Structured Sparsity Promoting Functions

Lixin Shen¹ · Bruce W. Suter² · Erin E. Tripp¹

Received: 1 August 2018 / Accepted: 29 June 2019 / Published online: 13 August 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract
Motivated by the minimax concave penalty-based variable selection in high-dimensional linear regression, we introduce a simple scheme to construct structured 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 and convex sets, piecewise quadratic functions, and linear combinations of the two. 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

Mathematics Subject Classification 90C26 · 90C30 · 65K99

Communicated by Lionel Thibault.

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

✉ Lixin Shen
lshen03@syr.edu

Bruce W. Suter
bruce.suter@us.af.mil

Erin E. Tripp
eetripp@syr.edu

¹ Syracuse University, Syracuse, NY 13244, USA
² Air Force Research Laboratory, Rome, NY 13441, USA
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 [1–3]. 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 ℓ₀-norm, which counts the nonzero components of a vector. However, minimizing an ℓ₀-penalized objective is a combinatorial optimization problem, which is known to be NP-hard.

To overcome these computational difficulties, regularization methods with the ℓ₁-norm as its penalty function like the least absolute shrinkage and selection operator (LASSO) [4] and Dantzig selectors [1] have been proposed. The convexity of the ℓ₁-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, nonconvex penalties including the ℓ_q penalty with 0 < q < 1 [5], the smoothly clipped absolute deviation penalty (SCAD) [6], and the minimax concave penalty (MCP) [7] have been proposed to replace the ℓ₁-norm penalty.

In this paper, we introduce a family of semiconvex sparsity promoting functions each of which 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 can be made convex by adding a convex quadratic function to it. Semiconvex functions possess useful structure and obey generalizations of many classical results from convex analysis (see, e.g., [8] or [9]).

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 refer to these functions as structured sparsity promoting functions. These properties enable us to make use of the fruitful results, for example in DC programming [10], to develop efficient algorithms for the associated regularized optimization problems. What is 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 ℓ₁-norm and its envelope. It has been shown (e.g., in [11]) that this closely approximates the ℓ₀-norm while preserving the continuity and subdifferentiability of the ℓ₁-norm.

The proximity operator, which was first introduced by Moreau in [12] as a generalization of the notion of projection onto a convex set, has been used extensively in nonlinear optimization (see, e.g., [13–15]). The desired features of the aforementioned regularization methods can be explained in terms of the proximity operators of the corresponding penalties. For example, the proximity operator of the ℓ₀-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. To determine the effectiveness of our proposed functions, we examine the behavior of their proximity operators. More generally, we provide sparsity guarantees in terms of thresholding behavior for the entire family of structured 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 Sect. 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 Sect. 5, several special sparsity promoting functions are presented and discussed thoroughly. Some examples of practical interest are provided in Sect. 6. We conclude by discussing applications and plans for future work in Sect. 7.

2 Motivation

Our work on semiconvex sparsity promoting functions is motivated mainly by the minimax concave penalty (MCP)-based variable selection in high-dimensional linear regression [7]. Variable selection is fundamental in statistical analysis of high-dimensional data, but 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 $\gamma$, and $n$-dimensional error vector $\epsilon$:

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

The goal of variable selection is to recover the true underlying model of the pattern $\{j : \gamma_j \neq 0\}$ and to estimate the nonzero regression coefficients $\gamma_j$, where $\gamma_j$ is the $j$th component of $\gamma$. For small $p$, subset selection methods can be used to find a good estimate of the pattern (see, e.g., [16]). However, subset selection becomes computationally infeasible for large $p$.

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

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

(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 $\ell_0$ penalty $\rho(|t|, \lambda) = \frac{\lambda^2}{2} 1_{|t| \neq 0}$ while LASSO [4] and basis pursuit [17] use the $\ell_1$ penalty $\rho(|t|, \lambda) = \lambda|t|$. Here, $1_{u \in E}$ denotes the characteristic function, which equals 1 if $u \in E$ and 0 otherwise. The estimator with the $\ell_0$ penalty (i.e., the hard thresholding operator) suffers from instability in model prediction while the estimator with the $\ell_1$ penalty (the soft thresholding operator) suffers from the bias issue, interfering with variable selection for large $p$ [6]. To remedy this issue, the SCAD penalty was introduced in [6]. The estimator with the SCAD penalty is continuous and leaves large components not excessively penalized. Introduced in [7], the MCP penalty is defined as follows

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

where the parameter $a > 0$. This penalty function (see [7]) minimizes the maximum concavity

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

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 [6,7] gave strong statistical evidence that the estimators from the nonconvex penalty functions SCAD and MCP are useful in variable selection. Recently, an application of MCP to signal processing has been reported in [18].

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 can be considered as a variation in the $\ell_1$ penalty function, that is, the absolution 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 certain 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, which 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 Euclidean space $\mathbb{R}^n$ 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 closure of a set $S$ is denoted by $\text{cl}(S)$. The boundary of a set $S$ denoted by $\text{bd}(S)$ is the set of points in the closure $\text{cl}(S)$, 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 $\delta$. 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

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

For any $x \notin \text{dom}(g)$, $\partial g(x) = \emptyset$. If $g \in \Gamma_0(\mathbb{R}^n)$, 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 $\partial g$ is a monotone operator; that is, for any $x, y \in \text{dom}(g)$, $\xi \in \partial g(x)$, and $\eta \in \partial g(y)$, we have $\langle \eta - \xi, y - x \rangle \geq 0$. Moreover, if $g$ is Fréchet differentiable, $\partial g(x) = \{\nabla g(x)\}$.

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

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

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

$$\text{prox}_{\alpha g}(x) = \text{argmin} \left\{ g(u) + \frac{1}{2\alpha} \|u - x\|^2 : u \in \mathbb{R}^n \right\}.$$
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
\[ g^*(x) = \sup \{ \langle u, x \rangle - g(u) : u \in \mathbb{R}^n \}. \]
The Fenchel conjugate is closely related to the Moreau envelope. Indeed, it is shown in [14] 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 3.1** 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.

Part (i) of the above definition and Fermat’s rule imply that $0 \in \partial f(0)$. It then follows from part (ii) that $f$ must be nondifferentiable at the origin. As pointed out in [6], 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. In fact, if $\| \cdot \|$ is any norm on $\mathbb{R}^n$, then $\| \cdot \|$ is a sparsity promoting function. It is obvious that the norm $\| \cdot \|$ is convex and $0 = \| 0 \| = \min_{x \in \mathbb{R}^n} \| x \|$. We further know that
\[ \partial \| \cdot \|(0) = \left\{ s \in \mathbb{R}^n : \max_{\| u \| \leq 1} \langle s, u \rangle \leq 1 \right\}, \]
which is the unit ball associated with the dual norm of $\| \cdot \|$ (see, e.g., [19]).

Another example of a sparsity promoting function is the indicator function $\iota_C$ 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 and convex set such that $0 \in \text{bd } C$ and $\{0\} \not\subseteq C$. For further discussion of this example, we refer to Sect. 5.

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., [14,20]): for any $\alpha > 0$
\[ x \in \alpha \partial f(y) \iff y = \text{prox}_{\alpha f}(x + y). \quad (4) \]
From this relationship, we get the following characterization of convex sparsity promoting functions.

**Lemma 3.1** 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 Eq. (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, \( \| \text{prox}_{\alpha f}(x) \| = \| \text{prox}_{\alpha f}(x) - \text{prox}_{\alpha f}(0) \| \leq \| x - 0 \| \) for all \( x \in \text{dom}(f) \). \( \square \)

It follows from Lemma 3.1 that the proximity operator of a convex sparsity promoting function shrinks all inputs toward the origin, and all inputs below a certain threshold are sent to zero. As an example, the proximity operator of \( | \cdot | \) is \( \text{prox}_{\alpha | \cdot |}(x) = \text{sgn}(x) \max\{ |x| - \alpha, 0 \} \), which is the well-known soft thresholding operator in the wavelet literature [21]. This exact behavior for the \( \ell_1 \) penalty is described by Tibshirani in the name LASSO: least absolute shrinkage and selection operator [4].

