A General Framework for Sparsity-Based Denoising and Inversion

Ali Gholami and S. Mohammad Hosseini

***Abstract*—Estimating a reliable and stable solution to many problems in signal processing and imaging is based on sparse regularizations, where the true solution is known to have a sparse representation in a given basis. Using different approaches, a large variety of regularization terms have been proposed in literature. While it seems that all of them have so much in common, a general potential function which fits most of them is still missing. In this paper, in order to propose an efficient reconstruction method based on a variational approach and involving a general regularization term (including most of the known potential functions, convex and nonconvex), we deal with i) the definition of such a general potential function, ii) the properties of the associated "proximity operator" (such as the existence of a discontinuity), and iii) the design of an approximate solution of the general "proximity operator" in a simple closed form. We also demonstrate that a special case of the resulting "proximity operator" is a set of shrinkage functions which continuously interpolate between the soft-thresholding and hard-thresholding. Computational experiments show that the proposed general regularization term performs better than ℓ_{p} -penalties for sparse approximation problems. Some numerical experiments are included to illustrate the effectiveness of the presented new potential function.**

***Index Terms*—Potential function, proximity operator, regularization, sparse approximation.**

I. Introduction

NOWADAYS, sparse approximation techniques are the most developing and promising methods for noise suppression and solving other inverse problems arising in various fields of study. Although sparsity is not a new concept and its success in the field of signal processing and inversion dates back to the 1970s and 1980s (e.g. [1]–[4]), since 1990s, it has been at the core of a rapidly evolving and very active area of research because of the developments in the field of computational harmonic analysis for multiscale sparsifying transforms (e.g., wavelets and curvelets). Sparsity has successfully been applied to many areas of signal/image processing and inversion, including noise suppression [5]–[8], deconvolution [2], [9], and tomography [10], to name only a few. Moreover, sparsity is the basic idea in compressed sensing theory, which allows signals and images to be reconstructed from a partial amount of observed information and provides an efficient framework for nonlinear sampling [11], [12]. The ill-posedness nature of such problems requires the regularization techniques to be used for making a meaningful solution [13], [14]. In other words, the solution that should have some required properties, while satisfying the observed data, can be formulated as the following general constrained optimization problem with an error bound $\varepsilon > 0$

$$\hat{\boldsymbol{x}} := \arg\min_{\boldsymbol{x}} \ \mathrm{Reg}(\boldsymbol{L}\boldsymbol{x}) \quad \text{s.t.} \quad \|\boldsymbol{G}\boldsymbol{x} - \boldsymbol{y}\|_2^2 < \varepsilon \tag{1}$$

if this bound is available, or

$$\hat{\boldsymbol{x}} := \arg\min_{\boldsymbol{x}} \ \|\boldsymbol{G}\boldsymbol{x} - \boldsymbol{y}\|_2^2 \quad \text{s.t.} \quad \mathrm{Reg}(\boldsymbol{L}\boldsymbol{x}) < \rho \tag{2}$$

if some information bound $\rho > 0$ on the regularity of the solution is known. Otherwise, such problem can be formulated as an unconstraint optimization problem

$$\hat{\boldsymbol{x}} := \arg\min_{\boldsymbol{x}} \left\{ \frac{1}{2}\|\boldsymbol{G}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \tau \mathrm{Reg}(\boldsymbol{L}\boldsymbol{x}) \right\} \tag{3}$$

with Lagrange multiplayer $\tau \in \mathsf{R}^+$, set of positive real numbers, known as the regularization parameter. In expressions (1)–(3), $\boldsymbol{y} \in \mathsf{R}^{\mathrm{M}}$ is the observed data vector, $\boldsymbol{G}$ is the observation matrix called forward operator, and $\boldsymbol{x} \in \mathsf{R}^{\mathrm{N}}$ is the vector of unknown parameters which is assumed to have a specific statistical property under known linear analysis operator $\boldsymbol{L}$ called regularization operator [15]–[18]. The set of $\boldsymbol{L}$ rows is typically a wavelet basis (orthogonal, $\boldsymbol{L}$ is square, or redundant, $\boldsymbol{L}$ is tall, i.e., has more rows than columns). It should be mentioned that (3) corresponds to an analysis approach and when $\boldsymbol{L}$ is a wavelet transform that is either orthonormal or tight frame, or more generally if $\boldsymbol{L}$ is invertible, it can easily be converted to a synthesis formulation [18]. The algorithm of this paper will be based on the synthesis form of (3). However, for a more general case where (3) is convex, the proximity operator of this algorithm can be used to solve it by the splitting Bregman algorithm [19]. In (3), $\mathrm{Reg}(\boldsymbol{v}) = \sum_i \varphi(v_i)$, where v_i is the ith entry of vector $\boldsymbol{v}$ of length equal to the number of $\boldsymbol{L}$ rows, is a function which measures some properties of the solution and φ, the so-called potential function [20], which introduces the statistical properties of the coefficients $\boldsymbol{L}\boldsymbol{x}$, is nondecreasing, non-negative and not necessarily convex on R^+. Although [20] analyzed continuously differentiable potential functions, here, we assume that φ is irregular (not differentiable) at zero point to produce sparse solutions. Many different forms of the potential function φ have been proposed in literature to express the sparse nature of the

Manuscript received May 18, 2011; revised July 28, 2011; accepted July 28, 2011. Date of publication August 12, 2011; date of current version October 12, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Jerome Idier.

A. Gholami is with the Institute of Geophysics, University of Tehran, 14155-6466, Tehran, Iran (e-mail: agholami@ut.ac.ir).

S. M. Hosseini is with the Faculty of Mathematical Sciences, Tarbiat Modares University, 14115-175, Tehran, Iran (e-mail: hossei_m@ modares.ac.ir).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2011.2164074

coefficient vector $\boldsymbol{Lx}$, for instance, ℓ_{p} (quasi-) norms [21]–[23], Cauchy-norm [4], [20], and the measure used by Geman and McClure [3]. Although all of these potential functions seem to have so much in common, a general potential function which fits most of them is still missing.

Recently, successful iterative shrinkage/thresholding algorithms have been proposed to efficiently solve optimization problems (1)–(3), in a convex and nonconvex formulations [16], [19], [23]–[29], by considering the following problem:

$$\mathrm{prox}_{\tau\varphi}(\boldsymbol{y}) := \arg\min_{\boldsymbol{x}} \left\{ \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + \tau \sum_i \varphi(x_i) \right\}. \quad (4)$$

This problem is proximity operator when φ is lower semicontinuous and proper (interested readers are referred to [28] and [29] for more details and properties of proximity operators of convex functionals). Furthermore, while successful applications of proximity operators caused it to receive much attention of researchers [23], [30], a closed form of such operators is only known for some specific forms of φ [27], [29], whereas, in the field of sparse reconstruction, it has been shown that good results can be obtained for some other forms of φ for which there are no simple closed form for the corresponding proximity operators [23], [31]. Thus, a fast approximation of $\mathrm{prox}_{\tau\varphi}$, in closed form, for a general φ is desirable and interesting.

In this paper, we deal with a generalization of ℓ_{p} -penalties by developing a general potential function which includes most of the known potential functions as its particular cases. The new potential function depends on two continuously varying parameters. Different choices of these parameters lead to a set of functions which continuously interpolate the well-known potential functions. Then, we try to find an approximate solution to the corresponding general proximity operator by truncating Taylor series expansion of potential function derivative. It is shown numerically that taking even the first term of the series results to a simple and efficient function in closed form, which approximates the original proximity operator with a high degree of accuracy. As a particular case, the proposed proximity operator for the general potential function is a set of shrinkage functions which continuously interpolate between the two most common thresholding operators: the so-called soft-thresholding and hard-thresholding operators. In addition, our general thresholding function converges to the well-known non-negative garrote shrinkage function for a specific parameter [6], [7]. The content of paper will appear in sections as follows.

In Section II, the new potential function is presented along with an illustrative discussion concerning its interesting properties. In Section III, general proximity operator and our iterative algorithm for solving sparse-regularized problems are analyzed. Illustrative examples are also included to show a clear performance of the new approach. Section IV is devoted to a generalized thresholding rule. In Section V, some numerical tests are experimented.

II. A Generalized Potential Function

In this section, we propose a new set of potential functions that brings many well-known potential functions adopted in many fields of science and engineering dealing with inverse problem theory under a single formula. The new potential function is formulated as

$$\varphi_{\mathrm{q}}^{\mathrm{p}}(x) = \begin{cases} \frac{1}{\mathrm{q}}\left(1-\left(|x|^{\mathrm{p}}+1\right)^{-\mathrm{q}}\right) & \text{if } \mathrm{q}\neq 0 \\ \ln\left(|x|^{\mathrm{p}}+1\right) & \text{if } \mathrm{q}=0 \end{cases} \quad (5)$$

for $\mathrm{p}\in(0,2]$, $\mathrm{q}\in[-1,\infty)$, and $x\in\mathsf{R}$. Note that $\varphi_{\mathrm{q}}^{\mathrm{p}}(0)=0$ as is required for practical reasons [20]. The properties of function $\varphi_{\mathrm{q}}^{\mathrm{p}}$ for different values of p and q, that can be proved easily, are expressed in three cases [Fig. 1(a)]:

Case 1) For $\mathrm{p}\in(0,1)$ and $\mathrm{q}\in[-1,\infty)$, $\frac{d^2}{dx^2}\varphi_{\mathrm{q}}^{\mathrm{p}}(x) < 0, \forall x>0$, so $\varphi_{\mathrm{q}}^{\mathrm{p}}(x)$ is strictly concave.

