# ANALYSIS AND ALGORITHMS FOR SOME COMPRESSED SENSING MODELS BASED ON L1/L2 MINIMIZATION*

LIAOYUAN ZENG[†], PEIRAN YU[†], AND TING KEI PONG[†]

**Abstract.** Recently, in a series of papers [Y. Rahimi, C. Wang, H. Dong, and Y. Lou, *SIAM J. Sci. Comput.*, 41 (2019), pp. A3649–A3672; C. Wang, M. Tao, J. Nagy, and Y. Lou, *SIAM J. Imaging Sci.*, 14 (2021), pp. 749–777; C. Wang, M. Yan, and Y. Lou, *IEEE Trans. Signal Process.*, 68 (2020), pp. 2660–2669; P. Yin, E. Esser, and J. Xin, *Commun. Inf. Syst.*, 14 (2014), pp. 87–109], the ratio of $\ell_1$ and $\ell_2$ norms was proposed as a sparsity inducing function for noiseless compressed sensing. In this paper, we further study properties of such model in the noiseless setting, and propose an algorithm for minimizing $\ell_1/\ell_2$ subject to noise in the measurements. Specifically, we show that the extended objective function (the sum of the objective and the indicator function of the constraint set) of the model in [Y. Rahimi, C. Wang, H. Dong, and Y. Lou, *SIAM J. Sci. Comput.*, 41 (2019), pp. A3649–A3672] satisfies the Kurdyka–Łojasiewicz (KL) property with exponent 1/2; this allows us to establish linear convergence of the algorithm proposed in [C. Wang, M. Yan, and Y. Lou, *IEEE Trans. Signal Process.*, 68 (2020), pp. 2660–2669] (see equation 11) under mild assumptions. We next extend the $\ell_1/\ell_2$ model to handle compressed sensing problems with noise. We establish the solution existence for some of these models under the spherical section property [S. A. Vavasis, *Derivation of Compressive Sensing Theorems from the Spherical Section Property*, University of Waterloo, 2009; Y. Zhang, *J. Oper. Res. Soc. China*, 1 (2013), pp. 79–105] and extend the algorithm in [C. Wang, M. Yan, and Y. Lou, *IEEE Trans. Signal Process.*, 68 (2020), pp. 2660–2669] (see equation 11) by incorporating moving-balls-approximation techniques [A. Auslender, R. Shefi, and M. Teboulle, *SIAM J. Optim.*, 20 (2010), pp. 3232–3259] for solving these problems. We prove the subsequential convergence of our algorithm under mild conditions and establish global convergence of the whole sequence generated by our algorithm by imposing additional KL and differentiability assumptions on a specially constructed potential function. Finally, we perform numerical experiments on robust compressed sensing and basis pursuit denoising with residual error measured by $\ell_2$ norm or Lorentzian norm via solving the corresponding $\ell_1/\ell_2$ models by our algorithm. Our numerical simulations show that our algorithm is able to recover the original sparse vectors with reasonable accuracy.

**Key words.** L1/L2 minimization, Kurdyka–Lojasiewicz exponent, moving balls approximation, linear convergence

**AMS subject classifications.** 90C26, 90C32, 90C55, 90C90

**DOI.** 10.1137/20M1355380

**1. Introduction.** In compressed sensing (CS), a high-dimensional sparse or approximately sparse signal $x_0 \in \mathbb{R}^n$ is compressed (linearly) as $Ax_0$ for transmission, where $A \in \mathbb{R}^{m \times n}$ is the sensing matrix. The CS problem seeks to recover the original signal $x_0$ from the possibly noisy low-dimensional measurement $b \in \mathbb{R}^m$. This problem is NP-hard in general; see [29].

When there is no noise in the transmission, i.e., $Ax_0 = b$, one can recover $x_0$ exactly by minimizing the $\ell_1$ norm over $A^{-1}\{b\}$ if $x_0$ is sufficiently sparse and the matrix $A$ satisfies certain assumptions [15, 19]. To empirically enhance the recovery ability, various nonconvex models like the $\ell_p$ $(0 < p < 1)$ minimization model [18] and the $\ell_{1-2}$ minimization model [27] have been proposed, in which the $\ell_p$ quasi-norm and

†Department of Applied Mathematics, Hong Kong Polytechnic University, Hong Kong, People's Republic of China (zengly@lsec.cc.ac.cn, peiran.yu@connect.polyu.hk, tk.pong@polyu.edu.hk).

the difference of $\ell_1$ and $\ell_2$ norms are minimized over $A^{-1}\{b\}$, respectively. Recently, a new nonconvex model based on minimizing the quotient of the $\ell_1$ and $\ell_2$ norms was introduced in [32, 41] and further studied in [38, 39]:

$$(1.1) \qquad \nu_{cs}^* := \min_{x \in \mathbb{R}^n} \frac{\|x\|_1}{\|x\|} \qquad \text{s.t.} \quad Ax = b,$$

where $A \in \mathbb{R}^{m \times n}$ has full row rank and $b \in \mathbb{R}^m \backslash \{0\}$. As discussed in [32], the above $\ell_1/\ell_2$ model has the advantage of being *scale-invariant* when reconstructing signals and images with high dynamic range. An efficient algorithm was proposed for solving (1.1) in [39, equation 11] and subsequential convergence was established under mild assumptions.

In practice, however, there is noise in the measurement, i.e., $b = Ax_0 + \epsilon$ for some noise vector $\epsilon$, and (1.1) is not applicable for (approximately) recovering $x_0$. To deal with noisy situations, it is customary to relax the equality constraint in (1.1) to an inequality constraint [14]. In this paper, we consider the following model that minimizes the $\ell_1/\ell_2$ objective over an inequality constraint:

$$(1.2) \qquad \nu_{ncs}^* = \min_{x \in \mathbb{R}^n} \frac{\|x\|_1}{\|x\|} \qquad \text{s.t. } q(x) \leq 0,$$

where $q(x) = P_1(x) - P_2(x)$ with $P_1 : \mathbb{R}^n \to \mathbb{R}$ being continuously differentiable with Lipschitz continuous gradient and $P_2 : \mathbb{R}^n \to \mathbb{R}$ being convex continuous, and we assume that $\{x : q(x) \leq 0\} \neq \emptyset$ and $q(0) > 0$. Our assumptions on $q$ are general enough to cover commonly used loss functions for modeling noise in various scenarios:

1. **Gaussian noise**: When the noise in the measurement follows the Gaussian distribution, the least squares loss function $y \mapsto \|y - b\|^2$ is typically employed [14, 19]. One may consider the following $\ell_1/\ell_2$ minimization problem:

$$(1.3) \qquad \min_{x \in \mathbb{R}^n} \frac{\|x\|_1}{\|x\|} \qquad \text{s.t. } \|Ax - b\|^2 - \sigma^2 \leq 0,$$

   where $\sigma > 0$, $A \in \mathbb{R}^{m \times n}$ has full row rank, and $b \in \mathbb{R}^m$ satisfies $\|b\| > \sigma$. Problem (1.3) corresponds to (1.2) with $q(x) = P_1(x) = \|Ax - b\|^2 - \sigma^2$ and $P_2 = 0$.

2. **Cauchy noise**: When the noise in the measurement follows the Cauchy distribution (a heavy-tailed distribution), the Lorentzian norm[1] $\|y\|_{LL_2,\gamma} := \sum_{i=1}^m \log\left(1 + \gamma^{-2} y_i^2\right)$ is used as the loss function [16, 17], where $\gamma > 0$. Note that the Lorentzian norm is continuously differentiable with Lipschitz continuous gradient. One may then consider the following $\ell_1/\ell_2$ minimization problem:

$$(1.4) \qquad \min_{x \in \mathbb{R}^n} \frac{\|x\|_1}{\|x\|} \qquad \text{s.t. } \|Ax - b\|_{LL_2,\gamma} - \sigma \leq 0,$$

   where $\sigma > 0$, $A \in \mathbb{R}^{m \times n}$ has full row rank, and $b \in \mathbb{R}^m$ with $\|b\|_{LL_2,\gamma} > \sigma$. Problem (1.4) corresponds to (1.2) with $q(x) = P_1(x) = \|Ax - b\|_{LL_2,\gamma} - \sigma$ and $P_2 = 0$.

---

[1]We refer readers to [17, equation (12)] for the definition and notation of the Lorentzian norm.

3. **Robust compressed sensing**: In this scenario, the measurement is corrupted by both Gaussian noise and electromyographic noise [17, 31]: the latter is sparse and may have large magnitude (outliers). Following [26, section 5.1.1], one may make use of the loss function $y \mapsto \mathrm{dist}^2(y, S)$, where $S := \{z \in \mathbb{R}^m : \|z\|_0 \le r\}$, $\|z\|_0$ is the number of nonzero entries in $z$, and $r$ is an estimate of the number of outliers. One may then consider the following $\ell_1/\ell_2$ minimization problem:

$$(1.5) \qquad \min_{x \in \mathbb{R}^n} \frac{\|x\|_1}{\|x\|} \qquad \text{s.t. } \mathrm{dist}^2(Ax - b, S) - \sigma^2 \le 0,$$

where $\sigma > 0$, $S = \{z \in \mathbb{R}^m : \|z\|_0 \le r\}$ with $r \ge 0$, $A \in \mathbb{R}^{m \times n}$ has full row rank and $b \in \mathbb{R}^m$ satisfies $\mathrm{dist}(b, S) > \sigma$. Notice that

$$(1.6) \quad \mathrm{dist}^2(Ax - b, S) - \sigma^2 = \underbrace{\|Ax - b\|^2 - \sigma^2}_{P_1(x)} - \underbrace{\max_{z \in S}\{\langle 2z, Ax - b\rangle - \|z\|^2\}}_{P_2(x)}$$

with $P_1$ being continuously differentiable with Lipschitz continuous gradient and $P_2$ being convex continuous. So this problem corresponds to (1.2) with $P_1$ and $P_2$ as in (1.6) and $q = P_1 - P_2$.

In the literature, algorithms for solving (1.3) with $\ell_1$ norm or $\ell_p$ quasi-norm in place of the quotient of the $\ell_1$ and $\ell_2$ norms have been discussed in [5, 20, 35], and [43] discussed an algorithm for solving (1.4) with $\ell_1$ norm in place of the quotient of the $\ell_1$ and $\ell_2$ norms. These existing algorithms, however, are not directly applicable for solving (1.2) due to the fractional objective and the possibly nonsmooth continuous function $q$ in the constraint.

In this paper, we further study properties of the $\ell_1/\ell_2$ models (1.1) and (1.2) and propose an algorithm for solving (1.2). In particular, we first argue that an optimal solution of (1.1) exists by making connections with the $s$-spherical section property [37, 44] of $\ker A$: a property which is known to hold with high probability when $n \gg m$ for Gaussian matrices. We then revisit the algorithm proposed in [39, equation 11] (see Algorithm 4.1 below) for solving (1.1). Specifically, we consider the following function:

$$(1.7) \qquad F(x) := \frac{\|x\|_1}{\|x\|} + \delta_{A^{-1}\{b\}}(x),$$

where $A \in \mathbb{R}^{m \times n}$ has full row rank and $b \in \mathbb{R}^m \backslash \{0\}$. We show in section 4.2 that $F$ is a Kurdyka–Łojasiewicz (KL) function with exponent $\frac{1}{2}$. This together with standard convergence analysis based on the KL property [1, 2, 3] allows us to deduce local linear convergence of the sequence $\{x^t\}$ generated by Algorithm 4.1 when $\{x^t\}$ is bounded. The KL exponent of $F$ is obtained based on a new calculus rule that deduces the KL exponent of a fractional objective from the difference between the numerator and (a suitable scaling of) the denominator.

