$L_{1/2}$ Regularization: A Thresholding Representation Theory and a Fast Solver

Zongben Xu, Xiangyu Chang, *Student Member, IEEE*, Fengmin Xu, and Hai Zhang

***Abstract*— The special importance of $L_{1/2}$ regularization has been recognized in recent studies on sparse modeling (particularly on compressed sensing). The $L_{1/2}$ regularization, however, leads to a nonconvex, nonsmooth, and non-Lipschitz optimization problem that is difficult to solve fast and efficiently. In this paper, through developing a threshoding representation theory for $L_{1/2}$ regularization, we propose an iterative $half$ thresholding algorithm for fast solution of $L_{1/2}$ regularization, corresponding to the well-known iterative $soft$ thresholding algorithm for L_1 regularization, and the iterative $hard$ thresholding algorithm for L_0 regularization. We prove the existence of the resolvent of gradient of $\|x\|_{1/2}^{1/2}$, calculate its analytic expression, and establish an alternative feature theorem on solutions of $L_{1/2}$ regularization, based on which a thresholding representation of solutions of $L_{1/2}$ regularization is derived and an optimal regularization parameter setting rule is formulated. The developed theory provides a successful practice of extension of the well-known Moreau's proximity forward-backward splitting theory to the $L_{1/2}$ regularization case. We verify the convergence of the iterative $half$ thresholding algorithm and provide a series of experiments to assess performance of the algorithm. The experiments show that the $half$ algorithm is effective, efficient, and can be accepted as a fast solver for $L_{1/2}$ regularization. With the new algorithm, we conduct a phase diagram study to further demonstrate the superiority of $L_{1/2}$ regularization over L_1 regularization.**

***Index Terms*— Compressive sensing, half, hard, L_q regularization, soft, sparsity, thresholding algorithms, thresholding representation theory.**

I. Introduction

THE sparsity problems have attracted a great deal of attention in recent years, which aim to find sparse solution(s) of a representation or an equation. Typically, the sparsity problems include those of variable selection [1], visual coding [2], error correction [3], matrix completion [4], and compressed sensing [5]– [8]. All these problems can be described as the following: Given a $M \times N$ matrix A, and a procedure of generating an observation y such as

$$y = Ax + \epsilon$$

where ϵ is the observation noise, we are asked to recover x from observation y such that x is of the sparsest structure (that is, x has the fewest nonzero components). The problem can be modeled as

$$\min_{x\in\mathbb{R}^N} \|x\|_0 \quad \text{subject to} \quad y = Ax + \epsilon$$

where $\|x\|_0$, formally called L_0 norm, is the number of nonzero components of x. Sparsity problems can be frequently transformed into the following so-called L_0 regularization problem:

$$\min_{x\in\mathbb{R}^N} \left\{\|y - Ax\|^2 + \lambda\|x\|_0\right\} \tag{1}$$

where (and henceforth) $\|\cdot\|$ denotes the Euclidean norm, $x = (x_1, \ldots, x_N)^T \in \mathbb{R}^N$, and $\lambda > 0$ is a regularization parameter.

The L_0 regularization can be understood as a penalized least squares with penalty $\|x\|_0$, in which the parameter λ functions as balancing the two objective terms. The complexity of the model is proportional with the number of variables, and solving the model generally is intractable, particularly when N is large (It is NP-hard, see [9]). In order to overcome such difficulty, many researchers [5], [6], [10], [11] have suggested to relax the L_0 regularization and, instead, to consider the following L_1 regularization:

$$\min_{x\in\mathbb{R}^N} \left\{\|y - Ax\|^2 + \lambda\|x\|_1\right\} \tag{2}$$

where $\|x\|_1$ is the L_1 norm of $\mathbb{R}^N$.

It is well known that the L_1 regularization has a very close relationship with the model Lasso and Basis Pursuit, two independent works of Tibshirani [1], and that of Chen, Donoho, and Saunders [12]. The L_1 regularization problem can be transformed into an equivalent convex quadratic optimization problem, and therefore, can be very efficiently solved. It can also result in sparse solution of the considered problem, with a promise that, under some mild conditions, the resultant solution coincides with one of the solutions of L_0 regularization (L_1/L_0 equivalence) [10], [11]. Because of this, the L_1 regularization gets its popularity and has been accepted as a very useful tool for the solution of sparsity problems. Nevertheless, the L_1 regularization may yield inconsistent selections [13] when applied to variable selection in some

Manuscript received December 8, 2011; revised March 8, 2012; accepted April 20, 2012. Date of publication May 25, 2012; date of current version June 8, 2012. This work was supported in part by the National Natural Science Foundation of China under Project 11171272, Project 60975036, Project 61075054, Project 11101325, and Project 11131006.

Z. Xu is with the Institute for Information and System Science and the MOE Key Laboratory for Intelligent Networks and Network Security, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: zbxu@mail.xjtu.edu.cn).

X. Chang and F. Xu are with the Institute for Information and System Science and Department of Applied Mathematics, School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: xiangyuchang@gmail.com; fengminxu@mail.xjtu.edu.cn).

H. Zhang is with the Department of Mathematics, Northwest University, Xi'an 710069, China (e-mail: zhanghai@nwu.edu.cn).

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/TNNLS.2012.2197412

situations. It often introduces extra bias in estimation [14], and cannot recover a signal with the least measurements when applied to compressed sensing [15], [7]. Thus, a further modification is required. Among such efforts, a very natural improvement is the suggestion of the use of L_q regularization [15], [16], [7], [17], [18]

$$\min_{x\in\mathbb{R}^N} \left\{\|y - Ax\|^2 + \lambda\|x\|_q^q\right\} \tag{3}$$

where $0 < q < 1$ and $\|x\|_q$ is the L_q quasi-norm of $\mathbb{R}^N$, defined by $\|x\|_q = (\sum_{i=1}^N |x_i|^q)^{1/q}$.

The L_q regularization is a nonconvex, nonsmooth, and non-Lipschitz optimization problem. It is difficult in general to have a thorough theoretical understanding and efficient algorithms for solutions. Moreover, even when solvable, which q should be selected to yield the best result is also a problem. Recent studies in [16], [19], [20], [7], and [21] have resolved partially these problems. In [16], Krishnan and Fergus demonstrated the very high efficiency of $L_{1/2}$ and $L_{2/3}$ regularization when applied to image deconvolution. In [20], we conducted a phase diagram study, and showed the representativeness of $L_{1/2}$ regularization among all L_q regularizations with q in (0, 1). The results basically revealed that the L_q regularizations can assuredly generate more sparse solutions than L_1 regularization, and, while so, the index 1/2 somehow plays a representative role: whenever $q \in [1/2, 1)$, the smaller the q, the sparser the solutions yielded by L_q regularizations, and, whenever $q \in (0, 1/2]$, the performance of L_q regularizations has no significant difference. From these studies, thus, the special importance of $L_{1/2}$ regularization

$$\min_{x\in\mathbb{R}^N} \left\{\|y - Ax\|^2 + \lambda\|x\|_{1/2}^{1/2}\right\} \tag{4}$$

is highlighted.

We continue such a study in this paper. Our aim is to expose a brand new feature of $L_{1/2}$ regularization: Its solutions can be analytically expressed in a thresholding form, distinguishing it from other $L_q (q \neq 2/3)$ regularizations, which permits then a fast algorithm for solutions, matching the iterative *hard* thresholding algorithm (the $hard$ algorithm in brief) [22]–[25] for L_0 regularization and the iterative *soft* thresholding algorithm (the $soft$ algorithm in brief) [26], [25], [27] for L_1 regularization.

There have been two approaches for the solution of L_0 regularization. One is the well-known greedy strategies, and the other is the iterative *hard* thresholding algorithm. The greedy strategies, such as the matching pursuit type of algorithms [28], OMP [29], ROMP [30], are very fast for non-high-dimensional problems [31], [32]. The iterative *hard* thresholding algorithm is the approach to approximately solve the L_0 regularization problem (1), which is efficient and applicable to high-dimensional problems.

The increasing popularity of L_1 regularization comes mainly from the fact that the problem can be solved very fast. There exist many exclusive and efficient algorithms, for instance, the piecewise linear method [33], LARs [34], the interior-point methods [5], [12], and the gradient boosting methods [35]. Another type algorithm called the iterative *soft* thresholding algorithm was suggested. It is a simple iteration procedure, with each iteration consisting of a Landweber update [36] followed by a thresholding operation. The $solf$ algorithm is convergent and of very low computational complexity. Consequently, it is adequate and fast even for very large-scale sparsity problems.

Inspecting the algorithms effective for L_0 and L_1 regularizations, we are particularly interested in finding of an iterative thresholding algorithm for $L_{1/2}$ regularization. This is promoted not only by the fact that the iterative thresholding-type algorithms are adequate and efficient for high-dimensional problems (this is crucial for compressed sensing application), but also by the advantage that it is relatively easy to specify the regularization parameter in the implementation of the algorithms. Apparently, as long as such a fast iterative thresholding algorithm is well developed, the $L_{1/2}$ regularization could be applied as powerfully as, or even more powerfully than, L_1 regularization, which then, hopefully, would be an essential step toward the better solution of sparsity problems.

The main contribution of this paper is to make the above expectation real. More precisely, by justifying the existence of the resolvent of gradient of penalty $\|x\|_{1/2}^{1/2}$, and uncovering a novel alterative feature of solutions of $L_{1/2}$ regularization, we derive a thresholding representation of solutions of $L_{1/2}$ regularization. With the representation, an iterative $half$ thresholding algorithm for fast solution of $L_{1/2}$ regularization is suggested, corresponding to the $soft$ for L_1 regularization, and the $hard$ for L_0 regularization. We prove the diagonal nonlinearity and analytical expressiveness of the thresholding operator by the Cartan formula. We derive an optimal regularization parameter setting rule based on the established alterative feature theorem, and verify the convergence of the $half$ algorithm when applied to k-sparsity problems. We provide also a series of experiments and applications to assess performance of the $half$ algorithm. The experiments and applications consistently show that the proposed $half$ algorithm is fast, effective, and very efficient for solving $L_{1/2}$ regularization. With the new algorithm, we finally conduct a phase diagram study to further demonstrate the superiority of $L_{1/2}$ regularization over L_1 regularization.

The reminder of this paper is organized as follows. In Section II, we develop the thresholding representation theory of $L_{1/2}$ regularization by showing the existence of the resolvent of gradient of penalty and the alternative feature theorem. In Section III, we derive the $half$ thresholding algorithm with the suggestion of an optimal regularization parameter setting strategy. In Section IV, we show the convergence of the $half$ algorithm when applied to k-sparsity problems. In Section V, we present the experiments with a series of sparse signal recovery applications, to demonstrate the robustness and effectiveness of the new algorithm. In Section VI, we conduct a phase diagram study with the help of the $half$ algorithm. We conclude this paper in Section VII with some useful remarks.

II. Thresholding Representation Theory

In this section, we establish a thresholding representation theory of $L_{1/2}$ regularization, which underlies the algorithm to be proposed.

A. Notion and Notation

For any q in [0, 1], we call an L_q regularization problem an L_q problem, and any solution of L_q problem an L_q solution (including local minimizers). A real function h is said to be a thresholding function if there is a value $t^* > 0$, called the threshold value, and a real function f_d, called the defining function, such that

$$h(t) = \begin{cases} f_d(t), & |t| > t^* \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

Clearly a thresholding function h is characterized by its threshold value t^* and defining function f_d. A mapping $H(x) = (h_1(x), h_2(x), \ldots, h_N(x))^T$ is said to be diagonally nonlinear if, for every i, $h_i(x)$ depends only on x_i and h_i is nonlinear, so that it can be simply represented as $H(x) = (h_1(x_1), h_2(x_2), \ldots, h_N(x_N))^T$. A diagonal nonlinear mapping H is said to be deduced from h if $h_1 = h_2 = \cdots = h_N = h$. When mapping H is deduced from a thresholding function h, we say that the mapping is a *thresholding operator*.

Definition 1: An L_q problem is said to permit a thresholding representation if there is a thresholding function h such that any of its L_q solutions, x, can be represented as

$$x = H(Bx) \tag{6}$$

where H is a thresholding operator deduced from h and B is an affine operator from $\mathbb{R}^N$ to $\mathbb{R}^N$.

