})$ -strongly convex if $\mu > \frac{1}{\alpha}$, convex if $\mu = \frac{1}{\alpha}$, and $(\frac{1}{\alpha} - \mu)$ -semiconvex if $\mu < \frac{1}{\alpha}$. \square

The following result is a direct consequence of Proposition 2.

Corollary 1. *Let f be a function in $\Gamma_0(\mathbb{R}^n)$ and let f_α be defined by (\mathcal{F}_α) . For any given $x \in \mathbb{R}^n$ and positive parameters α and β , we define*

$$F(u) = f_\alpha(u) + \frac{1}{2\beta}\|u - x\|^2, \quad (7)$$

where $u \in \mathbb{R}^n$. Then, F is $(\beta^{-1} - \alpha^{-1})$ -strongly convex if $\beta < \alpha$, convex if $\beta = \alpha$, and $(\alpha^{-1} - \beta^{-1})$ -semiconvex if $\beta > \alpha$. If, in addition, f is μ -strongly convex, then F is $(\mu - \alpha^{-1} + \beta^{-1})$ -strongly convex if $\mu > \alpha^{-1} - \beta^{-1}$, convex if $\mu = \alpha^{-1} - \beta^{-1}$, and $(\alpha^{-1} - \beta^{-1} - \mu)$ -semiconvex if $\mu < \alpha^{-1} - \beta^{-1}$.

As for convex sparsity promoting functions, we can further characterize the sparsity promotion of f_α by examining its proximity operator. Roughly speaking, we show that $\text{prox}_{\beta f_\alpha}(x) = 0$ for all $x \in \min\{\alpha, \beta\} \cdot \partial f(0)$. Towards this end, we present two technical lemmas. The first is a generalization of Lemma 1.

Lemma 2. *Let $f \in \Gamma_0(\mathbb{R}^n)$ be sparsity promoting and f_α as defined in (\mathcal{F}_α) .*

- (i) *For any $x \in \text{dom}(f)$, $\text{prox}_{\beta f_\alpha}(x) \subseteq \overline{B_{\|x\|}}(x)$.*
- (ii) *If $x \in \min\{\alpha, \beta\} \cdot \partial f(0)$, then $0 \in \text{prox}_{\beta f_\alpha}(x)$.*

Proof. For a fixed $x \in \mathbb{R}^n$, define F as in (7), so that $\text{prox}_{\beta f_\alpha}(x) = \text{argmin}_{u \in \mathbb{R}^n} F(u)$.

(i): Since $F(0) = \frac{1}{2\beta}\|x\|^2$ and $0 \in \overline{B_{\|x\|}}(x)$, to show $\text{prox}_{\beta f_\alpha}(x) \subseteq \overline{B_{\|x\|}}(x)$, we only need to show that for all $u \in \mathbb{R}^n \setminus \overline{B_{\|x\|}}(x)$, $F(u) > F(0)$. Actually, if $u \in \mathbb{R}^n \setminus \overline{B_{\|x\|}}(x)$, then $\|u - x\|^2 > \|x\|^2$. Since f_α is non-negative, it follows from (7) that $F(u) > \frac{1}{2\beta}\|x\|^2 = F(0)$. Thus the conclusion of Item (i) holds.

(ii): To prove Item (ii), from Item (i) and $F(0) = \frac{1}{2\beta}\|x\|^2$, it suffices to show $F(u) \geq \frac{1}{2\beta}\|x\|^2$ for all $u \in \overline{B_{\|x\|}}(x)$. From the assumption of $x \in \min\{\alpha, \beta\} \cdot \partial f(0)$, we have that for all $u \in \mathbb{R}^n$, $f(u) \geq \frac{1}{\min\{\alpha, \beta\}}\langle x, u \rangle$. Since $f(0) = 0$, we have $\text{env}_{\alpha f}(u) \leq \frac{1}{2\alpha}\|u\|^2$ for all $u \in \mathbb{R}^n$. Hence

$$f_\alpha(u) \geq \frac{1}{\min\{\alpha, \beta\}}\langle x, u \rangle - \frac{1}{2\alpha}\|u\|^2.$$

Therefore

$$\begin{aligned} F(u) &\geq \frac{1}{\min\{\alpha, \beta\}}\langle x, u \rangle - \frac{1}{2\alpha}\|u\|^2 + \frac{1}{2\beta}\|u - x\|^2 \\ &= \begin{cases} (\frac{1}{2\beta} - \frac{1}{2\alpha})\|u\|^2 + \frac{1}{2\beta}\|x\|^2, & \text{if } \beta \leq \alpha; \\ (\frac{1}{2\alpha} - \frac{1}{2\beta})(\|x\|^2 - \|u - x\|^2) + \frac{1}{2\beta}\|x\|^2, & \text{if } \alpha < \beta. \end{cases} \end{aligned}$$

So, $F(u) \geq \frac{1}{2\beta}\|x\|^2 = F(0)$ holds for all $u \in \overline{B_{\|x\|}}(x)$. This completes the proof of the lemma. \square

Remark 1. From item (i) of Lemma 2 we see for $x \in \mathbb{R}$, $\text{sgn}(x) = \text{sgn}(p)$ if $p \in \text{prox}_{\beta f_\alpha}(x)$ and both x and p are simultaneously nonzero. We note that this is also true for $\text{prox}_{\alpha f}(x)$.

The following technical lemma will greatly simplify the proof of Theorem 2, our main result. While the lemma may seem strange at first glance, the conditions therein arise naturally from the computation of the proximity operator.

Lemma 3. *Let $f \in \Gamma_0(\mathbb{R}^n)$ be a sparsity promoting function and $w \in \text{dom}(\partial f)$. If $w \in \partial f(0)$ and there exists a nonzero $\xi \in \text{ri}(\partial f(0)) \cap \partial f(w)$, then $w = 0$.*

Proof. Assume that $w \neq 0$. First, since $w \in \partial f(0)$ and $f(0) = 0$, we have $f(w) \geq \|w\|^2 > 0$.

Second, since $\xi \in \partial f(0) \cap \partial f(w)$, then $\xi \in \partial f(0)$ implies $f(0) + f^*(\xi) = \langle 0, \xi \rangle$ while $\xi \in \partial f(w)$ implies $f(w) + f^*(\xi) = \langle \xi, w \rangle$. Hence,

$$f(w) = \langle \xi, w \rangle. \quad (8)$$

By the monotonicity of ∂f , for any $\eta \in \partial f(0)$, $\langle \xi - \eta, w \rangle \geq 0$. Together with (8) we get

$$f(w) \geq \langle \eta, w \rangle. \quad (9)$$

Finally, since $\xi \in \text{ri}(\partial f(0))$ and $\partial f(0)$ is convex, there exists $\lambda > 1$ such that $\lambda \xi \in \partial f(0)$. By (8) and (9), we get

$$f(w) \geq \langle \lambda \xi, w \rangle = \lambda f(w)$$

which implies $f(w) \leq 0$. This is a contradiction, so $w = 0$. □

Now for our main result which characterizes the sparsity promoting structure of f_α in terms of the sparsity threshold of its proximity operator.

Theorem 2. *Let $f \in \Gamma_0(\mathbb{R}^n)$ be a sparsity promoting function. For any $x \in \text{dom}(f)$, the following statements hold:*

- (i) *If $\beta < \alpha$, then $\text{prox}_{\beta f_\alpha}(x) = 0$ for $x \in \beta \partial f(0)$;*
- (ii) *If $\beta = \alpha$, then $\text{prox}_{\beta f_\alpha}(x) = 0$ for $x \in \text{ri}(\alpha \partial f(0))$;*
- (iii) *If $\beta > \alpha$, then $\text{prox}_{\beta f_\alpha}(x) = 0$ for $x \in \alpha \partial f(0)$.*

Proof. Given $x \in \mathbb{R}^n$, define $F(u) = f_\alpha(u) + \frac{1}{2\beta} \|u - x\|^2$.

(i) We first consider the situation $\beta < \alpha$. From Corollary 1, we know that F is $\left(\frac{1}{\beta} - \frac{1}{\alpha}\right)$ -strongly convex and therefore has a unique minimizer. By Lemma 2, $x \in \beta \partial f(0)$ implies that $0 = \arg\min_{u \in \mathbb{R}^n} F(u)$. Together these imply that $\text{prox}_{\beta f_\alpha}(x) = 0$.

(ii) Next we consider $\alpha = \beta$. From Corollary 1, $F(u)$ is convex but not strongly, and the minimizer may no longer be unique. By Lemma 2, $0 \in \text{prox}_{\beta f_\alpha}(x)$ for $x \in \alpha \partial f(0)$.

Now suppose $x \in \text{ri}(\alpha \partial f(0))$ and let w^* be an element of $\text{prox}_{\beta f_\alpha}(x)$. To show that $w^* = 0$, by identifying αf , x , and w^* , respectively, as f , ξ , and w in Lemma 3, it suffices to show that $x \in \partial(\alpha f)(w^*)$ and $w^* \in \partial(\alpha f)(0)$. By Fermat's rule, $w^* \in \text{prox}_{\beta f_\alpha}(x)$ implies that $0 \in \partial f_\alpha(w^*) + \frac{1}{\beta}(w^* - x)$. As we saw earlier that $\partial f_\alpha(w^*) = \partial f(w^*) - \nabla \text{env}_{\alpha f}(w^*)$ and $\nabla \text{env}_{\alpha f}(w^*) = \frac{1}{\alpha}(w^* - \text{prox}_{\alpha f}(w^*))$, this can be rewritten as

$$\frac{1}{\beta}x + \left(\frac{1}{\alpha} - \frac{1}{\beta}\right)w^* - \frac{1}{\alpha}\text{prox}_{\alpha f}(w^*) \in \partial f(w^*). \quad (10)$$

From (10), we get $x - \text{prox}_{\alpha f}(w^*) \in \partial(\alpha f)(w^*)$. Therefore the conditions $x \in \partial(\alpha f)(w^*)$ and $w^* \in \partial(\alpha f)(0)$ hold if and only if $\text{prox}_{\alpha f}(w^*) = 0$.

Since $x \in \partial(\alpha f)(0)$, by the monotonicity of ∂f we have

$$\langle x - \text{prox}_{\alpha f}(w^*) - x, w^* \rangle \geq 0.$$

That is, $\langle \text{prox}_{\alpha f}(w^*), w^* \rangle \leq 0$. But due to the nonexpansiveness of $\text{prox}_{\alpha f}$ and the fact that $\text{prox}_{\alpha f}(0) = 0$,

$$\langle \text{prox}_{\alpha f}(w^*), w^* \rangle \geq \|\text{prox}_{\alpha f}(w^*)\|^2.$$

This implies that $\text{prox}_{\alpha f}(w^*) = 0$. Thus by Lemma 3, $w^* = 0$.

(iii) Finally, we consider the situation of $\beta > \alpha$. In this case, we assume that $0 \neq x \in \alpha \partial f(0)$. From Lemma 2, we know that $0 \in \text{prox}_{\beta f_\alpha}(x)$. We further show that the point 0 is the only element in $\text{prox}_{\beta f_\alpha}(x)$.

Recall from the proof of Lemma 2 that when $\beta > \alpha$,

$$F(u) \geq \left(\frac{1}{2\alpha} - \frac{1}{2\beta} \right) (\|x\|^2 - \|u - x\|^2) + \frac{1}{2\beta} \|x\|^2 \geq \frac{1}{2\beta} \|x\|^2.$$

Actually, if $w^* \in \text{prox}_{\beta f_\alpha}(x)$, then w^* must be on the boundary of $\overline{B_{\|x\|}}(x)$ and $F(w^*) = f_\alpha(w^*) + \frac{1}{2\beta} \|w^* - x\|^2 = \frac{1}{2\beta} \|x\|^2$. Thus, $f_\alpha(w^*) = 0$, that is, $f(w^*) = \text{env}_{\alpha f}(w^*)$. We also know that $f(w^*) \geq \frac{1}{\alpha} \langle x, w^* \rangle$ and $\text{env}_{\alpha f}(w^*) \leq \frac{1}{2\alpha} \|w^*\|^2$. Therefore, because $2\langle x, w^* \rangle = \|w^*\|^2$, we get

$$\text{env}_{\alpha f}(w^*) = \frac{1}{2\alpha} \|w^*\|^2,$$

which implies that $0 = \text{prox}_{\alpha f}(w^*)$. On the other hand, the identity $f(w^*) = \text{env}_{\alpha f}(w^*)$ indicates $w^* = \text{prox}_{\alpha f}(w^*)$. Therefore, $w^* = 0$. This completes the proof. \square

Remark 2. Item (iii) of the theorem is not tight. In fact in every example, when $\beta > \alpha$, $\text{prox}_{\beta f_\alpha}(x) = 0$ for all x in a set strictly larger than $\alpha \partial f(0)$. However, the exact form of this set depends entirely on the function in question.

