Accelerated Schemes for the L_1/L_2 Minimization

Chao Wang ⓘ, Ming Yan ⓘ, Yaghoub Rahimi, and Yifei Lou ⓘ

Abstract—In this paper, we consider the L_1/L_2 minimization for sparse recovery and study its relationship with the L_1-αL_2 model. Based on this relationship, we propose three numerical algorithms to minimize this ratio model, two of which work as adaptive schemes and greatly reduce the computation time. Focusing on the two adaptive schemes, we discuss their connection to existing approaches and analyze their convergence. The experimental results demonstrate that the proposed algorithms are comparable to state-of-the-art methods in sparse recovery and work particularly well when the ground-truth signal has a high dynamic range. Lastly, we reveal some empirical evidence on the exact L_1 recovery under various combinations of sparsity, coherence, and dynamic ranges, which calls for theoretical justification in the future.

Index Terms—Sparsity, L_0, adaptive scheme, dynamic range.

I. INTRODUCTION

IN VARIOUS science and engineering applications, one aims to seek for a low-dimensional representation from high-dimensional data, and sparsity is a crucial assumption. For example, it is reasonable to assume in machine learning [1] that only a few features correspond to the response. In image processing [2], the restored images are often piecewise constant, which means that gradients are sparse. In non-negative matrix factorization [3], the low-rank decomposition enforces sparsity with respect to singular values.

Sparse signal recovery is to find the sparsest solution of $A\mathbf{x} = \mathbf{b}$ where $A \in \mathbb{R}^{m \times n}$ ($m \ll n$), $\mathbf{x} \in \mathbb{R}^n$, and $\mathbf{b} \in \mathbb{R}^m$. We assume that A has a full row rank and $\mathbf{b}$ is nonzero. This problem is often referred to as *compressed sensing* (CS) [4], [5] in the sense that the sparse signal $\mathbf{x}$ is compressible. Mathematically, it can be formulated by the L_0 minimization,

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{x}\|_0 \quad \text{s.t.} \quad A\mathbf{x} = \mathbf{b}. \tag{1}$$

Unfortunately, the L_0 problem is known to be NP-hard [6]. Various approaches in sparse recovery have been investigated. Some

greedy methods include orthogonal matching pursuit (OMP) [7], orthogonal least squares (OLS) [8], and compressive sampling matching pursuit (CoSaMp) [9]. However, these greedy methods often lack of accuracy when n is large. Alternatively, approximations/relaxation approaches to the L_0 norm have been sought. For example, convex relaxation, referred to as *basis pursuit* (BP) [10], replaces L_0 in (1) with the L_1 norm. Recently, nonconvex models attract considerate amount of attentions due to their sharper approximations of L_0 compared to the L_1 norm. Some popular nonconvex models include L_p [11]–[13], L_1-L_2 [14], [15], transformed L_1 (TL1) [16]–[18], nonnegative garrote [19], and capped-L_1 [20]–[22]. Except for L_1-L_2, all of these nonconvex models involve one parameter to be determined and adjusted for different types of sparse recovery problems.

In this paper, we study the ratio of L_1 and L_2 as a scale-invariant and parameter-free metric to approximate the desired scale-invariant L_0 norm. The ratio of L_1 and L_2 can be traced back to [23] as a sparsity measure, and its scale-invariant property was explicitly mentioned in [24].

Esser *et al.* [14], [25] focused on nonnegative signals and established the equivalence between L_1/L_2 and L_0. The ratio model was later formulated as a nonlinear constraint that was solved by a lifted approach [26], [27]. Some applications of L_1/L_2 include blind deconvolution [28], [29] and sparse filtering [30], [31].

In our earlier work [32], we focused on a constrained minimization problem,

$$\min_{\mathbf{x} \in \mathbb{R}^n} \frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_2} \quad \text{s.t.} \quad A\mathbf{x} = \mathbf{b}. \tag{2}$$

Theoretically, we proved that any s-sparse vector is a local minimizer of the L_1/L_2 model provided with a strong null space property (sNSP) condition. Computationally, we considered to minimize (2) via the alternating direction method of multipliers (ADMM) [33]. In particular, we introduced two auxiliary variables and formed the augmented Lagrangian as

$$L(\mathbf{x}, \mathbf{y}, \mathbf{z}; \mathbf{v}, \mathbf{w}) = \frac{\|\mathbf{z}\|_1}{\|\mathbf{y}\|_2} + I(A\mathbf{x} - \mathbf{b}) + \frac{\rho_1}{2} \left\| \mathbf{x} - \mathbf{y} + \frac{1}{\rho_1}\mathbf{v} \right\|_2^2$$
$$+ \frac{\rho_2}{2} \left\| \mathbf{x} - \mathbf{z} + \frac{1}{\rho_2}\mathbf{w} \right\|_2^2, \tag{3}$$

where $I(\cdot)$ is defined as

$$I(\mathbf{t}) = \begin{cases} 0, & \mathbf{t} = \mathbf{0}, \\ +\infty, & \text{otherwise.} \end{cases} \tag{4}$$

There is a closed-form solution for each sub-problem. Please refer to [32] for more details.

This paper contributes three schemes to minimize (2). We demonstrate in experiments that the new schemes are computationally more efficiently compared to the previous ADMM approach. The novelties of the paper are three-fold:

1) Thanks to the new schemes, L_1/L_2 can effectively deal with sparse signals with a high dynamic range, which is not the case for the ADMM approach;
2) We reveal the connection of the proposed schemes to existing approaches, which helps to establish the convergence;
3) Our empirical results shed light about the effects of sparsity, coherence, and dynamic range on sparse recovery, which is new in the CS literature.

The rest of the paper is organized as follows. Section II is devoted to theoretical analysis on the relation between L_1/L_2 and L_1-αL_2, which motivates three numerical schemes to minimize L_1/L_2. We interpret the proposed schemes in line with some existing approaches in Section III, followed by convergence analysis in Section IV. We conduct extensive experiments in Section V to demonstrate the performance of the L_1/L_2 model with three minimizing algorithms over state-of-the-art methods in sparse recovery. Section VI presents how the classic L_1 approach behaves under different dynamic ranges and how sparsity, coherence, and dynamic range interplay on sparse recovery. Finally, conclusions and future works are given in Section VII.

II. NUMERICAL SCHEMES

We establish in Proposition 1 a link between the constrained L_1/L_2 formulation (2) and L_1-αL_2, where α is a positive parameter. Immediately following this proposition, we develop a numerical algorithm for minimizing the ratio model. We further discuss two accelerated approaches in Section II-B.

Proposition 1: Denote

$$\alpha^* := \inf_{\mathbf{x} \in \mathbb{R}^n} \left\{ \frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_2} \text{ s.t. } A\mathbf{x} = \mathbf{b} \right\}, \tag{5}$$

and

$$T(\alpha) := \inf_{\mathbf{x} \in \mathbb{R}^n} \{\|\mathbf{x}\|_1 - \alpha \|\mathbf{x}\|_2 \text{ s.t. } A\mathbf{x} = \mathbf{b}\}, \tag{6}$$

then we have

a) if $T(\alpha) < 0$, then $\alpha > \alpha^*$;
b) if $T(\alpha) \geq 0$, then $\alpha \leq \alpha^*$;
c) if $T(\alpha) = 0$, then $\alpha = \alpha^*$.

Proof: Denote the feasible set of (5) by $\mathbf{F} = \{\mathbf{x} \mid A\mathbf{x} = \mathbf{b}\}$. Since $\mathbf{b} \neq 0$ then $\mathbf{0} \notin \mathbf{F}$.

a) If $T(\alpha) < 0$, then there exists $\mathbf{x} \in \mathbf{F}$ such that $\|\mathbf{x}\|_1 - \alpha \|\mathbf{x}\|_2 < 0$, which implies that $\alpha > \frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_2}$. Therefore, we have $\alpha > \alpha^*$.
b) If $T(\alpha) \geq 0$, then for all $\mathbf{x} \in \mathbf{F}$ we have $\|\mathbf{x}\|_1 - \alpha \|\mathbf{x}\|_2 \geq 0$. So $\alpha \leq \frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_2}$ and hence $\alpha \leq \inf_{\mathbf{x} \in \mathbf{F}} \frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_2} = \alpha^*$, i.e., $\alpha \leq \alpha^*$.
c) If $T(\alpha) = 0$, then by part (b) we get $\alpha \leq \alpha^*$. Furthermore, there exists a sequence $\{\mathbf{x}_n\} \subset \mathbf{F}$ such that $\lim_{n \to \infty}(\|\mathbf{x}_n\|_1 - \alpha \|\mathbf{x}_n\|_2) = 0$. Since $\mathbf{x}_n \in \mathbf{F}$, we have $\|\mathbf{b}\| = \|A\mathbf{x}_n\| \leq \|A\| \|\mathbf{x}_n\|$. Hence, $\{\mathbf{x}_n\}$ has a lower bounded, i.e. $\|\mathbf{x}_n\| \geq \|\mathbf{b}\|/\|A\|$ for all n, then we get

$\lim_{n \to \infty}(\|\mathbf{x}_n\|_1 - \alpha \|\mathbf{x}_n\|_2)/\|\mathbf{x}_n\|_2 = 0$, which means $\alpha^* \leq \lim_{n \to \infty} \|\mathbf{x}_n\|_1/\|\mathbf{x}_n\|_2 = \alpha$. Therefore, we have $\alpha = \alpha^*$. ∎