When an L_q problem permits a thresholding representation, every L_q solution can be represented as a common fixed point of operators H and B. Thus, whenever L_q problem permits a thresholding representation, an iteration $x_{n+1} = H(B(x_n))$ can be naturally defined, which is called an *iterative thresholding algorithm* for the L_q problem.

Let $supp(x)$ be the support set of x, *i.e.*, $supp(x) = \{i : x_i \neq 0\}$. A vector x is said to be k-sparse whenever $supp(x)$ contains k elements. Without loss of generality, we assume $supp(x) = \{1, 2, ..., k\}$ whenever x is k-sparse. Let $[B(x)]_i$ be the ith row of $B(x)$. Then, by (5) and (6), we have

$$\begin{aligned} x_i &= f_d([B(x)]_i), \ \forall i \in \mathrm{supp}(x) \\ \left|[B(x)]_j\right| &\leq t^*, \quad \forall j \notin \mathrm{supp}(x). \end{aligned} \tag{7}$$

Furthermore, if we denote $z = (x_1, x_2, ..., x_k)^\mathsf{T}$, and let $F : \mathbb{R}^k \to \mathbb{R}^k$ be the diagonal nonlinear operator deduced from f_d, and $B_{k\times k}$ be the $k \times k$ principle submatrix of B, then (7) becomes $z = F(B_{k\times k}(z))$, where z is the vector whose components are all nonzero.

This suggests that, in order to justify the existence of thresholding representation of an L_q problem, we can try and look for the thresholding function in the following ways: The threshold value t^* is determined based on some kind of alternative features of components of L_q solution [such as (7)], and the defining function f_d is constructed through finding a common fixed point representation, when restricted to a specific region $\mathbb{R}_0^N$, where $\mathbb{R}_0^N = \{z = (z_1, z_2, ..., z_N)^\mathsf{T} \in \mathbb{R}^N : all\ z_i \neq 0\}$.

We will verify the existence of the thresholding representation of $L_{1/2}$ problem below, according to the methodology suggested above.

B. Resolvent Operator: Existence and Its Analytical Expression

We first show the existence of the defining function and formulate its analytic expression. We suppose that $x \in \mathbb{R}_0^N$ is a solution of $L_{1/2}$ (4). Then, the first-order optimality condition of x implies

$$0 = A^T(Ax - y) + \frac{\lambda}{2}\nabla\left(\|x\|_{1/2}^{1/2}\right) \tag{8}$$

where $\nabla(\|\cdot\|_{1/2}^{1/2})$ is the gradient of penalty $\|x\|_{1/2}^{1/2}$. Multiplying by any positive parameter μ both sides of (8) then gives $x + \mu A^T(y - Ax) = x + (\lambda\mu/2)\nabla(\|x\|_{1/2}^{1/2})$. Whenever the resolvent of $\nabla(\|\cdot\|_{1/2}^{1/2})$ exists, i.e., the operator

$$R_{\lambda,1/2}(\cdot) = (I + \frac{\lambda}{2}\nabla(\|\cdot\|_{1/2}^{1/2}))^{-1} \tag{9}$$

is well defined for any positive real λ, this then implies

$$\begin{aligned} x &= \left(I + \frac{\lambda\mu}{2}\nabla(\|\cdot\|_{1/2}^{1/2})\right)^{-1}(x + \mu A^T(y - Ax)) \\ &= R_{\lambda\mu,1/2}(x + \mu A^T(y - Ax)). \end{aligned}$$

Define $B_\mu(x) = x + \mu A^T(y - Ax))$. Then we have $x = R_{\lambda\mu,1/2}(B_\mu(x))$ which deduces to a common fixed-point representation of $L_{1/2}$ solution. We proceed to show that the resolvent operators $R_{\lambda\mu,1/2}$ can yield the defining function.

For any fixed $x = (x_1, x_2, \ldots, x_N) \in \mathbb{R}_0^N$, let

$$\mathbb{R}_{1/2}^N(x_i) = \arg\min_{y_i \neq 0}\{(y_i - x_i)^2 + \lambda|y_i|^{1/2}\},$$

$$\mathbb{R}_{1/2}^N(x) = \prod_{i=1}^N \mathbb{R}_{1/2}^N(x_i), \text{ and } \mathbb{R}_{1/2}^N = \bigcup_{x\in\mathbb{R}^N} \mathbb{R}_{1/2}^N(x).$$

We also let $\mathbb{D}_{1/2}^N = \{x \in \mathbb{R}^N : |x_i| > (3/4)\lambda^{(2/3)}\}$. Then we prove the following basic result on $R_{\lambda,1/2}(\cdot)$.

Theorem 1: As a mapping from $\mathbb{D}_{1/2}^N$ to $\mathbb{R}_{1/2}^N$, the resolvent operator $R_{\lambda,1/2}(\cdot)$ is well defined. It is a diagonally nonlinear analytically expressive operator, and can be specified by

$$R_{\lambda,1/2}(x) = \left((f_{\lambda,1/2}(x_1), f_{\lambda,1/2}(x_2), \ldots, f_{\lambda,1/2}(x_N))\right)^T \tag{10}$$

where

$$f_{\lambda,1/2}(x_i) = \frac{2}{3}x_i\left(1 + \cos\left(\frac{2\pi}{3} - \frac{2}{3}\varphi_\lambda(x_i)\right)\right) \tag{11}$$

with

$$\varphi_\lambda(x_i) = \arccos\left(\frac{\lambda}{8}\left(\frac{|x_i|}{3}\right)^{-\frac{3}{2}}\right). \tag{12}$$

Proof: Observe that $\|y\|_{1/2}^{1/2}$ is continuously differentiable at any y in $\mathbb{R}_0^N$ and its gradient is given by

$$\nabla(\|y\|_{1/2}^{1/2}) = \left(\frac{sign(y_1)}{2\sqrt{|y_1|}}, \frac{sign(y_2)}{2\sqrt{|y_2|}}, \ldots, \frac{sign(y_N)}{2\sqrt{|y_N|}}\right)^\mathsf{T}. \tag{13}$$

To show the existence of resolvent $R_{\lambda,1/2}$, we need to verify that, for any $x \in \mathbb{D}_{1/2}^N$, the equation, $x = y + (\lambda/2)\nabla(\|y\|_{1/2}^{1/2})$ has a unique solution y^* $(= R_{\lambda,1/2}(x))$ in $\mathbb{R}_{1/2}^N$. From (13), this requires us to show the existence of real solutions of the following algebraic equations:

$$y_i - x_i + \lambda \frac{sign(y_i)}{4\sqrt{|y_i|}} = 0, \;\; y_i \in \mathbb{R}^N_{1/2}(x_i) \tag{14}$$

for any fixed i. Note, from (14), that any solution y^* must be such that $y^* x_i > 0$. We only need to consider the case $y_i x_i > 0$ for any i.

Case 1: $x_i > (3/4)\lambda^{(2/3)}$: In this case, we can denote $\sqrt{|y_i|} = \eta$ and have $y_i = \eta^2$. Equation (14) then can be transformed into the following cubic algebraic equation:

$$\eta^3 - x_i \eta + \frac{\lambda}{4} = 0. \tag{15}$$

According to the Cartan's root-finding formula expressed in terms of hyperbolic functions (see [37]), the solutions of (15) vary with the sign of x_i and can be expressed as follows: Denote $r = \sqrt{(|x_i|/3)}$, $p = -(x_i/3)$ and $q = (\lambda/8)$.

Since $x_i > (3/4)\lambda^{(2/3)}$, with $\phi = \arccos(\frac{q}{r^3})$, the three roots of (15) are given by $\eta_1 = -2r\cos(\phi/3)$, $\eta_2 = 2r\cos(\pi/3 + \phi/3)$, and $\eta_3 = 2r\cos(\pi/3 - \phi/3)$. Since $\sqrt{|y_i|} = \eta > 0$, it can be tested that η_2 and η_3 in this case are solutions of (15). However, we can further check that $\eta_3 > \eta_2$ and η_3 is the unique solution of (14) satisfying $y \in \mathbb{R}^N_{1/2}(x) \subset \mathbb{R}^N_{1/2}$. Therefore, in this case, (14) has a unique solution $y_i^* \in \mathbb{R}^N_{1/2}$, which is given by $y_i^* = (2/3)x_i\left(1 + \cos\left(2\pi/3 - 2\varphi_\lambda(x_i)/3\right)\right)$, where $\varphi_\lambda(x_i)$ is defined as in (12).

Case 2. $x_i < -(3/4)\lambda^{(2/3)}$: In this case, let $\sqrt{|y_i|} = \eta$ and we have $y_i = -\eta^2$ ($\eta > 0$). Equation (14) is then transformed into the cubic equation $\eta^3 + x_i\eta + (\lambda/4) = 0$. By a similar analysis as in Case 1, we can justify that (14) has a unique solution $y_i^* = -(\eta_2)^2 = -(2/3)|x_i|\left(1 + \cos\left(2\pi/3 - 2\varphi_\lambda(x_i)/3\right)\right) \in \mathbb{R}^N_{1/2}(x_i)$.

To summarize, we have shown that

$$y_i^* = \begin{cases} \frac{2}{3}|x_i|\left(1 + \cos\left(\frac{2\pi}{3} - \frac{2\varphi_\lambda(x_i)}{3}\right)\right), & x_i > \frac{3}{4}\lambda^{\frac{2}{3}} \\ -\frac{2}{3}|x_i|\left(1 + \cos\left(\frac{2\pi}{3} - \frac{2\varphi_\lambda(x_i)}{3}\right)\right), & x_i < -\frac{3}{4}\lambda^{\frac{2}{3}} \end{cases} \tag{16}$$

is the unique solution of (14) that satisfies $y^* \in \mathbb{R}^N_{1/2}$. This leads to (11) and (12) and shows that operator $R_{\lambda,1/2}(\cdot)$ is well defined from $\mathbb{D}^N_{1/2}$ to $\mathbb{R}^N_{1/2}$, and $R_{\lambda,1/2}(\cdot)$ is a diagonally nonlinear and analytically expressive operator. The proof of Theorem 1 is completed.

Theorem 1 implies that $f_{\lambda,1/2}(x_i)$ defined as in (11) and (12) can be taken as a defining function, and so the resolvent operator $R_{\lambda,1/2}(\cdot)$ as a thresholding operator. In the next subsection we will explain why the resolvent $R_{\lambda,1/2}(\cdot)$ is defined from $\mathbb{D}^N_{1/2}$ to $\mathbb{R}^N_{1/2}$ and why the so-defined $R_{\lambda,1/2}(\cdot)$ is enough.

C. An Alternative Feature Theorem

We specify the threshold value t^* by proving the following alternative theorem on the solutions of $L_{1/2}$ regularization.

For any $\lambda, \mu \in (0, \infty)$ and $z \in \mathbb{R}^N$, let

$$C_\lambda(x) = \|y - Ax\|^2 + \lambda\|x\|^{1/2}_{1/2} \tag{17}$$

$$C_\mu(x, z) = \mu[C_\lambda(x) - \|Ax - Az\|^2] + \|x - z\|^2. \tag{18}$$

We first prove the following lemma.

Lemma 1: If $x^s = (x_1^s, x_2^s, \ldots, x_N^s)^{\mathsf{T}}$ is a local minimizer of $C_\mu(x, z)$ for any fixed λ, μ, and z, then

$$x_i^s = 0 \Leftrightarrow |[B_\mu(z)]_i| \le t^*_{1/2}, i = 1, 2, \ldots, N$$

where $t^*_{1/2}$ is the unique positive solution of

$$t^2 f_{\lambda\mu,1/2}(t) - \frac{9}{16}(\mu\lambda)^2 = 0. \tag{19}$$

Proof: We justify the lemma particularly through showing

$$x_i^s = \begin{cases} f_{\lambda\mu,1/2}([B_\mu(z)]_i), & |[B(z)]_i| > t^*_{1/2} \\ 0, & \text{otherwise.} \end{cases} \tag{20}$$

For this purpose, we first notice that, by definition, $C_\mu(x, z)$ can be reexpressed as

$$\begin{aligned} C_\mu(x, z) = \; & \|x - [(I - \mu A^\top A)z + \mu A^\top y]\|^2 \\ & + \mu\lambda\|x\|^{1/2}_{1/2} + \mu\|y\|^2 + \|z\|^2 - \mu\|Az\|^2 \\ & - \|(I - \mu A^\top A)z + \mu A^\top y\|^2 \end{aligned} \tag{21}$$

which implies that minimizing $C_\mu(x, z)$ for any fixed λ, μ, and z is equivalent to

$$\min_{x \in \mathbb{R}^N} \left\{ \sum_{i=1}^N (x_i - [B_\mu(z)]_i)^2 + \mu\lambda\|x\|^{1/2}_{1/2} \right\}. \tag{22}$$

So, $x^s = (x_1^s, x_2^s, \ldots, x_N^s)^{\mathsf{T}}$ is a local minimizer of $C_\mu(x, z)$ if and only if, for any i, x_i^s solves the problem

$$\min_{x_i \in \mathbb{R}} \{x_i^2 - 2x_i[B_\mu(z)]_i + \mu\lambda|x_i|^{1/2}\} \tag{23}$$

because the summation of (22) is separable.