5 Some Special Functions

The last section dealt primarily with behavior around the origin for general semiconvex sparsity promoting functions. In this section, we describe the structure of f_α on the entire domain for special classes of sparsity promoting functions, namely indicator functions, piecewise quadratic functions, and their linear combinations. The study of these particular functions is motivated by the thresholding behavior of their proximity operators.

5.1 Indicator Functions

Indicator functions are commonly used to include constraints in the objective of an optimization problem. We show in this section that they are not only fixed by the mapping $f \mapsto f_\alpha$ but they are the only functions that are fixed.

Throughout, we assume C is a closed convex set in \mathbb{R}^n with boundary $\text{bd}(C)$. Recall that the indicator function of C is

$$\iota_C(x) = \begin{cases} 0, & \text{if } x \in C; \\ +\infty, & \text{otherwise.} \end{cases} \quad (\mathcal{I})$$

We first determine when this is a sparsity promoting function.

Lemma 4. *The indicator function ι_C is sparsity promoting if and only if $0 \in \text{bd}(C)$ and $\{0\} \subsetneq C$.*

Proof. As long as $0 \in C$, $\iota_C(0) = 0$, but to be sparsity promoting, there must also be a nonzero element in $\partial\iota_C(0)$. Recall that for any x , $\partial\iota_C(x)$ is the normal cone to C at x . That is,

$$\partial\iota_C(x) = N_C(x) := \begin{cases} \{u : \sup\langle C - x, u \rangle \leq 0\}, & \text{if } x \in C \\ \emptyset, & \text{otherwise.} \end{cases}$$

Note that for $x \in C$, the normal cone is nonempty because $\{0\} \subseteq N_C(x)$. We further recall the following result from [2]:

$$x \in \text{int}(C) \iff N_C(x) = \{0\}.$$

If $0 \in \text{bd}(C)$, it follows that $N_C(0)$ is nonempty and contains a nonzero element. Conversely, if we assume $N_C(0)$ is nonempty, we must have $0 \in C$. If we further assume that $N_C(0)$ contains a nonzero element, then $0 \notin \text{int}(C)$. So we see that $0 \in \text{bd}(C)$ is equivalent to the sparsity promoting definition given in Section 3. \square

It is well known (see, e.g. [2]) that $\text{prox}_{\alpha\iota_C}(x) = P_C(x)$ and that $p = P_C(x)$ if and only if $x - p \in N_C(p)$. Here $P_C(x)$ is the unique operator such that $\|x - P_C(x)\|$ is the distance from x to C . In terms of the proximity operator, this becomes $0 = \text{prox}_{\alpha\iota_C}(x)$ if and only if $x \in N_C(0)$. Moreover $\text{env}_{\alpha\iota_C}(x) = \frac{1}{2\alpha}\|P_C(x) - x\|^2$ and

$$(\iota_C)_\alpha(x) := \iota_C(x) - \text{env}_{\alpha\iota_C}(x) = \iota_C(x). \quad (\mathcal{I}_\alpha)$$

This immediately implies that $\text{prox}_{\beta(\iota_C)_\alpha}(x) = P_C(x)$ as well. The converse of the above is also true.

Proposition 3. *Let $f \in \Gamma_0(\mathbb{R}^n)$ be sparsity promoting. If $f = f_\alpha$ as defined by (\mathcal{F}_α) , then $f = \iota_{\text{dom}(f)}$.*

Proof. Notice that $\text{dom}(\text{env}_{\alpha f}) = \mathbb{R}^n$ so $\text{dom}(f_\alpha) = \text{dom}(f)$. Hence $f = f_\alpha$ implies that $\text{env}_{\alpha f}(x) = 0$ for all $x \in \text{dom}(f)$. Because f is sparsity promoting, $f(x) \geq 0$ for all x . Hence, $0 = \text{env}_{\alpha f}(x) = \min_{u \in \mathbb{R}^n} \{f(u) + \frac{1}{2\alpha}\|u - x\|^2\}$ for all $x \in \text{dom}(f)$ implies that $f(x) = 0$ for all $x \in \text{dom}(f)$. \square

Remark 3. The proposition is true more generally if $f \in \Gamma_0(\mathbb{R}^n)$ is simply nonnegative.

5.2 Piecewise Quadratic Functions

Piecewise quadratic functions include a variety of important examples: absolute value, ReLU (rectified linear unit), and elastic net. We generalize the proximity-related properties of these functions and provide a framework for generating customized penalty functions.

The piecewise quadratic functions we consider here have the following form

$$f(x) = \begin{cases} \frac{1}{2}a_1x^2 + b_1x, & \text{if } x \leq 0; \\ \frac{1}{2}a_2x^2 + b_2x, & \text{if } x \geq 0, \end{cases} \quad (\mathcal{Q})$$

where the coefficients a_1, a_2, b_1 , and b_2 are real numbers. The characterization of sparsity promoting functions having a form given (\mathcal{Q}) is established in the following lemma.

Lemma 5. *Let f be a piecewise quadratic function defined by (\mathcal{Q}) . Then f is sparsity promoting if and only if*

$$a_1 \geq 0, \quad a_2 \geq 0, \quad b_1 \leq 0 \leq b_2, \quad \text{and} \quad b_2 - b_1 > 0. \quad (11)$$

Proof. “ \Rightarrow ”: Since f is sparsity promoting, then the assumption that f attains its minimum at 0 implies that $a_1 \geq 0$, $a_2 \geq 0$, $b_1 \leq 0$, and $b_2 \geq 0$. One can directly verify that $\partial f(0) = [b_1, b_2]$. This must contain at least one nonzero element, hence, $b_2 - b_1 > 0$.

“ \Leftarrow ”: One can see that f is nonincreasing on $(-\infty, 0]$ from $a_1 \geq 0$ and $b_1 \leq 0$ and that f is nondecreasing on $[0, \infty)$ from $a_2 \geq 0$ and $b_2 \geq 0$. So f achieves its global minimum at 0. The condition $b_2 - b_1 > 0$ implies that the set $\partial f(0) = [b_1, b_2]$ has nonzero elements. Therefore, f is a sparsity promoting function. \square

Remark 4. As a by-product of the above lemma, if f given by (Q) is a sparsity promoting function, then f must be convex, hence $f \in \Gamma_0(\mathbb{R})$.

In the rest of this section, we always assume that the coefficients in (Q) satisfy the conditions listed in (11). The proximity operator and Moreau envelope of f with index α at $x \in \mathbb{R}$ are

$$\text{prox}_{\alpha f}(x) = \begin{cases} \min \{0, \frac{1}{\alpha a_1 + 1}(x - \alpha b_1)\}, & \text{if } x \leq 0; \\ \max \{0, \frac{1}{\alpha a_2 + 1}(x - \alpha b_2)\}, & \text{if } x \geq 0; \end{cases}$$

and

$$\text{env}_{\alpha f}(x) = \begin{cases} \frac{1}{\alpha a_1 + 1}(f(x) - \frac{\alpha b_1^2}{2}), & \text{if } x \leq \alpha b_1; \\ \frac{1}{2\alpha}x^2, & \text{if } \alpha b_1 \leq x \leq \alpha b_2; \\ \frac{1}{\alpha a_2 + 1}(f(x) - \frac{\alpha b_2^2}{2}), & \text{if } x \geq \alpha b_2. \end{cases}$$

respectively. From the above two equations, we get

$$f_{\alpha}(x) = \begin{cases} \frac{\alpha a_1}{\alpha a_1 + 1}f(x) + \frac{\alpha b_1^2}{2(\alpha a_1 + 1)}, & \text{if } x \leq \alpha b_1; \\ f(x) - \frac{1}{2\alpha}x^2, & \text{if } \alpha b_1 \leq x \leq \alpha b_2; \\ \frac{\alpha a_2}{\alpha a_2 + 1}f(x) + \frac{\alpha b_2^2}{2(\alpha a_2 + 1)}, & \text{if } x \geq \alpha b_2, \end{cases} \quad (\mathcal{Q}_{\alpha})$$

which is a piecewise quadratic polynomial with possible breakpoints at αb_1 , 0, and αb_2 . We know this f_{α} is sparsity promoting by Theorem 1. Some other properties of this function which follow immediately from (Q_α) are collected in the following lemma.

Lemma 6. *Let $f \in \Gamma_0(\mathbb{R})$ be a sparsity promoting function defined by (Q). Then the following hold:*

- (i) f_{α} is nonincreasing on $(-\infty, 0]$ and is nondecreasing on $[0, \infty)$;
- (ii) f_{α} on $(-\infty, \alpha b_1]$ is convex and is a degree 2 polynomial if $a_1 > 0$ or constant if $a_1 = 0$;
- (iii) f_{α} on $[\alpha b_2, \infty)$ is convex and is a degree 2 polynomial if $a_2 > 0$ or a constant if $a_2 = 0$;
- (iv) f_{α} on $[\alpha b_1, \alpha b_2]$ is convex if $\min\{a_1, a_2\} \geq \frac{1}{\alpha}$.

Just as the sparsity promoting property corresponds to certain behavior in the proximity operator near the origin, this result in Lemma 6 guarantees special properties of the proximity operator away from the origin. To illustrate, we return to $f(x) = |x|$. This satisfies (Q) with $a_1 = a_2 = 0$, $b_1 = -1$, and $b_2 = 1$. We saw in Section 4 that $f_{\alpha}(x) = \min\{|x| - \frac{1}{2\alpha}x^2, \frac{\alpha}{2}\}$. Because this function is constant away from the origin, $\text{prox}_{\beta f_{\alpha}}(x)$ must be the identity for large values of x . For example, if $\beta > \alpha$, $\text{prox}_{\beta f_{\alpha}}(x) = x$ when $|x| \geq \sqrt{\alpha\beta}$. Some other detail can be found in Example 1 of Section 6.

In the rest of this subsection, we will give a general discussion on the proximity operator $\text{prox}_{\beta f_\alpha}$ for f defined by (\mathcal{Q}_α) . We assume that $x \geq 0$ for a moment. By Lemma 2, we know that $\text{prox}_{\beta f_\alpha}(x) \subseteq [0, \infty)$, therefore by the definition of the proximity operator,

$$\text{prox}_{\beta f_\alpha}(x) = \arg\min_{u \in [0, \infty)} E(x, u) := f_\alpha(u) + \frac{1}{2\beta}(u - x)^2.$$

In view of (\mathcal{Q}_α) , the objective function $E(x, u)$ with $(x, u) \in [0, \infty) \times [0, \infty)$ is

$$E(x, u) = \begin{cases} E_1(x, u), & \text{if } u \in [0, \alpha b_2]; \\ E_2(x, u), & \text{if } u \in [\alpha b_2, \infty), \end{cases} \quad (12)$$

where

$$E_1(x, u) = \frac{1}{2} \left(a_2 - \frac{1}{\alpha} + \frac{1}{\beta} \right) u^2 + \left(b_2 - \frac{1}{\beta} x \right) u + \frac{1}{2\beta} x^2, \quad (13)$$

$$E_2(x, u) = \frac{1}{2} \left(\frac{\alpha a_2^2}{\alpha a_2 + 1} + \frac{1}{\beta} \right) u^2 + \left(\frac{\alpha a_2 b_2}{\alpha a_2 + 1} - \frac{1}{\beta} x \right) u + \frac{\alpha b_2^2}{2(\alpha a_2 + 1)} + \frac{1}{2\beta} x^2. \quad (14)$$

These two functions match at the line $u = \alpha b_2$, that is, for all $x \geq 0$,

$$E_1(x, \alpha b_2) = E_2(x, \alpha b_2), \quad (15)$$

which will facilitate the proofs of technical lemmas given later.

