MINIMIZATION OF L1 OVER L2 FOR SPARSE SIGNAL RECOVERY WITH CONVERGENCE GUARANTEE*

MIN TAO†

Abstract. The ratio of the L_1 and L_2 norms, denoted by L_1/L_2, becomes attractive due to its scale-invariant property when approximating the L_0 norm to promote sparsity. In this paper, we incorporate the L_1/L_2 formalism into an unconstrained model in order to deal with both noiseless and noisy observations. To design an efficient algorithm, we derive an analytical solution for the proximal operator of the L_1/L_2 functional. Since the analytical solution depends on the sparsity of an unknown signal, we develop a bisection search method to find the desired sparsity and the corresponding solution to the proximal operator of L_1/L_2. With the new-developed solver of the proximal operator, we propose a specific variable-splitting scheme for the alternating direction method of multipliers (ADMM) so that we can establish its global convergence under mild assumptions and prove its linear convergence rate under suitable conditions. Experimentally, we conduct extensive numerical simulations to demonstrate the efficiency of the proposed approach over the state-of-the-art methods in sparse signal recovery with and without noise.

Key words. Proximal operator, sparsity, alternating direction method of multipliers, nonconvex optimization, global convergence, linear convergence

AMS subject classifications. 90C26, 90C90, 49N45

1. Introduction.

Compressive sensing (CS) is about finding a sparse signal from a set of linear measurements subject to some sort of noise. It has attracted tremendous interests and prompted developments in a variety of scientific disciplines, including mathematics [13], statistics [37], information science [4, 5], and geophysics [36]. The sparsity is measured by the L_0 norm, i.e., the number of nonzero elements of a vector. Mathematically, a fundamental problem in CS can be formulated as an unconstrained model,

$$\min_{\mathbf{x}\in\mathbb{R}^n} \left\{ \gamma\|\mathbf{x}\|_0 + \frac{1}{2}\|A\mathbf{x}-\mathbf{b}\|_2^2 \right\}, \tag{1.1}$$

where $A \in \mathbb{R}^{m\times n}$ $(m \ll n)$ is referred to as a sensing matrix, $\mathbf{b} = A\mathbf{x}^* + \boldsymbol{\epsilon}$ is an observation vector with an underlying signal $\mathbf{x}^*$ and an additive noise term $\boldsymbol{\epsilon}$, and $\gamma > 0$ is called a regularization parameter to balance these two terms. Unfortunately, the optimization (1.1) is NP-hard [32] to solve.

One popular approach in CS replaces the L_0 norm by the L_1 norm, leading to a so-called *basis pursuit* model [7]. The convex relaxation can be solved very efficiently by various optimization techniques [12, 14]. Theoretically, a sufficient condition to guarantee the exact recovery by the L_1 minimization is the restricted isometry property (RIP) [4], while a sufficient and necessary condition is given in terms of null space property (NSP) [8]. However, it is NP-hard to verify both RIP and NSP for a given matrix [2, 38]. A computable condition for L_1's exact recovery is based on *coherence*

*Submitted to the editors September 2, 2020.

Funding: Min Tao was partially supported by National Key Research and Development Program of China (2018AAA0101100), the Natural Science Foundation of China (No. 11971228) and the Jiangsu Provincial National Natural Science Foundation of China (No. BK20181257).

†Department of Mathematics, National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, 210093, China. (taom@nju.edu.cn)

[9] defined by

$$\mu(A) := \max_{i\neq j} \frac{|\langle \mathbf{a}_i, \mathbf{a}_j\rangle|}{\|\mathbf{a}_i\|_2\|\mathbf{a}_j\|_2}, \quad \text{with } A = [\mathbf{a}_1, \cdots, \mathbf{a}_n].$$

The L_1 minimization has been shown effective in sparse signal recovery when the coherence of the matrix A is small [5], while not working so well in the coherent regime. It was reported that L_1-L_2 [48, 27, 26] and transformed L_1 (TL$_1$) [49, 50] give superior performance over L_1, especially for coherent matrices. On the other hand, the resultant solution from the L_1 minimization often yields a biased estimation [31, 11]. To reduce the bias, there has been an increasing popularity to apply nonconvex terms as alternatives to L_1. For example, L_p for $p \in (0, 1)$ [42, 43] is a nonconvex regularizer with a representative case of $L_{1/2}$ [45, 46]. The optimization methods to minimize L_p include reweighted least squares [6, 22] and a half-thresholding algorithm for $L_{1/2}$ [45]. However, L_p only performs promisingly for incoherent matrices, and is even worse than L_1 in the coherent scenario.

This paper focuses on the ratio of L_1 and L_2 [10, 47, 34, 39], denoted by L_1/L_2, to approximate the L_0 norm, which is scale-invariant and parameter-free. L_1/L_2 was first proposed by Hoyer [16] in nonnegative matrix factorization, and later studied in [17] when discussing desirable properties of sparse promoting regularizations. In addition, the L_1/L_2 formulation has been successfully applied in blind deconvolution [20, 18] and limited-angle CT reconstruction [40]. Theoretically, Yin et al. [47] established the equivalence between L_1/L_2 and L_0 minimization when recovering nonnegative signals from linear constraints. Recently, Rahimi et al. [34] proved that a sparse solution is a local minimizer of the L_1/L_2 regularization subject to noiseless linear constraints under a strong NSP condition. The current optimization techniques for the L_1/L_2 minimization are limited to nonnegative constraints on the signal [10] as well as a noiseless setting [47, 34], all lacking convergence guarantees. One exception is a very recent work by Zeng et al. [51] that considered a series of constrained models involving the L_1/L_2 regularization with a linearly convergent algorithm based on a moving-ball approximation.

In order to deal with the noise in the data, we propose an unconstrained formulation by replacing the L_0 norm in (1.1) with the L_1/L_2 term, i.e.,

$$\min_{\mathbf{x}\in\mathcal{X}} F(\mathbf{x}) := \gamma\frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_2} + \frac{1}{2}\|A\mathbf{x}-\mathbf{b}\|_2^2. \tag{1.2}$$

By defining $\dfrac{\|\mathbf{0}\|_1}{\|\mathbf{0}\|_2} = 1$, the feasible set for (1.2) is $\mathcal{X} = \mathbb{R}^n$; and even so, $F(\mathbf{x})$ in (1.2) is not continuous at $\mathbf{0}$, but lower semi-continuous over $\mathcal{X}$. A constrained L_1/L_2 model was considered in [34] and solved via the alternating direction method of multipliers (ADMM). If we apply the same strategy as in [34, Algorithm 3.1] to minimize (1.2), we shall reformulate it as

$$\begin{aligned} \min\nolimits_{\mathbf{x},\mathbf{u},\mathbf{v}\in\mathcal{X}} \quad & \frac{\|\mathbf{u}\|_1}{\|\mathbf{v}\|_2} + \frac{1}{2}\|A\mathbf{x}-\mathbf{b}\|_2^2 \\ s.t. \quad & \mathbf{u}=\mathbf{x},\ \mathbf{v}=\mathbf{x}. \end{aligned} \tag{1.3}$$

As (1.3) violates standard convergence assumptions in nonconvex optimization for ADMM [23, 29, 15, 33, 41], the convergence of minimizing (1.3) via ADMM is hard to establish, which motivates us to design a different splitting scheme so that ADMM has guaranteed convergence.

In this paper, we derive a closed-form solution to the proximal operator of the L_1/L_2 functional, which provides a successful practice of extending the well-known Moreau's proximal theory [35] to a nonconvex setting. As the analytic solution of the proximal operator of L_1/L_2 depends on the true sparsity of the underlying signal, which is usually unknown, we develop a bisection search strategy to estimate the sparsity level. Equipped with the new-developed solver of the proximal operator, we propose a specific variable-splitting scheme to apply ADMM [23, 29, 33] for solving (1.2), which involves a subproblem that can be solved either by the derived proximal operator or by an inner-ADMM solver. We numerically demonstrate the efficiency of either approach for the subproblem. By verifying the Kurdyka-Łojasiewicz (KL) property of a merit function (defined later in (5.5)), we establish the global convergence of ADMM under a mild condition. We conduct extensive experiments on constructed examples, noisefree, and noisy cases to showcase the superior performance of the proposed approach over the state-of-the-art in sparse signal recovery, especially when dealing with noisy data. In summary, our contributions are threefold,

(1) We derive an analytic solution of the proximal operator of L_1/L_2 to enable an efficient algorithm for the L_1/L_2 minimization.
(2) We design a specific variable-splitting scheme of applying ADMM, and establish its global convergence together with a linear convergence rate under certain conditions.
(3) We compare several algorithms for minimizing the same L_1/L_2 model, shedding lights on computational efficiency for such a nonconvex problem.

The rest of the paper is organized as follows. We describe the notations and definitions of KL property/exponent in Section 2. The proximal operator for L_1/L_2 is derived in Section 3, including existence/uniqueness analysis and a practical solver. We elaborate on how to solve the unconstrained L_1/L_2 model (1.2) via ADMM in Section 4, where two ways to solve for an ADMM subproblem are discussed. The global convergence and convergence rate for the ADMM framework are analyzed in Section 5. Section 6 devotes to extensive experiments to showcase the superior performance of the proposed approach in sparse signal recovery with and without noise. Conclusions are given in Section 7.

2. Preliminary. We start with some notations and definitions that are used throughout the paper. We use a bold letter to denote a vector, e.g., $\mathbf{x} \in \mathbb{R}^n$, and x_i denotes the i-th entry of $\mathbf{x}$. We define $[n] := \{1, 2, \cdots, n\}$ as an index set up to n, and use the notation $\operatorname{supp}(\cdot)$ to present the support set of a vector. Give a vector $\mathbf{x} \in \mathbb{R}^n$ and an index set $\alpha \in [n]$, we use the notation of $\mathbf{x}_\alpha$ to denote a partial vector of $\mathbf{x}$ with its entries in the index set of α. Given a closed set $\mathcal{D} \in \mathbb{R}^n$, the notation $\iota_{\mathcal{D}}(\mathbf{x})$ represents the indicator function for the set $\mathcal{D}$, which takes the value 0 when $\mathbf{x} \in \mathcal{D}$ and $+\infty$ otherwise. Given an index set $\mathcal{I}$, we use $|\mathcal{I}|$ to denote the number of indices in this set. The distance from a point $\mathbf{x}$ to a closed set $\mathcal{D}$ is denoted by $\operatorname{dist}(\mathbf{x}, \mathcal{D})$, and the projection of a point $\mathbf{x}$ onto $\mathcal{D}$ is referred as $\operatorname{Proj}_{\mathcal{D}}(\mathbf{x})$. Given a matrix $A \in \mathbb{R}^{m\times n}$, we use $\mathcal{N}(A)$, $\mathcal{R}(A)$ to denote its null space, range space, respectively. We use I_n to denote the $n \times n$ identity matrix, and $\odot$ to present the componentwise product. An extended-real-valued function $f : \mathbb{R}^n \to (-\infty, +\infty]$ is said to be proper if its domain $\operatorname{dom} f := \{\mathbf{x} \mid f(\mathbf{x}) < \infty\}$ is nonempty. A proper function f is said to be closed if it is lower semi-continuous. For a proper closed function f and $\hat{\mathbf{x}} \in \operatorname{dom} f$, the regular subdifferential $\hat{\partial} f(\hat{\mathbf{x}})$ and the limiting subdifferential $\partial f(\mathbf{x})$ [35] are defined as

$$\hat{\partial} f(\hat{\mathbf{x}}) := \left\{ \mathbf{v} \Big| \lim_{\mathbf{x}\to\hat{\mathbf{x}}} \inf_{\mathbf{x}\neq\hat{\mathbf{x}}} \frac{f(\mathbf{x}) - f(\hat{\mathbf{x}}) - \langle \mathbf{v}, \mathbf{x} - \hat{\mathbf{x}} \rangle}{\|\mathbf{x} - \hat{\mathbf{x}}\|} \geq 0 \right\},$$

$$\partial f(\hat{\mathbf{x}}) := \left\{\mathbf{v} \middle| \exists\, \mathbf{x}^k \to \hat{\mathbf{x}},\ f(\mathbf{x}^k) \to f(\hat{\mathbf{x}}), \mathbf{v}^k \in \hat{\partial} f(\mathbf{x}^k) \text{ with } \mathbf{v}^k \to \mathbf{v}\right\},$$

respectively. As an abuse notation, the standard "sign" function can be applied to a vector as an elementwise calculation, i.e., $\text{sign}(\mathbf{x})$ is defined as a vector of the same length as $\mathbf{x}$ with its i-component equal to zero if $x_i = 0$; otherwise being the sign of each component of $\mathbf{x}$. We define the subgradient of the L_1 norm as

$$(\text{SGN}(\mathbf{x}))_i := \begin{cases} 1 & x_i > 0, \\ [-1, 1] & x_i = 0, \\ -1 & x_i < 0. \end{cases} \tag{2.1}$$

Note that $\text{sign}(\mathbf{x})$ and $\text{SGN}(\mathbf{x})$ are different in that if $x_i = 0$, we have $(\text{sign}(\mathbf{x}))_i = 0$ and $(\text{SGN}(\mathbf{x}))_i$ returns an arbitrary number between -1 and 1. We say $\bar{\mathbf{x}} \neq \mathbf{0}$ is a stationary point of (1.2), if it satisfies

$$\mathbf{0} \in \gamma \left(\frac{\text{SGN}(\bar{\mathbf{x}})}{\|\bar{\mathbf{x}}\|_2} - \frac{\|\bar{\mathbf{x}}\|_1}{\|\bar{\mathbf{x}}\|_2^3} \bar{\mathbf{x}} \right) + A^\top (A\bar{\mathbf{x}} - \mathbf{b}). \tag{2.2}$$

Next, we review on Kurdyka-Łojasiewicz (KL) property and KL exponent [3]. The KL property (also known as KL inequality) was first introduced by Łojasiewicz [25] for real analytic functions, and later extended to smooth functions [19]. Recently, Bolte et al. [3] extended the KL concepts to nonsmooth subanalytic functions. KL property is widely used in convergence analysis, while KL exponent plays a key role in analyzing convergence rate.

Definition 2.1. *(KL property & KL inequality) We say a proper closed function $h : \mathbb{R}^n \to (-\infty, +\infty]$ satisfies the KL property at a point $\hat{\mathbf{x}} \in \text{dom}\partial h$ if there exist a constant $\alpha \in (0, \infty]$, a neighborhood U of $\hat{\mathbf{x}}$, and a continuous concave function $\phi : [0, \nu) \to [0, \infty)$ with $\phi(0) = 0$ such that*

(i) ϕ is continuously differentiable on $(0, \nu)$ with $\phi' > 0$ on $(0, \nu)$;

(ii) for every $\mathbf{x} \in U$ with $h(\hat{\mathbf{x}}) < h(\mathbf{x}) < h(\hat{\mathbf{x}}) + \nu$, it holds that

$$\phi'(h(\mathbf{x}) - h(\hat{\mathbf{x}}))\text{dist}(\mathbf{0}, \partial h(\mathbf{x})) \geq 1.$$

We can further restrict the KL property on a closed set Ξ, if it holds for every $\mathbf{x} \in U \cap \Xi$ with $h(\hat{\mathbf{x}}) < h(\mathbf{x}) < h(\hat{\mathbf{x}}) + \nu$.

Definition 2.2. *(KL exponent) For a proper closed function h satisfying the KL property at $\hat{\mathbf{x}} \in \text{dom}\partial h$, if the corresponding function ϕ can be chosen as $\phi(s) = a_0 s^{1-\theta}$ for some $a_0 > 0$ and $\theta \in [0, 1)$, i.e., there exist $c,\ \varepsilon > 0$ and $\nu \in (0, \infty]$ so that*

$$\text{dist}(\mathbf{0}, \partial h(\mathbf{x})) \geq c(h(\mathbf{x}) - h(\hat{\mathbf{x}}))^\theta \tag{2.3}$$

whenever $\|\mathbf{x} - \hat{\mathbf{x}}\| \leq \varepsilon$ and $h(\hat{\mathbf{x}}) < h(\mathbf{x}) < h(\hat{\mathbf{x}}) + \nu$, then we say h has the KL property at $\hat{\mathbf{x}}$ with an exponent of θ. If h is a KL function and has the same exponent θ at any $\hat{\mathbf{x}} \in \text{dom}\partial h$, then we say that h is a KL function with an exponent of θ.

