SIAM J. SCI. COMPUT.
Vol. 41, No. 6, pp. A3649–A3672




# A SCALE-INVARIANT APPROACH FOR SPARSE SIGNAL RECOVERY*

YAGHOUB RAHIMI†, CHAO WANG‡, HONGBO DONG§, AND YIFEI LOU†

**Abstract.** In this paper, we study the ratio of the $L_1$ and $L_2$ norms, denoted as $L_1/L_2$, to promote sparsity. Due to the nonconvexity and nonlinearity, there has been little attention to this scale-invariant model. Compared to popular models in the literature, such as the $L_p$ model for $p \in (0, 1)$ and the transformed $L_1$, this ratio model is parameter free. Theoretically, we present a strong null space property (sNSP) and prove that any sparse vector is a local minimizer of the $L_1/L_2$ model provided with this sNSP condition. Computationally, we focus on a constrained formulation that can be solved via the alternating direction method of multipliers. Experiments show that the proposed approach is comparable to the state-of-the-art methods in sparse recovery. In addition, a variant of the $L_1/L_2$ model to apply on the gradient is also discussed with a proof-of-concept example of the MRI reconstruction.

**Key words.** sparsity, $L_0$, $L_1$, null space property, alternating direction method of multipliers, MRI reconstruction

**AMS subject classifications.** 90C90, 65K10, 49N45, 49M20

**DOI.** 10.1137/18M123147X

## 1. Introduction.

Sparse signal recovery is to find the sparsest solution of $A\mathbf{x} = \mathbf{b}$, where $\mathbf{x} \in \mathbb{R}^n, \mathbf{b} \in \mathbb{R}^m$, and $A \in \mathbb{R}^{m\times n}$ for $m \ll n$. This problem is often referred to as *compressed sensing* (CS) in the sense that the sparse signal $\mathbf{x}$ is compressible. Mathematically, this fundamental problem in CS can be formulated as

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{x}\|_0 \quad \text{s.t.} \quad A\mathbf{x} = \mathbf{b}, \tag{1.1}$$

where $\|\mathbf{x}\|_0$ is the number of nonzero entries in $\mathbf{x}$. Unfortunately, (1.1) is NP-hard [31] to solve. A popular approach in CS is to replace $L_0$ by the convex $L_1$ norm, i.e.,

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad A\mathbf{x} = \mathbf{b}. \tag{1.2}$$

Computationally, there are various $L_1$ minimization algorithms, such as primal dual [8], forward-backward splitting [34], and the alternating direction method of multipliers (ADMM) [4].

A major breakthrough in CS was the *restricted isometry property* (RIP) [6], which provides a sufficient condition of minimizing the $L_1$ norm to recover the sparse signal. There is a necessary and sufficient condition given in terms of null space of the matrix $A$, thus referred to as *null space property* (NSP); see Definition 1.1.

*Submitted to the journal's Methods and Algorithms for Scientific Computing section December 19, 2018; accepted for publication (in revised form) August 9, 2019; published electronically November 19, 2019.

https://doi.org/10.1137/18M123147X

**Funding:** The work of the first and fourth authors was partially supported by NSF grants DMS-1522786 and CAREER 1846690.

†Department of Mathematical Sciences, University of Texas at Dallas, Richardson, TX 75080 (yxr160430@utdallas.edu, yifei.lou@utdallas.edu).

‡Corresponding author. Department of Mathematical Sciences, University of Texas at Dallas, Richardson, TX 75080 (chaowang.hk@gmail.com).

§Department of Mathematics and Statistics, Washington State University, Pullman, WA 99164 (hongbo.dong@wsu.edu).

DEFINITION 1.1 (NSP [10]). *For any matrix $A \in \mathbb{R}^{m\times n}$, we say the matrix $A$ satisfies an NSP of order $s$ if*

$$\|\mathbf{v}_S\|_1 < \|\mathbf{v}_{\bar{S}}\|_1, \ \mathbf{v} \in \ker(A)\backslash\{\mathbf{0}\}, \ \forall S \subset [n], \ |S| \le s, \tag{1.3}$$

*where $[n] := \{1, \dots, n\}$; $\bar{S}$ is the complement of $S$, i.e., $[n]\backslash S$; and $\mathbf{x}_S$ is defined as*

$$(\mathbf{x}_S)_i = \begin{cases} x_i & \text{if } i \in S, \\ 0 & \text{otherwise.} \end{cases}$$

*The null space of $A$ is denoted by* $\ker(A) := \{\mathbf{x} \mid A\mathbf{x} = \mathbf{0}\}$.

Donoho and Huo [12] proved that every $s$-sparse signal $\mathbf{x} \in \mathbb{R}^n$ is the unique solution to the $L_1$ minimization (1.2) if and only if $A$ satisfies the NSP of order $s$. NSP quantifies the notion that vectors in the null space of $A$ should not be too concentrated on a small subset of indices. Since it is a necessary and sufficient condition, NSP is widely used in proving other exact recovery guarantees. Note that NSP is no longer necessary if "every $s$-sparse vector" is relaxed. A weaker[1] sufficient condition for the exact $L_1$ recovery was proved by Zhang [49]. It is stated that if a vector $\mathbf{x}^*$ satisfies $A\mathbf{x}^* = \mathbf{b}$ and

$$\sqrt{\|\mathbf{x}^*\|_0} < \frac{1}{2} \min_{\mathbf{v}} \left\{ \frac{\|\mathbf{v}\|_1}{\|\mathbf{v}\|_2} : \mathbf{v} \in \ker(A)\backslash\{\mathbf{0}\} \right\}, \tag{1.4}$$

then $\mathbf{x}^*$ is the unique solution to both (1.1) and (1.2). Unfortunately, neither RIP nor NSP can be numerically verified for a given matrix [1, 38].

Alternatively, a computable condition for $L_1$'s exact recovery is based on *coherence*, which is defined as

$$\mu(A) := \max_{i\neq j} \frac{|\mathbf{a}_i^T \mathbf{a}_j|}{\|\mathbf{a}_i\|\|\mathbf{a}_j\|} \tag{1.5}$$

for a matrix $A = [\mathbf{a}_1, \dots, \mathbf{a}_N]$. Donoho and Elad [11] and Gribonval [16] proved independently that if $\mathbf{x}^*$ satisfies $A\mathbf{x}^* = \mathbf{b}$ and

$$\|\mathbf{x}^*\|_0 < \frac{1}{2}\left(1 + \frac{2}{\mu(A)}\right), \tag{1.6}$$

then $\mathbf{x}^*$ is the optimal solution to both (1.1) and (1.2). Clearly, the coherence $\mu(A)$ is bounded by $[0, 1]$. The inequality (1.6) implies that $L_1$ may not perform well for highly coherent matrices, i.e., $\mu(A) \sim 1$, as $\|\mathbf{x}\|_0$ is then at most one, which seldom occurs simultaneously with $A\mathbf{x}^* = \mathbf{b}$.

Other than the popular $L_1$ norm, there are a variety of regularization functionals to promote sparsity, such as $L_p$ [9, 43, 23], $L_1$-$L_2$ [44, 26], capped $L_1$ (CL1) [48, 37], and transformed $L_1$ (TL1) [29, 46, 47]. Most of these models are nonconvex, leading to difficulties in proving exact recovery guarantees and algorithmic convergence, but they tend to give better empirical results compared to the convex $L_1$ approach. For example, it was reported in [44, 26] that $L_p$ gives superior results for incoherent matrices (i.e., $\mu(A)$ is small), while $L_1$-$L_2$ is the best for the coherent scenario. In addition, TL1 is always the second best no matter whether the matrix is coherent or not [46, 47].

[1] The sufficient condition of (1.4) is weaker than the one in (1.3).

In this paper, we study the ratio of $L_1$ and $L_2$ as a scale-invariant model to approximate the desired $L_0$, which is scale-invariant itself. In the one-dimensional case (i.e., $n = 1$), the $L_1/L_2$ model is exactly the same as the $L_0$ model if we use the convention $\frac{0}{0} = 0$. The ratio of $L_1$ and $L_2$ was first proposed by Hoyer [20] as a sparseness measure and later highlighted in [21] as a scale-invariant model. However, there has been little attention on it due to the computational difficulties that have arisen from its being nonconvex and nonlinear. There are some theorems that establish the equivalence between the $L_1/L_2$ and the $L_0$ models but only restricted to nonnegative signals [13, 44]. We aim to apply this ratio model to arbitrary signals. On the other hand, the $L_1/L_2$ minimization has an intrinsic drawback that tends to produce one erroneously large coefficient while suppressing the other nonzero elements, in which case the ratio is reduced. To compensate for this drawback, it is helpful to incorporate a box constraint, which will also be addressed in this paper.

Now we turn to a sparsity-related assumption that a signal is sparse after a given transform, as opposed to a signal itself being sparse. This assumption is widely used in image processing. For example, a natural image, denoted by $u$, is mostly sparse after taking a gradient, and hence it is reasonable to minimize the $L_0$ norm of the gradient, i.e., $\|\nabla u\|_0$. To bypass the NP-hard $L_0$ norm, the convex relaxation replaces $L_0$ by $L_1$, where the $L_1$ norm of the gradient is the well-known total variation (TV) [36] of an image. A weighted $L_1$-$\alpha L_2$ model (for $\alpha > 0$) on the gradient was proposed in [27], which suggested that $\alpha = 0.5$ yields better results than $\alpha = 1$ for image denoising, deblurring, and MRI reconstruction. The ratio of $L_1$ and $L_2$ on the image gradient was used in deconvolution and blind deconvolution [22, 35]. We further adapt the proposed ratio model from sparse signal recovery to imaging applications, specifically focusing on MRI reconstruction.

The rest of the paper is organized as follows. Section 2 is devoted to a theoretical analysis of the $L_1/L_2$ model. In section 3, we apply the ADMM to minimize the ratio of $L_1$ and $L_2$ with two variants of incorporating a box constraint as well as applying on the image gradient. We conduct extensive experiments in section 4 to demonstrate the performance of the proposed approaches over the state of the art in sparse recovery and MRI reconstruction. Section 5 is a fun exercise where we use the $L_1/L_2$ minimization to compute the right-hand side of the NSP condition (1.4), leading to an empirical upper bound of the exact $L_1$ recovery guarantee. Finally, conclusions and future works are given in section 6.

## 2. Rationales of the $L_1/L_2$ model.

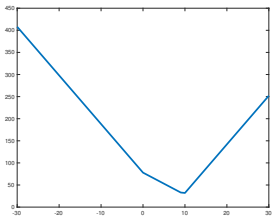
We begin with a toy example to illustrate the advantages of $L_1/L_2$ over other alternatives, followed by some theoretical properties of the proposed model.

### 2.1. Toy example.

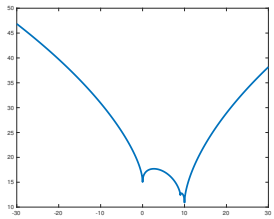
Define a matrix $A$ as

$$A := \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & -1 \end{bmatrix} \in \mathbb{R}^{5\times 6} \tag{2.1}$$

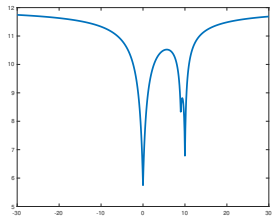
and $\mathbf{b} = (0, 0, 20, 40, 18)^T \in \mathbb{R}^5$. It is straightforward that any general solutions of $A\mathbf{x} = \mathbf{b}$ have the form of $\mathbf{x} = (t, t, t, 20 - 2t, 40 - 4t, 2(t - 9))^T$ for a scalar $t \in \mathbb{R}$. The sparsest solution occurs at $t = 0$, where the sparsity of $\mathbf{x}$ is 3 and some local solutions include $t = 10$ for sparsity being 4 and $t = 9$ for sparsity being 5. In Figure 1, we



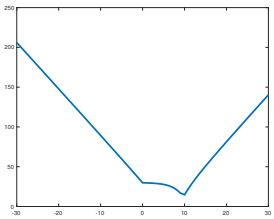
(a) $L_1$



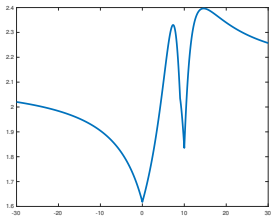
(b) $L_p$ (p=1/2)



(c) TL1



(d) $L_1$-$L_2$



(e) $L_1/L_2$

FIG. 1. *The objective functions of a toy example illustrate that only $L_1/L_2$ and TL1 can find $t = 0$ as the global minimizer, but TL1 has a very narrow basin of attraction (thus sensitive to initial guess and difficult to find the global solution).*

plot various objective functions with respect to $t$, including $L_1, L_p$ (for $p = 1/2$), $L_1$-$L_2$, and TL1 (for $a = 1$ as suggested in [47]). Note that all these functions are not differentiable at the values of $t = 0, 9$, and 10, where the sparsity of $\mathbf{x}$ is strictly smaller than 6. The sparsest vector $\mathbf{x}$ corresponding to $t = 0$ can only be found by minimizing TL1 and $L_1/L_2$, while the other models find $t = 10$ as a global minimum.