A. Bisection Search

It follows from Proposition 1 that the optimal value of L_1/L_2 equals to the value of α in the L_1-αL_2 model if the objective value of L_1-αL_2 is zero. That is to say, the optimal value of the ratio model is the root of $T(\alpha)$, which can be obtained by *bisection search*. Moreover, we have upper/lower bounds of α, i.e., $\alpha \in [1, \sqrt{n}]$, since $\|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1 \leq \sqrt{n}\|\mathbf{x}\|_2$, $\forall \mathbf{x} \in \mathbb{R}^n$ [34]. The procedure goes as follows: we start with an initial range of α to be $[1, \sqrt{n}]$ and an initial value of $\alpha^{(0)}$ in between. Then using this $\alpha^{(0)}$, we solve for the L_1-$\alpha^{(0)} L_2$ minimization via the difference-of-convex algorithm (DCA) [35]; more details on the DCA implementation will be given in Section II-B. Based on the objective value of $T(\alpha^{(0)})$, we update the range of α. Specifically if $T(\alpha^{(0)}) = 0$, then we find the minimum ratio and the corresponding minimizer $\mathbf{x}^*$ in the L_1-L_2 model is also the minimizer of the L_1/L_2 model. If $T(\alpha^{(0)}) > 0$, then we update the range as $[\alpha^{(0)}, \sqrt{n}]$. If $T(\alpha^{(0)}) < 0$, then the minimum ratio is smaller than $\alpha^{(0)}$, so we can shorten the range from $[1, \sqrt{n}]$ to $[1, \alpha^{(0)}]$. We can further shorten the internal as $[1, \frac{\|\mathbf{x}^{(k+1)}\|_1}{\|\mathbf{x}^{(k+1)}\|_2}]$, as the objective value of L_1-$\frac{\|\mathbf{x}^{(k+1)}\|_1}{\|\mathbf{x}^{(k+1)}\|_2} L_2$ would be less than or equal to zero in the next iteration. After the range is updated, we choose $\alpha^{(1)}$ using the middle point of two end points and iterate.

We summarize the entire process as Algorithm 1, in which the stopping criterion is that the error between two adjacent α values is small enough. As the algorithmic scheme follows directly from bisection search, we refer the algorithm as L_1/L_2-BS or BS if the context is clear. The convergence of BS can be obtained in the same way that the bisection method converges. However, due to the nonconvex nature of the L_1-αL_2 minimization (6), there is no guarantee to find its global minimizer and hence the solution to (5) may be suboptimal.

B. Adaptive Algorithms

The BS algorithm is computationally expensive, considering that the L_1-αL_2 minimization is conducted for multiple times. To speed up, we discuss two variants of L_1/L_2-BS by updating the parameter α iteratively while minimizing $\|\mathbf{x}\|_1 - \alpha \|\mathbf{x}\|_2$.

Following the DCA framework [36], [37] to minimize $\|\mathbf{x}\|_1 - \alpha \|\mathbf{x}\|_2$, we consider the objective function as the difference of two convex functions, i.e., $\min_{\mathbf{x} \in \mathbb{R}^n} g(\mathbf{x}) - h(\mathbf{x})$. By linearizing the second term $h(\cdot)$, the DCA iterates as follows,

$$\mathbf{x}^{(k+1)} = \arg\min_{\mathbf{x} \in \mathbb{R}^n} g(\mathbf{x}) - \left\langle \mathbf{x}, \nabla h(\mathbf{x}^{(k)}) \right\rangle. \tag{7}$$

Particularly for the L_1-αL_2 model, we have

$$g(\mathbf{x}) = \|\mathbf{x}\|_1 + I(A\mathbf{x} - b) \quad \text{and} \quad h(\mathbf{x}) = \alpha \|\mathbf{x}\|_2, \tag{8}$$

thus leading to the DCA update as

$$\mathbf{x}^{(k+1)} = \arg\min_{\mathbf{x} \in \mathbb{R}^n} g(\mathbf{x}) - \left\langle \mathbf{x}, \frac{\alpha \mathbf{x}^{(k)}}{\|\mathbf{x}^{(k)}\|_2} \right\rangle. \tag{9}$$

Algorithm 1: The L_1/L_2 Minimization Via Bisection Search (L_1/L_2-BS).

1:   Input: $A \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$, kMax, and $\epsilon \in \mathbb{R}$
2:   Initialize: $\mathbf{x}^{(0)}, \alpha^{(0)}, lb = 1, ub = \sqrt{n}$ and $k = 0$
3:   **while** $k <$ kMax or $|\alpha^{(k)} - \alpha^{(k-1)}| > \epsilon$ **do**
4:     $\mathbf{x}^{(k+1)} =$
     $\arg\min_{\mathbf{x} \in \mathbb{R}^n} \{\|\mathbf{x}\|_1 - \alpha^{(k)}\|\mathbf{x}\|_2 \text{ s.t. } A\mathbf{x} = \mathbf{b}\}$
5:     **if** $\|\mathbf{x}^{(k+1)}\|_1 - \alpha^{(k)}\|\mathbf{x}^{(k+1)}\|_2 < 0$ **then**
6:      $ub = \frac{\|\mathbf{x}^{(k+1)}\|_1}{\|\mathbf{x}^{(k+1)}\|_2}$
7:     **else if** $\|\mathbf{x}^{(k+1)}\|_1 - \alpha^{(k)}\|\mathbf{x}^{(k+1)}\|_2 > 0$ **then**
8:      $lb = \alpha^{(k)}$
9:     **else**
10:     break
11:     **end if**
      $\alpha^{(k+1)} = \frac{ub+lb}{2}$
12:     $k = k + 1$
13:   **end while**
14:   **return** $\mathbf{x}^{(k)}$

Algorithm 2: The L_1/L_2 Minimization via Adaptive Selection Method (L_1/L_2-A1 or A2).

1:   Input: $A \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$, kMax, and $\epsilon \in \mathbb{R}$
2:   initialization: $\mathbf{x}^{(0)}, \alpha^{(0)}$ and $k = 1$
3:   **while** $k <$ kMax or $\|\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\|_2 / \|\mathbf{x}^{(k)}\| > \epsilon$ **do**
4:

$$\begin{cases} \text{Update } \{\mathbf{x}^{(k+1)}, \alpha^{(k+1)}\} \text{ by (10)} & \text{for A1} \\ \text{Update } \{\mathbf{x}^{(k+1)}, \alpha^{(k+1)}\} \text{ by (11)} & \text{for A2} \end{cases}$$

5:     $k = k + 1$
6:   **end while**
7:   **return** $\mathbf{x}^{(k)}$

Now we consider to update α iteratively by the ratio of the current solution, leading to the following scheme,

$$\begin{cases} \mathbf{x}^{(k+1)} = \arg\min_{\mathbf{x}} \left\{ g(\mathbf{x}) - \left\langle \mathbf{x}, \frac{\alpha^{(k)}\mathbf{x}^{(k)}}{\|\mathbf{x}^{(k)}\|_2} \right\rangle \right\}, \\ \alpha^{(k+1)} = \|\mathbf{x}^{(k+1)}\|_1 / \|\mathbf{x}^{(k+1)}\|_2, \end{cases} \quad (10)$$

where g is defined in (8). Notice that the $\mathbf{x}$-subproblem in (10) is a linear programming (LP) problem, which unfortunately has no guarantee that the optimal solution exists (as the problem can be unbounded). To increase the robustness of the algorithm, we further incorporate a quadratic term into the linear problem, i.e.,

$$\begin{cases} \mathbf{x}^{(k+1)} = \arg\min_{\mathbf{x}} \left\{ g(\mathbf{x}) - \left\langle \mathbf{x}, \frac{\alpha^{(k)}\mathbf{x}^{(k)}}{\|\mathbf{x}^{(k)}\|_2} \right\rangle + \frac{\beta}{2}\|\mathbf{x} - \mathbf{x}^{(k)}\|_2^2 \right\}, \\ \alpha^{(k+1)} = \|\mathbf{x}^{(k+1)}\|_1 / \|\mathbf{x}^{(k+1)}\|_2. \end{cases}$$

$$(11)$$

We denote these two adaptive methods (10) and (11) as L_1/L_2-A1 and L_1/L_2-A2, respectively or A1 and A2 for short. Both algorithms are summarized in Algorithm 2.

For the $\mathbf{x}$ subproblem of L_1/L_2-A1, we convert it into an LP problem. Assume that $\mathbf{x} = \mathbf{x}^+ - \mathbf{x}^-$ where $\mathbf{x}^+ \geq \mathbf{0}$ and $\mathbf{x}^- \geq \mathbf{0}$. Denote $\bar{\mathbf{x}} = \begin{bmatrix} \mathbf{x}^+ \\ \mathbf{x}^- \end{bmatrix}$, then $A\mathbf{x} = \mathbf{b}$ becomes $\bar{A}\bar{\mathbf{x}} = \mathbf{b}$ with $\bar{A} = [A \quad -A]$. Therefore, the $\mathbf{x}$-subproblem becomes

$$\min_{\bar{\mathbf{x}} \geq \mathbf{0}} \mathbf{c}^T \bar{\mathbf{x}} \quad s.t. \quad \bar{A}\bar{\mathbf{x}} = \mathbf{b}, \quad (12)$$

where $\mathbf{c} = \left[\mathbf{1} + \frac{\alpha^{(k)}\mathbf{x}^{(k)}}{\|\mathbf{x}^{(k)}\|_2}; \mathbf{1} - \frac{\alpha^{(k)}\mathbf{x}^{(k)}}{\|\mathbf{x}^{(k)}\|_2} \right]$. We adopt the software Gurobi [38] to solve this LP problem.