Similarly, we can define the KL exponent of h restricted on Ξ when (2.3) holds for "$\|\mathbf{x} - \hat{\mathbf{x}}\| \leq \varepsilon$, $\mathbf{x} \in \Xi$ and $h(\hat{\mathbf{x}}) < h(\mathbf{x}) < h(\hat{\mathbf{x}}) + \nu$" instead of "$\|\mathbf{x} - \hat{\mathbf{x}}\| \leq \varepsilon$ and $h(\hat{\mathbf{x}}) < h(\mathbf{x}) < h(\hat{\mathbf{x}}) + \nu$".

3. Proximal operator.

Define a proximal operator of L_1/L_2 with a parameter $\rho > 0$ as

$$\text{Prox}^{\rho}_{L_1/L_2}(\mathbf{q}) := \arg\min_{\mathbf{x}} \Big(\frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_2} + \frac{\rho}{2} \|\mathbf{x} - \mathbf{q}\|_2^2 \Big). \tag{3.1}$$

Denote the objective function in (3.1) by

$$H(\mathbf{x}) := \frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_2} + \frac{\rho}{2}\|\mathbf{x}-\mathbf{q}\|^2. \tag{3.2}$$

Obviously if $\mathbf{q} = \mathbf{0}$, the solution of (3.1) is $\mathbf{0}$. For the rest of the paper, we assume $\mathbf{q} \neq \mathbf{0}$. In what follows, we establish the existence of solutions of (3.1) and then derive a closed-form solution of $\mathrm{Prox}^{\rho}_{L_1/L_2}(\mathbf{q})$ in Section 3.1. In Section 3.2, we analyze solution's property of (3.1) when the sparsity is given and provide an explicit formula to characterize the values of $a := \|\bar{\mathbf{x}}\|_1$, $r := \|\bar{\mathbf{x}}\|_2$ when $t := \|\bar{\mathbf{x}}\|_0$ is given where $\bar{\mathbf{x}}$ denotes any optimal solution of (3.1). Since the closed-form solution involves nonlinear equations and an (unknown) sparsity level of the underlying signal, Section 3.3 details a practical way to find one of the solutions of $\mathrm{Prox}^{\rho}_{L_1/L_2}(\mathbf{q})$ with given $\mathbf{q} \in \mathbb{R}^n$ and $\rho > 0$.

3.1. Analytical formula. We present the existence of a solution to $\mathrm{Prox}^{\rho}_{L_1/L_2}(\mathbf{q})$ in Theorem 3.1. In Theorem 3.2, we characterize any global solution $\bar{\mathbf{x}}$ with a closed-form by splitting into two cases while the classification depends on the unknown values of $a := \|\bar{\mathbf{x}}\|_1$ and $r := \|\bar{\mathbf{x}}\|_2$. Furthermore, we provide a characterization for *one of* the global solutions with a closed-form in Theorem 3.3 by dividing into two cases only relying on the proximal parameter ρ.

THEOREM 3.1. *Given $\mathbf{q}$ $(\neq \mathbf{0}) \in \mathbb{R}^n$ and $\rho > 0$, the solution set of (3.1) is nonempty.*

Proof. It follows from [35, Definition 1.23] and [35, Theorem 1.25] directly. ∎

THEOREM 3.2. *Given $\mathbf{q} \neq \mathbf{0} \in \mathbb{R}^n$ and $\rho > 0$. We can sort $\mathbf{q}$ in a descending order in a way of $|q_{\sigma(1)}| \geq |q_{\sigma(2)}| \geq \cdots \geq |q_{\sigma(n)}| \geq 0$, where $\{\sigma(i)\}_{i=1}^n$ is a proper permutation of $[n]$. Then the following assertions hold:*

(i) For any optimal solution $\bar{\mathbf{x}}$ of (3.1), it has the same descending order and sign pattern as $\mathbf{q}$, i.e.,

$$|\bar{x}_{\sigma(1)}| \geq |\bar{x}_{\sigma(2)}| \geq \cdots \geq |\bar{x}_{\sigma(n)}| \ \text{ and } \ \mathrm{sign}(\bar{x}_{\sigma(i)}) = \mathrm{sign}(q_{\sigma(i)}), \ \ \forall i \in [n].$$

(ii) Furthermore, we denote an integer s as the number of elements in $\mathbf{q}$ with the same largest magnitude, i.e., $|q_{\sigma(1)}| = \cdots = |q_{\sigma(s)}| > |q_{\sigma(s+1)}|$. For each optimal solution $\bar{\mathbf{x}}$ ($t = \|\bar{\mathbf{x}}\|_0$, $a = \|\bar{\mathbf{x}}\|_1$, $r = \|\bar{\mathbf{x}}\|_2$), one of the following assertions is satisfied with:

(a) If $0 < \rho \leq a/r^3$, then there exists one index $j \in [s]$ such that

$$\bar{x}_{\sigma(i)} = \begin{cases} \mathrm{sign}(q_{\sigma(i)}) \odot q_{\sigma(i)} & i = j; \\ 0 & \textit{otherwise}. \end{cases} \tag{3.3}$$

(b) If $\rho > a/r^3$, then there exists a pair of (a, r) such that ($Q^t = \sum_{i=1}^t |q_{\sigma(i)}|$)

$$\begin{cases} \dfrac{a^2}{r^3} - \rho a + \rho Q^t - \dfrac{t}{r} = 0, & \text{(3.4a)} \\ r^3 - \left(\displaystyle\sum_{i=1}^t q^2_{\sigma(i)}\right) r + \dfrac{Q^t - a}{\rho} = 0, & \text{(3.4b)} \end{cases}$$

and the r is also satisfied with

$$|q_{\sigma(t)}| > \frac{1}{\rho r} \ \ \textit{and} \ \ |q_{\sigma(t+1)}| \leq \frac{1}{\rho r}, \tag{3.5}$$

the solution $\bar{\mathbf{x}}$ is characterized by

$$\bar{x}_{\sigma(i)} = \begin{cases} \operatorname{sign}(q_{\sigma(i)}) \odot \dfrac{\rho|q_{\sigma(i)}| - \frac{1}{r}}{\rho - \frac{a}{r^3}} & 1 \le i \le t; \\ 0 & \textit{otherwise}. \end{cases} \tag{3.6}$$

Proof. (i) The assertion holds obviously. Therefore, we can assume without loss of generality that $q_1 \ge q_2 \ge \cdots \ge q_n \ge 0$, then $\bar{\mathbf{x}}$ has the same descending order as $\mathbf{q}$ and is all nonnegative, i.e.,

$$\bar{x}_1 \ge \cdots \ge \cdots \ge \bar{x}_n = 0.$$

Furthermore, we have the following assertion:

$$\text{Suppose } \bar{x}_i,\ \bar{x}_j > 0 \text{ and } q_i > q_j \ \Rightarrow\ \bar{x}_i > \bar{x}_j. \tag{3.7}$$

The proof is elementary and thus omitted here.
(ii) Since $\mathbf{q} \neq \mathbf{0}$, thus $\bar{\mathbf{x}} \neq \mathbf{0}$. By invoking [30, Theorem 3.2.2 and Theorem 3.2.12], the KKT conditions of (3.1) lead to

$$\frac{\boldsymbol{\zeta}}{\|\bar{\mathbf{x}}\|_2} - \frac{\|\bar{\mathbf{x}}\|_1}{\|\bar{\mathbf{x}}\|_2^3}\bar{\mathbf{x}} + \rho(\bar{\mathbf{x}} - \mathbf{q}) = \mathbf{0}, \tag{3.8}$$

where $\boldsymbol{\zeta} \in \mathrm{SGN}(\bar{\mathbf{x}})$. By the definitions of a, r, we express (3.8) for every component i,

$$\frac{1}{r}\zeta_i - \frac{a}{r^3}\bar{x}_i + \rho(\bar{x}_i - q_i) = 0. \tag{3.9}$$

Since $\bar{x}_i > 0$ for $1 \le i \le t$, we have $\zeta_i = 1$, thus leading to

$$\left(\rho - \frac{a}{r^3}\right)\bar{x}_i = \rho q_i - \frac{1}{r}, \quad 1 \le i \le t. \tag{3.10}$$

For the assertion (a), we split into two cases to verify: (I) $\rho - \frac{a}{r^3} < 0$; (II) $\rho - \frac{a}{r^3} = 0$.

(I) If $\rho - \frac{a}{r^3} < 0$, then we can verify that the nonzero entries of $\bar{\mathbf{x}}$ can take only one nonzero value. Suppose not. It means that there are two different nonzero entires in $\bar{\mathbf{x}}$, i.e., $\bar{x}_i,\ \bar{x}_j > 0$ and without loss of generality $q_i > q_j$. Then, it leads to $\bar{x}_i < \bar{x}_j$ due to $\rho - \frac{a}{r^3} < 0$ which contradicts to (3.7). Thus, $\bar{\mathbf{x}}$ can take only one nonzero value for its entries. Before verifying the assertion (a), we first show two facts:

Fact 1: If $\rho - a/r^3 < 0$, then $\rho < 1/q_1^2$.

Fact 2: Suppose $\rho \in (0, 1/q_1^2)$. Then, there exists a maximum integer $\nu_{\max} \in [s]$ such that $0 < \rho < \frac{1}{\sqrt{\nu_{\max}} q_1^2}$. For each $\nu \in [\nu_{\max}]$, the corresponding vector $\bar{\mathbf{x}}^{(\nu)}$ defined by

$$\bar{x}_i^{(\nu)} = \begin{cases} q_1 & 1 \le i \le \nu; \\ 0 & \text{otherwise}, \end{cases} \tag{3.11}$$

is satisfied with the first-order optimality condition (3.8) and $\rho < a/r^3$. If $\nu_{\max} > 1$, one has $H(\bar{\mathbf{x}}^{(1)}) < H(\bar{\mathbf{x}}^{(\nu)})$ for any ν $(1 < \nu \le \nu_{\max})$ where $H(\mathbf{x})$ is defined in (3.2).

First, we prove **Fact 1**. We have verified that $\bar{\mathbf{x}}$ has only one nonzero value in its entries since $\rho < a/r^3$. Thus, we assume $x_i = \kappa$ for $1 \le i \le \nu \le s$ and other entries are zero. It implies $a = \nu\kappa$ and $r = \sqrt{\nu}\kappa$. It follows from (3.10) with $i = 1$ that

$$(\rho - \frac{\nu\kappa}{\nu\sqrt{\nu}\kappa^3})\kappa = \rho q_1 - \frac{1}{\sqrt{\nu}\kappa}, \tag{3.12}$$

which leads to $\kappa = q_1$. Thus, $\rho < \frac{1}{\sqrt{\nu}q_1^2}$ due to $\rho < a/r^3$. Furthermore, we have $0 < \rho < 1/q_1^2$ due to $\nu \ge 1$.

Next, we prove **Fact 2**. For the given $\rho \in (0, \frac{1}{q_1^2})$, there exists a maximum integer $\nu_{\max} \in [s]$ such that $0 < \rho < \frac{1}{\sqrt{\nu_{\max}}q_1^2}$. Then, each $\nu \in [\nu_{\max}]$ is also satisfied with $0 < \rho < \frac{1}{\sqrt{\nu}q_1^2}$. Next, we show that

$$\bar{\mathbf{x}}^{(1)},\ \cdots,\ \bar{\mathbf{x}}^{(\nu_{\max}-1)},\ \bar{\mathbf{x}}^{(\nu_{\max})} \text{ are all satisfying (3.8) and } \rho < a/r^3,$$

where $\bar{\mathbf{x}}^{(\nu)}$ is defined in (3.11). In order to show that (3.8) holds for each $\bar{\mathbf{x}}^{(\nu)}$, we only need to verify (3.9) holds for $\bar{x}_i^{(\nu)}$ with $i = 1, \dots, n$. We divide the index set $[n]$ into three subsets: (1) $[\nu]$; (2) $\{\nu+1, \dots, s\}$; (3) $\{s+1, \dots, n\}$. First, (3.9) holds with $i = 1, \dots, \nu$ due to (3.12). Then, since there exist a scalar $\xi \in (0, 1]$ such that $\rho q_1 = \frac{\xi}{\sqrt{\nu}q_1}$ due to $\rho < \frac{1}{\sqrt{\nu}q_1^2}$, it leads to (3.9) holds with $i = \nu+1, \dots, s$. Similarly, we can show that (3.9) holds with $i = s+1, \dots, n$. Obviously, all these $\bar{\mathbf{x}}^{(\nu)}$ ($\nu \in [\nu_{\max}]$) are all satisfying $\rho < a/r^3$. Next, for any ν ($1 < \nu \le \nu_{\max}$), we have

$$\begin{aligned} & H(\bar{\mathbf{x}}^{(1)}) - H(\bar{\mathbf{x}}^{(\nu)}) \\ & = [1 + \frac{\rho}{2}\sum_{i=1}^{n}(\bar{\mathbf{x}}_i^{(1)} - q_i)^2] - [\sqrt{\nu} + \frac{\rho}{2}\sum_{i=1}^{n}(\bar{x}_i^{(\nu)} - q_i)^2] \\ & = \frac{\rho}{2}(\nu - 1)q_1^2 + 1 - \sqrt{\nu} \\ & < \frac{\nu - 1}{2\sqrt{\nu}} + 1 - \sqrt{\nu} < 0, \end{aligned}$$

where the last but one is due to $0 < \rho < \frac{1}{\sqrt{\nu}q_1^2}$. Thus,

$$H(\bar{\mathbf{x}}^{(1)}) < H(\bar{\mathbf{x}}^{(\nu)}),\ \ \forall\, 1 < \nu \le \nu_{\max}.$$

By combining Facts 1-2, we see that $\bar{\mathbf{x}}^{(1)}$ is a global solution of (3.1) when $0 < \rho < a/r^3$ [1].

(II) Suppose $\rho - \frac{a}{r^3} = 0$. Any $\hat{\mathbf{x}}$ satisfying (3.8) and $\rho = a/r^3$ where $a = \|\hat{\mathbf{x}}\|_1$ and $r = \|\hat{\mathbf{x}}\|_2$, it holds that

$$a = 1/(\rho^2 q_1^3) \ \text{ and } \ r = 1/(\rho q_1), \tag{3.13}$$

due to $\rho q_1 - \frac{1}{r} = 0$ and $\rho = \frac{a}{r^3}$. Next, we show that $\mathrm{supp}(\hat{\mathbf{x}}) \subseteq [s]$. In order to do so, we only need to show that $\hat{x}_{s+1} = 0$. Suppose not. Then,

[1] Indeed, any one-sparse solution with the support set on $\{j\}$ where $j \in [s]$ and nonzero entry q_1 is a global solution of (3.1) when $0 < \rho < a/r^3$. If $s \ge 2$, (3.1) has multiple one-sparse solutions. All of them are with the same objective function value H. For conciseness, we only consider one of them, i.e., $\bar{\mathbf{x}}^{(1)}$.

$\hat{x}_{s+1} > 0$. Furthermore, one has $\hat{x}_1 > 0$. Since $q_1 > q_{s+1}$, thus it yields that $\hat{x}_1 > \hat{x}_{s+1}$ due to (3.7). On the other hand, since (3.8) holds for $i = 1, s+1$, one has $\rho q_1 - \frac{1}{r} = \rho q_{s+1} - \frac{1}{r}$ due to $\rho - a/r^3 = 0$. It leads to $q_1 = q_{s+1}$ which contradicts to $q_1 > q_{s+1}$. Thus, $\mathrm{supp}(\hat{\mathbf{x}}) \subseteq [s]$. Since $a \geq r$, it leads to $\rho \leq 1/q_1^2$. If $\rho < 1/q_1^2$, as indicated in the proof to **Fact 2**, there exists a maximum integer $\nu_{\max} \in [s]$ such that $\rho < \frac{1}{\sqrt{\nu_{\max}} q_1^2}$, then

$$\bar{\mathbf{x}}^{(1)}, \ldots, \bar{\mathbf{x}}^{(\nu_{\max})} \text{ are satisfied with (3.8) and } \rho < a/r^3.$$

Also, $H(\bar{\mathbf{x}}^{(1)}) < H(\bar{\mathbf{x}}^{(\nu)})$, $\forall\, 1 < \nu \leq \nu_{\max}$ Therefore, we only need to compare the objective function values of H at $\bar{\mathbf{x}}^{(1)}$ and $\hat{\mathbf{x}}$. Since $\langle \hat{\mathbf{x}}, \mathbf{q} \rangle = aq_1$ due to $\mathrm{supp}(\hat{\mathbf{x}}) \subseteq [s]$, it leads to

$$\begin{aligned}
H(\hat{\mathbf{x}}) &= a/r + \frac{\rho}{2}\|\hat{\mathbf{x}} - \mathbf{q}\|^2 \\
&= 1/(\rho q_1^2) + \frac{\rho}{2}(r^2 - 2\langle \hat{\mathbf{x}}, \mathbf{q} \rangle + \|\mathbf{q}\|^2) \\
&= 1/(\rho q_1^2) + \frac{\rho}{2}(r^2 - 2aq_1 + \|\mathbf{q}\|^2) \\
&= 1/(\rho q_1^2) + \frac{\rho}{2}(-1/(\rho^2 q_1^2) + \|\mathbf{q}\|^2) \\
&= \frac{1}{2(\rho q_1^2)} + \frac{\rho}{2}\|\mathbf{q}\|^2.
\end{aligned}$$

Similarly, $H(\bar{\mathbf{x}}^{(1)}) = 1 + \frac{\rho}{2}(-q_1^2 + \|\mathbf{q}\|^2)$. Hence, one has $H(\bar{\mathbf{x}}^{(1)}) < H(\hat{\mathbf{x}})$ by using $\frac{1}{2(\rho q_1^2)} + \frac{\rho q_1^2}{2} > 1$. Thus, if $\rho < 1/q_1^2$, (3.1) does not have a global solution satisfying $\rho = a/r^3$. If $\rho = 1/q_1^2$, $\hat{\mathbf{x}}$ is a global solution of (3.1) since $H(\hat{\mathbf{x}}) = H(\bar{\mathbf{x}}^{(1)})$. For this case, $\hat{\mathbf{x}}$ is just one-sparse solution $\bar{\mathbf{x}}^{(1)}$.

