Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2016, Article ID 3567095, 11 pages
http://dx.doi.org/10.1155/2016/3567095

Research Article

ℓ_1- and ℓ_2-Norm Joint Regularization Based Sparse Signal Reconstruction Scheme

Chanzi Liu, Qingchun Chen, Bingpeng Zhou, and Hengchao Li

Southwest Jiaotong University, Chengdu, Sichuan 610031, China

Correspondence should be addressed to Chanzi Liu; liuchanzi@126.com

Received 10 May 2016; Revised 7 July 2016; Accepted 10 July 2016

Academic Editor: Nazrul Islam

Many problems in signal processing and statistical inference involve finding sparse solution to some underdetermined linear system of equations. This is also the application condition of compressive sensing (CS) which can find the sparse solution from the measurements far less than the original signal. In this paper, we propose ℓ_1- and ℓ_2-norm joint regularization based reconstruction framework to approach the original ℓ_0-norm based sparseness-inducing constrained sparse signal reconstruction problem. Firstly, it is shown that, by employing the simple conjugate gradient algorithm, the new formulation provides an effective framework to deduce the solution as the original sparse signal reconstruction problem with ℓ_0-norm regularization item. Secondly, the upper reconstruction error limit is presented for the proposed sparse signal reconstruction framework, and it is unveiled that a smaller reconstruction error than ℓ_1-norm relaxation approaches can be realized by using the proposed scheme in most cases. Finally, simulation results are presented to validate the proposed sparse signal reconstruction approach.

1. Introduction

Compressive sensing or compressive sampling (CS) [1–3] is a novel technique that enables efficient sampling below Nyquist rate, without (or with little) sacrificing reconstruction quality. CS converts the high-dimensional sparse signal into a significantly lower dimensional measurement signal. More precisely, let $\mathbf{x} \in \mathbb{R}^N$ be a sparse vector, let $\mathbf{A} \in \mathbb{R}^{M\times N}$ $(M < N)$ be a measurement matrix, and suppose the noisy observation vector $\mathbf{y}$ is given by

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}, \tag{1}$$

where $\mathbf{e} \in \mathbb{R}^M$ is the noise vector. Here sparsity means that $\|\mathbf{x}\|_0 = K$ and $K \ll N$; ℓ_0-norm counts the number of nonzero items of a vector. The goal is to obtain an estimate of $\mathbf{x}$ given $\mathbf{A}$ and $\mathbf{y}$.

Determining the sparse signal in $\mathbb{R}^N$ from $M < N$ measurements is typically an underdetermined problem; obviously, there will be no unique solution without any prior knowledge or constraint imposed on the solution $\mathbf{x}$. As the original signal $\mathbf{x}$ is sparse, the problem of finding the desired solution can be phrased as some optimization problem, where the objective is to minimize the noise between noisy observation vector and real measurement while satisfying the constraint of the sparsity of $\mathbf{x}$. As the sparsity of $\mathbf{x}$ is reflected by the number of its nonzero entries, equivalently its so-called ℓ_0-norm, in the noise-free case of (1), we can seek to solve the following P_0 problem:

$$\begin{aligned} P_0\colon \min_{\mathbf{x}} \quad & \|\mathbf{x}\|_0, \\ \text{s.t.} \quad & \mathbf{y} = \mathbf{A}\mathbf{x}. \end{aligned} \tag{2}$$

P_0 can recover $\mathbf{x}$ exactly if $\mathbf{x}$ is sufficiently sparse and the matrix $\mathbf{A}$ satisfies the requirement of Restricted Isometry Constant (RIC) δ_K, which can be defined as the smallest constant $0 \le \delta_K \le 1$ such that the matrix $\mathbf{A}$ satisfies the Restricted Isometry Property (RIP) of order K; namely,

$$(1-\delta_K)\|\mathbf{x}\|_2^2 \le \|\mathbf{A}\mathbf{x}\|_2^2 \le (1+\delta_K)\|\mathbf{x}\|_2^2, \tag{3}$$

whenever $\|\mathbf{x}\|_0 \le K$. However, the optimization problem in (2) is nondeterministic polynomial (NP) hard which is difficult to be solved in practice. A number of approaches are presented to solve (2). The related approaches can be briefly classified into the following three categories.

(i) Greedy Pursuits. Greedy algorithms attempt to determine the nonzero entry indices, based on the relationship between the columns of the measurement matrix $\mathbf{A}$ and the signal (or the residual signal), and then to estimate the non-zero entry amplitudes by using the least square method. Its objective is to pursue the minimization of ℓ_0-norm directly. The typical greedy approaches include Orthogonal Matching Pursuit (OMP) [4], Stagewise OMP (StOMP) [5], and Subspace Pursuit (SP) [6]. Basically, the greedy algorithms can only seek an approximate solution to (2) and are sensitive to noise.

(ii) Optimization. This kind of approach attempts to replace ℓ_0-norm minimization problem in (2) with ℓ_1- or ℓ_p-norm ($p \in (0,1)$), admitting tractable algorithms. These methods try to obtain the sparse solution by employing a newly reformulated problem as follows:

$$\begin{aligned} P_1\colon \min_{\mathbf{x}} \quad & \|\mathbf{x}\|_1, \\ \text{s.t.} \quad & \mathbf{y} = \mathbf{A}\mathbf{x}. \end{aligned} \tag{4}$$

In the noise-free case, P_1 can recover $\mathbf{x}$ exactly as soon as $\delta_{3K} + 3\delta_{4K} < 2$. In the noisy case, the above convex relaxation formulation leads to the following penalized least-square problem [7, 8]:

$$\min_{\mathbf{x}} \quad \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1. \tag{5}$$

Later, it is shown in [9] that the reformulated one like $\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \Sigma_i w_i |x_i|$ can better approach the original ℓ_0-norm problem in (2), where w_i denotes the nonnegative weighting coefficient. Iterative reweighting scheme was further investigated in [10]. ℓ_1-magic algorithm in [11], the gradient projection for sparse reconstruction algorithm (GPSR) in [8], and their variants also belong to the convex relaxation categories. ℓ_p-norm minimization problem can be recast as

$$\begin{aligned} P_p\colon \min_{\mathbf{x}} \quad & \|\mathbf{x}\|_p, \\ \text{s.t.} \quad & \mathbf{y} = \mathbf{A}\mathbf{x}, \end{aligned} \tag{6}$$

where $\|\mathbf{x}\|_p = (\sum_{i=1}^{N} |x_i|^p)^{1/p}$. Chartrand demonstrated that K-sparse vectors can be exactly recovered by solving P_p under the assumption that $\delta_{aK} + b\delta_{(a+1)K} < b - 1$ holds for some $b > 1$ and $a = b^{q/(2-q)}$ [12]. This kind of algorithm is capable of achieving better reconstruction quality and less error; however, it takes more time to solve the optimization problems.

(iii) Bayesian Framework. Given the unknown sparse coefficients *a priori* probability distribution, the Maximum A Posteriori (MAP) estimation mechanism can be employed to derive the Bayesian framework based sparse signal recovery schemes [13, 14]. It is unveiled in [15] that this framework may lead to a new mixture penalty of ℓ_0- and ℓ_1-norm functions.

ℓ_0-norm sparsity inducing item is very useful in image processing, for example, image restoration, image denoising, and image superresolution. Since ℓ_0-norm is nonconvex, image processing pursues ℓ_1- and ℓ_2-norm or total variation (TV) norm to replace the original ℓ_0-norm and solves the problems. Reference [16] studies a minimization problem where the objective includes a usual ℓ_2-norm data-fidelity term and an overlapping group sparsity total variation regularizer, and a fast algorithm is proposed to solve this problem. This method can avoid staircase effect and allow edges preserving. Reference [17] proposes ℓ_1-norm fidelity term with a total variation regularizer to recover blurred and salt-and-pepper impulse noisy image for higher visual quality. Reference [18] proposes an optimization model for noise and blur removal involving the generalized variation regularization term, the MAP (maximum posterior probability) based data fitting term, and a quadratic penalty term based on the statistical property of the noise. This minimization problem can be solved by a primal-dual algorithm. Reference [19] studies TV regularization in deblurring and sparse unmixing of hyperspectral images. Reference [20] investigates the modified linearized Bregman algorithm (MLBA) for image deblurring problems with a proper treatment of the boundary artifacts and ℓ_1- plus ℓ_2-norm.

Motivated by the utility of sparsity and the efforts in relaxing ℓ_0-norm function with tractable norm functions to get either convex or nonconvex relaxation problem, in this paper, we propose a novel ℓ_1- and ℓ_2-norm joint regularization based reconstruction framework to approach the original ℓ_0-norm sparseness-inducing constraint based reconstruction problem. Although [21] has shown that ℓ_p and $\ell_1 - \ell_2$ measures are theoretically better than ℓ_1-norm to promote sparsity, our analysis shows that the proposed sparse signal reconstruction model can also solve the original P_0 problem. Moreover, the upper error limit of the proposed sparse signal recovery model is derived. It is unveiled that the proposed signal reconstruction model can achieve the tradeoff between ℓ_1-norm relaxation and ℓ_p-norm relaxation techniques. Namely, it exhibits similar ℓ_0-norm approximation capability like ℓ_p-norm relaxation. At the same time, like ℓ_1-norm convex relaxation approaches, we can resort to a variety of feasible optimization algorithms to derive the solution.

The remainder of this paper is organized as follows. The proposed sparse signal recovery model will be introduced in Section 2. Nextly, we deduce the error bound of the sparse signal reconstruction in Section 3. The practical algorithm design and experimental results are demonstrated in Sections 4 and 5. Finally, we conclude this paper in Section 6.