Define $g(x_i, z) = x_i^2 - 2x_i[B_\mu(z)]_i + \mu\lambda|x_i|^{1/2}$. Let us first consider the solution of (23) in the setting $x_i \neq 0$. In this case, the solution then satisfies the following Euler equation:

$$x_i - [B_\mu(z)]_i + \mu\lambda \frac{sign(x_i)}{4\sqrt{|x_i|}} = 0. \tag{24}$$

Comparing (24) with (14), we immediately know from the proof of Theorem 1 that the solution of (23) is unique and given by $x_i^{ss} = f_{\lambda\mu,1/2}([B_\mu(z)]_i)$, wherever $|[B_\mu(z)]_i| > \frac{3}{4}(\lambda\mu)^{2/3}$. Consequently, the solution of (23) can be found to be $x_i^s = f_{\lambda\mu,1/2}([B_\mu(z)]_i)$ when $g(x_i^{ss}, z) < g(0, z)$, and $x_i^s = 0$ when $g(x_i^{ss}, z) \ge g(0, z)$. Note that $g(x_i^{ss}, z) < g(0, z)$ amounts to

$$|[B(z)]_i| > \frac{f^2_{\lambda\mu,1/2}([B_\mu(z)]_i) + \mu\lambda|f_{\lambda\mu,1/2}([B_\mu(z)]_i)|^{1/2}}{2|f_{\lambda\mu,1/2}([B_\mu(z)]_i)|}. \tag{25}$$

To show (20), let us define

$$G(\gamma) = |\gamma| - \frac{f^2_{\lambda\mu,1/2}(\gamma) + \mu\lambda|f_{\lambda\mu,1/2}(\gamma)|^{1/2}}{2|f_{\lambda\mu,1/2}(\gamma)|}, |\gamma| > \frac{3}{4}(\lambda\mu)^{\frac{2}{3}}.$$

Since, from (24), $f_{\lambda\mu,1/2}(\gamma)$ satisfies $f^2_{\lambda\mu,1/2}(\gamma) - \gamma f_{\lambda\mu,1/2}(\gamma) + (\lambda\mu/4)|f_{\lambda\mu,1/2}(\gamma)|^{1/2} = 0$, and

$\gamma f_{\lambda\mu,1/2}(\gamma) > 0$, $G(\gamma)$ can be further simplified to

$$G(\gamma) = |\gamma| - \frac{\gamma f_{\lambda\mu,1/2}(\gamma) + \frac{3}{4}\mu\lambda|f_{\lambda\mu,1/2}(\gamma)|^{1/2}}{2|f_{\lambda\mu,1/2}(\gamma)|} \tag{26}$$

$$= |\gamma| - \frac{|\gamma||f_{\lambda\mu,1/2}(\gamma)| + \frac{3}{4}\mu\lambda|f_{\lambda\mu,1/2}(\gamma)|^{1/2}}{2|f_{\lambda\mu,1/2}(\gamma)|}$$

$$= \frac{|\gamma|}{2} - \frac{3\lambda\mu}{8|f_{\lambda\mu,1/2}(\gamma)|^{1/2}}. \tag{27}$$

We can justify that $G(\gamma)$ is strictly increasing for $\gamma > 0$ and strictly decreasing for $\gamma < 0$. In effect, for $\gamma > 0$, for instance, $G(\gamma) = (\gamma/2) - (3\lambda\mu/8|f_{\lambda\mu,1/2}(\gamma)|^{1/2})$ is increasing because γ and $f_{\lambda\mu,1/2}(\gamma)$ both are strictly increasing (see Appendix A).

Without loss of generality, we further consider $G(\gamma)$ for $\gamma > 0$ below. It is obvious that $\lim_{\gamma\to+\infty} G(\gamma) = +\infty$, and by using $f_{\lambda\mu,1/2}(\gamma) < \gamma$, one has

$$\lim_{\gamma\to\frac{3}{4}(\lambda\mu)^{2/3}} G(\gamma) = \frac{3}{8}(\lambda\mu)^{2/3} - \frac{3\lambda\mu}{8 \lim_{\gamma\to\frac{3}{4}(\lambda\mu)^{2/3}} |f_{\lambda\mu,1/2}(\gamma)|^{1/2}} < 0. \tag{28}$$

By the strict increasing property of $G(\gamma)$, we thus know that $G(\gamma)$ has a unique positive root $t^*_{1/2}$, $t^*_{1/2} > (3/4)(\lambda\mu)^{2/3}$, and it is exactly the unique solution of (19). This implies

$$G(\gamma) > 0 (\le 0) \Leftrightarrow \gamma > t^*_{1/2} (\le t^*_{1/2}).$$

We thus conclude that $x_i^s = 0$ if and only if $G([B_\mu(z)]_i) \le 0$, and, further, if and only if $|[B_\mu(z)]_i)| \le t^*_{1/2}$. With reference to (25), this leads to (20) and the proof of Lemma 1 is completed.

Lemma 2: $t^*_{1/2} = (\sqrt[3]{54}/4)(\lambda\mu)^{(2/3)}$.

Proof: By definition, $t^*_{1/2}$ uniquely solves (19), and therefore, $t^{*2}_{1/2} f_{\lambda\mu,1/2}(t^*_{1/2}) = (9/16)\lambda^2\mu^2$. Form (11) and (12), we can test that $(t/2) \le f_{\lambda\mu,1/2} < t$ for any $t \ge 0$. Therefore, we have $(t^{*3}_{1/2}/2) \le t^{*2}_{1/2} f_{\lambda\mu,1/2}(t^*_{1/2}) = (9/16)\lambda^2\mu^2 < t^{*3}_{1/2}$, which implies $(\sqrt[3]{36}/4)(\lambda\mu)^{2/3} \le t^*_{1/2} < (\sqrt[3]{72}/4)(\lambda\mu)^{2/3}$. This shows that there is a positive constant $\beta \in [(\sqrt[3]{36}/4), (\sqrt[3]{72}/4))$ such that $t^*_{1/2} = \beta(\lambda\mu)^{2/3}$. The bound of β can be further tightened. For instance, by $G(t^*_{1/2}) = 0$ we can find

$$t^*_{1/2} = \frac{f^2_{\lambda\mu,1/2}(t^*_{1/2}) + \lambda\mu|f_{\lambda\mu,1/2}(t^*_{1/2})|^{1/2}}{2f_{\lambda\mu,1/2}(t^*_{1/2})} \tag{29}$$

$$= \frac{|f_{\lambda\mu,1/2}(t^*_{1/2})|}{2} + \frac{\lambda\mu}{4|f_{\lambda\mu,1/2}(t^*_{1/2})|^{1/2}} + \frac{\lambda\mu}{4|f_{\lambda\mu,1/2}(t^*_{1/2})|^{1/2}}$$

$$\ge 3\sqrt[3]{\frac{|f_{\lambda\mu,1/2}(t^*_{1/2})|}{2}\left(\frac{\lambda\mu}{4|f_{\lambda\mu,1/2}(t^*_{1/2})|^{1/2}}\right)^2}$$

$$= \frac{\sqrt[3]{54}}{4}(\lambda\mu)^{\frac{2}{3}}. \tag{30}$$

In consequence, $\beta \in [(\sqrt[3]{36}/4), (\sqrt[3]{72}/4))$ follows.

We now bring expression $t^*_{1/2} = \beta(\lambda\mu)^{2/3}$ into (19), resulting in the equation that parameter β should satisfy is

$$\beta^3\left[1 + \cos\left(\frac{2}{3}\left(\pi - \arccos\left(\frac{3\sqrt{3}}{8\beta^{3/2}}\right)\right)\right)\right] = \frac{27}{32}. \tag{31}$$

A direct check shows that $\beta = (\sqrt[3]{54}/4)$ is exactly the unique solution of (31). This completes the proof of Lemma 2.

By applying Lemmas 1 and 2, we now prove the following alternative theorem:

Theorem 2: If $x^* = (x_1^*, x_2^*, \ldots, x_N^*)^\mathsf{T}$ is an $L_{1/2}$ solution of (4) and μ is any fixed positive real number that satisfies $0 < \mu \le \|A\|^{-2}$, then

$$\text{either } x_i^* \ne 0 \text{ or } |[B_\mu(x^*)]_i| \le \frac{\sqrt[3]{54}}{4}(\lambda\mu)^{\frac{2}{3}}. \tag{32}$$

In particular, one can express

$$x_i^* = \begin{cases} f_{\lambda\mu,1/2}([B_\mu(x_i^*)]_i), & |[B_\mu(x_i^*)]_i| > \frac{\sqrt[3]{54}}{4}(\lambda\mu)^{\frac{2}{3}} \\ 0, & \text{otherwise.} \end{cases} \tag{33}$$

Proof: It follows directly from Lemmas 1 and 2 if we notice that $x^* = \arg\min_{x\in\mathbb{R}^N} C_\mu(x, x^*)$. This latter assertion can be justified from the fact that the assumption $\mu \in (0, \|A\|^{-2})$ implies

$$C_\mu(x, x^*) = \mu\left\{\|y - Ax\|^2 + \lambda\|x\|_{1/2}^{1/2}\right\} \tag{34}$$

$$+\left\{\|x - x^*\|^2 - \mu\|Ax - Ax^*\|^2\right\} \tag{35}$$

$$\ge \mu\left\{\|y - Ax\|^2 + \lambda\|x\|_{1/2}^{1/2}\right\} \tag{36}$$

$$\ge C_\mu(x^*, x^*) \tag{37}$$

for any $x \in \mathbb{R}^N$, which shows that x^* is a local minimizer of $C_\mu(x, x^*)$ as long as x^* is an $L_{1/2}$ solution of (22).

Theorem 2 shows that $t^*_{1/2} = (\sqrt[3]{54}/4)(\lambda\mu)^{2/3}$ is the threshold value of the thresholding representation of $L_{1/2}$ regularization. (It is not the value $(3/4)(\lambda\mu)^{2/3}$, as predicted in Theorem 1.) A lower-bound estimation on the nonzero components of L_q solution is presented in [38].

D. Thresholding Representation

Combining Theorems 1 and 2, we immediately conclude that the function

$$h_{\lambda\mu,1/2}(x) = \begin{cases} f_{\lambda\mu,1/2}(x), & |x| > \frac{\sqrt[3]{54}}{4}(\lambda\mu)^{2/3} \\ 0, & \text{otherwise} \end{cases} \tag{38}$$

define a thresholding function of the $L_{1/2}$ problem, and in this case, the thresholding representation of $L_{1/2}$ regularization is exactly given by

$$x = H_{\lambda\mu,1/2}(B_\mu(x)) \tag{39}$$

where $H_{\lambda\mu,1/2}$ is the operator deduced from $h_{\lambda\mu,1/2}(\cdot)$.

We call, henceforth, function (38) the *half* thresholding function, and operator $H_{\lambda\mu,1/2}$ the *half* thresholding operator.