Combining (I) and (II), the assertion (a) of (ii) holds.

For the assertion (b) of (ii), suppose $\rho > a/r^3$. By summing (3.10) from $i = 1, \ldots, t$ and using $\sum_{i=1}^{t} \bar{x}_i = a$, we obtain (3.4a). Together with $\|\bar{\mathbf{x}}\|_2 = r$, we further get

$$\begin{aligned}
\sum_{i=1}^{t}\left(\rho q_i - \frac{1}{r}\right)^2 &= r^2\rho^2\left((1 - \frac{a}{\rho r^3}) - \frac{a}{\rho r^3}(1 - \frac{a}{\rho r^3})\right) \\
&= r^2\rho^2\left((1 - \frac{a}{\rho r^3}) - \frac{1}{\rho r^3}(Q^t - \frac{t}{\rho r})\right),
\end{aligned}$$

which can be simplified as (3.4b). Using (3.10), we obtain an analytical solution of $\bar{\mathbf{x}}$ expressed in (3.6) as a general form (with sorting and sign). Note that the conditions in (3.5) is necessary for guaranteeing $\bar{x}_t > 0$ and $\bar{x}_{t+1} = 0$ from (3.9) and (3.10).

Therefore, the assertion (b) of (ii) follows directly. ∎

From the above results, it seems to be not practical to use Theorem 3.2 to characterize the optimal solutions of (3.1) since a and r are unknown. However, since we only need to find one of the optimal solutions of (3.1) instead of all of them, it turns to be possible to characterize one of them with a closed form in terms of the value of ρ. We present this result in the following theorem.

THEOREM 3.3. *Given $\mathbf{q} \neq \mathbf{0} \in \mathbb{R}^n$ and $\rho > 0$. We can sort $\mathbf{q}$ in a a descending order in a way of $|q_{\sigma(1)}| = \cdots = |q_{\sigma(s)}| > |q_{\sigma(s+1)}| \geq \cdots \geq |q_{\sigma(n)}| \geq 0$, where $\{\sigma(i)\}_{i=1}^n$ is a proper permutation of $[n]$ and s is an integer and denote the number of elements in $\mathbf{q}$ with the same largest magnitude. Then, one of the following assertions holds:*

(i) If $0 < \rho \le \frac{1}{q^2_{\sigma(1)}}$, (3.1) has a solution given by (3.3);

(ii) If $\rho > \frac{1}{q^2_{\sigma(1)}}$, (3.1) has a solution given by (3.6) such that (a, r) are satisfied with (3.4) and (3.5) simultaneously.

Proof. (a) If $0 < \rho \le \frac{1}{q^2_{\sigma(1)}}$, the assertion follows from the proof to Item (ii) of Theorem 3.2. (b) Suppose $\rho > \frac{1}{q^2_{\sigma(1)}}$. It implies that $\rho > a/r^3$. Suppose not. If $\rho \le a/r^3$, then $\rho \le 1/q^2_{\sigma(1)}$ (see the cases (I) and (II) in the proof to Theorem 3.2). Thus, $\rho > a/r^3$. Then, the conclusion follows immediately by invoking Item (ii) of Theorem 3.2. ∎

3.2. Solution's property when sparsity is given. As indicated by the assertion (ii) in Theorem 3.3, the solution of (3.1) may not be unique when the sparsity is unknown. We show in Example 3.4 that there are three solutions of (3.1) with two different sparsity levels.

EXAMPLE 3.4. *Let $n = 2$ and $q = \sqrt{2(\sqrt{2}-1)}$. Consider an objective function*

$$\frac{|x_1| + |x_2|}{\sqrt{x_1^2 + x_2^2}} + \frac{1}{2}(x_1 - q)^2 + \frac{1}{2}(x_2 - q)^2 .$$

It has three global optimal solutions $\mathbf{x}_1 = (q, 0)^\top$, $\mathbf{x}_2 = (0, q)^\top$ and $\mathbf{x}_3 = (q, q)^\top$, $\mathbf{x}_1$, $\mathbf{x}_2$ are satisfying the assertion (a) of (ii) in Theorem 3.2 and $\mathbf{x}_3$ is satisfying the assertion (b) of (ii) in Theorem 3.2. Clearly, they have different sparsity levels.

When the sparsity t is given, we can prove the uniqueness of (a, r), leading to the uniqueness of the solution to (3.1) in the sense of not considering inner rotation-invariant property for these entries of $\bar{\mathbf{x}}$ corresponding to these q_i in the same magnitude. For this purpose, we introduce Lemma 3.5 to characterize a cubic function for having a positive real root and Lemma 3.6 to guarantee there exists a unique pair of (a, r) satisfying the nonlinear system (3.4).

LEMMA 3.5. *Consider the following cubic polynomial for a single variable x,*

$$G(x) := x^3 - \mu x + \nu,$$

with $\mu > 0$ and $\nu \ge 0$. If $G(x) = 0$ admits a positive real solution, then its discriminant is nonpositive, i.e.,

$$\Delta^{(3)} := \frac{\nu^2}{4} - \frac{\mu^3}{27} \le 0, \tag{3.14}$$

where we use $\Delta^{(3)}$ to denote the discriminant for a 3rd-order polynomial. Then Cardano's Formula [44] shows that $G(x) = 0$ has three distinct real roots given by

$$x_1 = 2\varrho\cos(\pi/3 - \phi/3), \quad x_2 = -2\varrho\cos(\phi/3), \quad x_3 = 2\varrho\cos(\pi/3 + \phi/3), \tag{3.15}$$

where $\varrho := \sqrt{\mu/3}$ and $\phi := \arccos\frac{\nu}{2\varrho^3}$.

Proof. We prove (3.14) by contradiction. Suppose $\Delta^{(3)} > 0$, then we have

$$G(-\sqrt{\mu/3}) \cdot G(\sqrt{\mu/3}) = 4\Big((\nu/2)^2 - (\mu/3)^3\Big) > 0,$$

which implies that both $G(-\sqrt{\mu/3})$ and $G(\sqrt{\mu/3})$ have the same sign. If both are positive, then local maximum and minimum values of $G(x)$ are positive, as $\pm\sqrt{\mu/3}$ are roots of $G'(x) = 0$. Since $G(x)$ is a cubic polynomial and $\lim_{x\to\infty} G(x) = \infty$, $G(x)$ can only have one negative real root, which contradicts with the existence of a positive real solution. If both $G(-\sqrt{\mu/3})$ and $G(\sqrt{\mu/3})$ are negative, we get $G(-\sqrt{\mu/3}) + G(\sqrt{\mu/3}) = 2\nu < 0$, which contradicts with $\nu \geq 0$. Therefore, the assertion (3.14) holds, and the root formula (3.15) is elementary. ∎

LEMMA 3.6. *Under the same setting as in Theorem 3.3, there exists a unique pair of* (a, r) *satisfying* (3.4) *when the sparsity is given.*

Proof. We only need to consider the case of $\rho > \dfrac{1}{q^2_{\sigma(1)}}$. For (3.4b), we have $\sum_{i=1}^t q^2_{\sigma(i)} > 0$ as $\mathbf{q} \neq \mathbf{0}$. We can rewrite (3.4a) as

$$\rho(Q^t - a) = \frac{tr^2 - a^2}{r^3} \geq 0,$$

where the inequality is due to $tr^2 \geq a^2$. It further follows from Theorem 3.1 that there exists a solution to (3.1), which guarantees that (3.4b) has a positive solution $r > 0$. Then by Lemma 3.5, (3.4b) has three real roots, i.e.,

$$r_1 = 2\varrho\cos(\pi/3 - \phi/3),\ r_2 = -2\varrho\cos(\phi/3),\ r_3 = 2\varrho\cos(\pi/3 + \phi/3),$$

where $\varrho = \sqrt{\sum_{i=1}^t q^2_{\sigma(i)}/3}$ and $\phi = \arccos\dfrac{Q^t - a}{2\rho\varrho^3}$.

On the other hand, (3.4a) is a quadratic polynomial of variable a with two roots

$$a_\pm = \frac{r^3(\rho \pm \sqrt{\Delta^{(2)}})}{2},$$

where $\Delta^{(2)} := \rho^2 - \frac{4}{r^3}(\rho Q^t - \frac{t}{r})$ denotes its discriminant. Due to the existence of the solution, we have $\Delta^{(2)} \geq 0$, and we can further prove that $a_+ \geq a_- \geq 0$.

When minimizing the ratio of the L_1 and L_2 norms, i.e., $\dfrac{a}{r}$, we want to choose the largest value for r and the smallest value for a. By simple calculations, the largest positive root is r_1. We then choose a smaller value between a_+ and a_-, which should also be larger than r_1. Therefore, there is only one choice of (a, r) satisfying (3.4). ∎

3.3. Practical solver. Theorem 3.3 presents a closed-form solution of $\bar{\mathbf{x}} \in \mathrm{Prox}^{\rho}_{L_1/L_2}(\mathbf{q})$ in (3.1) that can be computed from (3.6), but it depends on the true sparsity t $(:= \|\bar{\mathbf{x}}\|_0)$. If t is unknown, we consider a bisection search to find its value. Suppose we have a lower bound and an upper bound, denoted by $t_1 < t_2$, respectively. We examine the middle point $t = \frac{1}{2}(t_1 + t_2)$, and adopt an alternating updated scheme to solve the nonlinear system (3.4) for (a, r) for the current sparsity t. Specifically starting from an initial guess of r_0, we define Q^t and $\varrho = \sqrt{\sum_{i=1}^t q^2_{\sigma(i)}/3}$ according to

Theorem 3.2, then the pair (a, r) can be found iteratively,

$$\begin{cases} \Delta_{\ell+1}^{(2)} = \rho^2 - \frac{4}{r_\ell^3}\big(\rho Q^t - \frac{t}{r_\ell}\big), & (3.16a)\\ a_{\ell+1} = \frac{r_\ell^3}{2}\big(\rho - \sqrt{\Delta_{\ell+1}^{(2)}}\big), & (3.16b)\\ \phi_{\ell+1} = \arccos\big(\frac{Q^t - a_{\ell+1}}{2\rho\varrho^3}\big), & (3.16c)\\ r_{\ell+1} = 2\varrho\cos(\pi/3 - \phi_{\ell+1}/3), & (3.16d)\end{cases}$$

where ℓ represents the iteration number. If $\Delta_{\ell+1}^{(2)} < 0$ at a certain iteration ℓ, we use a rough estimation of $\hat{r} = r_\ell$ to update t by examining whether $|q_{\sigma(t+1)}| > 1/(\rho\hat{r})$ or not. If $\Delta_{\ell+1}^{(2)}$ remains nonnegative, we iterate (3.16) until $\ell > \ell_{\max}$ or the relative error of r_ℓ and $r_{\ell+1}$ is less than $\varepsilon = 1e^{-3}$, leading to a pair of approximated solutions denoted by $(\hat{a}, \hat{r}) := (a_{\ell+1}, r_{\ell+1})$. If the conditions in (3.5) hold, then we can obtain a closed-form solution of $\bar{\mathbf{x}}$ via (3.6) with $(\hat{a}, \hat{r})$; otherwise, we shall update t. In summary, an overall algorithm for finding $\bar{\mathbf{x}} \in \mathrm{Prox}_{L_1/L_2}^{\rho}(\mathbf{q})$ with given $\mathbf{q} \in \mathbb{R}^n$ and $\rho > 0$ is presented in Algorithm 3.1. Heuristic though, Algorithm 3.1 works well in practice, which motivates us to seek for theoretical rationales in the future.

4. Computational approach.

We elaborate on how to minimize the unconstrained L_1/L_2 problem (1.2) via ADMM. In particular, we introduce an auxiliary variable $\mathbf{y}$ such that (1.2) is equivalent to the following constrained formulation,

$$\begin{array}{ll} \min_{\mathbf{x},\mathbf{y}\in\mathcal{X}} & \gamma\frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_2} + \Phi(\mathbf{y}) \\ s.t. & \mathbf{x} = \mathbf{y}, \end{array} \tag{4.1}$$

where $\Phi(\mathbf{x}) := \frac{1}{2}\|A\mathbf{x} - \mathbf{b}\|_2^2$. We define the augmented Lagrangian of (4.1) as

$$\mathcal{L}_{\mathcal{A}}(\mathbf{x}, \mathbf{y}, \mathbf{z}) = \gamma\frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_2} + \Phi(\mathbf{y}) + \langle \mathbf{z}, \mathbf{x} - \mathbf{y}\rangle + \frac{\beta}{2}\|\mathbf{x} - \mathbf{y}\|_2^2, \tag{4.2}$$

where $\mathbf{z}$ is the Lagrangian multiplier (or called dual variable) and $\beta > 0$ is a parameter. The ADMM scheme iterates as follows,

$$\begin{cases} \mathbf{x}^{k+1} \in \arg\min_{\mathbf{x}\in\mathcal{X}}\big(\gamma\frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_2} + \frac{\beta}{2}\left\|\mathbf{x} - \mathbf{y}^k + \frac{\mathbf{z}^k}{\beta}\right\|_2^2\big), & (4.3a)\\ \mathbf{y}^{k+1} = \arg\min_{\mathbf{y}\in\mathcal{X}}\big(\Phi(\mathbf{y}) + \frac{\beta}{2}\|\mathbf{y} - \mathbf{x}^{k+1} - \frac{\mathbf{z}^k}{\beta}\|_2^2\big), & (4.3b)\\ \mathbf{z}^{k+1} = \mathbf{z}^k + \beta(\mathbf{x}^{k+1} - \mathbf{y}^{k+1}). & (4.3c)\end{cases}$$

The $\mathbf{y}$-subproblem (4.3b) has a closed-form solution given by

$$\mathbf{y}^{k+1} = \left(I_n + \frac{1}{\beta}A^\top A\right)^{-1}\left(\frac{A^\top\mathbf{b}}{\beta} + \frac{\mathbf{z}^k}{\beta} + \mathbf{x}^{k+1}\right).$$

Since we often encounter an under-determined system in CS, we apply the Sherman-Morrison-Woodburg Theorem, leading to a more efficient $\mathbf{y}$-update

$$\mathbf{y}^{k+1} = \left(I_n - \frac{1}{\beta}A^\top(I_m + \frac{1}{\beta}AA^\top)^{-1}A\right)\left(\frac{A^\top b}{\beta} + \frac{\mathbf{z}^k}{\beta} + \mathbf{x}^{k+1}\right),$$

Algorithm 3.1 Finding the solution to $\mathrm{Prox}^{\rho}_{L_1/L_2}(\mathbf{q})$ in (3.1) via a bisection search

```
Require: ρ > 0 and q ∈ ℝ^n with |q_σ(1)| = ··· = |q_σ(s)| > |q_σ(s+1)| ≥ ··· ≥ |q_σ(n)|
  Parameters: ℓ_max (maximum iteration) and ε (error tolerance)
  if ρ ≤ 1/q_1^2 then
      x̄_σ(i) is computed by (3.3).
  else
      Set t_1 = 1, t_2 = n.
      while t_2 − t_1 > 1 do
          Set t = ⌊(t_1 + t_2)/2⌋, r_0 > 0 and Flag1 = 0.
          for ℓ = 0 : ℓ_max do
              Compute Δ^(2)_{ℓ+1} by (3.16a).
              if Δ^(2)_{ℓ+1} < 0 then
                  Set r̂ = r_ℓ; Flag1 = 1; break.
              end if
              Update (a_{ℓ+1}, r_{ℓ+1}) via (3.16b)-(3.16d)
              if |r_{ℓ+1} − r_ℓ|/|r_ℓ| < ε then
                  Set r̂ = r_{ℓ+1}; break;
              end if
          end for
          if Flag1 = 1 then
              if |q_σ(t+1)| > 1/(ρr̂), then
                          t_1 = t,
              else
                          t_2 = t,
              end if
          else
              Compute â via (3.16a)-(3.16b) with r_ℓ = r̂.
              Flag2 = (q_σ(t) − 1/(ρr̂))(1 − â/(ρr̂^3)) > 0,
              if Flag2, then
                  if (q_σ(t+1) − 1/(ρr̂))(1 − â/(ρr̂^3)) ≤ 0 then,
                                  compute x̄ via (3.6).
                  else
                                  t_1 = t.
                  end if
              else
                  t_2 = t
              end if
          end if
      end while
  end if
  Output x̄.
```

as $AA^\top$ has a smaller dimension than $A^\top A$.