**2.2. Theoretical properties.** Recently, Tran and Webster [39] generalized the NSP to deal with sparse promoting metrics that are symmetric, separable, and concave, which unfortunately does not apply to $L_1/L_2$ (not separable), but this work motivates us to consider a stronger form of the NSP, as defined in Definition 2.1.

DEFINITION 2.1. *For any matrix $A \in \mathbb{R}^{m\times n}$, we say the matrix $A$ satisfies a strong null space property (sNSP) of order $s$ if*

$$(s+1)\,\|\mathbf{v}_S\|_1 \le \|\mathbf{v}_{\bar{S}}\|_1\,,\ \mathbf{v} \in \ker(A)\backslash\{\mathbf{0}\},\ \forall S \subset [n],\ |S| \le s. \tag{2.2}$$

Note that Definition 2.1 is stronger than the original NSP in Definition 1.1 in the sense that if a matrix satisfies sNSP, then it also satisfies the original NSP. The

following theorem says that any $s$-sparse vector is a local minimizer of $L_1/L_2$ provided the matrix has the sNSP of order $s$. The proof is given in the appendix.

THEOREM 2.2. *Assume an $m \times n$ matrix $A$ satisfies the sNSP of order $s$. Then any $s$-sparse solution of $A\mathbf{x} = \mathbf{b}$ ($\mathbf{b} \neq \mathbf{0}$) is a local minimum for $L_1/L_2$ in the feasible space of $A\mathbf{x} = \mathbf{b}$; i.e., there exists a positive number $t^* > 0$ such that for every $\mathbf{v} \in \ker(A)$ with $0 < \|\mathbf{v}\|_2 \leq t^*$, we have*

$$\frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_2} \leq \frac{\|\mathbf{x}+\mathbf{v}\|_1}{\|\mathbf{x}+\mathbf{v}\|_2}. \tag{2.3}$$

Finally, we show the optimal value of the $L_1/L_2$ subject to $A\mathbf{x} = \mathbf{b}$ is upper bounded by the same ratio with $\mathbf{b} = \mathbf{0}$; see Proposition 1.

PROPOSITION 1. *For any $A \in \mathbb{R}^{m\times n}, \mathbf{x} \in \mathbb{R}^n$, we have*

$$\inf_{\mathbf{z}\in\mathbb{R}^n} \left\{ \frac{\|\mathbf{z}\|_1}{\|\mathbf{z}\|_2} \,\middle|\, A\mathbf{z} = A\mathbf{x} \right\} \leq \inf_{\mathbf{z}\in\mathbb{R}^n} \left\{ \frac{\|\mathbf{z}\|_1}{\|\mathbf{z}\|_2} \,\middle|\, \mathbf{z} \in \ker(A) \setminus \{\mathbf{0}\} \right\}. \tag{2.4}$$

*Proof.* Denote

$$\alpha^* = \inf_{\mathbf{z}\in\mathbb{R}^n} \left\{ \frac{\|\mathbf{z}\|_1}{\|\mathbf{z}\|_2} \,\middle|\, A\mathbf{z} = A\mathbf{x} \right\}. \tag{2.5}$$

For every $\mathbf{v} \in \ker(A) \setminus \{\mathbf{0}\}$ and $t \in \mathbb{R}$, we have that

$$\alpha^* \leq \frac{\|\mathbf{x}+t\mathbf{v}\|_1}{\|\mathbf{x}+t\mathbf{v}\|_2} \tag{2.6}$$

since $A(\mathbf{x}+t\mathbf{v}) = \mathbf{b}$. Then we obtain

$$\lim_{t\to\infty} \frac{\|\mathbf{x}+t\mathbf{v}\|_1}{\|\mathbf{x}+t\mathbf{v}\|_2} = \lim_{t\to\infty} \frac{\|\mathbf{x}/t+\mathbf{v}\|_1}{\|\mathbf{x}/t+\mathbf{v}\|_2} = \frac{\|\mathbf{v}\|_1}{\|\mathbf{v}\|_2}. \tag{2.7}$$

Therefore, for every $\mathbf{v} \in \ker(A) \setminus \{\mathbf{0}\}$,

$$\alpha^* \leq \frac{\|\mathbf{v}\|_1}{\|\mathbf{v}\|_2}, \tag{2.8}$$

which directly leads to the desired inequality (2.4). ∎

Proposition 1 implies that the left-hand side of the inequality involves both the underlying signal $\mathbf{x}$ and the system matrix $A$, which can be upper bounded by the minimum ratio that only involves $A$.

**3. Numerical schemes.** The proposed model is

$$\min_{\mathbf{x}\in\mathbb{R}^n} \left\{ \frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_2} + I_0(A\mathbf{x}-\mathbf{b}) \right\}, \tag{3.1}$$

where $I_S(\mathbf{t})$ is the function enforcing $\mathbf{t}$ into the feasible set $S$, i.e.,

$$I_S(\mathbf{t}) = \begin{cases} 0 & \mathbf{t} \in S, \\ +\infty & \text{otherwise.} \end{cases} \tag{3.2}$$

In subsection 3.1, we detail the ADMM algorithm for minimizing (3.1), followed by a minor change to incorporate an additional box constraint in subsection 3.2. We discuss in subsection 3.3 another variant of $L_1/L_2$ on the gradient to deal with imaging applications.

**3.1. The $L_1/L_2$ minimization via ADMM.** In order to apply the ADMM [4] to solve for (3.1), we introduce two auxiliary variables and rewrite (3.1) into an equivalent form:

$$\min_{\mathbf{x},\mathbf{y},\mathbf{z}} \left\{ \frac{\|\mathbf{z}\|_1}{\|\mathbf{y}\|_2} + I_0(A\mathbf{x}-\mathbf{b}) \right\} \quad \text{s.t.} \quad \mathbf{x}=\mathbf{y}, \quad \mathbf{x}=\mathbf{z}. \tag{3.3}$$

The augmented Lagrangian for (3.3) is

$$\begin{aligned} &L_{\rho_1,\rho_2}(\mathbf{x},\mathbf{y},\mathbf{z};\mathbf{v},\mathbf{w}) \\ &\quad = \frac{\|\mathbf{z}\|_1}{\|\mathbf{y}\|_2} + I_0(A\mathbf{x}-\mathbf{b}) + \langle \mathbf{v}, \mathbf{x}-\mathbf{y}\rangle + \frac{\rho_1}{2}\|\mathbf{x}-\mathbf{y}\|_2^2 + \langle \mathbf{w}, \mathbf{x}-\mathbf{z}\rangle + \frac{\rho_2}{2}\|\mathbf{x}-\mathbf{z}\|_2^2 \end{aligned} \tag{3.4}$$

The ADMM consists of the following five steps:

$$\begin{cases} \mathbf{x}^{(k+1)} := \arg\min_{\mathbf{x}} L_{\rho_1,\rho_2}(\mathbf{x},\mathbf{y}^{(k)},\mathbf{z}^{(k)};\mathbf{v}^{(k)},\mathbf{w}^{(k)}), \\ \mathbf{y}^{(k+1)} := \arg\min_{\mathbf{y}} L_{\rho_1,\rho_2}(\mathbf{x}^{(k+1)},\mathbf{y},\mathbf{z}^{(k)};\mathbf{v}^{(k)},\mathbf{w}^{(k)}), \\ \mathbf{z}^{(k+1)} := \arg\min_{\mathbf{z}} L_{\rho_1,\rho_2}(\mathbf{x}^{(k+1)},\mathbf{y}^{(k+1)},\mathbf{z};\mathbf{v}^{(k)},\mathbf{w}^{(k)}), \\ \mathbf{v}^{(k+1)} := \mathbf{v}^{(k)} + \rho_1(\mathbf{x}^{(k+1)} - \mathbf{y}^{(k+1)}), \\ \mathbf{w}^{(k+1)} := \mathbf{w}^{(k)} + \rho_2(\mathbf{x}^{(k+1)} - \mathbf{z}^{(k+1)}). \end{cases} \tag{3.5}$$

The update for $\mathbf{x}$ is a projection to the affine space of $A\mathbf{x}=\mathbf{b}$,

$$\begin{aligned} \mathbf{x}^{(k+1)} &:= \arg\min_{\mathbf{x}} L_{\rho_1,\rho_2}(\mathbf{x},\mathbf{y}^{(k)},\mathbf{z}^{(k)};\mathbf{v}^{(k)},\mathbf{w}^{(k)}) \\ &= \arg\min_{\mathbf{x}} \left\{ \frac{\rho_1+\rho_2}{2} \left\| \mathbf{x} - \mathbf{f}^{(k)} \right\|_2^2 \quad \text{s.t.} \quad A\mathbf{x}=\mathbf{b} \right\} \\ &= \left(I - A^T(AA^T)^{-1}A\right)\mathbf{f}^{(k)} + A^T(AA^T)^{-1}\mathbf{b}, \end{aligned}$$

where

$$\mathbf{f}^{(k)} = \frac{\rho_1}{\rho_1+\rho_2}\left(\mathbf{y}^{(k)} - \frac{1}{\rho_1}\mathbf{v}^{(k)}\right) + \frac{\rho_2}{\rho_1+\rho_2}\left(\mathbf{z}^{(k)} - \frac{1}{\rho_2}\mathbf{w}^{(k)}\right). \tag{3.6}$$

As for the $\mathbf{y}$-subproblem, let $c^{(k)} = \|\mathbf{z}^{(k)}\|_1$ and $\mathbf{d}^{(k)} = \mathbf{x}^{(k+1)} + \frac{\mathbf{v}^{(k)}}{\rho_1}$, and the minimization subproblem reduces to

$$\mathbf{y}^{(k+1)} = \arg\min_{\mathbf{y}} \frac{c^{(k)}}{\|\mathbf{y}\|_2} + \frac{\rho_1}{2}\|\mathbf{y}-\mathbf{d}^{(k)}\|_2^2. \tag{3.7}$$

If $\mathbf{d}^{(k)}=0$, then any vector $\mathbf{y}$ with $\|\mathbf{y}\|_2 = \sqrt[3]{\frac{c^{(k)}}{\rho_1}}$ is a solution to the minimization problem. If $c^{(k)}=0$, then $\mathbf{y}=\mathbf{d}^{(k)}$ is the solution. Now we consider $\mathbf{d}^{(k)} \neq 0$ and $c^{(k)} \neq 0$. By taking derivative of the objective function with respect to $\mathbf{y}$, we obtain

$$\left(-\frac{c^{(k)}}{\|\mathbf{y}\|_2^3} + \rho_1\right)\mathbf{y} = \rho_1 \mathbf{d}^{(k)}.$$

As a result, there exists a positive number $\tau^{(k)} \geq 0$ such that $\mathbf{y} = \tau^{(k)}\mathbf{d}^{(k)}$. Given $\mathbf{d}^{(k)}$, we denote $\eta^{(k)} = \|\mathbf{d}^{(k)}\|_2$. For $\eta^{(k)} > 0$, finding $\mathbf{y}$ becomes a one-dimensional

search for the parameter $\tau^{(k)}$. In other words, if we take $D^{(k)} = \frac{c^{(k)}}{\rho_1(\eta^{(k)})^3}$, then $\tau^{(k)}$ is a root of

$$\underbrace{\tau^3 - \tau^2 - D^{(k)}}_{F(\tau)} = 0.$$

The cubic-root formula suggests that $F(\tau) = 0$ has only one real root, which can be found by the following closed-form solution:

(3.8)
$$\tau^{(k)} = \frac{1}{3} + \frac{1}{3}\left(C^{(k)} + \frac{1}{C^{(k)}}\right) \text{ with } C^{(k)} = \sqrt[3]{\frac{27D^{(k)} + 2 + \sqrt{(27D^{(k)}+2)^2 - 4}}{2}}.$$

In summary, we have

$$\mathbf{y}^{(k+1)} = \begin{cases} \mathbf{e}^{(k)} & \mathbf{d}^{(k)} = 0, \\ \tau^{(k)}\mathbf{d}^{(k)} & \mathbf{d}^{(k)} \neq 0, \end{cases} \tag{3.9}$$

where $\mathbf{e}^{(k)}$ is a random vector with the $L_2$ norm to be $\sqrt[3]{\frac{c^{(k)}}{\rho_1}}$.