Define

$$s_1(x) = \arg\min_{u \in [0, \alpha b_2]} E_1(x, u) \quad \text{and} \quad s_2(x) = \arg\min_{u \in [\alpha b_2, \infty)} E_2(x, u).$$

Obviously,

$$\text{prox}_{\beta f_\alpha}(x) \subset s_1(x) \cup s_2(x). \quad (16)$$

Therefore, to figure out the expression of $\text{prox}_{\beta f_\alpha}(x)$, there is a need to know the structures of the sets $s_1(x)$ and $s_2(x)$.

Since the quadratic polynomial $E_2(x, \cdot)$ is strictly convex, then we have for each $x \geq 0$, $s_2(x)$ is a singleton set as follows:

$$\begin{aligned} s_2(x) &= \max \left\{ \alpha b_2, \frac{\alpha a_2 + 1}{\alpha a_2(a_2\beta + 1) + 1} \left(x - \frac{\alpha a_2 \beta b_2}{\alpha a_2 + 1} \right) \right\} \\ &= \begin{cases} \alpha b_2, & \text{if } 0 \leq x \leq \alpha b_2(a_2\beta + 1); \\ \frac{\alpha a_2 + 1}{\alpha a_2(a_2\beta + 1) + 1} \left(x - \frac{\alpha a_2 \beta b_2}{\alpha a_2 + 1} \right), & \text{if } x \geq \alpha b_2(a_2\beta + 1), \end{cases} \end{aligned} \quad (17)$$

which clearly is a piecewise linear function of x .

Lemma 7. *Let f be a piecewise quadratic sparsity promoting function as defined by (\mathcal{Q}) . If $b_2 = 0$, then $\text{prox}_{\beta f_\alpha}(x) = s_2(x)$ for all $x \geq 0$, where s_2 is given by (17).*

Proof. This follows from (12) and (14) that $E(x, u) = E_2(x, u)$ for $(x, u) \in [0, \infty) \times [0, \infty)$. \square

Next, we assume that $b_2 > 0$ by Lemma 5. In view of the form of $E_1(x, \cdot)$ in (13), we consider three cases: $a_2 - \frac{1}{\alpha} + \frac{1}{\beta} > 0$, $a_2 - \frac{1}{\alpha} + \frac{1}{\beta} = 0$, and $a_2 - \frac{1}{\alpha} + \frac{1}{\beta} < 0$ which are equivalently to (i) $\alpha b_2(a_2\beta + 1) > \beta b_2$, (ii) $\alpha b_2(a_2\beta + 1) = \beta b_2$, and (iii) $\alpha b_2(a_2\beta + 1) < \beta b_2$, respectively. Accordingly, $E_1(x, \cdot)$ is strongly convex, convex, or concave on $[0, \alpha b_2]$. The result for case (i) is stated in the following lemma.

Lemma 8. Let f be a piecewise quadratic sparsity promoting function as defined by (Q). If $b_2 > 0$ and $\alpha b_2(a_2\beta + 1) > \beta b_2$, then

$$\text{prox}_{\beta f_\alpha}(x) = \begin{cases} 0, & \text{if } 0 \leq x < \beta b_2; \\ \frac{\alpha}{(a_2\beta+1)\alpha-\beta}(x - \beta b_2), & \text{if } \beta b_2 \leq x \leq \alpha b_2(a_2\beta + 1); \\ \frac{\alpha a_2 + 1}{\alpha a_2(a_2\beta + 1) + 1} \left(x - \frac{\alpha a_2 \beta b_2}{\alpha a_2 + 1} \right), & \text{if } x > \alpha b_2(a_2\beta + 1). \end{cases} \quad (18)$$

Proof. From (16), we first find the set $s_1(x)$ since the set $s_2(x)$ is already given in (17). By the assumption of this lemma, for each $x \geq 0$, $s_1(x)$ contains only one element and is given as follows:

$$s_1(x) = \begin{cases} 0, & \text{if } 0 \leq x < \beta b_2; \\ \frac{\alpha}{(a_2\beta+1)\alpha-\beta}(x - \beta b_2), & \text{if } \beta b_2 \leq x \leq \alpha b_2(a_2\beta + 1); \\ \alpha b_2, & \text{if } x > \alpha b_2(a_2\beta + 1). \end{cases}$$

To determine the expression of $\text{prox}_{\beta f_\alpha}(x)$ from the sets $s_1(x)$ and $s_2(x)$, we look at the behaviours of the functions E_1 and E_2 in the first quadrant of the (x, u) -plane.

We use Figure 1 to visualize the minimizers of E_1 and E_2 . Three vertical lines $x = 0$, $x = \beta b_2$, and $x = \alpha b_2(a_2\beta + 1)$, and two horizontal lines $u = 0$ and $u = \alpha b_2$ partition the first quadrant into six rectangular regions (I to VI). The solid red line is the graph of $s_1(x)$ while the dashed blue line is the graph of $s_2(x)$.

We know $E_1(x, 0) \leq E_1(x, u)$ in region I and $E_2(x, \alpha b_2) \leq E_2(x, u)$ in region II, so $E_1(x, 0) < E_2(x, \alpha b_2)$ by Equation (15) for $0 \leq x \leq \beta b_2$. We observe $E_1(x, s_1(x)) \leq E_1(x, u)$ in region III and $E_2(x, \alpha b_2) \leq E_2(x, u)$ in region IV, so $E_1(x, s_1(x)) < E_2(x, \alpha b_2)$ by Equation (15) for $\beta b_2 \leq x \leq \alpha b_2(a_2\beta + 1)$; Finally, we know $E_1(x, \alpha b_2) \leq E_1(x, u)$ in region V and $E_2(x, s_2(x)) \leq E_2(x, u)$ in region VI, so $E_2(x, s_2(x)) < E_1(x, \alpha b_2)$ by Equation (15) for $x > \alpha b_2(a_2\beta + 1)$. Thus $\text{prox}_{\beta f_\alpha}$ is given by (18). \square

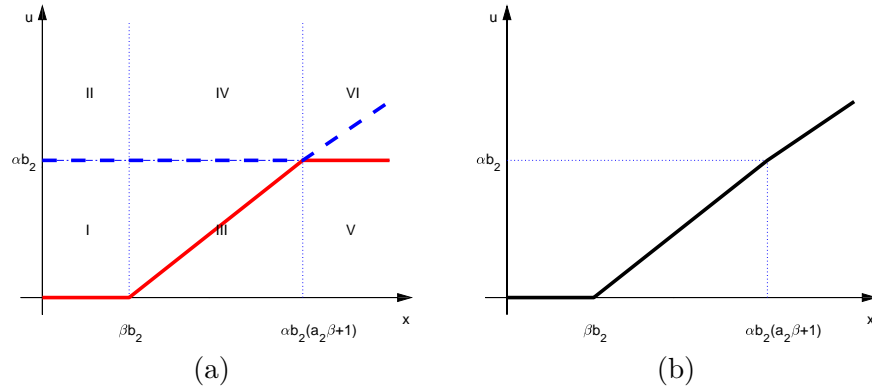


Figure 1: An illustration of case (i): $b_2 > 0$ and $\alpha b_2(a_2\beta + 1) > \beta b_2$. The graphs of (a) $s_1(x)$ (solid) and $s_2(x)$ (dashed) and (b) the resulting proximity operator $\text{prox}_{\beta f_\alpha}(x)$.

Next result is for case (ii).

Lemma 9. Let f be a piecewise quadratic sparsity promoting function as defined by (Q). If $b_2 > 0$ and $\alpha b_2(a_2\beta + 1) = \beta b_2$, then

$$\text{prox}_{\beta f_\alpha}(x) = \begin{cases} 0, & \text{if } 0 \leq x < \beta b_2; \\ [0, \alpha b_2], & \text{if } x = \beta b_2; \\ \frac{\alpha a_2 + 1}{\alpha a_2(a_2\beta + 1) + 1} \left(x - \frac{\alpha a_2 \beta b_2}{\alpha a_2 + 1} \right), & \text{if } x > \beta b_2. \end{cases} \quad (19)$$

Proof. Similar to the proof of Lemma 8, we first give the explicit form of the set $s_1(x)$:

$$s_1(x) = \begin{cases} 0, & \text{if } 0 \leq x < \beta b_2; \\ [0, \alpha b_2], & \text{if } x = \beta b_2; \\ \alpha b_2, & \text{if } x > \beta b_2. \end{cases}$$

We note that $\text{prox}_{\beta f_\alpha}$ can be set-valued only at βb_2 .

In Figure 2, two vertical lines $x = 0$ and $x = \beta b_2$, and two horizontal lines $u = 0$ and $u = \alpha b_2$ partition the first quadrant into four rectangular regions (I to IV). The solid red line is the graph of $s_1(x)$ while the dashed blue line is the graph of $s_2(x)$. It is identical to Figure 1 with the middle regions collapsed to a line. Following the same reasoning as in Lemma 8, we see that (19) holds. \square

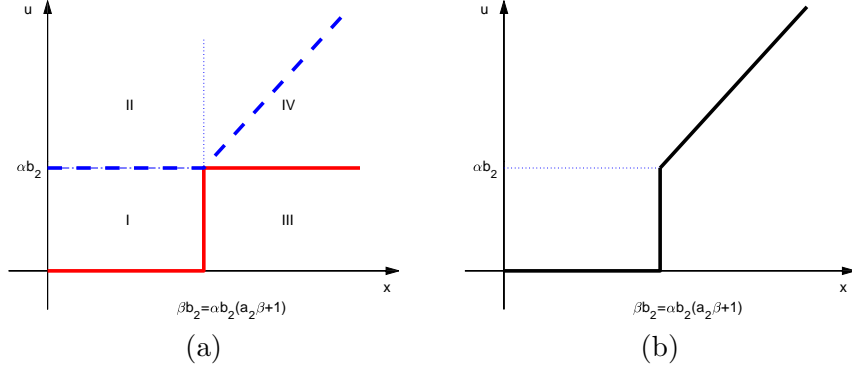


Figure 2: An illustration of case (ii): $b_2 > 0$ and $\alpha b_2(a_2\beta + 1) = \beta b_2$. The graphs of (a) $s_1(x)$ (solid) and $s_2(x)$ (dashed) and (b) the resulting proximity operator $\text{prox}_{\beta f_\alpha}(x)$.

Finally, we consider case (iii). Because βb_2 and $\alpha b_2(a_2\beta + 1)$ have now switched positions, we see that we must take care when dealing with the intermediate x values.

Lemma 10. *Let f be a piecewise quadratic sparsity promoting function as defined by (\mathcal{Q}) . Define*

$$\tau^+ = \frac{\alpha a_2 \beta b_2}{\alpha a_2 + 1} + \frac{\sqrt{\alpha \beta (\alpha a_2^2 \beta + \alpha a_2 + 1) b_2}}{\alpha a_2 + 1}.$$

If $b_2 > 0$ and $\alpha b_2(a_2\beta + 1) < \beta b_2$,

$$\text{prox}_{\beta f_\alpha}(x) = \begin{cases} 0, & \text{if } 0 \leq x < \tau^+; \\ \left\{ 0, \frac{\alpha a_2 + 1}{\alpha a_2(a_2\beta + 1) + 1} \left(\tau^+ - \frac{\alpha a_2 \beta b_2}{\alpha a_2 + 1} \right) \right\}, & \text{if } x = \tau^+; \\ \frac{\alpha a_2 + 1}{\alpha a_2(a_2\beta + 1) + 1} \left(x - \frac{\alpha a_2 \beta b_2}{\alpha a_2 + 1} \right), & \text{if } x > \tau^+. \end{cases} \quad (20)$$

Proof. Again, we first give the explicit form of the set $s_1(x)$. Note that $E_1(x, \cdot)$ is concave in this case, so the minimum occurs at the endpoints according to the position of the vertex. Thus,

$$s_1(x) = \begin{cases} 0, & \text{if } 0 \leq x < \frac{1}{2}(\alpha b_2(a_2\beta + 1) + \beta b_2); \\ \{0, \alpha b_2\}, & \text{if } x = \frac{1}{2}(\alpha b_2(a_2\beta + 1) + \beta b_2); \\ \alpha b_2, & \text{if } x > \frac{1}{2}(\alpha b_2(a_2\beta + 1) + \beta b_2). \end{cases}$$

This is set-valued at $\frac{1}{2}(\alpha b_2(a_2\beta + 1) + \beta b_2)$.