2. The Proposed Sparse Signal Recovery Model

2.1. Problem Reformulation. The reconstruction algorithm plays a central role in the sparse signal reconstruction. We propose ℓ_1- and ℓ_2-norm joint regularization to approximate ℓ_0-norm and the sparse signal recovery problem can be reformulated as follows:

$$\begin{aligned} P_{1,2}\colon \min_{\mathbf{x}} \quad & \lambda \|\mathbf{x}\|_1 - \tau \|\mathbf{x}\|_2, \\ \text{s.t.} \quad & \mathbf{y} = \mathrm{A}\mathbf{x}. \end{aligned} \tag{7}$$

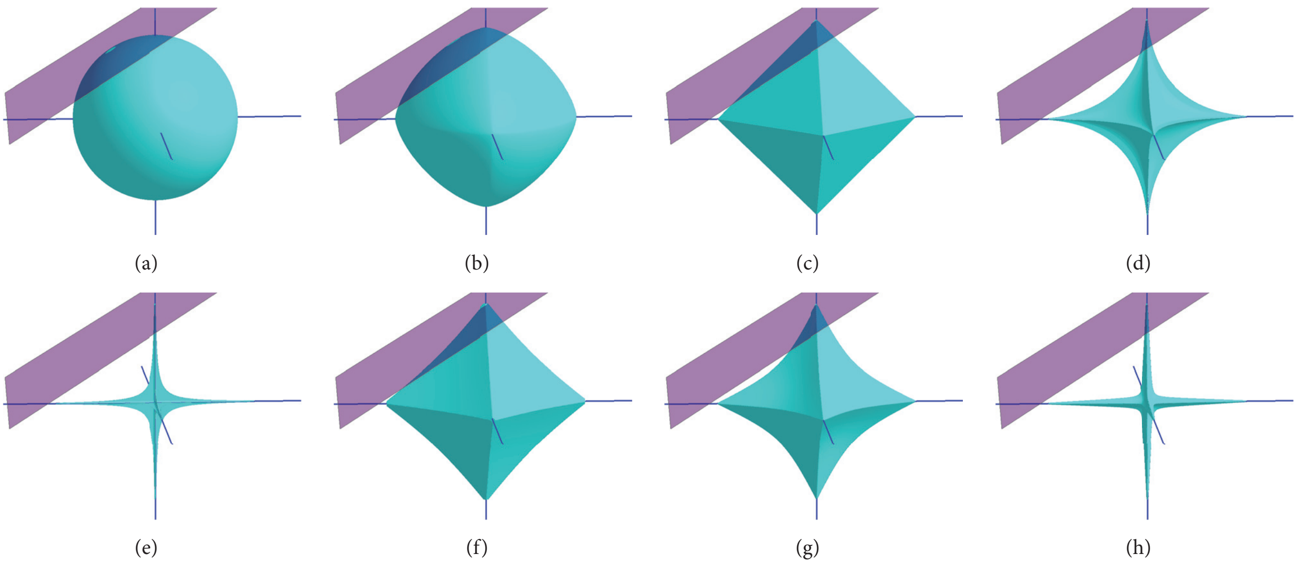

Figure 1: The illustration of the sparse signal reconstruction in 3D space, where the intersection between the unit-length norm ball and the hyperplane defines the solution. ℓ_p-norm illustration in (a)–(e); the proposed joint ℓ_1- and ℓ_2-norm in (f)–(h). (a) $p = 2$; (b) $p = 1.5$; (c) $p = 1$; (d) $p = 0.5$; (e) $p = 0.1$; (f) $\lambda/\tau = 5$; (g) $\lambda/\tau = 2$; (h) $\lambda/\tau = 1.05$. It can be noted that when $p \le 1$, the intersection takes place on the axes, leading to a sparse solution. The intersection of the proposed joint ℓ_1- and ℓ_2-norm case will take place on the axes as well.

We will try to explicate that the joint combination of ℓ_1- and ℓ_2-norm regularization provides a reasonable approximation to ℓ_0-norm, which can bring surprising benefits. Let us show this from the geometry point of view at first. The linear constraint of $\mathbf{y} = \mathbf{A}\mathbf{x}$ defines the feasible set of the problem. Geometrically, solving (4) (or (6)) is done by blowing ℓ_1- (or ℓ_p-) norm balloon centered around the origin and stopping its inflation once it touches the feasible set. Figure 1 illustrates some examples with different p values of 2, 1.5, 1, 0.5, and 0.1, respectively. One can see that the norms with $p \le 1$ tend to touch the feasible set at the axes, which leads to the sparse solutions. On the other hand, ℓ_2- or $\ell_{1.5}$-norm tends to derive the nonsparse solution. One can also observe from Figure 1 that, by using the joint ℓ_1- and ℓ_2-norm in (7) with the suitable ratio of λ/τ, the intersection will also take place at the axes, which leads to sparse solution as well. Namely, the proposed signal reconstruction model resembles ℓ_0-norm approximation capability like ℓ_p-norm relaxation scheme. The influences of different choices of λ and τ will be highlighted in the following discussions.

2.2. Sparse Solution Derivation Analysis. In this subsection, the analysis will validate that we can derive the same sparse solution by using the joint ℓ_1- and ℓ_2-norm instead of ℓ_0-norm function theoretically; that is to say, the solution to (7) is exactly $\mathbf{x}$, when the certain signal sparsity requirement can be satisfied. Fortunately, if the original problem actually has a sufficiently sparse solution, the success of the proposed model in addressing the original objective (2) can also be guaranteed. Some definitions are firstly introduced.

Definition 1 (see [22]). The mutual-coherence of a given matrix $\mathbf{A}$ is the largest absolute normalized inner product between different columns of $\mathbf{A}$. Let $\mathbf{A}_l$ denote the lth column in $\mathbf{A}$, $l = 1, 2, \ldots, N$; the mutual-coherence can be calculated as

$$\mu = \max_{1\le i,j\le N, i\ne j} \frac{\left|\mathbf{A}_i^T \mathbf{A}_j\right|}{\left\|\mathbf{A}_i\right\|_2 \left\|\mathbf{A}_j\right\|_2}. \tag{8}$$

The mutual-coherence provides a metric to assess the dependence between columns of the matrix $\mathbf{A}$.

Conclusion 1. For the linear system of equations $\mathbf{y} = \mathbf{A}\mathbf{x}$ ($\mathbf{A} \in \mathbb{R}^{M\times N}$ and $M < N$), $\mathbf{x}$ is assumed to be K-sparse; namely, $\|\mathbf{x}\|_0 = K$; if a solution $\mathbf{x}$ exists obeying

$$K \le \frac{\tau^2}{4\lambda^2} + \frac{1}{2}\left(1 + \frac{1}{\mu}\right) - \frac{\tau}{4\lambda}\sqrt{\frac{\tau^2}{\lambda^2} + 4\left(1 + \frac{1}{\mu}\right)}, \tag{9}$$

$\mathbf{x}$ will be the unique solution to (2) and (7).

Proof. Assume that there are possible solutions $\{\mathbf{w}\}$ different from $\mathbf{x}$, which have larger support and can satisfy $\mathbf{y} = \mathbf{A}\mathbf{w}$. Define the following set of alternative solutions:

$$\mathscr{C} = \left\{ \mathbf{w} \;\middle|\; \begin{array}{c} \mathbf{w} \ne \mathbf{x}, \\ \lambda\|\mathbf{w}\|_1 - \tau\|\mathbf{w}\|_2 \le \lambda\|\mathbf{x}\|_1 - \tau\|\mathbf{x}\|_2, \\ \|\mathbf{w}\|_0 \ge \|\mathbf{x}\|_0, \\ \mathbf{A}(\mathbf{w} - \mathbf{x}) = \mathbf{0}. \end{array} \right\}. \tag{10}$$

This set contains all the possible solutions $\{\mathbf{w}\}$ different from $\mathbf{x}$, which have larger support and can satisfy $\mathbf{y} = \mathbf{A}\mathbf{w}$. Nonempty $\mathscr{C}$ implies that there is at least one alternative solution to (7), rather than the desired $\mathbf{x}$.

Theorem 2.5 in [22] tells us that if $\|\mathbf{x}\|_0 \le 0.5(1 + 1/\mu)$, $\mathbf{x}$ is necessarily the unique sparsest solution; any other alternative solutions $\mathbf{w} \neq \mathbf{x}$ must be denser; that is to say, $\mathbf{w}$ has more nonzero items. Our requirement on sparsity is $\tau^2/4\lambda^2 + (1/2)(1+1/\mu) - (\tau/4\lambda)\sqrt{\tau^2/\lambda^2 + 4(1+1/\mu)}$, which is less than $0.5(1+1/\mu)$, no matter how to select the nonnegative parameters of λ and τ. It implies that $\mathbf{x}$ is also the unique sparsest possible solution, so $\|\mathbf{w}\|_0 \geq \|\mathbf{x}\|_0$ can be removed from $\mathscr{C}$. Let us define $\mathbf{v} = \mathbf{w} - \mathbf{x}$; we can rewrite $\mathscr{C}$ as the following shifted version around $\mathbf{x}$:

$$\mathscr{C}_s = \left\{ \mathbf{v} \;\middle|\; \begin{array}{c} \mathbf{v} \neq \mathbf{0}, \\ \mathbf{A}\mathbf{v} = \mathbf{0}, \\ \lambda \|\mathbf{v}+\mathbf{x}\|_1 - \tau \|\mathbf{v}+\mathbf{x}\|_2 - \lambda \|\mathbf{x}\|_1 + \tau \|\mathbf{x}\|_2 \leq 0 \end{array} \right\}. \quad (11)$$

The strategy of the proof we are going to present is to enlarge this set and show that the enlarged set is empty. This will prove that (7) indeed succeeds in recovering $\mathbf{x}$. We start with the requirement $\lambda\|\mathbf{v}+\mathbf{x}\|_1 - \tau\|\mathbf{v}+\mathbf{x}\|_2 - \lambda\|\mathbf{x}\|_1 + \tau\|\mathbf{x}\|_2 \leq 0$. Without loss of generality, the K nonzeros in $\mathbf{x}$ are moved to the first K positions through a simple column permutation, which will not change the aforementioned inequality; we may have

$$\lambda \|\mathbf{v}+\mathbf{x}\|_1 - \lambda \|\mathbf{x}\|_1 = \lambda \sum_{i=1}^{K} \left(|v_i + x_i| - |x_i|\right) + \lambda \sum_{i>K} |v_i|. \quad (12)$$