Case 2) For $\mathrm{p}\in[1,2]$ and $\mathrm{q}\in\left[-1,-\frac{1}{\mathrm{p}}\right]$, $\frac{d^2}{dx^2}\varphi_{\mathrm{q}}^{\mathrm{p}}(x)\geq 0, \forall x>0$, so $\varphi_{\mathrm{q}}^{\mathrm{p}}(x)$ is convex.

Case 3) For $\mathrm{p}\in[1,2]$ and $\mathrm{q}\in\left(-\frac{1}{\mathrm{p}},\infty\right)$, $\varphi_q^p(x)$ is convex for $x\leq\left(\frac{\mathrm{p}-1}{\mathrm{pq}+1}\right)^{\frac{1}{\mathrm{p}}}$ and strictly concave for $x>\left(\frac{\mathrm{p}-1}{\mathrm{pq}+1}\right)^{\frac{1}{\mathrm{p}}}$.

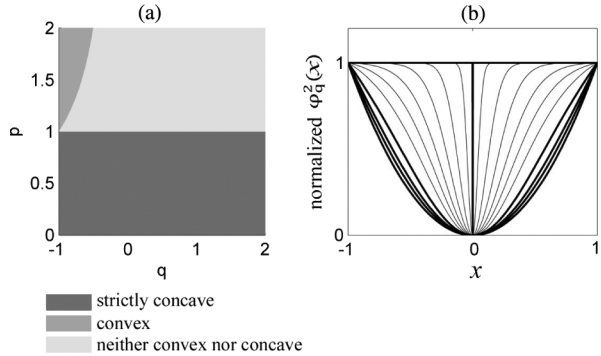


Fig. 1. (a) Convexity of potential function $\varphi_{\mathrm{q}}^{\mathrm{p}}$ for different values of p and q. Note that the line $\mathrm{q}=-1$ corresponds to the conventional functions $|x|^{\mathrm{p}}$. (b) The set of functions φ_{q}^2 for $\mathrm{q}\in[-1,10^{10}]$ shows that φ_{q}^2 converges to the Heaviside step function as q approaches infinity. Thicker curves represent conventional potential functions. For visual purposes, the curves are normalized by φ_{q}^2 (1).

TABLE I
Potential Function $\varphi_{\mathrm{q}}^{\mathrm{p}}$ for Some Specific Values of p and q

p	q	Expression of $\varphi_{\mathrm{q}}^{\mathrm{p}}(x)$	References
p	-1	$\lvert x\rvert^{\mathrm{p}}$	[12], [1], [22], [23], [16]
2	0	$\ln\left(x^2+1\right)$	[4], [20]
2	-0.5	$2\sqrt{x^2+1}-2$	[20]
2	1	$\frac{x^2}{x^2+1}$	[3], [20]
1	1	$\frac{\lvert x\rvert}{\lvert x\rvert+1}$	[32]

The potential function defined in (5) for some different choices of p and q leads to some known potential functions as presented in Table I. An important property of $\varphi_{\mathrm{q}}^{\mathrm{p}}$ is that when q is sufficiently large, the operator Reg performs as a scaled ℓ_0-norm regularizer independently of p, i.e., minimizing the number of nonzero entries of the solution. Examples of functions φ_{q}^2 for some different values of q in the interval $[-1,10^{10}]$ are shown in Fig. 1(b). For visual purposes, the curves are

normalized by φ_{q}^2 (1). Note how the potential functions φ_{q}^2 converge to Heaviside step function as q approaches infinity.

III. The General Proximity Operator

In this section, we determine a closed form for the proximity operator of the general potential function $\varphi_{\mathrm{q}}^{\mathrm{p}}$ defined in (5). Since the regularization term in (4) is separable and for every $x \geq 0$, $\varphi_{\mathrm{q}}^{\mathrm{p}}(x) \geq 0$, the minimization problem (4) is reduced to a one-dimensional minimization problem in which the cost function is

$$\mathrm{E}_y(x) = \frac{1}{2}(x-y)^2 + \tau\varphi_{\mathrm{q}}^{\mathrm{p}}(x). \tag{6}$$

Without loss of generality, we assume y is positive. For a given positive value y, the global minimizer of $\mathrm{E}_y(x)$, $\hat{x}$, occurs in the interval $[0, y]$ and its location depends on the parameter τ (when $\tau \to 0$, $\hat{x} \to y$, and when $\tau \to \infty$, $\hat{x} \to 0$). Clearly from E_y and the potential function it is easily seen that for a given τ, $\hat{x}$ is a solution of the following equation:

$$-y + x + \tau\frac{d}{dx}\varphi_{\mathrm{q}}^{\mathrm{p}}(x) = 0 \tag{7}$$

or has the value 0, the critical point of E_y, when $\varphi_{\mathrm{q}}^{\mathrm{p}}$ is not differentiable at the origin. It can also be shown that if there exists η_τ (the jumping point for which $\mathrm{E}_{\eta_\tau}(x)$ at the minima can be the same), it should satisfy the following equation (see [22]):

$$x^2 - 2\eta_\tau x + 2\tau\varphi_{\mathrm{q}}^{\mathrm{p}}(x) = 0. \tag{8}$$

Substituting η_τ from (7) into (8) yields

$$x^2 + 2\tau\left(x\frac{d}{dx}\varphi_{\mathrm{q}}^{\mathrm{p}}(x) - \varphi_{\mathrm{q}}^{\mathrm{p}}(x)\right) = 0 \tag{9}$$

and then inserting the maximum root of (9) into (7) gives η_τ.

Obviously, for $\mathrm{p} \in (0, 1)$, this discussion and the results in (7) and (9) are in accordance with those of Lemmas 3.2 and 3.10 of [23].

Now in order to obtain an approximation of the proximity operator, we need to have exact or some approximate roots of (7) and (9).

A. Determination of the Root of (7)

In view of the fact that the root of (7) is somewhere between 0 and y, it can be represented as

$$y - \omega, \quad 0 < \omega \leq y. \tag{10}$$

Substituting (10) into (7) implies

$$\omega = \tau\frac{d}{dx}\varphi_{\mathrm{q}}^{\mathrm{p}}(y-\omega) \tag{11}$$

which is to be solved for ω. To this end, we expand $\frac{d}{dx}\varphi_{\mathrm{q}}^{\mathrm{p}}(y-\omega)$ about y by the Taylor's formula, as

$$\frac{d}{dx}\varphi_{\mathrm{q}}^{\mathrm{p}}(y-\omega) = \frac{d}{dx}\varphi_{\mathrm{q}}^{\mathrm{p}}(y) - \omega\frac{d^2}{dx^2}\varphi_{\mathrm{q}}^{\mathrm{p}}(y) + \frac{\omega^2}{2}\frac{d^3}{dx^3}\varphi_{\mathrm{q}}^{\mathrm{p}}(y) - \cdots. \tag{12}$$

Substituting (12) into (11) yields

$$\omega = \tau\left(\frac{d}{dx}\varphi_{\mathrm{q}}^{\mathrm{p}}(y) - \omega\frac{d^2}{dx^2}\varphi_{\mathrm{q}}^{\mathrm{p}}(y) + \frac{\omega^2}{2}\frac{d^3}{dx^3}\varphi_{\mathrm{q}}^{\mathrm{p}}(y) - \cdots\right) \tag{13}$$

TABLE II
Analytic Solution of (18) for Different Values of p and q. In the Table, $\mathrm{u} = \sqrt[3]{\tau + \sqrt{\tau^2 - \left(\frac{\tau}{3}\right)^3}}$

p	q	Expression of $\varphi_{\mathrm{q}}^{\mathrm{p}}(x)$	Root of (19)
p	-1	$\lvert x\rvert^{\mathrm{p}}$	$\left(2\tau(1-\mathrm{p})\right)^{\frac{1}{2-\mathrm{p}}}$
1	1	$\frac{\lvert x\rvert}{\lvert x\rvert+1}$	$\sqrt{2\tau}-1$
1	2	$\frac{1}{2}\left(1-\frac{1}{(\lvert x\rvert+1)^2}\right)$	$\mathrm{u}+\frac{\tau}{3\mathrm{u}}-1$,
2	1	$\frac{x^2}{x^2+1}$	$\sqrt{\tau+\tau\sqrt{1-4\tau^{-1}}-1}$

or in other words

$$\sum_{\mathrm{n}=0}^{\infty}\frac{(-1)^{\mathrm{n}}\omega^{\mathrm{n}}}{\mathrm{n}!}\frac{d^{\mathrm{n}+1}}{dx^{\mathrm{n}+1}}\varphi_{\mathrm{q}}^{\mathrm{p}}(y) = \frac{\omega}{\tau}. \tag{14}$$