Next, for the model (1.2), we also relate existence of solutions to the $s$-spherical section property of $\ker A$ when $q$ takes the form in (1.3) and (1.4). We then propose an algorithm, which we call $\mathrm{MBA}_{\ell_1/\ell_2}$ (see Algorithm 6.1), for solving (1.2), which can be seen as an extension of Algorithm 4.1 by incorporating *moving-balls-approximation* (MBA) techniques. The MBA algorithm was first proposed in [4] for minimizing a smooth objective function subject to multiple smooth constraints and was further studied in [6, 8, 43] for more general objective functions. However, the existing convergence results of these algorithms cannot be applied to $\mathrm{MBA}_{\ell_1/\ell_2}$ because of the

possibly nonsmooth continuous function $q$ and the fractional objective in (1.2). Our convergence analysis of $\text{MBA}_{\ell_1/\ell_2}$ relies on a specially constructed potential function, which involves the indicator function of the lower level set of a *proper closed* function related to $q$ (see (6.8)). We prove that any accumulation point of the sequence generated by $\text{MBA}_{\ell_1/\ell_2}$ is a so-called Clarke critical point, under mild assumptions; Clarke criticality reduces to the usual notion of stationarity when $q$ is regular. Moreover, by imposing additional KL assumptions on this potential function and assuming $P_1$ is twice continuously differentiable, we show that the sequence generated by $\text{MBA}_{\ell_1/\ell_2}$ is globally convergent, and the convergence rate is related to the KL exponent of the potential function. Finally, we perform numerical experiments to illustrate the performance of our algorithm on solving (1.3), (1.4), and (1.5).

The rest of this paper is organized as follows. We present notation and some preliminaries in section 2. Existence of solutions for (1.1) is discussed in section 3. In section 4, we derive the KL exponent of $F$ in (1.7) and establish local linear convergence of the algorithm proposed in [39, equation 11]. Properties of (1.2) such as solution existence and optimality conditions are discussed in section 5, and our algorithm, $\text{MBA}_{\ell_1/\ell_2}$, for solving (1.2) is proposed and analyzed in section 6. Finally, numerical results are presented in section 7.

**2. Notation and preliminaries.** In this paper, we use $\mathbb{R}^n$ to denote the Euclidean space of dimension $n$ and use $\mathbb{N}_+$ to denote the set of nonnegative integers. For two vectors $x$ and $y \in \mathbb{R}^n$, we use $\langle x, y \rangle$ to denote their inner product, i.e., $\langle x, y \rangle = \sum_{i=1}^n x_i y_i$. The Euclidean norm, the $\ell_1$ norm, and the $\ell_0$ norm (i.e., the number of nonzero entries) of $x$ are denoted respectively by $\|x\|$, $\|x\|_1$, and $\|x\|_0$. We also use $B(x, r)$ to denote a closed ball centered at $x$ with radius $r \geq 0$, i.e., $B(x, r) = \{y : \|y - x\| \leq r\}$.

An extended-real-valued function $h : \mathbb{R}^n \to (-\infty, \infty]$ is said to be proper if its domain $\text{dom}\, h := \{x : h(x) < \infty\}$ is nonempty. A proper function $h$ is said to be closed if it is lower semicontinuous. For a proper closed function $h$, the regular subdifferential $\widehat{\partial} h(\bar{x})$ at $\bar{x} \in \text{dom}\, h$ and the limiting subdifferential $\partial h(\bar{x})$ are given respectively as

$$\widehat{\partial} h(\bar{x}) := \left\{ v : \liminf_{x \to \bar{x}, x \neq \bar{x}} \frac{h(x) - h(\bar{x}) - \langle v, x - \bar{x} \rangle}{\|x - \bar{x}\|} \geq 0 \right\},$$

$$\partial h(\bar{x}) := \left\{ v : \exists x^t \xrightarrow{h} \bar{x} \text{ and } v^t \in \widehat{\partial} h(x^t) \text{ with } v^t \to v \right\},$$

where $x^t \xrightarrow{h} \bar{x}$ means $x^t \to \bar{x}$ and $h(x^t) \to h(\bar{x})$. In addition, we set $\partial h(x) = \widehat{\partial} h(x) = \emptyset$ by convention when $x \notin \text{dom}\, h$, and we define $\text{dom}\, \partial h := \{x : \partial h(x) \neq \emptyset\}$. It is known that $\partial h(x) = \{\nabla h(x)\}$ if $h$ is continuously differentiable at $x$ [34, Exercise 8.8(b)]. For a proper closed convex function, the limiting subdifferential reduces to the classical subdifferential for convex functions [34, Proposition 8.12]. The convex conjugate of a proper closed convex function $h$ is defined as

$$h^*(y) = \sup_{x \in \mathbb{R}^n} \{\langle x, y \rangle - h(x)\}.$$

We recall the following relationship concerning convex conjugate and subdifferential of a proper closed convex function $h$; see [34, Proposition 11.3]:

(2.1)
$$y \in \partial h(x) \iff x \in \partial h^*(y) \iff h(x) + h^*(y) \leq \langle x, y \rangle \iff h(x) + h^*(y) = \langle x, y \rangle.$$

For a locally Lipschitz function $h$, its Clarke subdifferential at $\bar{x} \in \mathbb{R}^n$ is defined by

$$\partial^\circ h(\bar{x}) := \left\{ v : \limsup_{x \to \bar{x}, t \downarrow 0} \frac{h(x + tw) - h(x)}{t} \geq \langle v, w \rangle \text{ for all } w \in \mathbb{R}^n \right\};$$

it holds that $\partial h(\bar{x}) \subseteq \partial^\circ h(\bar{x})$ (see [11, Theorem 5.2.22]). Finally, for a proper closed function $f$, we say that $\bar{x}$ is a stationary point of $f$ if $0 \in \partial f(\bar{x})$. All local minimizers are stationary points according to [34, Theorem 10.1].

For a nonempty closed set $C$, we define the indicator function as

$$\delta_C(x) := \begin{cases} 0, & x \in C, \\ \infty, & x \notin C. \end{cases}$$

The normal cone (resp., regular normal cone) of $C$ at an $x \in C$ is given by $N_C(x) := \partial \delta_C(x)$ (resp., $\widehat{N}_C(x) := \widehat{\partial} \delta_C(x)$). The distance from a point $x$ to $C$ is denoted by $\text{dist}(x, C)$. The set of points in $C$ that are closest to $x$ is denoted by $\text{Proj}_C(x)$. The convex hull of $C$ is denoted by $\text{conv}\, C$.

We next recall the KL property. This property and the associated notion of KL exponent have been used extensively in the convergence analysis of various first-order methods; see [1, 2, 3, 9, 25].

DEFINITION 2.1 (KL property and exponent). *We say that a proper closed function $h : \mathbb{R}^n \to (-\infty, \infty]$ satisfies the KL property at an $\widehat{x} \in \text{dom}\, \partial h$ if there are $a \in (0, \infty]$, a neighborhood $U$ of $\widehat{x}$, and a continuous concave function $\varphi : [0, a) \to [0, \infty)$ with $\varphi(0) = 0$ such that*
(i) *$\varphi$ is continuously differentiable on $(0, a)$ with $\varphi' > 0$ on $(0, a)$;*
(ii) *for every $x \in U$ with $h(\widehat{x}) < h(x) < h(\widehat{x}) + a$, it holds that*

$$(2.2) \qquad\qquad \varphi'(h(x) - h(\widehat{x}))\text{dist}(0, \partial h(x)) \geq 1.$$

*If $h$ satisfies the KL property at $\widehat{x} \in \text{dom}\, \partial h$ and the $\varphi$ in (2.2) can be chosen as $\varphi(\nu) = a_0 \nu^{1-\theta}$ for some $a_0 > 0$ and $\theta \in [0, 1)$, then we say that $h$ satisfies the KL property at $\widehat{x}$ with exponent $\theta$.*

*A proper closed function $h$ satisfying the KL property at every point in $\text{dom}\, \partial h$ is called a KL function, and a proper closed function $h$ satisfying the KL property with exponent $\theta \in [0, 1)$ at every point in $\text{dom}\, \partial h$ is called a KL function with exponent $\theta$.*

KL functions arise naturally in various applications. For instance, proper closed semialgebraic functions are KL functions with some exponent $\theta \in [0, 1)$; see [2].

Another notion that will be needed in our discussion later is (subdifferential) regularity; see [34, Definition 6.4] and [34, Definition 7.25].

DEFINITION 2.2. *A nonempty closed set $C$ is regular at $x \in C$ if $N_C(x) = \widehat{N}_C(x)$, and a proper closed function $h$ is (subdifferentially) regular at $x \in \text{dom}\, h$ if its epigraph $\text{epi}\, h := \{(x, t) \in \mathbb{R}^n \times \mathbb{R} : h(x) \leq t\}$ is regular at $(x, h(x))$.*

According to [34, Example 7.28], continuously differentiable functions are regular everywhere. Thus, the constraint functions in (1.3) and (1.4) are regular everywhere. In addition, a nonsmooth regular function particularly relevant to our discussion is the objective function of (1.1). Indeed, in view of [28, Corollary 1.111(i)], it holds that

(2.3)     At any $\bar{x} \neq 0$, $\frac{\|\cdot\|_1}{\|\cdot\|}$ is regular and $\partial \frac{\|\bar{x}\|_1}{\|\bar{x}\|} = \frac{1}{\|\bar{x}\|} \partial \|\bar{x}\|_1 - \frac{\|\bar{x}\|_1}{\|\bar{x}\|^3} \bar{x}$.

We will also need the following auxiliary lemma concerning the subdifferential of a particular class of functions in our analysis in section 6.

LEMMA 2.3. *Let $q = P_1 - P_2$ with $P_1 : \mathbb{R}^n \to \mathbb{R}$ being continuously differentiable and $P_2 : \mathbb{R}^n \to \mathbb{R}$ being convex continuous. Then for any $x \in \mathbb{R}^n$, we have*

(2.4)                     $\partial^\circ q(x) = \nabla P_1(x) - \partial P_2(x).$

*Proof.* Note that for any $x \in \mathbb{R}^n$, we have

$$\partial^\circ q(x) \overset{\text{(a)}}{=} \nabla P_1(x) + \partial^\circ(-P_2)(x) \overset{\text{(b)}}{=} \nabla P_1(x) - \partial^\circ P_2(x) \overset{\text{(c)}}{=} \nabla P_1(x) - \partial P_2(x),$$

where (a) follows from Corollary 1 of [21, Proposition 2.3.3], (b) holds because of [21, Proposition 2.3.1], and (c) follows from [21, Proposition 2.2.7]. □

**3. Solution existence of model (1.1).** In this section, we establish the existence of optimal solutions to problem (1.1) under suitable assumptions. A similar discussion was made in [32, Theorem 2.2], where the existence of *local minimizers* was established under the strong null space property (see [32, Definition 2.1]) of the sensing matrix $A$. It was indeed shown that any sufficiently sparse solution of $Ax = b$ is a local minimizer for problem (1.1), under the strong null space property. Here, our discussion focuses on the existence of *globally optimal solutions*, and our analysis is based on the spherical section property [37, 44].

DEFINITION 3.1 (spherical section property [37, 44]). *Let $m$, $n$ be two positive integers such that $m < n$. Let $V$ be an $(n-m)$-dimensional subspace of $\mathbb{R}^n$ and $s$ be a positive integer. We say that $V$ has the $s$-spherical section property if*

$$\inf_{v \in V \setminus \{0\}} \frac{\|v\|_1}{\|v\|} \geq \sqrt{\frac{m}{s}}.$$

*Remark* 3.2. According to [44, Theorem 3.1], if $A \in \mathbb{R}^{m \times n}$ $(m < n)$ is a random matrix with independent and identically distributed (i.i.d.) standard Gaussian entries, then its $(n-m)$-dimensional null space has the $s$-spherical section property for $s = c_1(\log(n/m) + 1)$ with probability at least $1 - e^{-c_0(n-m)}$, where $c_0$ and $c_1$ are positive constants independent of $m$ and $n$.