Here we use $|\cdot|$ to denote the element-wise absolute operation for every element of the matrix (vector), and the result of the matrix (vector) is still a matrix (vector); only the elements of the matrix (vector) are now all positive. By using the common inequality $|\mathbf{a}+\mathbf{b}| - |\mathbf{b}| \geq -|\mathbf{a}|$, we can relax the above condition as follows:

$$-\lambda \sum_{i=1}^{K} |v_i| + \lambda \sum_{i>K} |v_i| \leq \lambda \sum_{i=1}^{K} \left(|v_i + x_i| - |x_i|\right) + \lambda \sum_{i>K} |v_i|. \quad (13)$$

As for $\tau\|\mathbf{x}\|_2 - \tau\|\mathbf{v}+\mathbf{x}\|_2$, because of the triangle inequality property of vector norm, we have $\|\mathbf{v}+\mathbf{x}\|_2 \leq \|\mathbf{v}\|_2 + \|\mathbf{x}\|_2$. Then

$$-\tau \|\mathbf{v}\|_2 \leq \tau \|\mathbf{x}\|_2 - \tau \|\mathbf{v}+\mathbf{x}\|_2. \quad (14)$$

We use $\mathbf{1}_K^T \cdot |\mathbf{v}|$ to denote the sum of the first K entries of the vector $|\mathbf{v}|$, where $\mathbf{I}_K$ denotes the column vector with K ones followed by N-K zeros; namely, $\sum_{i=1}^{K} |v_i|$. So we may have

$$\lambda \left(\|\mathbf{v}\|_1 - 2\mathbf{1}_K^T \cdot |\mathbf{v}| \right) - \tau \|\mathbf{v}\|_2 \leq 0. \quad (15)$$

Thus we have

$$\mathscr{C}_s \subseteq \left\{ \mathbf{v} \;\middle|\; \begin{array}{c} \mathbf{v} \neq \mathbf{0}, \\ \mathbf{A}\mathbf{v} = \mathbf{0}, \\ \lambda \left(\|\mathbf{v}\|_1 - 2\mathbf{1}_K^T \cdot |\mathbf{v}| \right) - \tau \|\mathbf{v}\|_2 \leq 0 \end{array} \right\} = \mathscr{C}_s^1. \quad (16)$$

As stated before, the relaxation of the new alternative $\mathscr{C}_s^1$ will effectively enlarge the initial set $\mathscr{C}_s$. We now turn to handle the requirement $\mathbf{A}\mathbf{v} = \mathbf{0}$ by replacing it with a relaxed requirement that expands the set $\mathscr{C}_s^1$ further. We can normalize the matrix $\mathbf{A}$ such that each column $\mathbf{A}_l$ is of unit ℓ_2-norm (this operation will preserve the mutual-coherence, as shown in [22]). Then the multiplication by $\mathbf{A}^T$ yields the condition $\mathbf{A}^T\mathbf{A}\mathbf{v} = \mathbf{0}$, which does not change the set $\mathscr{C}_s^1$. By letting $\mathbf{G} = \mathbf{A}^T\mathbf{A}$, this condition can be further rewritten as

$$-\mathbf{v} = \left(\mathbf{A}^T\mathbf{A} - \mathbf{I}\right)\mathbf{v}. \quad (17)$$

Taking an element-wise absolute value on both sides, we can relax the requirement on $\mathbf{v}$ to obtain

$$|\mathbf{v}| = \left|\left(\mathbf{A}^T\mathbf{A} - \mathbf{I}\right)\mathbf{v}\right| \leq \left|\mathbf{A}^T\mathbf{A} - \mathbf{I}\right| \cdot |\mathbf{v}|, \quad (18)$$

where we have used the inequality of $|\sum_i u_i v_i| \leq \sum_i |u_i||v_i|$. Because every column of $\mathbf{A}$ has been normalized, one may readily derive the properties of the entries $G_{i,j} = \mathbf{A}_i^T\mathbf{A}_j$ of the resulting Gram matrix $\mathbf{G} = \mathbf{A}^T\mathbf{A}$ as follows:

(1) $G_{i,i} = 1,\ 1 \leq i \leq N$;

(2) $G_{i,j} \leq \mu,\ 1 \leq i, j \leq N,\ i \neq j$.

By letting $\mathbf{I}$ denote an identity matrix and $\mathbf{P}$ denote a rank-1 matrix filled with ones, we can have

$$\begin{aligned} |\mathbf{v}| &\leq \left|\mathbf{A}^T\mathbf{A} - \mathbf{I}\right| \cdot |\mathbf{v}| \\ &\leq \left| \begin{pmatrix} 1 & \mu & \cdots & \mu \\ \mu & 1 & \cdots & \mu \\ \vdots & \vdots & \ddots & \vdots \\ \mu & \mu & \cdots & 1 \end{pmatrix} - \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \right| \cdot |\mathbf{v}| \\ &= \begin{pmatrix} 0 & \mu & \cdots & \mu \\ \mu & 0 & \cdots & \mu \\ \vdots & \vdots & \ddots & \vdots \\ \mu & \mu & \cdots & 0 \end{pmatrix} \cdot |\mathbf{v}| = \mu (\mathbf{P} - \mathbf{I}) \cdot |\mathbf{v}| \\ &= \mu\mathbf{P} \cdot |\mathbf{v}| - \mu |\mathbf{v}|. \end{aligned} \quad (19)$$

Based on the above inequality, we can derive that $(1+\mu)|\mathbf{v}| \leq \mu\mathbf{P} \cdot |\mathbf{v}|$. Returning to the set $\mathscr{C}_s^1$, we have

$$\mathscr{C}_s^1 \subseteq \left\{ \mathbf{v} \;\middle|\; \begin{array}{c} \mathbf{v} \neq \mathbf{0}, \\ |\mathbf{v}| \leq \dfrac{\mu}{1+\mu}\mathbf{P} \cdot |\mathbf{v}|, \\ \lambda \left(\|\mathbf{v}\|_1 - 2\mathbf{1}_K^T \cdot |\mathbf{v}| \right) - \tau \|\mathbf{v}\|_2 \leq 0 \end{array} \right\} = \mathscr{C}_s^2. \quad (20)$$

The resultant set $\mathscr{C}_s^2$ is unbounded since if $\mathbf{v} \in \mathscr{C}_s^2$, $a\mathbf{v} \in \mathscr{C}_s^2$ for all $a \neq 0$. Thus we can restrict our quest for the normalized vector $\|\mathbf{v}\|_1 = 1$, because we have $\mathbf{P}|\mathbf{v}| = \mathbf{1} \cdot \mathbf{1}^T|\mathbf{v}|$, where $\mathbf{1}$ denotes the all-ones column vector. Because $\mathbf{1}^T|\mathbf{v}| = \|\mathbf{v}\|_1 = 1$, this new set of $\mathscr{C}_r$ becomes

$$\mathscr{C}_r \subseteq \left\{ \mathbf{v} \;\middle|\; \begin{array}{c} \|\mathbf{v}\|_1 = 1, \\ |\mathbf{v}| \leq \dfrac{\mu}{1+\mu}\mathbf{1}, \\ \lambda \left(1 - 2\mathbf{1}_K^T \cdot |\mathbf{v}| \right) - \tau \|\mathbf{v}\|_2 \leq 0 \end{array} \right\}. \quad (21)$$

By substituting $|\mathbf{v}| = (\mu/(1+\mu))\mathbf{1}$ into $\lambda(1-2\mathbf{1}_K^T\cdot|\mathbf{v}|)-\tau\|\mathbf{v}\|_2 \le 0$, we have

$$\begin{aligned}&\lambda\left(1-2\mathbf{1}_K^T\cdot|\mathbf{v}|\right)-\tau\|\mathbf{v}\|_2\\&\quad=\lambda\left(1-\frac{2K\mu}{1+\mu}\right)-\tau\sqrt{2K}\left(\frac{\mu}{1+\mu}\right)\le 0,\end{aligned} \tag{22}$$

where we have utilized the fact that $\mathbf{v} = \mathbf{w} - \mathbf{x}$ is at most $2K$-sparse because it is assumed that both $\mathbf{w}$ and $\mathbf{x}$ are K-sparse solutions to $\mathbf{y} = \mathbf{A}\mathbf{x}$ and $\mathbf{w} \ne \mathbf{x}$. From the last inequality, we can obtain

$$\lambda(1+\mu)-2\lambda\mu K-\tau\mu\sqrt{2K}\le 0. \tag{23}$$

Let $\sqrt{2K} = t$; we get $\lambda(1+\mu)-\lambda\mu t^2-\tau\mu t \le 0$. Solving this inequality yields $t \ge (-\tau/\lambda+\sqrt{\tau^2/\lambda^2+4(1+1/\mu)})/2$ and $t \le (-\tau/\lambda-\sqrt{\tau^2/\lambda^2+4(1+1/\mu)})/2$. Since $t \ge 0$, we have $\sqrt{2K} \ge (-\tau/\lambda+\sqrt{\tau^2/\lambda^2+4(1+1/\mu)})/2$, leading to

$$K \ge \frac{\tau^2}{4\lambda^2}+\frac{1}{2}\left(1+\frac{1}{\mu}\right)-\frac{\tau}{4\lambda}\sqrt{\frac{\tau^2}{\lambda^2}+4\left(1+\frac{1}{\mu}\right)}. \tag{24}$$

This implies if K is less than $\tau^2/4\lambda^2 + (1/2)(1+1/\mu) - (\tau/4\lambda)\sqrt{\tau^2/\lambda^2+4(1+1/\mu)}$, the set $\mathscr{C}$ will be necessarily empty, which also implies that we can not find such $\mathbf{w}$ with larger ℓ_0-norm but smaller difference between ℓ_1- and ℓ_2-norm. So our model can find the desired solution. □

For BP algorithm, it is known that if a solution $\mathbf{x}$ exists obeying $\|\mathbf{x}\|_0 \le (1/2)(1+1/\mu)$, the solution will be the unique solution to (2) and (4) as well. In our proposed model, when $\tau = 0$, the requirement will be reduced to $(1/2)(1+1/\mu)$ as well. Otherwise, the proposed model in (7) can derive the same solution when $K \le \tau^2/4\lambda^2 + (1/2)(1+1/\mu) - (\tau/4\lambda)\sqrt{\tau^2/\lambda^2+4(1+1/\mu)}$, which is more stringent sparsity requirement. And this is the paid cost of the proposed ℓ_1- and ℓ_2-norm joint regularization based sparse signal reconstruction approach.