The $\mathbf{x}$ subproblem of L_1/L_2-A2 is a quadratic programming problem, which can be solved via ADMM. By introducing an auxiliary variable $\mathbf{y}$, we have the augmented Lagrangian,

$$L_\rho(\mathbf{x}, \mathbf{y}; \mathbf{u})$$
$$= \|\mathbf{y}\|_1 + I(A\mathbf{x} - \mathbf{b}) - \left\langle \mathbf{x}, \frac{\alpha^{(k)}\mathbf{x}^{(k)}}{\|\mathbf{x}^{(k)}\|_2} \right\rangle$$
$$+ \frac{\beta}{2}\|\mathbf{x} - \mathbf{x}^{(k)}\|_2^2 + \mathbf{u}^T(\mathbf{x} - \mathbf{y}) + \frac{\rho}{2}\|\mathbf{x} - \mathbf{y}\|_2^2. \quad (13)$$

Then the ADMM iteration goes as follows

$$\begin{cases} \mathbf{x}_{j+1} = \arg\min_{\mathbf{x}} L_\rho(\mathbf{x}, \mathbf{y}_j; \mathbf{u}_j), \\ \mathbf{y}_{j+1} = \arg\min_{\mathbf{y}} L_\rho(\mathbf{x}_{j+1}, \mathbf{y}; \mathbf{u}_j), \\ \mathbf{u}_{j+1} = \mathbf{u}_j + \rho(\mathbf{x}_{j+1} - \mathbf{y}_{j+1}), \end{cases} \quad (14)$$

where the subscript j indexes the inner loop, as opposed to the superscript k for outer iterations used in (11). The $\mathbf{x}$-subproblem of (14) is a projection problem to minimize

$$\left\| \mathbf{x} - \frac{\beta\mathbf{x}^{(k)} - \mathbf{u}_j + \rho\mathbf{y}_j + \frac{\alpha^{(k)}\mathbf{x}^{(k)}}{\|\mathbf{x}^{(k)}\|_2}}{\beta + \rho} \right\|_2^2,$$

under the constraint of $A\mathbf{x} = \mathbf{b}$. Since the closed-form solution of projecting a vector $\mathbf{z}$ to this constraint is

$$\mathbf{proj}(\mathbf{z}) = \mathbf{z} - A^T(AA^T)^{-1}(A\mathbf{z} - \mathbf{b}), \quad (15)$$

the $\mathbf{x}$-update is given by

$$\mathbf{x}_{j+1} = \mathbf{proj}\left(\frac{\beta\mathbf{x}^{(k)} - \mathbf{u}_j + \rho\mathbf{y}_j + \frac{\alpha^{(k)}\mathbf{x}^{(k)}}{\|\mathbf{x}^{(k)}\|_2}}{\beta + \rho} \right).$$

The $\mathbf{y}$-subproblem of (14) is equivalent to

$$\mathbf{y}_{j+1} = \arg\min_{\mathbf{y}} \left\{ \|\mathbf{y}\|_1 + \frac{\rho}{2} \left\| \mathbf{y} - \mathbf{x}_{j+1} - \frac{\mathbf{u}_j}{\rho} \right\|_2^2 \right\}.$$

It has a closed-form solution via *soft shrinkage*, i.e.,

$$\mathbf{y}_{j+1} = \mathbf{shrink}\left(\mathbf{x}_{j+1} + \frac{\mathbf{u}_j}{\rho}, \frac{1}{\rho} \right), \quad (16)$$

with $\mathbf{shrink}(\mathbf{v}, \mu) = \text{sign}(\mathbf{v})\max(|\mathbf{v}| - \mu, 0)$.

III. CONNECTIONS TO PREVIOUS WORKS

We try to interpret the proposed adaptive methods (A1 and A2) in line with some existing approaches: parameter selection, generalized inverse power, and gradient-based methods. Our efforts contribute to convergence analysis in Section IV.

A. Parameter Selection

Recall that in L_1/L_2-BS, the ratio L_1/L_2 is minimized when there exists a proper α^* such that $\|\mathbf{x}^*\|_1 - \alpha^*\|\mathbf{x}^*\|_2 = 0$ with $\mathbf{x}^* = \arg\min_{\mathbf{x}}\{\|\mathbf{x}\|_1 - \alpha^*\|\mathbf{x}\|_2 \text{ s.t. } A\mathbf{x} = \mathbf{b}\}$. We can regard this process as a root-finding problem for α^*, which often occurs in parameter selection. For example, in the discrepancy principle method [39]–[41], one aims to find a parameter α such that the resulting data-fitting term is close to the noise level. In particular, we represent this process by

$$\begin{cases} \mathbf{x}^{(k+1)} = \arg\min_{\mathbf{x}} f(\mathbf{x}, \alpha^{(k)}), \\ \alpha^{(k+1)} = l(\mathbf{x}^{(k+1)}, \alpha^{(k)}), \end{cases} \quad (17)$$

where $f(\cdot)$ is a general objective function to be minimized and $l(\cdot)$ is a certain scheme to update α so that discrepancy principle holds. Typically, an inner loop is required to find the solution of $\mathbf{x}$-subproblem, followed by updating this parameter in an outer iteration. We further present the j-th inner iteration at the k-th outer iteration by

$$\mathbf{x}_{j+1} = \Psi(\mathbf{x}_j, \alpha^{(k)}), \quad (18)$$

for the $\mathbf{x}$-subproblem in (17).

To speed-up the process, Wen and Chan [40] proposed an adaptive scheme that updates the parameter during the inner loop such that it renders the current data-fitting term equal to the noise level. In other words, instead of updating α after minimizing f, they directly iterated

$$\mathbf{x}_{j+1} = \Psi(\mathbf{x}_j, \alpha_{j+1}), \quad (19)$$

in a way that $\{\mathbf{x}_{j+1}, \alpha_{j+1}\}$ satisfies the discrepancy principle. In this way, only one loop is needed as opposed to inner/outer loops in (18). But it requires a closed-form solution for $\mathbf{x}_{j+1}$ so one can perform a one-dimensional search for α_{j+1}.

The proposed BS scheme falls into the framework of (17) in that the searching range of parameter is shorten every outer iteration. However, f in our BS method is the L_1-αL_2 minimization that does not have a closed-form solution. As opposed to (19), we consider to update

$$\mathbf{x}_{j+1} = \Psi(\mathbf{x}_j, \alpha_j) \quad (20)$$

prior to updating α. In other word, we update $\mathbf{x}_{j+1}$ based on α_j rather than α_{j+1}, the latter of which was adopted in the parameter-selection method [40]. The rationale of (20) is to guarantee that $\{\mathbf{x}_{j+1}, \alpha_{j+1}\}$ satisfies $\|\mathbf{x}_{j+1}\|_1 - \alpha_{j+1}\|\mathbf{x}_{j+1}\|_1 = 0$. The iterative scheme (20) is consistent with A1 or A2 (depending on the form of Ψ), if we change the notation from subscript j to superscript k.

B. Generalized Inverse Power Methods

A standard technique to find the smallest eigenvalue of a positive semi-definite symmetric matrix B is the inverse power method [34] that requires to iteratively solve the linear system,

$$B\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)}. \quad (21)$$

The iteration converges to the smallest eigenvector of B, denoted by $\mathbf{x}^*$. Then the smallest eigenvalue can be evaluated by $\lambda =$

$q(\mathbf{x}^*)$, where $q(\cdot)$ is Rayleigh quotient defined as

$$q(\mathbf{x}) = \frac{\langle \mathbf{x}, B\mathbf{x} \rangle}{\|\mathbf{x}\|_2^2}.$$

Note that (21) is equivalent to the minimization problem

$$\mathbf{x}^{(k+1)} = \arg\min_{\mathbf{x}} \left\{ \frac{1}{2}\langle \mathbf{x}, B\mathbf{x} \rangle - \langle \mathbf{x}^{(k)}, \mathbf{x} \rangle \right\}. \quad (22)$$

It is well known in linear algebra [34], [42] that eigenvectors of B are critical points of $\min_{\mathbf{x}} q(\mathbf{x})$ and the smallest eigenvalue/eigenvector can be found by (22). This idea is naturally extended to the nonlinear case in [43], where a general quotient is considered, $q(\mathbf{x}) = \frac{r(\mathbf{x})}{s(\mathbf{x})}$, with arbitrary functions $r(\cdot)$ and $s(\cdot)$. Similarly to (22), we have the corresponding scheme

$$\mathbf{x}^{(k+1)} = \arg\min_{\mathbf{x}} \left\{ r(\mathbf{x}) - \langle \nabla s(\mathbf{x}^{(k)}), \mathbf{x} \rangle \right\}.$$

Following [43], we consider to update the eigenvalue $\lambda^{(k)}$ at each iteration to guarantee the algorithm's descent. In particular, the iterative scheme is given by

$$\begin{cases} \mathbf{x}^{(k+1)} = \arg\min_{\mathbf{x}} \left\{ r(\mathbf{x}) - \lambda^{(k)} \langle \nabla s(\mathbf{x}^{(k)}), \mathbf{x} \rangle \right\}, \\ \lambda^{(k+1)} = \frac{r(\mathbf{x}^{(k+1)})}{s(\mathbf{x}^{(k+1)})}. \end{cases} \quad (23)$$

If we choose $r(\mathbf{x}) = g(\mathbf{x})$, $s(\mathbf{x}) = \|\mathbf{x}\|_2$, and denote λ as α, then the generalized inverse power method (23) is L_1/L_2-A1. In [44], a modified inverse power method was proposed via the steepest descent flow. The iteration scheme is to incorporate a quadratic term in the objective function of the $\mathbf{x}$-subproblem, which leads to L_1/L_2-A2.

C. Gradient-Based Methods

Definition 1: A *critical point* of a constrained optimization problem is a vector in the feasible set (satisfying the constraints) that is also a local maximum, minimum, or saddle point of the objective function.