**Remark 3.1** Item (i) does not directly depend on Definition 3.1. However, the definition guarantees that not only is some nonzero element sent to zero, but, by the convexity of \( \partial f(0) \), that all elements in the line segment \([0, x]\) are sent to zero. We refer to this as the thresholding behavior of the proximity operator.

### 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

\[
    f_\alpha(x) := f(x) - \text{env}_\alpha f(x). \quad (F_\alpha)
\]

Clearly, \( f_\alpha \) is in \( \Gamma(\mathbb{R}^n) \) and is the difference of two convex functions. As discussed in the previous section, when \( f \) is the absolute value function, this is the scaled minimax concave penalty (MCP) given in [7]. For details, see Sect. 6.

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., [14]), the subdifferentials of \( f_\alpha \) and \( f \) are related as follows (see [22]):

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

Due to this inherent relationship between \( \partial f_\alpha \) and \( \partial f \), we see immediately that \( f_\alpha \) must be sparsity promoting if \( f \) is.

**Theorem 4.1** Let \( f \in \Gamma_0(\mathbb{R}^n) \) be a sparsity promoting function. Then, the following statements hold:

}\( \Box \) Springer
(i) For any $\alpha > 0$, the function $f_\alpha$ defined by $(F_\alpha)$ is a sparsity promoting function. Moreover, $\partial f_\alpha(0) = \partial f(0)$;

(ii) Let $g : x \mapsto f(-x)$. Then, both $g$ and $g_\alpha$ are sparsity promoting. Moreover, it holds that $g_\alpha = f_\alpha(-\cdot)$ 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 $x \in \mathbb{R}^n$, $f_\alpha(x) \geq 0$ for all $x \in \text{dom}(f)$. Furthermore, since $f$ achieves a global minimum of 0 at the origin, it follows that $\min_{x \in \mathbb{R}^n} \text{env}_\alpha f(x) = \text{env}_\alpha f(0) = 0$ as well. Thus, we see that $f_\alpha(0) = 0$. Therefore, $\min_{x \in \mathbb{R}^n} f_\alpha(x) = f_\alpha(0) = 0$.

On the other hand, from (5) and the fact that 0 is a minimizer of $\text{env}_\alpha f$, we have $\partial f_\alpha(0) = -\partial f(0)$, which contains at least one nonzero element by assumption.

Hence, $f_\alpha$ 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, $g_\alpha$ is sparsity promoting and $\partial g_\alpha(0) = -\partial f(0)$ by Item (i). 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 4.1, we say $f_\alpha$ is a structured sparsity promoting function if $f$ is a convex sparsity promoting function. We now prove that $f_\alpha$ is semiconvex and show how its semiconvexity depends on the convexity of $f$. We remind the reader of the definition. Recall that, for $\sigma > 0$, a function $g \in \Gamma_0(\mathbb{R}^n)$ is $\sigma$-strongly convex if and only if the function $g - \frac{\sigma}{2} \| \cdot \|^2$ is convex. For $\rho > 0$, a function $g \in \Gamma(\mathbb{R}^n)$ is $\rho$-semiconvex if and only if $g + \frac{\rho}{2} \| \cdot \|^2$ is convex.

**Proposition 4.1** Let $f$ be a function in $\Gamma_0(\mathbb{R}^n)$, and let $f_\alpha$ be defined by $(F_\alpha)$. Then, $f_\alpha$ is $\frac{1}{\alpha}$-semiconvex. If, in addition, $f$ is $\mu$-strongly convex, then $f_\alpha$ 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 $f_\alpha$ is $\frac{1}{\alpha}$-semiconvex.

In addition, if $f$ is $\mu$-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 $f_\alpha$ 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 4.1.

**Corollary 4.1** Let $f$ be a function in $\Gamma_0(\mathbb{R}^n)$, and let $f_\alpha$ be defined by $(F_\alpha)$. For any given $x \in \mathbb{R}^n$ and positive parameters $\alpha$ and $\beta$, we define

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

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 $\mu$-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 $f_\alpha$ by examining its proximity operator. Roughly speaking, we show that $\text{prox}_{\beta f_\alpha}(x) = 0$ for all $x \in \min\{\alpha, \beta\} \cdot \partial f(0)$. Toward this end, we present two technical lemmas. The first is a generalization of Lemma 3.1.

**Lemma 4.1** Let $f \in \Gamma_0(\mathbb{R}^n)$ be sparsity promoting and $f_\alpha$ as defined in $(F_\alpha)$.

(i) For any $x \in \text{dom}(f)$, $\text{prox}_{\beta f_\alpha}(x) \subseteq \text{cl}(B_{\|x\|}(x))$.

(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 $\text{prox}_{\beta f_\alpha}(x) = \arg\min_{u \in \mathbb{R}^n} F(u)$.

(i) Since $F(0) = \frac{1}{2\beta} \|x\|^2$ and $0 \in \text{cl}(B_{\|x\|}(x))$, to show $\text{prox}_{\beta f_\alpha}(x) \subseteq \text{cl}(B_{\|x\|}(x))$, we only need to show that for all $u \in \mathbb{R}^n \setminus \text{cl}(B_{\|x\|}(x))$, $F(u) > F(0)$. Actually, if $u \in \mathbb{R}^n \setminus \text{cl}(B_{\|x\|}(x))$, then $\|u - x\|^2 > \|x\|^2$. Since $f_\alpha$ is nonnegative, 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 $u \in \text{cl}(B_{\|x\|}(x))$. Since $x \in \min\{\alpha, \beta\} \cdot \partial f(0)$, 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,

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

So, $F(u) \geq \frac{1}{2\beta} \|x\|^2 = F(0)$ holds for all $u \in \text{cl}(B_{\|x\|}(x))$. This completes the proof of the lemma. \qed
Remark 4.1 From item (i) of Lemma 4.1, we see for $x \in \mathbb{R}$, $\text{sgn}(x) = \text{sgn}(p)$ if $p \in \text{prox}_{\beta f_o}(x)$ and both $x$ and $p$ are simultaneously nonzero. We note that this is also true for $\text{prox}_{\alpha f}(x)$ when $f$ is a convex sparsity promoting function.

The following technical lemma will greatly simplify the proof of Theorem 4.2, 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 4.2 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 that $f(w) \geq \|w\|^2 > 0$.

Second, since $\xi \in \partial f(0)$ 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 $\partial f$, 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 $w = 0$. \qed

Now our main result which characterizes the sparsity promoting structure of $f_o$ in terms of the sparsity threshold of its proximity operator is presented in the following result.

**Theorem 4.2** 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 $\text{prox}_{\beta f_o}(x) = 0$ for $x \in \beta \partial f(0)$;
(ii) If $\beta = \alpha$, then $\text{prox}_{\beta f_o}(x) = 0$ for $x \in \text{ri}(\alpha \partial f(0))$;
(iii) If $\beta > \alpha$, then $\text{prox}_{\beta f_o}(x) = 0$ for $x \in \alpha \partial f(0)$.

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

(i) We first consider the situation $\beta < \alpha$. From Corollary 4.1, we know that $F$ is $\left(\frac{1}{\beta} - \frac{1}{\alpha}\right)$-strongly convex and therefore has a unique minimizer. By Lemma 4.1, $x \in \beta \partial f(0)$ implies that $\arg\min_{u \in \mathbb{R}^n} F(u) = 0$, i.e., $\text{prox}_{\beta f_o}(x) = 0$.