3. Signal Recovery via ℓ_1- and ℓ_2-Norm Joint Minimization

3.1. Noise-Free Signal Recovery. There exist lots of approaches to recover a sparse signal $\mathbf{x}$ from a small number of linear measurements; we attempt to recover $\mathbf{x}$ by solving (7). We begin by considering noise-free case. Our main theorem is similar in flavor to many previous ones; in fact, its proof is inspired by those previous works. We start by illustrating Theorem 2 in the special case of K-sparse vectors that are measured with infinite precision, which means that the signal $\mathbf{x}$ is exactly K-sparse and the relative measurement error is equal to zero. For noise-free sparse signal recovery, we have the following conclusion about the signal recovery error upper limit.

Theorem 2. *Given λ and τ, if $\mathbf{A}$ satisfies the RIP of order $2K$ with $\delta_{2K} < \min\{(2-\sqrt{2})/2, (\lambda\sqrt{K}-\tau)/((2+\sqrt{2})(\lambda\sqrt{K}-\tau)+2\tau)\}$; then every K-sparse vector is exactly recovered by solving $P_{1,2}$.*

Proof. Firstly, we introduce Lemma 3.

Lemma 3. *Suppose that $\mathbf{A}$ satisfies the RIP of order $2K$ with $\delta_{2K} < \min\{(2-\sqrt{2})/2, (\lambda\sqrt{K}-(2+\sqrt{2})\lambda\delta_{2K}\sqrt{K}-\tau)/\sqrt{2}\tau\}$; we obtain the measurement $\mathbf{y} = \mathbf{A}\mathbf{x}$. Let $\mathbf{x}$ and $\widehat{\mathbf{x}}$ be given, and define $\mathbf{h} = \widehat{\mathbf{x}} - \mathbf{x}$. Let Λ_0 denote the index set corresponding to K entries of $\mathbf{x}$ with the largest magnitude and Λ_1 the index set corresponding to K entries of $\mathbf{h}_{\Lambda_0^C}$ with the largest* magnitude, where Λ_0^C stands for the complement set of Λ_0. *Set $\Lambda = \Lambda_0 \cup \Lambda_1$. If $\lambda\|\widehat{\mathbf{x}}\|_1 - \tau\|\widehat{\mathbf{x}}\|_2 \le \lambda\|\mathbf{x}\|_1 - \tau\|\mathbf{x}\|_2$, then*

$$\|\mathbf{h}\|_2 \le \alpha\|\mathbf{A}\mathbf{h}\|_2 + \beta\sigma(\mathbf{x})_1, \tag{25}$$

where $\sigma(\mathbf{x})_1 = \|\mathbf{x}_{\Lambda_0^C}\|_1$, $\alpha = 2(1+\sqrt{2})\sqrt{1+\delta_{2K}}\lambda\sqrt{K}/(\lambda\sqrt{K}-(2+\sqrt{2})\delta_{2K}\lambda\sqrt{K}-(1+\sqrt{2}\delta_{2K})\tau)$, and $\beta = (2+2\sqrt{2}\delta_{2K})\lambda/(\lambda\sqrt{K}-(2+\sqrt{2})\delta_{2K}\lambda\sqrt{K}-(1+\sqrt{2}\delta_{2K})\tau)$.

Lemma 3 establishes an error bound for the proposed signal recovery model $P_{1,2}$ described by (7), when the measurement matrix $\mathbf{A}$ satisfies the RIP constraint. The detailed proof can be found in Appendix. In the case of noise-free measurement, that is, $\mathbf{y} = \mathbf{A}\mathbf{x}$, we can apply Lemma 3 to derive that, for $\widehat{\mathbf{x}} = \mathbf{x}+\mathbf{h}$, $\mathbf{y} = \mathbf{A}\mathbf{x} = \mathbf{A}\widehat{\mathbf{x}}$ yields $\mathbf{A}\mathbf{h} = \mathbf{A}(\widehat{\mathbf{x}}-\mathbf{x}) = \mathbf{y}-\mathbf{y} = \mathbf{0}$, then the term $\mathbf{A}\mathbf{h}$ vanishes, and $\mathbf{x}$ is K-sparse signal, leading to $\sigma(\mathbf{x})_1 = 0$; we can obtain the following result:

$$\|\widehat{\mathbf{x}}-\mathbf{x}\|_2 \le 0. \tag{26}$$

So every K-sparse vector can be exactly recovered by solving $P_{1,2}$ problem. □

Theorem 2 is an immediate consequence of Theorem 4 to be given in the next section. One may readily note that, in noise-free case, by setting appropriate parameter of λ and τ, we may achieve reasonable signal restoration performance, as will be illustrated by the numerical results in Section 4. Compared to the result of Theorem 1.8 in [23] with $\|\widehat{\mathbf{x}}-\mathbf{x}\|_2 \le C_0(\sigma(\mathbf{x})_1/\sqrt{K})$, where $C_0 = 2((1-(1-\sqrt{2})\delta_{2K})/(1-(1+\sqrt{2})\delta_{2K}))$, we can obtain the same result when $\sigma(\mathbf{x})_1 = 0$. We will reveal how the different choices of λ and τ influence the upper error bound of the reconstructed signal in the following section.

3.2. Noisy Signal Recovery. In practical applications, the measurements are likely to be contaminated by various types of noise, such as quantization error and random noise. We suppose

$$\begin{aligned}P_{1,2,\epsilon}\colon \min_{\mathbf{x}}\quad & \lambda\|\mathbf{x}\|_1-\tau\|\mathbf{x}\|_2,\\ \text{s.t.}\quad & \|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2 \le \epsilon.\end{aligned} \tag{27}$$

Note that ϵ represents a relative error between accurate and inaccurate measurements. In order to recover the original vector $\mathbf{x} \in \mathbb{R}^N$ from the measurement of $\mathbf{y}$, we can solve the minimization of the above objective function.

It is obvious that the minimization of $P_{1,2,\epsilon}$ is equivalent to

$$\begin{aligned} \min_{\mathbf{x}} \quad & \lambda \|\mathbf{x}\|_1 - \tau \|\mathbf{x}\|_2 , \\ \text{s.t.} \quad & \lambda \|\mathbf{x}\|_1 - \tau \|\mathbf{x}\|_2 \le \omega, \\ & \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \le \epsilon, \end{aligned} \tag{28}$$

where ω is the minimum of $P_{1,2,\epsilon}$. Because the set $\{\mathbf{x} \in \mathbb{R}^N : \lambda\|\mathbf{x}\|_1 - \tau\|\mathbf{x}\|_2 \le \omega,\ \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \le \epsilon\}$ is compact and $\lambda\|\mathbf{x}\|_1 - \tau\|\mathbf{x}\|_2$ is a continuous function, we can conclude that the minimization of $P_{1,2,\epsilon}$ must exist. So we have the following conclusion.

Theorem 4. *Given λ and τ, if $\mathbf{A}$ satisfies the RIP of order $2K$ with $\delta_{2K} < \min\{(2-\sqrt{2})/2, (\lambda\sqrt{K} - (2+\sqrt{2})\lambda\delta_{2K}\sqrt{K} - \tau)/\sqrt{2}\tau\}$, then the solution $\widehat{\mathbf{x}}$ of $P_{1,2,\epsilon}$ approximates the original vector $\mathbf{x}$ with error*

$$\|\widehat{\mathbf{x}} - \mathbf{x}\|_2 \le 2\alpha\epsilon + \beta\sigma(\mathbf{x})_1 , \tag{29}$$

where $\sigma(\mathbf{x})_1 = \|\mathbf{x}_{\Lambda_0^C}\|_1$, $\alpha = 2(1+\sqrt{2})\sqrt{1+\delta_{2K}}\lambda\sqrt{K}/(\lambda\sqrt{K} - (2+\sqrt{2})\delta_{2K}\lambda\sqrt{K} - (1+\sqrt{2}\delta_{2K})\tau)$, and $\beta = (2+2\sqrt{2}\delta_{2K})\lambda/(\lambda\sqrt{K} - (2+\sqrt{2})\delta_{2K}\lambda\sqrt{K} - (1+\sqrt{2}\delta_{2K})\tau)$.

Proof. Since $\|\mathbf{e}\|_2 \le \boldsymbol{\epsilon}$, we have

$$\begin{aligned} \|\mathbf{A}\mathbf{h}\|_2 &= \|\mathbf{A}(\widehat{\mathbf{x}} - \mathbf{x})\|_2 = \|\mathbf{A}\widehat{\mathbf{x}} - \mathbf{y} + \mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \\ &\le \|\mathbf{A}\widehat{\mathbf{x}} - \mathbf{y}\|_2 + \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \le 2\epsilon. \end{aligned} \tag{30}$$

So, with Lemma 3,

$$\|\widehat{\mathbf{x}} - \mathbf{x}\|_2 \le 2\alpha\epsilon + \beta\sigma(\mathbf{x})_1 . \tag{31}$$

□

Theorem 4 shows us that it is possible to stably recover sparse signals from the noisy measurements. Compared to the result of Theorem 1.9 in [23] with $\|\widehat{\mathbf{x}} - \mathbf{x}\|_2 \le C_0(\sigma(\mathbf{x})_1/\sqrt{K}) + C_2\epsilon$, where $C_0 = 2((1-(1-\sqrt{2})\delta_{2K})/(1-(1+\sqrt{2})\delta_{2K}))$ and $C_2 = 4(\sqrt{1+\delta_{2K}}/(1-(1+\sqrt{2})\delta_{2K}))$, we may achieve a smaller restoration error by using the proposed sparse signal recovery model when appropriate values of λ and τ are utilized.