It is interesting to compare the *half* thresholding function with the *hard* thresholding function in L_0 regularization, and the *soft* thresholding function in L_1 regularization. These latter two functions are defined respectively by ([23]) as

$$h_{\lambda,0}(x) = \begin{cases} x, & |x| > \lambda^{1/2} \\ 0, & \text{otherwise} \end{cases} \tag{40}$$

and by ([39])

$$h_{\lambda,1}(x) = \begin{cases} x - sgn(x)\lambda/2, & |x| > \lambda/2 \\ 0, & \text{otherwise.} \end{cases} \quad (41)$$

Remark 1: A natural question raised from the research here is: Is there another index q in (0,1), except 1/2, that permits an (39)-like thresholding representation for L_q regularization? If so, similar to (14), this amounts to asking that the algebraic equation

$$y_i - x_i + \left(\frac{q\lambda}{2}\right)\frac{sign(y_i)}{|y_i|^{1-q}} = 0, \ \ y_i \neq 0$$

has algebraic roots, or equivalently, so are the equation

$$|y_i|^{2-q} - sign(y_i)x_i|y_i|^{1-q} + \left(\frac{q\lambda}{2}\right) = 0, \ \ y_i \neq 0. \quad (42)$$

Let us express $q = n/m$, with n, m being two integers such that $n < m$, and let $t = |y_i|^{1/m}$. Then, (42) can be rewritten as

$$t^{2m-n} - sign(y_i)x_i t^{m-n} + \frac{q\lambda}{2} = 0.$$

By the classical Abel theorem, we then can test that this equation has algebraic root(s) only when $q = 1, 2/3, 1/2$. Thus, among L_q regularizations with $q \in (0, 1)$, only $L_{1/2}$ and $L_{2/3}$ regularizations permit an analytically expressive (39)-like thresholding representation. In view of the fact that $L_{2/3}$ regularization performs generally not as well as $L_{1/2}$ regularization [20], the special inportance of $L_{1/2}$ regularization can be concluded.

Remark 2: The analytic solvability of $q = 1/2$ and 2/3 for the L_q regularizations with $A = I$ has been already observed by Krishnan and Fergus [16] in the context of image deconvolution. Without giving an explicitly expressive analytic representation of the solution, however, they applied the solvability via formulating the closed-form solutions based on finding the roots of the resultant cubic or quadratic equations. In comparison, the contributions of this paper are the findings of the explicitly expressive analytic representation of the $L_{1/2}$ solutions (Theorem 1) and the alternative feature theorem (Theorem 2). These new findings make it possible not only to develop a precise thresholding representation theory of $L_{1/2}$ regularization that extends the classical Moreau's proximal forward-backward splitting theory for convex optimization [40], [41] but also to establish a rigorous theoretical analysis of $L_{1/2}$ regularization. The convergence analysis conducted in Section IV, for instance, shows such a benefit.

III. Half Thresholding Algorithm

In this section, we present an iterative *half* thresholding algorithm for performing $L_{1/2}$ regularization, with some useful parameter setting strategies.

A. Optimal Regularization Parameter

We begin with formulating an optimality condition on the regularization parameters λ, which then serves as the basis of the parameter setting strategies used in the algorithm to be proposed.

It is known that the quantity of the solutions of a regularization problem depends seriously on the setting of the regularization parameter λ. The selection of proper regularization parameters is, however, a very hard problem. There is no optimal rule in general, even when there exist various useful heuristics (see AIC [42], BIC [43]). In most and general cases, an "trial and error" method, say, the cross-validation method, is still an accepted, or even unique, choice. Nevertheless, when some prior information (e.g., sparsity) is known for a problem, it is realistic to set the regularization parameter more reasonably and intelligently.

To make this clear, let us suppose that we are considering a problem formulated as a regularization form (4), the solutions of which are of k-sparsity. Thus, we are required to solve the $L_{1/2}$ regularization problem restricted to the subregion $\Gamma_k = \{x = (x_1, x_2, \ldots, x_N) : supp(x) = k\}$ of $\mathbb{R}^N$. For any μ, denote $B_\mu(x) = x + \mu A^T(b - Ax)$. Assume x^* is an $L_{1/2}$ solution of the problem, and, without loss of generality, assume $[B_\mu(x^*)]_1 \geq [B_\mu(x^*)]_2 \geq .... \geq [B_\mu(x^*)]_N$. Then, by Theorems 1 and 2 [particularly, (38) and (39)], the following inequalities hold:

$$\left|[B_\mu(x^*)]_i\right| > \frac{\sqrt[3]{54}}{4}(\lambda^*\mu)^{\frac{2}{3}} \Leftrightarrow i \in \{1, 2, \ldots, k\}$$

$$\left|[B_\mu(x^*)]_j\right| \leq \frac{\sqrt[3]{54}}{4}(\lambda^*\mu)^{\frac{2}{3}} \Leftrightarrow j \in \{k+1, \ldots, N\}$$

which implies

$$\frac{\sqrt{96}}{9\mu}\left\{\left|[B_\mu(x^*)]_{k+1}\right|\right\}^{\frac{3}{2}} \leq \lambda^* < \frac{\sqrt{96}}{9\mu}\left\{\left|[B_\mu(x^*)]_k\right|\right\}^{\frac{3}{2}}$$

namely

$$\lambda^* \in \left[\frac{\sqrt{96}}{9\mu}\left\{\left|[B_\mu(x^*)]_{k+1}\right|\right\}^{3/2}, \ \frac{\sqrt{96}}{9\mu}\left\{\left|[B_\mu(x^*)]_k\right|\right\}^{\frac{3}{2}}\right). \quad (43)$$

The above estimation (43) provides an exact location of where an optimal regularization parameter should be. Remember that $\left|[B_\mu(x^*)]_k\right|$ is the kth largest component of $[B_\mu(x^*)]$ in magnitude. We can then take

$$\lambda^* = \frac{\sqrt{96}}{9\mu}\left[(1-\alpha)\left|[B_\mu(x^*)]_{k+1}\right|^{3/2} + \alpha\left|[B_\mu(x^*)]_k\right|^{\frac{3}{2}}\right] \quad (44)$$

with any $\alpha \in [0, 1)$. Taking $\alpha = 0$, this leads to a most reliable choice of λ^* specified by

$$\lambda^* = \frac{\sqrt{96}}{9\mu}\left|[B_\mu(x^*)]_{k+1}\right|^{\frac{3}{2}}. \quad (45)$$

(Note that the larger the λ^*, the larger the threshold value t^*, and the sparser the solution resulting by the thresholding algorithm.) Note that (45) is valid for any fixed μ. We will use this expression with a fixed $\mu_0 > 0$ below.

Instead of real solution, we may use any known approximations x_n of x^* in (44) and (45), say, we can take

$$\lambda_n^* = \frac{\sqrt{96}}{9\mu_0}\left|[B_{\mu_0}(x_n)]_{k+1}\right|^{\frac{3}{2}} \quad (46)$$

in applications. When so doing, an iteration algorithm will be adaptive and free from the choice of regularization parameter. This parameter-setting strategy will be adopted latter.

B. Half Thresholding Algorithm

With the thresholding representation (38) and (39), a thresholding algorithm for $L_{1/2}$ regularization can be naturally defined as

$$x_{n+1} = H_{\lambda_n \mu_n, \frac{1}{2}} \left(x_n + \mu_n A^T (y - A x_n) \right) \tag{47}$$

where $H_{\lambda\mu,1/2}$ is the *half* thresholding operator. We call this method the iterative *half* thresholding algorithm, or briefly, the *half* algorithm. The *half* algorithm can be seen as an extended proximal forward-backward splitting method ([40]).

Incorporated with different parameter-setting strategies, (47) defines different implementation schemes of the *half* algorithm. For example, we can have the following.

1) Scheme 1: $\mu_n = \mu_0$; λ_n is chosen by cross-validation.

2) Scheme 2:

$$\mu_n \in (0, \mu_0];\ \lambda_n = \frac{\sqrt{96}}{9} \|A\|^2 \left| [B_{\mu_0}(x_n)]_{k+1} \right|^{\frac{3}{2}}.$$

3) Scheme 3:

$$\mu_n = \mu_0;\ \lambda_n = \min \left\{ \lambda_{n-1}, \frac{\sqrt{96}}{9} \|A\|^2 \left| [B_{\mu_n}(x_n)]_{k+1} \right|^{\frac{3}{2}} \right\}.$$

Here

$$\mu_0 = \frac{(1-\varepsilon)}{\|A\|^2} \text{ with any small } \varepsilon \in (0, 1). \tag{48}$$

Scheme 1, when ignoring ε, actually is the *half* algorithm corresponding to taking $\mu_n = 1$ and normalizing matrix A, and *Scheme 2* corresponds to taking $\mu_n = 1$, normalizing A and λ_n being selected according to (46). *Scheme 3* is a variant of *Scheme 2* in the sense that the parameter λ_n is set to keep monotonically decreasing, in addition to the same setting as *Scheme 2*. We will test those half algorithms in the experiments given in Section V.

As a general suggestion, we recommend that *Scheme 1* be applied when a generic problem is considered, and *Scheme 2* and *Scheme 3* applied when a k-sparsity problem is tackled. Here, by a k-sparsity problem we mean the problem can be cast as

$$\min\{\|Ax - b\|^2\} \text{ subject to } \|x\|_0 \le k. \tag{49}$$

In applications, the parameter k can be set as an upper-bound estimation of the sparsity of the problem under consideration. We will show in Section V-B, that the *half* algorithm has certain robustness with the overestimation of the sparsity.

IV. Convergence Analysis

In this section we justify the convergence of the *half* algorithm when *Scheme 1* is adopted.

Theorem 3: Let $\{x_n\}$ be the sequence generated by the *half* algorithm with *Scheme 1.* Then:

1) $\{x_n\}$ is a minimization sequence, and $C_\lambda(x_n)$ converges to $C_\lambda(x^*)$, where x^* is a limit point of minimization sequence $\{x_n\}$;
2) $\{x_n\}$ is asymptotically regular, i.e., $\lim_{n\to\infty} \|x_{n+1} - x_n\| = 0$;
3) $\{x_n\}$ converges to a stationary point of (39), whenever μ is sufficiently small.

Proof: 1) As shown in (17) and (18), for any μ, λ, we define the function $C_\lambda(x)$ and $C_{\mu,\lambda}(x, z)$. Then from the proof of Lemma 1, we easily see that

$$C_{\mu,\lambda}(H_{\lambda\mu,1/2}(B_\mu(z)), z) = \min_x C_{\lambda,\mu}(x, z) \tag{50}$$

and therefore, $C_{\mu,\lambda}(x_{n+1}, x_n) = \min_x C_{\lambda,\mu}(x, x_n)$. Since $\mu < \|A\|^{-2}$, this implies

$$\begin{aligned} C_\lambda(x_{n+1}) &= \frac{1}{\mu}[C_{\mu,\lambda}(x_{n+1}, x_n) - \|x_{n+1} - x_n\|^2] \\ &\quad + \|Ax_{n+1} - Ax_n\|^2 \\ &\le \frac{1}{\mu}[C_{\mu,\lambda}(x_n, x_n) - \|x_{n+1} - x_n\|^2] \\ &\quad + \|Ax_{n+1} - Ax_n\|^2 \\ &= C_\lambda(x_n) \\ &\quad - \mu^{-1}[\|x_{n+1} - x_n\|^2 - \mu\|Ax_{n+1} - Ax_n\|^2] \\ &< C_\lambda(x_n). \end{aligned}$$

That is, x_n is a minimization sequence of function $C_\lambda(x)$, and $C_\lambda(x_n)$ is monotonically decreasing to a fixed value C^*. Since $x_n \in \mathbb{L}_0 = \{x : C_\lambda(x) \le C_\lambda(x_0)\}$, which is bounded, $\{x_n\}$ is bounded and, therefore, there is a limit point, say x^*. By continuity of $C_\lambda(x)$ and monotonicity of $C_\lambda(x_n)$, $C^* = C_\lambda(x^*)$ follows. This verifies *(i)* of Theorem 3.

2) Let $\epsilon = 1 - \mu\|A\|^2$. Then we have $\epsilon \in (0, 1)$

$$\mu\|Ax_{n+1} - Ax_n\|^2 \le (1 - \epsilon)\|x_{n+1} - x_n\|^2. \tag{51}$$

From (50), we have also that

$$\begin{aligned} C_\lambda(x_n) - C_\lambda(x_{n+1}) &= \mu^{-1} C_{\mu,\lambda}(x_n, x_n) - C_\lambda(x_{n+1}) \\ &\ge \mu^{-1} C_{\mu,\lambda}(x_{n+1}, x_n) - C_\lambda(x_{n+1}) \\ &= \mu^{-1}\|x_{n+1} - x_n\|^2 - \|Ax_{n+1} - Ax_n\|^2. \end{aligned}$$

This then implies

$$\begin{aligned} \sum_{n=1}^{N} \|x_{n+1} - x_n\|^2 &\le \frac{1}{\epsilon} \sum_{n=1}^{N} \{\|x_{n+1} - x_n\|^2\} \\ &\quad - \frac{1}{\epsilon} \sum_{n=1}^{N} \{\mu\|Ax_{n+1} - Ax_n\|^2\} \\ &\le \frac{\mu}{\epsilon} \sum_{n=1}^{N} \{C_\lambda(x_n) - C_\lambda(x_{n+1})\} \\ &\le \frac{\mu}{\epsilon} (C_\lambda(x_1) - C_\lambda(x_{N+1})) \\ &\le \frac{\mu}{\epsilon} f_\lambda(x_1). \end{aligned}$$

Thus, the series $\sum_{n=1}^{\infty} \|x_{n+1} - x_n\|^2$ is convergent, and so, $\|x_{n+1} - x_n\| \to 0$ as $n \to \infty$. This verifies 2) of Theorem 3.