Finally, the ADMM update for $\mathbf{z}$ is

$$\mathbf{z}^{(k+1)} = \mathbf{shrink}\left(\mathbf{x}^{(k+1)} + \frac{\mathbf{w}^{(k)}}{\rho_2}, \frac{1}{\rho_2\|\mathbf{y}^{(k+1)}\|_2}\right), \tag{3.10}$$

where **shrink** is often referred to as *soft shrinkage* operator,

$$\mathbf{shrink}(\mathbf{v}, \mu)_i = \text{sign}(v_i)\max\left(|v_i| - \mu, 0\right), \quad i = 1, 2, \ldots, n. \tag{3.11}$$

We summarize the ADMM algorithm for solving the $L_1/L_2$ minimization problem in Algorithm 3.1.

**Algorithm 3.1** The $L_1/L_2$ minimization via ADMM.

Input: $A \in \mathbb{R}^{m\times n}, \mathbf{b} \in \mathbb{R}^{m\times 1}$, Max and $\epsilon \in \mathbb{R}$
**while** $k <$ Max or $\|\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\|_2/\|\mathbf{x}^{(k)}\| > \epsilon$ **do**
$\mathbf{x}^{(k+1)} = \left(I - A^T(AA^T)^{-1}A\right)\mathbf{f}^{(k)} + A^T(AA^T)^{-1}\mathbf{b}$
$\mathbf{y}^{(k+1)} = \begin{cases} \mathbf{e}^{(k)} & \mathbf{d}^{(k)} = 0 \\ \tau^{(k)}\mathbf{d}^{(k)} & \mathbf{d}^{(k)} \neq 0 \end{cases}$
$\mathbf{z}^{(k+1)} = \mathbf{shrink}\left(\mathbf{x}^{(k+1)} + \frac{\mathbf{w}^{(k)}}{\rho_2}, \frac{1}{\rho_2\|\mathbf{y}^{(k+1)}\|_2}\right)$
$\mathbf{v}^{(k+1)} = \mathbf{v}^{(k)} + \rho_1(\mathbf{x}^{(k+1)} - \mathbf{y}^{(k+1)})$
$\mathbf{w}^{(k+1)} = \mathbf{w}^{(k)} + \rho_2(\mathbf{x}^{(k+1)} - \mathbf{z}^{(k+1)})$
$k = k + 1$
**end while**
**return** $\mathbf{x}^{(k)}$

*Remark* 1. We can precompute the matrix $I - A^T(AA^T)^{-1}A$ and the vector $A^T(AA^T)^{-1}\mathbf{b}$ in Algorithm 3.1. The complexity is $O(m^2n)$ for the precomputation including the matrix-matrix multiplication and Cholesky decomposition for solving the linear system. In each iteration, we need to do matrix-vector multiplication for the $\mathbf{x}$-subproblem, which is in the order of $O(n^2)$. In the $\mathbf{y}$-subproblem, the rooting finding is a one-dimensional search, whose cost can be neglected. The $\mathbf{z}$-subproblem is a pixelwise shrinkage operation and only takes $O(n)$. In summary, the computation complexity for each iteration is $O(n^2)$. We can consider the parallel computing to further speed up thanks to the separation of the $\mathbf{z}$-subproblem.

**3.2. $L_1/L_2$ with box constraint.** The $L_1/L_2$ model has an intrinsic drawback that tends to produce one erroneously large coefficient while suppressing the other nonzero elements, under which case the ratio is reduced. To compensate for this drawback, it is helpful to incorporate a box constraint if we know lower/upper bounds of the underlying signal a priori. Specifically, we propose

$$\min_{\mathbf{x}\in\mathbb{R}^n} \left\{ \frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_2} + I_0(A\mathbf{x}-\mathbf{b}) \,\middle|\, \mathbf{x}\in[c,d] \right\}, \tag{3.12}$$

which is referred to as $L_1/L_2$-box. Similar to (3.3), we look at the following form that enforces the box constraint on variable $\mathbf{z}$:

$$\min_{\mathbf{x},\mathbf{y},\mathbf{z}} \left\{ \frac{\|\mathbf{z}\|_1}{\|\mathbf{y}\|_2} + I_0(A\mathbf{x}-\mathbf{b}) \right\} \quad \text{s.t.} \quad \mathbf{x}=\mathbf{y}, \quad \mathbf{x}=\mathbf{z}, \quad \mathbf{z}\in[c,d]. \tag{3.13}$$

The only change we need to make by adapting Algorithm 3.1 to the $L_1/L_2$-box is the $\mathbf{z}$ update. The $\mathbf{z}$-subproblem in (3.5) with the box constraint is

$$\min_{\mathbf{z}} \frac{1}{\|\mathbf{y}^{(k+1)}\|_2}\|\mathbf{z}\|_1 + \frac{\rho_2}{2}\left\|\mathbf{x}^{(k+1)} - \mathbf{z} + \frac{1}{\rho_2}\mathbf{w}^{(k)}\right\|_2^2 \quad \text{s.t.} \quad \mathbf{z}\in[c,d]. \tag{3.14}$$

For a convex problem (3.14) involving the $L_1$ norm, it has a closed-form solution given by the soft shrinkage, followed by projection to the interval $[c,d]$. In particular, simple calculations show that

$$z_i^{(k+1)} = \min\left\{\max(\hat{z}_i, c), d\right\}, \quad i=1,2,\ldots,n, \tag{3.15}$$

where $\hat{\mathbf{z}} = \mathbf{shrink}\,(\mathbf{r},\nu)$, $\mathbf{r} = \mathbf{x}^{(k+1)} + \frac{\mathbf{w}^{(k)}}{\rho_2}$, and $\nu = \frac{1}{\rho_2\|\mathbf{y}^{(k+1)}\|_2}$. If the box constraint $[c,d]$ is symmetric, i.e., $c=-d$ and $d>0$, it follows from [2] that the update for $\mathbf{z}$ can be expressed as

$$z_i^{(k+1)} = \operatorname{sign}(v_i)\min\left\{\max(|r_i|-\nu,0), d\right\}, \quad i=1,2,\ldots,n. \tag{3.16}$$

*Remark* 2. The existing literature on the ADMM convergence [17, 19, 24, 33, 40, 41, 42] requires the existence of one separable function in the objective function, whose gradient is Lipschitz continuous. Obviously, $L_1/L_2$ does not satisfy this assumption no matter with or without the box constraint. Therefore, we have difficulties in analyzing the convergence theoretically. Instead, we show the convergence empirically in section 4 by plotting residual errors and objective functions, which gives strong supports for theoretical analysis in the future.

**3.3. $L_1/L_2$ on the gradient.** We adapt the $L_1/L_2$ model to apply on the gradient, which enables us to deal with imaging applications. Let $u\in\mathbb{R}^{n\times m}$ be an underlying image of size $n\times m$. Denote $A$ as a linear operator that models a certain degradation process to obtain the measured data $f$. For example, $A$ can be a subsampling operator in the frequency domain, and recovering $u$ from $f$ is called MRI reconstruction. In short, the proposed gradient model is given by

$$\min_{u\in\mathbb{R}^{n\times m}} \frac{\|\nabla u\|_1}{\|\nabla u\|_2} \quad \text{s.t.} \quad Au=f, \; u\in[0,1], \tag{3.17}$$

where $\nabla$ denotes discrete gradient operator $\nabla u := \{[u_{ij}-u_{(i+1)j}]_{i=1}^n\}_{j=1}^m, \{[u_{ij}-u_{i(j+1)}]_{j=1}^m\}_{i=1}^n$ with periodic boundary condition; hence, the model is referred to as $L_1/L_2$-grad. Note that the box constraint $0\le u\le 1$ is a reasonable assumption in the MRI reconstruction problem.

To solve for (3.17), we introduce three auxiliary variables $\mathbf{d}, \mathbf{h}$, and $v$, leading to an equivalent problem:

$$\min_{u\in\mathbb{R}^{n\times m}} \frac{\|\mathbf{d}\|_1}{\|\mathbf{h}\|_2} \quad \text{s.t. } Au = f,\ \mathbf{d} = \nabla u,\ \mathbf{h} = \nabla u, u = v, 0 \le v \le 1. \tag{3.18}$$

Note that we denote $\mathbf{d}$ and $\mathbf{h}$ in bold to indicate that they have two components corresponding to both $x$ and $y$ derivatives. The augmented Lagrangian is expressed as

$$\begin{aligned}\mathcal{L}(u,\mathbf{d},\mathbf{h},v;w,\mathbf{b}_1,\mathbf{b}_2,e) = \frac{\|\mathbf{d}\|_1}{\|\mathbf{h}\|_2} + \frac{\lambda}{2}\|Au-f-w\|_2^2 + \frac{\rho_1}{2}\|\mathbf{d}-\nabla u-\mathbf{b}_1\|_2^2 \\ + \frac{\rho_2}{2}\|\mathbf{h}-\nabla u-\mathbf{b}_2\|_2^2 + \frac{\rho_3}{2}\|v-u-e\|_2^2 + I_{[0,1]}(v),\end{aligned} \tag{3.19}$$

where $w, \mathbf{b}_1, \mathbf{b}_2, e$ are dual variables and $\lambda, \rho_1, \rho_2, \rho_3$ are positive parameters. The updates for $\mathbf{d}, \mathbf{h}$ are the same as Algorithm 3.1. Specifically for $\mathbf{h}$, we consider $D^{(k)} = \frac{\|\mathbf{d}\|_1}{\rho_2\|\nabla u^{(k+1)}+\mathbf{g}^{(k)}\|_2^2}$, and hence $\tau^{(k)}$ is the root of the same polynomial as in (3.8). By taking derivative of (3.19) with respect to $u$, we can obtain the $u$-update, i.e.,

$$\begin{aligned}u^{(k+1)} = \left(\lambda A^TA - (\rho_1+\rho_2)\triangle + \rho_3 I\right)^{-1}\Big(\lambda A^T(f+w^{(k)}) + \rho_1\nabla^T(\mathbf{d}^{(k)}-\mathbf{b}_1^{(k)}) \\ + \rho_2\nabla^T(\mathbf{h}^{(k)}-\mathbf{b}_2^{(k)}) + \rho_3(v^{(k)}-e^{(k)})\Big).\end{aligned} \tag{3.20}$$

Note for certain operator $A$, the inverse in the $u$-update (3.20) can be computed efficiently via the fast Fourier transform. The $v$-subproblem is a projection to an interval $[0,1]$, i.e.,

$$v_{ij}^{(k+1)} = \min\left\{\max(u_{ij}^{(k+1)} + e_{ij}^{(k)}, 0), 1\right\}, \quad i = 1,2,\dots,n, j = 1,2,\dots,m. \tag{3.21}$$

In summary, we present the ADMM algorithm for the $L_1/L_2$-grad model in Algorithm 3.2.

**Algorithm 3.2** The $L_1/L_2$-grad minimization via ADMM.

Input: $f \in \mathbb{R}^{n\times m}$, $A$, Max and $\epsilon \in \mathbb{R}$.
**while** $k <$ Max or $\|u^{(k)} - u^{(k-1)}\|_2/\|u^{(k)}\| > \epsilon$ **do**
  Solve $u^{(k+1)}$ via (3.20)
  Solve $v^{(k+1)}$ via (3.21)
  $\mathbf{h}^{(k+1)} = \begin{cases}\mathbf{e}^{(k)} & \nabla u^{(k+1)} + \mathbf{g}^{(k)} = 0, \\ \tau^{(k)}\left(\nabla u^{(k+1)} + \mathbf{g}^{(k)}\right) & \nabla u^{(k+1)} + \mathbf{g}^{(k)} \neq 0.\end{cases}$
  $\mathbf{d}^{(k+1)} = \mathbf{shrink}\left(\nabla u^{(k+1)} + \mathbf{b}^{(k)}, \frac{1}{\rho_1\|\mathbf{h}^{(k+1)}\|_2}\right)$
  $\mathbf{b}^{(k+1)} = \mathbf{b}^{(k)} + \nabla u^{(k+1)} - \mathbf{d}^{(k+1)}$
  $\mathbf{g}^{(k+1)} = \mathbf{g}^{(k)} + \nabla u^{(k+1)} - \mathbf{h}^{(k+1)}$
  $w^{(k+1)} = w^{(k)} + f - Au^{(k+1)}$
  $e^{(k+1)} = e^{(k)} + u^{(k+1)} - v^{(k+1)}$
  $k = k+1$
**end while**
**return** $u^{(k)}$

**4. Numerical experiments.** In this section, we carry out a series of numerical tests to demonstrate the performance of the proposed $L_1/L_2$ models together with its corresponding algorithms. All the numerical experiments are conducted on a standard desktop with CPU (Intel i7-6700, 3.4GHz) and MATLAB 9.2 (R2017a).