4. Realization Algorithm

For simplicity, in this paper we resort to the simple computational approach of conjugate gradient (CG) algorithm for computing the local minimizer of problem (7). The CG algorithm is an effective approach to solve the linear system equation $\mathbf{y} = \mathbf{A}\mathbf{x}$ from the measurements $\mathbf{y}$ for the given measurement matrix $\mathbf{A}$. It is straightforward to generalize the proposed model with other advanced algorithms.

Let us denote the initial guess of $\mathbf{x}$ as $\mathbf{x_0}$ (without loss of generality, we assume $\mathbf{x_0} = \mathbf{0}$ in this paper). The pseudocode of the CG algorithm is illustrated as follows.

The Conjugate Gradient Algorithm

(1) Initialization: Initial point λ, τ, $\mathbf{x}_n$, $\mathbf{g}_n = \nabla F(\mathbf{x}_n) = \mathbf{A}^T(\mathbf{y} - \mathbf{A}\mathbf{x}_n) + \lambda \cdot \text{sign}(\mathbf{x}_n) - \tau$, iterations $n = 0$, search direction $\mathbf{d}_n = -\mathbf{g}_n$;

(2) Loop until the specified accuracy or iteration numbers:

(3) Use the line search technique to determine the search step length η_n;

(4) Update the signal $\mathbf{x}_{n+1} = \mathbf{x}_n + \eta_n\mathbf{d}_n$;

(5) Update the gradient $\mathbf{g}_{n+1} = \mathbf{A}^T(\mathbf{y} - \mathbf{A}\mathbf{x}_{n+1}) + \lambda \cdot \text{sign}(\mathbf{x}_{n+1}) - \tau$;

(6) Update $\theta_n = \mathbf{g}_{n+1}^T(\mathbf{g}_{n+1} - \mathbf{g}_n)/\mathbf{g}_n^k\mathbf{g}_n$;

(7) Update the search direction $\mathbf{d}_{n+1} = -\mathbf{g}_{n+1} + \theta_n\mathbf{g}_n$;

(8) Iteration number increases by 1 as $n = n + 1$.

Here $\text{sign}(\mathbf{x}_n)$ denotes the element-wise sign function of the vector $\mathbf{x}_n$, which is defined as follows: for every element $\mathbf{x}_{n1}, \ldots, \mathbf{x}_{ni}, \ldots, \mathbf{x}_{nN}$, $i = 1, 2, \ldots, N$, if $\mathbf{x}_{ni} > 0$, $\text{sign}(\mathbf{x}_{ni}) = 1$; if $\mathbf{x}_{ni} = 0$, $\text{sign}(\mathbf{x}_{ni}) = 0$; and if $\mathbf{x}_{ni} < 0$, $\text{sign}(\mathbf{x}_{ni}) = -1$.

5. Numerical Results

Simulations are performed to testify the applicability of the proposed signal reconstruction model in different settings. Some representative sparse signal restoration algorithms, such as (1) the greedy algorithm of the OMP [4], SP [6], and MP [24] for the P_0 problem; (2) the ℓ_1-norm optimization of GPSR-Basic and GPSR-BB [8], LMS (Least Mean Square) [25], ℓ_1-*ls* [26], IRLS (Iteratively Reweighted Least Square) [27], and LARS (Least Angle Regression Stagewise) [28]; (3) the Bayesian framework of BCS (Bayesian Compressed Sensing) [14], are included in our simulations. For fair comparison, we considered the commonly assumed experimental setup in the literatures. All of our simulations are performed in MATLAB 2009 environment over Microsoft Window 7 operating system; the simulation platform is provisioned with Inter Core i5-480, ATI HD 5470 processor with 2 GB of memory. Gaussian measurement matrix $\mathbf{A}$ is assumed in all simulations for simplicity, which satisfies the RIP conditions with high probability. The entries of $\mathbf{A}$ are independently generated from normal distributed random variables with zero mean and variance of $1/M$. In all experiments, we consider the fixed signal length of $N = 1000$, while the number of measurements M and sparsity K may change in different experiment settings. The locations of K nonzero coefficients of sparse signal are assumed to be uniformly distributed within $[1, N]$. The corresponding nonzero coefficients are also generated from a Gaussian distributed random variable. Normalized reconstruction error $\|\widehat{\mathbf{x}} - \mathbf{x}\|_2/\|\mathbf{x}\|_2$ will be utilized as a metric to evaluate the reconstruction quality, where $\widehat{\mathbf{x}}$ and $\mathbf{x}$ represent the reconstructed and original signal vectors, respectively. For each experiment, 1000 simulations are conducted to calculate the normalized reconstruction error in different algorithms.

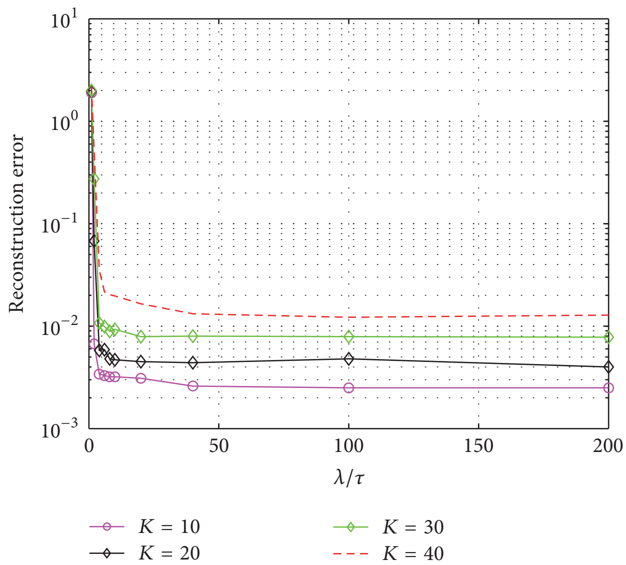

FIGURE 2: The reconstruction error as a logarithmic function of λ/τ.

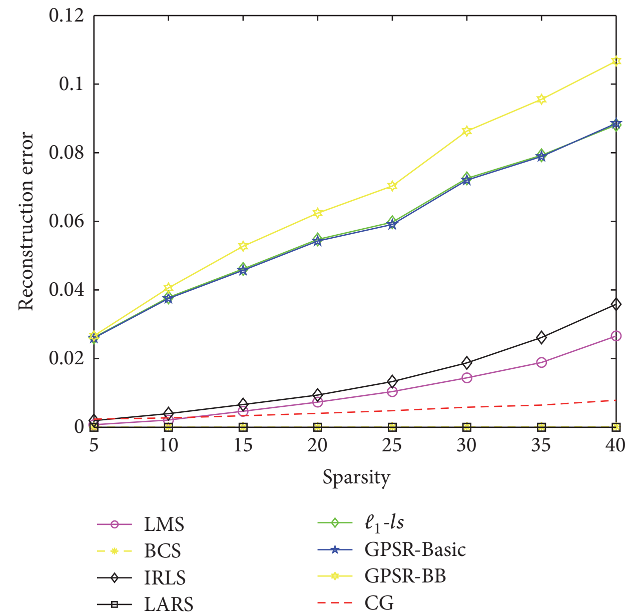

FIGURE 3: The reconstruction error with different sparsity degrees, $M = 200$, $N = 1000$.

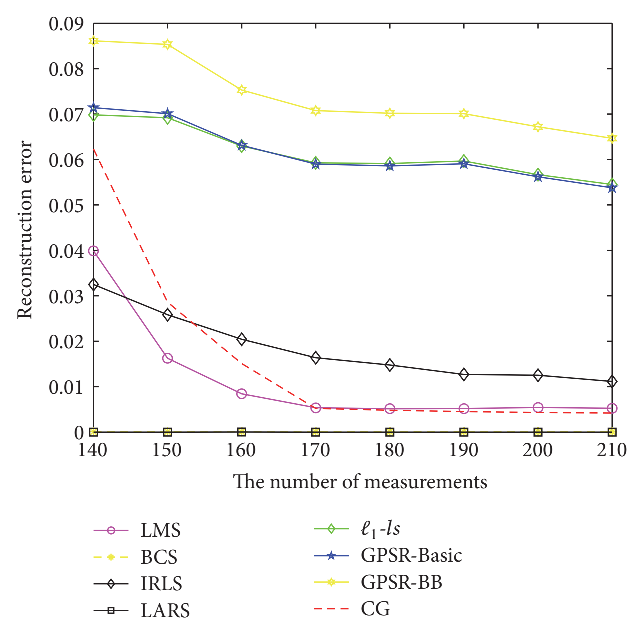

FIGURE 4: The reconstruction error with different number of measurements, $K = 20$, $N = 1000$.

5.1. The Influence of Parameters of λ and τ. The first problem we encounter is the choice of λ and τ. In this experiment, we fix $M = 200$ and the sparsity degree parameter $K = 10, 20, 30$, and 40, respectively. We use the simple CG algorithm to solve our proposed model. Figure 2 shows the normalized reconstruction error as a logarithmic function of the ratio λ/τ; each point in the curve is the outcome of 100 iterations of the CG algorithm. Figure 2 can help us determine the appropriate λ/τ; we can find that when $\lambda/\tau \leq 1$, the reconstruction results are terrible. A larger ratio leads to a more satisfying solution. And from the four curves, when $\lambda/\tau \geq 10$, the reconstruction performance tends to be stable, so we select $\lambda/\tau = 10$ in the following experiments.

5.2. Noise-Free Measurement Case. We start with the noise-free measurement case as $\mathbf{y} = \mathbf{A}\mathbf{x}$.

Experiment 1 (effect of sparsity on the performance). This experiment explores the answer to such a question on how the sparsity degree of the signal will affect the reconstruction quality with a given number of measurements. In this experiment, we fix $M = 200$, while the sparsity degree parameter K will change from 5 to 40. We assume the simple CG algorithm to solve our proposed model. One may readily observe from Figure 3 that the proposed scheme (labeled with CG) far exceed GPSR-Basic, GPSR-BB, LMS, ℓ_1-*ls*, and IRLS. With the increase of sparsity, there will be larger reconstruction error for all algorithms; however, the proposed reconstruction scheme can still achieve a comparatively low reconstruction error, which complies with our previous analysis. It can be observed that BCS and LARS algorithms are always better than our proposed scheme. Our scheme is comparable to ℓ_1-norm relaxation techniques but is inferior to the inference method.