As before, we plot $s_1(x)$ and $s_2(x)$ in Figure 3. Three vertical lines $x = 0$, $x = \alpha b_2(a_2\beta + 1)$, and $x = \frac{1}{2}(\alpha b_2(a_2\beta + 1) + \beta b_2)$, and two horizontal lines $u = 0$ and $u = \alpha b_2$ partition the first quadrant into six rectangular regions as shown in Figure 3(a). The solid red line is the graph of $s_1(x)$ while the dashed blue line is the graph of $s_2(x)$. From this figure and (15), it is easy to see that regions I, II, V, and VI behave as in the previous cases. That is, $\text{prox}_{\beta f_\alpha}(x) = s_1(x)$ for $0 \leq x \leq \alpha b_2(a_2\beta + 1)$ and $\text{prox}_{\beta f_\alpha}(x) = s_2(x)$ for $x \geq \frac{1}{2}(\alpha b_2(a_2\beta + 1) + \beta b_2)$.

To find the expression of $\text{prox}_{\beta f_\alpha}(x)$ for $\alpha b_2(a_2\beta + 1) < x < \frac{1}{2}(\alpha b_2(a_2\beta + 1) + \beta b_2)$, from the solid red line and the dashed blue line in regions III and IV, we need to compare the value of $E_1(x, 0)$ with $E_2(x, s_2(x))$. Using (17), a direct computation gives

$$E_2(x, s_2(x)) - E_1(x, 0) = -\frac{\alpha a_2 + 1}{2\beta(\alpha a_2(a_2\beta + 1) + 1)} \left(x - \frac{\alpha a_2 \beta b_2}{\alpha a_2 + 1} \right)^2 + \frac{\alpha b_2^2}{2(\alpha a_2 + 1)}.$$

Notice that $E_2(x, s_2(x)) - E_1(x, 0) > 0$ at $x = \alpha b_2(a_2\beta + 1)$ and $E_2(x, s_2(x)) - E_1(x, 0) < 0$ at $x = \frac{1}{2}(\alpha b_2(a_2\beta + 1) + \beta b_2)$. Hence, the quadratic polynomial $E_2(x, s_2(x)) - E_1(x, 0)$ has only one root at τ^+ that is between $\alpha b_2(a_2\beta + 1)$ and $\frac{1}{2}(\alpha b_2(a_2\beta + 1) + \beta b_2)$. So, the result of this lemma holds and is illustrated in Figure 3(c). \square

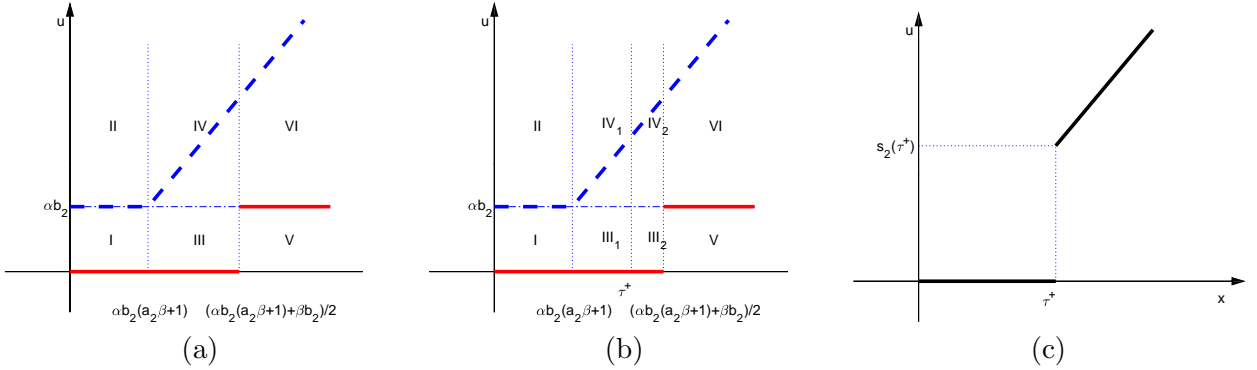


Figure 3: An illustration of case (iii): $b_2 > 0$ and $\alpha b_2(a_2\beta + 1) < \beta b_2$. The graphs of (a), (b) $s_1(x)$ (solid) and $s_2(x)$ (dashed) and (c) the resulting proximity operator $\text{prox}_{\beta f_\alpha}(x)$.

With the above results, we know $\text{prox}_{\beta f_\alpha}(x)$ for $x \geq 0$. The following lemma extends these results to $x \leq 0$.

Lemma 11. *Let f be a piecewise quadratic sparsity promoting function as defined by (Q). Define $g : x \mapsto f(-x)$. Then for $x \leq 0$ and any positive numbers α and β , we have $\text{prox}_{\beta f_\alpha}(x) = -\text{prox}_{\beta g_\alpha}(-x)$ where $\text{prox}_{\beta g_\alpha}(-x)$ can be evaluated using the results in Lemmas 7-10.*

Proof. Since f is sparsity promoting, so is g by Theorem 1. Moreover, $f_\alpha = g_\alpha(-\cdot)$ which leads to $\text{prox}_{\beta f_\alpha}(x) = -\text{prox}_{\beta g_\alpha}(-x)$ for all x . Note that

$$g(x) = \begin{cases} \frac{1}{2}a_2x^2 - b_2x, & \text{if } x \leq 0; \\ \frac{1}{2}a_1x^2 - b_1x, & \text{if } x \geq 0, \end{cases}$$

which is a piecewise quadratic sparsity promoting function. All results developed in Lemmas 7-10 can be applied for g . Therefore, the results of this lemma follow immediately. \square

In summary, we have the following result.

Theorem 3. If $f \in \Gamma_0(\mathbb{R})$ is a quadratic sparsity promoting function as defined by (Q), then the following statements hold.

- (i) $\text{prox}_{\beta f_\alpha}$ is set-valued at at most one point at each side of the origin. Moreover, $\text{prox}_{\beta f_\alpha}$ is piecewise linear on any interval not containing these possible set-valued points.
- (ii) For any $p \in \text{prox}_{\beta f_\alpha}(x)$, $|p| \leq |x|$. Furthermore, $\text{sgn}(p) = \text{sgn}(x)$ if both p and x are nonzero.

Proof. All results follows directly from the expressions of $\text{prox}_{\beta f_\alpha}(x)$ given in Lemma 7-Lemma 11. \square

Remark 5. Theorem 3 guarantees that $\text{prox}_{\beta f_\alpha}$ will be a thresholding operator for any f_α given by (Q_α). Furthermore, Lemmas 7-10 provide detailed and easily customizable forms which can be tailored to applications.

5.3 Piecewise Quadratic on Intervals

Let C be a closed interval containing the origin and f a piecewise quadratic function defined by (Q). We consider a function \tilde{f} that is the restriction of f on the interval C as follows:

$$\tilde{f} = f + \iota_C. \quad (\tilde{Q})$$

Lemma 12. Let f be a piecewise quadratic sparsity promoting function defined by (Q) and let C be a closed interval on \mathbb{R} such that $\{0\} \subsetneq \partial f(0) \cap C$. Then \tilde{f} defined in (\tilde{Q}) is a sparsity promoting function. Moreover,

$$\tilde{f}_\alpha = f_\alpha + \iota_C. \quad (\tilde{Q}_\alpha)$$

Proof. Since f is sparsity promoting, $\min_{x \in \mathbb{R}} f(x) = f(0) = 0$. Because $\{0\} \subsetneq \partial f(0) \cap C$, we know that $\tilde{f}(0) = \min_{x \in C} f(x) = \min_{x \in \mathbb{R}} f(x) = 0$. That is, \tilde{f} achieves its minimum at the origin. We further know that $\partial f(0) = \partial \tilde{f}(0)$, hence $\{0\} \subsetneq \partial f(0) \cap C = \partial \tilde{f}(0) \cap C$. Therefore, \tilde{f} is sparsity promoting.

By Lemma 1 and Lemma 2, $\text{prox}_{\alpha f}(x) \in C$ if $x \in C$. This indicates that for $x \in C$

$$\text{env}_{\alpha f}(x) = \min_{u \in \mathbb{R}} \left\{ f(u) + \frac{1}{2\alpha}(u - x)^2 \right\} = \min_{u \in C} \left\{ f(u) + \frac{1}{2\alpha}(u - x)^2 \right\} = \text{env}_{\alpha \tilde{f}}(x).$$

The above identities yield $\tilde{f}_\alpha = f_\alpha + \iota_C$. This completes the proof of the result. \square

By the above lemma, for \tilde{f} defined in (\tilde{Q}) we always assume that the coefficients in f satisfy (11) and that $C = [\lambda_1, \lambda_2]$ with $\lambda_1 \leq 0 \leq \lambda_2$ and $\lambda_2 - \lambda_1 > 0$.

Theorem 4. Let \tilde{f} be defined in (\tilde{Q}) , let $x \in \mathbb{R}$, and let α and β be two positive numbers. Then the following statements hold.

- (i) If the set $\text{prox}_{\beta f_\alpha}(x) \cap C$ is not empty, then $\text{prox}_{\beta f_\alpha}(x) \cap C \subseteq \text{prox}_{\beta \tilde{f}_\alpha}(x)$;
- (ii) If $\lambda_2 \in \text{prox}_{\beta \tilde{f}_\alpha}(x)$, then $\lambda_2 \in \text{prox}_{\beta \tilde{f}_\alpha}(y)$ for all $y > x$;
- (iii) If $\lambda_1 \in \text{prox}_{\beta \tilde{f}_\alpha}(x)$, then $\lambda_1 \in \text{prox}_{\beta \tilde{f}_\alpha}(y)$ for all $y < x$;

Proof. (i): Assume p is an element in $\text{prox}_{\beta f_\alpha}(x) \cap C$. We have

$$\begin{aligned} f_\alpha(p) + \frac{1}{2\beta}(p-x)^2 &= \min_{u \in \mathbb{R}} \left\{ f_\alpha(u) + \frac{1}{2\beta}(u-x)^2 \right\} \\ &= \min_{u \in C} \left\{ f_\alpha(u) + \frac{1}{2\beta}(u-x)^2 \right\} \\ &= \min_{u \in \mathbb{R}} \left\{ \tilde{f}_\alpha(u) + \frac{1}{2\beta}(u-x)^2 \right\}, \end{aligned}$$

where the first equation is due to $p \in \text{prox}_{\beta f_\alpha}(x)$, the second equation is due to $p \in C$, the last one is due to Lemma 12, hence, $p \in \text{prox}_{\beta \tilde{f}_\alpha}(x)$.

(ii): Since $\lambda_2 \geq 0$, the inclusion $\lambda_2 \in \text{prox}_{\beta \tilde{f}_\alpha}(x)$ together with Lemma 2 implies that $x \geq 0$ and for all $u \in [\lambda_1, \lambda_2]$,

$$\tilde{f}_\alpha(u) + \frac{1}{2\beta}(u-x)^2 \geq \tilde{f}_\alpha(\lambda_2) + \frac{1}{2\beta}(\lambda_2-x)^2.$$

With the above inequality, when $y > x$, we have that

$$\begin{aligned} \tilde{f}_\alpha(\lambda_2) + \frac{1}{2\beta}(\lambda_2-y)^2 &= \tilde{f}_\alpha(\lambda_2) + \frac{1}{2\beta}(\lambda_2-x)^2 + \frac{1}{2\beta}(y-x)(y+x-2\lambda_2) \\ &\leq \tilde{f}_\alpha(u) + \frac{1}{2\beta}(u-x)^2 + \frac{1}{2\beta}(y-x)(y+x-2u) \\ &= \tilde{f}_\alpha(u) + \frac{1}{2\beta}(u-y)^2 \end{aligned}$$

hold for all $u \in [\lambda_1, \lambda_2]$. This yields $\lambda_2 \in \text{prox}_{\beta \tilde{f}_\alpha}(y)$.

(iii): The proof is similar to (ii). □

Theorem 4 tells us that the for \tilde{f} as in (\tilde{Q}) , $\text{prox}_{\beta \tilde{f}_\alpha}$ will resemble the proximity operator of f_α around the origin and the proximity operator of ι_C elsewhere. Due to the number of parameters, there are a huge number of possible combinations. Rather than list all of the combinations here, we provide the details for a specific function in Example 4 of Section 6.

We have shown that sparsity promoting quadratic and indicator functions have thresholding proximity operators. The results essentially rely on the fact that $\text{env}_{\alpha f}$ is quadratic for these functions. In fact, quadratic and indicator functions are the only ones with this property [18], so our discussion is a comprehensive method for obtaining thresholding rules.