According to Karush-Kuhn-Tucker (KKT) conditions, $\mathbf{x}^* \neq \mathbf{0}$ is a critical point of (2) if and only if there exists a vector $\mathbf{s}$ such that

$$\begin{cases} 0 \in \frac{\partial\|\mathbf{x}^*\|_1}{\|\mathbf{x}^*\|_2} - \frac{\|\mathbf{x}^*\|_1}{\|\mathbf{x}^*\|_2^2}\frac{\mathbf{x}^*}{\|\mathbf{x}^*\|_2} + A^T\mathbf{s}, \\ 0 = A\mathbf{x}^* - \mathbf{b}. \end{cases} \quad (24)$$

By introducing $\hat{\mathbf{s}} = \|\mathbf{x}^*\|_2 \cdot \mathbf{s}$, we have

$$\begin{cases} 0 \in \partial\|\mathbf{x}^*\|_1 - \frac{\|\mathbf{x}^*\|_1}{\|\mathbf{x}^*\|_2}\frac{\mathbf{x}^*}{\|\mathbf{x}^*\|_2} + A^T\hat{\mathbf{s}}, \\ 0 = A\mathbf{x}^* - \mathbf{b}. \end{cases} \quad (25)$$

The condition (25) is also an optimality condition to another optimization problem:

$$\min_{\mathbf{x}} g(\mathbf{x}) + w(\mathbf{x}), \quad (26)$$

where $g(\mathbf{x})$ is from (8) and $w(\mathbf{x})$ is some function satisfying

$$\nabla w(\mathbf{x}) = -\frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_2}\frac{\mathbf{x}}{\|\mathbf{x}\|_2}. \quad (27)$$

Note that $w(\cdot)$ can not be explicitly determined from (27).

By applying a proximal gradient method (PGM) [45]–[47] on the model (26), we obtain the following scheme

$$\mathbf{x}^{(k+1)} = \mathbf{prox}_{\frac{1}{\beta} g}\left(\mathbf{x}^{(k)} - \frac{1}{\beta}\nabla w(\mathbf{x}^{(k)})\right), \qquad (28)$$

where $\mathbf{prox}_g(\mathbf{y}) = \arg\min_{\mathbf{z}}\{g(\mathbf{z}) + \frac{1}{2}\|\mathbf{z} - \mathbf{y}\|_2^2\}$. This iterative scheme is the same as L_1/L_2-A2.

As for L_1/L_2-A1, we can interpret it as a generalized conditional gradient method [48] that minimizes $g(\mathbf{x}) + w(\mathbf{x})$ by $\mathbf{x}^{(k+1)} = \min_{\mathbf{y}}\langle\nabla w(\mathbf{x}^{(k)}), \mathbf{y}\rangle + g(\mathbf{y})$.

IV. CONVERGENCE ANALYSIS

Following the discussion in Section III-C, we present the convergence analysis. We start with the convergence of A2, which is characterized in Theorem 1. To prove it, we need four lemmas, whose proofs are given in Appendix.

Lemma 1: (Sufficient decreasing) The sequence $\{\mathbf{x}^{(k)}, \alpha^{(k)}\}$ produced by L_1/L_2-A2 satisfies

$$\alpha^{(k)} - \alpha^{(k+1)} \geq \frac{\beta}{2\|\mathbf{x}^{(k+1)}\|_2}\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|_2^2, \quad \forall k > 0.$$

The next two lemmas (Lemma 2 and Lemma 3) discuss the Lipschitz properties.

Lemma 2: Define $L = \frac{1}{\|A^T(AA^T)^{-1}\mathbf{b}\|_2}$. Then for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ satisfying $A\mathbf{x} = A\mathbf{y} = \mathbf{b}$, we have

$$\left\|\frac{\mathbf{x}}{\|\mathbf{x}\|_2} - \frac{\mathbf{y}}{\|\mathbf{y}\|_2}\right\|_2 \leq L\|\mathbf{x} - \mathbf{y}\|_2.$$

Since the gradient of the L_2 norm is $\nabla\|\mathbf{x}\|_2 = \frac{\mathbf{x}}{\|\mathbf{x}\|_2}$, Lemma 2 implies that the gradient of Euclidean norm is Lipschitz-continuous in the domain $\{\mathbf{x} \mid A\mathbf{x} = \mathbf{b}\}$. The next lemma is about the Lipschitz property for the implicit function $w(\cdot)$ that satisfies (27).

Lemma 3: Given L defined in Lemma 2. For any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ satisfying $A\mathbf{x} = A\mathbf{y} = \mathbf{b}$, then

$$\|\nabla w(\mathbf{x}) - \nabla w(\mathbf{y})\|_2 \leq L_w\|\mathbf{x} - \mathbf{y}\|_2, \qquad (29)$$

for w satisfying (27) and $L_w = 2\sqrt{n}L$.

Lemma 4: Given $g(\cdot)$ defined in (8) and suppose $w(\cdot)$ satisfies (27), we denote

$$\Phi(\mathbf{x}) := \beta\left(\mathbf{x} - \mathbf{prox}_{\frac{1}{\beta} g}\left(\mathbf{x} - \frac{1}{\beta}\nabla w(\mathbf{x})\right)\right), \qquad (30)$$

for an arbitrary $\beta > 0$. Then we have

a) $\Phi(\mathbf{x}^*) = \mathbf{0}$ if and only if $\mathbf{x}^*$ is a critical point of (2);
b) $\|\Phi(\mathbf{x}) - \Phi(\mathbf{y})\|_2 \leq L_\Phi\|\mathbf{x} - \mathbf{y}\|_2$ with $L_\Phi = L_w + 2\beta$, for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ satisfying $A\mathbf{x} = A\mathbf{y} = \mathbf{b}$.

It is stated in (28) that L_1/L_2-A2 can be expressed as $\mathbf{x}^{(k+1)} = \mathbf{prox}_{\frac{1}{\beta} g}(\mathbf{x}^{(k)} - \frac{1}{\beta}\nabla w(\mathbf{x}^{(k)}))$. By the definition of $\Phi(\cdot)$ in (30) and the decreasing property of $\|\mathbf{x}\|_1/\|\mathbf{x}\|_2$ in Lemma 1, we can interpret A2 as a gradient descent method

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \frac{1}{\beta}\Phi(\mathbf{x}^{(k)}).$$

In the following theorem, we rely on Lemma 4 to show that the descent direction along $\Phi(\cdot)$ leads to convergence.

Theorem 1: Given a sequence $\{\mathbf{x}^{(k)}, \alpha^{(k)}\}$ generated by L_1/L_2-A2. If $\{\mathbf{x}^{(k)}\}$ is bounded, there exists a subsequence that converges to a critical point of the ratio model (2).

Proof: According to Lemma 1, we know that $\alpha^{(k)}$ is decreasing and bounded from below, so there exists a scalar α^* such that $\alpha^{(k)} \to \alpha^*$. With the boundedness assumption of $\mathbf{x}$, we get $\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|_2 \to 0$ from Lemma 1, which implies that $\|\Phi(\mathbf{x}^{(k)})\|_2 \to 0$. The boundedness of $\mathbf{x}^{(k)}$ also leads to a convergent subsequence, i.e., $\mathbf{x}^{(k_i)} \to \mathbf{x}^*$. Therefore, we have

$$\|\Phi(\mathbf{x}^*)\|_2 = \left\|\Phi(\mathbf{x}^*) - \Phi\left(\mathbf{x}^{(k_i)}\right) + \Phi\left(\mathbf{x}^{(k_i)}\right)\right\|_2$$
$$\leq \left\|\Phi(\mathbf{x}^*) - \Phi\left(\mathbf{x}^{(k_i)}\right)\right\|_2 + \left\|\Phi\left(\mathbf{x}^{(k_i)}\right)\right\|_2$$
$$\leq L_\Phi\|\mathbf{x}^{(k_i)} - \mathbf{x}^*\|_2 + \left\|\Phi\left(\mathbf{x}^{(k_i)}\right)\right\|_2.$$

As $k_i \to \infty$, we get $\|\Phi(\mathbf{x}^*)\|_2 = 0$ and hence $\Phi(\mathbf{x}^*) = \mathbf{0}$. By Lemma 4, $\{\mathbf{x}^{(k_i)}\}$ converges to a critical point. ∎

Remark 1: Theorem 1 does not require that the step-size $\frac{1}{\beta}$ is small, which is typically for gradient-based methods. In our numerical tests, we can choose small β and get good results.

Theorem 2: Given a sequence $\{\mathbf{x}^{(k)}, \alpha^{(k)}\}$ generated by L_1/L_2-A1. If $\{\mathbf{x}^{(k)}\}$ is bounded, it has a convergent subsequence.

Proof: Denote

$$z(\mathbf{x}, \mathbf{x}^{(k)}) := \|\mathbf{x}\|_1 - \left\langle\mathbf{x}, \frac{\alpha^{(k)}\mathbf{x}^{(k)}}{\|\mathbf{x}^{(k)}\|_2}\right\rangle.$$

Since $z(\mathbf{x}^{(k)}, \mathbf{x}^{(k)}) = 0$ by the definition of $\alpha^{(k)}$, the minimal value of $z(\mathbf{x}, \mathbf{x}^{(k)})$ subject to the constraint $\{\mathbf{x} \mid A\mathbf{x} = \mathbf{b}\}$ is less than or equal to zero. Specifically, $z(\mathbf{x}^{(k+1)}, \mathbf{x}^{(k)}) \leq 0$. As a result, by Cauchy-Schwarz inequality, we have

$$\|\mathbf{x}^{(k+1)}\|_1 \leq \left\langle\mathbf{x}^{(k+1)}, \frac{\alpha^{(k)}\mathbf{x}^{(k)}}{\|\mathbf{x}^{(k)}\|_2}\right\rangle \leq \alpha^{(k)}\|\mathbf{x}^{(k+1)}\|_2, \quad (31)$$

which implies $\alpha^{(k+1)} \leq \alpha^{(k)}$. Since $\alpha^{(k)} \in [1, \sqrt{n}]$, the decreasing sequence of $\alpha^{(k)}$ converges, i.e., $\alpha^{(k)} \to \alpha^*$.