We now present our analysis. We first characterize the existence of unbounded minimizing sequences of (1.1): recall that $\{x^t\}$ is called a minimizing sequence of (1.1) if $Ax^t = b$ for all $t$ and $\lim_{t \to \infty} \frac{\|x^t\|_1}{\|x^t\|} = \nu_{cs}^*$. Our characterization is related to the following auxiliary problem, where $A$ is as in (1.1):

(3.1)                     $\nu_d^* := \inf \left\{ \frac{\|d\|_1}{\|d\|} : Ad = 0, d \neq 0 \right\}.$

LEMMA 3.3. *Consider (1.1) and (3.1). Then $\nu_{cs}^* = \nu_d^*$ if and only if there exists a minimizing sequence of (1.1) that is unbounded.*

*Proof.* We first suppose that there exists an unbounded minimizing sequence $\{x^t\}$ of (1.1). By passing to a subsequence if necessary, we may assume without loss of

Downloaded 05/10/22 to 158.132.161.181 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

generality that $\|x^t\| \to \infty$ and that $\lim_{t\to\infty} \frac{x^t}{\|x^t\|} = x^*$ for some $x^*$ with $\|x^*\| = 1$. Then we have $\|x^*\|_1 = \nu_{cs}^*$ using the definition of minimizing sequence, and

$$(3.2) \qquad Ax^* = \lim_{t\to\infty} \frac{Ax^t}{\|x^t\|} = \lim_{t\to\infty} \frac{b}{\|x^t\|} = 0.$$

One can then see that

$$(3.3) \qquad \nu_d^* \le \frac{\|x^*\|_1}{\|x^*\|} = \|x^*\|_1 = \nu_{cs}^* < \infty.$$

Next, fix any $x$ such that $Ax = b$ and choose any $d \ne 0$ satisfying $Ad = 0$ (these exist thanks to $\nu_d^* \le \nu_{cs}^* < \infty$). Then it holds that

$$\nu_{cs}^* \le \frac{\|x + sd\|_1}{\|x + sd\|}$$

for any $s \in \mathbb{R}$. It follows from the above display that

$$\nu_{cs}^* \le \lim_{s\to\infty} \frac{\|x + sd\|_1}{\|x + sd\|} = \frac{\|d\|_1}{\|d\|}.$$

Then we have $\nu_{cs}^* \le \nu_d^*$ by the arbitrariness of $d$. This together with (3.3) shows that $\nu_{cs}^* = \nu_d^*$.

We next suppose that $\nu_{cs}^* = \nu_d^*$. Since $\nu_{cs}^* < \infty$ (thanks to $A^{-1}\{b\} \ne \emptyset$), there exists a sequence $\{d^k\}$ satisfying $Ad^k = 0$ and $d^k \ne 0$ such that $\lim_{k\to\infty} \frac{\|d^k\|_1}{\|d^k\|} = \nu_d^*$. Passing to a further subsequence if necessary, we may assume without loss of generality that $\lim_{k\to\infty} \frac{d^k}{\|d^k\|} = d^*$ for some $d^*$ with $\|d^*\| = 1$. It then follows that

$$Ad^* = \lim_{k\to\infty} \frac{Ad^k}{\|d^k\|} = 0 \ \text{ and } \ \|d^*\|_1 = \lim_{k\to\infty} \left\|\frac{d^k}{\|d^k\|}\right\|_1 = \nu_d^*.$$

Now, choose any $x^0$ such that $Ax^0 = b$ and define $x^t = x^0 + td^*$ for each $t = 1, 2, \ldots$. Then we have $Ax^t = b$ for all $t$. Moreover $\|x^t\| \to \infty$ as $t \to \infty$ and

$$\lim_{t\to\infty} \frac{\|x^t\|_1}{\|x^t\|} = \frac{\|d^*\|_1}{\|d^*\|} = \nu_d^* = \nu_{cs}^*.$$

Thus, $\{x^t\}$ is an unbounded minimizing sequence for (1.1). This completes the proof. $\square$

We are now ready to present the theorem on solution existence for (1.1).

THEOREM 3.4 (solution existence for (1.1)). *Consider* (1.1). *Suppose that* $\ker A$ *has the s-spherical section property for some* $s > 0$ *and there exists* $\widetilde{x} \in \mathbb{R}^n$ *such that* $\|\widetilde{x}\|_0 < m/s$ *and* $A\widetilde{x} = b$. *Then the optimal value* $\nu_{cs}^*$ *of* (1.1) *is attainable, i.e., the set of optimal solutions of* (1.1) *is nonempty.*

*Proof.* According to the *s*-spherical property of $\ker A$ and the definition of $\nu_d^*$ in (3.1), we see that $\nu_d^* \ge \sqrt{\frac{m}{s}}$. It then follows that

$$\nu_{cs}^* \overset{(a)}{\le} \frac{\|\widetilde{x}\|_1}{\|\widetilde{x}\|} \overset{(b)}{\le} \sqrt{\|\widetilde{x}\|_0} \overset{(c)}{<} \sqrt{\frac{m}{s}} \le \nu_d^*,$$

where (a) follows from the definition of $\nu^*_{cs}$ and the fact that $A\widetilde{x} = b$, (b) follows from the Cauchy–Schwarz inequality, and (c) holds by our assumption. Invoking Lemma 3.3 and noting $\nu^*_{cs} < \infty$, we see that there is a bounded minimizing sequence $\{x^t\}$ for (1.1). We can then pass to a convergent subsequence $\{x^{t_j}\}$ so that $\lim_{j\to\infty} x^{t_j} = x^*$ for some $x^*$ satisfying $Ax^* = b$. Since $b \neq 0$, this means in particular that $x^* \neq 0$. We then have upon using the continuity of $\frac{\|\cdot\|_1}{\|\cdot\|}$ at $x^*$ and the definition of minimizing sequence that

$$\frac{\|x^*\|_1}{\|x^*\|} = \lim_{j\to\infty} \frac{\|x^{t_j}\|_1}{\|x^{t_j}\|} = \nu^*_{cs}.$$

This shows that $x^*$ is an optimal solution of (1.1). This completes the proof. $\qquad\square$

**4. KL exponent of $F$ in (1.7) and global convergence of Algorithm 4.1.** In this section, we discuss the KL exponent of (1.7) and its implication on the convergence rate of the algorithm proposed in [39, equation 11] for solving (1.1). For ease of reference, this algorithm is presented as Algorithm 4.1. It was shown in [39] that if the sequence $\{x^t\}$ generated by this algorithm is bounded, then any accumulation point is a stationary point of $F$ in (1.7).

---

**Algorithm 4.1** The algorithm proposed in [39, equation 11] for (1.1).

---

**Step 0.** Choose $x^0$ with $Ax^0 = b$ and $\alpha > 0$. Set $\omega_0 = \|x^0\|_1/\|x^0\|$ and $t = 0$.
**Step 1.** Solve the subproblem

$$(4.1) \qquad \begin{aligned} x^{t+1} = \underset{x\in\mathbb{R}^n}{\arg\min} \quad & \|x\|_1 - \frac{\omega_t}{\|x^t\|}\langle x, x^t\rangle + \frac{\alpha}{2}\|x - x^t\|^2 \\ \text{s.t.} \quad & Ax = b. \end{aligned}$$

**Step 2.** Compute $\omega_{t+1} = \|x^{t+1}\|_1/\|x^{t+1}\|$. Update $t \leftarrow t + 1$ and go to Step 1.

---

Here, we first remark that if the sequence $\{x^t\}$ generated by Algorithm 4.1 is bounded, then it converges to a stationary point $x^*$ of $F$ in (1.7). The argument is standard (see [2, 3, 9]), making use of H1, H2, H3 in [3, section 2.3]. We include the proof for the ease of readers.

PROPOSITION 4.1 (global convergence of Algorithm 4.1). *Consider* (1.1). *Let* $\{x^t\}$ *be the sequence generated by Algorithm* 4.1 *and suppose that* $\{x^t\}$ *is bounded. Then* $\{x^t\}$ *converges to a stationary point of $F$ in* (1.7).

*Proof.* First, according to [39, Lemma 1], the sequence $\{\omega_t\}$ generated by Algorithm 4.1 enjoys the following sufficient descent property:

$$(4.2) \qquad \omega_t - \omega_{t+1} \geq \frac{\alpha}{2\|x^{t+1}\|}\|x^{t+1} - x^t\|^2.$$

Now, if we let $\lambda^t$ denote a Lagrange multiplier of the subproblem (4.1) at iteration $t$, one can then see from the first-order optimality condition that

$$(4.3) \qquad -A^T\lambda^t + \frac{\|x^t\|_1}{\|x^t\|^2}x^t - \alpha(x^{t+1} - x^t) \in \partial\|x^{t+1}\|_1.$$

On the other hand, using (2.3) and noting that $x^t \neq 0$ for all $t$, we have

$$\frac{1}{\|x^{t+1}\|}\partial\|x^{t+1}\|_1 - \frac{\|x^{t+1}\|_1}{\|x^{t+1}\|^3}x^{t+1} + \frac{A^T\lambda^t}{\|x^{t+1}\|}$$
$$= \partial\frac{\|x^{t+1}\|_1}{\|x^{t+1}\|} + \frac{A^T\lambda^t}{\|x^{t+1}\|} \subset \partial\frac{\|x^{t+1}\|_1}{\|x^{t+1}\|} + N_{A^{-1}\{b\}}(x^{t+1}) = \partial F(x^{t+1}),$$

where the last equality follows from [34, Corollary 10.9], the regularity at $x^{t+1}$ of $\frac{\|\cdot\|_1}{\|\cdot\|}$ (see (2.3)) and $\delta_{A^{-1}\{b\}}(\cdot)$ (see [34, Theorem 6.9]), and the definition of $F$ in (1.7). Combining (4.3) and the above display, we obtain that

$$\frac{1}{\|x^{t+1}\|}\left(\frac{\|x^t\|_1}{\|x^t\|^2}x^t - \frac{\|x^{t+1}\|_1}{\|x^{t+1}\|^2}x^{t+1}\right) - \frac{\alpha}{\|x^{t+1}\|}(x^{t+1} - x^t) \in \partial F(x^{t+1}).$$

On the other hand, since $\|x^t\| \geq \inf_{y \in A^{-1}\{b\}}\|y\| > 0$ for all $t$ (thanks to $Ax^t = b$ and $b \neq 0$) and $\{x^t\}$ is bounded, we see that there exists $C_0 > 0$ so that

$$\left\|\frac{\|x^t\|_1}{\|x^t\|^2}x^t - \frac{\|x^{t+1}\|_1}{\|x^{t+1}\|^2}x^{t+1}\right\| \leq C_0\|x^{t+1} - x^t\| \text{ for all } t.$$

Thus, in view of the above two displays, we conclude that

$$\text{dist}(0, \partial F(x^{t+1})) \leq \frac{C_0 + \alpha}{\inf_{y \in A^{-1}\{b\}}\|y\|}\|x^{t+1} - x^t\| \text{ for all } t.$$

Using the boundedness of $\{x^t\}$, (4.2), the above display and the continuity of $F$ on its domain, we see that the conditions H1, H2, H3 in [3, section 2.3] are satisfied. Since $F$ is clearly proper closed semialgebraic and hence a KL function, we can then invoke [3, Theorem 2.9] to conclude that $\{x^t\}$ converges to a stationary point of $F$. □

While it is routine to show that the sequence $\{x^t\}$ generated by Algorithm 4.1 is convergent when it is bounded, it is more challenging to deduce the asymptotic convergence rate: the latter typically requires an estimate of the KL exponent of $F$ in (1.7), which was used in the above analysis. In what follows, we will show that the KL exponent of $F$ is $\frac{1}{2}$. To do this, we will first establish a calculus rule for deducing the KL exponent of a fractional objective from the difference between the numerator and (a suitable scaling of) the denominator: this is along the line of the calculus rules for KL exponents developed in [25, 26, 42] and can be of independent interest.