Equation (14) shows that ω depends only on parameter τ and derivatives of $\varphi_{\mathrm{q}}^{\mathrm{p}}$ at point y. For most values of p and q, in sparse reconstruction schemes, a truncation of series (14) up to the first order derivative provides satisfactory results. However, one may also use truncation up to a higher order term, say, second order derivative, to obtain more accurate solutions. In former case, using linear terms only, the solution of (14) results to

$$\omega \cong \tau\frac{d}{dx}\varphi_{\mathrm{q}}^{\mathrm{p}}(y) \tag{15}$$

where one should have $0 < \omega \leq y$. Consequently, the root of (7) is approximated as

$$\left(y - \tau\frac{d}{dx}\varphi_{\mathrm{q}}^{\mathrm{p}}(y)\right)_+ \tag{16}$$

where $(\cdot)_+$ stands for "the positive part," i.e., $(x)_+ = x$, if $x > 0$, and $(x)_+ = 0$, if $x \leq 0$.

B. Determination of the Root of (9)

Obviously, zero is an inevitable root of (9). However, if (9) has also a real positive root, it should be a solution of the following equation:

$$\mathrm{r}_\tau(x) = 1 + 2\tau\frac{d}{dx}\left(\frac{\varphi_{\mathrm{q}}^{\mathrm{p}}(x)}{x}\right) = 0, \quad x > 0 \tag{17}$$

or

$$\frac{d}{dx}\left(\frac{\varphi_{\mathrm{q}}^{\mathrm{p}}(x)}{x}\right) = \frac{-1}{2\tau} \quad x > 0 \tag{18}$$

which can be found numerically. Analytic solutions of (18) for some specific forms of $\varphi_{\mathrm{q}}^{\mathrm{p}}$ are presented in Table II. The same result was obtained in [22] and [23] for the case $\varphi_{-1}^{\mathrm{p}} = |x|^{\mathrm{p}}$. The main concern here is when the existence of a root for r_τ is unclear or when $\varphi_{\mathrm{q}}^{\mathrm{p}}$ is neither convex nor concave.

Proposition 1: Expression r_τ defined in (17) for $\varphi_{\mathrm{q}}^{\mathrm{p}}$ defined as in (5) and $\tau \in \mathrm{R}^+$, a) has a single real, positive root for $\mathrm{p} \in (0, 1)$ and $\mathrm{q} \in [-1, \infty)$, b) does not have a real, positive root for $\mathrm{p} \in [1, 2]$ and $\mathrm{q} \in \left[-1, -\frac{1}{\mathrm{p}}\right]$, and c) either has two

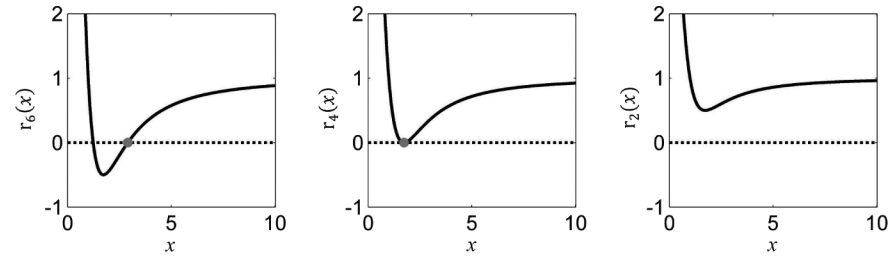


Fig. 2. The plot of function r_τ for φ_1^2 and three different values of τ showing that the existence of root of r_τ depends on the values of τ when $\varphi_{\mathrm{q}}^{\mathrm{p}}$ is neither convex nor concave.

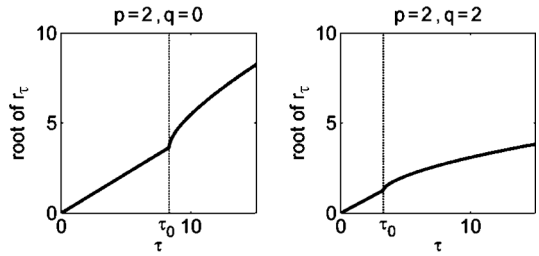


Fig. 3. Root of r_τ as a function of parameter τ for two different potential functions $\varphi_{\mathrm{q}}^{\mathrm{p}}$, extended for $\tau < \tau_0$ using the proposed approach.

positive roots or has no positive root at all for $\mathrm{p} \in [1, 2]$ and $\mathrm{q} \in (-\frac{1}{\mathrm{p}}, \infty)$.

Proof: See the Appendix.

For neither convex nor concave $\varphi_{\mathrm{q}}^{\mathrm{p}}$, when r_τ has a root, it must have two positive roots and hence a local minimum occurs between these roots at, say $\bar{x}$, which is a solution of

$$\frac{d^2}{dx^2}\varphi_{\mathrm{q}}^{\mathrm{p}}(x) = 2\frac{d}{dx}\left(\frac{\varphi_{\mathrm{q}}^{\mathrm{p}}(x)}{x}\right) \tag{19}$$

(see the Appendix). Note that $\bar{x}$ is independent of τ; however, the value of $\mathrm{r}_\tau(x)$ at this point varies as τ changes. From (18) and (19), for this value of $\bar{x}$, we obtain τ_0 such that

$$\frac{1}{\tau_0} = -\frac{d^2}{dx^2}\varphi_{\mathrm{q}}^{\mathrm{p}}(\overline{x}). \tag{20}$$

For $\tau < \tau_0$, $\mathrm{r}_\tau(x) > 0, \forall x > 0$, so it does not have a real root. Fig. 2 shows this situation corresponding to φ_1^2 using three different values of τ. As it is seen from Fig. 2, r_6 has two distinct positive roots, as expected, r_4 has a double root at $\bar{x}$, and r_2 does not have any root. This behavior can be justified by the fact that we already have assumed irregularity of $\varphi_{\mathrm{q}}^{\mathrm{p}}$ at the origin when we were generating expression (8). This assumption is needed to produce sparse solutions.

For potential functions regular at zero, usually in practice, the undesired extremum of E_y (occurring near zero) is moved toward origin by contracting the horizontal axis of the potential function using a scaling parameter, say $\varphi_{\mathrm{q}}^{\mathrm{p}}(\gamma^{-1}x)$, for a small non-negative parameter γ (see [20]). However, here we treat this difficulty more specifically. What we need is only to find a root of r_τ for all values of τ for a given potential function.

Here, for $\tau < \tau_0$, we extend the range of the values of the roots by scaling the horizontal axis of the $\mathrm{r}_{\tau_0}(x)$ by $\frac{\tau_0}{\tau}$ to get its root given by

$$\frac{\overline{x}}{\tau_0}\tau = [1 - \mathrm{r}_\tau(\overline{x})]\,\overline{x} \quad \tau < \tau_0 \tag{21}$$

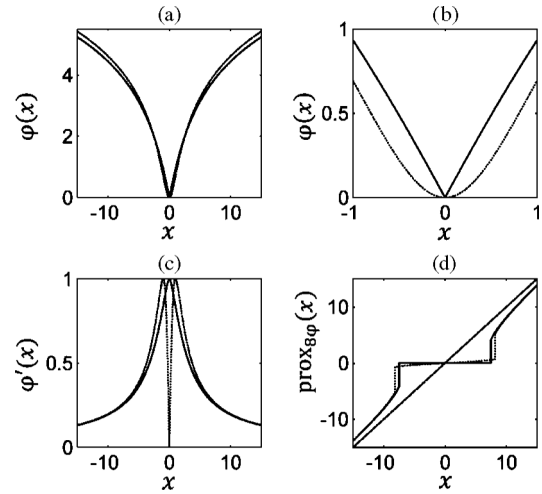


Fig. 4. Two potential functions (a) that are exaggerated near the origin (b), their absolute value of derivative (c), and the corresponding proximity operators (d). In all plots, the dashed line corresponds to the original function φ_0^2 and the solid line corresponds to its alternative obtained by the proposed method. In plot (d), the dashed line shows the input x.

which is obtained using (19) and the definition of r_τ. In (21), $\bar{x}$ is a solution of (19) and τ_0 is obtained from (20). Fig. 3 shows roots of r_τ as a function of parameter τ extended for $\tau < \tau_0$ using the proposed approach for two different potential functions.