We consider two types of sensing matrices: One is called an oversampled discrete cosine transform (DCT), and the other is a Gaussian matrix. Specifically for the oversampled DCT, we follow the works of [14, 26, 45] to define $A = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n] \in \mathbb{R}^{m\times n}$ with

$$\mathbf{a}_j := \frac{1}{\sqrt{m}} \cos\left(\frac{2\pi \mathbf{w} j}{F}\right), \quad j = 1, \ldots, n, \tag{4.1}$$

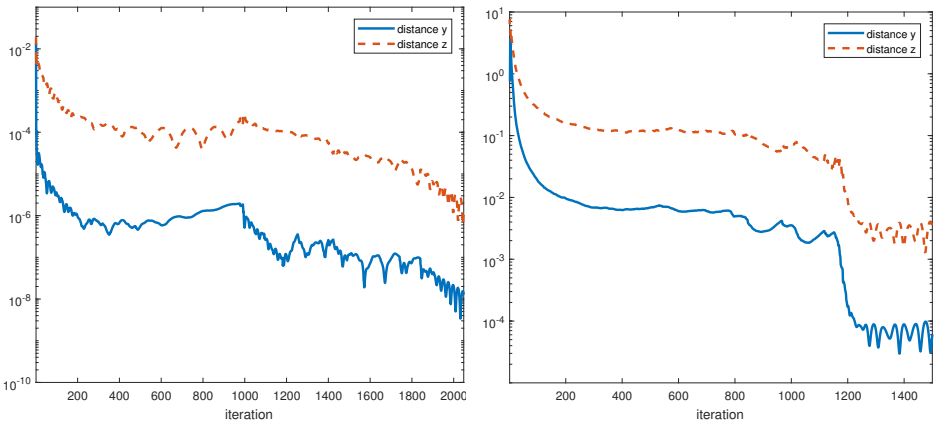
where $\mathbf{w}$ is a random vector uniformly distributed in $[0, 1]^m$ and $F \in \mathbb{R}_+$ controls the coherence in a way that a larger value of $F$ yields a more coherent matrix. In addition, we use $\mathcal{N}(\mathbf{0}, \Sigma)$ (the multivariable normal distribution) to generate a Gaussian matrix, where the covariance matrix is $\Sigma = \{(1-r)*I(i = j)+r\}_{i,j}$ with a positive parameter $r$. This type of matrix is used in the TL1 paper [47], which mentioned that a larger $r$ value indicates a more difficult problem in sparse recovery. Throughout the experiments, we consider sensing matrices of size $64 \times 1024$. The ground truth $\mathbf{x} \in \mathbb{R}^n$ is simulated as an $s$-sparse signal, where $s$ is the total number of nonzero entries. The support of $\mathbf{x}$ is a random index set, and the values of nonzero elements follow a Gaussian normal distribution, i.e., $(\mathbf{x}_s)_i \sim \mathcal{N}(0, 1), \quad i = 1, 2, \ldots, s$. We then normalize the ground-truth signal to have maximum magnitude as 1 so that we can examine the performance of an additional $[-1, 1]$ box constraint.

Due to the nonconvex nature of the proposed $L_1/L_2$ model, the initial guess $\mathbf{x}^{(0)}$ is very important and should be well chosen. A typical choice is the $L_1$ solution (1.2), which is used here. We adopt a commercial optimization software called Gurobi [32] to minimize the $L_1$ norm via linear programming for the sake of efficiency. The stopping criterion is when the relative error of $\mathbf{x}^{(k)}$ to $\mathbf{x}^{(k-1)}$ is smaller than $10^{-8}$ or the iterative number exceeds $10n$.

**4.1. Algorithmic behaviors.** We empirically demonstrate the convergence of the proposed ADMM algorithms in Figure 2. Specifically, we examine the $L_1/L_2$ minimization problem (3.1), where the sensing matrix is an oversampled DCT matrix with $F = 10$ and the ground-truth sparse vector has 12 nonzero elements. We also study the MRI reconstruction from 6 radical lines as a particular sparse gradient problem that involves the $L_1/L_2$-grad minimization of (3.17) by Algorithm 3.2.

There are two auxiliary variables $\mathbf{y}$ and $\mathbf{z}$ in $L_1/L_2$ such that $\mathbf{x} = \mathbf{y} = \mathbf{z}$, while two auxiliary variables $\mathbf{d}, \mathbf{h}$ are in $L_1/L_2$-grad for $\nabla u = \mathbf{d} = \mathbf{h}$. We show in the top row of Figure 2 the values of $\|\mathbf{x}^{(k)} - \mathbf{y}^{(k)}\|_2$ and $\|\mathbf{x}^{(k)} - \mathbf{z}^{(k)}\|_2$ as well as $\|\nabla u^{(k)} - \mathbf{d}^{(k)}\|_2$ and $\|\nabla u^{(k)} - \mathbf{h}^{(k)}\|_2$; all are plotted with respect to the iteration counter $k$. The bottom row of Figure 2 is for objective functions, i.e., $\|\mathbf{x}^{(k)}\|_1/\|\mathbf{x}^{(k)}\|_2$ and $\|\nabla u^{(k)}\|_1/\|\nabla u^{(k)}\|_2$ for $L_1/L_2$ and $L_1/L_2$-grad, respectively. All the plots in Figure 2 decrease rapidly with respect to iteration counters, which serve as heuristic evidence of algorithmic convergence. On the other hand, the objective functions in Figure 2 look oscillatory. This phenomenon implies difficulties in theoretically proving the convergence, as one key step in the convergence proof requires showing that the objective function decreases monotonically [3, 42].

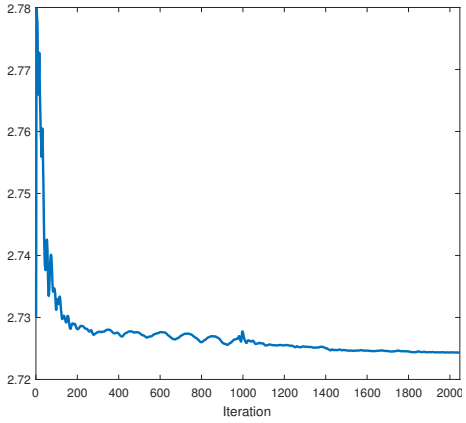
**4.2. Comparison on various models.** We now compare the proposed $L_1/L_2$ approach with other sparse recovery models: $L_1$, $L_p$ [9], $L_1$-$L_2$ [45, 26], and TL1 [47]. We choose $p = 0.5$ for $L_p$ and $a = 1$ for TL1. The initial guess for all the



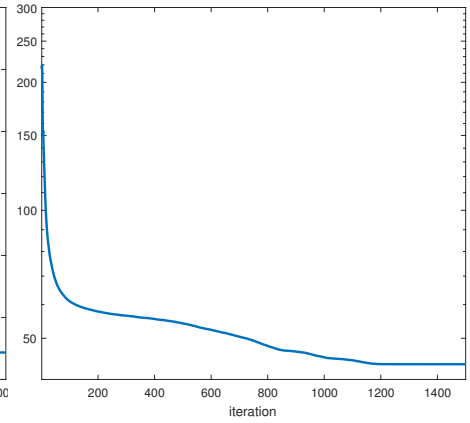

(a) Residual errors in $L1/L2$




(b) Residual errors in $L1/L2$-grad




(c) Objective functions of $L1/L2$

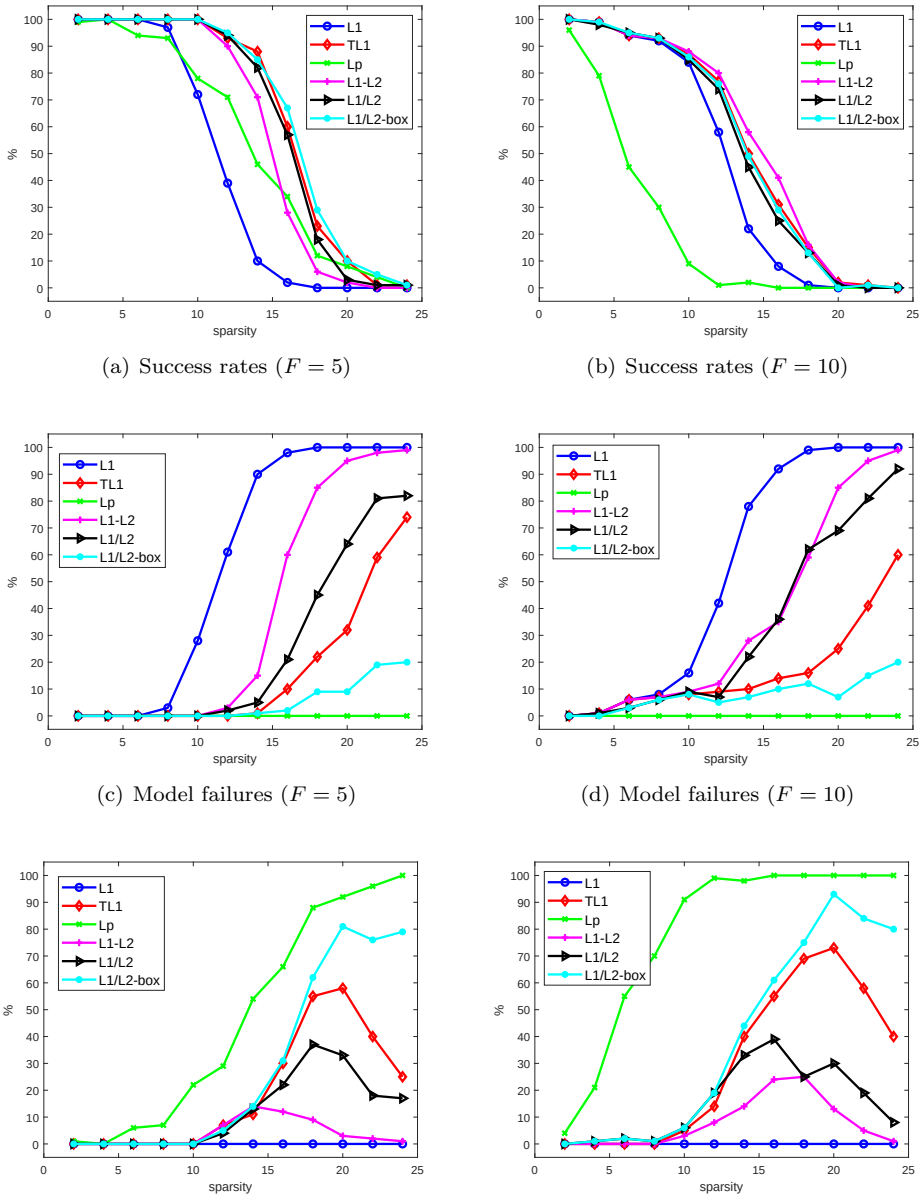



(d) Objective functions of $L1/L2$-grad

FIG. 2. *Plots of residual errors and objective functions for empirically demonstrating the convergence of the proposed algorithms: $L_1/L_2$ in signal processing and $L_1/L_2$-grad with a box constraint for MRI reconstruction.*

algorithms is the solution of the $L_1$ model. Both $L_1$-$L_2$ and TL1 are solved via the difference of convex algorithm (DCA) with the same stopping criterion as $L_1/L_2$, i.e., $\frac{\|\mathbf{x}^{(k)}-\mathbf{x}^{(k-1)}\|_2}{\|\mathbf{x}^{(k)}\|_2} \leq 10^{-8}$. As for $L_p$, we follow the default setting in [9].

We evaluate the performance of sparse recovery in terms of *success rate*, defined as the number of successful trials over the total number of trials. A success is declared if the relative error of the reconstructed solution $\mathbf{x}^*$ to the ground truth $\mathbf{x}$ is less than $10^{-3}$, i.e., $\frac{\|\mathbf{x}^*-\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq 10^{-3}$. We further categorize the failure of not recovering the ground truth as *model failure* and *algorithm failure*. In particular, we compare the objective function $\mathcal{F}(\cdot)$ at the ground truth $\mathbf{x}$ and at the reconstructed solution $\mathbf{x}^*$. If $\mathcal{F}(\mathbf{x}) > \mathcal{F}(\mathbf{x}^*)$, then $\mathbf{x}$ is not a global minimizer of the model, in which case we call it *model failure*. On the other hand, $\mathcal{F}(\mathbf{x}) < \mathcal{F}(\mathbf{x}^*)$ implies that the algorithm does not reach a global minimizer, which is referred to as *algorithm failure*. Although this type of analysis is not deterministic, it sheds some light on which direction to improve: model or algorithm. For example, it was reported in [30] that $L_1$ has the highest model-failure rates, which justifies the need for nonconvex models.