6 Examples

In this section, we illustrate our theory by presenting several examples that are of practical interest.

For the first example, we collect and expand upon the previous discussion of $f(x) = \|x\|_1 = \sum_{i=1}^n |x_i|$ for $x \in \mathbb{R}^n$. The ℓ_1 -norm has been extensively used in myriad applications for promoting sparsity.

The second example is the ReLU (Rectified Linear Unit) function. It is the most commonly used activation function in convolutional neural networks or deep learning. The ReLU function on \mathbb{R}^n is defined as follows: $f(x) = \sum_{i=1}^n \max\{0, x_i\}$, where $x \in \mathbb{R}^n$.

The third example is the elastic net penalty function which is widely used in statistics (see [26]). The general form of the elastic net is the linear combination of the ℓ_1 and ℓ_2 norms as follows:

$f(x) = \frac{\lambda_1}{2}\|x\|^2 + \lambda_2\|x\|_1$, where λ_1 and λ_2 are two nonnegative parameters. In our discussion, we will simply choose $\lambda_1 = \lambda_2 = 1$. This is known as the naive elastic net.

The last example is similar to the first one, but restricted to a cube centered at the origin. The function f is given as follows: $f(x) = \|x\|_1 + \iota_C(x)$, where $C = [-\lambda, \lambda]^n$. Generally speaking, this function promotes the sparsity on C .

We notice that the function f in the above four examples can be written as

$$f(x) = \sum_{i=1}^n g(x_i)$$

for $x \in \mathbb{R}^n$ and some specific function g . For example, g is $|\cdot|$, $\max\{0, \cdot\}$, $\frac{1}{2}|\cdot|^2 + |\cdot|$, or $|\cdot| + \iota_{[-\lambda, \lambda]}$, in examples 1, 2, 3, or 4, an analogue of f when \mathbb{R}^n reduces to \mathbb{R} . We further have that $\text{prox}_{\alpha f}(x) = \text{prox}_{\alpha g}(x_1) \times \text{prox}_{\alpha g}(x_2) \times \cdots \times \text{prox}_{\alpha g}(x_n)$, $\text{env}_{\alpha f}(x) = \sum_{i=1}^n \text{env}_{\alpha g}(x_i)$, $\text{prox}_{\beta f_{\alpha}}(x) = \text{prox}_{\beta g_{\alpha}}(x_1) \times \text{prox}_{\beta g_{\alpha}}(x_2) \times \cdots \times \text{prox}_{\beta g_{\alpha}}(x_n)$, and $\text{env}_{\beta f_{\alpha}}(x) = \sum_{i=1}^n \text{env}_{\beta g_{\alpha}}(x_i)$. Therefore, in the following discussion we will restrict ourself on $n = 1$.

6.1 Example 1: The absolute value function

The first example is the absolute value function $f : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto |x|$, which is a special case of the piecewise quadratic function in (\mathcal{Q}) with $a_1 = a_2 = 0$, $b_1 = -1$, and $b_2 = 1$. This function is nondifferentiable at the origin.

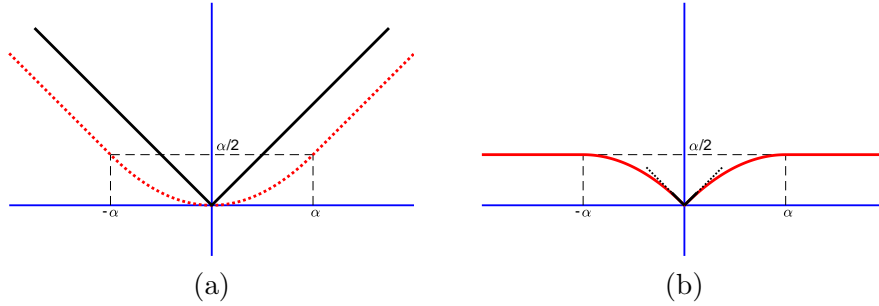


Figure 4: Example 1. (a) The graphs of f (solid), $\text{env}_{\alpha f}$ (dotted), and (b) the graph of $f_{\alpha} = f(x) - \text{env}_{\alpha f}(x)$. Near the origin f_{α} retains the structure of f , which is emphasized in black (solid-dotted).

The proximity operator and the Moreau envelope of f with parameter $\alpha > 0$ are

$$\text{prox}_{\alpha|\cdot|}(x) = \text{sgn}(x) \max\{0, |x| - \alpha\} \quad \text{and} \quad \text{env}_{\alpha|\cdot|}(x) = \begin{cases} \frac{1}{2\alpha}x^2, & \text{if } |x| \leq \alpha; \\ |x| - \frac{1}{2}\alpha, & \text{otherwise,} \end{cases}$$

respectively. It is well known that $\text{prox}_{\alpha|\cdot|}$ is called the soft thresholding in literature of wavelet [8] and $\text{env}_{\alpha|\cdot|}$ is Huber's function in robust statistics [14]. Figure 5 shows the typical shape of the proximity operator of f .

As defined in (\mathcal{F}_{α}) , for the absolute value function f ,

$$f_{\alpha}(x) := |x| - \text{env}_{\alpha|\cdot|}(x) = \begin{cases} |x| - \frac{1}{2\alpha}x^2, & \text{if } |x| \leq \alpha; \\ \frac{1}{2}\alpha, & \text{otherwise.} \end{cases}$$

This function f_{α} (see Figure 4(b)) is identical to the minimax convex penalty (MCP) function given in [25], but motivated from statistic perspective.

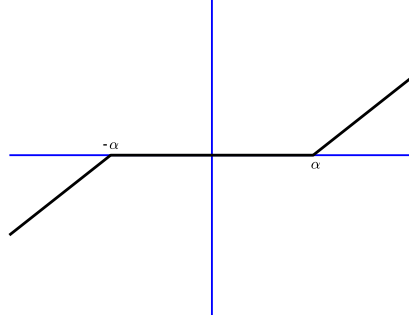


Figure 5: Example 1. The typical shape of $\text{prox}_{\alpha f}$.

The expression of $\text{prox}_{\beta f_{\alpha}}$ depends on the relative values of α and β . If $\beta < \alpha$, Lemma 8 gives

$$\text{prox}_{\beta f_{\alpha}}(x) = \begin{cases} 0, & \text{if } |x| \leq \beta; \\ \frac{\alpha}{\alpha - \beta}(|x| - \beta)\text{sgn}(x), & \text{if } \beta < |x| \leq \alpha; \\ x, & \text{if } |x| \geq \alpha. \end{cases} \quad (21)$$

This is the firm thresholding operator [12]. If $\beta = \alpha$, Lemma 9 gives

$$\text{prox}_{\beta f_{\alpha}}(x) = \begin{cases} 0, & \text{if } |x| < \alpha; \\ [0, \alpha], & \text{if } |x| = \alpha; \\ x, & \text{if } |x| > \alpha, \end{cases} \quad (22)$$

Finally, if $\beta > \alpha$, Lemma 10 gives

$$\text{prox}_{\beta f_{\alpha}}(x) = \begin{cases} 0, & \text{if } |x| < \sqrt{\alpha\beta}; \\ \{0, x\}, & \text{if } |x| = \sqrt{\alpha\beta}; \\ x, & \text{if } |x| > \sqrt{\alpha\beta}; \end{cases} \quad (23)$$

The proximity operator $\text{prox}_{\beta f_{\alpha}}$ for different values of α and β is plotted in Figure 6.

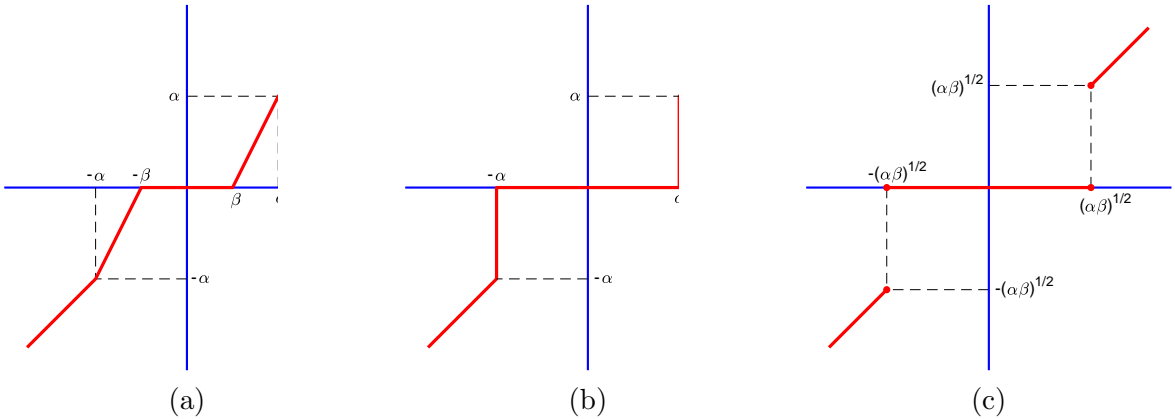


Figure 6: Typical shapes of the proximity operator of $|\cdot|_{\alpha}$ for (a) $\beta < \alpha$, (b) $\beta = \alpha$, (c) $\beta > \alpha$. The sparsity threshold and the thresholding behavior depend on the relationship between α and β .

To end this example, we give several remarks on the proximity operators of $\text{prox}_{\alpha f}$ and $\text{prox}_{\beta f_{\alpha}}$ as follows:

- Note that $\partial f(0) = [-1, 1]$. The results given in (21) (for $\beta < \alpha$) and (22) (for $\beta = \alpha$) exactly match the first two statements of Theorem 2. For $\beta > \alpha$, the $\text{prox}_{\beta f_\alpha}(x) = 0$ for all $x \in [-\sqrt{\alpha\beta}, \sqrt{\alpha\beta}]$ which includes the interval $[-\alpha, \alpha] = \alpha\partial f(0)$ as indicated in the third statement of Theorem 2.
- The operator $\text{prox}_{\alpha f}$ forces its variable to zero when the absolute value is less than a given threshold, and otherwise reduces the variable, in absolute value, by the amount of the threshold. Like $\text{prox}_{\alpha f}$, $\text{prox}_{\beta f_\alpha}$ forces its variable to zero when the absolute value is less than a given threshold, but it fixes variables whose absolute value exceeds a certain threshold.
- For $\beta \geq \alpha$ the proximity operator $\text{prox}_{\beta f_\alpha}$ is almost identical to the hard threshold operator. Let $|\cdot|_0$ be the ℓ_0 “norm” on \mathbb{R} , that is, $|x|_0$ equals 1 if x is nonzero, 0 otherwise. The proximity operator of $|\cdot|_0$ with parameter γ at x is

$$\text{prox}_{\gamma|\cdot|_0}(x) = \begin{cases} \{0\}, & \text{if } |x| < \sqrt{2\gamma}; \\ \{0, x\}, & \text{if } |x| = \sqrt{2\gamma}; \\ \{x\}, & \text{if } |x| > \sqrt{2\gamma}, \end{cases}$$

which is called the hard thresholding operator with threshold $\sqrt{2\gamma}$. We can see that $\text{prox}_{\gamma|\cdot|_0} = \text{prox}_{\beta f_\alpha}$ as long as $2\gamma = \alpha\beta$ and $\beta > \alpha$. It is interesting that although $|\cdot|_0$ is discontinuous and f_α is continuous, they have the same proximity operator. Moreover, by fixing α and varying the parameter β , the proximity operator $\text{prox}_{\beta f_\alpha}$ changes from the firm thresholding operator to the hard thresholding operator.

6.2 Example 2: ReLU function

The ReLU (Rectified Linear Unit) function on \mathbb{R} is

$$f(x) := \max\{0, x\},$$

which is a special case of the piecewise quadratic function in (\mathcal{Q}) with $a_1 = b_1 = a_2 = 0$ and $b_2 = 1$. The proximity operator and the Moreau envelope of f with parameter $\alpha > 0$ are

$$\text{prox}_{\alpha f}(x) = \min\{x, \max\{0, x - \alpha\}\} \quad \text{and} \quad \text{env}_{\alpha f}(x) = \begin{cases} 0 & \text{if } x \leq 0; \\ \frac{1}{2\alpha}x^2, & \text{if } 0 \leq x \leq \alpha; \\ x - \frac{1}{2}\alpha, & \text{if } x \geq \alpha, \end{cases}$$

respectively. By (\mathcal{F}_α) , $f_\alpha(x) = f(x) - \text{env}_{\alpha f}(x)$ is

$$f_\alpha(x) = \begin{cases} 0, & \text{if } x < 0 \\ x - \frac{1}{2\alpha}x^2, & \text{if } 0 \leq x \leq \alpha \\ \frac{\alpha}{2}, & \text{if } x > \alpha \end{cases}$$

Figure 7(a) depicts the graphs of f and $\text{env}_{\alpha f}$ while Figure 7(b) presents the function f_α . The graph of $\text{prox}_{\alpha f}$ is given in Figure 8.