As for the $\mathbf{x}$-subproblem (4.3a), we adopt two ways to solve it. One is based on the proximal operator (3.1), referred to as ADMM_p, while the other uses an iterative scheme, referred to as ADMM_i. Specifically by letting $\mathbf{q} = \mathbf{y}^k - \mathbf{z}^k/\beta$ and $\rho = \beta/\gamma$, we observe that the $\mathbf{x}$-subproblem (4.3a) coincides with $\mathrm{Prox}^{\rho}_{L_1/L_2}(\mathbf{q})$, which has a closed-form solution (Theorem 3.3) and a practical solver (Algorithm 3.1).

Algorithm 4.1 ADMM

Require: $A \in \mathbb{R}^{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$, $\gamma, \beta, \varepsilon > 0$, and $k_{\max}$.
  Initialize: $\mathbf{y}^0 = \mathbf{z}^0$.
  while $k < k_{\max}$ or $\|\mathbf{x}^{k-1} - \mathbf{x}^k\|/\|\mathbf{x}^k\| > \varepsilon$ **do**
    Solving the $\mathbf{x}$-subproblem (4.3a) via Algorithm 3.1 or (4.5).
    Computing $\mathbf{y}^{k+1} = (I_n - \frac{1}{\beta}A^\top(I_m + \frac{1}{\beta}AA^\top)^{-1}A)(\frac{1}{\beta}(A^\top\mathbf{b}) + \frac{1}{\beta}\mathbf{z}^k + \mathbf{x}^{k+1})$.
    Updating $\mathbf{z}^{k+1} = \mathbf{z}^k + \beta(\mathbf{x}^{k+1} - \mathbf{y}^{k+1})$.
  end while

For ADMM$_i$, we introduce a variable $\mathbf{u}$ and rewrite (4.3a) as

$$\begin{array}{lll} (4.4) & \min_{\mathbf{x},\mathbf{u}\in\mathcal{X}} & \gamma\frac{\|\mathbf{x}\|_1}{\|\mathbf{u}\|_2} + \frac{\beta}{2}\|\mathbf{x}-\mathbf{q}\|_2^2 \\ & s.t. & \mathbf{x} = \mathbf{u}. \end{array}$$

The ADMM scheme to minimize (4.4) reads as

$$\begin{cases} \mathbf{x}_{j+1} = \arg\min_{\mathbf{x}\in\mathcal{X}} \left(\gamma\frac{\|\mathbf{x}\|_1}{\|\mathbf{u}_j\|_2} + \frac{\beta}{2}\|\mathbf{x}-\mathbf{q}\|_2^2 + \frac{\delta}{2}\left\|\mathbf{x} - \mathbf{u}_j + \mathbf{v}_j\right\|_2^2\right), & (4.5a) \\ \mathbf{u}_{j+1} = \arg\min_{\mathbf{u}\in\mathcal{X}} \left(\gamma\frac{\|\mathbf{x}_{j+1}\|_1}{\|\mathbf{u}\|_2} + \frac{\delta}{2}\|\mathbf{u} - \mathbf{x}_{j+1} - \mathbf{v}_j\|_2^2\right), & (4.5b) \\ \mathbf{v}_{j+1} = \mathbf{v}_j + (\mathbf{x}_{j+1} - \mathbf{u}_{j+1}), & (4.5c) \end{cases}$$

where $\mathbf{v}$ is the dual variable, δ is a positive parameter, and we use the subscript j to index the inner iterations, as opposed to superscript k for the outer ones. The $\mathbf{x}$-subproblem (4.5a) can be solved by *soft shrinkage*, i.e.,

$$\mathbf{x}_{j+1} = \mathbf{shrink}\left(\frac{\beta\mathbf{q} + \delta(\mathbf{u}_j - \mathbf{v}_j)}{\beta+\delta},\ \frac{\gamma}{(\delta+\beta)\|\mathbf{u}_j\|_2}\right),$$

where $\mathbf{shrink}(\mathbf{x}, \mu) = \mathrm{sign}(\mathbf{x}) \odot \max(|\mathbf{x}| - \mu, 0)$. There is a closed-form solution for the $\mathbf{u}$-subproblem given by the same root formula of cubic polynomials as in [34]. We incorporate a warm-start strategy to accelerate its performance. Warm-start means that we use $\mathbf{x}^k$ as the initial point of $\mathbf{x}_0$ in (4.5) in order to obtain $\mathbf{x}^{k+1}$. The convergence of (4.5) to a stationary point of (4.3a) can be established, which is omitted as it is similar to the convergence analysis in Section 5. In practice, we use a stopping condition of a maximum iteration number, denoted by $j_{\max}$.

We summarize the overall ADMM framework (4.3) for minimizing the unconstrained L_1/L_2 model (1.2) in Algorithm 4.1. In particular, we can use either Algorithm 3.1 or (4.5) for solving the $\mathbf{x}$-subproblem (4.3a). However, both ADMM$_p$ and ADMM$_i$ may go to stationary points that are not global solution, as illustrated in Example 4.1.

Example 4.1. *Let $m = 2$ and $n = 3$, and the objective function be*

$$\frac{|x_1| + |x_2| + |x_3|}{\sqrt{x_1^2 + x_2^2 + x_3^2}} + \frac{1}{2}(x_1 + x_2 - 1)^2 + \frac{1}{2}(x_2 + x_3 - 1)^2. \quad (4.6)$$

We can verify that $\mathbf{x}^* = (0, 1, 0)^\top$ is a global minimizer to this toy problem (4.6). We run ADMM$_p$ and ADMM$_i$ ($j_{\max} = 10$) with five different initial points of (x_1^0, x_2^0, x_3^0)

TABLE 1
Comparison between $ADMM_p$ and $ADMM_i$ under different initial points.

Initial points	Solution		AbErr	
	ADMM_p	ADMM_i	ADMM_p	ADMM_i
$(-1,1,-1)^\top$	$(0,1.00,0)^\top$	$(0,1.00,0)^\top$	8.35e-07	4.60e-04
$(0,1,-1)^\top$	$(0,1.00,0)^\top$	$(0,1.00,0)^\top$	4.33e-07	8.62e-06
$(-0.2,0,0.2)^\top$	$(0,1.00,0)^\top$	$(1.00,0,1.00)^\top$	2.10e-07	1.73e+00
$(10,-10,0)^\top$	$(6.75,-5.74,6.75)^\top$	$(6.99,-5.98,6.99)^\top$	1.17e+01	1.21e+01
$\mathtt{rand}(3,1)$	$(0,1.00,0)^\top$	$(0,1.00,0)^\top$	1.95e-08	3.70e-05

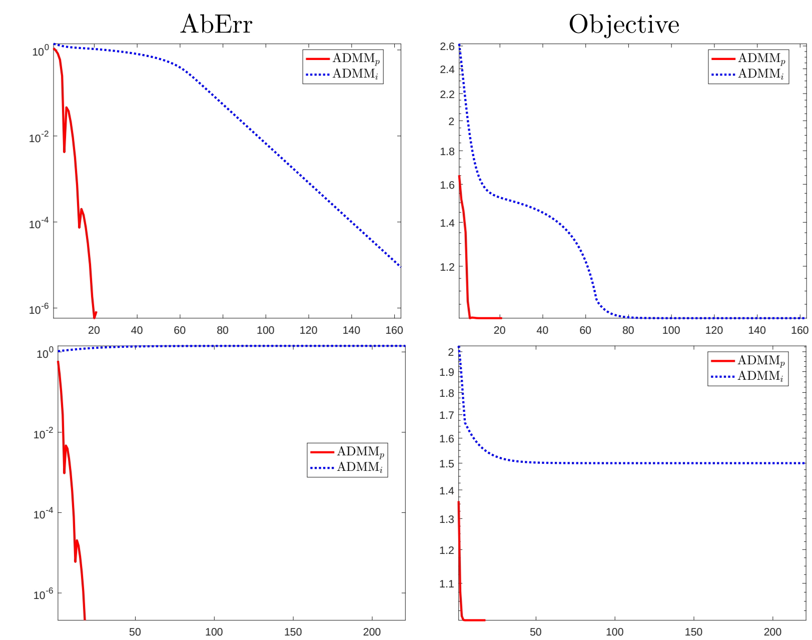

FIG. 1. *Comparison between $ADMM_p$ and $ADMM_i$ for minimizing (4.6) with two initial points: $(-1,1,-1)^\top$ and $(-0.2,0,0.2)^\top$, as shown on the top and bottom rows, respectively.*

listed in Table 1 and report the results in terms of the final solution ("Solution") and the absolute error between $\mathbf{x}^k$ and $\mathbf{x}^*$ (ground-truth) ("AbErr"), showing that both of ADMM_p and ADMM_i reach stationary points, and even converge to the global solution for some initial points. We also plot the absolute error between $\mathbf{x}^k$ and $\mathbf{x}^*$ and the objective function of (4.6) evaluated at $\mathbf{x}^k$ with respect to iteration count k in Figure 1, which illustrates that ADMM_p converges much faster than ADMM_i. Please refer to Section 6 for more comparison between these two methods.

5. Theoretical analysis. This section focuses on the analysis of the proposed model (1.2) and the ADMM scheme (4.3). In particular, we establish the existence of solutions to (1.2) in Section 5.1. We prove the subsequential convergence of the ADM-M scheme (4.3) when minimizing the unconstrained L_1/L_2 model (4.1) in Section 5.2. Finally, the global convergence and convergence rate are analyzed in Section 5.3.

5.1. Solution's existence. Since $1 \le \frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_2} \le \sqrt{n}$, we have

$$F^* := \inf_{\mathbf{x}} \left(\gamma \frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_2} + \frac{1}{2}\|A\mathbf{x} - \mathbf{b}\|_2^2 \right) < \infty.$$

We define $\{\mathbf{x}^k\}$ as a minimizing sequence of (1.2) if $\lim_{k\to\infty} F(\mathbf{x}^k) = F^*$. Inspired by [51], we only need to show that any minimizing sequence of (1.2) has an accumulation point, which implies that the solution set of our model (1.2) is nonempty.

THEOREM 5.1. *If $\mathbf{b} \notin \mathcal{N}(A^\top)$, the optimal solution of (1.2) can not be $\mathbf{0}$.*

Proof. Suppose that $A = [\mathbf{a}_1, \cdots, \mathbf{a}_m]$ with $\mathbf{a}_i \in \mathbb{R}^n$. By the assumption $\mathbf{b} \notin \mathcal{N}(A^\top)$, there exists at least one index j such that $\mathbf{a}_j^\top \mathbf{b} \neq 0$ and a scalar $\alpha \neq 0$ such that $\|\alpha \mathbf{a}_j - \mathbf{b}\|_2 < \|\mathbf{b}\|_2$. With such α, we define a vector $\hat{\mathbf{x}} \in \mathbb{R}^n$ as

$$\hat{x}_i = \begin{cases} \alpha & i = j; \\ 0 & \text{otherwise.} \end{cases}$$

As $F(\hat{\mathbf{x}}) < F(\mathbf{0})$, it implies that $\mathbf{0}$ can not be the global solution of (1.2). ∎

THEOREM 5.2. *If $A \in \mathbb{R}^{m\times n}$ is full row rank, $\mathbf{b} \neq \mathbf{0}$, and $0 < \gamma < \frac{\|\mathbf{b}\|_2^2}{2(\sqrt{n}-1)}$, the optimal value of (1.2) is nonempty.*

Proof. Thanks to Theorem 5.1, there exists a minimizing sequence of (1.2), denoted by $\{\mathbf{x}^k\}$, with $\mathbf{x}^k \neq \mathbf{0}$. Since A is full row rank, we define $\tilde{\mathbf{x}} := A^\top (AA^\top)^{-1}\mathbf{b}$. As $A\tilde{\mathbf{x}} = \mathbf{b}$, then we have $F(\tilde{\mathbf{x}}) \le \gamma\sqrt{n}$. In the following, we split into two cases to verify:

(I) Suppose that there are a finite number of k such that $\mathbf{x}^k \notin \mathcal{N}(A)$. We remove these terms from the sequence and still denote the remaining sequence by $\{\mathbf{x}^k\}$ without loss of generality. We see that

$$\gamma\sqrt{n} \ge F(\tilde{\mathbf{x}}) \ge F^* = \lim_{k\to\infty} F(\mathbf{x}^k) \ge \gamma + \frac{1}{2}\|\mathbf{b}\|_2^2,$$

which contradicts to $0 < \gamma < \frac{\|\mathbf{b}\|_2^2}{2(\sqrt{n}-1)}$.

(II) Otherwise, there are infinite number of k such that $\mathbf{x}^k \notin \mathcal{N}(A)$. It implies there exists a subsequence $\{\mathbf{x}^{k_j}\} \subseteq \{\mathbf{x}^k\}$ such that $\mathbf{x}^{k_j} \notin \mathcal{N}(A)$. Using the fact of $A\tilde{\mathbf{x}} = \mathbf{b}$, it leads to

$$\sqrt{\sigma_{\min}(A^\top A)}\|\mathbf{x}^{k_j}\|_2 - \|\mathbf{b}\|_2 \le \|A\mathbf{x}^{k_j}\|_2 - \|\mathbf{b}\|_2 \le \|A\mathbf{x}^{k_j} - \mathbf{b}\|_2 \le 2\sqrt{F^*} \tag{5.1}$$

where $\sigma_{\min}(A^\top A)$ denotes the smallest nonzero eigenvalue of the matrix $A^\top A$. The first and the last inequalities in (5.1) are due to $\mathbf{x}^{k_j} \notin \mathcal{N}(A)$ and the sequence $\{\mathbf{x}^{k_j}\}$ such that $\|A\mathbf{x}^{k_j} - \mathbf{b}\|_2^2 \le 2F(\mathbf{x}^{k_j}) \le 4F^*$ when j is sufficient large. Then it follows from (5.1) that $\{\mathbf{x}^{k_j}\}$ is bounded, and hence there exists a subsequence of $\{\mathbf{x}^{k_{j_t}}\}$ convergent to a vector $\hat{\mathbf{x}}$, i.e., $\lim_{t\to\infty} \mathbf{x}^{k_{j_t}} = \hat{\mathbf{x}}$ and

$$\lim_{t\to\infty} F(\mathbf{x}^{t_{j_t}}) = F^*,$$

which implies that $\hat{\mathbf{x}} \in \arg\min_{\mathbf{x}} F(\mathbf{x})$. Therefore, $\hat{\mathbf{x}}$ is a global solution to (1.2). ∎

From Theorem 5.1, we see that $\mathbf{0}$ can not be a global solution of (1.2) under the assumptions of Theorem 5.2 as $A \in \mathbb{R}^{m\times n}$ being full row rank and $\mathbf{b} \neq \mathbf{0}$ imply that $\mathbf{b} \notin \mathcal{N}(A^\top)$.

5.2. Subsequential convergence. Following the convergence analysis [41, 33, 28], we prove in Lemma 5.3 that augmented Lagrangian function (4.2) is sufficiently decreasing. For succinctness, we denote $\mathcal{L}_{\mathcal{A}}^k := \mathcal{L}_{\mathcal{A}}(\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k)$ and $\mathbf{w}^k = (\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k)$. We then prove in Theorem 5.4 that the sequence generated by (4.3) converges to a stationary point of (4.1) when the sequence $\{\mathbf{x}^k\}$ is bounded.

LEMMA 5.3. *Let $\{\mathbf{w}^k\}$ be the sequence generated by (4.3), then we have*

$$\mathcal{L}_{\mathcal{A}}^{k+1} \leq \mathcal{L}_{\mathcal{A}}^k - \left(\frac{\beta}{2} - \frac{L^2}{\beta}\right) \|\mathbf{y}^k - \mathbf{y}^{k+1}\|_2^2, \tag{5.2}$$

where $L := \sigma_{\max}(A^\top A)$ is the largest eigenvalue of $A^\top A$.