**4.1. KL exponent of fractional functions.** Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ be proper closed and $g : \mathbb{R}^n \to \mathbb{R}$ be a continuous nonnegative function that is continuously differentiable on an open set containing $\text{dom } f$. Suppose that $\inf f \geq 0$ and $\inf_{\text{dom } f} g > 0$. We consider the following fractional programming problem:

$$(4.4) \qquad \min_x G(x) := \frac{f(x)}{g(x)}.$$

In algorithmic developments for solving (4.4) (see, for example, [22, 23]), it is customary to consider functions of the following form:

$$(4.5) \qquad H_u(x) := f(x) - \frac{f(u)}{g(u)}g(x),$$

where $u$ typically carries information from the previous iterate. In the literature, KL-type assumptions are usually imposed on $G$ or $H_u$ for establishing the global convergence of the sequence generated by first-order methods for solving (4.4); see, for example, the discussions in [12, Theorem 16] and [13, Theorem 5.5]. Here, we study a relationship between the KL exponent of $G$ in (4.4) and that of $H_{\bar{x}}$ in (4.5) when $\bar{x}$ is a stationary point of $G$.

THEOREM 4.2 (KL exponent of fractional functions). *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ be a proper closed function with $\inf f \geq 0$ and $g : \mathbb{R}^n \to \mathbb{R}$ be a continuous nonnegative function that is continuously differentiable on an open set containing $\mathrm{dom}\, f$ with $\inf_{\mathrm{dom}\, f} g > 0$. Assume that one of the following conditions holds:*

(i) *$f$ is locally Lipschitz.*
(ii) *$f = h + \delta_D$ for some continuously differentiable function $h$ and nonempty closed set $D$.*
(iii) *$f = h + \delta_D$ for some locally Lipschitz function $h$ and nonempty closed set $D$, and $h$ and $D$ are regular at every point in $D$.*

*Let $\bar{x}$ be such that $0 \in \partial G(\bar{x})$, where $G$ is defined as in (4.4). Then $\bar{x} \in \mathrm{dom}\, \partial H_{\bar{x}}$. If $H_{\bar{x}}$ defined as in (4.5) satisfies the KL property with exponent $\theta \in [0, 1)$ at $\bar{x}$, then so does $G$.*

*Proof.* It is clear that $\mathrm{dom}\, H_{\bar{x}} = \mathrm{dom}\, f = \mathrm{dom}\, G$. We first argue that under the assumptions on $f$ and $g$, we have for any $x \in \mathrm{dom}\, G$ that

$$(4.6) \quad \partial H_{\bar{x}}(x) = \partial f(x) - G(\bar{x})\nabla g(x) \quad \text{and} \quad \partial G(x) = \frac{1}{g(x)}\left(\partial f(x) - G(x)\nabla g(x)\right).$$

Indeed, in all cases, the first relation in (4.6) follows from [34, Exercise 8.8(c)]. When $f$ is locally Lipschitz, the second relation in (4.6) follows from [28, Corollary 1.111(i)]. When $f = h + \delta_D$ for some continuously differentiable function $h$ and nonempty closed set $D$, the second relation in (4.6) follows by first applying [34, Exercise 8.8(c)] to $G = \frac{h}{g} + \delta_D$, then applying the usual quotient rule to the differentiable function $\frac{h}{g}$, and subsequently using $\partial f = \nabla h + \partial \delta_D$ (thanks to [34, Exercise 8.8(c)]). Finally, when $f = h + \delta_D$ for some locally Lipschitz function $h$ and nonempty closed set $D$ with $h$ and $D$ being regular at every point in $D$, we have that the function $\frac{h}{g}$ is regular for all $x \in D$ in view of [28, Corollary 1.111(i)]. This together with the regularity of $D$ gives

$$\begin{aligned}
\partial G(x) &= \partial \left(\frac{h}{g}\right)(x) + \partial \delta_D(x) \\
&= \frac{g(x)\partial h(x) - h(x)\nabla g(x)}{g(x)^2} + \partial \delta_D(x) \\
&= \frac{g(x)\partial f(x) - f(x)\nabla g(x)}{g(x)^2},
\end{aligned}$$

where the first and last equalities follow from [34, Corollary 10.9] and [34, Exercise 8.14], and the second equality follows from [28, Corollary 1.111(i)].

Now, in view of (4.6), we have $\mathrm{dom}\, \partial H_{\bar{x}} = \mathrm{dom}\, \partial f = \mathrm{dom}\, \partial G$. In addition, in all three cases, it holds that $\mathrm{dom}\, f = \mathrm{dom}\, \partial f$. Indeed, when $f$ is locally Lipschitz, this claim follows from Exercise 8(c) of [10, section 6.4]. When $f = h + \delta_D$ as in (ii), the claim follows from [34, Exercise 8.8(c)], while for case (iii), we have $\mathrm{dom}\, f = \mathrm{dom}\, \partial f = D$ in view of [34, Corollary 10.9], [34, Exercise 8.14], and Exercise 8(c) of [10, section 6.4]. Consequently, in all three cases, we have

$$\Xi := \mathrm{dom}\, G = \mathrm{dom}\, \partial G = \mathrm{dom}\, H_{\bar{x}} = \mathrm{dom}\, \partial H_{\bar{x}} = \mathrm{dom}\, f = \mathrm{dom}\, \partial f,$$

and $H_{\bar{x}}$ is continuous relative to $\Xi$. In particular, $\bar{x} \in \mathrm{dom}\, \partial G = \mathrm{dom}\, \partial H_{\bar{x}}$.

Let $U$ be the open set containing $\mathrm{dom}\, f$ on which $g$ is continuously differentiable. Since $H_{\bar{x}}$ satisfies the KL property with exponent $\theta$ at $\bar{x}$ and is continuous relative to $\Xi$, there exist $\epsilon > 0$ and $c > 0$ so that $B(\bar{x}, 2\epsilon) \subseteq U$ and

$$(4.7) \qquad \text{dist}(0, \partial H_{\bar{x}}(x)) \geq c(H_{\bar{x}}(x) - H_{\bar{x}}(\bar{x}))^{\theta} = c(H_{\bar{x}}(x))^{\theta}$$

whenever $x \in \Xi$, $H_{\bar{x}}(x) > 0$ and $\|x - \bar{x}\| \leq \epsilon$. Let $M := \sup_{\|x - \bar{x}\| \leq \epsilon} \max\{g(x), \|\nabla g(x)\|\}$, which is finite as $g$ is continuously differentiable on $U \supseteq B(\bar{x}, 2\epsilon)$. Using the facts that $\theta \in [0, 1)$, $H_{\bar{x}}$ is continuous relative to $\Xi$, $H_{\bar{x}}(\bar{x}) = 0$, and $\inf_{\text{dom} f} g > 0$, we deduce that there exists $\epsilon' \in (0, \epsilon)$ such that

$$(4.8) \qquad |H_{\bar{x}}(x)|^{1-\theta} \leq \frac{c \inf_{\text{dom} f} g}{2M} \quad \text{whenever } \|x - \bar{x}\| \leq \epsilon' \text{ and } x \in \Xi,$$

where $c$ is given in (4.7).

Now, consider any $x \in \Xi$ satisfying $\|x - \bar{x}\| \leq \epsilon'$ and $G(\bar{x}) < G(x) < G(\bar{x}) + \epsilon'$. Then we have from (4.6) that

$$\text{dist}(0, \partial G(x)) = \frac{1}{g(x)} \inf_{\xi \in \partial f(x)} \|\xi - G(x)\nabla g(x)\| \overset{(a)}{\geq} \frac{1}{M} \inf_{\xi \in \partial f(x)} \|\xi - G(x)\nabla g(x)\|$$

$$\overset{(b)}{\geq} \frac{1}{M} \inf_{\xi \in \partial f(x)} \|\xi - G(\bar{x})\nabla g(x)\| - \frac{1}{M}|G(x) - G(\bar{x})|\|\nabla g(x)\|$$

$$\overset{(c)}{\geq} \frac{1}{M} \inf_{\xi \in \partial f(x)} \|\xi - G(\bar{x})\nabla g(x)\| - (G(x) - G(\bar{x}))$$

$$= \frac{1}{M}\text{dist}(0, \partial H_{\bar{x}}(x)) - \frac{1}{g(x)}H_{\bar{x}}(x) \overset{(d)}{\geq} \frac{1}{M}\text{dist}(0, \partial H_{\bar{x}}(x)) - \frac{1}{\inf_{\text{dom} f} g}H_{\bar{x}}(x)$$

$$\overset{(e)}{\geq} \frac{c}{M}(H_{\bar{x}}(x))^{\theta} - \frac{1}{\inf_{\text{dom} f} g}H_{\bar{x}}(x) \overset{(f)}{\geq} \frac{c}{2M}(H_{\bar{x}}(x))^{\theta}$$

$$= \frac{c(g(x))^{\theta}}{2M}(G(x) - G(\bar{x}))^{\theta} \overset{(g)}{\geq} \frac{c(\inf_{\text{dom} f} g)^{\theta}}{2M}(G(x) - G(\bar{x}))^{\theta},$$

where (a) holds because $g(x) \leq M$, (b) follows from the triangle inequality, (c) holds because $\|\nabla g(x)\| \leq M$ and $G(x) > G(\bar{x})$, (d) holds because $H_{\bar{x}}(x) > 0$ (thanks to $G(x) > G(\bar{x})$), (e) then follows from (4.7), and (f) follows from (4.8) and the fact that $H_{\bar{x}}(x) > 0$. Finally, (g) holds because $G(x) > G(\bar{x})$. This completes the proof. $\square$

**4.2. KL exponent of $F$ in (1.7).** Before proving our main result concerning the KL exponent of $F$ in (1.7), we also need the following simple proposition.

PROPOSITION 4.3. *Let $p$ be a proper closed function, and let $\bar{x} \in \text{dom} p$ be such that $p(\bar{x}) > 0$. Then the following statements hold:*
(i) *We have $\partial(p^2)(x) = 2p(x)\partial p(x)$ for all $x$ sufficiently close to $\bar{x}$.*
(ii) *Suppose in addition that $\bar{x} \in \text{dom} \partial(p^2)$ and $p^2$ satisfies the KL property at $\bar{x}$ with exponent $\theta \in [0, 1)$. Then $p$ satisfies the KL property at $\bar{x}$ with exponent $\theta \in [0, 1)$.*

*Proof.* Since $p(\bar{x}) > 0$ and $p$ is closed, there exists $\epsilon > 0$ so that

$$0 < p(x) < \infty$$

whenever $\|x - \bar{x}\| \leq \epsilon$ and $x \in \text{dom} p$. Then we deduce from [30, Lemma 1] that

$$(4.9) \qquad \widehat{\partial}(p^2)(x) = 2p(x)\widehat{\partial}p(x) \quad \text{whenever } x \in \text{dom} p \text{ and } \|x - \bar{x}\| \leq \epsilon.$$

Using (4.9), and invoking the definition of limiting subdifferential and by shrinking $\epsilon$ if necessary, we deduce that

(4.10)            $\partial(p^2)(x) = 2p(x)\partial p(x)$   whenever $x \in \operatorname{dom} p$ and $\|x - \bar{x}\| \leq \epsilon$.