3) By 1) and 2), the sequence $\{x_n\}$ is asymptotically regular and, for its any limit point x^*, there holds $\lim_{n\to\infty} C_\lambda(x_n) = C_\lambda(x^*)$. We further show that x^* must be a stationary state of (39), and x_n itself is convergent.

We justify this through the following three steps.

Step 1: We prove that any limit point of $\{x_n\}$ is a stationary state of (39). Assume, say, x^* is a limit point of $\{x_n\}$. Then there is a subsequence x_{n_i}, such that $x_{n_i} \to x^*(i \to \infty)$. By definition (47), it then implies

$$x_{n_i+1} = H_{\lambda\mu,1/2}(B_\mu(x_{n_i})) \to H_{\lambda\mu,1/2}(B_\mu(x^*))(i \to \infty).$$

Note that $H_{\lambda\mu,1/2}$ is discontinuous, but the convergence of $x_{n_i} \to x^*$ implies the existence of the subsequence of $|[B_\mu(x_{n_i})]_p|$ for each p, say, $|[B_\mu(x_{n_i})]_p|$ itself, satisfies either $|[B_\mu(x_{n_i})]_p| > (\sqrt[3]{54}/4)(\lambda\mu)^{2/3}$ and $\lim_{i\to\infty} |[B_\mu(x_{n_i})]_p| > (\sqrt[3]{54}/4)(\lambda\mu)^{2/3}$, or $|[B_\mu(x_{n_i})]_p| \le (\sqrt[3]{54}/4)(\lambda\mu)^{2/3}$, based on which we can show the validity of the above limit (for details, refer to Appendix B). However, by Theorem 3 2), we have $\lim_{i\to\infty} x_{n_i+1} = \lim_{i\to\infty} x_{n_i} = x^*$. It then follows that $x^* = H_{\lambda\mu,1/2}(B_\mu(x^*))$. That is, x^* is a stationary state of (39), as claimed.

Step 2: We show that the stationary states of (39) are finite. Actually, let Υ be the stationary state set of (39), and let $\Upsilon_k = \{x^* \in \Upsilon : |supp(x^*)| = k\}$. Then it is obvious that $\Upsilon = \cup_{k=1}^N \Upsilon_k$. We only need to show that Υ_k contains a finite number of elements. Let $x^* = (x_1^*, x_2^*, \cdots, x_N^*) \in \Upsilon_k$. Then, for any $i \in supp(x^*)$, $x_i^* = h_{\lambda\mu,1/2}([B_\mu(x^*)]_i)$ and $[B_\mu(x^*)]_i > (\sqrt[3]{54}/4)(\lambda\mu)^{2/3}$. By Theorem 1, such x_i^* is unique. So Υ_k at most contains the vectors whose components consist of k definite values and $N-k$ zeros. All those vectors are finite. This shows the finiteness of Υ.

Step 3: We justify the convergence of $\{x_n\}$. This follows directly from the facts that $\{x_n\}$ has limit point(s), every limit point is in Υ which is a finite discrete set, and $\{x_n\}$ is asymptotically regular.

With this, the proof of Theorem 3 is completed. Theorem 3 implies immediately that the $half$ algorithm with *Scheme 1* is sure to converge if μ is sufficiently small. It has not, however, concluded the convergence of the $half$ algorithm with *Scheme 2* and *Scheme 3*. This latter cases need more technical skills. A more detailed convergence analysis of the algorithm will be presented in [44].

V. Simulations and Applications

In this section, we provide a series of simulations and applications to demonstrate the high performance of the $half$ thresholding algorithm.

The simulations and applications are conducted by applying the algorithm to a typical compressed sensing problem, i.e., signal recovery. The purpose of the simulations is to assess the effectiveness, robustness, and convergence of the algorithm. The effectiveness is measured by how few measurements (samples) are required to exactly recover a signal. The fewer the measurements used by an algorithm, the better it is. To compare performance, some other competitive algorithms such as L_1-Magic, $soft$, $hard$, and reweighed L_1-methods have been also applied, together with the $half$ algorithm.

The stimulations and applications were all conducted on a personal computer (2.67 GHz, 4 Gb RAM) with MATLAB 7.9 programming platform (R2009b). The error precision was set to 10^{-8}.

A. Signal Recovery

The sparse signal recovery problem has been studied extensively in the past few years (see [45], [6], [7]). According to Donoho [6], the problem can be cast as an L_0-problem

$$\min \|x\|_0 \ subject\ to\ y = Ax + \varepsilon \tag{52}$$

and solved by L_1 regularization. We propose to solve the problem via $L_{1/2}$ regularization. Here, $A \in \mathbb{R}^{M\times N}$ is a sensing (sampling in this case) matrix, y is an observation, ε is observation noise, and x is the signal we would like to recover.

We present two experiments to compare the performance of L_0, L_1 and $L_{1/2}$ regularizations when they are applied to the problem. In the experiments, the $hard$ algorithm ([22], [23]), the $soft$ algorithm ([26], [39]), the $half$ algorithm, together with the representative L_1-algorithm, L_1-Magic ([45]), and the newly suggested $L_{1/2}$-algorithm, reweighed L_1-algorithm ($RL1$ in brief, [7]), were simulated. The sensing matrix A was taken as the Gaussian random matrix, as suggested in [6]. We have used *Scheme 2* and *Scheme 3* (the simulation results are almost the same) of the $half$ algorithm in the simulations. In each case, the mean square error (MSE) between the recovered signal and the original signal was computed, and the CPU time [times (s)] for running the algorithm was recorded.

1) Signal Without Noise: We considered a real-valued N-length ($N = 512$) signal x without noise, shown as in Fig.1, where x is k-sparse with $k = 130$. The simulations then aim to recover $x \in \mathbb{R}^{512}$ through M measurements determined by sampling A, where M is much less than 512. The sampling was taken in the Gaussian random way in $[0, 512]$. The five regularization algorithms were applied with a variable number (M) of measurements. Some of the simulation results are listed in Table I.

From Table I, we can see that all the five algorithms can accurately recover the signal when $M = 330$, and, in this case, the $half$ algorithm attains the highest accuracy among the algorithms (This was also observed for $M > 330$). When $M = 250$, however, the first three algorithms, i.e., L_1-Magic, $soft$, and $hard$, all failed, but the $L_{1/2}$-algorithms, i.e., $half$ and $RL1$, still succeed in recovering the signal, and the $half$ algorithm still has the highest accuracy and keeps the lowest computation time. Furthermore, when the measurements are reduced to 240, it is seen that there is no other algorithm except $half$ that can accurately recover the signal. We found that $M = 240$ is a phase transition point in the sense that the $L_{1/2}$-algorithm, $half$, cannot recover the signal any more if the measurements are less than 240, as indicated for $M = 239$ in Table I. In this critical case ($M = 239$), it is seen that the $half$ algorithm still succeeds in recovery with the highest precision and lowest cost.

This experiment shows that the $half$ algorithm outperforms all the other four algorithms.

2) Signal With Noise: The signal in Fig.1 is considered again, but with noise, say, with the white noise $\varepsilon \in N(0, \sigma^2)$ ($\sigma = 0.1$). Such noisy signal is designed to simulate a real measurement or observation in which noise is inevitably involved. Our simulations aim to assess the capability of all

TABLE I

RECOVERY RESULTS OF A SIGNAL WITH DIFFERENT NUMBER OF SAMPLES

M	Method(s)	MSE	CPU time (s)	M	Method(s)	MSE	CPU time (s)
330	L_1-Magic	5.02E-05	1.91	**250**	L_1-Magic	**6.74**	11.01
	soft	2.63E-05	1.11		soft	**7.33**	0.16
	hard	4.20E-07	0.07		hard	**10.04**	0.17
	RL1	9.67E-06	34.07		RL1	1.27E-05	56.56
	half	4.27E-08	0.18		half	1.198E-07	0.47
240	L_1-Magic	**4.36**	1.05	**239**	L_1-Magic	**1.35**	1.41
	soft	**4.83**	0.51		soft	**5.96**	0.35
	hard	**12.07**	0.18		hard	**11.9**	0.24
	RL1	**3.71**	1.13		RL1	**9.87**	2.85
	half	1.98E-07	0.67		half	**9.77**	0.82

TABLE II

RECOVERY RESULTS OF A NOISY SIGNAL WITH DIFFERENT NUMBER OF SAMPLES

M	Method(s)	MSE	Ratio	CPU time (s)	M	Method(s)	MSE	Ratio	CPU time (s)
330	L_1-Magic	3.71	1.46	2.14	**274**	L_1-Magic	6.34	1.86	1.75
	RL1	3.58	1.41	47.07		RL1	6.10	1.79	60.83
	soft	3.42	1.34	0.16		soft	6.43	1.89	0.28
	hard	2.96	1.16	0.13		hard	8.52	2.50	0.22
	half	2.58	**1.02**	0.85		half	3.21	**0.94**	0.83
	Oracle	2.54				*Oracle*	3.41		
300	L_1-Magic	5.30	1.77	1.86	**239**	L_1-Magic	6.77	1.46	1.49
	RL1	4.04	1.35	54.23		RL1	6.53	1.41	54.03
	soft	5.21	1.75	0.20		soft	7.07	1.53	0.30
	hard	4.21	1.41	0.21		hard	13.89	3.01	0.19
	half	3.49	**1.17**	0.83		half	5.73	**1.24**	0.94
	Oracle	2.98				Oracle	4.61		
275	L_1-Magic	6.10	1.75	1.91	**238**	L_1-Magic	11.07	2.47	1.34
	RL1	5.05	1.45	61.98		RL1	11.11	2.48	49.70
	soft	6.27	1.80	0.30		soft	11.06	2.47	0.24
	hard	10.46	3.00	0.20		hard	16.70	3.72	0.19
	half	3.22	**0.92**	0.83		half	13.22	**2.95**	0.98
	Oracle	3.49				Oracle	4.49		

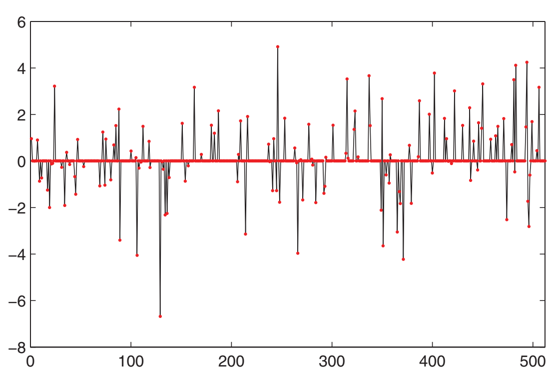

Fig. 1. Sparse signal with length $N = 512$ and sparsity $k = 130$.

the regularization algorithms in recovering the signal from a noisy circumstance and with fewer samplings. The robustness of the algorithms is thus observed in the simulations.

In order to understand the effect of noise, we have used the oracle MSE to examine the recovery capability of the algorithms in the simulations. For each algorithm, we calculated the ratio of the MSE generated from the algorithm and the oracle, listed as "Ratio" in Table II. Thus, the more close the ratio is to 1, the better the algorithm, and, correspondingly, the stronger the robustness of the algorithm.

The simulation results with algorithms L_1-Magic, $RL1$, $soft$, $hard$, and $half$ are shown in Table II, as the number of measurements (M) varies from 330 to 238. We found that $M = 274$ is a phase transition point above which the performance of $L_{1/2}$ regularization changes dramatically.