Proof. It follows from the optimality condition of (4.3b) that $\nabla\Phi(\mathbf{y}^{k+1}) = \mathbf{z}^{k+1}$, and similarly $\nabla\Phi(\mathbf{y}^k) = \mathbf{z}^k$. It is straightforward that $\nabla\Phi(\cdot)$ is Lipschitz continuous with the constant L, which implies that

$$\|\mathbf{z}^k - \mathbf{z}^{k+1}\|_2 = \|\nabla\Phi(\mathbf{y}^k) - \nabla\Phi(\mathbf{y}^{k+1})\|_2 \leq L\|\mathbf{y}^k - \mathbf{y}^{k+1}\|_2. \tag{5.3}$$

The convexity of Φ also yields that

$$\Phi(\mathbf{y}^k) \geq \Phi(\mathbf{y}^{k+1}) - \langle \mathbf{y}^{k+1} - \mathbf{y}^k, \nabla\Phi(\mathbf{y}^{k+1})\rangle.$$

Simple calculations lead to

$$\begin{aligned}
&\mathcal{L}_{\mathcal{A}}(\mathbf{x}^{k+1}, \mathbf{y}^k, \mathbf{z}^k) - \mathcal{L}_{\mathcal{A}}(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}^{k+1}) \\
&= \Phi(\mathbf{y}^k) - \Phi(\mathbf{y}^{k+1}) + \langle \mathbf{z}^k, \mathbf{x}^{k+1} - \mathbf{y}^k\rangle + \frac{\beta}{2}\|\mathbf{x}^{k+1} - \mathbf{y}^k\|_2^2 \\
&\quad - \left(\langle \mathbf{z}^{k+1}, \mathbf{x}^{k+1} - \mathbf{y}^{k+1}\rangle + \frac{\beta}{2}\|\mathbf{x}^{k+1} - \mathbf{y}^{k+1}\|_2^2\right) \\
&= \Phi(\mathbf{y}^k) - \Phi(\mathbf{y}^{k+1}) - \langle \mathbf{z}^{k+1}, \mathbf{y}^k - \mathbf{y}^{k+1}\rangle + \frac{\beta}{2}\|\mathbf{y}^{k+1} - \mathbf{y}^k\|_2^2 - \frac{1}{\beta}\|\mathbf{z}^k - \mathbf{z}^{k+1}\|_2^2 \\
&\geq \frac{\beta}{2}\|\mathbf{y}^{k+1} - \mathbf{y}^k\|_2^2 - \frac{1}{\beta}\|\mathbf{z}^k - \mathbf{z}^{k+1}\|_2^2 \geq (\frac{\beta}{2} - \frac{L^2}{\beta})\|\mathbf{y}^{k+1} - \mathbf{y}^k\|_2^2.
\end{aligned}$$

Since $\mathcal{L}_{\mathcal{A}}(\mathbf{x}^{k+1}, \mathbf{y}^k, \mathbf{z}^k) \leq \mathcal{L}_{\mathcal{A}}(\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k)$ by (4.3a), the inequality (5.2) holds. ∎

THEOREM 5.4. *Let $\{\mathbf{w}^k\}$ be the sequence generated by (4.3). If $\{\mathbf{x}^k\}$ is bounded and $\beta > \sqrt{2}L$, we have the following statements:*
- *(i) The sequence $\{\mathbf{w}^k\}$ has at least one accumulation point;*
- *(ii) $\lim_{k\to\infty} \|\mathbf{x}^k - \mathbf{x}^{k+1}\| = 0$, $\lim_{k\to\infty} \|\mathbf{y}^k - \mathbf{y}^{k+1}\| = 0$, and $\lim_{k\to\infty} \|\mathbf{z}^k - \mathbf{z}^{k+1}\| = 0$;*
- *(iii) Any accumulation point of $\{\mathbf{x}^k, \mathbf{y}^k\}$ is a stationary point of (4.1).*

Proof. (i) It follows from Lemma 5.3 that

$$\mathcal{L}_{\mathcal{A}}^{k+1} \leq \mathcal{L}_{\mathcal{A}}^k - \tilde{c}\|\mathbf{y}^k - \mathbf{y}^{k+1}\|_2^2, \tag{5.4}$$

with $\tilde{c} := \frac{\beta}{2} - \frac{L^2}{\beta}$. By the assumption of $\beta > \sqrt{2}L$, we have $\tilde{c} > 0$. Consequently, we get $\mathcal{L}_{\mathcal{A}}^k \leq \mathcal{L}_{\mathcal{A}}^0$, and hence $\mathcal{L}_{\mathcal{A}}^k$ is upper bounded. On the other hand, we have

$$\begin{aligned}
\mathcal{L}_{\mathcal{A}}^k &= \gamma\frac{\|\mathbf{x}^k\|_1}{\|\mathbf{x}^k\|_2} + \Phi(\mathbf{y}^k) + \frac{\beta}{2}\|\mathbf{x}^k - \mathbf{y}^k\|_2^2 + \langle \nabla\Phi(\mathbf{y}^k), \mathbf{x}^k - \mathbf{y}^k\rangle \\
&\geq \gamma\frac{\|\mathbf{x}^k\|_1}{\|\mathbf{x}^k\|_2} + \Phi(\mathbf{x}^k) + \frac{\beta - L}{2}\|\mathbf{x}^k - \mathbf{y}^k\|_2^2 \geq 0.
\end{aligned}$$

If $\{\mathbf{x}^k\}$ is bounded and $\beta > \sqrt{2}L > L$, then $\|\mathbf{x}^k - \mathbf{y}^k\|_2$ is bounded, leading to the boundedness of $\{\mathbf{y}^k\}$. Note that

$$\|\mathbf{z}^k\|_2^2 = \|\nabla\Phi(\mathbf{y}^k) - \nabla\Phi(\mathbf{0}) + \Phi(\mathbf{0})\|_2^2 \le 2\left(L^2\|\mathbf{y}^k\|_2^2 + \|\Phi(\mathbf{0})\|_2^2\right),$$

which implies that $\{\mathbf{z}^k\}$ is bounded. Therefore, the sequence of $\{\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k\}$ is bounded and hence it has at least one accumulation point.

(ii) By summing (5.4) from $k = 0$ to ∞ and using $\mathcal{L}_{\mathcal{A}}^k \ge 0$, we get that $\|\mathbf{y}^k - \mathbf{y}^{k+1}\|_2 \to 0$ as $k \to \infty$. Then it follows from (5.3) that $\|\mathbf{z}^k - \mathbf{z}^{k+1}\|_2 \to 0$ as $k \to \infty$. By the $\mathbf{z}$-update (4.3c), we have $\mathbf{x}^k - \mathbf{x}^{k+1} = \frac{1}{\beta}(\mathbf{z}^{k+1} - \mathbf{z}^k) - \frac{1}{\beta}(\mathbf{z}^k - \mathbf{z}^{k-1}) + (\mathbf{y}^{k+1} - \mathbf{y}^k)$, leading to $\|\mathbf{x}^{k+1} - \mathbf{x}^k\|_2 \to 0$ as well.

(iii) The proof is similar to [28, Theorem 2], thus we omit here. ∎

REMARK 5.5. *In standard convergence analysis, one often encounters coercive functionals such as L_p and $L_1 - L_2$ to guarantee that the solution is bounded. Unfortunately, L_1/L_2 is not coercive, so we have to assume the sequence is bounded in order to prove for convergence.*

5.3. Global convergence and convergence rate. Theorem 5.4 establishes the subsequential convergence of ADMM. To ease the global convergence and convergence rate analysis, we introduce an auxiliary function, often referred to as a merit function,

$$M(\mathbf{x}, \mathbf{y}) := \gamma\frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_2} + \Phi(\mathbf{x}) + \frac{\beta}{2}\|\mathbf{x} - \mathbf{y}\|_2^2. \tag{5.5}$$

Using Lemma 5.3, we show the sufficient decay of the function M; see Lemma 5.6. Then we provide an upper bound for the distance between the subgradient of M and the zero vector in terms of residual between two adjacent iterates in Lemma 5.7.

LEMMA 5.6. *Let $\{\mathbf{w}^k\}$ be the sequence generated by (4.3) and $M(\mathbf{x}, \mathbf{y})$ defined in (5.5). If $\beta > 2L$, then there exists a constant $c_1 > 0$ such that*

$$M(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}) \le M(\mathbf{x}^k, \mathbf{y}^k) - c_1\|\mathbf{y}^k - \mathbf{y}^{k+1}\|_2^2. \tag{5.6}$$

Proof. By using Lipschitz continuity of $\nabla\Phi$ and $\mathbf{z}^{k+1} = \nabla\Phi(\mathbf{y}^{k+1})$, we have

$$\begin{aligned}
\mathcal{L}_{\mathcal{A}}^{k+1} &= \gamma\frac{\|\mathbf{x}^{k+1}\|_1}{\|\mathbf{x}^{k+1}\|_2} + \Phi(\mathbf{y}^{k+1}) + \langle\nabla\Phi(\mathbf{y}^{k+1}), \mathbf{x}^{k+1} - \mathbf{y}^{k+1}\rangle + \frac{\beta}{2}\|\mathbf{x}^{k+1} - \mathbf{y}^{k+1}\|_2^2 \\
&\ge F(\mathbf{x}^{k+1}) + \frac{\beta - L}{2}\|\mathbf{x}^{k+1} - \mathbf{y}^{k+1}\|_2^2.
\end{aligned}$$

Additionally, due to $\mathbf{z}^k = \nabla\Phi(\mathbf{y}^k)$, we get

$$\begin{aligned}
\mathcal{L}_{\mathcal{A}}^{k} &= \gamma\frac{\|\mathbf{x}^{k}\|_1}{\|\mathbf{x}^{k}\|_2} + \Phi(\mathbf{y}^{k}) + \langle\nabla\Phi(\mathbf{y}^{k}), \mathbf{x}^{k} - \mathbf{y}^{k}\rangle + \frac{\beta}{2}\|\mathbf{x}^{k} - \mathbf{y}^{k}\|_2^2 \\
&\le \gamma\frac{\|\mathbf{x}^{k}\|_1}{\|\mathbf{x}^{k}\|_2} + \Phi(\mathbf{x}^{k}) + \frac{\beta}{2}\|\mathbf{x}^{k} - \mathbf{y}^{k}\|_2^2.
\end{aligned}$$

Together with (5.2), the definition of M and (4.3c), the above two inequalities lead to

$$M(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}) \le M(\mathbf{x}^k, \mathbf{y}^k) - (\frac{\beta}{2} - \frac{L^2}{2})\|\mathbf{y}^k - \mathbf{y}^{k+1}\|_2^2 + \frac{L}{2\beta^2}\|\mathbf{z}^k - \mathbf{z}^{k+1}\|_2^2.$$

It further follows from (5.3) that the inequality (5.6) holds with $c_1 = \frac{\beta}{2} - \frac{L^2}{\beta} - \frac{L^3}{2\beta^2}$, which is positive due to $\beta > 2L$. ∎

LEMMA 5.7. *Let $\{\mathbf{w}^k\}$ be the sequence generated by (4.3). Then there exists a constant $c_2 > 0$ such that*

$$\mathrm{dist}(\mathbf{0}, \partial M(\mathbf{x}^{k+1}, \mathbf{y}^{k+1})) \leq c_2 \|\mathbf{y}^{k+1} - \mathbf{y}^k\|_2.$$

Proof. The optimality condition of $\mathbf{x}$-subproblem (4.3a) leads to

$$\mathbf{0} \in \gamma\partial\left(\frac{\|\mathbf{x}^{k+1}\|_1}{\|\mathbf{x}^{k+1}\|_2}\right) + \mathbf{z}^{k+1} + \beta(\mathbf{y}^{k+1} - \mathbf{y}^k).$$

Denote $\boldsymbol{\xi}_1^{k+1} := \mathbf{z}^{k+1} - \mathbf{z}^k + \beta(\mathbf{y}^k - \mathbf{y}^{k+1}) + \nabla\Phi(\mathbf{x}^{k+1}) - \nabla\Phi(\mathbf{y}^{k+1})$, then we obtain

$$\boldsymbol{\xi}_1^{k+1} \in \gamma\partial\left(\frac{\|\mathbf{x}^{k+1}\|_1}{\|\mathbf{x}^{k+1}\|_2}\right) + \nabla\Phi(\mathbf{x}^{k+1}) + \beta(\mathbf{x}^{k+1} - \mathbf{y}^{k+1}),$$

which means that $\boldsymbol{\xi}_1^{k+1} \in \partial_{\mathbf{x}} M(\mathbf{x}^{k+1}, \mathbf{y}^{k+1})$. In addition, we define $\boldsymbol{\xi}_2^{k+1} := \beta(\mathbf{y}^{k+1} - \mathbf{x}^{k+1}) \in \partial_{\mathbf{y}} M(\mathbf{x}^{k+1}, \mathbf{y}^{k+1})$, and get $\boldsymbol{\xi}_2^{k+1} = \mathbf{z}^k - \mathbf{z}^{k+1}$ based on the $\mathbf{z}$-update of (4.3c). Denoting $\boldsymbol{\xi}^{k+1} = ((\boldsymbol{\xi}_1^{k+1})^\top, (\boldsymbol{\xi}_2^{k+1})^\top)^\top$, there exists a positive scalar c_2 such that

$$\|\boldsymbol{\xi}^{k+1}\|_2^2 = \|\boldsymbol{\xi}_1^{k+1}\|_2^2 + \|\boldsymbol{\xi}_2^{k+1}\|_2^2 \leq c_2^2 \|\mathbf{y}^{k+1} - \mathbf{y}^k\|_2^2.$$ ∎

We are now ready to establish the global convergence and the linear convergence rate of (4.3) in Theorems 5.8 and 5.9, respectively. The proofs of these two theorems are given in the appendix.

THEOREM 5.8. *Let $\{\mathbf{w}^k\}$ be the sequence generated by (4.3). If $\mathbf{b} \notin \mathcal{N}(A^\top)$, $\beta > 2L$, and $\{\mathbf{x}^k\}$ is bounded, then $\{\mathbf{w}^k\}$ has finite length, i.e., $\sum_{k=1}^\infty \|\mathbf{w}^{k+1} - \mathbf{w}^k\| < \infty$, and hence $\{\mathbf{w}^k\}$ converges to a stationary point of (4.1).*

THEOREM 5.9. *Let $\{\mathbf{w}^k\}$ be the sequence generated by (4.3). If $\mathbf{b} \notin \mathcal{N}(A^\top)$, $\beta > 2L$, and $\{\mathbf{x}^k\}$ is bounded, then $\{\mathbf{w}^k\}$ converges to a stationary point $\mathbf{w}^\infty = (\mathbf{x}^\infty, \mathbf{y}^\infty, \mathbf{z}^\infty)$. Let $\bar\Lambda := \mathrm{supp}(\mathbf{x}^\infty)$. Define*

$$\Xi := \{\mathbf{x} \in \mathbb{R}^n \mid \mathrm{supp}(\mathbf{x}) = \bar\Lambda,\ \mathrm{sign}(\mathbf{x}) = \mathrm{sign}(\mathbf{x}^\infty)\}.$$

We have the following statements:

(i) There exists an index K such that $\mathbf{x}^k \in \Xi$ and $\mathbf{y}^k \in \Xi$ when $k \geq K$.

(ii) Suppose that $A_{\bar\Lambda}^\top A_{\bar\Lambda} \succ 0$. Then, there exists a scalar $\bar\gamma > 0$ such that the sequence of $\{\mathbf{w}^k\}$ converges linearly when $0 < \gamma < \bar\gamma$, i.e., there exist $c_1 > 0$ and $\nu \in [0, 1)$ such that

$$\|\mathbf{w}^k - \mathbf{w}^\infty\| \leq c_1 \nu^k. \tag{5.7}$$

6. Experiments. We conduct extensive experiments to demonstrate the performance of the proposed L_1/L_2 algorithms. Specifically in Section 6.1, we construct a stationary point of the model (1.2) to show the efficiency of the proposed model (1.2), as the ground-truth signal may not be the optimal solution to (1.2). We then compare the L_1/L_2 model (1.2) with the state-of-the-art models in sparse recovery, including L_1, L_p [6] ($L_{1/2}$ [45]), and $L_1 - L_2$ [28], for noiseless and noisy cases in Sections 6.2

and 6.3, respectively. The initial value for all the nonconvex models is chosen to be the L_1 solution, i.e.,

$$\mathbf{x}^0 = \arg\min_{\mathbf{x}} \left(\gamma \|\mathbf{x}\|_1 + \frac{1}{2}\|A\mathbf{x} - \mathbf{b}\|_2^2 \right). \tag{6.1}$$

The stopping criterion is set to meet one of the conditions:

$$\texttt{RelChg} := \frac{\|\mathbf{x}^k - \mathbf{x}^{k-1}\|_2}{\max\{\|\mathbf{x}^k\|_2, \|\mathbf{x}^{k-1}\|_2, \varepsilon\}} < \texttt{Tol} \quad \text{or} \quad k_{\max} > 5n, \tag{6.2}$$

for $\varepsilon = 10^{-16}$. We set $\texttt{Tol}$ as

$$\texttt{Tol} = \begin{cases} 10^{-5} & \text{if } \sigma = 0, \\ 0.01 * \sigma & \text{if } \sigma > 0, \end{cases}$$

where σ is the variance of the noise ($\sigma = 0$ means the noiseless case).