(ii) Next, we consider $\alpha = \beta$. From Corollary 4.1, $F(u)$ is convex but not strongly, and the minimizer may no longer be unique. By Lemma 4.1, $0 \in \text{prox}_{\beta f_o}(x)$ 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 $w^* = 0$, identify $\alpha f, x,$ and $w^*$, respectively, as $f, \xi,$ and $w$ in Lemma 4.2. Then, 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

$$0 \in \partial f_\alpha(w^*) + \frac{1}{\beta}(w^* - x).$$

Because $\partial f_\alpha(w^*) = \partial f(w^*) - \nabla \text{env}_f(w^*)$ and $\nabla \text{env}_f(w^*) = \frac{1}{\alpha}(w^* - \text{prox}_{\alpha f}(w^*))$ (see, e.g., [14]), the above inclusion can be rewritten as

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

From (10), we get $x - \text{prox}_{\alpha f}(w^*) \in \partial(\alpha f)(w^*)$. Therefore, both $x \in \partial(\alpha f)(w^*)$ 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 $\partial f$ 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 4.2, $w^* = 0$.

(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 4.1, 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 4.1 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 $\text{cl}(B\|x\|(x))$ 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,

$$f(w^*) = \text{env}_f(w^*).$$

We also know that $f(w^*) \geq \frac{1}{\alpha}\langle x, w^* \rangle$ and $\text{env}_f(w^*) \leq \frac{1}{2\alpha}\|w^*\|^2$. Therefore, because $2\langle x, w^* \rangle = \|w^*\|^2$, we get

$$\text{env}_f(w^*) = \frac{1}{2\alpha}\|w^*\|^2.$$  

Springer
which implies that \(0 = \prox_{\alpha f}(w^*)\). On the other hand, the identity (11) indicates \(w^* = \prox_{\alpha f}(w^*)\). Therefore, \(w^* = 0\). This completes the proof. \(\square\)

**Remark 4.2** Item (iii) of the theorem is not tight. In fact, in every example, when \(\beta > \alpha\), \(\prox_{\beta f_{\alpha}}(x) = 0\) 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 previous section deals primarily with behavior around the origin for general semi-convex sparsity promoting functions. In this section, we describe the structure of \(f_{\alpha}\) on its 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 not only are they fixed by the mapping \(f \mapsto f_{\alpha}\), but they are the only functions that are fixed.

Throughout, we assume \(C\) is a closed and convex set in \(\mathbb{R}^n\) with boundary \(\text{bd}(C)\).

Recall that the indicator function of \(C\) is

\[
\iota_C(x) = \begin{cases} 
0, & \text{if } x \in C, \\
+\infty, & \text{otherwise}. 
\end{cases} \tag{I}
\]

We first determine when this is a sparsity promoting function.

**Lemma 5.1** The indicator function \(\iota_C\) 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 for \(\iota_C\) 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 [14]:

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

If \(0 \in \text{bd}(C)\), it follows that \(N_C(x)\) 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 Sect. 3. □

Let $P_C(x)$ be the projection of $x$ onto $C$ such that $\|x - P_C(x)\|$ is the distance from $x$ to $C$. It is well known (see, e.g., [14]) that $\text{prox}_{\alpha f_C}(x) = P_C(x)$ and, further, that $p = P_C(x)$ if and only if $x - p \in N_C(p)$. In terms of the proximity operator, this states $\text{prox}_{\alpha f_C}(x) = 0$ if and only if $x \in N_C(0)$. It is straightforward to show that $f_{\alpha} \iota_C(x) = \frac{1}{2\alpha} \|P_C(x) - x\|^2$. Therefore,

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

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

**Theorem 5.1** Let $f \in \Gamma_0(\mathbb{R}^n)$ be sparsity promoting. If $f = f_\alpha$ as defined by $(\mathcal{F}_\alpha)$, 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 $x \in \text{dom}(f)$. □

**Remark 5.1** The theorem is true more generally if $f \in \Gamma_0(\mathbb{R}^n)$ is simply nonnegative.

The previous results show that indicator functions are fixed under this process.

**Proposition 5.1** Let $f \in \Gamma_0(\mathbb{R}^n)$ be a sparsity promoting function. Suppose that $C$ is a closed and convex subset of $\mathbb{R}^n$ and $\{0\} \subset \partial f(0) \cap C$. Then, the sum $\tilde{f} := f + \iota_C$ is sparsity promoting and $\tilde{f}_\alpha = f_\alpha + \iota_C$.

**Proof** Since $f$ is sparsity promoting, $\min_{x \in \mathbb{R}^n} f(x) = f(0) = 0$. Now, because $\{0\} \subset \partial f(0) \cap C$, we know that $\tilde{f}(0) = \min_{x \in C} f(x) = 0$. That is, $\tilde{f}$ achieves its minimum at the origin.

Note that $0 \in N_C(0)$ and $\partial f(0) + N_C(0) = \partial f(0) + \partial \iota_C(0) \subset \partial \tilde{f}(0)$. We have $\partial f(0) \subset \partial \tilde{f}(0)$. Hence, $\partial \tilde{f}(0)$ must contain a nonzero element. Therefore, $\tilde{f}$ is sparsity promoting.

By Lemma 3.1, $\text{prox}_{\alpha f}(x) \in C$ if $x \in C$. This indicates that for $x \in C$,

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

It follows that $\tilde{f}_\alpha = f_\alpha + \iota_C$. This completes the proof of the result. □

**5.2 Piecewise Quadratic Functions**

Piecewise quadratic functions include a variety of important examples (see Sect. 6) and are of particular interest in the study of Moreau envelopes and proximity operators.
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

$$ f(x) = \begin{cases} \frac{1}{2}a_1 x^2 + b_1 x, & \text{if } x \leq 0, \\ \frac{1}{2}a_2 x^2 + b_2 x, & \text{if } x \geq 0, \end{cases} $$

(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 by (Q) is established in the following lemma.

**Lemma 5.2** Let $f$ be a piecewise quadratic function defined by (Q). Then, $f$ is sparsity promoting if and only if

$$ a_1 \geq 0, \quad a_2 \geq 0, \quad b_1 \leq 0 \leq b_2, \quad \text{and} \quad b_2 - b_1 > 0. \quad (12) $$

**Proof** “⇒”: Since $f$ is sparsity promoting, 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, and hence, $b_2 - b_1 > 0$.

“⇐”: 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. ☐

**Remark 5.2** 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, assume that the coefficients in (Q) satisfy the conditions listed in (12). The proximity operator and Moreau envelope of $f$ with index $\alpha$ at $x \in \mathbb{R}$ are

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

and

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

 Springer
respectively. From the above two equations, we get

\[
f_\alpha(x) = \begin{cases} \frac{\alpha a_1}{\alpha a_1 + 1} f(x) + \frac{\alpha b_1^2}{2(\alpha a_1 + 1)}, & \text{if } x \leq \alpha b_1, \\ f(x) - \frac{1}{2\alpha} x^2, & \text{if } \alpha b_1 \leq x \leq \alpha b_2, \\ \frac{\alpha a_2}{\alpha a_2 + 1} f(x) + \frac{\alpha b_2^2}{2(\alpha a_2 + 1)}, & \text{if } x \geq \alpha b_2, \end{cases} \quad (Q_\alpha)
\]

which is a piecewise quadratic polynomial with possible breakpoints at \(\alpha b_1, 0,\) and \(\alpha b_2\). We know this \(f_\alpha\) is sparsity promoting by Theorem 4.1. Some other properties of this function which follow immediately from \((Q_\alpha)\) are collected in the following lemma.