From Table II, we can see that the ratio of $half$ algorithm is always very close to 1 from $M = 330$ to 239, and, in every case before all the algorithms failed (namely, $M = 238$), it is the closest to the oracle among the five algorithms. Moreover, if we regard an algorithm to have failed in the exact recovery when its ratio is larger than 1.5, then it is seen from Table II that two L_1-algorithms, i.e., L_1-Magic and $soft$, fail when $M \leq 300$, while the $L_{1/2}$ algorithm $RL1$ fails until $M = 274$, and $soft$ fails until $M = 239$ (this is even later than the failure number of the $hard$ algorithm). Observing the MSE values in Table II, we can also find that, for all cases, $half$ always yields the most accurate recovery results. This shows that

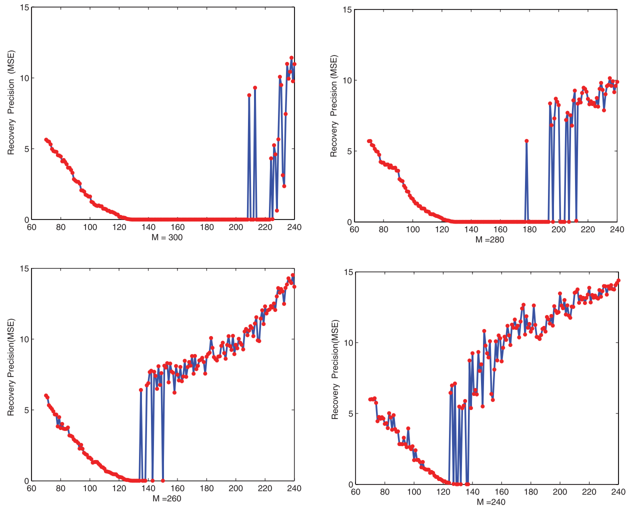

Fig. 2. Robustness of the half algorithm to sparsity overestimation when applied to the signal recovery problem.

$L_{1/2}$-regularization, the *half* algorithm in particular, provides the best signal recovery at the same noise level before the phase transition value is reached. It is again the *half* algorithm that requires the least number of samplings among the five algorithms. From this experiment, the robustness of the *half* algorithm is established.

B. Robustness on Sparsity Overestimation

In previous subsections, we have implemented the half thresholding algorithm when the sparsity of a problem is known. In real applications, the sparsity, however, is hardly known, and, instead, we may only have a rough estimation of the value. How about the performance of the *half* algorithm when the exact sparsity value is replaced by its rough estimation? We provide simulations in this subsection to show that the *half* algorithm actually shows robustness with respect to the sparsity estimation.

Let us again consider the signal recovery application made in Section V-A. Instead of using the exact sparsity value $k = 130$, we resimulated the *half* algorithm with variable estimations on sparsity k, from an underestimated value (70) to an overestimated value (240).

Fig. 2 shows the simulation results for different measurements, in which the abscissa is the sparsity estimation value and the ordinate is recovery precision (MSE) the *half* algorithm has reached. From Fig. 2, we can see that the underestimation of sparsity generally leads to an unsatisfactory performance (the closer to the true sparsity, the better the performance); nevertheless, a wide range of overestimations yield the same perfect recovery as the true sparsity value is used. This stability of the algorithm varies with the number M of the measurements. That is, the more number of measurements one uses, the wider the stable range (e.g., Fig. 2), and the stable range shrinks to zero when the measurements get too few (say, $M = 240$, as indicated in Fig. 2); the algorithm can still recover the signal when the exact sparsity value $k = 130$ is used.

These experiments reveal that the *half* thresholding algorithm has certain robustness with the overestimation of sparsity value. The robustness degree, however, is proportional to the number of measurements used. Thus, a more precise estimation of the sparsity value can generally lead to fewer measurements required for exact recovery, and conversely, a rougher estimation of the sparsity must be compensated with more measurements.

Remark 3: In the past studies, the $L_q(0 < q < 1)$ regularization problems have been solved either by approximation techniques (e.g., [15]) or by reweighting skills (e.g., [46], [44]). The former approximates the nonconvex L_q quasi-norm terms by appropriate convex functions and yields approximate solutions of the problems, while the latter finds the solution through successively transforming the problem into a reweighted convex regularization (L_1 or L_2) problem. Different from those approaches, Krishnan and Fergus [16], Chartrand [17], [18] developed a closed-form-solution-based approach. Their approach is non-convexification and thresholding based, but the thresholding operator needs to be computed at each iteration step by implementing a root-finding procedure. Different from those approaches, the *half* algorithm proposed in this paper is also a thresholding-representation-based nonconvexification method, which is as explicitly expressive a thresholding operator and can be computed faster. This feature makes the *half* algorithm applicable more conveniently and efficiently very often. For example, FOCUSS proposed by Gorodnitsky and Rao [46] is a latest efficient algorithm for sparse signal reconstruction, which is a reweighted L_2 regularization method. Because, for each L_2 regularization subproblem, it has the same closed-form analytic solution as the estimator of ridge regression, FOCUSS handles sparse signal reconstruction fast. We compared FOCUSS with the *half* algorithm when both the algorithms were applied to the signal recovery problems in 1) of this section. The comparison shows that, when signal length is not large (say, less than 1000), both the algorithms perform with nearly the same speed, but when the signal length is large (say, large than 1000), FOCUSS performs much slower than the *half* algorithm. The reason for this is that FOCUSS needs calculation of the Moore–Penrose inverse at each iteration step while the *half* algorithm only needs to calculate the half operator [the expressions (10)–(12)]. This is the reason for the high efficiency of the *half* algorithm for $L_{1/2}$ regularization.

VI. Phase Diagram Research

In this section we conduct a phase transition study of the *half* thresholding algorithm. The purpose is to further demonstrate the stronger sparsity-promoting capability of $L_{1/2}$ regularization over L_1 regularization.

Phase diagram, as a succinct tool in studying the equivalence of L_0 and L_1 regularization, was first introduced by Donoho [10], [11]. Using high-dimensional geometry, he provided a necessary and sufficient condition for any $M \times N$ matrix A in (2) such that every $k-$sparse solutions of L_1 problem (2) are k-sparse solutions of L_0 problem (1). (Such a solution is called a point of L_1/L_0 equivalence in [10], [11].) This makes the performance of L_1 regularization exhibit a two-phase (success/failure) structure in a diagram by a phase

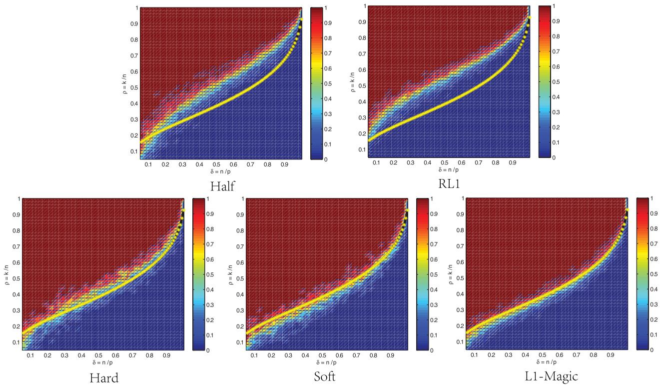

Fig. 3. Phase diagram of signal recovery using different algorithms. Horizontal axis: undersampling fraction $\delta = (\mathrm{M/N})$. Vertical axis: sparsity fraction $\rho = (s/M)$.

transition curve. Above the phase transition curve, the L_1-solution is not found to be an L_0-solution; below, the curve the L_1-solution is an L_0-solution. Using phase diagrams, Donoho and Stodden [47] conducted a series of simulation studies to assess how features of a problem (number of nonzeros in x/number of rows in A) and indeterminacy (number of rows in A/number of columns in A) determine the performance of an L_1-algorithm, or, equivalently, how the capability of an L_1-algorithm varies with the features of a problem. For this reason, the phase diagram provides a useful methodology (but in dependable way) to compare the abilities of various L_1 algorithms. We extend such a methodology to $L_{1/2}/L_0$ equivalence study, and utilize the phase diagram to compare different regularization approaches.

In this paper, we have taken the 512-length signal recovery problem considered in Section V as a prototype (thus $N = 512$) with which the variable features of the problem could be constructed. More specifically, for each fixed M (abscissa $\delta = M/N$ is fixed), we vary the sparsity level k from 1 to M (ordinate $\rho = k/M$ varies from $1/M$ to 1) by considering 100 equidistributed values $k_i = iM/100$ in the interval [0, 1], and then increase M from 0 to N (so δ from 0 to 1) in a way such that 100 discrete values $M_j = jN/100$ are considered. This constitutes a testing situation with 10000 models. For each model, a fixed k-sparse solution is computed by the various algorithms to be tested.

We have applied the $half$ thresholding algorithm with *Scheme 2* to this paper. For comparison, we have also applied $hard$, L_1-Magic, $soft$, and $RL1$ at the same time. The recovery was accepted as a "success" whenever the normalized root-mean-square error (nRMSE), $\|\hat{x} - x\|/\|x\|$, was smaller than 10^{-5}; otherwise, it was regarded as a "failure," where x is the true signal and $\hat{x}$ is the signal recovered from the fewer measurements by an algorithm. Also, we embody a pixel (corresponding to a specific model feature) blue whenever the point is the case of "success," otherwise, red when "failure." In this way, a phase diagram (color image) of an algorithm was drawn.

Fig. 3 shows the phase diagrams of the involved five regularization algorithms. In Fig. 3, the commonly appearing yellow curves are the phase transition curves of L_1 regularization, which consists of the theoretical thresholds at which L_1/L_0 equivalence breaks down.

From Fig. 3, we can see that the phase transition phenomenon does appear for all the algorithms. As can be seen, the theoretical L_1 phase transition curve has been recovered again with the L_1 algorithm L_1-Magic. The phase diagram of $soft$ almost coincides with the theoretical one, demonstrating the power of the iterative thresholding-type algorithms. It is very pleasing to observe that the phase diagrams of $half$ and $RL1$ in Fig. 3 show that the phase transition curves are all above the L_1 curve. As predicted, it shows the stronger sparsity-promoting property of $L_{1/2}$ regularization over L_1 regularization.

VII. Conclusion

We have conducted a continuation study of a specific regularization framework, i.e., $L_{1/2}$ regularization, for better solution of sparsity problems. The main contribution is the establishment of a precise thresholding representation theory, and based on the theory a fast and effective iterative thresholding algorithm for the implementation of $L_{1/2}$ regularization has been developed. The convergence analysis the proposed algorithm was also conducted.

It has been shown in the previous studies [16], [19], [20], [7], [17], [18] that $L_{1/2}$ regularization provides a potentially powerful new approach for sparsity problems, which is capable of yielding sparser solutions than L_1 regularization, and it can be taken as a representative of L_q. The $L_{1/2}$ regularization, however, leads to a nonconvex, nonsmooth, and non-Lipschitz optimization problem which is difficult solve fast and efficiently. Accordingly, finding a fast algorithm to solve $L_{1/2}$ regularization is imperative and of fundamental importance.

In this paper, by developing a thresholding representation theory of $L_{1/2}$ regularization, we have demonstrated a new advantage of $L_{1/2}$ regularization over other $L_q (q \neq 2/3)$ regularization: its solution can be analytically expressed in a thresholding fixed-point form. Based on the developed theory, an iterative $half$ thresholding algorithm was proposed for fast solution of $L_{1/2}$ regularization. The thresholding representation was constructed through a novel finding of alternative feature of solutions of $L_{1/2}$ regularization, and by the resolvent gradient of penalty $\|x\|_{1/2}^{1/2}$. We have verified the existence of the resolvent, and shown that the resolvent is a diagonally nonlinear analytically expressive operator deduced from a defining function. With the help of the alternative feature theorem, we have resolved the problem as to where the optimal regularization parameter should be when a k-sparsity problem is tackled. This defines a very powerful strategy of parameter setting. Incorporated with such strategy, the iterative $half$ thresholding algorithm is adaptive and free from choice of parameters. We have verified the convergence of the proposed algorithm, and applied the algorithm, together with other competitive regularization algorithms, to a series of problems in signal processing. The applications consistently show the following.

1) The algorithm is fast, effective, and very efficient for k-sparsity problems.

2) It is simple, is very convenient in use, and can be applied to large scale problems.
3) It is robust to observation noise and overestimation of sparsity value.
4) $L_{1/2}$ regularization shows a significantly stronger sparsity-promoting property than L_1 regularization in the sense that it allows getting more sparse solutions of a problem and recovering a sparse signal from fewer samplings, as compared with L_1 regularization.

Consequently, we conclude that the iterative *half* thresholding algorithm provides a fast and effective solver for $L_{1/2}$ regularization, particularly for large-scale problems, and also, that the *half* algorithm outperforms L_1 regularization in solution of sparsity problems. This is justified further by a phase diagram study.