In particular, if $\bar{x} \in \operatorname{dom} \partial(p^2)$, then $\bar{x} \in \operatorname{dom} \partial p$.

When $p^2$ also satisfies the KL property at $\bar{x}$ with exponent $\theta$, by shrinking $\epsilon$ further if necessary, we see that there exists $c > 0$ so that

(4.11)            $\operatorname{dist}(0, \partial(p^2)(x)) \geq c(p^2(x) - p^2(\bar{x}))^\theta,$

whenever $p^2(\bar{x}) < p^2(x) < p^2(\bar{x}) + \epsilon(2p(\bar{x}) + \epsilon)$ and $\|x - \bar{x}\| \leq \epsilon$. Thus, for $x \in \operatorname{dom} \partial p$ satisfying $\|x - \bar{x}\| \leq \epsilon$ and $p(\bar{x}) < p(x) < p(\bar{x}) + \epsilon$, we have from (4.10) that

$$\operatorname{dist}(0, \partial p(x)) = \frac{1}{2p(x)}\operatorname{dist}(0, \partial(p^2)(x)) \geq \frac{1}{2p(\bar{x}) + 2\epsilon}\operatorname{dist}(0, \partial(p^2)(x))$$

$$\overset{(a)}{\geq} \frac{c}{2p(\bar{x}) + 2\epsilon}(p^2(x) - p^2(\bar{x}))^\theta = \frac{c}{2p(\bar{x}) + 2\epsilon}(p(x) + p(\bar{x}))^\theta(p(x) - p(\bar{x}))^\theta$$

$$\geq \frac{c[p(\bar{x})]^\theta}{2^{1-\theta}(p(\bar{x}) + \epsilon)}(p(x) - p(\bar{x}))^\theta,$$

where (a) follows from (4.11). This completes the proof.                    □

We are now ready to show that the KL exponent of $F$ in (1.7) is $\frac{1}{2}$. We remark that if the set $\mathcal{X} := \{x : 0 \in \partial F(x)\}$ is empty, then this claim holds trivially in view of [25, Lemma 2.1]. However, in general, one can have $\mathcal{X} \neq \emptyset$. Indeed, according to Theorem 3.4 and [34, Theorem 10.1], we have $\mathcal{X} \neq \emptyset$ with high probability when $A$ is generated in a certain way.

THEOREM 4.4. *The function $F$ in (1.7) is a KL function with exponent $\frac{1}{2}$.*

*Proof.* In view of [25, Lemma 2.1], it suffices to look at the KL exponent at a stationary point $\bar{x}$ of $F$. For any $\bar{x}$ satisfying $0 \in \partial F(\bar{x})$, we have $F(\bar{x}) > 0$ since $b \neq 0$. Moreover, we have $0 \in \partial(F^2)(\bar{x})$ in view of Proposition 4.3(i). Next, note that the function

$$F_1(x) := \|x\|_1^2 - \frac{\|\bar{x}\|_1^2}{\|\bar{x}\|^2}\|x\|^2 + \delta_{A^{-1}\{b\}}(x)$$

can be written as $\min_{\sigma \in \mathfrak{R}}\{Q_\sigma(x) + P_\sigma(x)\}$, where $\mathfrak{R} = \{u \in \mathbb{R}^n : u_i \in \{1, -1\} \text{ for all } i\}$, $Q_\sigma$ are quadratic functions (nonconvex), and $P_\sigma$ are polyhedral functions indexed by $\sigma$: indeed, for each $\sigma \in \mathfrak{R}$, one can define $P_\sigma$ as the indicator function of the set $\{x : Ax = b, \sigma \circ x \geq 0\}$, where $\circ$ denotes the Hadamard product, and $Q_\sigma(x) := (\langle \sigma, x \rangle)^2 - \frac{\|\bar{x}\|_1^2}{\|\bar{x}\|^2}\|x\|^2$. Then, in view of [25, Corollary 5.2], $F_1$ is a KL function with exponent $\frac{1}{2}$. Since the convex function $\|\cdot\|_1^2$ is regular everywhere and the convex set $A^{-1}\{b\}$ is regular at every $x \in A^{-1}\{b\}$ (thanks to [34, Theorem 6.9]), we deduce using Theorem 4.2 that the function

$$x \mapsto \frac{\|x\|_1^2}{\|x\|^2} + \delta_{A^{-1}\{b\}}(x)$$

satisfies the KL property at $\bar{x}$ with exponent $\frac{1}{2}$. The desired conclusion now follows from this and Proposition 4.3(ii).                    □

Equipped with the result above, by following the line of arguments in [1, Theorem 2], one can conclude further that the sequence $\{x^t\}$ generated by Algorithm 4.1 converges locally linearly to a stationary point of $F$ in (1.7) if the sequence is bounded. The proof is standard and we omit it here for brevity.

THEOREM 4.5 (convergence rate of Algorithm 4.1). *Consider* (1.1). *Let* $\{x^t\}$ *be the sequence generated by Algorithm* 4.1 *and suppose that* $\{x^t\}$ *is bounded. Then* $\{x^t\}$ *converges to a stationary point* $x^*$ *of* $F$ *in* (1.7) *and there exist* $\underline{t} \in \mathbb{N}_+$, $a_0 \in (0,1)$ *and* $a_1 > 0$ *such that*

$$\|x^t - x^*\| \le a_1 a_0^t \quad \text{whenever } t > \underline{t}.$$

**5. Compressed sensing with noise based on $\ell_1/\ell_2$ minimization.** In the previous sections, we had been focusing on the model (1.1), which corresponds to noiseless CS problems. In this section and the next, we will be looking at (1.2). We will discuss conditions for existence of solutions and derive some first-order optimality conditions for (1.2) in this section. An algorithm for solving (1.2) will be proposed in the next section and will be shown to generate sequences that cluster at "critical" points in the sense defined in this section, under suitable assumptions.

**5.1. Solution existence.** Clearly, if $q$ in (1.2) is in addition level-bounded, then the feasible set is compact and hence the set of optimal solutions is nonempty. However, in applications such as (1.3), (1.4), and (1.5), the corresponding $q$ is not level-bounded. Here, we discuss solution existence for (1.3) and (1.4). Our arguments are along the same line as those in section 3. We first present a lemma that establishes a relationship between the problems (1.3), (1.4), and (3.1).

LEMMA 5.1. *Consider* (3.1) *and* (1.2) *with q given as in* (1.3) *or* (1.4). *Then* $\nu_{ncs}^* = \nu_d^*$ *if and only if there exists a minimizing sequence of* (1.2) *that is unbounded.*

The proof of this lemma is almost identical to that of Lemma 3.3. Here we omit the details and only point out a slight difference concerning the derivation of (3.2). Take (1.3) as an example and let $\{x^t\}$ be an unbounded minimizing sequence of it with $\lim_{t\to\infty} \frac{x^t}{\|x^t\|} = x^*$ for some $x^*$ satisfying $\|x^*\| = 1$. Then one can prove $Ax^* = 0$ by using the facts that $\|Ax^t - b\| \le \sigma$ for all $t$ and $\|x^t\| \to \infty$. Similar deductions can be done for (1.4).

Using Lemma 5.1, we can deduce solution existence based on the spherical section property of $\ker A$ and the existence of a sparse feasible solution to (1.3) (or (1.4)). The corresponding arguments are the same as those in Theorem 3.4 and we omit the proof for brevity.

THEOREM 5.2 (solution existence for (1.3) and (1.4)). *Consider* (1.2) *with q given as in* (1.3) *or* (1.4). *Suppose that* $\ker A$ *has the s-spherical section property and there exists* $\widetilde{x} \in \mathbb{R}^n$ *such that* $\|\widetilde{x}\|_0 < m/s$ *and* $q(\widetilde{x}) \le 0$. *Then the optimal value of* (1.2) *is attainable.*

**5.2. Optimality conditions.** We discuss first-order necessary optimality conditions for local minimizers. Our analysis is based on the following standard constraint qualifications.

DEFINITION 5.3 (generalized Mangasarian–Fromovitz constraint qualifications). *Consider* (1.2). *We say that the generalized Mangasarian–Fromovitz constraint qualifications (GMFCQ) holds at an* $x^*$ *satisfying* $q(x^*) \le 0$ *if the following statement holds:*

- *If* $q(x^*) = 0$, *then* $0 \notin \partial^\circ q(x^*)$.

The GMFCQ reduces to the standard MFCQ when $q$ is smooth. One can then see from sections 5.1 and 5.2 of [43] that the GMFCQ holds at every $x$ feasible for (1.3) and (1.4) for all positive $\sigma$ and $\gamma$, because $A$ is surjective. We next study the GMFCQ for (1.5), in which $A$ is also surjective.

PROPOSITION 5.4. *The GMFCQ holds in the whole feasible set of* (1.5).

*Proof.* It is straightforward to see that the GMFCQ holds for those $x \in \{x : q(x) < 0\}$. Then it remains to consider those $x$ satisfying $q(x) = 0$. Let $q$ be as in (1.5) and let $\bar{x}$ satisfy $q(\bar{x}) = 0$. Notice that a $\xi \in \mathrm{Proj}_S(A\bar{x} - b)$ takes the following form:

$$\xi_j = \begin{cases} [A\bar{x} - b]_j & \text{if } j \in I^*, \\ 0 & \text{otherwise,} \end{cases}$$

where $I^*$ is an index set corresponding to the $r$-largest entries (in magnitude). Then for any $\xi \in \mathrm{Proj}_S(A\bar{x} - b)$, we have

$$\langle A\bar{x} - b, \xi \rangle = \|\xi\|^2,$$

(5.1)
$$\|A\bar{x} - b\|^2 = \|\xi\|^2 + \|A\bar{x} - b - \xi\|^2$$

$$= \|\xi\|^2 + \mathrm{dist}^2(A\bar{x} - b, S) \stackrel{\text{(a)}}{=} \|\xi\|^2 + \sigma^2,$$

where (a) holds because $0 = q(\bar{x}) = \mathrm{dist}^2(A\bar{x} - b, S) - \sigma^2$. Furthermore, since $A$ is surjective, we can deduce from [34, Example 8.53], [34, Exercise 10.7], and [34, Theorem 8.49] that

$$\partial^\circ q(\bar{x}) = \mathrm{conv}\{2A^T(A\bar{x} - b - \xi) : \xi \in \mathrm{Proj}_S(A\bar{x} - b)\}.$$

Now, suppose to the contrary that $0 \in \partial^\circ q(\bar{x})$. Using Carathéodory's theorem, we see that there exist $\lambda_i \geq 0$ and $\xi_i \in \mathrm{Proj}_S(A\bar{x} - b)$, $i = 1, \ldots, m + 1$, such that $\sum_{i=1}^{m+1} \lambda_i = 1$ and $\sum_{i=1}^{m+1} \lambda_i A^T(A\bar{x} - b - \xi_i) = 0$. Since $A$ is surjective, we then have

$$\sum_{i=1}^{m+1} \lambda_i(A\bar{x} - b - \xi_i) = 0.$$

Multiplying both sides of the above equality by $(A\bar{x} - b)^T$, we obtain further that

$$0 = \sum_{i=1}^{m+1} \lambda_i \langle A\bar{x} - b, A\bar{x} - b - \xi_i \rangle = \sum_{i=1}^{m+1} \lambda_i[\|A\bar{x} - b\|^2 - \langle A\bar{x} - b, \xi_i \rangle]$$

$$\stackrel{\text{(a)}}{=} \sum_{i=1}^{m+1} \lambda_i[\|\xi_i\|^2 + \sigma^2 - \|\xi_i\|^2] = \sigma^2 > 0,$$

where (a) follows from (5.1) and the fact that $\xi_i \in \mathrm{Proj}_S(A\bar{x} - b)$ for each $i$, and the last equality holds because $\sum_{i=1}^{m+1} \lambda_i = 1$. This is a contradiction and thus we must have $0 \notin \partial^\circ q(\bar{x})$. This completes the proof. $\square$

In the next definition, we consider some notions of criticality. The first one is the standard notion of stationarity while the second one involves the Clarke subdifferential.