C. An Approximation of the Proximity Operator

Knowing the threshold η_τ for $\varphi_{\mathrm{q}}^{\mathrm{p}}$ and using (16), the proximity operator of $\varphi_{\mathrm{q}}^{\mathrm{p}}$ is now approximated as follows:

$$\mathrm{prox}_{\tau\varphi_{\mathrm{q}}^{\mathrm{p}}}(y) \cong \begin{cases} \left(1 - \tau\mathrm{p}\frac{|y|^{\mathrm{p}-2}}{[|y|^{\mathrm{p}}+1]^{\mathrm{q}+1}}\right)_+ y, & \text{if } |y| > \eta_\tau \\ 0, & \text{otherwise.} \end{cases} \tag{22}$$

We show the performance of this approximation by considering two examples.

Example 1: We show how our approach moves the first local minimum of a smooth potential function toward origin to produce a sparse solution. Let us consider the proximity operator of nonconvex and smooth potential function $\varphi_0^2(x) = \ln(x^2 + 1)$. Fig. 4 shows the plots of φ_0^2, its absolute value of derivative, and the corresponding proximity operator for input ramp function $\boldsymbol{y}$ and $\tau = 8$. Fig. 4(b) is an exaggeration of Fig. 4(a) to better illustrate the behavior of φ_0^2 at the vicinity of origin. The true functions (dotted lines) are plotted with their approximations by our method (solid lines) for comparison. It is interesting to note how the original potential function φ_0^2 is modified adaptively at

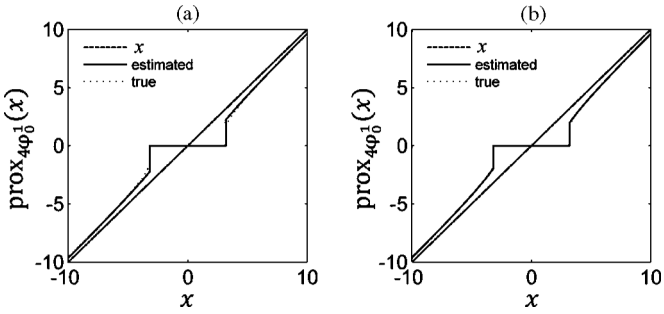


Fig. 5. The proximity operators of φ_0^1 approximated by the proposed scheme using (a) only the first term of the series (14), and (b) the first two terms of the series. In both plots, the input identity function is shown by a dashed line for comparison.

the origin to suppress small coefficients. The modified potential function was inversely calculated from the corresponding proximity operator using the method in [17].

Example 2: This example shows the effect of truncating series (14) in approximating the proximity operator. The approximate proximity operator of the nonconvex potential function $\varphi_0^1(x)$ with $\tau = 4$ was calculated by considering the truncation of the series up to its first [Fig. 5(a), $\eta_4 = 3.22$] and second order terms [Fig. 5(b), $\eta_4 = 3.19$]. As it is seen, the truncation of the series up to its first order term still gives an approximate solution with a high degree of accuracy. Higher order truncation only resulted to an improvement in approximation of the coefficients at the vicinity of jump η_4.

D. Application to the Main Problem (3)

In this section, we use the calculated proximity operator to solve unconstraint optimization problem (3) with the general $\varphi_{\mathrm{q}}^{\mathrm{p}}$ and $\boldsymbol{L} = \boldsymbol{I}$ (where $\boldsymbol{I}$ is the identity matrix).

Our algorithm is based on the proximal forward-backward splitting recursion [23], [30], [33], summarized in Algorithm 1. In the algorithm, T denotes transpose, $\text{prox}_{\tau s_n \varphi_q^p}(\boldsymbol{u})$ is applied entrywise to the vector $\boldsymbol{u}$ and appropriate choice of step-size s_{n} is necessary to guarantee its convergence. Under appropriate conditions on the operator $\boldsymbol{G}$, it has been shown that for the case of convex $\varphi_{\mathrm{q}}^{\mathrm{p}}$, the sequence $s_{\mathrm{n}} \in (0, 2/|||\boldsymbol{G}|||^2)$, where $|||\boldsymbol{G}|||$ is the spectral radius of operator $\boldsymbol{G}$, guarantees a linear convergence of $\{\boldsymbol{x}^{\mathrm{n}+1}\}$ [34], [30]. Furthermore, in the case of nonconvex $\varphi_{\mathrm{q}}^{\mathrm{p}}$, Algorithm 1 gives subsequent convergence to a stationary point that satisfies necessary conditions of a global minimizer for $s_{\mathrm{n}} \in (0, 1/|||\boldsymbol{G}|||^2)$ [23], [35]. In other words, if $\text{prox}_{\tau\varphi_{\mathrm{q}}^{\mathrm{p}}}(\cdot)$ defined by (22) provides a good approximation of the global minimizer of (6), then descent of the corresponding objective function is guaranteed by Algorithm 1.

Algorithm 1: Proximal Forward–Backward Splitting to Solve (3) for $\varphi_{\mathrm{q}}^{\mathrm{p}}$ and $\boldsymbol{L} = \boldsymbol{I}$

Set $\boldsymbol{x}^0 = \boldsymbol{G}^{\mathrm{T}}\boldsymbol{y}$, $\mathrm{n} = 0$

Choose the number of iterations, N_{it}, and a sequence s_{n}

for $\mathrm{n} = 1$ to $\mathrm{N_{it}}$ **do**

$$\boldsymbol{u}^{n+1} = \boldsymbol{x}^n + s_n \boldsymbol{G}^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{G}\boldsymbol{x}^n)$$
$$\boldsymbol{x}^{n+1} = \text{prox}_{\tau s_n \varphi_{\mathrm{q}}^{\mathrm{p}}}(\boldsymbol{u}^{n+1})$$

end for.

IV. A Generalized Shrinkage/Thresholding Rule

In this section we consider a special case of our proximity operator by setting $\eta_\tau = 0$ to develop a generalized continuous thresholding function. Following [36] function $\delta(x, \tau)$ from $\mathsf{R} \times \mathsf{R}^+ \to \mathsf{R}$ is a strict shrinkage rule if there are constants $\mathrm{C}_1, \varepsilon > 0$ such that for $\tau \in \mathsf{R}^+$ and $x \in \mathsf{R}$

$$|x| - |\delta(x,\tau)| \le \mathrm{C}_1 \min(|x|, \tau) \tag{23}$$

and

$$|\delta(x,\tau)| \le |x| \min\left(1, \left|\frac{x}{\tau}\right|^{\varepsilon}\right). \tag{24}$$

Furthermore, shrinkage $\delta(x, \tau)$ is called a thresholding rule if it satisfies

$$\delta(x,\tau) = 0 \quad \text{for} \quad |x| \le \mathrm{C}_2\tau \tag{25}$$

for another non-negative constant C_2. The thresholding condition implies that the coefficient x subjected to thresholding is replaced by zero if it is smaller, in absolute value, than a fixed threshold value. With this in mind, we design a general continuous thresholding function as

$$\delta(x,\tau) = \text{sgn}(x)\left(|x| - \frac{\tau}{\phi(\tau)}\phi(|x|)\right)_+ \tag{26}$$

where $\text{sgn}(\cdot)$ is the sign function and $\phi(x) = \frac{d}{dx}\varphi_{\mathrm{q}}^{\mathrm{p}}(x)$. Parameter τ is normalized by $\phi(\tau)$ to make the effective threshold equal to τ. Note that $\phi(x) \ge 0$ and is continuous for $x \ge 0$, and decreasing on $x \in [\tau, \infty)$ for $\mathrm{q} \ge -\min\left(1, \frac{\tau^{\mathrm{p}}+1-\mathrm{p}}{\mathrm{p}\tau^{\mathrm{p}}}\right)$. In addition, $\delta(x,\tau) = 0$ for $|x| \le \tau$, because $\tau\phi(|x|) \ge |x|\phi(\tau)$ for $|x| \le \tau$, which enables $\delta(x,\tau)$ to place itself between soft- and hard-thresholding functions. Substituting $\phi(x) = \frac{d}{dx}\varphi_{\mathrm{q}}^{\mathrm{p}}(x)$ into (26) gives

$$\delta_{\mathrm{q}}^{\mathrm{p}}(x,\tau) = \left(1 - \left|\frac{\tau}{x}\right|^{2-\mathrm{p}}\left(\frac{\tau^{\mathrm{p}}+1}{|x|^{\mathrm{p}}+1}\right)^{\mathrm{q}+1}\right)_+ x. \tag{27}$$

Proposition 2: Function $\delta_{\mathrm{q}}^{\mathrm{p}}(x,\tau)$ defined in (27) for $\mathrm{p} \in (0,1]$, $\mathrm{q} \in [-1, \infty)$, $\tau \in \mathsf{R}^+$ and $x \in \mathsf{R}$ is a thresholding rule and converges almost everywhere towards x as τ tends to zero.