(a) Success rates ($F = 5$)

(b) Success rates ($F = 10$)

(c) Model failures ($F = 5$)

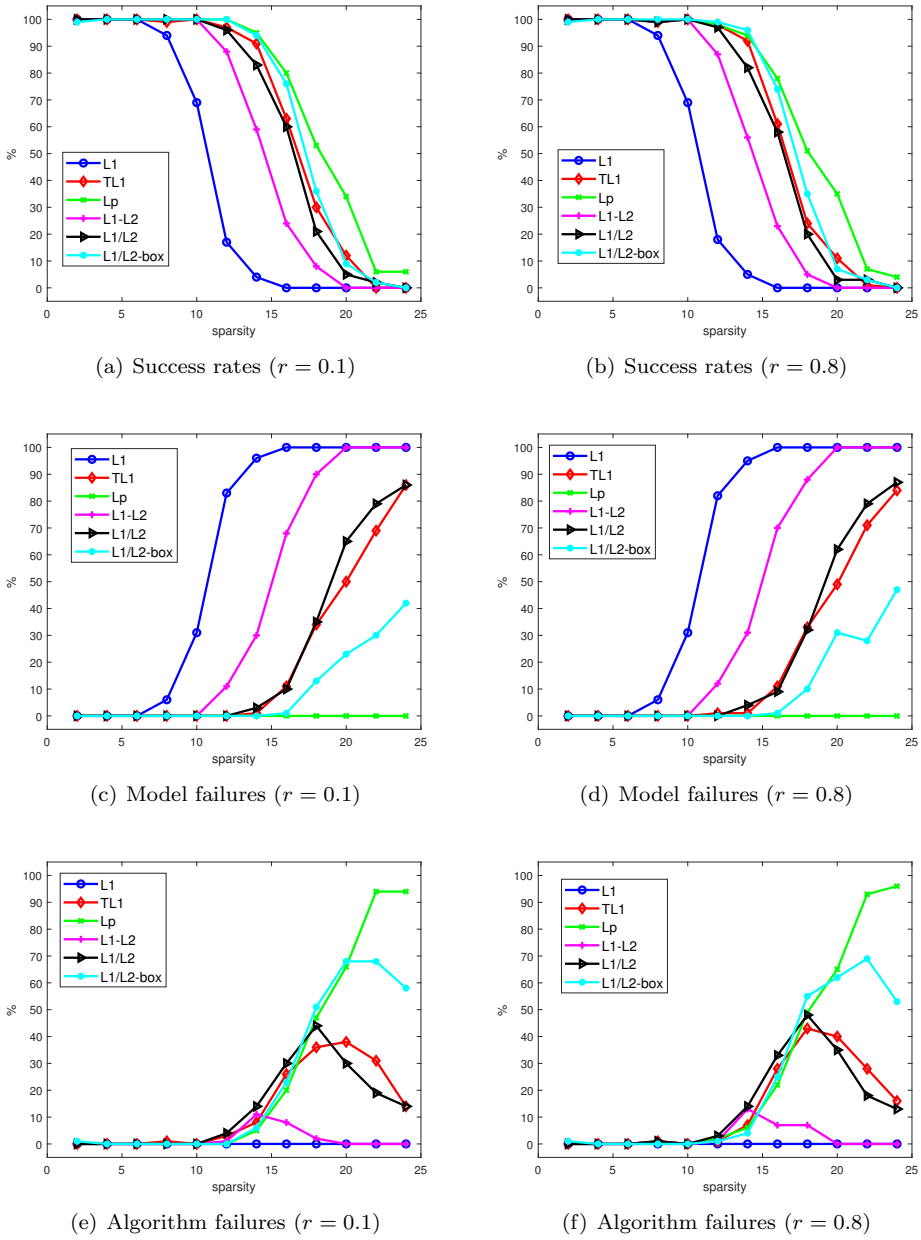
(d) Model failures ($F = 10$)

(e) Algorithm failures ($F = 5$)

(f) Algorithm failures ($F = 10$)

FIG. 3. *Success rates, model failures, and algorithm failures for* 6 *algorithms in the case of oversampled DCT matrices.*

In Figure 3, we examine two coherence levels: $F = 5$ corresponds to relatively low coherence and $F = 20$ to higher coherence. The success rates of various models reveal that $L_1/L_2$-box performs the best at $F = 5$ and is comparable to $L_1$-$L_2$ for the highly coherent case of $F = 20$. We look at a Gaussian matrix with $r = 0.1$ and $r = 0.8$ in Figure 4, both of which exhibit very similar performance of various models. In particular, the $L_p$ model gives the best results for the Gaussian case,




(a) Success rates ($r = 0.1$) (b) Success rates ($r = 0.8$)

(c) Model failures ($r = 0.1$) (d) Model failures ($r = 0.8$)

(e) Algorithm failures ($r = 0.1$) (f) Algorithm failures ($r = 0.8$)

FIG. 4. *Success rates, model failures, and algorithm failures for* 6 *algorithms in the Gaussian matrix case.*

which is consistent with the literature [44, 26]. The proposed model of $L_1/L_2$-box is the second best for such incoherent matrices.

By comparing $L_1/L_2$ with and without box among the plots for success rates and model failures, we can draw the conclusion that the box constraint can mitigate the inherent drawback of the $L_1/L_2$ model, thus improving the recovery rates. In addition, $L_1/L_2$ is the second lowest in terms of model failure rates, and simply adding a box

TABLE 1
*Computation time (sec.) in 5 algorithms.*

(a) DCT matrix

| $F = 5$ | | | | | | | |
|---|---|---|---|---|---|---|---|
| Sparsity | 2 | 6 | 10 | 14 | 18 | 22 | mean |
| TL1 | **0.049** | **0.050** | **0.066** | **0.207** | 0.618 | 0.795 | 0.298 |
| $L_p$ | 0.061 | 0.137 | 0.209 | 0.355 | 0.515 | 0.565 | 0.307 |
| $L_1$-$L_2$ | **0.049** | **0.050** | 0.071 | 0.260 | 0.550 | 0.625 | 0.267 |
| $L_1/L_2$ | 0.276 | 0.279 | 0.311 | 0.353 | 0.358 | 0.366 | 0.324 |
| $L_1/L_2$-box | 0.102 | 0.183 | 0.247 | 0.313 | **0.325** | **0.332** | **0.250** |
| $F = 10$ | | | | | | | |
| Sparsity | 2 | 6 | 10 | 14 | 18 | 22 | mean |
| TL1 | **0.048** | **0.069** | **0.092** | 0.330 | 0.654 | 0.755 | 0.325 |
| $L_p$ | 0.094 | 0.254 | 0.423 | 0.472 | 0.530 | 0.534 | 0.385 |
| $L_1$-$L_2$ | 0.049 | 0.070 | 0.093 | **0.272** | 0.598 | 0.677 | 0.293 |
| $L_1/L_2$ | 0.263 | 0.272 | 0.295 | 0.340 | 0.355 | 0.356 | 0.314 |
| $L_1/L_2$-box | 0.090 | 0.179 | 0.239 | 0.301 | **0.324** | **0.322** | **0.243** |

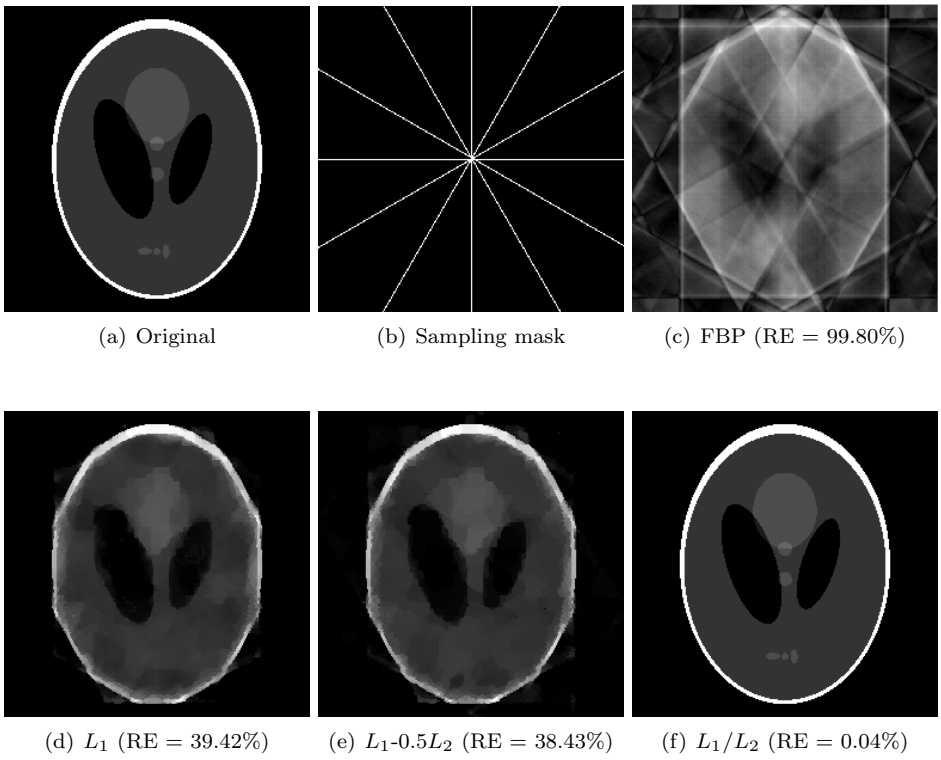
(b) Gaussian matrix

| $r = 0.1$ | | | | | | | |
|---|---|---|---|---|---|---|---|
| Sparsity | 2 | 6 | 10 | 14 | 18 | 22 | mean |
| TL1 | **0.070** | **0.069** | **0.117** | 0.295 | 1.101 | 1.633 | 0.548 |
| $L_p$ | 0.079 | 0.128 | 0.229 | **0.261** | **0.742** | 1.218 | **0.443** |
| $L_1$-$L_2$ | **0.070** | **0.069** | 0.122 | 0.399 | 0.877 | **1.161** | 0.450 |
| $L_1/L_2$ | 0.864 | 0.866 | 1.175 | 1.130 | 1.210 | 1.458 | 1.117 |
| $L_1/L_2$-box | 0.324 | 0.625 | 1.039 | 1.060 | 1.146 | 1.385 | 0.930 |
| $r = 0.8$ | | | | | | | |
| Sparsity | 2 | 6 | 10 | 14 | 18 | 22 | mean |
| TL1 | **0.050** | **0.053** | **0.071** | 0.239 | 0.613 | 0.750 | 0.296 |
| $L_p$ | 0.061 | 0.094 | 0.140 | **0.207** | 0.426 | 0.613 | 0.257 |
| $L_1$-$L_2$ | 0.051 | 0.054 | 0.077 | 0.306 | 0.497 | 0.576 | 0.260 |
| $L_1/L_2$ | 0.277 | 0.277 | 0.324 | 0.358 | 0.364 | 0.363 | 0.327 |
| $L_1/L_2$-box | 0.102 | 0.192 | 0.265 | 0.321 | **0.332** | **0.327** | **0.256** |

constraint also increases the occurrence of algorithm failure compared to the no-box version. These two observations suggest a need to further improve on algorithms of minimizing $L_1/L_2$.

Finally, we provide the computation time for all the competing algorithms in Table 1 with the shortest time in each case highlighted in bold. The time for the $L_1$ method is not included, as all the other methods use the $L_1$ solution as an initial guess. It is shown that TL1 is the fastest for relatively lower sparsity levels and that the proposed $L_1/L_2$-box is the most efficient at higher sparsity levels. The computational times for all these methods seem consistent with DCT and Gaussian matrices.