We generate an s-sparse ground truth signal $\mathbf{x}^* \in \mathbb{R}^n$ with each nonzero entry following a Gaussian normal distribution. We consider two types of sensing matrices:

(1) Oversampled DCT: $A = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n] \in \mathbb{R}^{m \times n}$ with each column

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

where $\mathbf{w} \in \mathbb{R}^m$ is a random vector following a uniformly distribution on $[0,1]$ and $F \in \mathbb{R}_+$ controls the coherence. A larger value of F yields a more coherent matrix A.

(2) Gaussian matrix: we generate A subject to $\mathcal{N}(\mathbf{0}, \Sigma)$ with the covariance matrix $\Sigma = \{(1-r)I_n(i=j) + r\}_{i,j}$, for a positive parameter r. A larger value of r is more difficult for recovery.

All these algorithms are implemented on MATLAB R2016a and the experiments are performed on a desktop with Windows 10 and an Intel Core i7-7600U CPU processor (2.80GH) with 16GB memory.

6.1. Constructed stationary points. Given a dataset of $(A, \gamma, \mathbf{x}^*)$, we want to generate a vector $\mathbf{b}$ such that the sparse vector $\mathbf{x}^*$ is a stationary point of the unconstrained L_1/L_2 model (1.2). The vector $\mathbf{b}$ can be constructed in a similar way as in [21] for L_1 and in [28] for $L_1 - L_2$. In particular, we examine the first-order optimality condition of (1.2),

$$\gamma \left(\frac{\mathbf{w}^*}{\|\mathbf{x}^*\|_2} - \frac{\mathbf{x}^*\|\mathbf{x}^*\|_1}{\|\mathbf{x}^*\|_2^3} \right) + A^\top(A\mathbf{x}^* - \mathbf{b}) = \mathbf{0}, \tag{6.3}$$

where $\mathbf{w}^* \in \mathrm{SGN}(\mathbf{x}^*)$ defined in (2.1). We need to find $\mathbf{w}^*$ such that $\mathbf{w}^* - \frac{\|\mathbf{x}^*\|_1 \mathbf{x}^*}{\|\mathbf{x}^*\|_2^2} \in \mathcal{R}(A^\top)$. If there exists a vector $\mathbf{y}$ satisfying $A^\top \mathbf{y} = \mathbf{w}^* - \frac{\mathbf{x}^*\|\mathbf{x}^*\|_1}{\|\mathbf{x}^*\|_2^2}$ and we define $\mathbf{b} = \frac{\gamma \mathbf{y}}{\|\mathbf{x}^*\|_2} + A\mathbf{x}^*$, then $\mathbf{x}^*$ is a stationary point to (1.2). In order to find such a vector $\mathbf{w}^* \in \mathbb{R}^n$ numerically, we use an alternating projection method (APM). More specifically, the projection onto the set of $\mathcal{R}(A^\top)$ can be obtained by computing the orthogonal basis U of $A^\top$. In short, the iteration starts at $\mathbf{w}^0 \in \mathrm{SGN}(\mathbf{x}^*)$ and iterates as follows:

$$\mathbf{w}^{k+1} = \mathrm{Proj}_{\mathrm{SGN}(\mathbf{x}^*)}\left(UU^\top(\mathbf{w}^k - \frac{\|\mathbf{x}^*\|_1 \mathbf{x}^*}{\|\mathbf{x}^*\|_2^2}) + \frac{\|\mathbf{x}^*\|_1 \mathbf{x}^*}{\|\mathbf{x}^*\|_2^2} \right).$$

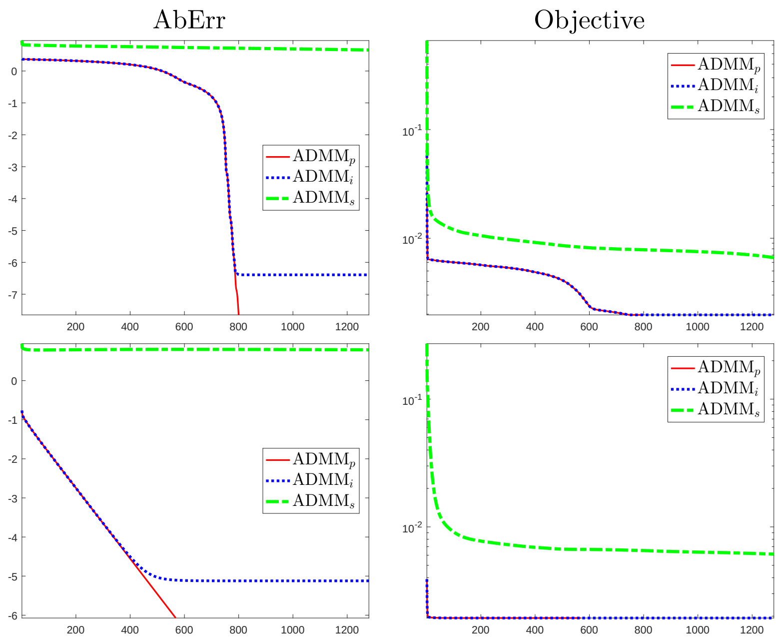

FIG. 2. *Comparison of ADMM$_p$, ADMM$_i$, and ADMM$_s$ in terms of AbErr (left) and objective values (right) using constructed examples based on oversampled DCT (top) and Gaussian matrix (bottom).*

If APM does not converge, we discard this trial. We can check whether APM is successful by computing the following errors:

$$\begin{aligned}&\texttt{err1} = \texttt{norm}(\gamma * (\texttt{w}/\texttt{norm}(\texttt{x}) - \texttt{x} * \texttt{norm}(\texttt{x}, 1)/\texttt{norm}(\texttt{x})^3) + \texttt{A}' * (\texttt{A} * \texttt{x} - \texttt{b}));\\ &\texttt{err2} = \texttt{max}([\texttt{max}(\texttt{w}) - 1, \texttt{min}(\texttt{w}) + 1, \texttt{norm}(\texttt{w}(\texttt{x} > 0) - 1), \texttt{norm}(\texttt{w}(\texttt{x} < 0) + 1)]).\end{aligned}$$

If `err1` is small, it means that the optimality condition (6.3) holds (numerically), and $\texttt{err2} = 0$ confirms that $\mathbf{w}^* \in \mathrm{SGN}(\mathbf{x}^*)$.

With the constructed datasets defined by $(A, \gamma, \mathbf{x}^*, \mathbf{b})$, we compare three different algorithms for solving (1.2): ADMM$_p$, ADMM$_i$, and ADMM$_s$, last of which involves using ADMM to solve the model (1.3). Since that ADMM$_s$ is similar to [34, Algorithm 3.1], the details are omitted here.

We test on two different types of matrices (oversampled DCT and Gaussian matrix) with ground-truth signals of sparsity 5. The size of the sensing matrix is 64×256. We set $\gamma = 0.001, \beta = 0.2$, and the stopping condition given by (6.2) with $\texttt{Tol} := 10^{-8}$ due to unknown noise level. In Table 2, we provide the values of `err1` and `err2` to confirm the constructed solution $\mathbf{x}^*$ is indeed a stationary point of (1.2) with a constructed dataset of $(A, \mathbf{b}, \gamma)$. We then compare three algorithms (ADMM$_p$, ADMM$_i$ and ADMM$_s$) in terms of "AbErr" defined by $\|\mathbf{x}^k - \mathbf{x}^*\|_2$ and "Time" measured in seconds. Table 2 clearly shows that both ADMM$_p$ and ADMM$_i$ (with inner iteration number $j_{\max} = 10$) perform much better than ADMM$_s$ in terms of accuracy and efficiency. Furthermore, we plot "AbErr" and objective values with respect to iteration count k in Figure 2, which shows that ADMM$_p$ converges to the ground-truth solution with the least number of iterations.

Since ADMM$_s$ is not the focus of this work and is worse than ADMM$_p$/ADMM$_i$, we exclude ADMM$_s$ in the rest of experimental comparisons.

Table 2
Results for constructing stationary

Type.1	`err1`: 6.27e-18	`err2`: 0	
	ADMM_p	ADMM_i	ADMM_s
AbErr	1.33e-06	6.02e-06	1.20e+01
Time	0.33	0.68	0.15
Type.2	`err1`: 2.07e-17	`err2`: 0	
	ADMM_p	ADMM_i	ADMM_s
AbErr	4.27e-08	4.13e-07	1.03e+01
Time	0.03	0.67	0.15

6.2. Noisefree. In this section, we show the efficiency of the proposed algorithm, i.e., L_1/L_2 via ADMM_p in the noisefree case. We compare with other sparse recovery models: L_1, $L_{1/2}$ [6], and L_1-L_2 [28] (note that L_1-L_2 is solved by the same ADMM framework as the proposed approach), all in an unconstrained formulation. We use the default setting for these methods, except that the stopping criteria are unified by (6.2). We choose a very small regularization parameter $\gamma = 10^{-6}$ for all these models, since we shall put more weights on the fitting term in the noisefree case. We consider over-sampled DCT matrix with $F = 10$ and Gaussian matrix with $r = 0.9$ of size 64×1024, and the sparsity level ranging from 2 to 24 with an equal increment of 2. As adopted in [48, 34], we evaluate the performance of sparse recovery in terms of success rate, the rates of *model failure* and *algorithm failure*. Success rate is defined as the ratio of the number of successful trials to the total number of trials, in which a trial declared to be successful when the relative error of the reconstructed solution $\hat{\mathbf{x}}$ to the ground truth $\mathbf{x}^*$ is less than 10^{-3}. Furthermore, we compare the objective function $F(\cdot)$ at the ground truth $\mathbf{x}^*$ and the reconstructed solution $\hat{\mathbf{x}}$. If $F(\mathbf{x}^*) > F(\hat{\mathbf{x}})$, then $\mathbf{x}^*$ is not a global minimizer, which is referred to as *model failure*. Otherwise, we have *algorithm failure*, as the algorithm does not achieve a global solution. Note that model/algorithm failure is evaluated by the corresponding objective function of a model. We present success rates and rates of model/algorithm failure for various models in Figure 3. For the oversampled DCT matrices with $F = 10$, L_1-L_2 performs the best, while L_1/L_2 is the second best, and very close to L_1-L_2. For Gaussian matrix with $r = 0.9$, $L_{1/2}$ performs the best, while L_1/L_2 is comparable to $L_{1/2}$. The experimental results are consistent with the literature [34, 27, 47], but it is worth noting that L_1/L_2 is always the second best for both types of the sensing matrices. Based on model/algorithm failure rates in Figure 3, we hypothesize that L_1/L_2 is not sufficient to promote sparsity for the oversampled-DCT matrices, while we need to work on algorithmic improvements of L_1/L_2 for Gaussian matrices.

6.3. Noisy data. Here we provide a series of simulations to demonstrate sparse recovery under a noisy setting. In particular, we consider two types of sensing matrices: over-sampled DCT with $F = 5/10$ and Gaussian matrices with $r = 0.2/0.8$. We fix the ambient dimension and sparsity of the ground-truth signal $\mathbf{x}^*$ as $n = 1024$ and $s = 15$, while varying the number of measurements m from 64 to 128 with an equal increment of 4. The linear measurements are obtained by $\mathbf{b} = A\mathbf{x}^* + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon}$ follows the random Gaussian distribution with standard deviation $\sigma = 0.05$. For all the competing methods, we set the regularization parameter $\gamma = 0.08$, while using the default setting for other algorithmic parameters. In Figure 4, we plot the absolute error between the reconstructed solution ($\bar{\mathbf{x}}$) of a certain method to the ground-truth

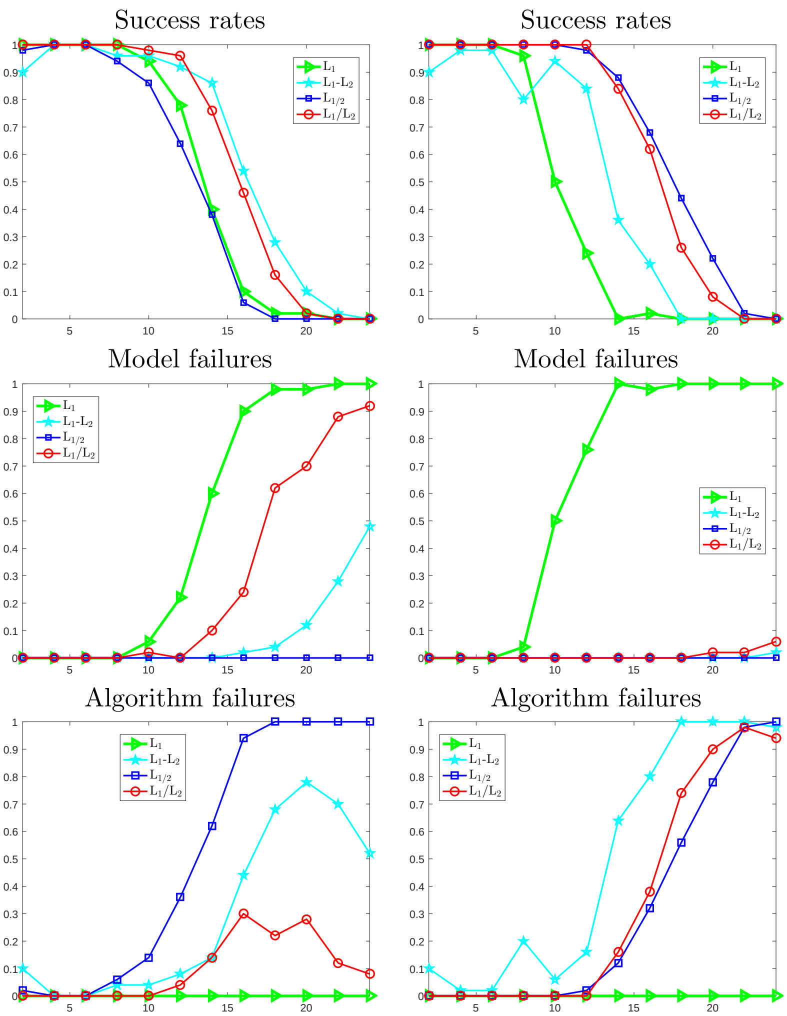

FIG. 3. *Comparison results in the noisefree case based on the oversampled DCT matrix with $F = 10$ (left) and Gaussian matrix with $r = 0.9$ (right). From top to bottom: success rates, model failures, and algorithm failures.*

($\mathbf{x}^*$), i.e., AbErr $= \|\bar{\mathbf{x}} - \mathbf{x}^*\|_2$, with respect to m (the number of measurements). If the support $\Lambda := \text{supp}(\mathbf{x}^*)$ of the ground-truth solution $\mathbf{x}^*$ is known *a priori*, we use the oracle solution $\sigma^2 \text{tr}(A_\Lambda^\top A_\Lambda)^{-1}$ as a benchmark. For the oversampled DCT case, both ADMM$_p$ and ADMM$_i$ perform much better than the other competing methods. As for the Gaussian case of $r = 0.2$, ADMM$_p$/ADMM$_i$ and L_1-L_2 are nearly the same. ADMM$_i$ is slightly better when $r = 0.2$, while ADMM$_p$ is better for $r = 0.8$.

Finally, we list the means and standard deviations of computation time in Table 3 for both oversampled DCT and Gaussian matrices with different m. The proposed ADMM$_p$ approach is comparable to the popular methods of $L_{1/2}$ and L_1-L_2 in time, while ADMM$_i$ is nearly 2-3 times slower than ADMM$_p$. We also observe that ADMM$_p$ requires less iterations to satisfy the stopping criteria (6.2) when m increases, and as a result, the computation time by ADMM$_p$ is inversely proportional to m. The same behavior is observed for L_1-L_2 in the case of Gaussian matrices.