Proof: $\delta_{\mathrm{q}}^{\mathrm{p}}(x,\tau)$ satisfies conditions (23) to (25) for $\mathrm{C}_1 = \mathrm{C}_2 = 1$, and $\varepsilon > 0$, so it is a thresholding rule and according to [37] converges almost everywhere towards x as τ tends to zero. ∎

Thresholding function $\delta_{\mathrm{q}}^{\mathrm{p}}(x,\tau)$ defined in (27) converges to the well-known thresholding rules for different values of p and q: it is the soft-thresholding rule for $\mathrm{p} = 1$ and $\mathrm{q} = -1$, it converges to the hard-thresholding when $\mathrm{q} \to \infty$, and for $\mathrm{p} = 0$, it is the well-known non-negative garrote thresholding rule [6], [7].

V. Numerical Results

A. Signal Denoising

We have illustrated the performance of proposed general thresholding function versus the standard SureShrink and VisuShrink (based on soft-thresholding) methods using the well-known test signals: "Blocks," "Bumps," "HeaviSine," and

TABLE III
INPUT AND OUTPUT SNR FOR VARIOUS WAVELET-BASED DENOISING SCHEMES APPLIED TO FOUR STANDARD TEST SIGNALS: "BLOCKS", "BUMPS", "HEAVISINE", AND "DOPPLER" USING DB2, DB6, DB8, AND DB8 WAVELET SYSTEMS, RESPECTIVELY

	Output SNR (db)											
Input SNR (db)	"Blocks"			"Bumps"			"HeaviSine"			"Doppler"		
	$\delta_0^1(y,\tau)$	Sure	Visu	$\delta_0^1(y,\tau)$	Sure	Visu	$\delta_0^1(y,\tau)$	Sure	Visu	$\delta_0^1(y,\tau)$	Sure	Visu
0	9.640	10.91	7.67	8.22	8.84	4.19	18.44	18.94	18.43	10.96	11.84	8.12
5	14.33	14	10.05	12.15	11.45	7.03	20.45	21.23	19.87	15.36	15.62	10.44
10	17.71	17.19	13.08	16.69	15.48	10.69	21.81	24.75	20.28	20.79	19.89	14.38
15	23.04	19.84	16.54	21.83	20.17	14.54	26.3	25.92	23.34	24.42	23.34	18.29
20	28.05	24.82	20.62	26.3	24.46	18.61	30.33	30.84	26.16	29.64	27.38	22.81
25	33.76	31.45	24.85	30.87	29.49	22.85	35.67	33.41	29.34	34.32	31.91	26.86
30	38.26	35.97	29.25	36.6	34.22	27.3	40.4	38.19	33.45	39.15	37.08	30.99
35	43.43	40.9	34.07	41.36	39.25	32.04	44.83	43.21	37.55	43.01	42.34	36.01
40	48.95	46.59	38.95	45.72	44.37	36.35	50.52	47.32	42.07	48.47	46.09	40.23
45	53.54	50.83	43.83	50.98	48.75	40.83	54.73	52.57	46.74	53.26	51.09	44.56
50	59.88	55.92	48.63	55.44	53.18	45.78	60.5	57.12	50.6	58.59	55.59	49.17

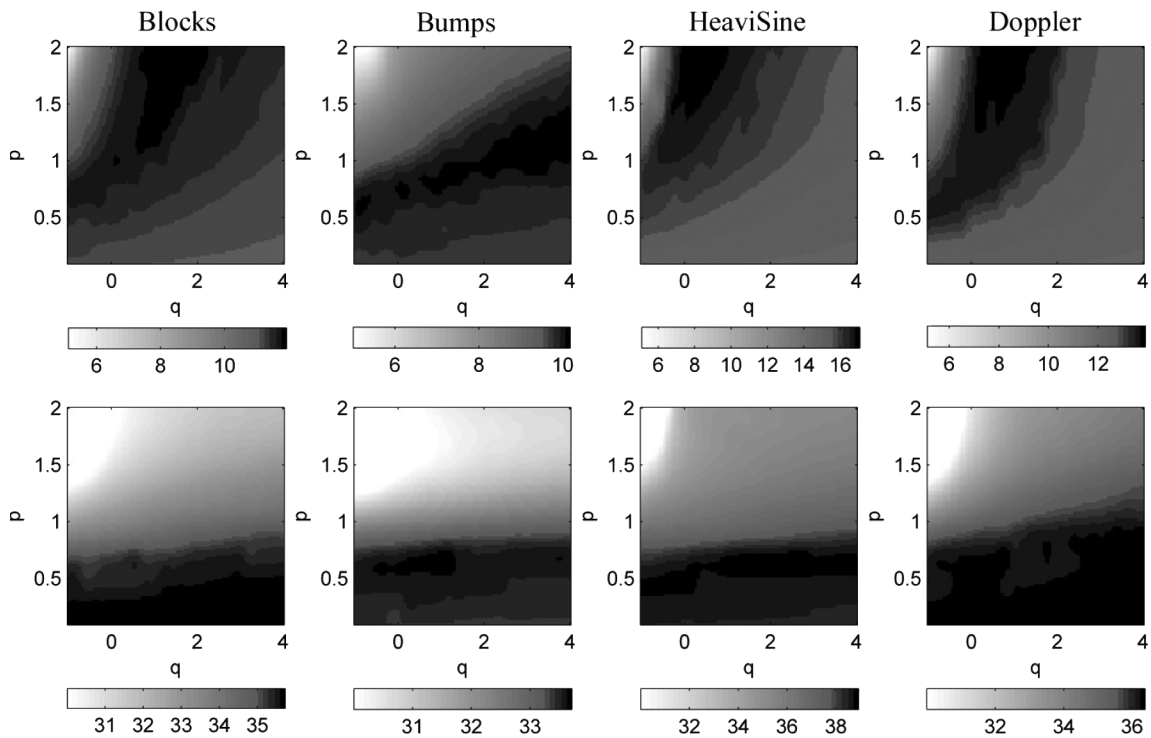


Fig. 6. The role of p and q for variational denoising based on potential function $\varphi_{\mathrm{q}}^{\mathrm{p}}$. The SNR of the input signal corresponding to top and bottom rows figures are 5 and 30 dB, respectively. In all figures, the colorbars show the SNR of the denoised signal.

"Doppler," introduced in [5]. The test signals are contaminated by additive white Gaussian noise of zero mean and standard deviation σ. Then, we tried to estimate the original signal by separating its wavelet coefficients from those of noise in a wavelets domain which corresponds to solving problem (3) for $\boldsymbol{G} = \boldsymbol{I}$ and $\boldsymbol{L}$ as orthogonal wavelet decomposition. For each method, the signal-to-noise ratio (SNR) of the output (denoised) signal was calculated as a function of SNR of the input (noisy) signal. The results for all of the tested signals are presented in Table III. We used $\delta_0^1(x,\tau)$ as the new thresholding function. Since here we wanted to show the performance of our solution approach, threshold parameter τ was taken to be optimum by minimizing the mean-square error (MSE) between denoised and original signal. The results show that, for a properly selected τ, the shrinkage function $\delta_0^1(x,\tau)$ performs better than SureShrink scheme for most of the cases and of course VisuShrink scheme for all of the cases. Here, we just tested $\delta_0^1(x,\tau)$ as the case $\delta_{\mathrm{q}}^0(x,\tau)$ has already been completely analyzed in [7]. However, function $\delta_{\mathrm{q}}^{\mathrm{p}}(x,\tau)$ takes the advantage of designing desirable thresholding function using different values of p and q for any given practical problem.

To examine the functionalities of p and q, the proposed proximity operator $\mathrm{prox}_{\tau\varphi_{\mathrm{q}}^{\mathrm{p}}}$ was applied to above test signals for different values of p and q for random noise removal. The amount of sparsity and data fidelity was balanced by choosing a regularization parameter that satisfies the discrepancy $\|\boldsymbol{y} - \mathrm{prox}_{\tau\varphi_{\mathrm{q}}^{\mathrm{p}}}(\boldsymbol{y})\|_2 \leq \sigma\sqrt{M}$. Fig. 6 shows the output SNR for two different input SNRs. The SNR of input signals

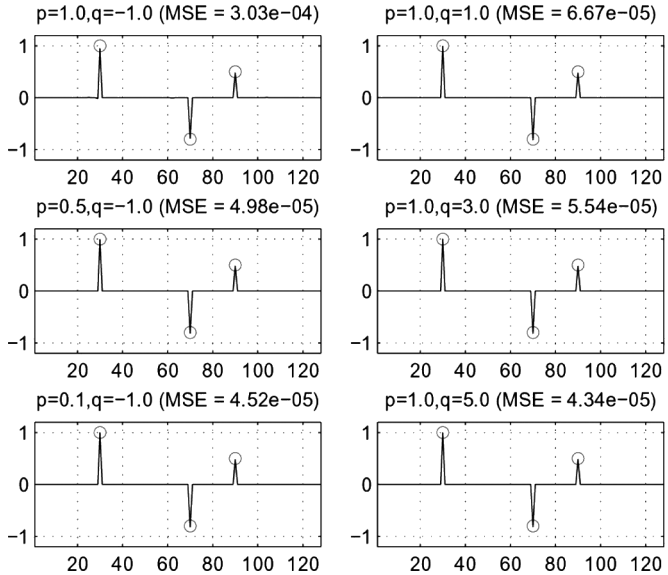


Fig. 7. Reconstructions of a sparse signal (length-128, three nonzeroes) from its 15 random projections via Algorithm 1 and the potential function $\varphi_{\mathrm{q}}^{\mathrm{p}}$, for different values of p and q. Circles show nonzero samples of the original signal.