As in example 1, the expression of $\text{prox}_{\beta f_\alpha}$ depends on the relative values of α and β . If $\beta < \alpha$,

$$\text{prox}_{\beta f_\alpha}(x) = \begin{cases} x, & \text{if } x \leq 0 \text{ or } x \geq \alpha; \\ 0, & \text{if } 0 \leq x \leq \beta; \\ \frac{\alpha(x-\beta)}{\alpha-\beta}, & \text{if } \beta \leq x \leq \alpha. \end{cases} \quad (24)$$

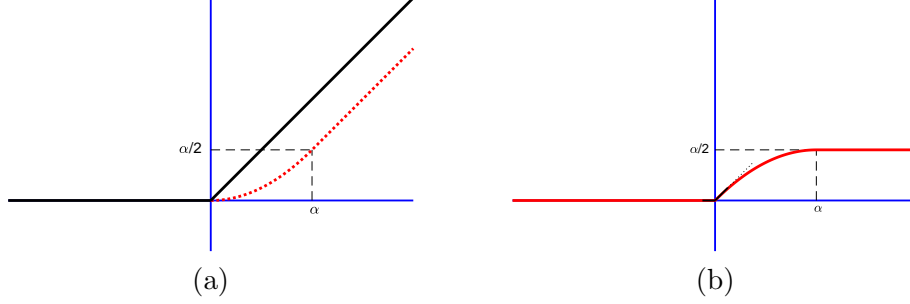


Figure 7: Example 2. (a) The graphs of f (solid), $\text{env}_\alpha f$ (dotted), and (b) their difference $f_\alpha = f - \text{env}_\alpha f$. The singularity of f_α at zero is emphasized in black (solid-dotted).

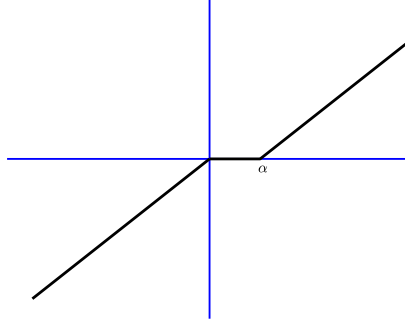


Figure 8: Example 2. The typical shape of $\text{prox}_\alpha f$. The parameter α is the sparsity threshold.

If $\beta = \alpha$,

$$\text{prox}_{\beta f_\alpha}(x) = \begin{cases} x, & \text{if } x \leq 0 \text{ or } x > \alpha; \\ 0, & \text{if } 0 \leq x < \alpha; \\ [0, \alpha] & \text{if } x = \alpha. \end{cases} \quad (25)$$

Finally, if $\beta > \alpha$,

$$\text{prox}_{\beta f_\alpha}(x) = \begin{cases} x, & \text{if } x \leq 0 \text{ or } x > \sqrt{\alpha\beta}; \\ 0, & \text{if } 0 \leq x < \sqrt{\alpha\beta}; \\ \{0, \sqrt{\alpha\beta}\}, & \text{if } x = \sqrt{\alpha\beta}. \end{cases} \quad (26)$$

Note that $\partial f(0) = [0, 1]$. The results given in (24) (for $\beta < \alpha$) and (25) (for $\beta = \alpha$) exactly match the first two statements of Theorem 2. For $\beta > \alpha$, equation (26) shows that $\text{prox}_{\beta f_\alpha}(x) = 0$ for all $x \in [0, \sqrt{\alpha\beta}]$, which includes the interval $[0, \alpha] = \alpha \partial f(0)$ as indicated in the third statement of Theorem 2.

6.3 Example 3: Elastic Net

The elastic net is a regularized regression method in data analysis that linearly combines the ℓ_1 and ℓ_2 penalties of the LASSO and ridge methods. In this example, we consider a special case of the elastic net in \mathbb{R} :

$$f(x) = \frac{1}{2}x^2 + |x|.$$

This is an instance of the piecewise quadratic function given in (\mathcal{Q}) with $a_1 = a_2 = 1$, $b_1 = -1$ and $b_2 = 1$. Clearly, f is nondifferentiable at the origin with $\text{argmin}_{x \in \mathbb{R}} f(x) = \{0\}$. Moreover,

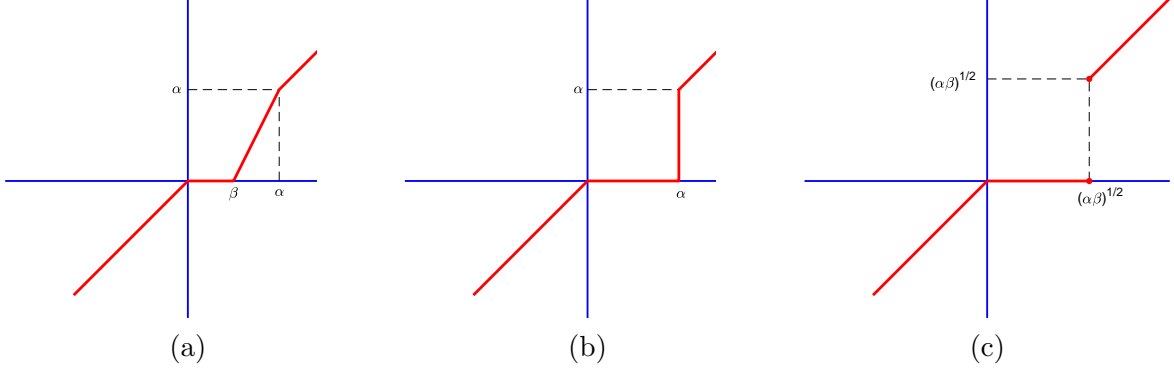


Figure 9: Example 2. Typical shapes of the proximity operator of f_α for (a) $\beta < \alpha$; (b) $\beta = \alpha$; and (c) $\beta > \alpha$.

$$\partial f(0) = \partial |\cdot| (0) = [-1, 1].$$

The proximity operator and the Moreau envelope of f with parameter $\alpha > 0$ are

$$\text{prox}_{\alpha f}(x) = \max \left\{ 0, \frac{1}{\alpha + 1}(|x| - \alpha) \right\} \text{sgn}(x) \quad \text{and} \quad \text{env}_{\alpha f}(x) = \begin{cases} \frac{1}{2\alpha}x^2, & \text{if } |x| \leq \alpha; \\ \frac{1}{\alpha+1}(\frac{1}{2}x^2 + |x| - \frac{\alpha}{2}), & \text{if } |x| \geq \alpha, \end{cases}$$

respectively.

The graphs of f and $\text{env}_{\alpha f}$ are plotted in Figure 10 (a). The graph of $\text{prox}_{\alpha f}$ is plotted in Figure 10 (b). As in the case of the absolute value function, $\text{prox}_{\alpha f}$ sends all values between α and $-\alpha$ to zero. Unlike the absolute value, it also contracts elements outside of this interval toward the origin.

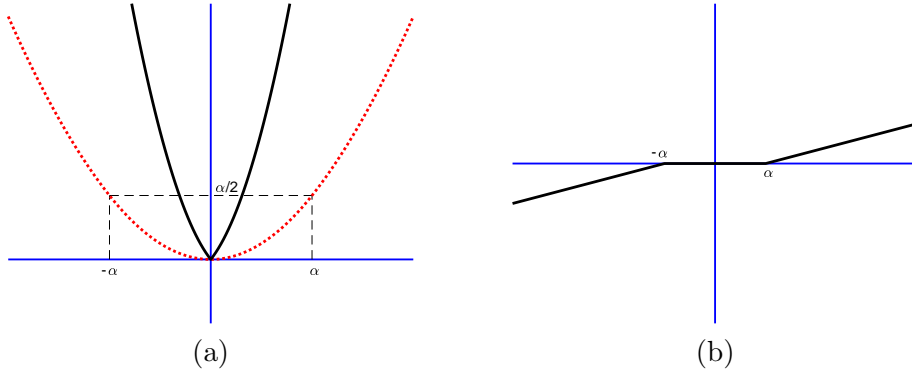


Figure 10: Example 3. (a) The graphs of f (solid) and $\text{env}_{\alpha f}$ (dotted); and (b) the graph of $\text{prox}_{\alpha f}$.

Now f_α , the difference between f and its Moreau envelope $\text{env}_{\alpha f}$, is

$$f_\alpha(x) = \begin{cases} \frac{\alpha-1}{2\alpha}x^2 + |x|, & \text{if } |x| \leq \alpha; \\ \frac{\alpha}{2(\alpha+1)}x^2 + \frac{\alpha}{\alpha+1}|x| + \frac{\alpha}{2(\alpha+1)}, & \text{if } |x| \geq \alpha. \end{cases}$$

We remark that f_α is convex when $\alpha \geq 1$ and nonconvex when $\alpha < 1$. The graph of f_α for $\alpha \geq 1$ and $\alpha < 1$ are shown in Figure 11(a) and (b), respectively.

According to the discussion given in subsection 5.2, we consider three cases: $\beta(\alpha - 1) + \alpha > 0$, $\beta(\alpha - 1) + \alpha = 0$, and $\beta(\alpha - 1) + \alpha < 0$. These cases are equivalent to $\alpha(\beta + 1) > \beta$, $\alpha(\beta + 1) = \beta$,

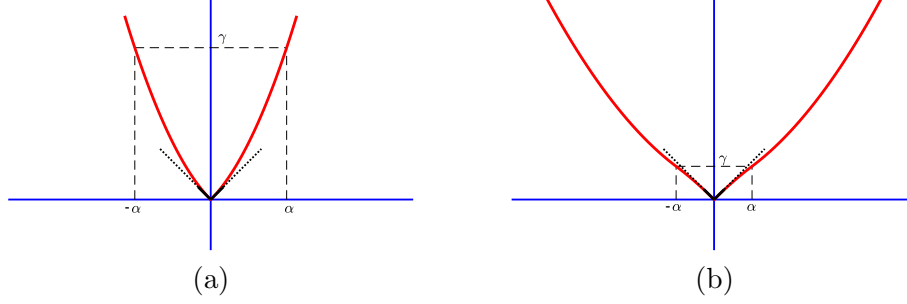


Figure 11: Example 3. The graph of f_α when (a) $\alpha \geq 1$ and (b) $\alpha < 1$. The singularity of f_α at zero is emphasized in black (solid-dotted).

and $\alpha(\beta + 1) < \beta$ respectively. Recall that these cases correspond to the convexity (or lack thereof) of $f_\alpha(u) + \frac{1}{2\beta}(u - x)^2$ for u close to zero.

Case 1: $\alpha(\beta + 1) > \beta$. In this case, by Lemma 8 we have

$$\text{prox}_{\beta f_\alpha}(x) = \begin{cases} 0 & \text{if } |x| \leq \beta; \\ \frac{\alpha}{\alpha\beta - \beta + \alpha}(x - \beta \text{sgn}(x)) & \text{if } \beta \leq |x| \leq \alpha(\beta + 1); \\ \frac{\alpha + 1}{\alpha\beta + \alpha + 1}(x - \frac{\alpha\beta}{\alpha + 1} \text{sgn}(x)) & \text{if } \alpha(\beta + 1) \leq |x|. \end{cases} \quad (27)$$

Case 2: $\alpha(\beta + 1) = \beta$. By Lemma 9 we have

$$\text{prox}_{\beta f_\alpha}(x) = \begin{cases} 0 & \text{if } |x| \leq \beta; \\ [0, \alpha] \text{sgn}(x) & \text{if } |x| = \beta; \\ \frac{\alpha + 1}{\alpha\beta + \alpha + 1}(x - \frac{\alpha\beta}{\alpha + 1} \text{sgn}(x)) & \text{if } \beta \leq |x|. \end{cases} \quad (28)$$

Case 3: $\alpha(\beta + 1) < \beta$. Define

$$\tau = \frac{\alpha\beta}{\alpha + 1} + \frac{\sqrt{\alpha\beta(\alpha\beta + \alpha + 1)}}{\alpha + 1}. \quad (29)$$

as in Lemma 10. Then we have

$$\text{prox}_{\beta f_\alpha}(x) = \begin{cases} 0 & \text{if } |x| \leq \tau; \\ \{0, \omega\} & \text{if } |x| = \tau; \\ \frac{(\alpha + 1)x - \alpha\beta \text{sgn}(x)}{\alpha\beta + \alpha + 1}, & \text{if } |x| > \tau, \end{cases} \quad (30)$$

where $\omega = \frac{(\alpha + 1)\tau - \alpha\beta}{\alpha\beta + \alpha + 1}$. The graphs of $\text{prox}_{\beta f_\alpha}$ in the above three cases are plotted in Figure 6.3.