DEFINITION 5.5. *Consider* (1.2). *We say that an* $\bar{x} \in \mathbb{R}^n$ *satisfying* $q(\bar{x}) \leq 0$ *is*
(i) *a stationary point of* (1.2) *if*

(5.2)
$$0 \in \partial \left( \frac{\|\cdot\|_1}{\|\cdot\|} + \delta_{[q \leq 0]}(\cdot) \right)(\bar{x});$$

(ii) *a Clarke critical point of* (1.2) *if there exists* $\bar{\lambda} \geq 0$ *such that*

(5.3)
$$0 \in \partial \frac{\|\bar{x}\|_1}{\|\bar{x}\|} + \bar{\lambda} \partial^\circ q(\bar{x}) \quad \text{and} \quad \bar{\lambda} q(\bar{x}) = 0.$$

As mentioned above, Definition 5.5(i) is standard and it is known that every local minimizer of (1.2) is a stationary point; see [34, Theorem 10.1]. We next study some relationships between these notions of criticality and show in particular that every local minimizer is Clarke critical when the GMFCQ holds.

PROPOSITION 5.6 (stationarity versus Clarke criticality). *Consider* (1.2) *and let* $\bar{x}$ *be such that* $q(\bar{x}) \leq 0$. *Then the following statements hold.*
 (i) *If* $\bar{x}$ *is a stationary point of* (1.2) *and the GMFCQ holds at* $\bar{x}$, *then* $\bar{x}$ *is a Clarke critical point.*
 (ii) *If* $\bar{x}$ *is a Clarke critical point of* (1.2) *and* $q$ *is regular at* $\bar{x}$, *then* $\bar{x}$ *is stationary.*

*Remark* 5.7. Since local minimizers of (1.2) are stationary points, we see from Proposition 5.6(i) that when the GMFCQ holds in the whole feasible set, local minimizers are also Clarke critical.

*Proof.* Suppose that $\bar{x}$ is a stationary point of (1.2) at which the GMFCQ holds. Then (5.2) holds and we consider two cases.

**Case 1**: $q(\bar{x}) < 0$. Since $q$ is continuous, (5.2) implies $0 \in \partial \frac{\|\bar{x}\|_1}{\|\bar{x}\|}$ and hence (5.3) holds with $\bar{\lambda} = 0$. Thus, $\bar{x}$ is a Clarke critical point.

**Case 2**: $q(\bar{x}) = 0$. Since the GMFCQ holds for (1.2) at $\bar{x}$, we see that $0 \notin \partial^\circ q(\bar{x})$. Then we can deduce from (5.2) and [34, Exercise 10.10] that

$$0 \in \partial \frac{\|\bar{x}\|_1}{\|\bar{x}\|} + N_{[q\leq 0]}(\bar{x}) \stackrel{(a)}{\subseteq} \partial \frac{\|\bar{x}\|_1}{\|\bar{x}\|} + \bigcup_{\lambda \geq 0} \lambda \partial^\circ q(\bar{x}),$$

where (a) follows from [11, Theorem 5.2.22], the first corollary to [21, Theorem 2.4.7], and the fact that $0 \notin \partial^\circ q(\bar{x})$. Thus, (5.3) holds with some $\bar{\lambda} \geq 0$ (recall that $q(\bar{x}) = 0$), showing that $\bar{x}$ is a Clarke critical point. This proves item (i).

We now prove item (ii). Suppose that $\bar{x}$ is a Clarke critical point and that $q$ is regular at $\bar{x}$. Then there exists $\bar{\lambda} \geq 0$ so that (5.3) holds. We again consider two cases.

**Case 1:** $\bar{\lambda} = 0$. In this case, we see from (5.3) that $0 \in \partial \frac{\|\bar{x}\|_1}{\|\bar{x}\|}$, which implies

$$0 \in \partial \frac{\|\bar{x}\|_1}{\|\bar{x}\|} \stackrel{(a)}{=} \widehat{\partial} \frac{\|\bar{x}\|_1}{\|\bar{x}\|} \subseteq \widehat{\partial} \frac{\|\bar{x}\|_1}{\|\bar{x}\|} + \widehat{N}_{[q\leq 0]}(\bar{x})$$

$$\stackrel{(b)}{\subseteq} \widehat{\partial} \left( \frac{\|\cdot\|_1}{\|\cdot\|} + \delta_{[q\leq 0]}(\cdot) \right)(\bar{x}) \stackrel{(c)}{\subseteq} \partial \left( \frac{\|\cdot\|_1}{\|\cdot\|} + \delta_{[q\leq 0]}(\cdot) \right)(\bar{x}),$$

where (a) follows from (2.3) and [34, Corollary 8.11], (b) holds thanks to [34, Corollary 10.9], and (c) follows from [34, Theorem 8.6]. Thus, $\bar{x}$ is a stationary point.

**Case 2:** $\bar{\lambda} > 0$. In this case, we have from (5.3) that $q(\bar{x}) = 0$. Since $q$ is regular at $\bar{x}$, we see from [34, Corollary 8.11] and the discussion right after [34, Theorem 8.49] that

$$(5.4) \qquad \widehat{\partial} q(\bar{x}) = \partial q(\bar{x}) = \partial^\circ q(\bar{x}).$$

Now, in view of (5.4), $q(\bar{x}) = 0$, and [34, Proposition 10.3], we have

$$(5.5) \qquad \widehat{N}_{[q\leq 0]}(\bar{x}) \supseteq \bigcup_{\lambda \geq 0} \lambda \widehat{\partial} q(\bar{x}) = \bigcup_{\lambda \geq 0} \lambda \partial^\circ q(\bar{x}).$$

We then deduce that

$$\partial \left( \frac{\|\cdot\|_1}{\|\cdot\|} + \delta_{[q \leq 0]}(\cdot) \right)(\bar{x}) \overset{(a)}{\supseteq} \widehat{\partial} \frac{\|\bar{x}\|_1}{\|\bar{x}\|} + \widehat{N}_{[q \leq 0]}(\bar{x}) \overset{(b)}{\supseteq} \partial \frac{\|\bar{x}\|_1}{\|\bar{x}\|} + \bigcup_{\lambda \geq 0} \lambda \partial^\circ q(\bar{x}),$$

where (a) follows from [34, Theorem 8.6] and [34, Corollary 10.9], and (b) follows from (5.5), (2.3), and [34, Corollary 8.11]. This together with the definition of Clarke criticality shows that (5.2) holds. This completes the proof. $\qquad \square$

**6. A moving-balls-approximation based algorithm for solving (1.2).** In this section, we propose and analyze an algorithm for solving (1.2), which is an extension of Algorithm 4.1 by incorporating MBA techniques [4]. Our algorithm, which we call MBA$_{\ell_1/\ell_2}$, is presented as Algorithm 6.1.

---

**Algorithm 6.1** MBA$_{\ell_1/\ell_2}$: Moving-balls-approximation based algorithm for (1.2).

---

**Step 0.** Choose $x^0$ with $q(x^0) \leq 0$, $\alpha > 0$, and $0 < l_{\min} < l_{\max}$. Set $\omega_0 = \|x^0\|_1/\|x^0\|$ and $t = 0$.
**Step 1.** Choose $l_t^0 \in [l_{\min}, l_{\max}]$ arbitrarily and set $l_t = l_t^0$. Choose $\zeta^t \in \partial P_2(x^t)$.
(1a) Solve the subproblem

$$(6.1) \qquad \begin{aligned} \widetilde{x} = \underset{x \in \mathbb{R}^n}{\arg\min} \quad & \|x\|_1 - \frac{\omega_t}{\|x^t\|}\langle x, x^t \rangle + \frac{\alpha}{2}\|x - x^t\|^2 \\ \text{s.t.} \quad & q(x^t) + \langle \nabla P_1(x^t) - \zeta^t, x - x^t \rangle + \frac{l_t}{2}\|x - x^t\|^2 \leq 0. \end{aligned}$$

(1b) If $q(\widetilde{x}) \leq 0$, go to Step 2. Else, update $l_t \leftarrow 2l_t$ and go to Step (1a).
**Step 2.** Set $x^{t+1} = \widetilde{x}$ and compute $\omega_{t+1} = \|x^{t+1}\|_1/\|x^{t+1}\|$. Set $\bar{l}_t := l_t$. Update $t \leftarrow t + 1$ and go to Step 1.

---

Unlike previous works [6, 8, 43] that made use of MBA techniques, our algorithm deals with a *fractional* objective and a possibly *nonsmooth* continuous constraint function. Thus, the convergence results in [6, 8, 43] cannot be directly applied to analyze our algorithm. Indeed, as we shall see later in section 6.2, we need to introduce a new potential function for our analysis to deal with the possibly nonsmooth $q$ in the constraint.

We will show that Algorithm 6.1 is well defined later, i.e., for each $t \in \mathbb{N}_+$, the subproblem (6.1) has a unique solution for every $l_t$ and the inner loop in Step 1 terminates finitely. Here, it is worth noting that (6.1) can be efficiently solved using a root-finding procedure outlined in [43, Appendix A] since (6.1) takes the form of

$$\min_x \|x\|_1 + \frac{\alpha}{2}\|x - c^t\|^2 \quad \text{s.t.} \quad \|x - s^t\|^2 \leq R_t$$

for some $c^t \in \mathbb{R}^n$, $s^t \in \mathbb{R}^n$, and $R_t \geq 0$.

**6.1. Convergence analysis.** In this subsection, we establish subsequential convergence of MBA$_{\ell_1/\ell_2}$ under suitable assumptions. We start with the following auxiliary lemma that concerns well-definedness and sufficient descent. The proof of the sufficient descent property in item (iii) below is essentially the same as [39, Lemma 1]. We include it here for completeness.

LEMMA 6.1 (well-definedness and sufficient descent). *Consider* (1.2). *Then the following statements hold:*

(i) $MBA_{\ell_1/\ell_2}$ *is well defined, i.e., for each* $t \in \mathbb{N}_+$, *the subproblem* (6.1) *has a unique solution for every* $l_t$ *and the inner loop in* Step 1 *terminates finitely.*

(ii) *The sequence* $\{\bar{l}_t\}$ *is bounded.*

(iii) *Let* $\{(x^t, \omega_t)\}$ *be the sequence generated by* $MBA_{\ell_1/\ell_2}$. *Then there exists* $\delta > 0$ *such that* $\|x^t\| \geq \delta$ *for every* $t \in \mathbb{N}_+$, *and the sequence* $\{\omega_t\}$ *satisfies*

$$(6.2) \qquad \omega_t - \omega_{t+1} \geq \frac{\alpha}{2\|x^{t+1}\|} \|x^t - x^{t+1}\|^2, \quad t \in \mathbb{N}_+.$$

*Proof.* Suppose that an $x^t$ satisfying $q(x^t) \leq 0$ is given for some $t \in \mathbb{N}_+$. Then $x^t \neq 0$ since $q(0) > 0$. Moreover, for any $l_t > 0$, $x^t$ is feasible for (6.1) and the feasible set is thus nonempty. Since (6.1) minimizes a strongly convex continuous function over a nonempty closed convex set, it has a unique optimal solution, i.e., $\widetilde{x}$ exists.