**Lemma 5.3** Let \(f \in \Gamma_0(\mathbb{R})\) be a sparsity promoting function defined by \((Q)\). Then, the following hold:

(i) \(f_\alpha\) is nonincreasing on \([-\infty, 0]\) and is nondecreasing on \([0, \infty[;\)

(ii) \(f_\alpha\) on \([-\infty, \alpha b_1]\) is convex and is a degree 2 polynomial if \(a_1 > 0\) or constant if \(a_1 = 0;\)

(iii) \(f_\alpha\) on \([\alpha b_1, \alpha b_2, \infty[\) is convex and is a degree 2 polynomial if \(a_2 > 0\) or a constant if \(a_2 = 0;\)

(iv) \(f_\alpha\) 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, the result in Lemma 5.3 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 \(b_2 = 1.\) Then, \(f_\alpha(x)\) equals \(|x| - \frac{1}{2\alpha} x^2\) if \(|x| \leq \alpha\) and \(\frac{\alpha}{2}\) otherwise. 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 \(\beta > \alpha,\)

\[
\text{prox}_{\beta f_\alpha}(x) = x \quad \text{when } |x| > \sqrt{\alpha \beta}.\]

Some other details can be found in Example 1 of Sect. 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 \((Q_\alpha).\) We assume that \(x \geq 0\) for a moment. By Lemma 4.1, we know that \(\text{prox}_{\beta f_\alpha}(x) \subseteq [0, \infty[\), and therefore, by the definition of the proximity operator,

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

Let \(E(u, x) := f_\alpha(u) + \frac{1}{2\beta} (u - x)^2.\) In view of \((Q_\alpha),\) the objective function \(E(x, u)\) with \((x, u) \in [0, \infty[ \times [0, \infty[\) is

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

where

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

\[
E_2(x, u) = \frac{1}{2\beta} (u - x)^2.
\]
These two functions match at the line $u = \alpha b_2$, that is, for all $x \geq 0$,

$$E_1(x, \alpha b_2) = E_2(x, \alpha b_2), \quad (16)$$

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,

$$\text{prox}_{\beta f_\alpha}(x) \subseteq s_1(x) \cup s_2(x). \quad (17)$$

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, we have for each $x \geq 0$, $s_2(x)$ is a singleton set as follows:

$$s_2(x) = \max \left\{ \alpha b_2, \frac{aa_2 + 1}{aa_2(2\beta + 1) + 1} \left( x - \frac{aa_2 b_2}{aa_2 + 1} \right) \right\}$$

$$= \begin{cases} 
\alpha b_2, & \text{if } 0 \leq x \leq \alpha b_2(2\beta + 1), \\
\frac{aa_2 + 1}{aa_2(2\beta + 1) + 1} \left( x - \frac{aa_2 b_2}{aa_2 + 1} \right), & \text{if } x \geq \alpha b_2(2\beta + 1),
\end{cases} \quad (18)$$

which clearly is a piecewise linear function of $x$.

**Lemma 5.4** Let $f$ be a piecewise quadratic sparsity promoting function as defined by ($Q$). 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 (18).

**Proof** When $b_2 = 0$, it follows from (13) to (15) that $E(x, u) = E_2(x, u)$ for $(x, u) \in [0, \infty] \times [0, \infty]$.

Next, we assume that $b_2 > 0$ by Lemma 5.2. In view of the form of $E_1(x, \cdot)$ in (14), 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 equivalent to (i) $\alpha b_2(2\beta + 1) > \beta b_2$, (ii) $\alpha b_2(2\beta + 1) = \beta b_2$, and (iii) $\alpha b_2(2\beta + 1) < \beta b_2$, 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 5.5 Let \( f \) be a piecewise quadratic sparsity promoting function as defined by (Q). If \( b_2 > 0 \) and \( ab_2(a_2\beta + 1) > \beta b_2 \), then
\[
\text{prox}_{\beta f_a}(x) = \begin{cases} 
0, & \text{if } 0 \leq x < \beta b_2, \\
\frac{\alpha}{a_2\beta + 1} (x - \beta b_2), & \text{if } \beta b_2 \leq x \leq ab_2(a_2\beta + 1), \\
\frac{\alpha}{a_2a_2\beta + 1} \left( x - \frac{a_2b_2}{a_2a_2\beta + 1} \right), & \text{if } x > ab_2(a_2\beta + 1).
\end{cases}
\]

Proof From (17), we first find the set \( s_1(x) \) since the set \( s_2(x) \) is already given in (18). By the assumption of this lemma, for each \( x \geq 0 \), \( s_1(x) \) contains only one element and is given as follows:
\[
s_1(x) = \begin{cases} 
0, & \text{if } 0 \leq x < \beta b_2, \\
\frac{\alpha}{a_2\beta + 1} (x - \beta b_2), & \text{if } \beta b_2 \leq x \leq ab_2(a_2\beta + 1), \\
\frac{\alpha}{a_2b_2}, & \text{if } x > ab_2(a_2\beta + 1).
\end{cases}
\]

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

We use Fig. 1 to visualize the minimizers of \( E_1 \) and \( E_2 \). Three vertical lines \( x = 0 \), \( x = \beta b_2 \), and \( x = ab_2(a_2\beta + 1) \) and two horizontal lines \( u = 0 \) and \( u = ab_2 \) 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 \( E_2(x, ab_2) \leq E_2(x, u) \) in region II, so \( E_1(x, 0) < E_2(x, ab_2) \) by Eq. (16) 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 \( E_2(x, ab_2) \leq E_2(x, u) \) in region IV, so \( E_1(x, s_1(x)) < E_2(x, ab_2) \) by Eq. (16) for \( \beta b_2 \leq x \leq ab_2(a_2\beta + 1) \); Finally, we know \( E_1(x, ab_2) \leq E_1(x, u) \) in region V and \( E_2(x, s_2(x)) \leq E_2(x, u) \) in region VI, so \( E_2(x, s_2(x)) < E_1(x, ab_2) \) by Eq. (16) for \( x > ab_2(a_2\beta + 1) \). Thus, \( \text{prox}_{\beta f_a} \) is given by (19).

Next result is for case (ii).

---

Fig. 1 Illustration of case (i): \( b_2 > 0 \) and \( ab_2(a_2\beta + 1) > \beta b_2 \). The graphs of \( a \) \( s_1(x) \) (solid) and \( s_2(x) \) (dashed) and \( b \) the resulting proximity operator \( \text{prox}_{\beta f_a}(x) \)
\[
\text{prox}_{\beta f_a}(x) = \begin{cases} 
0, & \text{if } 0 \leq x < \beta b_2, \\
[0, \alpha b_2], & \text{if } x = \beta b_2, \\
\alpha b_2, & \text{if } x > \beta b_2.
\end{cases}
\]

**Proof** Similar to the proof of Lemma 5.5, 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 \(\text{prox}_{\beta f_a}\) can be set-valued only at \(\beta b_2\).

In Fig. 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 Fig. 1 with the middle regions collapsed to a line. Following the same reasoning as in Lemma 5.5, we see that (20) holds.

Finally, we consider case (iii). Because \(\beta b_2\) and \(\alpha b_2(\alpha_2 + 1)\) have now switched positions, we see that we must take care when dealing with the intermediate \(x\) values.

**Lemma 5.7** Let \(f\) be a piecewise quadratic sparsity promoting function as defined by (Q). Define

\[
\tau^+ = \frac{\alpha a_2 \beta b_2}{\alpha a_2 + 1} + \sqrt{\frac{\alpha \beta (\alpha a_2^2 + \alpha a_2 + 1) b_2}{\alpha a_2 + 1}}.
\]
If \( b_2 > 0 \) and \( ab_2(a_2\beta + 1) < \beta b_2 \),

\[
\text{prox}_{\beta f_a}(x) = \begin{cases} 
0, & \text{if } 0 \leq x < \frac{1}{2}(ab_2(a_2\beta + 1) + \beta b_2), \\
0, & \text{if } x = \tau^+, \\
\frac{a\alpha_a}{a\alpha_a(a_2\beta + 1) + 1} \left( \frac{\alpha}{a\alpha_a(a_2\beta + 1) + 1} + \frac{\beta b_2}{a\alpha_a(a_2\beta + 1) + 1} \right), & \text{if } x > \tau^+. 
\end{cases}
\]

(21)

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,

\[
s_1(x) = \begin{cases} 
0, & \text{if } 0 \leq x < \frac{1}{2}(ab_2(a_2\beta + 1) + \beta b_2), \\
[0, ab_2], & \text{if } x = \tau^+, \\
\frac{ab_2}{a\alpha_a(a_2\beta + 1) + 1}, & \text{if } x > \tau^+. 
\end{cases}
\]

This is set-valued at \( \frac{1}{2}(ab_2(a_2\beta + 1) + \beta b_2) \).