Experiment 2 (effect of measurement number on the performance). This experiment tries to answer the question on how different number of measurements will affect the achieved reconstruction performance. In the simulations, we fix $K = 20$, while the number of measurements M changes from 140 to 210. The signal reconstruction errors are illustrated in Figure 4. We may observe similar results in the first experiment. BCS and LARS algorithms always achieve the best performance. The proposed method is comparable to the LMS algorithm, both of which are superior to other algorithms.

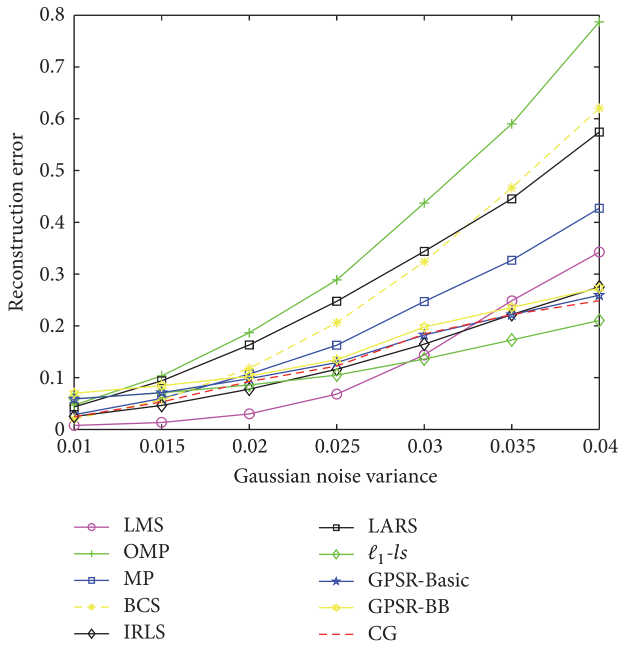

Figure 5: The reconstruction error with different Gaussian noise variance, $K = 20$, $M = 200$, $N = 1000$.

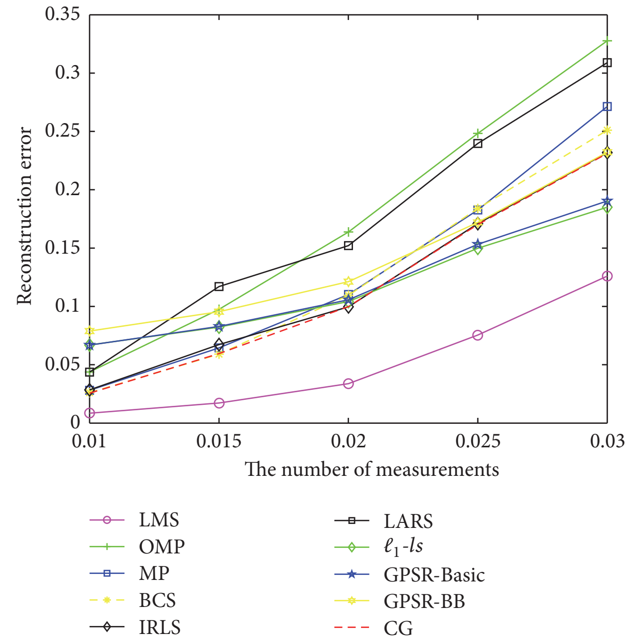

Figure 6: The reconstruction error with measurement matrix $\mathbf{A}$ disturbed by different Gaussian noise variance, $K = 20$, $M = 200$, $N = 1000$.

5.3. Noisy Measurement Case

Experiment 3 (robustness against noise). In this experiment, we testify the noisy sparse signal reconstruction performance by fixing $K = 20$, $M = 200$. The noisy measurement model is given by $\mathbf{y} = \mathbf{Ax} + \mathbf{e}$, where $\mathbf{e}$ is an additive white Gaussian noise with covariance matrix $\sigma^2 \mathbf{I}_M$ ($\mathbf{I}_M$ is $M \times M$ identity matrix). From Figure 5, we may observe that the proposed reconstruction algorithm can achieve superior restoration quality compared to greedy approaches, such as OMP and MP, in terms of the reconstruction errors. The performance of BCS and LARS algorithms is satisfactory in the first two experiments, but their antinoise performance is no match to our scheme. In addition, our scheme is slightly inferior to ℓ_1-*ls*. Obviously, with the increase of noise variance, our scheme has smaller reconstructed error than LMS which shows the good antinoise performance.

Experiment 4 (robustness against noisy measurement matrix $\mathbf{A}$). In this experiment, we testify the stability of sparse signal reconstruction performance when the measurement matrix $\mathbf{A}$ is disturbed by Gaussian noise with fixed $K = 20$, $M = 200$. It is assumed that the measurement matrix $\mathbf{A}$ is contaminated by zero-mean Gaussian noise with different variances at the receiver. From Figure 6, we can observe that the proposed reconstruction model exhibits reasonable ability against the noisy measurement matrix except for LMS algorithm. With the increase of noise, our antinoise ability is worse compared to ℓ_1-*ls* and GPSR-Basic algorithms; however, when the noise is small, we can show competitive power. Our proposed scheme outperforms OMP, MP, BCS, LARS, and GPSR-BB.

By comparing the simulations results in Figures 3–6, we may observe that the proposed scheme outperforms the approaches of GPSR-BB, GPSR-Basic, and LRLS in terms of the achieved sparse signal reconstruction performance under different signal sparsity, different measurements, and robustness against background noise as well as the contamination of the measurement matrix. Although BCS and LARS algorithms exhibit better performance against different sparsity K and number of measurements M, it is sensitive to noise. LMS algorithm shows stable antinoise performance but is inferior to our proposed model with different sparsity K and number of measurements M. Greedy algorithms can achieve satisfactory performance given different signal sparsity and different measurements, but our proposed model exhibits better robustness against noise. Meanwhile, our proposed model outperforms most optimization based algorithms. In a word, the proposed reconstruction model provides an effective choice in the sparse signal reconstruction.

6. Concluding Remarks

In this paper, motivated by the efforts in relaxing ℓ_0-norm function with tractable norm functions to obtain either convex or nonconvex problem, we developed a new ℓ_1- and ℓ_2-norm joint regularization reconstruction framework to approach the original ℓ_0-norm sparseness-inducing constraint reconstruction problem. Our analysis shows that the proposed sparse signal reconstruction model can derive the same solution of the original sparse problem with ℓ_0-norm regularization. Moreover, the upper error bound of the proposed sparse signal recovery model is derived. It is shown that the proposed signal reconstruction model can achieve an interesting tradeoff between both ℓ_1-norm relaxation and ℓ_p-norm relaxation techniques. For instance, it exhibits the similar ℓ_0-norm approximation capability like ℓ_p-norm relaxation, while, on the other hand, straightforward and simple algorithms (such as the CG algorithm considered in this paper) can be utilized to derive the solution, just like ℓ_1-norm convex relaxation approaches. Numerical analysis is presented to validate that the proposed sparse signal reconstruction framework provides the exciting and satisfying method for the sparse signal reconstruction problem. Nextly, we will find other effective algorithm to solve this new model for better performance and popularize the proposed new norm model to image processing, such as image restoration in [16].

Appendix

Proof of Lemma 3. Firstly, we introduce Lemma A.1.

Lemma A.1 (see [23]). *Let Λ_0 be an arbitrary subset of $\{1,2,\ldots,N\}$ such that $|\Lambda_0|\le K$. For any vector $\mathbf{u}\in\mathbb{R}^N$, define Λ_1 as the index set corresponding to K largest entries of $\mathbf{u}_{\Lambda_0^C}$ (in absolute value), Λ_2 as the index set corresponding to the next K largest entries, and so on. Then*

$$\sum_{i\ge2}\left\|\mathbf{u}_{\Lambda_i}\right\|_2\le\frac{\left\|\mathbf{u}_{\Lambda_0^C}\right\|_1}{\sqrt{K}}. \tag{A.1}$$

The proof is shown in [23]. Lemma A.1 immediately yields

$$\|\mathbf{h}_{\Lambda^C}\|_2=\left\|\sum_{i\ge2}\mathbf{h}_{\Lambda_i}\right\|_2\le\sum_{i\ge2}\left\|\mathbf{h}_{\Lambda_i}\right\|_2\le\frac{\left\|\mathbf{h}_{\Lambda_0^C}\right\|_1}{\sqrt{K}}, \tag{A.2}$$

where Λ_i are defined as in Lemma A.1; that is, Λ_1 is the index set corresponding to K largest entries of $\mathbf{h}_{\Lambda_0^C}$ (in absolute value), Λ_2 is the index set corresponding to the next K largest entries, and so on.