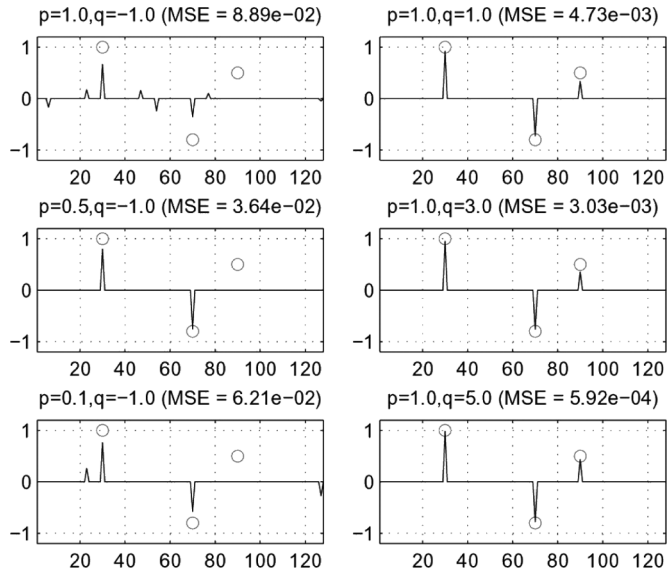


Fig. 8. Reconstructions of the same signal in Fig. 7 from its only eight random projections via Algorithm 1 and the potential function $\varphi_{\mathrm{q}}^{\mathrm{p}}$, for different values of p and q. Circles show nonzero samples of the original signal.

corresponding to the top row of the figure was 5 (dB), and the SNR of input signals corresponding to the bottom row of the same figure was 30 (dB). From these figures, it can be seen that when input SNR is low, potential function $\varphi_{\mathrm{q}}^{\mathrm{p}}$ with higher value of p performs better, and vice versa. About q, it can be stated that those values of q that correspond to nonconvex $\varphi_{\mathrm{q}}^{\mathrm{p}}$ performed better for the entire test signals with different input SNR.

B. Compressed Sensing

In this example, we have considered a typical reconstruction problem from the applications of compressed sensing (CS) theory, where the goal is to reconstruct an N-point signal which is known to be sparse from M observations, where $M < N$ (see, e.g. [11] and [12]). We generated a simple synthetic sparse signal $\boldsymbol{x}$ with dimension $N = 128$, containing three isolated nonzero samples. The observation matrix $\boldsymbol{G}$ consists of 128 random column vectors, of Gaussian components, each of length M. We generated observation vector $\boldsymbol{y}$ by adding white Gaussian noise of standard deviation 0.05 to $\boldsymbol{Gx}$. The goal was to approximate the original sparse signal using underdetermined linear system of equations $\boldsymbol{y} = \boldsymbol{Gx} + \boldsymbol{e}$, where $\boldsymbol{e}$ is the noise component. We considered $\boldsymbol{x}$ to be a solution of unconstraint optimization problem (3) with $\boldsymbol{L} = \boldsymbol{I}$ and ran the algorithm for a set of potential functions $\varphi_{\mathrm{q}}^{\mathrm{p}}$: $(\mathrm{p},\mathrm{q}) = \{(1,-1),(0.5,-1),(0.1,-1),(1,1),(1,3),(1,5)\}$. We minimized generated cost functions using Algorithm 1. The conditions for all of the cases were the same: $N_{\mathrm{it}} = 4000$, regularization parameter was selected based on the discrepancy principle $\|\boldsymbol{y} - \boldsymbol{Gx}\|_2 \leq \sigma\sqrt{M}$, and increasing step-size rule $s_{\mathrm{n}} = \frac{\mathrm{n}}{\mathrm{n}\|\boldsymbol{G}\|^2+1}$ [22]. We repeated the experiment for two sets of observations ($M = 15$ and $M = 8$). The final reconstructed signals for $M = 15$ and $M = 8$ are shown in Figs. 7 and 8, respectively. In each figure, the true nonzero samples are shown by circles. From these figures one may note that keeping $\mathrm{p} = 1$ and increasing the value of q produces better results in the sense of MSE than keeping $\mathrm{q} = -1$ and decreasing the value of p (conventional ℓ_{p}-penalties). Note also that in the case of $M = 8$, the ℓ_{p}-penalties failed to recover one of the true spikes while generating additional false spikes. In contrast, potential functions $\{\varphi_1^1, \varphi_3^1, \varphi_5^1\}$ recovered all of the spikes with a high degree of accuracy and higher value of q led to more accurate reconstruction. Fig. 9 shows the evolution of the objective function along the iterations associated to the simulations of Figs. 7 and 8. This figure confirms high accuracy of the proposed proximity operator because the objective functions are clearly decreasing functions of iteration [23]. Fig. 9 also shows better behavior of the curves corresponding to the potential functions $\{\varphi_1^1, \varphi_3^1, \varphi_5^1\}$.

To examine the functionalities of p and q, the above CS problem was simulated again for sparse signal $\boldsymbol{x}$ with dimension $N = 512$, containing ten nonzero samples with positions selected at random. Again white Gaussian noise of standard deviation 0.05 was added to observation vector $\boldsymbol{Gx}$ of length M. We solved the problem for different sets of observations: $M = 200$, 100, 75, 50 each for a rage of values of $\mathrm{p} \in (0, 2]$ and $\mathrm{q} \in [-1, 4]$. Fig. 10 shows the MSE of reconstructions. This figure can better illustrate the performance of the set of potential functions $\{\varphi_{\mathrm{q}}^{\mathrm{p}}\}$ in sparse reconstruction problems. The dashed lines in each figure show that the boundary of p and q values resulted in a good reconstruction. It can be seen that for $M = 200$, most of the potential functions performed well. However, as the number of observations decreases, the values of p and q that result to a good reconstruction are those concentrated close to the line $\mathrm{p} = 1$. These results can provide a new approach to regularization theory, that is, instead of decreasing p value in conventional ℓ_{p}-penalties, one can increase q value in the so called $\ell_{\mathrm{p},\mathrm{q}}$-penalties for a specific value of p such as $\mathrm{p} = 1$ to get more accurate and more stable solutions.

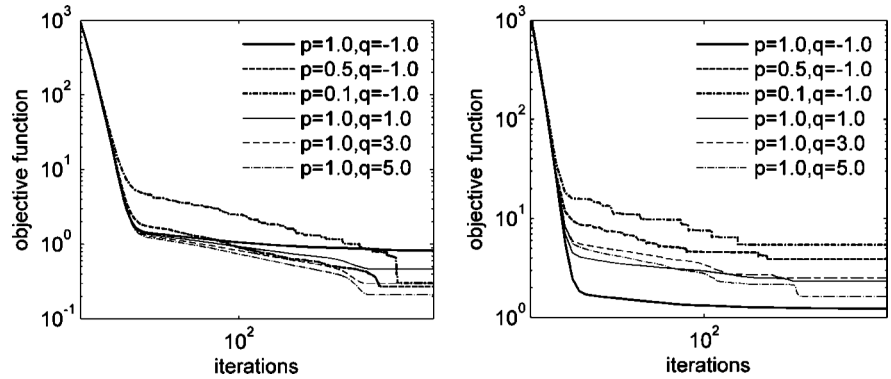


Fig. 9. Evolution of the objective function plotted against iteration number, via Algorithm 1 and the potential function $\varphi_{\mathrm{q}}^{\mathrm{p}}$ for different values of p and q solving CS problem corresponding to simulations of Fig. 7 (left) and Fig. 8 (right).

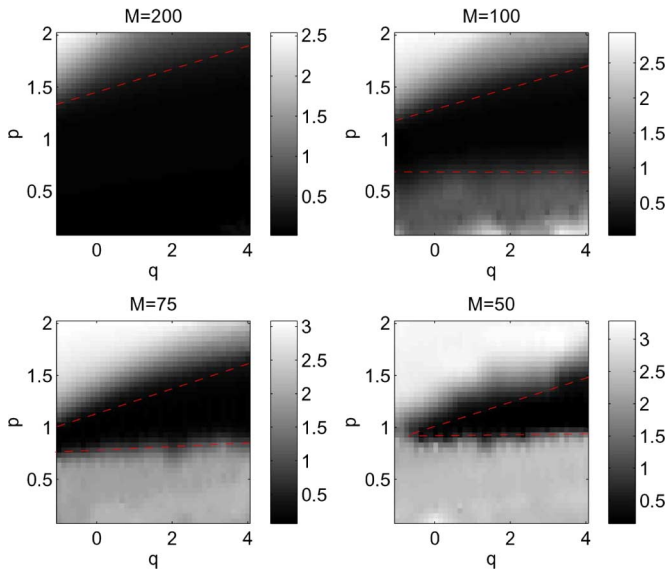


Fig. 10. The MSE as a function of p and q in potential function $\varphi_{\mathrm{q}}^{\mathrm{p}}$, for reconstruction of a sparse signal (length-512, 10 nonzeros) from its M random projections via Algorithm 1.