By the boundedness of $\mathbf{x}^{(k)}$, it has a convergent subsequence, i.e, there exists a vector $\mathbf{x}^*$ such that $\mathbf{x}^{(k_i)} \to \mathbf{x}^*$. ∎

Remark 2: The sufficient decrease property (Lemma 1) does not hold for $\beta = 0$ when L_1/L_2-A2 reduces to A1. So, we cannot show that A1 converges to a critical point.

Remark 3: According to Theorem 1 and Theorem 2, we prove that either both algorithms diverge due to unboundedness or there exists a convergent subsequence. It is possible that the solution can be unbounded. For example, A has a zero-column, then the corresponding entry can take $+\infty$ so that the ratio of L_1 and L_2 is minimized. In the numerical tests, we demonstrate empirically that $\{\mathbf{x}^{(k)}\}$ is always bounded and hence convergent for general (random) matrices A.

V. NUMERICAL EXPERIMENTS

In this section, we compare the proposed algorithms with state-of-the-art methods in sparse recovery. All the numerical experiments are conducted on a desktop with CPU (Intel i7-6700, 3.4 GHz) and MATLAB 9.2 (R2017a).

We focus on the sparse recovery problem with highly coherent matrices, on which standard L_1 models fail. Following the works of [15], [49], [50], we consider an oversampled discrete cosine transform (DCT), defined as $A = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n] \in \mathbb{R}^{m \times n}$ with

$$\mathbf{a}_j := \frac{1}{\sqrt{m}} \cos\left(\frac{2\pi \mathbf{w}j}{F}\right), \quad j = 1, \ldots, n, \quad (32)$$

where $\mathbf{w}$ is a random vector that is uniformly distributed in $[0, 1]^m$ and $F \in \mathbb{R}$ is a positive parameter to control the coherence in a way that a larger F yields a more coherent matrix. Throughout the experiments, we consider over-sampled DCT matrices of size 64×1024. The ground truth $\mathbf{x} \in \mathbb{R}^n$ is simulated as an s-sparse signal, where s is the number of nonzero entries. As suggested in [50], we require a minimum separation at least $2F$ in the support of $\mathbf{x}$. As for the values of non-zero elements, we follow the work of [51] to consider sparse signals with a high dynamic range. Define the dynamic range of a signal $\mathbf{x}$ as $\Theta(\mathbf{x}) = \frac{\max\{|\mathbf{x}_s|\}}{\min\{|\mathbf{x}_s|\}}$, which can be controlled by an exponential factor D. In particular, we simulate $\mathbf{x}_s$ by the following MATLAB command,

```
xs = sign(randn(s,1)).*10.^D*rand(s,1))
```

In the experiments, we set $D = 3$ and 5, corresponding to $\Theta \approx 10^3$ and 10^5, respectively. Note that `randn` and `rand` are the MATLAB commands for the Gaussian distribution $\mathcal{N}(0, 1)$ and the uniform distribution $\mathcal{U}(0, 1)$, respectively. To compare with our previous work [32] of the L_1/L_2 minimization, we also consider that the nonzero elements follow the Gaussian distribution, i.e., $(\mathbf{x}_s)_i \sim \mathcal{N}(0, 1), i = 1, 2, \ldots, s$.

The fidelity of sparse signal recovery is assessed in terms of *success rate*, defined as the number of successful trials over the total number of trials. When the relative error between the ground truth $\mathbf{x}$ and the reconstructed solution $\mathbf{x}^*$, i.e., $\frac{\|\mathbf{x}^* - \mathbf{x}\|_2}{\|\mathbf{x}\|_2}$, is less than 10^{-3}, we declare it as a *success*. Moreover, we categorize the failure of not recovering the ground-truth signal as *model/algorithm failures* and by comparing the objective function $f(\cdot)$ at the ground truth $\mathbf{x}$ and at the restored solution $\mathbf{x}^*$. If $f(\mathbf{x}) > f(\mathbf{x}^*)$, then $\mathbf{x}$ is not a global minimizer of the model, so we regard it as a *model failure*. If $f(\mathbf{x}) < f(\mathbf{x}^*)$, then the algorithm does not reach a global minimizer. It is referred to as an *algorithm failure*. Similarly to success rates, we can define model-failure rates and algorithm-failure rates.

A. Algorithmic Comparison

We present various computational aspects of the proposed algorithms, i.e., BS, A1, and A2, together with comparison to our previous ADMM approach [32]. First of all, we attempt to demonstrate the convergence of all proposed algorithms using an example of $s = 15, F = 15$ (so the minimal separation is 30), and nonzero elements following Gaussian distribution. Since the ratio model is solved via the $L_1 - \alpha L_2$ model, we plot the values of $\|\mathbf{x}^{(k)}\|_1 - \alpha^{(k-1)}\|\mathbf{x}^{(k)}\|_2$ and $\alpha^{(k)}$ versus iteration counter k in Figure 1. For L_1/L_2-BS, we record the value at each outer iteration and the stopping conditions are either the maximum outer iteration reaches 10 or $|\alpha^{(k)} - \alpha^{(k-1)}| \leq 10^{-2}$. For each
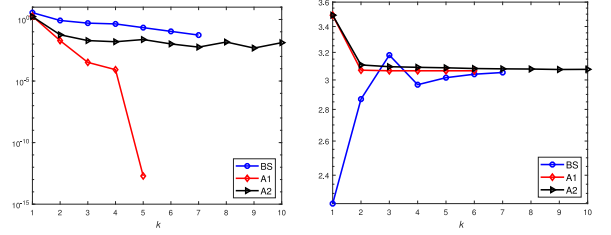

Fig. 1. Empirical analysis on convergence: $\|\mathbf{x}^{(k)}\|_1 - \alpha^{(k-1)}\|\mathbf{x}^{(k)}\|_2$ (left) and $\alpha^{(k)}$ (right) versus iteration counter k for BS, A1, and A2.
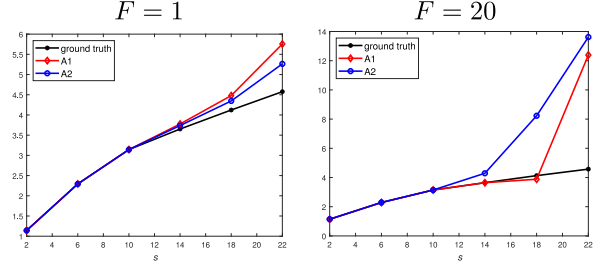

Fig. 2. The L_2 norm of the ground truth vectors as well as the reconstructed solutions by A1 and A2.

iteration of A1, A2, and the inner loop of BS, the stopping criterions are the relative error $\|\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\|_2/\|\mathbf{x}^{(k)}\|_2 \leq 10^{-8}$. The left plot in Figure 1 illustrates the convergence of the three algorithms in the sense that $\|\mathbf{x}^{(k)}\|_1 - \alpha^{(k)}\|\mathbf{x}^{(k)}\|_2$ goes down. Both A1 and A2 are faster than BS as BS starts with a larger range of α as $[1, \sqrt{n}] = [1, 32]$, while A1 and A2 start with a good initial value of $\alpha^{(0)} = \frac{\|\mathbf{x}^{(0)}\|_1}{\|\mathbf{x}^{(0)}\|_2}$, which is very close to the final optimal value α^*.

The right plot in Figure 1 examines the evolution of $\alpha^{(k)}$, which gradually becomes stable and approaches to a similar value around 3.06 for all three algorithms. Figure 1 confirms the decrease property of $\alpha^{(k)}$ proved in Lemma 1.

In Theorem 1, we require the sequence $\{\mathbf{x}^{(k)}\}$ to be bounded for the convergence analysis. Here we aim at an empirical verification on the boundedness. In particular, we test on various kinds of linear systems with $F \in \{1, 20\}$ and sparsity ranging from 2 to 22. In each setting, we randomly generate 50 pairs of ground-truth signals and linear systems to compute the L_2 norm of solutions obtained by A1 and A2, along with the L_2 norm of ground-truth signals. The mean values of these L_2 norms are plotted in Figure 2. As the maximum values are finite numbers, it means that the reconstructed signal is always bounded. Figure 2 also shows that the L_2 norms of A1 and A2 align quite well with the ground truth when the sparsity is below 14, no matter the system is coherent or not. When the matrix is highly coherent with more nonzero elements, both A1 and A2 give much larger values of the L_2 norm compared to the ground truth. It is because a larger L_2 norm gives rise to a smaller value in the ratio sof L_1/L_2 that we try to minimize. In any cases, the solutions of both A1 and A2 are shown to be bounded.