Since $\lambda\|\widehat{\mathbf{x}}\|_1-\tau\|\widehat{\mathbf{x}}\|_2\le\lambda\|\mathbf{x}\|_1-\tau\|\mathbf{x}\|_2$, and $\widehat{\mathbf{x}}=\mathbf{x}+\mathbf{h}$, we have $\lambda\|\mathbf{x}\|_1\ge\lambda\|\mathbf{x}+\mathbf{h}\|_1+\tau(\|\mathbf{x}\|_2-\|\mathbf{x}+\mathbf{h}\|_2)$. Because of the triangle inequality property of norm, we have $\|\mathbf{x}+\mathbf{h}\|_2\le\|\mathbf{x}\|_2+\|\mathbf{h}\|_2$; then $\lambda\|\mathbf{x}\|_1\ge\lambda\|\mathbf{x}+\mathbf{h}\|_1-\tau\|\mathbf{h}\|_2$; we can have

$$\begin{aligned}\lambda\|\mathbf{x}\|_1&=\lambda\left(\left\|\mathbf{x}_{\Lambda_0}\right\|_1+\left\|\mathbf{x}_{\Lambda_0^C}\right\|_1\right)\\&\ge\lambda\left(\left\|\mathbf{x}_{\Lambda_0}+\mathbf{h}_{\Lambda_0}\right\|_1+\left\|\mathbf{x}_{\Lambda_0^C}+\mathbf{h}_{\Lambda_0^C}\right\|_1\right)-\tau\|\mathbf{h}\|_2\\&\ge\lambda\left(\left\|\mathbf{x}_{\Lambda_0}\right\|_1-\left\|\mathbf{h}_{\Lambda_0}\right\|_1+\left\|\mathbf{h}_{\Lambda_0^C}\right\|_1-\left\|\mathbf{x}_{\Lambda_0^C}\right\|_1\right)\\&\quad-\tau\|\mathbf{h}\|_2.\end{aligned} \tag{A.3}$$

Let $\sigma(\mathbf{x})_1=\|\mathbf{x}_{\Lambda_0^C}\|_1$; it results in

$$\left\|\mathbf{h}_{\Lambda_0^C}\right\|_1\le2\sigma(\mathbf{x})_1+\left\|\mathbf{h}_{\Lambda_0}\right\|_1+\frac{\tau}{\lambda}\|\mathbf{h}\|_2. \tag{A.4}$$

Combining this with (A.2), we can obtain

$$\begin{aligned}\|\mathbf{h}_{\Lambda^C}\|_2&\le\frac{\left\|\mathbf{h}_{\Lambda_0^C}\right\|_1}{\sqrt{K}}\le\frac{2\sigma(\mathbf{x})_1}{\sqrt{K}}+\frac{\left\|\mathbf{h}_{\Lambda_0}\right\|_1}{\sqrt{K}}+\frac{\tau}{\lambda\sqrt{K}}\|\mathbf{h}\|_2\\&\le\frac{2\sigma(\mathbf{x})_1}{\sqrt{K}}+\left\|\mathbf{h}_{\Lambda_0}\right\|_2+\frac{\tau}{\lambda\sqrt{K}}\|\mathbf{h}\|_2,\end{aligned} \tag{A.5}$$

where we have used the fact that, for any K-sparse signal $\mathbf{u}$, there is $\|\mathbf{u}\|_1/\sqrt{K}\le\|\mathbf{u}\|_2$.

Because $\mathbf{h}=\mathbf{h}_\Lambda+\mathbf{h}_{\Lambda^C}$, from the triangle inequality, we get

$$\|\mathbf{h}\|_2\le\left\|\mathbf{h}_\Lambda\right\|_2+\left\|\mathbf{h}_{\Lambda^C}\right\|_2. \tag{A.6}$$

Combining this with (A.5) and observing that $\|\mathbf{h}_{\Lambda_0}\|_2\le\|\mathbf{h}_\Lambda\|_2$ will yield

$$\begin{aligned}\|\mathbf{h}\|_2&\le\left\|\mathbf{h}_\Lambda\right\|_2+\left\|\mathbf{h}_{\Lambda^C}\right\|_2\\&\le2\left\|\mathbf{h}_\Lambda\right\|_2+\frac{2\sigma(\mathbf{x})_1}{\sqrt{K}}+\frac{\tau}{\lambda\sqrt{K}}\|\mathbf{h}\|_2.\end{aligned} \tag{A.7}$$

Because $1-\tau/\lambda\sqrt{K}>0$, let $a=1-\tau/\lambda\sqrt{K}$; we obtain

$$\|\mathbf{h}\|_2\le\frac{2}{a}\left\|\mathbf{h}_\Lambda\right\|_2+\frac{2\sigma(\mathbf{x})_1}{a\sqrt{K}}. \tag{A.8}$$

What we need to do is to establish a bound for $\|\mathbf{h}_\Lambda\|_2$. Consider

$$\begin{aligned}&\left\|\mathbf{h}_{\Lambda_0}\right\|_2^2+\left\|\mathbf{h}_{\Lambda_1}\right\|_2^2=\left\|\mathbf{h}_{\Lambda_0}+\mathbf{h}_{\Lambda_1}\right\|_2^2\\&\quad\le\frac{1}{1-\delta_{2K}}\left\|\mathbf{A}\left(\mathbf{h}_{\Lambda_0}+\mathbf{h}_{\Lambda_1}\right)\right\|_2^2\\&\quad=\frac{1}{1-\delta_{2K}}\left\langle\mathbf{A}\left(\mathbf{h}-\mathbf{h}_{\Lambda_2}-\cdots\right),\mathbf{A}\left(\mathbf{h}_{\Lambda_0}+\mathbf{h}_{\Lambda_1}\right)\right\rangle\\&\quad=\frac{1}{1-\delta_{2K}}\left\langle\mathbf{Ah},\mathbf{A}\left(\mathbf{h}_{\Lambda_0}+\mathbf{h}_{\Lambda_1}\right)\right\rangle+\frac{1}{1-\delta_{2K}}\\&\quad\cdot\sum_{t\ge2}\left[\left\langle\mathbf{A}\left(-\mathbf{h}_{\Lambda_t}\right),\mathbf{Ah}_{\Lambda_0}\right\rangle+\left\langle\mathbf{A}\left(-\mathbf{h}_{\Lambda_t}\right),\mathbf{Ah}_{\Lambda_1}\right\rangle\right],\end{aligned} \tag{A.9}$$

where $\langle\cdot,\cdot\rangle$ denotes the inner product. We can renormalize the vectors $-\mathbf{h}_{\Lambda_t}$ and $\mathbf{h}_{\Lambda_0}$ so that their ℓ_2-norms equal one by setting $\mathbf{u}_t=-\mathbf{h}_{\Lambda_t}/\|\mathbf{h}_{\Lambda_t}\|_2$ and $\mathbf{u}_0=\mathbf{h}_{\Lambda_0}/\|\mathbf{h}_{\Lambda_0}\|_2$. We can obtain

$$\begin{aligned}&\frac{\left\langle\mathbf{A}\left(-\mathbf{h}_{\Lambda_t}\right),\mathbf{Ah}_{\Lambda_0}\right\rangle}{\left\|\mathbf{h}_{\Lambda_t}\right\|_2\left\|\mathbf{h}_{\Lambda_0}\right\|_2}=\left\langle\mathbf{Au}_t,\mathbf{Au}_0\right\rangle\\&\quad=\frac{1}{4}\left[\left\|\mathbf{A}\left(\mathbf{u}_t+\mathbf{u}_0\right)\right\|_2^2-\left\|\mathbf{A}\left(\mathbf{u}_t-\mathbf{u}_0\right)\right\|_2^2\right]\\&\quad\le\frac{1}{4}\left[\left(1+\delta_{2K}\right)\left\|\mathbf{u}_t+\mathbf{u}_0\right\|_2^2-\left(1-\delta_{2K}\right)\left\|\mathbf{u}_t-\mathbf{u}_0\right\|_2^2\right]\\&\quad\le\delta_{2K}.\end{aligned} \tag{A.10}$$

Substituting Λ_0 by Λ_1, we can obtain similar results:

$$\frac{\left\langle\mathbf{A}\left(-\mathbf{h}_{\Lambda_t}\right),\mathbf{Ah}_{\Lambda_1}\right\rangle}{\left\|\mathbf{h}_{\Lambda_t}\right\|_2\left\|\mathbf{h}_{\Lambda_1}\right\|_2}=\left\langle\mathbf{Au}_t,\mathbf{Au}_1\right\rangle\le\delta_{2K}. \tag{A.11}$$

Then

$$\begin{aligned}&\left\langle\mathbf{A}\left(-\mathbf{h}_{\Lambda_t}\right),\mathbf{Ah}_{\Lambda_0}\right\rangle+\left\langle\mathbf{A}\left(-\mathbf{h}_{\Lambda_t}\right),\mathbf{Ah}_{\Lambda_1}\right\rangle\\&\quad\le\delta_{2K}\left\|\mathbf{h}_{\Lambda_t}\right\|_2\left(\left\|\mathbf{h}_{\Lambda_0}\right\|_2+\left\|\mathbf{h}_{\Lambda_1}\right\|_2\right).\end{aligned} \tag{A.12}$$

Because

$$\begin{aligned}\left\langle \mathbf{Ah}, \mathbf{A}\left(\mathbf{h}_{\Lambda_0}+\mathbf{h}_{\Lambda_1}\right)\right\rangle &\le \|\mathbf{Ah}\|_2 \left\|\mathbf{A}\left(\mathbf{h}_{\Lambda_0}+\mathbf{h}_{\Lambda_1}\right)\right\|_2 \\ &\le \sqrt{1+\delta_{2K}}\,\|\mathbf{Ah}\|_2 \left\|\mathbf{h}_{\Lambda_0}+\mathbf{h}_{\Lambda_1}\right\|_2 \\ &\le \sqrt{1+\delta_{2K}}\,\|\mathbf{Ah}\|_2 \left[\left\|\mathbf{h}_{\Lambda_0}\right\|_2+\left\|\mathbf{h}_{\Lambda_1}\right\|_2\right],\end{aligned} \tag{A.13}$$

substituting (A.12) and (A.13) into (A.9) will yield

$$\begin{aligned}&\left\|\mathbf{h}_{\Lambda_0}\right\|_2^2+\left\|\mathbf{h}_{\Lambda_1}\right\|_2^2 \\ &\quad\le \left(\frac{\sqrt{1+\delta_{2K}}}{1-\delta_{2K}}\|\mathbf{Ah}\|_2+\frac{\delta_{2K}\sum_{t\ge 2}\left\|\mathbf{h}_{\Lambda_t}\right\|_2}{1-\delta_{2K}}\right) \\ &\quad\cdot\left(\left\|\mathbf{h}_{\Lambda_0}\right\|_2+\left\|\mathbf{h}_{\Lambda_1}\right\|_2\right).\end{aligned} \tag{A.14}$$