Below are some comments on this example.

- The function f_α in the first two examples is nonconvex for any $\alpha > 0$, however, by Proposition 2 it is convex if $\alpha \geq 1$ due to our elastic net function f being 1-strongly convex.
- The computation of the proximity operator $\text{prox}_{\beta f_\alpha}$ is discussed under three different situations, namely, $\alpha(\beta + 1) > \beta$, $\alpha(\beta + 1) = \beta$, and $\alpha(\beta + 1) < \beta$. These situations are quite nature from Proposition 2. Since f is 1-strongly convex, hence, the function $f_\alpha + \frac{1}{2\beta}(\cdot - x)^2$ is $(1 + \beta^{-1} - \alpha^{-1})$ -strongly convex if $\alpha(\beta + 1) > \beta$, convex if $\alpha(\beta + 1) = \beta$, and $(\alpha^{-1} - 1 - \beta^{-1})$ -semiconvex if $\alpha(\beta + 1) < \beta$.

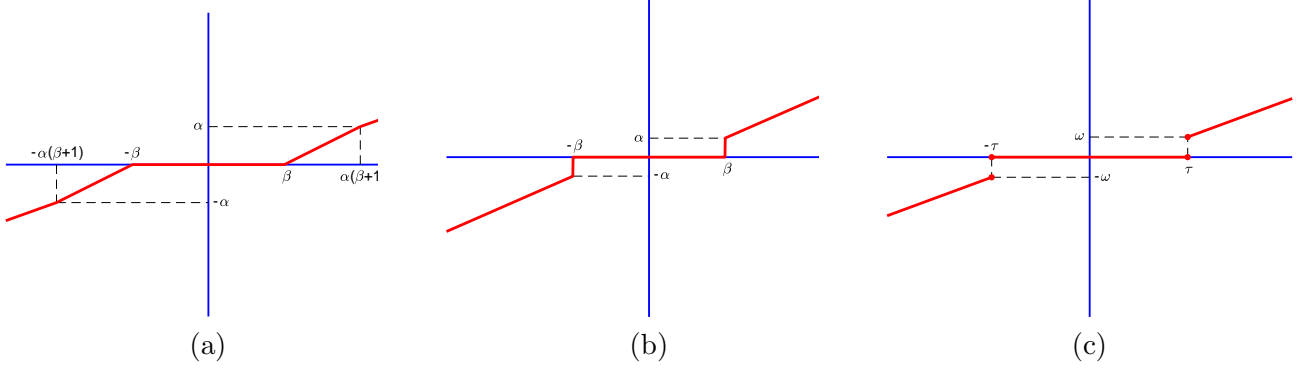


Figure 12: Example 3. Typical shapes of $\text{prox}_{\beta f_\alpha}$ when (a) $\alpha(\beta + 1) > \beta$, (b) $\alpha(\beta + 1) = \beta$, and (c) $\alpha(\beta + 1) < \beta$.

- For the case of $\beta \leq \alpha$, we know that $\alpha(1 + \beta) > \beta$, so the proximity operator given (27) covers both statements 1 and 2 in Theorem 2.
- For the case of $\beta > \alpha$, there are three possible related cases. If $\alpha < \beta < \alpha(\beta + 1)$ (resp. $\alpha < \beta = \alpha(\beta + 1)$), the proximity operator given (27) (resp. (28)) shows that this operator vanishes all elements in $\beta \partial f(0) = [-\beta, \beta] \supset \alpha \partial f(0)$, fulfilling the third statement of Theorem 2. If $\beta > \alpha(\beta + 1)$, we know that $\alpha < 1$, $\beta > \frac{\alpha}{1-\alpha}$, and τ defined in (29) satisfying

$$\tau = \frac{\alpha\beta}{\alpha+1} + \frac{\sqrt{\alpha\beta(\alpha\beta + \alpha + 1)}}{\alpha+1} > \frac{\alpha^2}{1-\alpha^2} + \frac{\alpha}{1-\alpha^2} > \alpha.$$

Hence, the proximity operator given (30) annihilates all elements in $\tau \partial f(0) \supset \alpha \partial f(0)$, once again fulfilling the third statement of Theorem 2.

6.4 Example 4: Absolute value on an interval centered at the origin

Let λ be a positive parameter. The absolute function on the interval $[-\lambda, \lambda]$ centered at the origin is

$$f(x) := |x| + \iota_{[-\lambda, \lambda]}(x),$$

which is a special case given in $(\tilde{\mathcal{Q}})$ with $a_1 = a_2 = 0$, $b_1 = -1$, $b_2 = 1$, and $C = [-\lambda, \lambda]$. Its proximity operator and Moreau envelope with parameter α at point x , respectively, are

$$\text{prox}_{\alpha f}(x) = \begin{cases} 0, & \text{if } |x| \leq \alpha; \\ \text{sgn}(x)(|x| - \alpha), & \text{if } \alpha < |x| \leq \alpha + \lambda; \\ \lambda \text{sgn}(x), & \text{if } \alpha + \lambda < |x|; \end{cases}$$

and

$$\text{env}_{\alpha f}(x) = \begin{cases} |x| - \frac{\alpha}{2} + \frac{1}{2\alpha}(|x| - \alpha)^2, & \text{if } |x| \leq \alpha; \\ |x| - \frac{\alpha}{2}, & \text{if } \alpha < |x| \leq \alpha + \lambda; \\ |x| - \frac{\alpha}{2} + \frac{1}{2\alpha}(|x| - (\lambda + \alpha))^2, & \text{if } \alpha + \lambda < |x|. \end{cases}$$

Figure 6.4 depicts the graphs of f , $\text{env}_{\alpha f}$, and $\text{prox}_{\alpha f}$. We observe that on the interval $[-\lambda, \lambda]$ (the domain of f_α) the envelope $\text{env}_{\alpha f}$ is piecewise quadratic polynomial (Figure 6.4(a)) if $\alpha < \lambda$ and is simply quadratic polynomial (Figure 6.4(b)) if $\alpha \geq \lambda$. It turns out that the expression of $\text{prox}_{\beta f_\alpha}$ for $\alpha < \lambda$ is much more complicated than that for $\alpha \geq \lambda$ as we will see below.

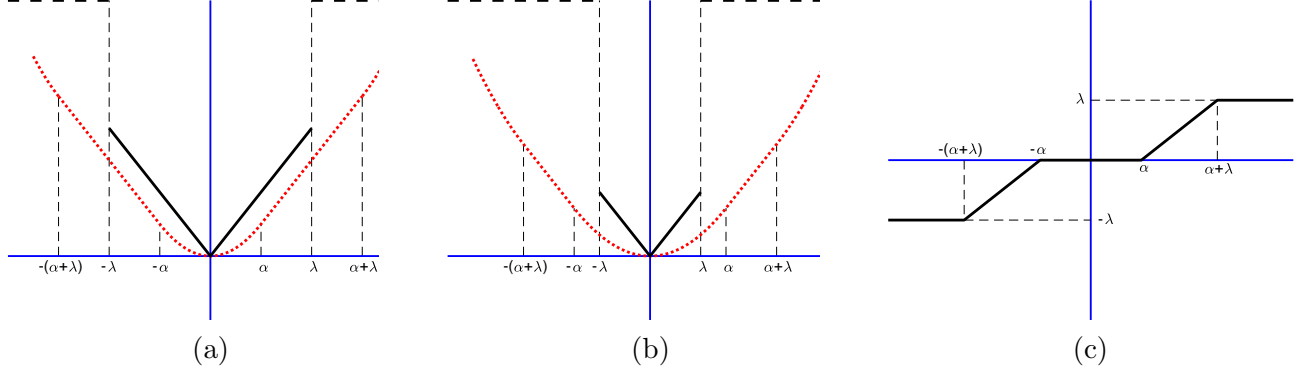


Figure 13: Example 4. The graphs of f (solid, dashed) and $\text{env}_{\alpha f}$ (dotted) when (a) $\alpha < \lambda$ and (b) $\alpha > \lambda$. The graph of $\text{prox}_{\alpha f}$ is shown in (c). Between $-(\alpha + \lambda)$ and $\alpha + \lambda$, $\text{prox}_{\alpha f}$ is the soft thresholding operator with sparsity parameter α ; otherwise it projects onto this interval.

As both f and $\text{env}_{\alpha f}$ depend on α and λ , the explicit expression for f_{α} will depend on the values of these parameters. To compute the proximity operator $\text{prox}_{\beta f_{\alpha}}$, we consider separately two main cases: $\alpha < \lambda$ and $\alpha \geq \lambda$.

Case 1: $\alpha < \lambda$. In this case, we get (see Figure 6.4)

$$f_{\alpha}(x) = f(x) - \text{env}_{\alpha f}(x) = \begin{cases} \frac{\alpha}{2} - \frac{1}{2\alpha}(|x| - \alpha)^2, & \text{if } |x| \leq \alpha; \\ \frac{\alpha}{2}, & \text{if } \alpha \leq |x| \leq \lambda; \\ \lambda, & \text{if } \lambda < |x|. \end{cases} \quad (31)$$

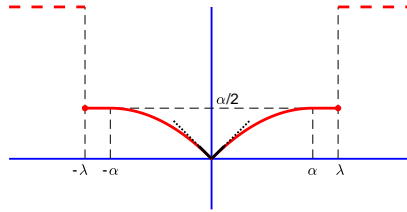


Figure 14: Example 4. The graph of f_{α} when $\alpha < \lambda$ with the singularity of f_{α} at zero emphasized in black (solid-dotted). Further, we see that f_{α} agrees with Example 1 on $[-\lambda, \lambda]$.

Depending on the values of α, β , and λ , we consider four possible cases: $\beta < \alpha < \lambda$, $\beta = \alpha < \lambda$, $\alpha < \beta \leq \lambda$, and $\lambda < \beta$.

Case 1.1: $\beta < \alpha < \lambda$. In this case, we have

$$\text{prox}_{\beta f_{\alpha}}(x) = \begin{cases} \max\{0, \frac{\alpha(|x| - \beta)}{\alpha - \beta}\} \text{sgn}(x), & \text{if } |x| \leq \alpha; \\ \min\{|x|, \lambda\} \text{sgn}(x), & \text{if } |x| > \alpha. \end{cases} \quad (32)$$

Case 1.2: $\beta = \alpha < \lambda$. In this case, we have

$$\text{prox}_{\beta f_{\alpha}}(x) = \begin{cases} 0, & \text{if } |x| < \alpha; \\ \text{sgn}(x)[0, \alpha], & \text{if } |x| = \alpha; \\ \text{sgn}(x) \min\{|x|, \lambda\}, & \text{if } \alpha < |x|, \end{cases} \quad (33)$$

Case 1.3: $\alpha < \beta \leq \lambda$. In this case, we have

$$\text{prox}_{\beta f_\alpha}(x) = \begin{cases} 0, & \text{if } |x| < \sqrt{\alpha\beta}; \\ \{0, \text{sgn}(x)\sqrt{\alpha\beta}\}, & \text{if } |x| = \sqrt{\alpha\beta}; \\ \min\{|x|, \lambda\} \text{sgn}(x), & \text{if } \sqrt{\alpha\beta} < |x|, \end{cases} \quad (34)$$

Case 1.4: $\alpha < \lambda < \beta$. We have

$$\text{prox}_{\beta f_\alpha}(x) = \begin{cases} \{0\}, & \text{if } |x| < \frac{\alpha\beta + \lambda^2}{2\lambda}; \\ \{0, \lambda \text{sgn}(x)\}, & \text{if } |x| = \frac{\alpha\beta + \lambda^2}{2\lambda}; \\ \{\lambda \text{sgn}(x)\}, & \text{if } \frac{\alpha\beta + \lambda^2}{2\lambda} < |x|, \end{cases} \quad (35)$$

We now move on to the second main case.