C. Sparse Spike Deconvolution

Deconvolution is one of the main steps in seismic data processing for recovering geological reflection coefficients (reflectivity series). This processing step can be posed as an inverse problem in which we attempt to remove the source signature and therefore increase the temporal resolution. In the simulated deconvolution, the observed seismogram $\boldsymbol{y}\,(M = 256)$ consists of unknown desired sparse reflectivity series $\boldsymbol{x}\,(N = 256)$ first blurred by 256×256 circular convolution matrix $\boldsymbol{G}$ containing a Ricker wavelet with mean frequency 20 Hz and then corrupted by zero mean additive white Gaussian noise with standard deviation 0.05. Algorithm 1 was used to recover the reflectivity series and the parameters were set as those in the CS experiment. From left to right, the first row of Fig. 11 shows the original reflectivity, noise free seismogram, and noisy seismogram, respectively. The second row of the same figure shows the recovered reflectivity series using potential functions $\{\varphi_{-1}^{1}, \varphi_{-1}^{0.5}, \varphi_{-1}^{0.1}\}$ and the third row shows the results obtained using potential functions $\{\varphi_{1}^{1}, \varphi_{3}^{1}, \varphi_{5}^{1}\}$. The evolution of the objective function along the iterations associated to the simulations of the results is also shown in Fig. 12. From Figs. 11 and 12, one may note that again under similar conditions the results obtained by using potential functions $\{\varphi_{1}^{1}, \varphi_{3}^{1}, \varphi_{5}^{1}\}$ are more accurate and the corresponding optimization problems are more stable than those of potential functions $\{\varphi_{-1}^{1}, \varphi_{-1}^{0.5}, \varphi_{-1}^{0.1}\}$.

VI. Conclusion

We have proposed a general regularization term for the solution of ill-posed inverse problems which looks at most of the known regularizers through a common view by a single formula. It is a generalization of ℓ_{p}-penalties including an additional variable q such that for a sufficiently large q, behaves like ℓ_0-penalty. We have also provided a complete analysis of finding an approximate solution to the corresponding proximity operator in a simple closed form. This provides the possibility of solving the corresponding general (convex and nonconvex) sparse regularized problem using a simple recursion algorithm. As a specific case of our proximity operator, we have also developed a general thresholding rule that continuously interpolates between the well-known soft- and hard-thresholding rules, including nonnegative garrote thresholding. Various numerical experiments confirm the effectiveness of the proposed general regularization term and show that it performs much better than conventional ℓ_{p}-penalties for sparse reconstruction problems.

Appendix
Proof of Proposition 1

Function r_τ defined in (17) has horizontal asymptote 1 or $1 + 2\tau$ for different values of p and q:

$$\lim_{x\to\infty} \mathrm{r}_\tau(x) = \begin{cases} 1+2\tau, & \mathrm{p}=2, \mathrm{q}=-1, \\ 1 & \text{otherwise} \end{cases} \tag{28}$$

and

$$\lim_{x\to 0^+} \mathrm{r}_\tau(x) = \begin{cases} -\infty, & 0<\mathrm{p}<1 \\ 1-\tau(1+\mathrm{q}), & \mathrm{p}=1 \\ \infty, & 1<\mathrm{p}<2 \\ 1+2\tau, & \mathrm{p}=2. \end{cases} \tag{29}$$

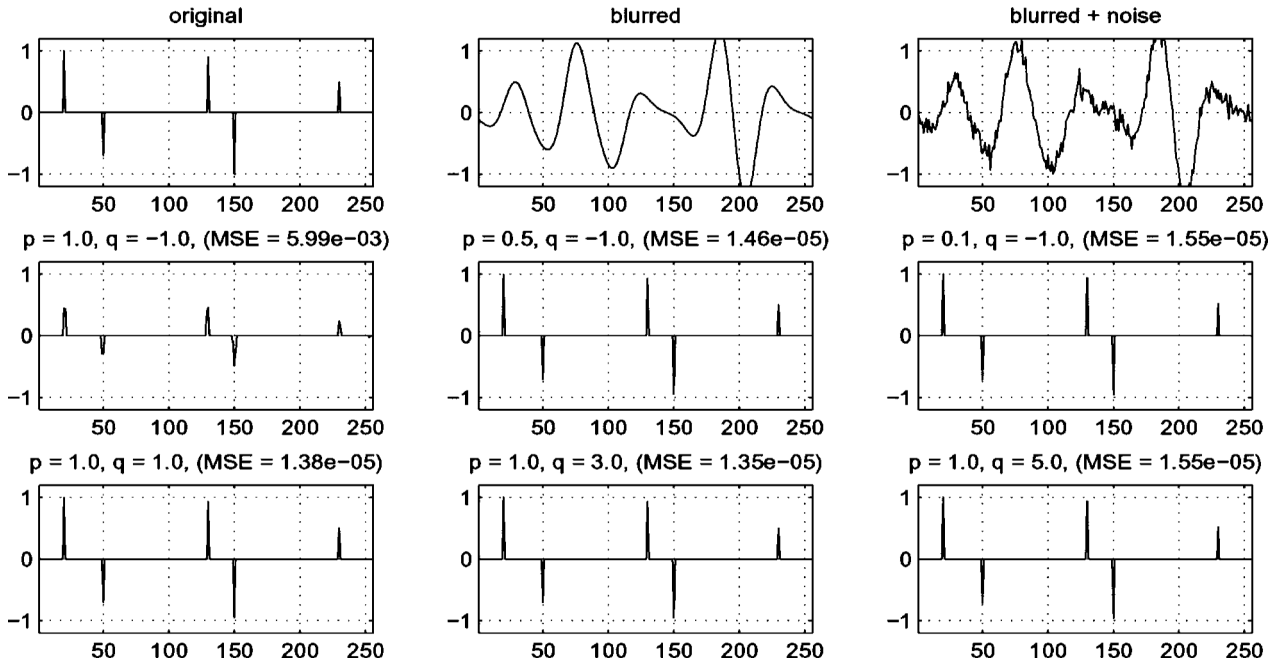


Fig. 11. Sparse spike deconvolution test. Top row shows the original, blurred, and blurred noisy signals. The figures in the second and third rows are reconstructions via Algorithm 1 and the potential function $\varphi_{\mathrm{q}}^{\mathrm{p}}$, for different values of p and q.

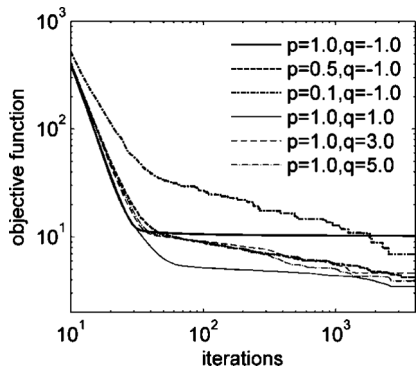


Fig. 12. Evolution of the objective function plotted against iteration number, for Algorithm 1 and the potential function $\varphi_{\mathrm{q}}^{\mathrm{p}}$, for different values of p and q, solving sparse spike deconvolution problem explained in the text.