As before, we plot \( s_1(x) \) and \( s_2(x) \) in Fig. 3. Three vertical lines \( x = 0, x = ab_2(a_2\beta + 1), \) and \( x = \frac{1}{2}(ab_2(a_2\beta + 1) + \beta b_2) \) and two horizontal lines \( u = 0 \) and \( u = ab_2 \) partition the first quadrant into six rectangular regions as shown in Fig. 3a. 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 (16), it is easy to see that regions I, II, V, and VI behave as in the previous cases. That is, \( \text{prox}_{\beta f_a}(x) = s_1(x) \) for \( 0 \leq x \leq ab_2(a_2\beta + 1) \) and \( \text{prox}_{\beta f_a}(x) = s_2(x) \) for \( x \geq \frac{1}{2}(ab_2(a_2\beta + 1) + \beta b_2) \).

For \( ab_2(a_2\beta + 1) < x < \frac{1}{2}(ab_2(a_2\beta + 1) + \beta b_2) \), to find the expression of \( \text{prox}_{\beta f_a}(x) \) from the solid red line and the dashed blue in regions III and IV, we need to compare the value of \( E_1(x, 0) \) with \( E_2(x, s_2(x)) \). For notational convenience, we let \( E_{12}(x) := E_2(x, s_2(x)) - E_1(x, 0) \). Using (18), a direct computation gives

\[
E_{12}(x) = -\frac{\alpha a_2 + 1}{2\beta(a\alpha_a(a_2\beta + 1) + 1)} \left( x - \frac{\alpha a_2 \beta b_2}{a\alpha_a + 1} \right)^2 + \frac{\alpha b_2^2}{2(a\alpha_a + 1)},
\]

Fig. 3 Illustration of case (iii): \( b_2 > 0 \) and \( ab_2(a_2\beta + 1) < \beta b_2 \). The graphs of \( a s_1(x) \) (solid) and \( s_2(x) \) (dashed) with \( b \) the line \( x = \tau^+ \) added.
Illustration of case (iii):

\[ b_2 > 0 \quad \text{and} \quad ab_2(a_2\beta + 1) < \beta b_2. \]

The graph of the resulting proximity operator \( \text{prox}_{\beta f_{\alpha}}(x) \)

which is positive at \( x = \alpha b_2(a_2\beta + 1) \) and negative at \( x = \frac{1}{2}(\alpha b_2(a_2\beta + 1) + \beta b_2). \) Hence, the quadratic polynomial \( E_{12}(x) \) has only one root \( \tau^+ \) that is between \( \alpha b_2(a_2\beta + 1) \) and \( \frac{1}{2}(\alpha b_2(a_2\beta + 1) + \beta b_2) \) as shown in Fig. 3b. In other words, \( E_1(\tau^+, 0) = E_2(\tau^+, s_2(\tau^+)), \) giving us \( \text{prox}_{\beta f_{\alpha}}(\tau^+) = \{0, s_2(\tau^+)\}. \) So, the result of this lemma holds and is illustrated in Fig. 4. \( \square \)

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 5.8** 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 \( \alpha \) and \( \beta, \) 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 5.4–5.7.

**Proof** By Theorem 4.1, \( g \) is sparsity promoting because \( f \) is, and \( f_{\alpha} = g_{\alpha}(-\cdot). \) This implies that \( \text{prox}_{\beta f_{\alpha}}(x) = -\text{prox}_{\beta g_{\alpha}}(-x) \) 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 the results developed in Lemmas 5.4–5.7 can be applied for \( g. \) Therefore, the results of this lemma follow immediately. \( \square \)

In summary, we have the following result.

**Theorem 5.2** 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 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 the results follow directly from the expressions of \( \text{prox}_{\beta f_{\alpha}}(x) \) given in Lemmas 5.4–5.8. \( \square \)
Remark 5.3  Theorem 5.2 guarantees that prox$\beta f_\alpha$ will be a thresholding operator for any $f_\alpha$ given by $(Q_\alpha)$. Furthermore, Lemmas 5.4–5.7 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 $\widetilde{f}$ that is the restriction of $f$ on the interval $C$ as follows:

$$\widetilde{f} = f + \iota_C.$$

Let $f$ be a piecewise quadratic sparsity promoting function defined by $(Q)$ and let $C$ be a closed interval on $\mathbb{R}$ such that $\{0\} \subseteq \partial f(0) \cap C$. By Proposition 5.1, $\widetilde{f}$ defined above is a sparsity promoting function, and

$$\widetilde{f}_\alpha = f_\alpha + \iota_C.$$

For $\widetilde{f}$ defined in $(\widetilde{Q})$, we assume that the coefficients in $f$ satisfy (12) and that $C = [\lambda_1, \lambda_2]$ with $\lambda_1 \leq 0 \leq \lambda_2$ and $\lambda_2 - \lambda_1 > 0$.

Theorem 5.3  Let $\widetilde{f}$ be defined in $(\widetilde{Q})$, let $x \in \mathbb{R}$ and let $\alpha$ and $\beta$ be two positive numbers. Then, the following statements hold.

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

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

$$f_\alpha(p) + \frac{1}{2\beta}(p - x)^2 = \min \left\{ f_\alpha(u) + \frac{1}{2\beta}(u - x)^2 : u \in \mathbb{R} \right\}$$

$$= \min \left\{ f_\alpha(u) + \frac{1}{2\beta}(u - x)^2 : u \in C \right\}$$

$$= \min \left\{ \widetilde{f}_\alpha(u) + \frac{1}{2\beta}(u - x)^2 : u \in \mathbb{R} \right\},$$

where the first equation is due to $p \in$ prox$\beta f_\alpha(x)$, the second equation is due to $p \in C$, and the last one is due to Proposition 5.1, and hence, $p \in$ prox$\beta \widetilde{f}_\alpha(x)$.

(ii) Since $\lambda_2 \geq 0$, the inclusion $\lambda_2 \in$ prox$\beta \widetilde{f}_\alpha(x)$ together with Lemma 4.1 implies that $x \geq 0$ and for all $u \in [\lambda_1, \lambda_2]$,

$$\widetilde{f}_\alpha(u) + \frac{1}{2\beta}(u - x)^2 \geq \widetilde{f}_\alpha(\lambda_2) + \frac{1}{2\beta}(\lambda_2 - x)^2.$$
With the above inequality, when \( y > x \), we have that

\[
\tilde{f}_\alpha(\lambda_2) + \frac{1}{2\beta}(\lambda_2 - y)^2 = \tilde{f}_\alpha(\lambda_2) + \frac{1}{2\beta}(\lambda_2 - x)^2 + \frac{1}{2\beta}(y - x)(y + x - 2\lambda_2) \\
\leq \tilde{f}_\alpha(u) + \frac{1}{2\beta}(u - x)^2 + \frac{1}{2\beta}(y - x)(y + x - 2u) \\
= \tilde{f}_\alpha(u) + \frac{1}{2\beta}(u - y)^2
\]

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). \( \square \)

Theorem 5.3 tells us that for \( \tilde{f} \) as in (\( \mathcal{O} \)), \( \text{prox}_\beta \tilde{f}_\alpha \) will resemble the proximity operator of \( f_\alpha \) around the origin and the proximity operator of \( \iota_C \) elsewhere. Due to the number of parameters, there are a huge number of possible combinations. Rather than listing all of the combinations here, we provide the details for a specific function in Example 4 of Sect. 6.