Next, we compare the three algorithms with our previous ADMM approach [32]. We consider $F = 1$ and 20 with nonzero
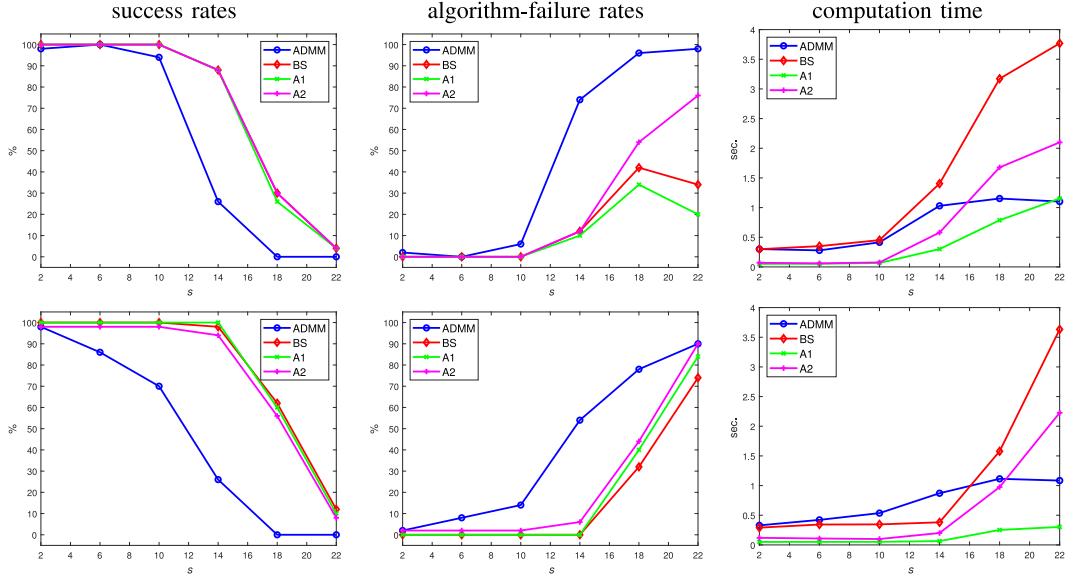
Fig. 3. Algorithmic comparison in the Gaussian distribution case with $F = 1$ (top) and $F = 20$ (bottom) in terms of success rates (left), algorithm-failure rates (middle), and computation time (right).
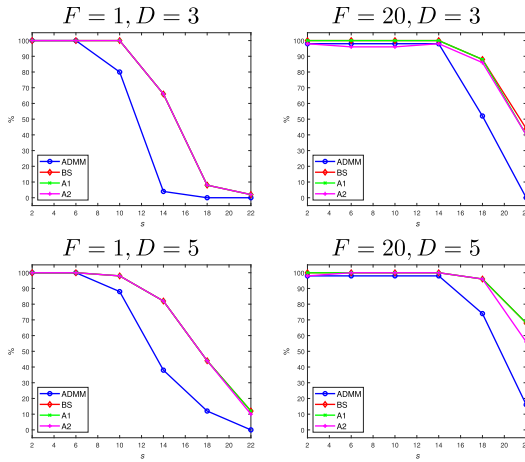

Fig. 4. Success rates of different L_1/L_2 minimizing algorithms versus sparsity at coherence levels $F = 1$ (left) and $F = 20$ (right) as well as high dynamic ranges of $D = 3$ (top) and $D = 5$ (bottom).
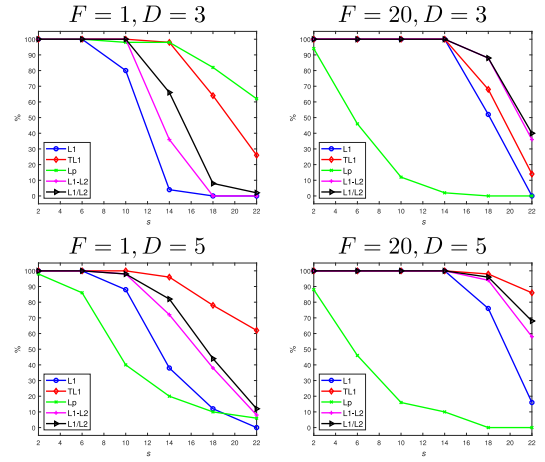

Fig. 5. Success rates of different models versus sparsity at coherence levels $F = 1$ (left) and $F = 20$ (right) as well as high dynamic ranges of $D = 3$ (top) and $D = 5$ (bottom).

elements following the Gaussian distribution or having high dynamic ranges. We randomly simulate 50 trials for each sparsity level and compute the average of success rates, algorithm-failure rates, and computation time. The Gaussian case is illustrated in Figure 3, showing that ADMM is the worst in terms of success rates partly due to high algorithm failure rates. Here, $\rho_1 = \rho_2 = 2000$ for ADMM and $\beta = 1, \rho = 20$ for A2. In addition, BS achieves the highest success rates but is the slowest. Both A1 and A2 have similar performance to BS with much reduced computation time. Figure 4 examines the case of the dynamic range for the non-zero values in x with $D = 3$ and 5. Here we set $\beta = 10^{-5}$ and $\rho = 0.3$ for A2, while $\rho_1 = \rho_2 = 100$ for ADMM. Similar performance is observed as the Gaussian case.

In summary, we rate A1 as the most efficient algorithm for minimizing the ratio model with a balanced performance
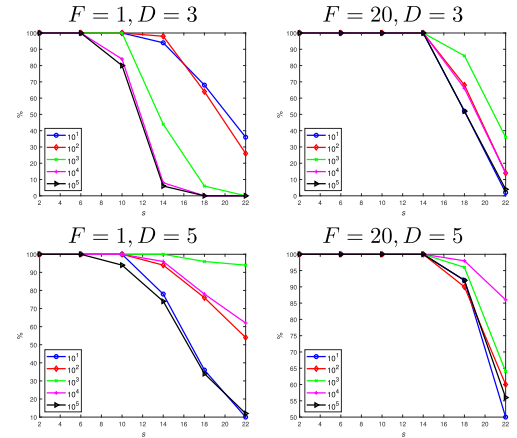

Fig. 6. Success rates of different a values in the TL1 model at coherence levels $F = 1$ (left) and $F = 20$ (right) as well as high dynamic ranges of $D = 3$ (top) and $D = 5$ (bottom).
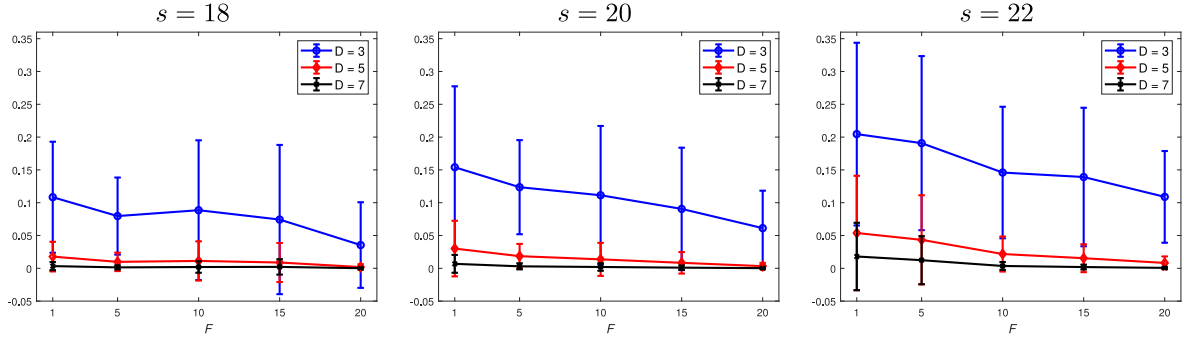
Fig. 7. The relative errors $\|\mathbf{x}^* - \hat{\mathbf{x}}\|_2/\|\hat{\mathbf{x}}\|_2$ produced by the L_1 approach for sparse signal recovery.

between accuracy and computational costs. We also observe that all the algorithms tend to give better performance in terms of success rates with higher dynamic ranges, which seems counter-intuitive. We will revisit this phenomenon in Section VI.

B. Model Comparison

We intend to compare various sparse promoting models. Since the Gaussian case was conducted in our previous work [32], we focus on the dynamic range here. We compare the proposed L_1/L_2 model with the following models: L_1 [10], L_p [11], L_1-L_2 [15], [49], and TL1 [18]. We adopt L_1/L_2-A1 to solve for the ratio model, as it is the most efficient algorithm from the discussion in Section V-A. The initial guess for all non-convex models is the L_1 solution obtained by Gurobi. We choose $p = 1/2$ for L_p and $a = 10^{D-1}$ for TL1 when the range factor D is known *a priori*.

Figure 5 plots the success rates of $F = 1, 20$ and $D = 3, 5$. We observe that TL1 is the best except for the low coherence and the low dynamic case, where L_p is the best. But L_p is the worst in the other cases. The L_1/L_2 model is always the second best. Note that the ratio model is parameter-free, while the performance of TL1 largely relies on the parameter a. Figure 6 examines the success rate of TL1 with different values of a. We choose $a = 10^{D-1}$ in the model comparison, which is almost the best among these testing values of a. If no such prior information of the dynamic range were available to tune a, the performance of TL1 might be worse than L_1/L_2.

VI. DISCUSSIONS

Candés and Wakin [52] presented two principles in compressed sensing, i.e., *sparsity* and *incoherence*. We reported in our previous work [32] that higher coherence leads to better sparse recovery, which seems to contradict with the current belief in CS. In this paper, we discuss the dynamic range and reveal its effect on the exact recovery via the L_1 approach. To our best of our knowledge, there has been little discussion on the dynamic range in the CS literature, except for [51]. We consider low-coherent matrices with $F = 1$ and high-coherent ones with $F = 20$. We record the success rates of different combinations of sparsity levels ($s = 2 : 4 : 22$) and dynamic ranges $D = 0 : 5$ in Table I. It shows that a higher dynamic range leads a better

TABLE I
SUCCESS RATE (%) IN SOLVING DIFFERENT DYNAMIC RANGES VIA THE L_1 MODEL AT TWO COHERENCE LEVELS $F = 1$ AND $F = 20$

$F = 1$						
s	2	6	10	14	18	22
$D = 0$	100	100	80	4	0	0
$D = 1$	100	100	80	4	0	0
$D = 2$	100	100	80	4	0	0
$D = 3$	100	100	80	4	0	0
$D = 4$	100	100	86	16	0	0
$D = 5$	100	100	88	38	12	0
$F = 20$						
s	2	6	10	14	18	22
$D = 0$	100	100	100	100	50	0
$D = 1$	100	100	100	100	52	0
$D = 2$	100	100	100	100	52	0
$D = 3$	100	100	100	100	52	0
$D = 4$	100	100	100	100	54	0
$D = 5$	100	100	100	100	76	16

performance. It seems that the L_1 approach is independent on D for relatively sparser signals.