It is worth remarking finally that the threshoding representation theory developed for $L_{1/2}$ regularization in this paper has provided a successful extension of the well-known Moreau's proximal forward-backward splitting theory for convex optimization ([40], [41]). The success of such an extension sheds light on the possibility to generalize more well-developed convex optimization theory to nonconvex cases.

We note that a related thresholding theory for the nonconvex L_q regularization problems has been presented in [16], [17], and [18].

Appendix A
Strict Increasing Property of Half Thresholding Function $f_{\lambda,1/2}$

In this appendix, the strict increasing property of the half threshoding function $f_{\lambda,1/2}(x)$ is verified. In particular, we prove that $f_{\lambda,1/2}(x)$ is strictly increasing for any $|x| > (\sqrt[3]{54}/4)\lambda^{2/3}$. Since $f_{\lambda,1/2}(x)$ is an odd function, we only justify this for $x > (\sqrt[3]{54}/4)\lambda^{2/3}$.

By definition, for any $x > (\sqrt[3]{54}/4)\lambda^{2/3}$, $f_{\lambda,1/2}(x)$ is defined by

$$f_{\lambda,1/2}(x) = \frac{2}{3}x\left[1+\cos\left(\frac{2}{3}\pi - \frac{2}{3}\varphi_\lambda(x)\right)\right]$$

where

$$\varphi_\lambda(x) = \arccos\left(\frac{\lambda}{8}\left(\frac{x}{3}\right)^{\frac{-3}{2}}\right) \in [0,\pi].$$

Note that $0 \le (\lambda/8)(x/3)^{-3/2} \le (\sqrt{2}/2)$, and we have $(\pi/4) \le \varphi_\lambda(x) \le (\pi/2)$, and therefore, $0 \le (\pi/3) \le (2\pi/3) - (2/3)\varphi_\lambda(x) \le (\pi/2)$, which implies $0 \le \cos(2\pi/3) - (2/3)\varphi_\lambda(x)) \le (1/2)$ and $(\sqrt{3}/2) \le \sin(2\pi/3) - (2/3)\varphi_\lambda(x)) \le 1$. Also, since

$$\begin{aligned}\sin\varphi_\lambda(x) &= \sqrt{1-\cos^2\varphi_\lambda(x)}\\ &= \sqrt{1-\frac{27\lambda^2}{64x^3}} = x^{-2/3}\sqrt{x^3-\frac{27\lambda^2}{64}}\end{aligned}$$

and $\sin\varphi_\lambda(x).\varphi_\lambda(x) = (9\sqrt{3}/16)\lambda x^{-5/2}$, we deduce

$$\acute{\varphi}(x) = \frac{\frac{9\sqrt{3}}{16}\lambda x^{-5/2}}{\sin\varphi_\lambda(x)} = \frac{\frac{9\sqrt{3}}{16}\lambda}{x\sqrt{x^3-\frac{27}{64}\lambda^2}}. \quad (53)$$

By using (53), we thus calculate

$$\begin{aligned}\acute{f}_{\lambda,1/2}(x) &= \frac{2}{3}\left[1+\cos\left(\frac{2}{3}\pi-\frac{2}{3}\varphi_\lambda(x)\right)\right]\\ &\quad+\frac{2}{3}x\left[-\sin\left(\frac{2}{3}\pi-\frac{2}{3}\varphi_\lambda(x)\right)\right]\left[-\frac{2}{3}\acute{\varphi}_\lambda(x)\right]\\ &= \frac{2}{3}\left[1+\cos\left(\frac{2}{3}\pi-\frac{2}{3}\varphi_\lambda(x)\right)\right]\\ &\quad+\frac{4}{9}x\sin\left(\frac{2}{3}\pi-\frac{2}{3}\varphi_\lambda(x)\right)\acute{\varphi}_\lambda(x)\\ &= \frac{2}{3}\left[1+\cos\left(\frac{2}{3}\pi-\frac{2}{3}\varphi_\lambda(x)\right)\right]\\ &\quad+\frac{\frac{\sqrt{3}}{4}\lambda\sin\left(\frac{2}{3}\pi-\frac{2}{3}\varphi_\lambda(x)\right)}{\sqrt{x^3-\frac{27}{64}\lambda^2}}\\ &\ge \frac{2}{3}\left[1+\cos\left(\frac{2}{3}\pi-\frac{2}{3}\varphi_\lambda(x)\right)\right]\\ &\ge \frac{2}{3}\end{aligned}$$

and

$$\begin{aligned}\acute{f}_{\lambda,1/2}(x) &= \frac{2}{3}\left[1+\cos\left(\frac{2}{3}\pi-\frac{2}{3}\varphi_\lambda(x)\right)\right]\\ &\quad+\frac{\frac{\sqrt{3}}{4}\lambda\sin\left(\frac{2}{3}\pi-\frac{2}{3}\varphi_\lambda(x)\right)}{\sqrt{x^3-\frac{27}{64}\lambda^2}}\\ &\le \frac{2}{3}\left[1+\cos\left(\frac{2}{3}\pi-\frac{2}{3}\varphi_\lambda(x)\right)\right]\\ &\quad+\frac{\frac{\sqrt{3}}{4}\lambda\sin\left(\frac{2}{3}\pi-\frac{2}{3}\varphi_\lambda(x)\right)}{\frac{3\sqrt{3}\lambda}{8}}\\ &\le \frac{5}{3}.\end{aligned}$$

Therefore, $(2/3) \le \acute{f}_{\lambda,1/2}(x) \le (5/3)$ follows. This justifies the strictly increasing property and Lipschitz property of $f_{\lambda,1/2}(x)$ whenever $x > (\sqrt[3]{54}/4)\lambda^{2/3}$.

Appendix B
Any Limit Point of Half Algorithm is Its Stationary Point

We prove here that any limit point of the half thresholding algorithm (47) with *Scheme 1* must be its stationary point. More specifically, we prove the following theorem.

Theorem 4: Let $\{x_n\}$ be the sequence defined by *half* thresholding algorithm (47) with *Scheme 1*, and

$$\mu < \left\{\frac{\lambda^{2/3}}{\sqrt[3]{32}\left\{\frac{\sqrt{N}\|A^\top A\|_F[C_\lambda(x_0)]^2}{\lambda^2}+\frac{9\lambda^{2/3}\sqrt{N}\|A^\top A\|_F}{(\sqrt[3]{54})^2\|A\|_2^{4/3}}+\|A^\top y\|_F\right\}}\right\}^3. \quad (54)$$

Then any limit point x^* of $\{x_n\}$ satisfies

$$x^* = H_{\lambda\mu,1/2}(B_\mu(x^*)). \quad (55)$$

Proof: Without loss of generality, we assume $\{x_n\}$ itself converges to x^*. Write $x_n = (x_1^{(n)}, x_2^{(n)}, ..., x_N^{(n)})^\mathsf{T}$ and $x^* = (x_1^*, x_2^*, ..., x_N^*)^\mathsf{T}$. Then $x_i^{(n)} \to x_i^*$ as $n \to \infty$. Since $f_{\lambda\mu,1/2}(.)$ is an odd function (i.e., $f_{\lambda\mu,1/2}(t) = -f_{\lambda\mu,1/2}(-t)$), it suffices to consider the case $[x_n]_i \geq 0$ for any i (Note that $[B_\mu(x_n)]_i > 0$ if $[x_n]_i > 0$). Let $supp(x^*) = \{1, 2, ..., k\}$ be the support set of x^*. Denote

$$\delta_{\lambda\mu} = \frac{\sqrt[3]{54}}{4}(\lambda\mu)^{2/3}.$$

We observe that each x_i^* can be obtained from one of the following three possible ways.

1) There are infinitely many indices $\{n_p\}$ such that $[B_\mu(x_{n_p})]_i \leq \delta_{\lambda\mu}$ $(x_i^* = 0)$.
2) There are infinitely many indices $\{n_p\}$ such that $[B_\mu(x_{n_p})]_i > \delta_{\lambda\mu}$ and $\lim_{p\to\infty}[B_\mu(x_{n_p})]_i = [B_\mu(x^*)]_i > \delta_{\lambda\mu}.(x_i^* \neq 0)$.
3) There are infinitely many indices $\{n_p\}$ such that $[B_\mu(x_{n_p})]_i > \delta_{\lambda\mu}$ but $\lim_{p\to\infty}[B_\mu(x_{n_p})]_i = [B_\mu(x^*)]_i = \delta_{\lambda\mu}$ $(x_i^* \neq 0)$.

Let us express x^* as

$$x^* = (X_1^*, X_2^*, X_3^*)^\mathsf{T} \tag{56}$$

where $X_1^* = (x_i^* : x_i^*$ is deduced from the case 1), $X_2^* = (x_i^* : x_i^*$ is deduced from the case 2), and $X_3^* = (x_i^* : x_i^*$ is deduced from the case 3), with the dimensions of X_1^*, X_2^*, X_3^* being, respectively, N_1, N_2 and N_3.(Here and henceforth, X_j^*, $j = 1, 2, 3$, is understood both as a vector and as a set.) It is obvious that $X_1^* = 0$ and $k = N_2 + N_3$.

Below, we proceed by showing that X_3^* is an empty set, i.e., the 3) case can never happen.

To this end, let us notice that, from the property $f_{\lambda\mu,1/2}(t) < t$, whenever $t > (3/4)(\lambda\mu)^{2/3}$, we have $[x_{n_p+1}]_i = f_{\lambda\mu,1/2}([B_\mu(x_{n_p})]_i) < [B_\mu(x_{n_p})]_i$ and therefore, for any $x_i^* \in X_3^*$

$$x_i^* = \zeta = \lim_{p\to\infty} f_{\lambda\mu,1/2}([B_\mu(x_{n_p})]_i) < [B_\mu(x^*)]_i = \delta_{\lambda\mu} \tag{57}$$

where

$$\zeta = \lim_{t\to\delta_{\lambda\mu}+} f_{\lambda\mu,1/2}(t) = \frac{9}{(\sqrt[3]{54})^2}(\lambda\mu)^{2/3} = \frac{2}{3}\delta_{\lambda\mu} \tag{58}$$

and

$$\delta_{\lambda\mu} - \zeta = \frac{1}{\sqrt[3]{32}}(\lambda\mu)^{2/3}. \tag{59}$$

So, we can further express $X_3^* = \zeta E_3$,, with E_3 being the row vector of dimension N_3 whose components are all 1. We correspondingly split A as

$$A = [A_1, A_2, A_3], A^\mathsf{T} = [A_1^\mathsf{T}, A_2^\mathsf{T}, A_3^\mathsf{T}]^\mathsf{T}.$$

Then we can calculate $B_\mu(x^*)$ by

$$\begin{aligned} B_\mu(x^*) &= x^* - \mu A^\mathsf{T}(Ax^* - y) \\ &= \begin{pmatrix} 0 - \mu A_1^\mathsf{T}(A_2X_2^* + A_3X_3^* - y) \\ X_2^* - \mu A_2^\mathsf{T}(A_2X_2^* + A_3X_3^* - y) \\ X_3^* - \mu A_3^\mathsf{T}(A_2X_2^* + A_3X_3^* - y) \end{pmatrix}. \end{aligned} \tag{60}$$

Now if $X_3^* \neq \varnothing$, then, by assumption, we have

$$X_3^* - \mu A_3^\mathsf{T}(A_2X_2^* + A_3X_3^* - y) = \delta_{\lambda\mu}E_3.$$

Thus

$$\zeta E_3 - \mu A_3^\mathsf{T}(A_2X_2^* + A_3\zeta E_3 - y) = \delta_{\lambda\mu}E_3$$

or equivalently

$$(\delta_{\lambda\mu} - \zeta)E_3 = -\mu A_3^\mathsf{T}(A_2X_2^* + A_3\zeta E_3 - y) \tag{61}$$

which implies

$$\begin{aligned} \sqrt{N_3}(\delta_{\lambda\mu} - \zeta) &= (\delta_{\lambda\mu} - \zeta)\,\|E_3\|_F \\ &= \mu\left\|A_3^\mathsf{T}(A_2X_2^* + A_3\zeta E_3 - y)\right\|_F \\ &\leq \mu\left\{\left\|A_3^\mathsf{T}A_2\right\|_F\left\|X_2^*\right\|_F + \sqrt{N_3}\zeta\left\|A_3^\mathsf{T}A_3\right\|_F\right\} \\ &\quad + \mu\left\{\left\|A_3^\mathsf{T}y)\right\|_F\right\}. \end{aligned} \tag{62}$$