Let $L_p$ be the Lipschitz continuity modulus of $\nabla P_1$. Then we have

$$(6.3) \quad \begin{aligned} q(\widetilde{x}) &= P_1(\widetilde{x}) - P_2(\widetilde{x}) \leq P_1(x^t) + \langle \nabla P_1(x^t), \widetilde{x} - x^t \rangle + \frac{L_p}{2}\|\widetilde{x} - x^t\|^2 - P_2(\widetilde{x}) \\ &\overset{(a)}{\leq} P_1(x^t) - P_2(x^t) + \langle \nabla P_1(x^t) - \zeta^t, \widetilde{x} - x^t \rangle + \frac{L_p}{2}\|\widetilde{x} - x^t\|^2 \overset{(b)}{\leq} \frac{L_p - l_t}{2}\|\widetilde{x} - x^t\|^2, \end{aligned}$$

where (a) holds because of the convexity of $P_2$ and the definition of $\zeta^t$, and (b) follows from the feasibility of $\widetilde{x}$ for (6.1). Let $k_0 \in \mathbb{N}_+$ be such that $L_p - 2^{k_0}l_{\min} \leq 0$. Then by (6.3) and the definition of $l_t$ we see that $q(\widetilde{x}) \leq 0$ after at most $k_0$ calls of Step (1b). Moreover, it holds that $\bar{l}_t \leq 2^{k_0}l_{\max}$. Therefore, if $q(x^t) \leq 0$, then the inner loop of Step 1 stops after at most $k_0$ iterations and outputs an $x^{t+1}$ satisfying $q(x^{t+1}) \leq 0$ (in particular, $x^{t+1} \neq 0$) with $\bar{l}_t \leq 2^{k_0}l_{\max}$. Since we initialize our algorithm at an $x^0$ satisfying $q(x^0) \leq 0$, the conclusions in items (i) and (ii) now follow from an induction argument.

Next, we prove item (iii). Since $q(0) > 0$, we see immediately from the continuity of $q$ that there exists some $\delta > 0$ such that $\|x\| \geq \delta$ whenever $q(x) \leq 0$. Thus, $\|x^t\| \geq \delta$ for all $t \in \mathbb{N}_+$, thanks to $q(x^t) \leq 0$. Now consider (6.1) with $l_t = \bar{l}_t$. Then $x^t$ is feasible and $x^{t+1}$ is optimal. This together with the definition of $\omega_t$ yields

$$\|x^{t+1}\|_1 - \frac{\|x^t\|_1}{\|x^t\|^2}\langle x^{t+1}, x^t \rangle + \frac{\alpha}{2}\|x^{t+1} - x^t\|^2 \leq \|x^t\|_1 - \frac{\|x^t\|_1}{\|x^t\|^2}\langle x^t, x^t \rangle + \frac{\alpha}{2}\|x^t - x^t\|^2 = 0.$$

Dividing both sides of the above inequality by $\|x^{t+1}\|$ and rearranging terms, we have

$$\frac{\|x^{t+1}\|_1}{\|x^{t+1}\|} + \frac{\alpha}{2\|x^{t+1}\|}\|x^t - x^{t+1}\|^2 \leq \frac{\|x^t\|_1}{\|x^t\|^2}\frac{\langle x^{t+1}, x^t \rangle}{\|x^{t+1}\|} \leq \frac{\|x^t\|_1}{\|x^t\|^2}\frac{\|x^{t+1}\|\|x^t\|}{\|x^{t+1}\|} = \frac{\|x^t\|_1}{\|x^t\|}.$$

This proves (iii) and completes the proof. $\square$

We next introduce the following assumption.

ASSUMPTION 6.2. *The GMFCQ for* (1.2) *holds at every point in* $[q \leq 0]$.

Recall from Proposition 5.4 and the discussions preceding it that Assumption 6.2 holds for (1.3), (1.4), and (1.5) since $A$ is surjective. We next derive the Karush–Kuhn–Tucker (KKT) conditions for (6.1) at every iteration $t$ under Assumption 6.2, which will be used in our subsequent analysis.

LEMMA 6.3 (KKT conditions for (6.1)). *Consider* (1.2) *and suppose that Assumption* 6.2 *holds. Let* $\{x^t\}$ *be the sequence generated by* $MBA_{\ell_1/\ell_2}$. *Then the following statements hold:*

(i) *The Slater's condition holds for the constraint of* (6.1) *at each* $t \in \mathbb{N}_+$.
(ii) *For each* $t \in \mathbb{N}_+$, $\zeta^t \in \partial P_2(x^t)$ *and* $l_t > 0$, *the subproblem* (6.1) *has a Lagrange multiplier* $\lambda_t \geq 0$. *Moreover, if* $\widetilde{x}$ *is as in* (6.1), *then it holds that*

$$(6.4) \qquad \lambda_t \left( q(x^t) + \left\langle \nabla P_1(x^t) - \zeta^t, \widetilde{x} - x^t \right\rangle + \frac{l_t}{2} \|\widetilde{x} - x^t\|^2 \right) = 0,$$

$$(6.5) \qquad 0 \in \partial \|\widetilde{x}\|_1 - \frac{\omega_t x^t}{\|x^t\|} + \lambda_t (\nabla P_1(x^t) - \zeta^t) + (\alpha + \lambda_t l_t)(\widetilde{x} - x^t).$$

*Proof.* Notice that we can rewrite the feasible set of (6.1) as $B\left(s^t, \sqrt{R_t}\right)$ with $s^t := x^t - \frac{1}{l_t}(\nabla P_1(x^t) - \zeta^t)$ and $R_t := \frac{1}{l_t^2}\|\nabla P_1(x^t) - \zeta^t\|^2 - \frac{2}{l_t}q(x^t)$, where $R_t \geq 0$ because $q(x^t) \leq 0$. Suppose to the contrary that $R_t = 0$. Then we have $q(x^t) = 0$ and $\nabla P_1(x^t) - \zeta^t = 0$. The latter relation together with (2.4) implies $0 \in \partial^\circ q(x^t)$, contradicting the GMFCQ assumption at $x^t$. Thus, we must have $R_t > 0$ and hence the Slater's condition holds for (6.1) at the $t$th iteration.

Since the Slater's condition holds for (6.1), we can apply [33, Corollary 28.2.1] and [33, Theorem 28.3] to conclude that there exists a Lagrange multiplier $\lambda_t$ such that the relation (6.4) holds at the $t$th iteration and $\widetilde{x}$ minimizes the following function:

$$\mathfrak{L}_t(x) := \|x\|_1 - \frac{\omega_t}{\|x^t\|}\langle x, x^t \rangle + \frac{\alpha}{2}\|x - x^t\|^2$$

$$+ \lambda_t \left( q(x^t) + \left\langle \nabla P_1(x^t) - \zeta^t, x - x^t \right\rangle + \frac{l_t}{2}\|x - x^t\|^2 \right).$$

This fact together with [34, Exercise 8.8] and [34, Theorem 10.1] implies that (6.5) holds at the $t$th iteration. This completes the proof. $\square$

Now we are ready to establish the subsequential convergence of Algorithm 6.1. In our analysis, we assume that the GMFCQ holds and that the $\{x^t\}$ generated by $\mathrm{MBA}_{\ell_1/\ell_2}$ is bounded. The latter boundedness assumption was also used in [39] for analyzing the convergence of Algorithm 4.1. We remark that this assumption is not too restrictive. Indeed, for the sequence $\{x^t\}$ generated by $\mathrm{MBA}_{\ell_1/\ell_2}$, in view of Lemma 6.1(i), we know that $q(x^t) \leq 0$ for all $t$. Thus, if $q$ is level-bounded, then $\{x^t\}$ is bounded. On the other hand, if $q$ is only known to be bounded from below (as in (1.3), (1.4), and (1.5)) but the corresponding (1.2) is known to have an optimal solution, then one may replace $q(x)$ by the level-bounded function $q_M(x) := q(x) + (\|x\| - M)_+^2$ for a sufficiently large $M$. As long as $M > \|x^*\|$ for some optimal solution $x^*$ of (1.2), replacing $q$ by $q_M$ in (1.2) will not change the optimal value.

THEOREM 6.4 (subsequential convergence of $\mathrm{MBA}_{\ell_1/\ell_2}$). *Consider* (1.2) *and suppose that Assumption* 6.2 *holds. Let* $\{(x^t, \zeta^t, \bar{l}_t)\}$ *be the sequence generated by* $\mathrm{MBA}_{\ell_1/\ell_2}$ *and* $\lambda_t$ *be a Lagrange multiplier of* (6.1) *with* $l_t = \bar{l}_t$. *Suppose in addition that* $\{x^t\}$ *is bounded. Then the following statements hold:*
(i) $\lim_{t\to\infty} \|x^{t+1} - x^t\| = 0$.
(ii) *The sequences* $\{\lambda_t\}$ *and* $\{\zeta^t\}$ *are bounded.*
(iii) *Let* $\bar{x}$ *be an accumulation point of* $\{x^t\}$. *Then* $\bar{x}$ *is a Clarke critical point of* (1.2). *If* $q$ *is also regular at* $\bar{x}$, *then* $\bar{x}$ *is a stationary point.*

*Proof.* Since $\{x^t\}$ is bounded, there exists $M > 0$ such that $\|x^t\| \leq M$ for all $t \in \mathbb{N}_+$. Using (6.2), we obtain

$$\sum_{t=0}^{\infty} \frac{\alpha}{2M}\|x^t - x^{t+1}\|^2 \leq \omega_0 - \liminf_{t\to\infty}\omega_t \leq \omega_0,$$

which proves item (i).

Now we turn to item (ii). The boundedness of $\{\zeta^t\}$ follows from the boundedness of $\{x^t\}$ and [36, Theorem 2.6]. We next prove the boundedness of $\{\lambda_t\}$. Suppose to the contrary that $\{\lambda_t\}$ is unbounded. Then there exists a subsequence $\{\lambda_{t_k}\}$ such that $\lim_{k\to\infty}\lambda_{t_k}=\infty$. Passing to a subsequence if necessary, we can find subsequences $\{x^{t_k}\}$ and $\{\lambda_{t_k}\}$ such that $\lim_{k\to\infty}x^{t_k}=x^*$ and $\lambda_{t_k}>0$ for all $k\in\mathbb{N}_+$, where the existence of $x^*$ is due to the boundedness of $\{x^t\}$. According to (6.4) and the definition of $x^{t_k+1}$, we obtain

$$q(x^{t_k})+\langle\nabla P_1(x^{t_k})-\zeta^{t_k},x^{t_k+1}-x^{t_k}\rangle+\frac{\bar{l}_{t_k}}{2}\left\|x^{t_k+1}-x^{t_k}\right\|^2=0.$$

Since $\{x^t\}$ is bounded and $\nabla P_1$ is Lipschitz continuous, we then see that $\{\nabla P_1(x^t)\}$ is bounded. Moreover, $\{\bar{l}_{t_k}\}$ is bounded thanks to Lemma 6.1(ii) and we also know that $\{\zeta^t\}$ is bounded. Using these facts, item (i), and the continuity of $q$, we have upon passing to the limit in the above display that $q(x^*)=0$. Since the GMFCQ holds for (1.2) at $x^*$, we then have $0\notin\partial^\circ q(x^*)$.