Case 2: $\lambda \leq \alpha$. In this case, we get (see Figure 6.4)

$$f_\alpha(x) = \begin{cases} \frac{\alpha}{2} - \frac{1}{2\alpha}(|x| - \alpha)^2, & \text{if } |x| \leq \lambda; \\ \end{cases}$$

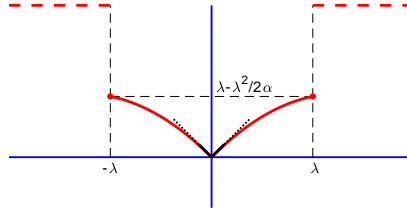


Figure 15: Example 4. The graph of f_α when $\lambda \leq \alpha$ with the singularity of f_α at zero emphasized in black (solid-dotted). As before, f_α agrees with Example 1 on $[-\lambda, \lambda]$, but is cut off before it plateaus.

To compute $\text{prox}_{\beta f_\alpha}$, we consider three situations: $\beta < \alpha$, $\beta = \alpha$, and $\beta > \alpha$.

Case 2.1: $\beta < \alpha$. In this case, we have that

$$\text{prox}_{\beta f_\alpha}(x) = \begin{cases} 0, & \text{if } |x| \leq \beta; \\ \frac{\alpha(|x| - \beta)}{\alpha - \beta} \text{sgn}(x), & \text{if } \beta \leq |x| \leq \beta + \frac{\alpha - \beta}{\alpha} \lambda; \\ \lambda \text{sgn}(x), & \text{if } \beta + \frac{\alpha - \beta}{\alpha} \lambda \leq |x|, \end{cases} \quad (36)$$

Case 2.2: $\beta = \alpha$. In this case, we have

$$\text{prox}_{\beta f_\alpha}(x) = \begin{cases} 0, & \text{if } |x| < \alpha; \\ \text{sgn}(x)[0, \lambda], & \text{if } |x| = \alpha; \\ \lambda \text{sgn}(x), & \text{if } \alpha < |x|, \end{cases} \quad (37)$$

Case 2.3: $\beta > \alpha$. Similar to Case 1.4, we get

$$\text{prox}_{\beta f_\alpha}(x) = \begin{cases} 0, & \text{if } |x| \leq \beta - \frac{\beta - \alpha}{2\alpha} \lambda; \\ \text{sgn}(x)\{0, \lambda\}, & \text{if } |x| = \beta - \frac{\beta - \alpha}{2\alpha} \lambda; \\ \lambda \text{sgn}(x), & \text{if } \beta - \frac{\beta - \alpha}{2\alpha} \lambda < |x|, \end{cases} \quad (38)$$

To end up this example, we comment on this example in a comparison with Theorem 2.

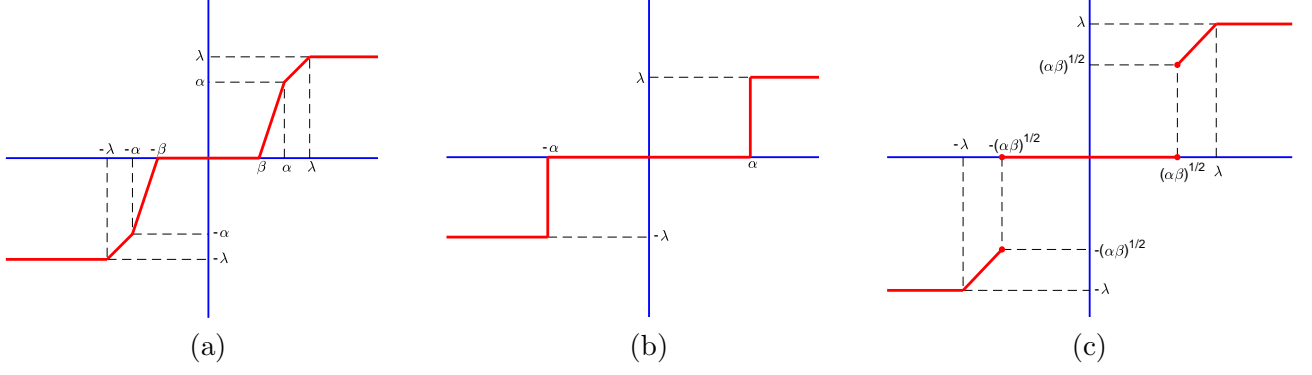


Figure 16: Example 4. Typical shapes of $\text{prox}_{\beta f_\alpha}$ in (a) Case 1.1: $\beta < \alpha < \lambda$, (b) Case 2.2: $\beta = \alpha \geq \lambda$, and (c) Case 1.3: $\alpha < \beta \leq \lambda$. In each case, we see that the absolute threshold is λ , while the sparsity threshold and thresholding behavior depend on α and β .

- Note that $\partial f(0) = [-1, 1]$. For $\beta < \alpha$, both equations (32) and (36) show that the operator $\text{prox}_{\beta f_\alpha}$ vanishes all elements in $\beta \partial f(0) = [-\beta, \beta]$ as required by the first statement of Theorem 2.
- For $\beta = \alpha$, both equations (33) and (37) show that the operator $\text{prox}_{\beta f_\alpha}$ vanishes all elements in $\text{ri}(\alpha \partial f(0)) = (-\alpha, \alpha)$ as described in the second statement of Theorem 2.
- For $\beta > \alpha$, since $\sqrt{\alpha\beta} > \alpha$, $\frac{\alpha\beta + \lambda^2}{2\lambda} > \alpha$ when $\alpha < \lambda < \beta$, and $\beta - \frac{\beta - \alpha}{2\alpha}\lambda \geq \alpha$ when $\alpha \geq \lambda$, then equations (34), (34), and (38) shows that the operator $\text{prox}_{\beta f_\alpha}$ vanishes all elements in $\text{ri}(\alpha \partial f(0)) = (-\alpha, \alpha)$ as described in the third statement of Theorem 2.

To close this section, Table 1 lists the proximity operators $\text{prox}_{\beta f_\alpha}$ of all examples.

Table 1: Proximity operators for all examples

Function	$\beta < \alpha$	$\beta = \alpha$	$\beta > \alpha$
$f(x) = x $	(21)	(22)	(23)
$f(x) = \max\{0, x\}$	(24)	(25)	(26)
$f(x) = \frac{1}{2}x^2 + x $	(27)	(27)	$\beta < \alpha(\beta + 1)$ $\beta = \alpha(\beta + 1)$ $\beta > \alpha(\beta + 1)$ (27) (28) (30)
$f(x) = x + \iota_{[-\lambda, \lambda]}$	$\alpha < \lambda$ $\alpha \geq \lambda$ (32) (36)	$\alpha < \lambda$ $\alpha \geq \lambda$ (33) (37)	$\beta \leq \lambda$ $\beta > \lambda$ $\alpha \geq \lambda$ (34) (35) (38)

7 Conclusions

We presented a simple scheme to construct a family of semiconvex structured sparsity promoting functions from any convex sparsity promoting function. Theoretical guarantees of sparsity promotion were proved in Section 4, among other properties related to the structure of these functions. In Section 5, we expanded upon these results in the case of indicator and piecewise quadratic functions. We demonstrated that the classical MCP can be derived under this framework, while also providing several other examples motivated by a variety of applications.

Because of the structure of the proposed functions, we can use convex, nonconvex, and difference of convex algorithms in practice. We plan on testing these examples on problems such as signal denoising and variable selection. Furthermore, we hope to use the unique properties of these functions to develop new algorithms. Other future work will also expand upon the theoretical properties of these functions.

Disclaimer and Acknowledgment of Support

Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of AFRL (Air Force Research Laboratory).

Lixin Shen is partially supported by the US National Science Foundation under grant DMS-1522332.

References

- [1] H. ATTOUCH, J. BOLTE, AND B. SVAITER, *Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized gauss-seidel methods*, Mathematical Programming, Ser. A, 137 (2013), pp. 91–129.
- [2] H. L. BAUSCHKE AND P. L. COMBETTES, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, AMS Books in Mathematics, Springer, New York, 2011.
- [3] J. BOLTE, A. DANIILIDIS, O. LEY, AND L. MAZET, *Characterizations of łojasiewicz inequalities: Subgradient flows, talweg, convexity*, Transactions of the American Mathematical Society, 362 (2010), pp. 3319–3363.
- [4] E. CANDÈS AND T. TAO, *Near optimal signal recovery from random projections: Universal encoding strategies?*, IEEE Transactions on Information Theory, 52 (2006), pp. 5406–5425.
- [5] S. CHEN, D. DONOHO, AND M. SAUNDERS, *Atomic decomposition by basis pursuit*, SIAM Journal of Scientific Computing, 20 (1998), pp. 33–61.
- [6] F. H. CLARKE, *Optimization and Nonsmooth Analysis*, Wiley, New York, 1983.
- [7] P. COMBETTES AND V. WAJS, *Signal recovery by proximal forward-backward splitting*, Multiscale Modeling and Simulation: A SIAM Interdisciplinary Journal, 4 (2005), pp. 1168–1200.
- [8] D. DONOHO, *Compressed sensing*, IEEE Transactions on Information Theory, 52 (2006), pp. 1289–1306.
- [9] D. DONOHO AND I. JOHNSTONE, *Ideal spatial adaptation by wavelet shrinkage*, Biometrika, 81 (1994), pp. 425–455.

- [10] J. FAN AND R. LI, *Variable selection via nonconcave penalized likelihood and its oracle properties*, Journal of the American Statistical Association, 96 (2001), pp. 1348–1360.
- [11] I. E. FRANK AND J. H. FRIEDMAN, *A statistical view of some chemometrics regression tools (with discussion)*, Technometrics, 35 (1993), pp. 109–148.
- [12] H.-Y. GAO AND A. G. BRUCE, *WaveShrink with firm shrinkage*, Statistica Sinica, 7 (1997), pp. 855–874.
- [13] J.-B. HIRIART-URRUTY AND C. LEMARECHAL, *Convex analysis and minimization algorithms I*, vol. 305 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer-Verlag, Berlin, 1993.
- [14] P. HUBER, *Robust Statistics*, John Wiley & Sons Inc., Hoboken, New Jersey, second ed., 2009.
- [15] S. MALLAT, *A Wavelet Tour of Signal Processing*, Academic Press, New York, 2nd ed., 1999.
- [16] C. A. MICCHELLI, L. SHEN, AND Y. XU, *Proximity algorithms for image models: Denoising*, Inverse Problems, 27 (2011), p. 045009(30pp).
- [17] J.-J. MOREAU, *Proximité et dualité dans un espace hilbertien*, Bull. Soc. Math. France, 93 (1965), pp. 273–299.
- [18] C. PLANIDEN AND X. WANG, *Epi-convergence: The moreau envelope and generalized linear-quadratic functions*, Journal of Optimization Theory and Applications, 177 (2018), pp. 21–63.
- [19] G. SCHWARZ, *Estimating the dimension of a model*, The Annals of Statistics, 6 (1978), pp. 461–464.
- [20] I. SELESNICK, *Total variation denoising via the moreau envelope*, IEEE Signal Processing Letters, 24 (2017), pp. 216–220.
- [21] E. SOUBIES, L. BLANC-FERAUD, AND G. AUBERT, *A unified view of exact continuous penalties for ℓ_2 - ℓ_0 minimization*, SIAM Journal on Optimization, 27 (2017), pp. 2034–2060.
- [22] B. W. SUTER, *Multirate and wavelet signal processing*, Academic Press, 1997.
- [23] P. TAO AND H. LE THI, *Convex analysis approach to dc programming: Theory, algorithms and applications*, Acta Mathematica Vietnamica, 22 (1997), pp. 289–355.
- [24] R. TIBSHIRANI, *Regression shrinkage and selection via the LASSO*, Journal of the Royal Statistical Society, Series B, 58 (1996), pp. 267–288.
- [25] C.-H. ZHANG, *Nearly unbiased variable selection under minimax concave penalty*, Annals of Statistics, 38 (2010), pp. 894–942.
- [26] H. ZOU AND T. HASTIE, *Regularization and variable selection via the elastic net*, Journal of the Royal Statistical Society, Series B, 67 (2005), pp. 301–320.