**4.3. MRI reconstruction.** As a proof-of-concept example, we study an MRI reconstruction problem [28] to compare the performance of $L_1$, $L_1$-$L_2$, and $L_1/L_2$ on the gradient. The $L_1$ on the gradient is the celebrated TV model [36], while $L_1$-$L_2$ on the gradient was recently proposed in [27]. We use a standard Shepp–Logan phantom as a testing image, as shown in Figure 5(a). The MRI measurements are obtained by several radical lines in the frequency domain (i.e., after taking the Fourier transform); an example of such a sampling scheme using 6 lines is shown in Figure 5(b). As this paper focuses on the constrained formulation, we do not consider noise, following the same setting as in the previous works [45, 27]. Since all the



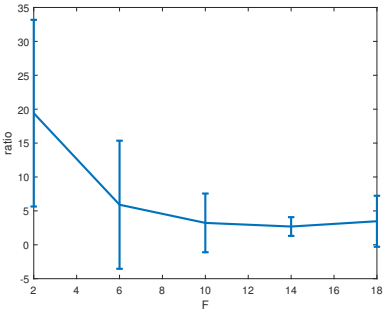
(a) Original (b) Sampling mask (c) FBP (RE = 99.80%)

(d) $L_1$ (RE = 39.42%) (e) $L_1$-0.5$L_2$ (RE = 38.43%) (f) $L_1/L_2$ (RE = 0.04%)

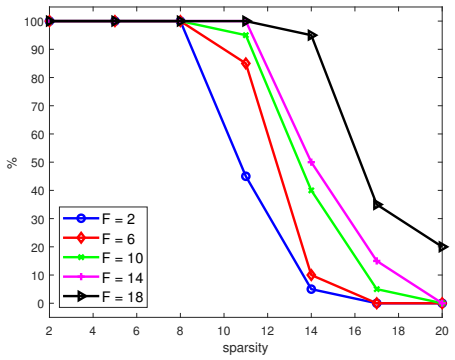
FIG. 5. *MRI reconstruction results from* 6 *radical lines in the frequency domain (*2.57% *measurements). The relative errors (RE) are provided for each method.*

competing methods ($L_1$, $L_1$-0.5$L_2$, and $L_1/L_2$) yield an exact recovery with 8 radical lines, with accuracy in the order of $10^{-8}$, we present the reconstructions results of 6 radical lines in Figure 5, which illustrates that the ratio model ($L_1/L_2$) gives much better results than the difference model ($L_1$-0.5$L_2$). Figure 5 also includes quantitative measures of the performance by relative error (RE) between the reconstructed and ground-truth images, which shows significantly improvement of the proposed $L_1/L_2$-grad over a classic method in MRI reconstruction, called filter-back projection (FBP), and two recent works of using $L_1$ [15] and $L_1$-0.5$L_2$ [27] on the gradient. Note that the state-of-the-art methods in MRI reconstruction are from [18, 30], who have reported exact recovery from 7 radical lines.

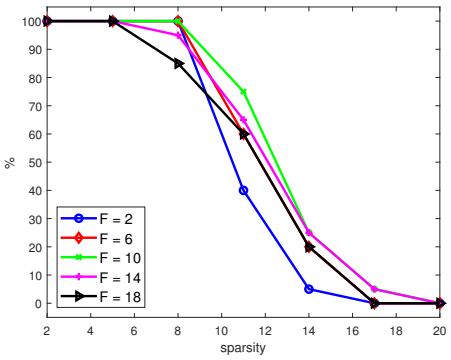
**5. Empirical validations.** A review article [7] indicated that two principles in CS are *sparsity* and *incoherence*, leading to an impression that a sensing matrix with smaller coherence is easier for sparse recovery. However, we observe through numerical results [25] (also given in Figure 6(b)) that a more coherent matrix gives higher recovery rates. This contradiction motivates us to collect empirical evidence regarding either proving or refusing whether coherence is relevant to sparse recovery. Here we examine one case of such evidence by minimizing the ratio of $L_1$ and $L_2$, which gives an upper bound for a sufficient condition of $L_1$ exact recovery; see (1.4). To avoid the trivial solution of $\mathbf{x} = \mathbf{0}$ to the problem of $\min_{\mathbf{x}}\{\frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_2} : A\mathbf{x} = \mathbf{0}\}$, we incorporate a sum-to-one constraint. In other words, we define an expanded matrix




(a) The mean of ratios




(b) Success rates with MS




(c) Success rates without MS

FIG. 6. *The use of* $\min_{\mathbf{x}}\{\frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_2} : A\mathbf{x} = \mathbf{0}\}$ *as an upper bound for the* $L_1$ *recovery.* (a)*: Plots of the mean of ratios over* 50 *realizations with the standard deviation indicated as vertical bars.* (b), (c)*: The success rates of* $L_1$ *recovery with and without minimum separation.*

$\tilde{A} = [A; \text{ones}(n, 1)]$ (following Matlab's notation) and an expanded vector $\tilde{\mathbf{b}} = [\mathbf{0}; 1]$. We then adapt the proposed method to solve for $\min_{\mathbf{x}}\{\frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_2} : \tilde{A}\mathbf{x} = \tilde{\mathbf{b}}\}$. In Figure 6(a), we plot the mean value of ratios from 50 random realizations of matrices $A$ at each coherence level (controlled by $F$), which shows that the ratio actually decreases[2] with respect to $F$. As the $L_0$ norm is bounded by the ratio (1.4), a smaller ratio indicates it is more difficult to recover the signals. Therefore, Figure 6(a) is consistent with the common belief in CS.

We postulate that an underlying reason of more coherent matrices giving better results is minimum separation (MS), as formally introduced in [5]. In Figure 6(b), we enforce the MS of two neighboring spikes to be 40, following the suggestion of $2F$ in [14] (we consider $F$ up to 20). In comparison, we also give the success rates of the $L_1$ recovery without any restrictions on MS in Figure 6(c). Note that we use the exactly same matrices in both cases (with and without MS). Figure 6(c) does not have a clear pattern regarding how coherence affects the exact recovery, which supports our hypothesis that MS plays an important role in sparse recovery. It will be our future work to analyze it throughly.

[2]We also observe that the ratio stagnates for larger $F$, which is probably because of instability of the proposed method when the matrix becomes more coherent.

**6. Conclusions and future works.** In this paper, we have studied a novel $L_1/L_2$ minimization to promote sparsity. Two main benefits of $L_1/L_2$ are scale invariant and parameter free. Two numerical algorithms based on the ADMM are formulated for the assumptions of sparse signals and sparse gradients, together with a variant of incorporating an additional box constraint. The experimental results demonstrate the performance of the proposed approaches in comparison to the state-of-the-art methods in sparse recovery and MRI reconstruction. As a by-product, minimizing the ratio also gives an empirical upper bound toward $L_1$'s exact recovery, which motivates further investigations on exact recovery theories. Other future works include algorithmic improvement and convergence analysis. In particular, it is shown in Table 1 and Figures 3 and 4 that $L_1/L_2$ is not as fast as the competing methods in CS and also has certain algorithmic failures, which call for a more robust and more efficient algorithm. In addition, we have provided heuristic evidence of the ADMM's convergence in Figure 2, and it will be interesting to analyze it theoretically.

**Appendix: Proof of Theorem 2.2.** In order to prove Theorem 2.2, we study the function

$$g(t) = \frac{\|\mathbf{x} + t\mathbf{v}\|_1^2}{\|\mathbf{x} + t\mathbf{v}\|_2^2}, \tag{A.1}$$

where $\mathbf{x} \neq \mathbf{0}$[3] and

$$\mathbf{v} \in \ker(A) \backslash \{\mathbf{0}\} \text{ with } \|\mathbf{v}\|_2 = 1. \tag{A.2}$$

Notice that the denominator of the function $g$ is nonzero for all $t \in \mathbb{R}$. Otherwise, we have $\mathbf{x} + t\mathbf{v} = \mathbf{0}$, and hence $A\mathbf{x} + A(t\mathbf{v}) = A(\mathbf{0})$. Since $A\mathbf{x} = \mathbf{b}$ and $A\mathbf{v} = \mathbf{0}$, we get $\mathbf{b} = \mathbf{0}$, which is a contradiction. Therefore, the function $g$ is continuous everywhere. Next, we introduce the following lemma to discuss the $L_1$ term in the numerator of $g$.

LEMMA A.1. *For any* $\mathbf{x} \in \mathbb{R}^n$ *and* $\mathbf{v} \in \mathbb{R}^n$ *satisfying* (A.2)*, denote* $S$ *as the support of* $\mathbf{x}$ *and* $t_0 := \min_{i \in S} |x_i|$*. We have*

$$\|\mathbf{x} + t\mathbf{v}\|_1 = \|\mathbf{x}\|_1 + t\sigma_t(\mathbf{v}) > 0 \quad \forall |t| < t_0, \tag{A.3}$$

*where*

$$\sigma_t(\mathbf{v}) = \sum_{i \in S} v_i sign(x_i) + sign(t)\|\mathbf{v}_{\bar{S}}\|_1. \tag{A.4}$$

*Proof.* Since $\mathbf{x} + t\mathbf{v} \neq \mathbf{0}$ for all $t \in \mathbb{R}$, we have $\|\mathbf{x} + t\mathbf{v}\|_1 > 0$. It follows from (A.2) that $|v_i| \leq 1 \, \forall i$. Then we get $\text{sign}(x_i + tv_i) = \text{sign}(x_i) \, \forall i \in S$, as $|tv_i| < |x_i|$ for

[3] We assume that $\mathbf{b} \neq 0$, so $\mathbf{x} = \mathbf{0}$ is not a solution to $A\mathbf{x} = \mathbf{b}$.

$|t| < t_0$. Therefore, we have

$$\begin{aligned}\|\mathbf{x}+t\mathbf{v}\|_1 &= \sum_{i\in S}|x_i+tv_i| + \sum_{i\notin S}|t||v_i| \\ &= \sum_{i\in S}(x_i+tv_i)\text{sign}(x_i) + |t|\|\mathbf{v}_{\bar{S}}\|_1 \\ &= \sum_{i\in S}x_i\text{sign}(x_i) + t\sum_{i\in S}v_i\text{sign}(x_i) + |t|\|\mathbf{v}_{\bar{S}}\|_1 \\ &= \|\mathbf{x}\|_1 + t\sum_{i\in S}v_i\text{sign}(x_i) + |t|\|\mathbf{v}_{\bar{S}}\|_1 \\ &= \|\mathbf{x}\|_1 + t\left(\sum_{i\in S}v_i\text{sign}(x_i) + \text{sign}(t)\|\mathbf{v}_{\bar{S}}\|_1\right),\end{aligned}$$

which implies (A.3), and hence Lemma A.1 holds. ∎

Notice that $\sigma_t(\mathbf{v})$ only relies on the sign of $t$; i.e., it is constant for $t > 0$ and $t < 0$. Therefore, $g(t)$ is differentiable on $0 < t < t_0$ and $-t_0 < t < 0$ (note that when $t \neq 0$, $g$ is not differentiable at the points where $x_i + tv_i = 0$). Some simple calculations lead to the derivative of $g$ for $0 < t < t_0$ and $-t_0 < t < 0$:

(A.5)

$$\begin{aligned}g'(t) &= \frac{d}{dt}\left(\frac{(\|\mathbf{x}\|_1 + t\sigma_t(\mathbf{v}))^2}{\|\mathbf{x}\|_2^2 + 2t\langle\mathbf{v}_S,\mathbf{x}\rangle + t^2\|\mathbf{v}\|_2^2}\right) \\ &= \frac{2\sigma_t(\mathbf{v})\left(\|\mathbf{x}\|_1 + t\sigma_t(\mathbf{v})\right)\left(\|\mathbf{x}\|_2^2 + 2t\langle\mathbf{v}_S,\mathbf{x}\rangle + t^2\|\mathbf{v}\|_2^2\right) - \left(2\langle\mathbf{v}_S,\mathbf{x}\rangle + 2t\|\mathbf{v}\|_2^2\right)\left(\|\mathbf{x}\|_1 + t\sigma_t(\mathbf{v})\right)^2}{\left(\|\mathbf{x}\|_2^2 + 2t\langle\mathbf{v}_S,\mathbf{x}\rangle + t^2\|\mathbf{v}\|_2^2\right)^2} \\ &= \frac{2\left(\|\mathbf{x}\|_1 + t\sigma_t(\mathbf{v})\right)\left[\sigma_t(\mathbf{v})\left(\|\mathbf{x}\|_2^2 + 2t\langle\mathbf{v}_S,\mathbf{x}\rangle + t^2\|\mathbf{v}\|_2^2\right) - \left(\langle\mathbf{v}_S,\mathbf{x}\rangle + t\|\mathbf{v}\|_2^2\right)\left(\|\mathbf{x}\|_1 + t\sigma_t(\mathbf{v})\right)\right]}{\left(\|\mathbf{x}\|_2^2 + 2t\langle\mathbf{v}_S,\mathbf{x}\rangle + t^2\|\mathbf{v}\|_2^2\right)^2} \\ &= \frac{2\left(\|\mathbf{x}\|_1 + t\sigma_t(\mathbf{v})\right)\left[\left(\sigma_t(\mathbf{v})\|\mathbf{x}\|_2^2 - \langle\mathbf{v}_S,\mathbf{x}\rangle\|\mathbf{x}\|_1\right) + \left(\sigma_t(\mathbf{v})\langle\mathbf{v}_S,\mathbf{x}\rangle - \|\mathbf{x}\|_1\|\mathbf{v}\|_2^2\right)t\right]}{\left(\|\mathbf{x}\|_2^2 + 2t\langle\mathbf{v}_S,\mathbf{x}\rangle + t^2\|\mathbf{v}\|_2^2\right)^2}.\end{aligned}$$

It follows from Lemma A.1 that the first term in the numerator of (A.5) is strictly positive, i.e., $\|\mathbf{x}\|_1 + t\sigma_t(\mathbf{v}) > 0$. Therefore, the sign of $g'$ depends on the second term in the numerator. We further introduce two lemmas (Lemmas A.2 and A.3) to study this term.