We have shown that sparsity promoting quadratic and indicator functions have piecewise linear proximity operators which act as thresholding 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 [23], so our discussion is a comprehensive method for obtaining such 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 \( \ell_1 \)-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 [24]). The general form of the elastic net is the linear combination of the \( \ell_1 \)- and \( \ell_2 \)-norms as follows: \( f(x) = \frac{\lambda_1}{2} \|x\|^2 + \lambda_2 \|x\|_1 \), where \( \lambda_1 \) and \( \lambda_2 \) 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 \( C \).
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 $|\cdot| + \lambda [\cdot - \lambda]$, in Example 1, 2, 3, or 4, respectively. Then, $\text{env}_\alpha f(x) = \sum_{i=1}^{n} \text{env}_\alpha g(x_i)$, $\text{env}_\beta f_\alpha(x) = \sum_{i=1}^{n} \text{env}_\beta g_\alpha(x_i)$, and

$$\text{prox}_\alpha f(x) = \text{prox}_{\alpha g}(x_1) \times \text{prox}_{\alpha g}(x_2) \times \cdots \times \text{prox}_{\alpha g}(x_n),$$

$$\text{prox}_\beta f_\alpha(x) = \text{prox}_{\beta g_\alpha}(x_1) \times \text{prox}_{\beta g_\alpha}(x_2) \times \cdots \times \text{prox}_{\beta g_\alpha}(x_n),$$

where $\times$ denotes the Cartesian product. Therefore, in the following discussion we restrict our attention to $n = 1$.

6.1 Example 1: The Absolute Value Function

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

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

$$\text{prox}_{\alpha |\cdot|}(x) = \text{sgn}(x) \max\{0, |x| - \alpha\}$$

and

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

respectively. It is well known that $\text{prox}_{\alpha |\cdot|}$ is called the soft thresholding in the literature of wavelet [25] and $\text{env}_\alpha |\cdot|$ is Huber’s function in robust statistics [26]. Figure 5 shows the typical shape of the proximity operator of $f$.

![Figure 5](image-url)
As defined in $\mathcal{F}_\alpha$, for the absolute value function $f$,

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

This function $f_\alpha$ (see Fig. 6b) is identical to the minimax convex penalty (MCP) function given in [7], motivated from a statistics perspective.

The expression of $\text{prox}_{\beta f_\alpha}$ depends on the relative values of $\alpha$ and $\beta$. If $\beta < \alpha$, Lemma 5.5 gives

$$\text{prox}_{\beta f_\alpha}(x) = \begin{cases} 0, & \text{if } |x| \leq \beta, \\ \frac{\alpha}{\alpha - \beta} (|x| - \beta) \text{sgn}(x), & \beta < |x| \leq \alpha, \\ x, & \text{if } |x| \geq \alpha. \end{cases}$$ (22)

This is the firm thresholding operator [27]. If $\beta = \alpha$, Lemma 5.6 gives

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

Finally, if $\beta > \alpha$, Lemma 5.7 gives

$$\text{prox}_{\beta f_\alpha}(x) = \begin{cases} 0, & \text{if } |x| < \sqrt{\alpha\beta}, \\ [0, x], & \text{if } |x| = \sqrt{\alpha\beta}, \\ x, & \text{if } |x| > \sqrt{\alpha\beta}. \end{cases}$$ (24)

The proximity operator $\text{prox}_{\beta f_\alpha}$ for different values of $\alpha$ and $\beta$ is plotted in Fig. 7.

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

- Note that $\partial f(0) = [-1, 1]$. The results given in (22) (for $\beta < \alpha$) and (23) (for $\beta = \alpha$) exactly match the first two statements of Theorem 4.2. For $\beta > \alpha$, $\text{prox}_{\beta f_\alpha}(x) = 0$ for all $x \in [-\sqrt{\alpha\beta}, \sqrt{\alpha\beta}]$, which includes $[-\alpha, \alpha] = \alpha \partial f(0)$ as indicated in the third statement of Theorem 4.2.
Fig. 7  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 $\alpha$ and $\beta$.

- The operator $\text{prox}_{\alpha f}$ forces its argument to zero when the absolute value of the argument is less than a given threshold, and otherwise reduces the argument, in absolute value, by the amount of the threshold. Like $\text{prox}_{\alpha f}$, $\text{prox}_{\beta f_{\alpha}}$ forces its argument to zero when the absolute value is less than a given threshold, but it fixes arguments 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 $\ell_{0}$-norm on $\mathbb{R}$, that is, $|x|_{0}$ equals 1 if $x$ is nonzero, 0 otherwise. The proximity operator of $|\cdot|_{0}$ with parameter $\gamma$ 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 $f_{\alpha}$ is continuous, they have the same proximity operator. Moreover, by fixing $\alpha$ and varying the parameter $\beta$, 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 rectified linear unit (ReLU) function on $\mathbb{R}$ is

$$f(x) := \max\{0, x\},$$
Fig. 8 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 \( f_\alpha \) at zero is emphasized in black (solid-dotted).

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

Figure 8a depicts the graphs of \( f \) and \( \text{env} \alpha f \), while Fig. 8b presents the function \( f_\alpha \). The graph of \( \text{prox} \alpha f \) is given in Fig. 9.

As in Example 1, the expression of \( \text{prox} \beta f_\alpha \) depends on the relative values of \( \alpha \) and \( \beta \). If \( \beta < \alpha \),
\[
\text{prox}_{\beta f_\alpha}(x) = \begin{cases} 
  x, & \text{if } x \leq 0 \text{ or } x \geq \alpha, \\
  0, & \text{if } 0 \leq x \leq \beta, \\
  \frac{\alpha - \beta}{\alpha - \beta} (x - \beta), & \text{if } \beta \leq x \leq \alpha.
\end{cases}
\] 

(25)

If \( \beta = \alpha \),

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

(26)

Finally, if \( \beta > \alpha \),

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

(27)

Note that \( \partial f(0) = [0, 1] \). The results given in Eqs. (25) (for \( \beta < \alpha \)) and (26) (for \( \beta = \alpha \)) exactly match the first two statements of Theorem 4.2. For \( \beta > \alpha \), equation (27) shows that \( \text{prox}_{\beta f_\alpha}(x) = 0 \) for all \( x \in [0, \sqrt{\alpha \beta}] \), which includes \( [0, \alpha] = \alpha \partial f(0) \) as indicated in the third statement of Theorem 4.2 (Fig. 10).

Fig. 10 Example 2. Typical shapes of the proximity operator of \( f_\alpha \) for a \( \beta < \alpha \), b \( \beta = \alpha \), and c \( \beta > \alpha \)
6.3 Example 3: Elastic Net

The elastic net is a regularized regression method in data analysis that linearly combines the $\ell_1$ and $\ell_2$ 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 $(Q)$ with $a_1 = 1$, $a_2 = 1$, $b_1 = -1$, and $b_2 = 1$. Clearly, $f$ is nondifferentiable at the origin with $\arg\min_{x \in \mathbb{R}} f(x) = \{0\}$. Moreover, $\partial f(0) = \partial |\cdot|(0) = [-1, 1]$.

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

$$\text{prox}_\alpha f(x) = \max \left\{ 0, \frac{1}{\alpha + 1} (|x| - \alpha) \right\} \text{sgn}(x)$$

and

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

respectively.

The graphs of $f$ and $\text{env}_\alpha f$ are plotted in Fig. 11a. The graph of $\text{prox}_\alpha f$ is plotted in Fig. 11b. As in the case of the absolute value function, $\text{prox}_\alpha f$ sends all values between $\alpha$ and $-\alpha$ to zero. Unlike the absolute value, it also contracts elements outside of this interval toward the origin.

Now, $f_\alpha$ is

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

![Example 3. a The graphs of $f$ (solid) and $\text{env}_\alpha f$ (dotted); and b the graph of $\text{prox}_\alpha f$](image)
We remark that $f_{\alpha}$ is convex when $\alpha \geq 1$ and nonconvex when $\alpha < 1$. The graph of $f_{\alpha}$ for $\alpha \geq 1$ and $\alpha < 1$ is shown in Fig. 12a and b, respectively.