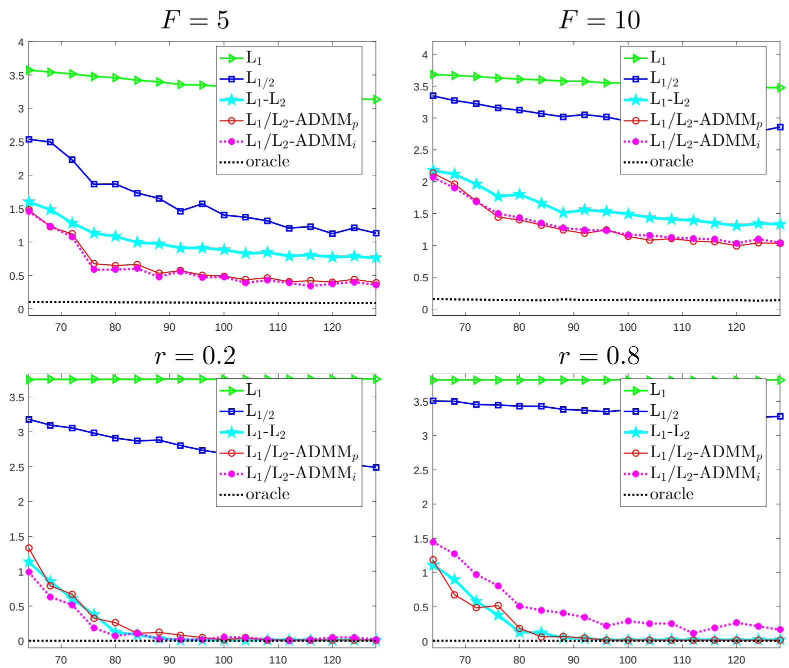

FIG. 4. *Comparison of sparse recovery from noisy data obtained by oversampled DCT (top) and Gaussian matrix (bottom) in terms of AbErr. Each reported value is an average based on 100 random realizations.*

TABLE 3
Average computation time (standard deviation) for sparse recovery from noisy measurements.

m	$L_{1/2}$ [45]	$L_1 - L_2$ [28]	ADMM$_p$	ADMM$_i$
		oversampled DCT matrix with $F = 10$		
64	0.26 (0.10)	0.15 (0.04)	0.26 (0.11)	0.58 (0.08)
76	0.25 (0.08)	0.15 (0.03)	0.23 (0.08)	0.58 (0.05)
88	0.27 (0.10)	0.17 (0.18)	0.21 (0.07)	0.60 (0.07)
100	0.28 (0.10)	0.18 (0.04)	0.20 (0.07)	0.61 (0.05)
112	0.28 (0.10)	0.19 (0.06)	0.21 (0.09)	0.63 (0.10)
124	0.31 (0.13)	0.21 (0.04)	0.20 (0.07)	0.64 (0.07)
		Gaussian matrix with $r = 0.8$		
64	0.31 (0.24)	0.13 (0.05)	0.42 (0.46)	0.68 (0.11)
76	0.33 (0.25)	0.11 (0.04)	0.33 (0.44)	0.71 (0.12)
88	0.37 (0.27)	0.08 (0.03)	0.16 (0.22)	0.71 (0.12)
100	0.35 (0.28)	0.07 (0.02)	0.10 (0.04)	0.72 (0.11)
112	0.41 (0.31)	0.07 (0.01)	0.09 (0.03)	0.71 (0.08)
124	0.47 (0.35)	0.07 (0.01)	0.08 (0.03)	0.74 (0.09)

7. Conclusions. We proposed an unconstrained L_1/L_2 model to deal with both noiseless and noisy observations. We derived a closed-form solution of the proximal operator of L_1/L_2, which involves a solution of a two-dimension nonlinear equations. By verifying the solution's uniqueness of (a, r) when the sparsity is given, we further characterize a solution of the resultant nonlinear system. Thanks to the analytical formula, we developed a practical solver for the proximal operator of L_1/L_2 and designed a specific splitting scheme to apply the ADMM framework when solving the proposed unconstrained model. Under some mild conditions, the global convergence of ADMM for such a nonconvex problem was established, and its linear convergence rate was also analyzed under certain conditions. We demonstrated numerically that the proposed model outperforms some state-of-the-art models in sparse recovery for both noisefree and noisy data, and the proposed ADMM framework converges faster to solutions with higher accuracy than other existing algorithms.

Appendix A. Proof of Theorem 5.8.

Proof. Clearly, $\frac{1}{2}\|A\mathbf{x}-\mathbf{b}\|_2^2$ is a real analytic function. On the other hand, the function $\frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_2}$ is subanalytic, because its graph is a subanalytic subset in $\mathbb{R}^{n+1}$. If at least one of the two subanalytic functions maps bounded sets to bounded sets, then their sum is also subanalytic [3], which guarantees that the objective function F defined in (1.2) is subanalytic. Similarly, the merit function M defined in (5.5) is also subanalytic.

Invoking Theorem 5.1, $\mathbf{0}$ cannot be the global solution of (1.2). Furthermore, any accumulation point $\mathbf{x}^\infty$ of the sequence $\{\mathbf{x}^k\}$ generated by (4.3) cannot be $\mathbf{0}$. Suppose not. Then, $\mathbf{x}^{k_j}\to\mathbf{0}$. Thus, $\mathbf{x}^{k_j+1}\to\mathbf{0}$ due to $\|\mathbf{x}^k-\mathbf{x}^{k+1}\|\to\mathbf{0}$. In view of (4.3a) and Theorem 3.2, one has $\mathbf{y}^{k_j}-\frac{1}{\beta}\mathbf{z}^{k_j}\to\mathbf{0}$. On the other hand, $\mathbf{y}^{k_j}\to\mathbf{0}$ due to $\mathbf{x}^{k_j}-\mathbf{y}^{k_j}\to\mathbf{0}$. Thus, $\mathbf{z}^{k_j}\to\mathbf{0}$. Invoking $\mathbf{z}^{k_j}=A^\top(A\mathbf{y}^{k_j}-\mathbf{b})$ and letting $j\to\infty$, it leads to $\mathbf{0}=A^\top\mathbf{b}$ which contradicts to $\mathbf{b}\notin\mathcal{N}(A^\top)$. Invoking Lemma 5.7, any accumulation point $(\mathbf{x}^\infty,\mathbf{y}^\infty)$ of $\{(\mathbf{x}^k,\mathbf{y}^k)\}$ generated from (4.3) is satisfied with

$$\mathbf{0}\in(\partial_{\mathbf{x}}M(\mathbf{x}^\infty,\mathbf{y}^\infty),\partial_{\mathbf{y}}M(\mathbf{x}^\infty,\mathbf{y}^\infty)),\ \ \mathbf{x}^\infty=\mathbf{y}^\infty.$$

Thus, $(\mathbf{x}^\infty,\mathbf{y}^\infty)$ can not be $(\mathbf{0},\mathbf{0})$.

Thus, we claim that there exists a sufficient small parameter $\epsilon>0$ such that $\mathbf{x}^\infty$ not in the open ball $\mathcal{B}_\epsilon=\{\mathbf{x}|\|\mathbf{x}\|_2<\epsilon\}$. By defining $\tilde{\mathcal{X}}=\{\mathbf{x}\in\mathbb{R}^n|\|\mathbf{x}\|_2\geq\epsilon\}$ and $\tilde{\mathcal{Y}}=\{\mathbf{y}\in\mathbb{R}^n|\|\mathbf{y}\|_2\geq\epsilon\}$, the merit function $M(x,y)|_{\tilde{\mathcal{X}}\times\tilde{\mathcal{Y}}}$ satisfies the KL property due to $M(x,y)|_{\tilde{\mathcal{X}}\times\tilde{\mathcal{Y}}}$ is continuous and closed on its domain. Therefore, the function $M(\mathbf{x},\mathbf{y})$ satisfies the KL property at the stationary point $(\mathbf{x}^\infty,\mathbf{x}^\infty)$. The remaining proof is standard and similar to [23, Theorem 4], thus omitted here. ∎

Appendix B. Proof of Theorem 5.9.

Proof. (i) As proved in Theorem 5.8, the sequence of $\{\mathbf{w}^k\}$ converges to a stationary point $\mathbf{w}^\infty$. First, we prove the assertion of $\text{supp}(\mathbf{x}^k)\equiv\bar{\Lambda}$ when k is sufficient large. Since $\mathbf{x}^k\to\mathbf{x}^\infty$, there exists an index $\tilde{K}$ such that $\text{supp}(\mathbf{x}^k)\equiv\bar{\Lambda}$ when $k\geq\tilde{K}$. Suppose not. Then, there exists a subsequence $\{\mathbf{x}^{k_j}\}_{j=1}^\infty$ such that $\text{supp}(\mathbf{x}^{k_j})\neq\bar{\Lambda}$ for all j. Next, we consider the support set sequence of $\{\text{supp}(\mathbf{x}^{k_j})\}_{j=1}^\infty$. Since the choices of the support set is finite, by invoking the Pigeonhole Principle and restricting the subsequence on hand, we have $\text{supp}(\mathbf{x}^{k_{j_t}})\equiv\hat{\Lambda}$ when $t\in\kappa$ where κ represents the subsequence. Consequently, $\hat{\Lambda}=\bar{\Lambda}$ due to $\mathbf{x}^{k_{j_t}}\to\mathbf{x}^\infty$ which follows from $\mathbf{x}^k\to\mathbf{x}^\infty$. It contradicts to $\text{supp}(\mathbf{x}^{k_j})\neq\bar{\Lambda}$ for all j. Next, we prove the assertion of $\text{sign}(\mathbf{x}^k)=\text{sign}(\mathbf{x}^\infty)$ when k is sufficiently large. Define the scalar

$$\omega:=\inf_{i\in\bar{\Lambda}}|x_i^\infty|, \tag{B.1}$$

and the open ball of $\mathcal{B}_\omega(\mathbf{x}^\infty):=\{\mathbf{x}\mid\|\mathbf{x}-\mathbf{x}^\infty\|_2<\omega\}$. Since $\mathbf{x}^k\to\mathbf{x}^\infty$, there exists an index $K_1(\geq\tilde{K})$ such that $\|\mathbf{x}^k-\mathbf{x}^\infty\|_2<\omega$ when $k\geq K_1$. For these $\mathbf{x}^k$, it holds that $\mathrm{supp}(\mathbf{x}^k) = \mathrm{supp}(\mathbf{x}^\infty) = \bar{\Lambda}$. Thus, for any $i \in \bar{\Lambda}$, $\mathrm{sign}(x_i^k) = \mathrm{sign}(x_i^\infty)$ which follows from $|x_k^k - x_i^\infty| < \omega$ and the definition of ω in (B.1). Consequently, $\mathbf{x}^k \in \Xi$ when $k \geq K_1$. Similarly, there exists an index K_2 such that $\mathbf{y}^k \in \Xi$ when $k \geq K_2$ since $\mathbf{x}^\infty = \mathbf{y}^\infty$. By setting $K := \max\{K_1, K_2\}$, both of $\mathbf{x}^k,\ \mathbf{y}^k \in \Xi$ when $k \geq K$.

(ii) First, we verify the following assertion: For the stationary point $\mathbf{x}^\infty$, there exists a scalar $\bar{\gamma} > 0$ such that F is a KL function restricted on Ξ with an exponent of $1/2$ when $0 < \gamma < \bar{\gamma}$. In order to do so, we introduce $\tilde{F} : \mathbb{R}^\alpha \to \mathbb{R}$ ($\alpha = |\bar{\Lambda}|$), defined by

$$\tilde{F}(\mathbf{u}) := \gamma \frac{\|\mathbf{u}\|_1}{\|\mathbf{u}\|_2} + \frac{1}{2}\|A_{\bar{\Lambda}}\mathbf{u} - \mathbf{b}\|^2.$$

Obviously, $\tilde{F}(\mathbf{x}_{\bar{\Lambda}}) = F(\mathbf{x})$ when $\mathbf{x} \in \Xi$. Define $\Xi_{\bar{\Lambda}} := \{\mathbf{u} \in \mathbb{R}^\alpha \mid \mathbf{u} = \mathbf{x}_{\bar{\Lambda}},\ \mathbf{x} \in \Xi\}$. Since $\mathrm{sign}(\mathbf{u})$ is single-value when $\mathbf{u} \in \Xi_{\bar{\Lambda}}$, $\tilde{F}$ is differentiable on the set $\Xi_{\bar{\Lambda}}$, i.e.,

$$\nabla \tilde{F}(\mathbf{u}) = \gamma \left(\frac{\mathrm{sign}(\mathbf{u})}{\|\mathbf{u}\|_2} - \frac{\|\mathbf{u}\|_1}{\|\mathbf{u}\|_2^3}\mathbf{u} \right) + A_{\bar{\Lambda}}^\top (A_{\bar{\Lambda}}\mathbf{x}_{\bar{\Lambda}} - \mathbf{b}),\ \ \forall\, \mathbf{u} \in \Xi_{\bar{\Lambda}}.$$

Next, define the set $\tilde{\mathcal{X}} = \{\mathbf{x} \in \mathbb{R}^n | \|\mathbf{x}_{\bar{\Lambda}}\|_2 \geq \epsilon\}$ where $\epsilon < \omega$ and ω is defined in (B.1). Obviously, $\mathbf{x}^\infty \in \tilde{\mathcal{X}}$. In the following, we show that there exists a parameter $\bar{\gamma} > 0$ such that when $0 < \gamma < \bar{\gamma}$, it holds that

$$\text{(B.2)} \qquad \|\nabla \tilde{F}(\mathbf{x}_{\bar{\Lambda}}) - \nabla \tilde{F}((\mathbf{x}^\infty)_{\bar{\Lambda}})\| \geq \hat{\tau}\|\mathbf{x}_{\bar{\Lambda}} - (\mathbf{x}^\infty)_{\bar{\Lambda}}\|,\ \ \forall\, \mathbf{x} \in \Xi \cap \tilde{\mathcal{X}},$$

with $\hat{\tau} > 0$. First, we illustrate that the function $f_1(\mathbf{u}) := \dfrac{\mathrm{sign}(\mathbf{u})}{\|\mathbf{u}\|_2} - \dfrac{\|\mathbf{u}\|_1 \mathbf{u}}{\|\mathbf{u}\|_2^3}$ is Lipschitz continuous with constant $\tilde{L} := \frac{6\sqrt{\alpha}}{\epsilon^2}$ on the set of $\Xi_{\bar{\Lambda}} \cap \tilde{\mathcal{X}}_{\bar{\Lambda}}$ where

$$\tilde{\mathcal{X}}_{\bar{\Lambda}} := \{\mathbf{u} \in \mathbb{R}^\alpha \mid \mathbf{u} = \mathbf{x}_{\bar{\Lambda}},\ \mathbf{x} \in \tilde{\mathcal{X}}\}.$$

For any $\mathbf{u},\ \mathbf{v} \in \Xi_{\bar{\Lambda}} \cap \tilde{\mathcal{X}}_{\bar{\Lambda}}$, we denote the scalars of $a_1 := \|\mathbf{u}\|_1$, $a_2 := \|\mathbf{v}\|_1$, $r_1 := \|\mathbf{u}\|_2$, $r_2 := \|\mathbf{v}\|_2$ and the vector $\mathbf{e}_\alpha = \mathrm{sign}((\mathbf{x}^\infty)_{\bar{\Lambda}}) \in \mathbb{R}^\alpha$ and it leads to

$$\begin{aligned}
&\|f_1(\mathbf{u}) - f_1(\mathbf{v})\| \\
&= \left\| \left(\frac{\mathbf{e}_\alpha}{r_1} - \frac{\mathbf{e}_\alpha}{r_2} \right) - \left(\frac{a_1}{r_1^3}\mathbf{u} - \frac{a_2}{r_2^3}\mathbf{v} \right) \right\| \\
&= \left\| \frac{r_2 - r_1}{r_1 r_2}\mathbf{e}_\alpha - \frac{a_1 r_2^3 \mathbf{u} - a_2 r_1^3 \mathbf{v}}{r_1^3 r_2^3} \right\| \\
&\leq \sqrt{\alpha}\frac{\|\mathbf{u} - \mathbf{v}\|_2}{r_1 r_2} + \frac{r_2^3 r_1 |a_2 - a_1| + r_1 a_2 (r_2^3 - r_1^3) + a_2 r_1^3 \|\mathbf{u} - \mathbf{v}\|}{(r_1 r_2)^3} \\
&\leq \frac{6\sqrt{\alpha}}{\epsilon^2}\|\mathbf{u} - \mathbf{v}\|_2.
\end{aligned}$$

Next, we prove (B.2). For any $\mathbf{x} \in \Xi \cap \tilde{\mathcal{X}}$, it leads to

$$\begin{aligned}
&\|\nabla \tilde{F}(\mathbf{x}_{\bar{\Lambda}}) - \nabla \tilde{F}((\mathbf{x}^\infty)_{\bar{\Lambda}})\| \\
&= \|\gamma(f_1(\mathbf{x}_{\bar{\Lambda}}) - f_1((\mathbf{x}^\infty)_{\bar{\Lambda}})) + A_{\bar{\Lambda}}^\top A_{\bar{\Lambda}}(\mathbf{x}_{\bar{\Lambda}} - (\mathbf{x}^\infty)_{\bar{\Lambda}})\| \\
&\geq \|A_{\bar{\Lambda}}^\top A_{\bar{\Lambda}}(\mathbf{x}_{\bar{\Lambda}} - (\mathbf{x}^\infty)_{\bar{\Lambda}})\| - \gamma\|f_1(\mathbf{x}_{\bar{\Lambda}}) - f_1((\mathbf{x}^\infty)_{\bar{\Lambda}})\| \\
&\geq \|A_{\bar{\Lambda}}^\top A_{\bar{\Lambda}}(\mathbf{x}_{\bar{\Lambda}} - (\mathbf{x}^\infty)_{\bar{\Lambda}})\| - \gamma\tilde{L}\|\mathbf{x}_{\bar{\Lambda}} - (\mathbf{x}^\infty)_{\bar{\Lambda}}\| \\
&\geq (\sigma_{\min}(A_{\bar{\Lambda}}^\top A_{\bar{\Lambda}}) - \gamma\tilde{L})\|\mathbf{x}_{\bar{\Lambda}} - (\mathbf{x}^\infty)_{\bar{\Lambda}}\|.
\end{aligned}$$

Let $\bar{\gamma} = \frac{\sigma_{\min}(A_{\bar{\Lambda}}^\top A_{\bar{\Lambda}})}{\tilde{L}}$. Since $A_{\bar{\Lambda}}^\top A_{\bar{\Lambda}} \succ 0$, thus $\sigma_{\min}(A_{\bar{\Lambda}}^\top A_{\bar{\Lambda}}) > 0$ where $\sigma_{\min}(\cdot)$ represents the smallest eigenvalue. Consequently, $\bar{\gamma} > 0$. Let $0 < \gamma < \bar{\gamma}$, the inequality (B.2) holds with $\hat{\tau} := \sigma_{\min}(A_{\bar{\Lambda}}^\top A_{\bar{\Lambda}}) - \gamma \tilde{L} > 0$.