Now that there are three quantities that may contribute to the success of sparse recovery, i.e., sparsity, coherence, and dynamic range, we try to give a comprehensive analysis by using the relative error $\|\mathbf{x}^* - \mathbf{x}\|_2/\|\mathbf{x}\|_2$ instead of the success rates, as the latter depends on the successful threshold.

We plot in Figure 7 the mean and the standard deviation of the relative errors from 50 random trails versus coherence levels ($F = 1, 5, 10, 15, 20$). Based on Table I, we only consider the number of non-zeros value larger than 18 and $D \geq 3$. In each subfigure of Figure 7, the curves decrease when F increases, which means that higher coherence leads to better performance. This is consistent with the observation in [32]. As for the dynamic range, we discover in Figure 7 that a larger value of D leads to a smaller relative error.

Finally, the sparsity affects the performance in the way that smaller relative errors can be achieved for sparser signals. These numerical phenomena have not been reported in the CS literature, which motivate for future theoretical justifications.

VII. CONCLUSIONS AND FUTURE WORKS

We studied the scale-invariant and parameter-free minimization L_1/L_2 to promote sparsity. We presented three numerical algorithms to minimize this nonconvex model based on the

relationship between L_1/L_2 and L_1-αL_2 for certain α. Experimental results compared the proposed algorithms with state-of-the-art methods in sparse recovery. Particularly important is the proposed algorithm works well when the ground-truth signal has a high dynamic range. Last but not least, we analyzed the behaviors of the L_1 approach towards the exact recovery with varying sparsity, coherence, and dynamic range. Future works include the theoretical analysis on the effect of the high dynamic range towards sparse recovery as well as the applications of the ratio model in image processing such as blind deconvolution [28], [29].

APPENDIX

A. Proof of Lemma 1

Proof: Based on the $\mathbf{x}$-subproblem in (11), we get

$$\left\|\mathbf{x}^{(k+1)}\right\|_1 - \left\langle \mathbf{x}^{(k+1)}, \frac{\alpha^{(k)}\mathbf{x}^{(k)}}{\|\mathbf{x}^{(k)}\|_2} \right\rangle + \frac{\beta}{2}\left\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\right\|_2^2$$

$$\leq \left\|\mathbf{x}^{(k)}\right\|_1 - \left\langle \mathbf{x}^{(k)}, \frac{\alpha^{(k)}\mathbf{x}^{(k)}}{\|\mathbf{x}^{(k)}\|_2} \right\rangle + \frac{\beta}{2}\|\mathbf{x}^{(k)} - \mathbf{x}^{(k)}\|_2^2$$

$$= \left\|\mathbf{x}^{(k)}\right\|_1 - \left\langle \mathbf{x}^{(k)}, \frac{\alpha^{(k)}\mathbf{x}^{(k)}}{\|\mathbf{x}^{(k)}\|_2} \right\rangle.$$

After rearranging, we get the following inequality

$$\|\mathbf{x}^{(k+1)}\|_1 + \frac{\beta}{2}\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|_2^2$$

$$\leq \|\mathbf{x}^{(k)}\|_1 + \alpha^{(k)}\left\langle \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}, \frac{\mathbf{x}^{(k)}}{\|\mathbf{x}^{(k)}\|_2} \right\rangle$$

$$\leq \|\mathbf{x}^{(k)}\|_1 + \alpha^{(k)}\left(\|\mathbf{x}^{(k+1)}\|_2 - \|\mathbf{x}^{(k)}\|_2\right)$$

$$= \alpha^{(k)}\|\mathbf{x}^{(k+1)}\|_2. \tag{33}$$

The second inequality is owing to the convexity of Euclidean norm and the definition of $\alpha^{(k)}$. Lemma 1 is then obtained by dividing $\|\mathbf{x}^{(k+1)}\|_2$ on both sides of (33). ∎

B. Proof of Lemma 2

Proof: Simple calculations lead to

$$\left\|\frac{\mathbf{x}}{\|\mathbf{x}\|_2} - \frac{\mathbf{y}}{\|\mathbf{y}\|_2}\right\|_2^2 = 1 - \frac{2\langle\mathbf{x},\mathbf{y}\rangle}{\|\mathbf{x}\|_2\|\mathbf{y}\|_2} + 1$$

$$= \frac{1}{\|\mathbf{x}\|_2\|\mathbf{y}\|_2}\left(2\|\mathbf{x}\|_2\|\mathbf{y}\|_2 - 2\langle\mathbf{x},\mathbf{y}\rangle\right)$$

$$\leq \frac{1}{\|\mathbf{x}\|_2\|\mathbf{y}\|_2}\left(\|\mathbf{x}\|_2^2 + \|\mathbf{y}\|_2^2 - 2\langle\mathbf{x},\mathbf{y}\rangle\right)$$

$$= \frac{1}{\|\mathbf{x}\|_2\|\mathbf{y}\|_2}\|\mathbf{x} - \mathbf{y}\|_2^2. \tag{34}$$

For any $\mathbf{x}$ satisfying $A\mathbf{x} = \mathbf{b}$, the minimal L_2 norm is reached by projecting the origin $\mathbf{0}$ onto the feasible set of $\{\mathbf{x} \mid A\mathbf{x} = \mathbf{b}\}$. It follows from the projection operator defined in (15) that

$$\|\mathbf{x}\|_2 \geq \|\mathbf{proj}(\mathbf{0})\|_2 = \|A^T(AA^T)^{-1}\mathbf{b}\|_2. \tag{35}$$

Combining (34) and (35), we get Lemma 2. ∎

C. Proof of Lemma 3

Proof: It is straightforward to have

$$\|\nabla w(\mathbf{x}) - \nabla w(\mathbf{y})\|_2 = \left\|\frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_2^2}\mathbf{x} - \frac{\|\mathbf{y}\|_1}{\|\mathbf{y}\|_2^2}\mathbf{y}\right\|_2$$

$$= \left\|\frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_2^2}\mathbf{x} - \frac{\|\mathbf{x}\|_1}{\|\mathbf{y}\|_2^2}\mathbf{y} + \frac{\|\mathbf{x}\|_1}{\|\mathbf{y}\|_2^2}\mathbf{y} - \frac{\|\mathbf{y}\|_1}{\|\mathbf{y}\|_2^2}\mathbf{y}\right\|_2$$

$$\leq \|\mathbf{x}\|_1\left\|\frac{\mathbf{x}}{\|\mathbf{x}\|_2^2} - \frac{\mathbf{y}}{\|\mathbf{y}\|_2^2}\right\|_2 + \frac{1}{\|\mathbf{y}\|_2}\Big|\|\mathbf{x}\|_1 - \|\mathbf{y}\|_1\Big|. \tag{36}$$

We simplify the first term in (36) by calculating

$$\left\|\frac{\mathbf{x}}{\|\mathbf{x}\|_2^2} - \frac{\mathbf{y}}{\|\mathbf{y}\|_2^2}\right\|_2^2 = \frac{1}{\|\mathbf{x}\|_2^2} + \frac{1}{\|\mathbf{y}\|_2^2} - \frac{2\langle\mathbf{x},\mathbf{y}\rangle}{\|\mathbf{x}\|_2^2\|\mathbf{y}\|_2^2}$$

$$= \frac{\|\mathbf{x}\|_2^2 + \|\mathbf{y}\|_2^2 - 2\langle\mathbf{x},\mathbf{y}\rangle}{\|\mathbf{x}\|_2^2\|\mathbf{y}\|_2^2} = \left(\frac{\|\mathbf{x} - \mathbf{y}\|_2}{\|\mathbf{x}\|_2\|\mathbf{y}\|_2}\right)^2,$$

and using $\|\mathbf{x}\|_1 \leq \sqrt{n}\|\mathbf{x}\|_2$. Therefore, we get

$$\|\mathbf{x}\|_1\left\|\frac{\mathbf{x}}{\|\mathbf{x}\|_2^2} - \frac{\mathbf{y}}{\|\mathbf{y}\|_2^2}\right\|_2 \leq \sqrt{n}L\|\mathbf{x} - \mathbf{y}\|_2. \tag{37}$$

As for the second term in (36), we have it bounded by

$$\frac{1}{\|\mathbf{y}\|_2}\Big|\|\mathbf{x}\|_1 - \|\mathbf{y}\|_1\Big| \leq \frac{1}{\|\mathbf{y}\|_2}\|\mathbf{x} - \mathbf{y}\|_1$$

$$\leq \frac{\sqrt{n}}{\|\mathbf{y}\|_2}\|\mathbf{x} - \mathbf{y}\|_2 \leq \sqrt{n}L\|\mathbf{x} - \mathbf{y}\|_2. \tag{38}$$

Combining (37) and (38), we obtain (29). ∎

D. Proof of Lemma 4

Proof: It is straightforward that

$$\Phi(\mathbf{x}^*) = \mathbf{0} \iff \mathbf{x}^* = \mathrm{prox}_{\frac{1}{\beta}g}\left(\mathbf{x}^* - \frac{1}{\beta}\nabla w(\mathbf{x}^*)\right).$$

By the optimality condition [47], the latter relation holds if and only if there exists a vector $\mathbf{s}$ such that

$$0 \in \partial\|\mathbf{x}^*\|_1 + \nabla w(\mathbf{x}^*) + \beta(\mathbf{x}^* - \mathbf{x}^*) + A^T\mathbf{s}$$

$$= \partial\|\mathbf{x}^*\|_1 + \nabla w(\mathbf{x}^*) + A^T\mathbf{s},$$

which implies that $\mathbf{x}^*$ is a critical point of (26). It follows from (28) that (26) is equivalent to (2) and hence $\mathbf{x}^*$ is also a critical point of (2). According to the nonexpansiveness of the proximal operator and the Lipschitz continuousness of ∇w, we have

$$\|\Phi(\mathbf{x}) - \Phi(\mathbf{y})\|_2$$

$$\leq \beta\left\|\mathrm{prox}_{\frac{1}{\beta}g}\left(\mathbf{x} - \frac{1}{\beta}\nabla w(\mathbf{x})\right) - \mathrm{prox}_{\frac{1}{\beta}g}\left(\mathbf{y} - \frac{1}{\beta}\nabla w(\mathbf{y})\right)\right\|_2$$

$$+ \beta\|\mathbf{x} - \mathbf{y}\|_2$$

$$\leq \beta\left\|\left(\mathbf{x} - \frac{1}{\beta}\nabla w(\mathbf{x})\right) - \left(\mathbf{y} - \frac{1}{\beta}\nabla w(\mathbf{y})\right)\right\|_2 + \beta\|\mathbf{x} - \mathbf{y}\|_2$$

$$\leq \|\nabla w(\mathbf{x}) - \nabla w(\mathbf{y})\|_2 + 2\beta\|\mathbf{x} - \mathbf{y}\|_2$$

$$\leq (L_w + 2\beta)\|\mathbf{x} - \mathbf{y}\|_2.$$

The Lemma follows. ∎

REFERENCES

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

[2] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, no. 1-4, pp. 259–268, 1992.