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

Case 1: $\alpha(\beta + 1) > \beta$. In this case, by Lemma 5.5 we have

$$\text{prox}_{\beta f_{\alpha}}(x) = \begin{cases} 0, & \text{if } |x| \leq \beta, \\ \frac{\alpha}{\alpha^{2} + \beta + \alpha} (x - \beta \text{sgn}(x)), & \text{if } \beta \leq |x| \leq \alpha(\beta + 1), \\ \frac{\alpha + 1}{\alpha^{2} + \alpha + 1} \left( x - \frac{\alpha \beta}{\alpha + 1} \text{sgn}(x) \right), & \text{if } \alpha(\beta + 1) \leq |x|. \end{cases}$$

(28)

Case 2: $\alpha(\beta + 1) = \beta$. By Lemma 5.6 we have

$$\text{prox}_{\beta f_{\alpha}}(x) = \begin{cases} 0, & \text{if } |x| \leq \beta, \\ [0, \alpha] \text{sgn}(x), & \text{if } |x| = \beta, \\ \frac{\alpha + 1}{\alpha^{2} + \alpha + 1} \left( x - \frac{\alpha \beta}{\alpha + 1} \text{sgn}(x) \right), & \text{if } \beta \leq |x|. \end{cases}$$

(29)

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

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

(30)

as in Lemma 5.7. Then, we have

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

(31)

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 Fig. 13.

Fig. 12 Example 3. The graph of $f_{\alpha}$ when $a \alpha \geq 1$ and $b \alpha < 1$. The singularity of $f_{\alpha}$ at zero is emphasized in black (solid-dotted).
Below are some comments on this example.

– In Examples 1 and 2, \( f_\alpha \) is nonconvex for any \( \alpha > 0 \). By Proposition 4.1, the function \( f_\alpha \) in Example 3 is convex if \( \alpha \geq 1 \) due to the 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 natural from Proposition 4.1. Since \( f \) is 1-strongly convex, 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 \( \alpha(\beta + 1) < \beta \).

– For the case of \( \beta \leq \alpha \), we know that \( \alpha(1 + \beta) > \beta \), so the proximity operator given (28) covers both statements 1 and 2 in Theorem 4.2.

– 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 (28) (resp. (29)) shows that this operator sends all elements in \( \beta \partial f(0) = [-\beta, \beta] \supset \alpha \partial f(0) \) to zero, fulfilling the third statement of Theorem 4.2. If \( \beta > \alpha(\beta + 1) \), we know that \( \alpha < 1 \), \( \beta > \alpha \frac{\alpha}{1 - \alpha} \), and \( \tau \) defined in (30) satisfying

\[
\tau = \frac{\alpha \beta}{\alpha + 1} + \sqrt{\alpha \beta \left(\frac{\alpha \beta + \alpha + 1}{\alpha + 1}\right)} > \frac{\alpha^2}{1 - \alpha^2} + \frac{\alpha}{1 - \alpha^2} > \alpha.
\]

Hence, the proximity operator given by (31) annihilates all elements in \( \tau \partial f(0) \), which strictly contains \( \alpha \partial f(0) \), once again fulfilling the third statement of Theorem 4.2.
6.4 Example 4: Absolute Value on an Interval Centered at the Origin

Let \( \lambda \) 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 of \((\tilde{Q})\) with \(a_1 = a_2 = 0, b_2 = -b_1 = 1\), and \(C = [-\lambda, \lambda]\). Its proximity operator and Moreau envelope with parameter \(\alpha\) 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 14 depicts the graphs of \(f\) and \(\text{env}_\alpha f\). We observe that on the interval \([-\lambda, \lambda]\) (the domain of \(f_\alpha\)), the envelope \(\text{env}_\alpha f\) is piecewise quadratic polynomial (Fig. 14a) if \(\alpha < \lambda\) and is simply a quadratic polynomial (Fig. 14b) if \(\alpha \geq \lambda\). Figure 15 depicts the graph of \(\text{prox}_\alpha f\). It turns out that the expression of \(\text{prox}_\alpha f\) for \(\alpha < \lambda\) is much more complicated than that for \(\alpha \geq \lambda\) as we will see below.

As both \(f\) and \(\text{env}_\alpha f\) depend on \(\alpha\) and \(\lambda\), the explicit expression for \(f_\alpha\) 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 \(\alpha \geq \lambda\).

![Figure 14](image-url)  

*Fig. 14* Example 4. The graphs of \(f\) (solid, dashed) and \(\text{env}_\alpha f\) (dotted) when a \(\alpha < \lambda\) and b \(\alpha > \lambda\).
Fig. 15 Example 4. The graph of \( \text{prox}_\alpha f \). Between \(-\alpha-\lambda\) and \(\alpha+\lambda\), \( \text{prox}_\alpha f \) is the soft thresholding operator with sparsity parameter \(\alpha\); otherwise, it projects onto this interval.

Fig. 16 Example 4. The graph of \( f_\alpha \) when \(\alpha<\lambda\) with the singularity of \( f_\alpha \) at zero emphasized in black (solid-dotted). Further, we see that \( f_\alpha \) agrees with Example 1 on \([\lambda, \lambda]\).

Case 1: \( \alpha < \lambda \). In this case, we get (see Fig. 16)

\[
f_\alpha(x) = f(x) - \text{env}_\alpha f(x) = \begin{cases} 
\frac{\alpha}{2} - \frac{1}{2\alpha}(|x| - \alpha)^2, & \text{if } |x| \leq \alpha, \\
\frac{\alpha}{2}, & \text{if } \alpha \leq |x| \leq \lambda, \\
+\infty, & \text{if } \lambda < |x|. 
\end{cases}
\] (32)

Depending on the values of \(\alpha, \beta,\) and \(\lambda\), 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

\[
\text{prox}_{\beta f_\alpha}(x) = \begin{cases} 
\max \left\{ 0, \frac{\alpha|x| - \beta}{\alpha - \beta} \right\} \text{sgn}(x), & \text{if } |x| \leq \alpha, \\
\text{sgn}(x)[0, \alpha], & \text{if } |x| = \alpha, \\
\text{sgn}(x) \min\{|x|, \lambda\} \text{sgn}(x), & \text{if } \alpha < |x|. 
\end{cases}
\] (33)

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

\[
\text{prox}_{\beta f_\alpha}(x) = \begin{cases} 
0, & \text{if } |x| < \alpha, \\
\text{sgn}(x)[0, \alpha], & \text{if } |x| = \alpha, \\
\text{sgn}(x) \min\{|x|, \lambda\}, & \text{if } \alpha < |x|. 
\end{cases}
\] (34)

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

\[
\text{prox}_{\beta f_\alpha}(x) = \begin{cases} 
0, & \text{if } |x| < \sqrt{\alpha\beta}, \\
0, \text{sgn}(x)\sqrt{\alpha\beta}, & \text{if } |x| = \sqrt{\alpha\beta}, \\
\min\{|x|, \lambda\} \text{sgn}(x), & \text{if } \sqrt{\alpha\beta} < |x|. 
\end{cases}
\] (35)
Example 4. The graph of \( f_\alpha \) when \( \lambda \leq \alpha \) with the singularity of \( f_\alpha \) at zero emphasized in black (solid-dotted). As before, \( f_\alpha \) agrees with Example 1 on \([-\lambda, \lambda]\), but is cut off before it plateaus.

**Case 1.4 : \( \alpha < \lambda < \beta \)** We have

\[
\text{prox}_{f_\alpha} (x) = \begin{cases} 
0, & \text{if } |x| < \frac{\alpha \beta + \lambda^2}{2\lambda}, \\
\{0, \lambda \text{ sgn}(x)\}, & \text{if } |x| = \frac{\alpha \beta + \lambda^2}{2\lambda}, \\
\lambda \text{ sgn}(x), & \text{if } \frac{\alpha \beta + \lambda^2}{2\lambda} < |x|. 
\end{cases}
\] (36)

We now move on to the second main case.