LEMMA A.2. *For any $\mathbf{x}, \mathbf{v} \in \mathbb{R}^n$ and $i \in [n]$, we have*

$$n\|\mathbf{x}\|_2^2 - |x_i|\|\mathbf{x}\|_1 \geq (n-1)\left(\sum_{j\neq i}x_j^2\right), \tag{A.6}$$

$$n\|\mathbf{v}\|_1\|\mathbf{x}\|_2^2 \geq \|\mathbf{x}\|_1|\langle\mathbf{v},\mathbf{x}\rangle|. \tag{A.7}$$

*Furthermore, if $\|\mathbf{x}\|_0 = s$, then the constant $n$ in the inequalities can be reduced to $s$.*

*Proof.* Simple calculations show that

$$\begin{aligned} n\|\mathbf{x}\|_2^2 - |x_i|\|\mathbf{x}\|_1 &= n\left(\sum_j x_j^2\right) - |x_i|\left(\sum_j |x_j|\right) \\ &= (n-1)\left(\sum_{j\neq i} x_j^2\right) + \sum_{j\neq i} x_j^2 + (n-1)x_i^2 - \sum_{j\neq i} |x_i||x_j| \\ &= (n-1)\left(\sum_{j\neq i} x_j^2\right) + \sum_{j\neq i} \left((|x_i| - |x_j|)^2 + |x_i||x_j|\right) \\ &\geq (n-1)\left(\sum_{j\neq i} x_j^2\right) \geq 0. \end{aligned} \tag{A.8}$$

Therefore, we have $\sum_i (n\|\mathbf{x}\|_2^2 - |x_i|\|\mathbf{x}\|_1)|v_i| \geq 0$, which implies that

$$n\|\mathbf{v}\|_1\|\mathbf{x}\|_2^2 \geq \|\mathbf{x}\|_1\left(\sum_i |x_i||v_i|\right) \geq \|\mathbf{x}\|_1 |\langle \mathbf{v}, \mathbf{x}\rangle|. \tag{A.9}$$

Similarly, we can reduce the constant $n$ to $s$ if we know $\|\mathbf{x}\|_0 = s$. ∎

LEMMA A.3. *Suppose that an $s$-sparse vector $\mathbf{x}$ satisfies $A\mathbf{x} = \mathbf{b}$ $(\mathbf{b} \neq 0)$ with its support on an index set $S$ and the matrix $A$ satisfies the sNSP of order $s$. Define*

$$t_1 := \inf_{\mathbf{v},t} \left\{ \frac{|\sigma_t(\mathbf{v})\|\mathbf{x}\|_2^2 - \langle \mathbf{v}_S, \mathbf{x}\rangle \|\mathbf{x}\|_1|}{|\sigma_t(\mathbf{v})\langle \mathbf{v}_S, \mathbf{x}\rangle - \|\mathbf{x}\|_1\|\mathbf{v}\|_2^2|} \,\middle|\, \mathbf{v} \in \ker(A), \|\mathbf{v}\|_2 = 1, t \neq 0 \right\}, \tag{A.10}$$

*where $\sigma_t(\mathbf{v})$ is defined as* (A.4). *Then $t_1 > 0$.*

*Proof.* For any $\mathbf{v} \in \ker(A)$ and $\|\mathbf{v}\|_2 = 1$, it is straightforward that

$$\begin{aligned} |\sigma_t(\mathbf{v})\langle \mathbf{v}_S, \mathbf{x}\rangle - \|\mathbf{x}\|_1\|\mathbf{v}\|_2^2| &\leq |\sigma_t(\mathbf{v})|\|\mathbf{v}\|_2\|\mathbf{x}\|_2 + \|\mathbf{x}\|_1\|\mathbf{v}\|_2^2 \\ &\leq \|\mathbf{v}\|_1\|\mathbf{v}\|_2\|\mathbf{x}\|_2 + \|\mathbf{x}\|_1\|\mathbf{v}\|_2^2 \\ &= \|\mathbf{v}\|_1\|\mathbf{x}\|_2 + \|\mathbf{x}\|_1 \\ &\leq \sqrt{n}\|\mathbf{x}\|_2 + \|\mathbf{x}\|_1 \end{aligned} \tag{A.11}$$

and

$$|\sigma_t(\mathbf{v})| \geq |\mathrm{sign}(t)\|\mathbf{v}_{\bar{S}}\|_1| - \left|\sum_{i\in S} v_i \mathrm{sign}(x_i)\right| \geq \|\mathbf{v}_{\bar{S}}\|_1 - \sum_{i\in S}|v_i| = \|\mathbf{v}_{\bar{S}}\|_1 - \|\mathbf{v}_S\|_1. \tag{A.12}$$

It follows from the sNSP that $\|\mathbf{v}_{\bar{S}}\|_1 \geq (s+1)\|\mathbf{v}_S\|_1$, thus leading to the following two inequalities:

$$\begin{aligned} &|\sigma_t(\mathbf{v})| \geq \|\mathbf{v}_{\bar{S}}\|_1 - \|\mathbf{v}_S\|_1 \geq s\|\mathbf{v}_S\|_1 \\ &|\sigma_t(\mathbf{v})| \geq \|\mathbf{v}_{\bar{S}}\|_1 - \|\mathbf{v}_S\|_1 \geq \left(1 - \frac{1}{s+1}\right)\|\mathbf{v}_{\bar{S}}\|_1 = \frac{s}{s+1}\|\mathbf{v}_{\bar{S}}\|_1. \end{aligned} \tag{A.13}$$

Next we will discuss two cases: $s = 1$ and $s > 1$.

(i) For $s = 1$. Without loss of generality, we assume the only nonzero element is $x_n \neq 0$, and hence we have

$$\begin{aligned}|\sigma_t(\mathbf{v})\|\mathbf{x}\|_2^2 - \langle \mathbf{v}_S, \mathbf{x}\rangle \|\mathbf{x}\|_1| &= |(v_n \text{sign}(x_n) + \text{sign}(t)\|\mathbf{v}_{\bar{S}}\|_1)x_n^2 - (v_n x_n)|x_n||\\ &= \|\mathbf{v}_{\bar{S}}\|_1 x_n^2.\end{aligned}$$

We further discuss two cases: $|v_n| \geq \frac{1}{\sqrt{n}}$ and $|v_n| < \frac{1}{\sqrt{n}}$. If $|v_n| \geq \frac{1}{\sqrt{n}}$, then $\|\mathbf{v}_{\bar{S}}\|_1 \geq (s+1)|v_n| \geq \frac{s+1}{\sqrt{n}}$, and hence

$$|\sigma_t(\mathbf{v})\|\mathbf{x}\|_2^2 - \langle \mathbf{v}_S, \mathbf{x}\rangle \|\mathbf{x}\|_1| = \|\mathbf{v}_{\bar{S}}\|_1\|\mathbf{x}\|_2^2 \geq \frac{s+1}{\sqrt{n}}\|\mathbf{x}\|_2^2. \tag{A.14}$$

If $|v_n| < \frac{1}{\sqrt{n}}$, then we have $\|\mathbf{v}_{\bar{S}}\|_1 \geq 1 - |v_n| = 1 - \frac{1}{\sqrt{n}} = \frac{\sqrt{n}-1}{\sqrt{n}}$ and

$$|\sigma_t(v)\|\mathbf{x}\|_2^2 - \langle \mathbf{v}_S, \mathbf{x}\rangle \|\mathbf{x}\|_1| = \|\mathbf{v}_{\bar{S}}\|_1\|\mathbf{x}\|_2^2 \geq \frac{\sqrt{n}-1}{\sqrt{n}}\|\mathbf{x}\|_2^2. \tag{A.15}$$

Combining (A.14) and (A.15), we have

$$t_1 \geq \frac{\min\left\{\frac{s+1}{\sqrt{n}}\|\mathbf{x}\|_2^2, \frac{\sqrt{n}-1}{\sqrt{n}}\|\mathbf{x}\|_2^2\right\}}{\sqrt{n}\|\mathbf{x}\|_2 + \|\mathbf{x}\|_1} > 0. \tag{A.16}$$

(ii) For $s > 1$. We split into two cases. The first case is $\forall j \in S,\ v_j < c$ (we will determine the value of $c$ shortly). As a result, we get $\|\mathbf{v}_S\|_1 < sc$ and $\|\mathbf{v}_{\bar{S}}\|_1 \geq 1 - sc$ since $\|\mathbf{v}\|_1 \geq \|\mathbf{v}\|_2 = 1$. Some simple calculations lead to

$$\begin{aligned}&\big|\sigma_t(\mathbf{v})\|\mathbf{x}\|_2^2 - \langle \mathbf{v}_S, \mathbf{x}\rangle \|\mathbf{x}\|_1\big|\\ &\geq |\sigma_t(\mathbf{v})|\|\mathbf{x}\|_2^2 - |\langle \mathbf{v}_S, \mathbf{x}\rangle|\|\mathbf{x}\|_1\\ &\geq \frac{s}{s+1}\|\mathbf{v}_{\bar{S}}\|_1\|\mathbf{x}\|_2^2 - \sum_{i\in S}|v_i||x_i|\|x\|_1 \quad \text{(based on (A.13))}\\ &\geq \frac{s}{s+1}(1-sc)\|\mathbf{x}\|_2^2 - \sum_{i\in S} c|x_i|\|\mathbf{x}\|_1\\ &= \frac{s}{s+1}(1-sc)\|\mathbf{x}\|_2^2 - c\|\mathbf{x}\|_1^2\\ &\geq \frac{s}{s+1}(1-su_0)\|\mathbf{x}\|_2^2 - sc\|\mathbf{x}\|_2^2\\ &= \frac{s}{s+1}\big(1-(2s+1)c\big)\|\mathbf{x}\|_2^2.\end{aligned}$$

If we choose $c = \frac{1}{2s+2}$, then the above quantity is larger than $\frac{s\|\mathbf{x}\|_2^2}{(s+1)(2s+2)} > 0$.

In the second case, we have that there exist $j \in S$ such that $v_j \geq c$, leading to

$$
\begin{aligned}
&\left|\sigma_t(\mathbf{v})\|\mathbf{x}\|_2^2 - \langle \mathbf{v}_S, \mathbf{x}\rangle \|\mathbf{x}\|_1\right| \\
&\quad \geq |\sigma_t(\mathbf{v})|\|\mathbf{x}\|_2^2 - |\langle \mathbf{v}_S, \mathbf{x}\rangle| \|\mathbf{x}\|_1 \\
&\quad \geq s\|\mathbf{v}_S\|_1\|\mathbf{x}\|_2^2 - \left(\sum_{i\in S} |x_i||v_i|\right)\|\mathbf{x}\|_1 \\
&\quad = \sum_{i\in S} \left(s\|\mathbf{x}\|_2^2 - |x_i|\|\mathbf{x}\|_1\right)|v_i| \\
&\quad \geq \left(s\|\mathbf{x}\|_2^2 - |x_j|\|\mathbf{x}\|_1\right)|v_j| \quad \text{(based on } \textit{Lemma} \text{ A.2)} \\
&\quad \geq c\left(s\|\mathbf{x}\|_2^2 - |x_j|\|\mathbf{x}\|_1\right) \\
&\quad \geq c(s-1)\sum_{i\neq j} x_i^2 \quad \text{(based on } \textit{Lemma} \text{ A.2)} \\
&\quad \geq c(s-1)\min_{j\in S}\sum_{i\neq j} x_i^2.
\end{aligned} \tag{A.17}
$$

These two cases guarantee that $t_1 > 0$, i.e.,

$$
t_1 \geq \frac{\min\left\{c(s-1)\min_{j\in S}\sum_{i\neq j} x_i^2,\ \frac{s\|\mathbf{x}\|_2^2}{(s+1)(2s+2)}\right\}}{\sqrt{n}\|\mathbf{x}\|_2 + \|\mathbf{x}\|_1} > 0. \tag{A.18}
$$

By (A.16) and (A.18), we get Lemma A.3. ∎

Now, we are ready to prove Theorem 2.2.