[3] A. Berman and R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*. Philadelphia, PA, USA: SIAM, 1994.

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

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

[6] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Comput.*, vol. 24, no. 2, pp. 227–234, 1995.

[7] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, 1993, pp. 40–44.

[8] S. Chen, S. A. Billings, and W. Luo, "Orthogonal least squares methods and their application to non-linear system identification," *Int. J. Control*, vol. 50, no. 5, pp. 1873–1896, 1989.

[9] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, 2009.

[10] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, 2001.

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

[12] Z. Xu, X. Chang, F. Xu, and H. Zhang, "$L_{1/2}$ regularization: A thresholding representation theory and a fast solver," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 7, pp. 1013–1027, Jul. 2012.

[13] M.-J. Lai, Y. Xu, and W. Yin, "Improved iteratively reweighted least squares for unconstrained smoothed ℓ_q minimization," *SIAM J. Numer. Anal.*, vol. 51, no. 2, pp. 927–957, 2013.

[14] P. Yin, E. Esser, and J. Xin, "Ratio and difference of l_1 and l_2 norms and sparse representation with coherent dictionaries," *Commun. Inf. Syst.*, vol. 14, no. 2, pp. 87–109, 2014.

[15] 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," *J. Sci. Comput.*, vol. 64, no. 1, pp. 178–196, 2015.

[16] J. Lv *et al.*, "A unified approach to model selection and sparse recovery using regularized least squares," *Ann. Stat.*, vol. 37, no. 6A, pp. 3498–3528, 2009.

[17] S. Zhang and J. Xin, "Minimization of transformed L_1 penalty: Closed form representation and iterative thresholding algorithms," *Comm. Math. Sci.*, vol. 15, pp. 511–537, 2017.

[18] S. Zhang and J. Xin, "Minimization of transformed L_1 penalty: Theory, difference of convex function algorithm, and robust application in compressed sensing," *Math. Program.*, vol. 169, no. 1, pp. 307–336, 2018.

[19] L. Breiman, "Better subset regression using the nonnegative garrote," *Technometrics*, vol. 37, no. 4, pp. 373–384, 1995.

[20] D. Peleg and R. Meir, "A bilinear formulation for vector sparsity optimization," *Signal Process.*, vol. 88, no. 2, pp. 375–389, 2008.

[21] T. Zhang, "Multi-stage convex relaxation for learning with sparse regularization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 1929–1936.

[22] X. Shen, W. Pan, and Y. Zhu, "Likelihood-based selection and sharp parameter estimation," *J. Amer. Statistical Assoc.*, vol. 107, no. 497, pp. 223–232, 2012.

[23] P. O. Hoyer, "Non-negative sparse coding," in *Proc. 12th IEEE Workshop Neural Netw. for Signal Process.*, 2002, pp. 557–565.

[24] N. Hurley and S. Rickard, "Comparing measures of sparsity," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4723–4741, Oct. 2009.

[25] E. Esser, Y. Lou, and J. Xin, "A method for finding structured sparse solutions to nonnegative least squares problems with applications," *SIAM J. Imag. Sci.*, vol. 6, no. 4, pp. 2010–2046, 2013.

[26] E. Esser, T. Lin, R. Wang, and F. J. Herrmann, "A lifted ℓ_1/ℓ_2 constraint for sparse blind deconvolution," in *Proc. 77th EAGE Conf. Exhib.*, 2015, no. 1, pp. 1–5.

[27] E. Esser, T. T. Lin, F. J. Herrmann, and R. Wang, "Resolving scaling ambiguities with the ℓ_1/ℓ_2 norm in a blind deconvolution problem with feedback," in *Proc. IEEE 6th Int. Workshop Comput. Adv. Multi-Sensor Adapt. Process.*, 2015, pp. 365–368.

[28] D. Krishnan, T. Tay, and R. Fergus, "Blind deconvolution using a normalized sparsity measure," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, 2011, pp. 233–240.

[29] A. Repetti, M. Q. Pham, L. Duval, E. Chouzenoux, and J. C. Pesquet, "Euclid in a taxicab: Sparse blind deconvolution with smoothed ℓ_1/ℓ_2 regularization," *IEEE Signal Process Lett.*, vol. 22, no. 5, pp. 539–543, May 2015.

[30] M. Q. Pham, B. Oudompheng, J. I. Mars, and B. Nicolas, "A noise-robust method with smoothed ℓ_1/ℓ_2 regularization for sparse moving-source mapping," *Signal Process.*, vol. 135, pp. 96–106, 2017.

[31] X. Jia, M. Zhao, Y. Di, P. Li, and J. Lee, "Sparse filtering with the generalized ℓ_p/ℓ_q norm and its applications to the condition monitoring of rotating machinery," *Mech. Syst. Signal Process.*, vol. 102, pp. 198–213, 2018.

[32] Y. Rahimi, C. Wang, H. Dong, and Y. Lou, "A scale invariant approach for sparse signal recovery," *SIAM J. Sci. Comput.*, vol. 41, no. 6, pp. A3649–A3672, 2019.

[33] S. Boyd *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[34] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD, USA: *The Johns Hopkins Univ. Press*, 1996, pp. 374–426.

[35] Y. Lou, S. Osher, and J. Xin, "Computational aspects of constrained l_1-l_2 minimization for compressive sensing," in *Modelling, Computation and Optimization in Information Systems and Management Sciences*. Berlin, Germany: Springer, 2015, pp. 169–180.

[36] T. Pham-Dinh and H. A. Le-Thi, "A D.C. optimization algorithm for solving the trust-region subproblem," *SIAM J. Optim.*, vol. 8, no. 2, pp. 476–505, 1998.

[37] T. Pham-Dinh and H. A. Le-Thi, "The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems," *Ann. Oper. Res.*, vol. 133, no. 1-4, pp. 23–46, 2005.

[38] "Gurobi optimizer reference manual, 2015," *Gurobi Optimization, Inc.*, 2014. [Online]. Available: https://www.gurobi.com

[39] V. A. Morozov, *Methods for Solving Incorrectly Posed Problems*. Berlin, Germany: Springer, 2012.

[40] Y. Wen and R. H. Chan, "Parameter selection for total-variation-based image restoration using discrepancy principle," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1770–1781, Apr. 2012.

[41] T. Teuber, G. Steidl, and R. H. Chan, "Minimization and parameter estimation for seminorm regularization models with I-divergence constraints," *Inverse Probability*, vol. 29, no. 3, 2013, Art. no. 035007.

[42] L. N. Trefethen and D. Bau III, *Numerical Linear Algebra*, vol. 50. Philadelphia, PA, USA: SIAM, 1997.

[43] M. Hein and T. Bühler, "An inverse power method for nonlinear eigenproblems with applications in 1-spectral clustering and sparse PCA," in *Proc. Adv. Neural Inf. Proces. Syst.*, 2010, pp. 847–855.

[44] X. Bresson, T. Laurent, D. Uminsky, and J. V. Brecht, "Convergence and energy landscape for cheeger cut clustering," in *Proc. Adv. Neural Inf. Proces. Syst.*, 2012, pp. 1385–1393.

[45] N. Parikh *et al.*, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, no. 3, pp. 127–239, 2014.

[46] M. Fukushima and H. Mine, "A generalized proximal point algorithm for certain non-convex minimization problems," *Int. J. Syst. Sci.*, vol. 12, no. 8, pp. 989–1000, 1981.

[47] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*. Berlin, Germany: Springer, 2011, pp. 185–212.

[48] K. Bredies, D. A. Lorenz, and P. Maass, "A generalized conditional gradient method and its connection to an iterative shrinkage method," *Comput. Opt. Appl.*, vol. 42, no. 2, pp. 173–193, 2009.

[49] P. Yin, Y. Lou, Q. He, and J. Xin, "Minimization of ℓ_{1-2} for compressed sensing," *SIAM J. Sci. Comput.*, vol. 37, no. 1, pp. A536–A563, 2015.

[50] A. Fannjiang and W. Liao, "Coherence pattern–guided compressive sensing with unresolved grids," *SIAM J. Imag. Sci.*, vol. 5, no. 1, pp. 179–202, 2012.

[51] D. A. Lorenz, "Constructing test instances for basis pursuit denoising," *IEEE Trans. Signal Process.*, vol. 61, no. 5, pp. 1210–1214, Mar. 2013.

[52] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.