Thus, depending on the values of p, q, and τ, r_τ may or may not have a real, positive root.

a) For $\mathrm{p} \in (0,1)$ and $\mathrm{q} \in [-1,\infty)$, r_τ is continuous and has horizontal asymptote 1 and vertical asymptote 0, where $\lim_{x\to 0^+} \mathrm{r}_\tau(x) \to -\infty$ [see (28) and (29)]. Therefore, there always exists a positive root for r_τ. We need to show this is the only root of $\mathrm{r}_\tau(x) = 0$ for $x \in \mathsf{R}^+$. If $\mathrm{r}_\tau(x)$ has more positive roots, then it must have at least two local extrema at some points $x \in \mathsf{R}^+$, where $\mathrm{r}'_\tau(x) = 0$. In other words, $\mathrm{r}'_\tau(x) = 0$ has at least two positive roots. However, $\mathrm{r}'_\tau(x) = 0$ gives the equation $\varphi'' x^2 - 2\varphi' x + 2\varphi = 0$, where $\varphi = \varphi_{\mathrm{q}}^{\mathrm{p}}(x)$. Since $\varphi \geq 0$, $\varphi' > 0$, and $\varphi'' < 0$ for $x \in \mathsf{R}^+$, this equation has two real roots with opposite signs, which is a contradiction.

b) For $\mathrm{p} \in [1,2]$ and $\mathrm{q} \in \left[-1, -\frac{1}{\mathrm{p}}\right]$, $\lim_{x\to 0^+} \mathrm{r}_\tau(x) \geq 1$ and $\lim_{x\to\infty} \mathrm{r}_\tau(x) \geq 1$. Since r_τ is continuous and differentiable on $(0,\infty)$, if there exists a root, then r_τ must have a local extremum at some point $\bar{x} \in (0,\infty)$ at which $\mathrm{r}'_\tau(\bar{x}) = 0$, that is we must have

$$\frac{d^2}{dx^2}\varphi_{\mathrm{q}}^{\mathrm{p}}(x) = 2\frac{d}{dx}\left(\frac{\varphi_{\mathrm{q}}^{\mathrm{p}}(x)}{x}\right) \tag{30}$$

at $x = \bar{x}$ which is impossible, because (18) implies that $\frac{d}{dx}\left(\frac{\varphi_{\mathrm{q}}^{\mathrm{p}}(x)}{x}\right) < 0$, while $\frac{d^2}{dx^2}\varphi_{\mathrm{q}}^{\mathrm{p}}(x) \geq 0$, as the function $\varphi_{\mathrm{q}}^{\mathrm{p}}(x)$ is convex for $x \in \mathsf{R}^+$.

c) For $\mathrm{p} \in [1,2]$ and $\mathrm{q} \in (-\frac{1}{\mathrm{p}}, \infty)$, $\lim_{x\to 0^+} \mathrm{r}_\tau(x) \geq 1$ and it has horizontal asymptote 1, so $\mathrm{r}_\tau(x)$ either has two positive roots or has no positive root at all. ■

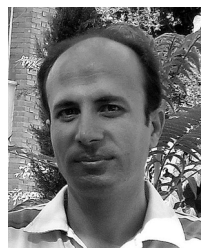
ACKNOWLEDGMENT

The authors would like to thank the editor and the three anonymous referees for their constructive comments.

REFERENCES

[1] H.-L. Taylor, S.-C. Banks, and J.-F. McCoy, "Deconvolution with the lp-norm," *Geophys.*, vol. 44, pp. 39–52, 1979.
[2] S. Levy and P.-K. Fullager, "Reconstruction of a sparse spike train from a portion of its spectrum and application to high-resolution deconvolution," *Geophys.*, vol. 46, no. 9, pp. 1235–1243, 1981.
[3] S. Geman and D.-L. McClure, "Bayesian image analysis: An application to single photon emission tomography," in *Proc. Stat. Comput. Section, Amer. Stat. Assoc.*, 1985, pp. 12–18.
[4] T. Hebert and R. Leahy, "A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors," *IEEE Trans. Med. Imag.*, vol. 8, no. 2, pp. 194–202, 1989.
[5] D.-L. Donoho and I.-M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.
[6] H. Gao, "Wavelet shrinkage denoising using the nonnegative garrote," *J. Comput. Graph. Stat.*, vol. 7, pp. 469–488, 1998.
[7] M.-A. Figueiredo and R.-D. Nowak, "Wavelet-based image estimation: An empirical Bayes approach using Jeffrey's noninformative prior," *IEEE Trans. Image Process.*, vol. 10, no. 9, pp. 1322–1331, 2001.
[8] A. Antoniadis and J. Fan, "Regularization of wavelet approximations," *J. Amer. Stat. Assoc.*, vol. 96, pp. 939–967, 2001.

[9] M. O'Brien, A. Sinclair, and S. Kramer, "Recovery of a sparse spike time series by l_1-norm deconvolution," *IEEE Trans. Signal Process.*, vol. 42, no. 12, pp. 3353–3365, 1994.
[10] A. Gholami and H.-R. Siahkoohi, "Regularization of linear and non-linear geophysical ill-posed problems with joint sparsity constraints," *Geophys. J. Int.*, vol. 180, pp. 871–882, 2010.
[11] D.-L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
[12] E. Candes, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, pp. 1207–1223, 2006.
[13] A.-N. Tikhonov and V.-Y. Arsenin, *Solutions of Ill-Posed Problems*. Washington: Winston, 1977.
[14] P.-C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*. Philadelphia, PA: SIAM, 1999.
[15] G. Archer and D. Titterington, "On Bayesian/regularization methods for image restoration," *IEEE Trans. Image Process.*, vol. 4, no. 3, pp. 989–995, 1995.
[16] J.-M. Bioucas-Dias and M. Figueiredo, "A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration," *IEEE Trans. Image Process.*, vol. 16, pp. 2980–2991, 2007.
[17] J.-M. Bioucas-Dias, "Bayesian wavelet-based image deconvolution: A GEM algorithm exploiting a class of heavy-tailed priors," *IEEE Trans. Image Process.*, vol. 15, no. 4, pp. 937–951, 2006.
[18] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse Problems*, vol. 23, no. 3, pp. 947–968, 2007.
[19] T. Goldstein and S. Osher, "The Split Bregman method for L1 regularized problems," *SIAM J. Imag. Sci.*, vol. 2, no. 2, pp. 223–243, 2009.
[20] P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE Trans. Image Process.*, vol. 6, pp. 298–311, 1997.
[21] D.-L. Donoho, "For most large underdetermined systems of linear equations, the minimal l_1 solution is also the sparsest solution," *Commun. Pure Applied Mathematics*, vol. 59, pp. 907–934, 2006.
[22] D. A. Lorenz, "Non-convex variational denoising of images: Interpolation between hard and soft wavelet shrinkage," *Current Development in Theory and Application of Wavelets*, vol. 1, no. 1, pp. 31–56, 2007.
[23] K. Bredies and D. Lorenz, Minimization of Non-Smooth, Non-Convex Functionals by Iterative Thresholding 2009, Preprint.
[24] I. Daubechies, M. Defriese, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. LVII, pp. 1413–1457, 2004.
[25] S.-J. Wright, R.-D. Nowak, and M.-A. Figueiredo, "Sparse reconstruction by separable approximation," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2479–2493, Jul. 2009.
[26] L. Montefusco, D. Lazzaro, and S. Papi, "Nonlinear filtering for sparse signal recovery from incomplete measurements," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2494–2402, Jul 2009.
[27] T. Blumensath and M.-E. Davis, "Iterative hard thresholding for compressed sensing," *App. Comput. Harmon. Anal.*, vol. 27, no. 3, pp. 265–274, 2009.
[28] P.-L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Model. Simul.*, vol. 4, no. 4, pp. 1168–1200, 2005.
[29] C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs, "A variational formulation for frame-based inverse problems," *Inverse Problems*, vol. 23, no. 4, pp. 1495–1518, 2007.
[30] J.-L. Starck, F. Murtagh, and J. M. Fadili, *Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity*. New York: Cambridge Univ. Press, 2010.
[31] P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using generalized-Gaussian and complexity priors," *IEEE Trans. Inf. Theory*, vol. 45, no. 3, pp. 909–919, 1999.
[32] S. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, no. 3, pp. 367–383, 1992.
[33] K. Bredies, "A forward-backward splitting algorithm for the minimization of non-smooth convex functionals in banach space," *Inverse Problems*, vol. 25, p. 015005, 2009.
[34] K. Bredies and D.-A. Lorenz, "Linear convergence of iterative soft-thresholding," *J. Fourier Anal. Appl.*, vol. 14, no. 5–6, pp. 813–837, 2008.
[35] H. Attouch, J. Bolte, and B.-F. Svaiter, Convergence of Descent Methods for Semi-Algebraic and Tame Problems: Proximal Algorithms, Forward-Backward Splitting, and Regularized Gauss–Seidel Methods [Online]. Available: http://www.optimization-online.org/DB_HTML/2010/12/2864.html, preprint
[36] L. Borup and M. Nielsen, Some Remarks on Shrinkage Operators preprint, 2003 [Online]. Available: http://people.math.aau.dk/~mnielsen/reprints/R-2003-07.ps
[37] T. Tao and B. Vidakovic, "Almost everywhere convergence of general wavelet shrinkage estimators," *Appl. Comput. Harmon. Anal.*, vol. 9, pp. 72–82, 2000.



Ali Gholami was born in Hamedan, Iran, in 1978. He received the Ph.D. degree in geophysics (exploration seismology) from the University of Tehran, Tehran, Iran, in 2010.

He is currently an Assistant Professor of Geophysics in the Institute of Geophysics, University of Tehran, Tehran, Iran. His research interests include inverse problems, seismic signal processing, and seismic imaging.

S. Mohammad Hosseini was born in Kashan, Iran, in 1956. He received the Ph.D. degree in mathematics from the Imperial College of Science, Technology, and Medicine, London, U.K., in 1988.

He is currently a Professor of Applied Mathematics in Tarbiat Modares University, Tehran, Iran. One of his current research interests is numerical optimization.