*Proof.* According to (A.3), the first term in the numerator is strictly positive, i.e., $\|\mathbf{x}\|_1 + t\sigma_t(\mathbf{v}) = \|\mathbf{x} + t\mathbf{v}\|_1 > 0\ \forall |t| < t_0$. As for the second one, there exists a positive number $t_1$ defined in Lemma A.3 such that

$$
\left|\sigma_t(\mathbf{v})\langle \mathbf{v}_S, \mathbf{x}\rangle - \|\mathbf{x}\|_1\|\mathbf{v}\|_2^2\right||t| < \left|\sigma_t(\mathbf{v})\|\mathbf{x}\|_2^2 - \langle \mathbf{v}_S, \mathbf{x}\rangle\|\mathbf{x}\|_1\right|
$$

for all $|t| < t_1$ and $\mathbf{v} \in \ker(A)$ with $\|\mathbf{v}\|_2 = 1$. Moreover, we have

$$
\begin{aligned}
&\operatorname{sign}\Big[\left(\sigma_t(\mathbf{v})\|\mathbf{x}\|_2^2 - \langle \mathbf{v}_S, \mathbf{x}\rangle\|\mathbf{x}\|_1 + \left(\sigma_t(\mathbf{v})\langle \mathbf{v}_S, \mathbf{x}\rangle - \|\mathbf{x}\|_1\|\mathbf{v}\|_2^2\right)t\Big] \\
&\quad = \operatorname{sign}\Big(\sigma_t(\mathbf{v})\|\mathbf{x}\|_2^2 - \langle \mathbf{v}_S, \mathbf{x}\rangle\|\mathbf{x}\|_1\Big).
\end{aligned}
$$

Letting $t^* = \min\{t_0, t_1\}$, we have for any $t \in (0, t^*)$ and $\mathbf{v} \neq \mathbf{0}$ that $\sigma_t(\mathbf{v}) > 0$ as

$$
\begin{aligned}
\sigma_t(\mathbf{v}) &= \sum_{i\in S} v_i \operatorname{sign}(x_i) + \operatorname{sign}(t)\|\mathbf{v}_{\bar{S}}\|_1 \\
&= \sum_{i\in S} v_i \operatorname{sign}(x_i) + \|\mathbf{v}_{\bar{S}}\|_1 \\
&\geq \|\mathbf{v}_{\bar{S}}\|_1 - \|\mathbf{v}_S\|_1 \\
&\geq \max\left\{s\|\mathbf{v}_S\|_1, \frac{s}{s+1}\|\mathbf{v}_{\bar{S}}\|_1\right\} > 0 \quad \text{(based on (A.13))}.
\end{aligned} \tag{A.19}
$$

Also, (A.13) implies that

$$
|\sigma_t(\mathbf{v})|\|\mathbf{x}\|_2^2 \geq s\|\mathbf{v}_S\|_1\|\mathbf{x}\|_2^2 \geq \|\mathbf{v}_S\|_1\|\mathbf{x}\|_1^2 \geq |\langle \mathbf{v}_S, \mathbf{x}\rangle|\|\mathbf{x}\|_1, \tag{A.20}
$$

thus leading to

$$\sigma_t(\mathbf{v})\|\mathbf{x}\|_2^2 - \langle \mathbf{v}_S, \mathbf{x}\rangle \|\mathbf{x}\|_1 \geq 0 \tag{A.21}$$

for $\sigma_t(\mathbf{v}) > 0$. As a result, we have $g'(t) \geq 0$ if $0 < t < t^*$. The function $g(t)$ is not differentiable at zero, but we can compute the subderivative as follows:

$$g'(0^+) = \lim_{t\to 0^+} \frac{g(t)-g(0)}{t-0} = \frac{2\|\mathbf{x}\|_1\left(\sigma_{+1}(\mathbf{v})\|\mathbf{x}\|_2^2 - \langle \mathbf{v}_S, \mathbf{x}\rangle \|\mathbf{x}\|_1\right)}{\|\mathbf{x}\|_2^4} \geq 0. \tag{A.22}$$

Similarly, we can get $g'(t) \leq 0$ if $-t^* < t < 0$ and $g'(0^-) \leq 0$. Therefore, for any $0 < |t| < t^*$, we have $g(0) \leq g(t)$, which implies that

$$\frac{\|\mathbf{x}+t\mathbf{v}\|_1}{\|\mathbf{x}+t\mathbf{v}\|_2} \geq \frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_2} \qquad \forall |t| < t^*. \tag{A.23}$$

Notice that $t^*$ does not depend on the choice of $\mathbf{v}$; therefore, the inequality is true for any $\mathbf{v}$ satisfying (A.2), which will imply the result. ∎

**Acknowledgments.** We would like to thank the editor and two reviewers for careful and thoughtful comments, which helped us greatly improve our paper. We also acknowledge the help of Dr. Min Tao from Nanjing University, who suggested the reference on the ADMM convergence.

REFERENCES

[1] A. S. Bandeira, E. Dobriban, D. G. Mixon, and W. F. Sawin, *Certifying the restricted isometry property is hard*, IEEE Trans. Inform. Theory, 59 (2013), pp. 3448–3450.
[2] A. Beck, *First-Order Methods in Optimization*, Vol. 25, SIAM, Philadelphia, 2017.
[3] J. Bolte, S. Sabach, and M. Teboulle, *Proximal alternating linearized minimization for nonconvex and nonsmooth problems*, Math. Program., 146 (2014), pp. 459–494.
[4] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Found. Trends Mach. Learn., 3 (2011), pp. 1–122.
[5] E. J. Candès and C. Fernandez-Granda, *Towards a mathematical theory of super-resolution*, Comm. Pure Appl. Math., 67 (2014), pp. 906–956.
[6] E. J. Candès, J. Romberg, and T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, Comm. Pure Appl. Math., 59 (2006), pp. 1207–1223.
[7] E. J. Candès and M. B. Wakin, *An introduction to compressive sampling*, IEEE Signal Process. Mag., 25 (2008), pp. 21–30.
[8] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vision, 40 (2011), pp. 120–145.
[9] R. Chartrand, *Exact reconstruction of sparse signals via nonconvex minimization*, IEEE Signal Process. Lett., 10 (2007), pp. 707–710.
[10] A. Cohen, W. Dahmen, and R. DeVore, *Compressed sensing and the best k-term approximation*, J. Amer. Math. Soc., 22 (2009), pp. 211–231.
[11] D. Donoho and M. Elad, *Optimally sparse representation in general (nonorthogonl) dictionaries via $l_1$ minimization*, Proc. Natl. Acad. Sci. USA, 100 (2003), pp. 2197–2202.
[12] D. L. Donoho and X. Huo, *Uncertainty principles and ideal atomic decomposition*, IEEE Trans. Inform. Theory, 47 (2001), pp. 2845–2862.
[13] E. Esser, Y. Lou, and J. Xin, *A method for finding structured sparse solutions to non-negative least squares problems with applications*, SIAM J. Imaging Sci., 6 (2013), pp. 2010–2046.
[14] A. Fannjiang and W. Liao, *Coherence pattern–guided compressive sensing with unresolved grids*, SIAM J. Imaging Sci., 5 (2012), pp. 179–202.
[15] T. Goldstein and S. Osher, *The split Bregman method for $L_1$-regularized problems*, SIAM J. Imaging Sci., 2 (2009), pp. 323–343.
[16] R. Gribonval and M. Nielsen, *Sparse representations in unions of bases*, IEEE Trans. Inform. Theory, 49 (2003), pp. 3320–3325.

[17] K. Guo, D. Han, and T.-T. Wu, *Convergence of alternating direction method for minimizing sum of two nonconvex functions with linear constraints*, Int. J. Comput. Math., 94 (2017), pp. 1653–1669.
[18] W. Guo and W. Yin, *Edge guided reconstruction for compressive imaging*, SIAM J. Sci. Imaging, 5 (2012), pp. 809–834.
[19] M. Hong, Z.-Q. Luo, and M. Razaviyayn, *Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems*, SIAM J. Optim., 26 (2016), pp. 337–364.
[20] P. O. Hoyer, *Non-negative sparse coding*, in Proceedings of the IEEE Workshop on Neural Networks for Signal Processing, Martigny, Switzerland, 2002, pp. 557–565.
[21] N. Hurley and S. Rickard, *Comparing measures of sparsity*, IEEE Trans. Inform. Theory, 55 (2009), pp. 4723–4741.
[22] D. Krishnan, T. Tay, and R. Fergus, *Blind deconvolution using a normalized sparsity measure*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, New York, 2011, pp. 233–240.
[23] M. J. Lai, Y. Xu, and W. Yin, *Improved iteratively reweighted least squares for unconstrained smoothed lq minimization*, SIAM J. Numer. Anal., 5 (2013), pp. 927–957.
[24] G. Li and T. K. Pong, *Global convergence of splitting methods for nonconvex composite optimization*, SIAM J. Optim., 25 (2015), pp. 2434–2460.
[25] Y. Lou, S. Osher, and J. Xin, *Computational aspects of $L_1$-$L_2$ minimization for compressive sensing*, in Modeling, Computation and Optimization in Information Systems Management Sciences, Vol. 359, Springer, Cham, Switzerland, 2015, pp. 169–180
[26] 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., 64 (2015), pp. 178–196.
[27] Y. Lou, T. Zeng, S. Osher, and J. Xin, *A weighted difference of anisotropic and isotropic total variation model for image processing*, SIAM J. Imaging Sci., 8 (2015), pp. 1798–1823.
[28] M. Lustig, D. L. Donoho, and J. M. Pauly, *Sparse MRI: The application of compressed sensing for rapid MR imaging*, Magnet. Reson. Med., 58 (2007), pp. 1182–1195.
[29] J. Lv and Y. Fan, *A unified approach to model selection and sparse recovery using regularized least squares*, Ann. Statist., 37 (2009), pp. 3498–3528.
[30] T. Ma, Y. Lou, and T. Huang, *Truncated $L_1$-$L_2$ models for sparse recovery and rank minimization*, SIAM J. Imaging Sci., 10 (2017), pp. 1346–1380.
[31] B. K. Natarajan, *Sparse approximate solutions to linear systems*, SIAM J. Comput., 24 (1995), pp. 227–234.
[32] Gurobi Optimization, *Gurobi Optimizer Reference Manual*, 2015, http://www.gurobi.com.
[33] J.-S. Pang and M. Tao, *Decomposition methods for computing directional stationary solutions of a class of nonsmooth nonconvex optimization problems*, SIAM J. Optim., 28 (2018), pp. 1640–1669.
[34] H. Raguet, J. Fadili, and G. Peyré, *A generalized forward-backward splitting*, SIAM J. Imaging Sci., 6 (2013), pp. 1199–1226.
[35] A. Repetti, M. Q. Pham, L. Duval, E. Chouzenouxe, and J.-C. Pesquet, *Euclid in a taxicab: Sparse blind deconvolution with smoothed $\ell_1/\ell_2$ regularization*, IEEE Signal Process. Lett., 22 (2015), pp. 539–543.
[36] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.
[37] X. Shen, W. Pan, and Y. Zhu, *Likelihood-based selection and sharp parameter estimation*, J. Amer. Statist. Assoc., 107 (2012), pp. 223–232.
[38] A. M. Tillmann and M. E. Pfetsch, *The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing*, IEEE Trans. Inform. Theory, 60 (2014), pp. 1248–1259.
[39] H. Tran and C. Webster, *Unified Sufficient Conditions for Uniform Recovery of Sparse Signals via Nonconvex Minimizations*, preprint, arXiv:1710.07348, 2017.
[40] F. Wang, W. Cao, and Z. Xu, *Convergence of multi-block Bregman ADMM for nonconvex composite problems*, Sci. China Info. Sci., 61 (2018), pp. 122101:1–12.
[41] F. Wang, Z. Xu, and H.-K. Xu, *Convergence of Bregman Alternating Direction Method with Multipliers for Nonconvex Composite Problems*, preprint, arXiv:1410.8625, 2014.
[42] Y. Wang, W. Yin, and J. Zeng, *Global convergence of ADMM in nonconvex nonsmooth optimization*, J. Sci. Comput., 78 (2019), pp. 29–63.
[43] 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., 23 (2012), pp. 1013–1027.
[44] 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., 14 (2014), pp. 87–109.

[45] P. Yin, Y. Lou, Q. He, and J. Xin, *Minimization of $\ell_{1-2}$ for compressed sensing*, SIAM J. Sci. Comput., 37 (2015), pp. A536–A563.
[46] S. Zhang and J. Xin, *Minimization of transformed $l_1$ penalty: Closed form representation and iterative thresholding algorithms*, Commun. Math. Sci., 15 (2017), pp. 511–537.
[47] 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. Ser. B, 169 (2018), pp. 307–336.
[48] T. Zhang, *Multi-stage convex relaxation for learning with sparse regularization*, in Advances in Neural Information Processing Systems, 2009, pp. 1929–1936.
[49] Y. Zhang, *Theory of compressive sensing via $L_1$-minimization: A non-RIP analysis and extensions*, J. Oper. Res. Soc. China, 1 (2013), pp. 79–105.