Theorem 3 (i) has shown that $\{C_\lambda(x_n)\}$ is monotonically decreasing, and particularly

$$\lambda\,\|x_n\|_{1/2}^{1/2} = \lambda\sum_{i=1}^N |[x_n]_i|^{1/2} \leq C_\lambda(x_n) \leq C_\lambda(x_0).$$

So $\{x_n\}$ must be bounded, and, in particular, we can deduce that for each x_i^*,

$$\left|x_i^*\right| \leq (\left|x_i^*\right|^{1/2})^2 \leq (\lim_{n\to\infty}\|x_n\|_{1/2}^{1/2})^2 \leq \frac{1}{\lambda^2}[C_\lambda(x_0)]^2.$$

From (59) and (62), this then implies

$$\begin{aligned} \frac{1}{\sqrt[3]{32}}(\lambda\mu)^{2/3} &= (\delta_{\lambda\mu} - \zeta) \\ &\leq \frac{\mu}{\sqrt{N_3}}\left\{\frac{\sqrt{N_2}\left\|A_3^\mathsf{T}A_2\right\|_F[C_\lambda(x_0)]^2}{\lambda^2}\right\} \\ &\quad + \frac{\mu}{\sqrt{N_3}}\left\{\frac{9(\lambda\mu)^{2/3}\sqrt{N_3}\left\|A_3^\mathsf{T}A_3\right\|_F}{(\sqrt[3]{54})^2} + \left\|A_3^\mathsf{T}y\right\|_F\right\} \\ &\leq \mu\left\{\frac{\sqrt{N}\,\|A^\mathsf{T}A\|_F\,[C_\lambda(x_0)]^2}{\lambda^2}\right\} \\ &\quad + \mu\left\{\frac{9\lambda^{2/3}\sqrt{N}\,\|A^\mathsf{T}A\|_F}{(\sqrt[3]{54})^2\,\|A\|_2^{4/3}} + \left\|A^\mathsf{T}y\right\|_F\right\}. \end{aligned}$$

That is

$$\begin{aligned} \frac{1}{\sqrt[3]{32}}\lambda^{2/3} &\leq \mu^{1/3}\left\{\frac{\sqrt{N}\,\|A^\mathsf{T}A\|_F\,[C_\lambda(x_0)]^2}{\lambda^2} + \left\|A^\mathsf{T}y\right\|_F\right\} \\ &\quad + \mu^{1/3}\left\{\frac{9\lambda^{2/3}\sqrt{N}\,\|A^\mathsf{T}A\|_F}{(\sqrt[3]{54})^2\,\|A\|_2^{4/3}}\right\}. \end{aligned} \tag{63}$$

However, by assumption

$$\mu^{1/3} < \frac{\lambda^{2/3}}{\sqrt[3]{32}\left\{\frac{\sqrt{N}\|A^\mathsf{T}A\|_F[C_\lambda(x_0)]^2}{\lambda^2} + \frac{9\lambda^{2/3}\sqrt{N}\|A^\mathsf{T}A\|_F}{(\sqrt[3]{54})^2\|A\|_2^{4/3}} + \|A^\mathsf{T}y\|_F\right\}}$$

which implies

$$\mu^{1/3}\left\{\frac{\sqrt{N}\,\|A^\mathsf{T}A\|_F\,[C_\lambda(x_0)]^2}{\lambda^2} + \frac{9\lambda^{2/3}\sqrt{N}\,\|A^\mathsf{T}A\|_F}{(\sqrt[3]{54})^2\,\|A\|_2^{4/3}}\right\} + \mu^{1/3}\left\{\left\|A^\mathsf{T}y\right\|_F\right\} < \frac{\lambda^{2/3}}{\sqrt[3]{32}}.$$

This obviously contradicts (63). This contradiction shows that $X_3^* = \varnothing$.

From (56), thus, we have $x^* = (X_1^*, X_2^*)^{\mathsf{T}}$, and, for each $x_i^* \in X_1^* \cup X_2^*$, the continuity of $f_{\lambda\mu}(.)$ can be directly applied to derive

$$x_i^* = \lim_{p\to\infty} f_{\lambda\mu}([B_\mu(x_{n_p})]_i) = f_{\lambda\mu}([B_\mu(x^*)]_i).$$

That is, x^* is a stationary point of (55). This completes the proof of Theorem 1.

ACKNOWLEDGMENT

The authors would like to thank the anonymous referees for their insightful comments, as well as Prof. J. Chen and Dr. C. Xu, Department of Statistics, University of British Columbia, Vancouver, BC, Canada, for a critical reading of the manuscript, which led to a significant improvement of this paper.

REFERENCES

[1] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Royal Stat. Soc. Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.

[2] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.

[3] E. Candes and P. Randall, "Highly robust error correction by Convex programming," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 2829–2840, Jul. 2008.

[4] E. Candes and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, 2008.

[5] E. Candes, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, Aug. 2006.

[6] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[7] Z. B. Xu, H. Zhang, Y. Wang, and X. Y. Chang, "$L_{1/2}$ regularization," *Sci. China*, vol. 53, no. 6, pp. 1159–1169, 2010.

[8] D. Merhej, C. Diab, M. Khalil, and R. Prost, "Embedding prior knowledge within compressed sensing by neural networks," *IEEE Trans. Neural Netw.*, vol. 22, no. 10, pp. 1638–1649, Oct. 2011.

[9] B. K. Natraajan, "Sparse approximation to linear systems," *SIAM J. Comput.*, vol. 24, no. 2, pp. 227–234, 1995.

[10] D. L. Donoho, "Neighborly polytopes and the sparse solution of underdetermined systems of linear equations," Stat. Dept., Stanford Univ., Stanford, CA, Tech. Rep. 2005-4, 2005.

[11] D. L. Donoho, "High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension," *Discrete Comput. Geometry*, vol. 35, no. 4, pp. 617–652, 2006.

[12] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.

[13] H. Zou, "The adaptive lasso and its oracle properties," *J. Amer. Stat. Assoc.*, vol. 101, no. 476, pp. 1418–1429, 2006.

[14] N. Meinshausen and B. Yu, "Lasso-type recovery of sparse representations for high-dimensional data," *Ann. Stat.*, vol. 37, no. 1, pp. 246–270, 2009.

[15] R. Chartrand and V. Staneva, "Restricted isometry properties and nonconvex compressive sensing," *Inverse Problems*, vol. 24, no. 3, pp. 20–35, 2008.

[16] D. Krishnan and R. Fergus, "Fast image deconvolution using hyper-Laplacian priors," in *Neural Information Processing Systems*. Cambridge, MA: MIT Press, 2009.

[17] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimizaion," *IEEE Signal Process. Lett.*, vol. 14, no. 10, pp. 707–710, Oct. 2007.

[18] R. Chartrand, "Fast algorithms for nonconvex compressive sensing: MRI reconstruction from very few data," in *Proc. IEEE Int. Symp. Biomed. Imag.*, Jun.–Jul. 2009, pp. 262–265.

[19] Z. B. Xu, "Data modeling: Visual psychology approach and $L_{1/2}$ regularization theory," in *Proc. Int. Congr. Math.*, vol. 4. 2010, pp. 3151–3184.

[20] Z. B. Xu, H. Guo, Y. Wang, and H. Zhang, "The representation of $L_{1/2}$ regularizer among L_q $(0 < q < 1)$ regularizer: An experimental study based on phase diagram," *Acta Autom. Sinica*, 2011, to be published.

[21] A. Rakotomamonjy, R. Flamary, G. Gasso, and S. Canu, "ℓ_p-ℓ_q penalty for sparse linear and sparse multiple kernel multitask learning," *IEEE Trans. Neural Netw.*, vol. 22, no. 1, pp. 1307–1320, Aug. 2011.

[22] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Appl. Comput. Harmon. Anal.*, vol. 27, no. 3, pp. 265–274, 2009.

[23] T. Blumensath and M. E. Davies, "Iterative thresholding for sparse approximations," *J. Fourier Anal. Appl.*, vol. 14, nos. 5–6, pp. 629–654, 2008.

[24] K. Bredies and D. A. Lorenz, "Iterated hard shrinkage for minimization problems with sparsity constraints," *SIAM J. Sci. Comput.*, vol. 30, no. 2, pp. 657–683, 2008.

[25] K. K. Herrity, A. C. Gilbert, and J. A. Tropp, "Sparse approximation via iterative thresholding," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, vol. 3. May 2006, pp. 624–627.

[26] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 613–627, May 1995.

[27] N. G. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *J. Appl. Comput. Harmon. Anal.*, vol. 10, no. 3, pp. 234–253, 2001.

[28] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[29] J. A. Tropp and A. C. Gilbert, "Signal recovery from partial information via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[30] D. Needell and R. Vershynin, "Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 310–316, Apr. 2010.

[31] R. Gribonval and P. Vandergheynst, "On the exponential convergence of matching pursuits in quasi-incoherent dictionaries," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 255–261, Jan. 2006.

[32] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory,* vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[33] J. Friedman, T. Hastie, H. Höfling, and R. Tibshirani, "Pathwise coordinate optimization," *Ann. Appl. Stat.*, vol. 1, no. 2, pp. 302–332, 2007.

[34] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Ann. Stat.*, vol. 32, no. 2, pp. 407–499, 2004.

[35] J. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Stat.*, vol. 29, no. 5, pp. 1189–1232, 2001.

[36] M. Bertero and P. Boccacci, *Introduction to Inverse Problems in Imaging*. New York: Taylor & Francis, 1998.

[37] F. C. Xing, "Investigation on solutions of cubic equations with one unknown," *J. Central Univ. Nat. (Natural Sci. Ed.)*, vol. 12, no. 3, pp. 207–218, 2003.

[38] X. Chen, F. Xu, and Y. Ye, "Lower bound theory of nonzero entries in solutions of l_2-l_p minimization," *SIAM J. Sci. Comput.*, vol. 32, no. 5, pp. 2832–2852, 2010.

[39] I. Daubechies, M. Defrise, and D. M. Christine, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 1, pp. 1413–1457, 2004.

[40] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *SIAM J. Multiscale Model. Simul.*, vol. 4, no. 4, pp. 1168–1200, 2005.

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

[42] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Proc. 2nd Int. Symp. Inf. Theory*, 1973, pp. 267–281.

[43] G. Schwarz, "Estimating the dimension of a model," *Ann. Stat.*, vol. 6, no. 2, pp. 461–464, 1978.

[44] Z. B. Xu, J. J. Wang, and Z. S. Zhang, "Convergence of iterative half thresholding algorithm for $L_{1/2}$ regularization," to be published.

[45] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[46] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 600–616, Mar. 1997.

[47] V. Stodden, "Model selection when the number of variables exceeds the number of observations," Ph.D. dissertation, Dept. Stat., Stanford Univ., Stanford, CA, 2006.

Zongben Xu received the Ph.D. degree in mathematics from Xi'an Jiaotong University, Xi'an, China, in 1987.

He currently serves as a Vice President with Xi'an Jiaotong University, the Academician of the Chinese Academy of Sciences, the Chief Scientist of the National Basic Research Program of China (973 Project), and the Director of the Institute for Information and System Sciences of the University. His current research interests include nonlinear functional analysis and intelligent information processing.

Dr. Xu was a recipient of the National Natural Science Award of China in 2007 and was a winner of the CSIAM Su Buchin Applied Mathematics Prize in 2008. He delivered a talk at the International Congress of Mathematicians 2010.

Xiangyu Chang (S'12) received the B.Sc. degree in applied mathematics from Xi'an Jiaotong University, Xi'an, China, in 2007. He is currently pursuing the Ph.D. degree with the School of Mathematics and Statistics, Xi'an Jiaotong University.

He has been with the Statistics Department, University of California, Berkeley, as a Visiting Research Student since December 2010. His current research interests include statistical machine learning, learning theory, high-dimensional statistics, social network analysis, and social media analysis.

Fengmin Xu received the Ph.D. degree in mathematics from Xi'an Jiaotong University, Xi'an, China, in 2005.

She is currently an Associate Professor with Xi'an Jiaotong University. Her current research interests include nonlinear algorithm analysis and intelligent information processing.

Hai Zhang received the Masters degree in mathematics from Northwest University, Xi'an, China, in 2003.

He is currently an Associate Professor with the Department of Mathematics, Northwest University. His current research interests include sparse machine learning and complex systems.