Let $c=(\sqrt{1+\delta_{2K}}/(1-\delta_{2K}))\|\mathbf{Ah}\|_2$, $d=\delta_{2K}/(1-\delta_{2K})$ and $e=\sum_{t\ge2}\|\mathbf{h}_{\Lambda_t}\|_2$; the above inequality implies

$$\begin{aligned}&\left(\left\|\mathbf{h}_{\Lambda_0}\right\|_2-\frac{c+de}{2}\right)^2+\left(\left\|\mathbf{h}_{\Lambda_1}\right\|_2-\frac{c+de}{2}\right)^2 \\ &\quad\le\frac{(c+de)^2}{2},\end{aligned} \tag{A.15}$$

This inequality leads to the following two results:

$$\left\|\mathbf{h}_{\Lambda_0}\right\|_2 \le \frac{c+de}{2}+\frac{c+de}{\sqrt{2}}=\frac{1+\sqrt{2}}{2}\cdot(c+de), \tag{A.16}$$

$$\left\|\mathbf{h}_{\Lambda_1}\right\|_2 \le \frac{c+de}{2}+\frac{c+de}{\sqrt{2}}=\frac{1+\sqrt{2}}{2}\cdot(c+de). \tag{A.17}$$

So combining (A.2) and (A.4), the fact that $\|\mathbf{h}_{\Lambda_0}\|_1/\sqrt{K}\le\|\mathbf{h}_{\Lambda_0}\|_2$ and $\|\mathbf{h}_{\Lambda_0}\|_2\le\|\mathbf{h}_{\Lambda}\|_2$ will yield

$$\begin{aligned}\left\|\mathbf{h}_{\Lambda}\right\|_2 &\le \left\|\mathbf{h}_{\Lambda_0}\right\|_2+\left\|\mathbf{h}_{\Lambda_1}\right\|_2 \le \left(1+\sqrt{2}\right)\cdot(c+de) \\ &= \left(1+\sqrt{2}\right)\left(\frac{\sqrt{1+\delta_{2K}}}{1-\delta_{2K}}\|\mathbf{Ah}\|_2\right. \\ &\left.+\frac{\delta_{2K}}{1-\delta_{2K}}\sum_{t\ge2}\left\|\mathbf{h}_{\Lambda_t}\right\|_2\right)\le\left(1+\sqrt{2}\right) \\ &\cdot\left(\frac{\sqrt{1+\delta_{2K}}}{1-\delta_{2K}}\|\mathbf{Ah}\|_2+\frac{\delta_{2K}}{1-\delta_{2K}}\frac{\left\|\mathbf{h}_{\Lambda_0^C}\right\|_1}{\sqrt{K}}\right) \\ &\le\frac{\left(1+\sqrt{2}\right)\sqrt{1+\delta_{2K}}}{1-\delta_{2K}}\|\mathbf{Ah}\|_2+\frac{\left(1+\sqrt{2}\right)\delta_{2K}}{1-\delta_{2K}} \\ &\cdot\frac{2\sigma(\mathbf{x})_1+\left\|\mathbf{h}_{\Lambda_0}\right\|_1+(\tau/\lambda)\|\mathbf{h}\|_2}{\sqrt{K}} \\ &\le\frac{\left(1+\sqrt{2}\right)\sqrt{1+\delta_{2K}}}{1-\delta_{2K}}\|\mathbf{Ah}\|_2 \\ &+\frac{\left(1+\sqrt{2}\right)\delta_{2K}}{1-\delta_{2K}}\left(\frac{2\sigma(\mathbf{x})_1}{\sqrt{K}}+\left\|\mathbf{h}_{\Lambda_0}\right\|_2\right. \\ &\left.+\frac{\tau}{\lambda\sqrt{K}}\|\mathbf{h}\|_2\right)\le\frac{\left(1+\sqrt{2}\right)\sqrt{1+\delta_{2K}}}{1-\delta_{2K}}\|\mathbf{Ah}\|_2 \\ &+\frac{\left(1+\sqrt{2}\right)\delta_{2K}}{1-\delta_{2K}}\left(\frac{2\sigma(\mathbf{x})_1}{\sqrt{K}}+\left\|\mathbf{h}_{\Lambda}\right\|_2\right. \\ &\left.+\frac{\tau}{\lambda\sqrt{K}}\|\mathbf{h}\|_2\right).\end{aligned} \tag{A.18}$$

From (A.18), we get

$$\begin{aligned}&\left(1-\frac{\left(1+\sqrt{2}\right)\delta_{2K}}{1-\delta_{2K}}\right)\left\|\mathbf{h}_{\Lambda}\right\|_2 \\ &\quad\le\frac{\left(1+\sqrt{2}\right)\sqrt{1+\delta_{2K}}}{1-\delta_{2K}}\|\mathbf{Ah}\|_2 \\ &\quad+\frac{\left(1+\sqrt{2}\right)\delta_{2K}}{1-\delta_{2K}}\frac{2\sigma(\mathbf{x})_1}{\sqrt{K}} \\ &\quad+\frac{\left(1+\sqrt{2}\right)\delta_{2K}\tau}{\left(1-\delta_{2K}\right)\lambda\sqrt{K}}\|\mathbf{h}\|_2.\end{aligned} \tag{A.19}$$

Combining this with (A.8) and because $\delta_{2K}<\min\{(2-\sqrt{2})/2,(\lambda\sqrt{K}-(2+\sqrt{2})\lambda\delta_{2K}\sqrt{K}-\tau)/\sqrt{2}\tau\}$, we can obtain

$$\|\mathbf{h}\|_2\le\alpha\|\mathbf{Ah}\|_2+\beta\sigma(\mathbf{x})_1, \tag{A.20}$$

where $\alpha=2(1+\sqrt{2})\sqrt{1+\delta_{2K}}\lambda\sqrt{K}/(\lambda\sqrt{K}-(2+\sqrt{2})\delta_{2K}\lambda\sqrt{K}-(1+\sqrt{2}\delta_{2K})\tau)$ and $\beta=(2+2\sqrt{2}\delta_{2K})\lambda/(\lambda\sqrt{K}-(2+\sqrt{2})\delta_{2K}\lambda\sqrt{K}-(1+\sqrt{2}\delta_{2K})\tau)$. □

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This work was supported by the NSFC under Grants no. 61271246 and no. 61371165.

References

[1] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[2] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[3] R. G. Baraniuk, "Compressive sensing," *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 118–124, 2007.

[4] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[5] D. L. Donoho, Y. Tsaig, and J. L. Starck, "Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit," Tech. Rep., Stanford University, 2006.

[6] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing: closing the gap between performance and complexity," *IEEE Transactions on Information Theory*, vol. 55, no. 5, pp. 2230–2249, 2009.

[7] E. van den Berg and M. P. Friedlander, "In pursuit of a root," Tech. Rep. TR-2007-19, Department of Computer Science, University of British Columbia, 2007.

[8] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems," *IEEE Journal on Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 586–597, 2007.

[9] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, 2008.

[10] D. Wipf and S. Nagarajan, "Iterative reweighted ℓ_1 and ℓ_2 methods for finding sparse solution," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 6, pp. 317–329, 2010.

[11] E. J. Candès and J. Romberg, "l1-Magic: a collection of MATLAB routines for solving the convex optimization programs central to compressive," 2006, http://www.acm.caltech.edu/ l1magic/.

[12] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Processing Letters*, vol. 14, no. 10, pp. 707–710, 2007.

[13] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2346–2356, 2008.

[14] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Bayesian compressive sensing using Laplace priors," *IEEE Transactions on Image Processing*, vol. 19, no. 1, pp. 53–63, 2010.

[15] X. Lu, Y. Wang, and Y. Yuan, "Sparse coding from a Bayesian perspective," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 6, pp. 929–939, 2013.

[16] J. Liu, T.-Z. Huang, I. W. Selesnick, X.-G. Lv, and P.-Y. Chen, "Image restoration using total variation with overlapping group sparsity," *Information Sciences*, vol. 295, pp. 232–246, 2015.

[17] G. Liu, T.-Z. Huang, and J. Liu, "High-order TVL1-based images restoration and spatially adapted regularization parameter selection," *Computers and Mathematics with Applications*, vol. 67, no. 10, pp. 2015–2026, 2014.

[18] M.-G. Shama, T.-Z. Huang, J. Liu, and S. Wang, "A convex total generalized variation regularized model for multiplicative noise and blur removal," *Applied Mathematics and Computation*, vol. 276, pp. 109–121, 2016.

[19] X.-L. Zhao, F. Wang, T.-Z. Huang, M. K. Ng, and R. J. Plemmons, "Deblurring and sparse unmixing for hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 7, pp. 4045–4058, 2013.

[20] Y. Cai, M. Donatelli, D. Bianchi, and T.-Z. Huang, "Regularization preconditioners for frame-based image deblurring with reduced boundary artifacts," *SIAM Journal on Scientific Computing*, vol. 38, no. 1, pp. B164–B189, 2016.

[21] Y. Lou, P. Yin, Q. He, and J. Xin, "Computing sparse representation in a highly coherent dictionary based on difference of L_1 and L_2," *Journal of Scientific Computing*, vol. 64, no. 1, pp. 178–196, 2015.

[22] M. Elad, *Sparse and Redundant Representations*, Springer, New York, NY, USA, 2010.

[23] M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok, "Introduction to compressed sensing," in *Compressed Sensing Theory and Applications*, chapter 1, pp. 1–64, Cambridge University Press, Cambridge, UK, 2012.

[24] G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations," *Constructive Approximation*, vol. 13, no. 1, pp. 57–98, 1997.

[25] J. Jin, Y. Gu, and S. Mei, "A stochastic gradient approach on compressive sensing signal reconstruction based on adaptive filtering framework," *IEEE Journal on Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 409–420, 2010.

[26] K. Koh, S. Kim, and S. Boyd, "A method for large-scale ℓ_1-regularized logistic regression," in *Proceedings of the National Conference on Artificial Intelligence*, pp. 565–571, 2007.

[27] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '08)*, pp. 3869–3872, Las Vegas, Nev, USA, April 2008.

[28] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *The Annals of Statistics*, vol. 32, no. 2, pp. 407–499, 2004.