Furthermore, since the function $f_1(\mathbf{u})$ is Lipschitz continuous with constant $\tilde{L}$ on the set of $\Xi_{\bar{\Lambda}} \cap \tilde{\mathcal{X}}_{\bar{\Lambda}}$, it leads to

$$
\begin{aligned}
&\|\nabla \tilde{F}(\mathbf{u}) - \nabla \tilde{F}(\mathbf{v})\| \\
&= \|\gamma(f_1(\mathbf{u}) - f_1(\mathbf{v})) + A_{\bar{\Lambda}}^\top A_{\bar{\Lambda}}(\mathbf{u} - \mathbf{v})\| \\
&\leq \gamma \tilde{L}\|\mathbf{u} - \mathbf{v}\| + \|A_{\bar{\Lambda}}^\top A_{\bar{\Lambda}}(\mathbf{u} - \mathbf{v})\| \\
&\leq (\gamma \tilde{L} + \sigma_{\max}(A_{\bar{\Lambda}}^\top A_{\bar{\Lambda}}))\|\mathbf{u} - \mathbf{v}\|, \;\; \forall\, \mathbf{u}, \mathbf{v} \in \Xi_{\bar{\Lambda}} \cap \tilde{\mathcal{X}}_{\bar{\Lambda}}.
\end{aligned}
$$

Thus, $\nabla \tilde{F}(\mathbf{u})$ is Lipschitz continuous on the set of $\Xi_{\bar{\Lambda}} \cap \tilde{\mathcal{X}}_{\bar{\Lambda}}$ with constant $L_{\tilde{F}} := \gamma \tilde{L} + \sigma_{\max}(A_{\bar{\Lambda}}^\top A_{\bar{\Lambda}})$, i.e.,

$$
\|\nabla \tilde{F}(\mathbf{x}_{\bar{\Lambda}}) - \nabla \tilde{F}((\mathbf{x}^\infty)_{\bar{\Lambda}})\| \leq L_{\tilde{F}} \|\mathbf{x}_{\bar{\Lambda}} - (\mathbf{x}^\infty)_{\bar{\Lambda}}\|, \quad \forall\, \mathbf{x} \in \Xi \cap \tilde{\mathcal{X}}.
$$

On the other hand, for any $\mathbf{x} \in \Xi \cap \tilde{\mathcal{X}}$, we have

$$
\begin{aligned}
&|F(\mathbf{x}) - F(\mathbf{x}^\infty)| = |\tilde{F}(\mathbf{x}_{\bar{\Lambda}}) - \tilde{F}((\mathbf{x}^\infty)_{\bar{\Lambda}})| \\
&\leq (\mathbf{x}_{\bar{\Lambda}} - (\mathbf{x}^\infty)_{\bar{\Lambda}})^\top \nabla \tilde{F}((\mathbf{x}^\infty)_{\bar{\Lambda}}) + \frac{L_{\tilde{F}}}{2} \|\mathbf{x}_{\bar{\Lambda}} - (\mathbf{x}^\infty)_{\bar{\Lambda}}\|^2.
\end{aligned}
$$

Since $\mathbf{x}^\infty$ is a stationary point of (1.2) in the sense of (2.2), it leads to that

(B.3)
$$
\nabla \tilde{F}((\mathbf{x}^\infty)_{\bar{\Lambda}}) = \mathbf{0}.
$$

Consequently,

$$
|F(\mathbf{x}) - F(\mathbf{x}^\infty)| \leq \frac{L_{\tilde{F}}}{2} \|\mathbf{x}_{\bar{\Lambda}} - (\mathbf{x}^\infty)_{\bar{\Lambda}}\|^2, \;\; \forall\, \mathbf{x} \in \Xi \cap \tilde{\mathcal{X}}.
$$

By combining the above inequality with (B.2) and invoking (B.3), we have that

(B.4)
$$
|F(\mathbf{x}) - F(\mathbf{x}^\infty)| \leq \tilde{c} \|\nabla \tilde{F}(\mathbf{x}_{\bar{\Lambda}})\|^2, \;\; \forall\, \mathbf{x} \in \Xi \cap \tilde{\mathcal{X}},
$$

with $\tilde{c} := \frac{L_{\tilde{F}}}{2(\hat{\tau})^2}$. Finally, for $\mathbf{x} \in \mathcal{B} := \{\mathbf{x} \in \mathbb{R}^n \;\mid\; \mathbf{x} \in \Xi \cap U(\mathbf{x}^\infty) \text{ and } F(\mathbf{x}^\infty) < F(\mathbf{x}) < F(\mathbf{x}^\infty) + \nu\}$, where $U(\mathbf{x}^\infty)$ is a neighborhood of $\mathbf{x}^\infty$ and $U(\mathbf{x}^\infty) \subseteq \tilde{\mathcal{X}}$, it yields that

$$
\begin{aligned}
\operatorname{dist}^2(\mathbf{0}, \partial F(\mathbf{x})) &= \inf_{\substack{\xi \,\in\, \partial F(\mathbf{x}) \\ \mathbf{x} \,\in\, \mathcal{B}}} \|\xi\|^2 \\
&\geq \inf_{\substack{\xi \,\in\, \partial F(\mathbf{x}) \\ \mathbf{x} \,\in\, \mathcal{B}}} \|\xi_{\bar{\Lambda}}\|^2 = \inf_{\mathbf{x} \in \mathcal{B}} \|\nabla \tilde{F}(\mathbf{x}_{\bar{\Lambda}})\|^2.
\end{aligned}
$$

By combining the above inequality with (B.4), one has that F is a KL function restricted on Ξ with an exponent of $1/2$.

Finally, we have that $M(\mathbf{x}, \mathbf{y})$ is a KL function with an exponent of $1/2$ at $(\mathbf{x}^\infty, \mathbf{x}^\infty)$ when it is restricted on $\Xi \times \Xi$ by following the proof to [24, Theorem 3.6]. In view of the assertion (i), both of $\mathbf{x}^k$ and $\mathbf{y}^k$ will enter into the set of Ξ when $k \geq K$. Thus, the assertion (5.7) follows directly by following the proof in [1]. We omit the proof here for brevity. ∎

REFERENCES

[1] H. Attouch and J. Bolte, *On the convergence of the proximal algorithm for nonsmooth functions involving analytic features*, Math. Program., 116 (2009), pp. 5-16.

[2] A. S. Bandeira, E. Dobriban, D. G. Mixon and W. F. Sawin, *Certifying the restricted isometry property is hard*, IEEE Trans. Inf. Theory, 59 (2013), pp. 3448-3450.

[3] J. Bolte, A. Daniilidis and A. Lewis, *The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems*, SIAM J. Optim., 17 (2007), pp. 1205-1223.

[4] 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.

[5] E. J. Candès and M. B. Wakin, *An introduction to compressive sampling*, IEEE Signal Process. Mag., 25 (2008), pp. 21–30.

[6] R. Chartrand, *Exact reconstruction of sparse signals via nonconvex minimization*, IEEE Signal Process. Lett., 10 (2007), pp. 707–710.

[7] S. S. Chen, D. L. Donoho and M. A. Saunders, *Atomic decomposition by basis pursuit*, SIAM J. Sci. Comput., 20 (1998), pp. 33-61.

[8] D. L. Donoho and X. Huo, *Uncertainty principles and ideal atomic decomposition*, IEEE Transactions on Information Theory, 47 (2001), pp. 2845-2862.

[9] D. L. Donoho and M. Elad, *Optimally sparse representation in general (nonorthogonl) dictionaries via ℓ_1 minimization*, Proc. Nat. Acad. Scien. USA, 100 (2003), pp. 2197–2202.

[10] 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.

[11] J. Fan and R. Li, *Variable selection via nonconcave penalized likelihood and its oracle properties*, J. Am. Stat. Assoc., 96 (2001), pp. 1348–1360.

[12] M. Figueiredo, R. Nowak and S. J. Wright, *Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems*, IEEE J. Sel. Top. Signa., 1 (2007), pp. 586–597.

[13] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, Springer, 2013.

[14] T. Goldstein and S. Osher,*The split Bregman method for ℓ_1-regularized problems*, SIAM J. Imaging Sci., 2 (2009), pp. 323-343.

[15] K. Guo, D. Han and 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.

[16] P. O. Hoyer, *Non-negative sparse coding*, in Proc. IEEE Workshop on Neural Networks for Signal Processing, (2002), pp. 557–565.

[17] N. Hurley and S. Rickard, *Comparing measures of sparsity*, IEEE Trans. on Inform. Theory, 55 (2009), pp. 4723–4741.

[18] H. Ji, J. Li, Z. Shen and K. Wang, *Image deconvolution using a characterization of sharp images in wavelet domain*, Appl. Comput. Harmon. A., 32 (2012), pp. 295–304.

[19] K. Kurdyka, On gradients of functions definable in o-minimal structures, Annales de linstitut Fourier (Grenoble) 48(1998), pp. 769-783.

[20] D. Krishnan, T. Tay and R. Fergus, *Blind deconvolution using a normalized sparsity measure*, in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (2011), pp. 233–240.

[21] D. A. Lorenz, *Constructing test instances for basis pursuit denoising*, Trans. Signal Process, 61 (2012), pp. 1210–1214.

[22] M.-J. Lai, Y. Xu and W. Yin, *Improved iteratively reweighted least squares for unconstrained smoothed ℓ_q minimization*, SIAM J. Numer. Anal., 5 (2013), pp. 927-957.

[23] G. Li and T. K. Pong, *Global convergence of splitting methods for nonconvex composite optimization*, SIAM J. Optim., 25 (2015), pp. 2434-2460.

[24] G. Li and T. K. Pong, *Calculus of the exponent of Kurdyka-Łojasiewicz inequality and its applications to linear convergence of first-order methods*, Found Comput Math, 18 (2018), pp. 1199-1232.

[25] S. Łojasiewicz, *Une propri été topologique des sous-ensembles analytiques reels, Les Équations aux Dérivées Partielles, Éditions du Centre National de la Recherche Scientifique Paris*, 87-89, 1963.

[26] Y. Lou, S. Osher and J. Xin, *Computational aspects of constrained L1-L2 minimization for compressive sensing*, in Model. Comput. & Optim. Inf. Syst. & Manage. Sci., Advances in Intelligent Systems and Computing, 359 (2015), pp. 169-180.

[27] Y. Lou, P. Yin, Q. He and J. Xin, *Computing sparse representation in a highly coherent*

dictionary based on difference of L1 and L2, J. Sci. Comput., 64 (2015), pp. 178-196.
[28] Y. Lou and M. Yan, *Fast L_1-L_2 minimization via a proximal operator*, J. Sci. Comput., 74 (2017), pp. 767-785.
[29] M. Hong, Z.-Q. Luo and M. Razaviyayn, *Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems*, SIAM J. Optimiz., 26 (2016), pp. 337-364.
[30] M. M. Mäkelä and P. Neittaanmäki, *Nonsmooth Optimization: Analysis and Algorithms with Applications to Optimal Control*, World Scientific Publishing, 1992.
[31] N. Meinshausen and B. Yu, *Lasso-type recovery of sparse representations for high-dimensional data*, Ann. Stat., 37 (2009), pp. 246–270.
[32] B. K. Natarajan, *Sparse approximate solutions to linear systems*, SIAM J. Comput., 24 (1995), pp. 227–234.
[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] Y. Rahimi, C. Wang, H. Dong and Y. Lou, *A scale-invariant approach for sparse signal recovery*, SIAM J. Sci. Comput., 41 (2019), A3649–A3672.
[35] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer, 1998.
[36] F. Santosa and W. W. Symes, *Linear inversion of band-limited reflection seismograms*, SIAM J. Sci. Stat. Comput., 7 (1986), pp. 1307–1330.
[37] R. Tibshirani, *Regression shrinkage and selection via the lasso*, J. Royal Stat. Soc. Ser. B, 58 (1996), pp. 267–288.
[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. Inf. Theory, 60 (2014), pp. 1248–1259.
[39] C. Wang, M. Yan, Y. Rahimi and Y. Lou, *Accelerated schemes for the L_1/L_2 minimization*, IEEE Trans. Signal Process., 68 (2020), pp. 2660–2669.
[40] C. Wang, M. Tao, J. Nagy and Y. Lou, *Limited-angle CT reconstruction via the L1/L2 minimization*, available at https://arxiv.org/abs/2006.00601.
[41] Y. Wang, W. Yin and J. Zeng, *Global convergence of ADMM in nonconvex nonsmooth optimization*, J. Sci. Comput., 78 (2019), pp. 1-35.
[42] J. Woodworth and R. Chartrand, *Compressed sensing recovery via nonconvex shrinkage penalties*, Inverse Probl., 32 (2016), 75004, 25 pp.
[43] L. Wu, Z. Sun and D.-H. Li, *A Barzilai-Borwein-like iterative half thresholding algorithm for the $\ell_{1/2}$ regularized problem*, J. Sci. Comput., 67 (2016), pp. 581–601.
[44] F.-C. Xing, *Investigation on solutions of cubic equations with one unknown*, J. Central Univ. Nat. (Natural Sci. Ed.), 12 (2003), pp. 207–218.
[45] 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.
[46] Z. Xu, H. Guo, Y. Wang and H. Zhang, *Representative of $L_{1/2}$ regularization among L_q ($0 < q \leq 1$) regularizations: an experimental study based on phase diagram*, Acta Autom. Sinica, 38 (2012), pp. 1225-1228.
[47] P. Yin, E. Esser and J. Xin, *Ratio and difference of ℓ_1 and ℓ_2 norms and sparse representation with coherent dictionaries*, Comm. Info. Systems, 14 (2014), pp. 87-109.
[48] P. Yin, Y. Lou, Q. He and J. Xin, *Minimization of ℓ_{1-2} for compressed sensing*, SIAM J. Sci. Comput., 37 (2015), pp. A536-A563.
[49] S. Zhang and J. Xin, *Minimization of transformed ℓ_1 penalty: closed form representation and iterative thresholding algorithms*, Comm. Math. Sci., 15 (2017), pp. 511-537.
[50] 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.
[51] L. Zeng, P. Yu and T. K. Pong, *Analysis and algorithms for some compressed sensing models based on L1/L2 minimization*, available at https:/arxiv.org/abs/2007.12821.