**Case 2 : \( \lambda \leq \alpha \)** In this case, we get (see Fig. 17)

\[
f_\alpha(x) = \begin{cases} 
\frac{\alpha}{2} - \frac{1}{2\alpha} (|x| - \alpha)^2, & \text{if } |x| \leq \lambda, \\
+\infty, & \text{otherwise.}
\end{cases}
\]

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

**Case 2.1 : \( \beta < \alpha \)** In this case, we have that

\[
\text{prox}_{f_\alpha} (x) = \begin{cases} 
0, & \text{if } |x| \leq \beta, \\
\frac{\alpha(|x| - \beta)}{\alpha - \beta} \text{ sgn}(x), & \text{if } \beta \leq |x| \leq \beta + \frac{\alpha - \beta}{\alpha} \lambda, \\
\lambda \text{ sgn}(x), & \text{if } \beta + \frac{\alpha - \beta}{\alpha} \lambda < |x|. 
\end{cases}
\] (37)

**Case 2.2 : \( \beta = \alpha \)** In this case, we have

\[
\text{prox}_{f_\alpha} (x) = \begin{cases} 
0, & \text{if } |x| < \alpha, \\
\text{sgn}(x)[0, \alpha], & \text{if } |x| = \alpha, \\
\lambda \text{ sgn}(x), & \text{if } \alpha < |x|. 
\end{cases}
\] (38)

**Case 2.3 : \( \beta > \alpha \)** Similar to Case 1.4, we get

\[
\text{prox}_{f_\alpha} (x) = \begin{cases} 
0, & \text{if } |x| \leq \beta - \frac{\beta - \alpha}{2\alpha} \lambda, \\
\text{sgn}(x)[0, \alpha], & \text{if } |x| = \beta - \frac{\beta - \alpha}{2\alpha} \lambda, \\
\lambda \text{ sgn}(x), & \text{if } \beta - \frac{\beta - \alpha}{2\alpha} \lambda < |x|. 
\end{cases}
\] (39)

To end up this example, we comment on this example in comparison with Theorem 4.2.
Some typical shapes of \( \text{prox}_{\beta f_\alpha} \) are depicted in Fig. 18.

- Note that \( \partial f(0) = [-1, 1] \). For \( \beta < \alpha \), both Eqs. (33) and (37) show that the operator \( \text{prox}_{\beta f_\alpha} \) annihilates all elements in \( \beta \partial f(0) = [-\beta, \beta] \) as required by the first statement of Theorem 4.2.

- For \( \beta = \alpha \), both Eqs. (34) and (38) show that the operator \( \text{prox}_{\beta f_\alpha} \) annihilates all elements in \( \text{ri}(\alpha \partial f(0)) = (-\alpha, \alpha) \) as described in the second statement of Theorem 4.2.

- For \( \beta > \alpha \), since \( \sqrt{\alpha \beta} > \alpha \) when \( \alpha < \beta < \lambda \), \( \frac{\alpha \beta + \lambda^2}{2 \alpha} > \alpha \) when \( \alpha < \lambda < \beta \), and \( \beta - \frac{\beta - \alpha}{2 \alpha} \lambda > \alpha \) when \( \beta > \alpha \geq \lambda \), Eqs. (35), (36), and (39) show that the operator \( \text{prox}_{\beta f_\alpha} \) annihilates all elements in \( \alpha \partial f(0) = [-\alpha, \alpha] \) as described in the third
statement of Theorem 4.2. In each case, all elements in an open set containing \([-\alpha, \alpha]\) are sent to zero.

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

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 Sect. 4, among other properties related to the structure of these functions. In Sect. 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.

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

References

1. Candès, E., Tao, T.: Near optimal signal recovery from random projections: universal encoding strategies? IEEE Trans. Inf. Theory 52(12), 5406–5425 (2006)
2. Mallat, S.: A Wavelet Tour of Signal Processing, 2nd edn. Academic Press, New York, USA (1999)
3. Suter, B.W.: Multirate and Wavelet Signal Processing. Academic Press, San Diego, USA (1997)
4. Tibshirani, R.: Regression shrinkage and selection via the LASSO. J. Roy. Stat. Soc. B 58, 267–288 (1996)
5. Frank, I.E., Friedman, J.H.: A statistical view of some chemometrics regression tools (with discussion). Technometrics 35, 109–148 (1993)
6. Fan, J., Li, R.: Variable selection via nonconcave penalized likelihood and its oracle properties. J. Am. Stat. Assoc. 96(456), 1348–1360 (2001)
7. Zhang, C.H.: Nearly unbiased variable selection under minimax concave penalty. Ann. Stat. 38(2), 894–942 (2010)
8. Bolte, J., Daniilidis, A., Ley, O., Mazet, L.: Characterizations of lojasiewicz inequalities: subgradient flows, talweg, convexity. Trans. Am. Math. Soc. 362, 3319–3363 (2010)
9. Cannarsa, P., Sinestrari, C.: Semiconcave Functions, Hamilton–Jacobi Equations, and Optimal Control, Progress in Nonlinear Differential Equations and Their Applications, vol. 58. Birkhauser, Boston, USA (2004)
10. Tao, P., Le Thi, H.: Convex analysis approach to dc programming: theory, algorithms and applications. Acta Mathematica Vietnamica 22(1), 289–355 (1997)
11. Soubies, E., Blanc-Feraud, L., Aubert, G.: A unified view of exact continuous penalties for \(\ell_2-\ell_0\) minimization. SIAM J. Optim. 27(3), 2034–2060 (2017)
12. Moreau, J.J.: Proximité et dualité dans un espace hilbertien. Bull. Soc. Math. France 93, 273–299 (1965)
13. Attouch, H., Bolte, J., Svaiter, B.: Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods. Math. Program. Ser. A 137, 91–129 (2013)
14. Bauschke, H.L., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. AMS Books in Mathematics. Springer, New York (2011)
15. Combettes, P., Wajs, V.: Signal recovery by proximal forward–backward splitting. Multiscale Model. Simul. SIAM Interdiscip. J. 4, 1168–1200 (2005)
16. Schwarz, G.: Estimating the dimension of a model. Ann. Stat. 6(2), 461–464 (1978)
17. Chen, S., Donoho, D., Saunders, M.: Atomic decomposition by basis pursuit. SIAM J. Sci. Comput. 20, 33–61 (1998)
18. Selesnick, I.: Total variation denoising via the moreau envelope. IEEE Signal Process. Lett. 24(2), 216–220 (2017)
19. Hiriart-Urruty, J.B., Lemarechal, C.: Convex Analysis and Minimization Algorithms I, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 305. Springer-Verlag, Berlin (1993)
20. Micchelli, C.A., Shen, L., Xu, Y.: Proximity algorithms for image models: denoising. Inverse Prob. 27, 045,009 (2011). (30pp)
21. Donoho, D., Johnstone, I.: Ideal spatial adaptation by wavelet shrinkage. Biometrika 81, 425–455 (1994)
22. Clarke, F.H.: Optimization and Nonsmooth Analysis. Wiley, New York (1983)
23. Planiden, C., Wang, X.: Epi-convergence: the Moreau envelope and generalized linear-quadratic functions. J. Optim. Theory Appl. 177(1), 21–63 (2018)
24. Zou, H., Hastie, T.: Regularization and variable selection via the elastic net. J. Roy. Stat. Soc. B 67, 301–320 (2005)
25. Donoho, D.: Compressed sensing. IEEE Trans. Inf. Theory 52(4), 1289–1306 (2006)
26. Huber, P.: Robust Statistics, 2nd edn. John Wiley & Sons Inc., Hoboken, New Jersey (2009)
27. Gao, H.Y., Bruce, A.G.: WaveShrink with firm shrinkage. Stat. Sin. 7, 855–874 (1997)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.