Let $t=t_k$, $l_t=\bar{l}_{t_k}$, $\tilde{x}=x^{t_k+1}$ in (6.5), and divide both sides of (6.5) by $\lambda_{t_k}$. Then

$$\nabla P_1(x^{t_k})-\zeta^{t_k}\in-\frac{1}{\lambda_{t_k}}\partial\|x^{t_k+1}\|_1+\frac{\omega_{t_k}x^{t_k}}{\lambda_{t_k}\|x^{t_k}\|}-\left(\bar{l}_{t_k}+\frac{\alpha}{\lambda_{t_k}}\right)(x^{t_k+1}-x^{t_k}).$$

Thus, there exists a sequence $\{\eta^k\}$ satisfying $\eta^k\in\partial\|x^{t_k+1}\|_1$ and

$$\nabla P_1(x^{t_k})-\zeta^{t_k}=-\frac{1}{\lambda_{t_k}}\eta^k+\frac{\omega_{t_k}x^{t_k}}{\lambda_{t_k}\|x^{t_k}\|}-\left(\bar{l}_{t_k}+\frac{\alpha}{\lambda_{t_k}}\right)(x^{t_k+1}-x^{t_k}).$$

Note that $\{\eta^k\}$ is bounded since $\partial\|x\|_1\subseteq[-1,1]^n$ for any $x\in\mathbb{R}^n$. Moreover, $\{\omega_{t_k}\}$ is bounded since $\|x\|\leq\|x\|_1\leq\sqrt{n}\|x\|$ for any $x\in\mathbb{R}^n$. Furthermore, we have the boundedness of $\{\bar{l}_{t_k}\}$ from Lemma 6.1(ii). Also recall that $\lim_{k\to\infty}\lambda_{t_k}=\infty$ and $\zeta^t\in\partial P_2(x^t)$. Using these together with item (i), we have upon passing to the limit in the above display and invoking the closedness of $\partial P_2$ (see Exercise 8 of [10, section 4.2]) that $\nabla P_1(x^*)\in\partial P_2(x^*)$. This together with (2.4) further implies $0\in\partial^\circ q(x^*)$, leading to a contradiction. Thus, the sequence $\{\lambda_t\}$ is bounded.

We now turn to item (iii). Suppose $\bar{x}$ is an accumulation point of $\{x^t\}$ with $\lim_{j\to\infty}x^{t_j}=\bar{x}$ for some convergent subsequence $\{x^{t_j}\}$. Since $\{\lambda_t,\bar{l}_t\}$ and $\{\zeta^t\}$ are bounded (thanks to Lemma 6.1(ii) and item (ii)), passing to a further subsequence if necessary, we may assume without loss of generality that

$$(6.6)\quad\lim_{j\to\infty}(\lambda_{t_j},\bar{l}_{t_j})=(\bar{\lambda},\bar{l})\text{ for some }\bar{\lambda},\,\bar{l}\geq0,\quad\lim_{j\to\infty}\zeta^{t_j}=\bar{\zeta}\text{ for some }\bar{\zeta}\in\partial P_2(\bar{x});$$

here, $\bar{\zeta}\in\partial P_2(\bar{x})$ because of the closedness of $\partial P_2$ (see Exercise 8 of [10, section 4.2]). On the other hand, according to Lemma 6.1(iii), we have $\|x^t\|\geq\delta>0$ for all $t\in\mathbb{N}_+$. This together with the definition of $\bar{x}$ yields $\|\bar{x}\|\neq0$. It then follows that $\frac{\|\cdot\|_1}{\|\cdot\|}$ is continuous at $\bar{x}$. Thus, we have, upon using this fact, the definition of $\omega_t$, the continuity of $\nabla P_1$, the closedness of $\partial\|\cdot\|_1$, item (i), (6.6), and passing to the limit as $j\to\infty$ in (6.5) with $(\tilde{x},\lambda_t,l_t)=(x^{t_j+1},\lambda_{t_j},\bar{l}_{t_j})$ and $t=t_j$ that

$$0\in\partial\|\bar{x}\|_1-\frac{\|\bar{x}\|_1}{\|\bar{x}\|^2}\bar{x}+\bar{\lambda}(\nabla P_1(\bar{x})-\bar{\zeta}).$$

We then divide both sides of the above inclusion by $\|\bar{x}\|$ and obtain

$$(6.7) \quad 0 \in \frac{1}{\|\bar{x}\|}\partial\|\bar{x}\|_1 - \frac{\|\bar{x}\|_1}{\|\bar{x}\|^3}\bar{x} + \frac{\bar{\lambda}}{\|\bar{x}\|}(\nabla P_1(\bar{x}) - \bar{\zeta}) = \partial\frac{\|\bar{x}\|_1}{\|\bar{x}\|} + \frac{\bar{\lambda}}{\|\bar{x}\|}(\nabla P_1(\bar{x}) - \bar{\zeta}),$$

where the equality holds due to (2.3). In addition, using (6.4) with $(\widetilde{x}, \lambda_t, l_t) = (x^{t_j+1}, \lambda_{t_j}, \bar{l}_{t_j})$ and $t = t_j$, we have

$$\lim_{j\to\infty} \lambda_{t_j}\left[q(x^{t_j}) + \langle\nabla P_1(x^{t_j}) - \zeta^{t_j}, x^{t_j+1} - x^{t_j}\rangle + \frac{\bar{l}_{t_j}}{2}\|x^{t_j+1} - x^{t_j}\|^2\right] = 0.$$

This together with item (i) and (6.6) shows that $\bar{\lambda}q(\bar{x}) = 0$. Combining this with (6.7), $\bar{\zeta} \in \partial P_2(\bar{x})$ (see (6.6)), (2.4), and the fact that $q(\bar{x}) \leq 0$ (because $q(x^t) \leq 0$ for all $t$) shows that $\bar{x}$ is a Clarke critical point. Finally, the claim concerning stationarity follows immediately from Proposition 5.6. This completes the proof. $\square$

**6.2. Global convergence under KL assumption.** We now discuss global convergence of the sequence $\{x^t\}$ generated by Algorithm 6.1. Our analysis follows the line of analysis in [1, 2, 3, 6, 8, 43] and is based on the following auxiliary function:

$$(6.8) \qquad \widetilde{F}(x, y, \zeta, w) := \frac{\|x\|_1}{\|x\|} + \delta_{[\widetilde{q}\leq 0]}(x, y, \zeta, w) + \delta_{\|\cdot\|\geq\rho}(x)$$

with

$$(6.9) \qquad \widetilde{q}(x, y, \zeta, w) := P_1(y) + \langle\nabla P_1(y), x - y\rangle + P_2^*(\zeta) - \langle\zeta, x\rangle + \frac{w}{2}\|x - y\|^2,$$

where $P_1$ and $P_2$ are as in (1.2), and $\rho > 0$ is chosen such that $\{x : q(x) \leq 0\} \subset \{x : \|x\| > \rho\}$. Some comments on $\widetilde{F}$ are in place. First, recall that in the potential function used in [8] for analyzing their MBA variant, the authors replaced $P_1(x)$ by a quadratic majorant $P_1(y) + \langle\nabla P_1(y), x - y\rangle + \frac{L_p}{2}\|x - y\|^2$, where $L_p$ is the Lipschitz continuity modulus of $\nabla P_1$. Here, as in [43], we further introduce the variable $w$ to handle the varying $\bar{l}_t$. Finally, to deal with the possibly nonsmooth $-P_2$, we replaced $-P_2(x)$ by its majorant $P_2^*(\zeta) - \langle\zeta, x\rangle$ as in [26].

The next proposition concerns the subdifferential of $\widetilde{F}$ and will be used for deriving global convergence of the sequence generated by $\text{MBA}_{\ell_1/\ell_2}$.

PROPOSITION 6.5. *Consider* (1.2) *and assume that* $P_1$ *is twice continuously differentiable. Suppose that Assumption 6.2 holds. Let* $\{(x^t, \zeta^t, \bar{l}_t)\}$ *be the sequence generated by* $\text{MBA}_{\ell_1/\ell_2}$ *and suppose that* $\{x^t\}$ *is bounded. Let* $\widetilde{F}$ *and* $\widetilde{q}$ *be given in* (6.8) *and* (6.9)*, respectively. Then the following statements hold:*
  (i) *For any* $t \in \mathbb{N}_+$*, we have* $\widetilde{q}(x^{t+1}, x^t, \zeta^t, \bar{l}_t) \leq 0$.
 (ii) *There exist* $\kappa > 0$ *and* $\underline{t} \in \mathbb{N}_+$ *such that*

$$\text{dist}(0, \partial\widetilde{F}(x^{t+1}, x^t, \zeta^t, \bar{l}_t)) \leq \kappa\|x^{t+1} - x^t\| \quad \text{for all } t > \underline{t}.$$

*Proof.* We first observe that

$$
\widetilde{q}(x^{t+1}, x^t, \zeta^t, \bar{l}_t)
$$

$$
= P_1(x^t) + \langle \nabla P_1(x^t), x^{t+1} - x^t \rangle + P_2^*(\zeta^t) - \langle \zeta^t, x^{t+1} \rangle + \frac{\bar{l}_t}{2} \|x^{t+1} - x^t\|^2
$$

$$
(6.10) \qquad = P_1(x^t) + \langle \nabla P_1(x^t) - \zeta^t, x^{t+1} - x^t \rangle + P_2^*(\zeta^t) - \langle \zeta^t, x^t \rangle + \frac{\bar{l}_t}{2} \|x^{t+1} - x^t\|^2
$$

$$
\overset{\text{(a)}}{=} P_1(x^t) - P_2(x^t) + \langle \nabla P_1(x^t) - \zeta^t, x^{t+1} - x^t \rangle + \frac{\bar{l}_t}{2} \|x^{t+1} - x^t\|^2
$$

$$
= q(x^t) + \langle \nabla P_1(x^t) - \zeta^t, x^{t+1} - x^t \rangle + \frac{\bar{l}_t}{2} \|x^{t+1} - x^t\|^2 \le 0,
$$

where (a) follows from (2.1) because $\zeta^t \in \partial P_2(x^t)$, and the last inequality holds because $x^{t+1}$ is feasible for (6.1) with $l_t = \bar{l}_t$. This proves item (i).

Now, note that $N_{\mathbb{R}_-}(\widetilde{q}(x^{t+1}, x^t, \zeta^t, \bar{l}_t)) = \{0\}$ if $\widetilde{q}(x^{t+1}, x^t, \zeta^t, \bar{l}_t) < 0$. Using this together with [34, Proposition 10.3], we conclude that at any $(x^{t+1}, x^t, \zeta^t, \bar{l}_t)$ (regardless of whether $\widetilde{q}(x^{t+1}, x^t, \zeta^t, \bar{l}_t) < 0$ or $\widetilde{q}(x^{t+1}, x^t, \zeta^t, \bar{l}_t) = 0$), the relation

$$
\widehat{N}_{[\widetilde{q} \le 0]}(x^{t+1}, x^t, \zeta^t, \bar{l}_t) \supseteq \lambda \widehat{\partial} \widetilde{q}(x^{t+1}, x^t, \zeta^t, \bar{l}_t)
$$

holds for any $\lambda \in N_{\mathbb{R}_-}(\widetilde{q}(x^{t+1}, x^t, \zeta^t, \bar{l}_t))$. Thus, for any $\lambda \in N_{\mathbb{R}_-}(\widetilde{q}(x^{t+1}, x^t, \zeta^t, \bar{l}_t))$, we have that

$$
\widehat{N}_{[\widetilde{q} \le 0]}(x^{t+1}, x^t, \zeta^t, \bar{l}_t) \supseteq \lambda \widehat{\partial} \widetilde{q}(x^{t+1}, x^t, \zeta^t, \bar{l}_t)
$$

$$
(6.11) \qquad \overset{\text{(a)}}{=}
\begin{bmatrix}
\lambda[\nabla P_1(x^t) - \zeta^t + \bar{l}_t(x^{t+1} - x^t)] \\
\lambda[\nabla^2 P_1(x^t)(x^{t+1} - x^t) - \bar{l}_t(x^{t+1} - x^t)] \\
\lambda \partial P_2^*(\zeta^t) - \lambda x^{t+1} \\
\frac{\lambda}{2} \|x^{t+1} - x^t\|^2
\end{bmatrix}
\overset{\text{(b)}}{\ni}
\begin{bmatrix}
\lambda V_1^t \\
\lambda V_2^t \\
\lambda(x^t - x^{t+1}) \\
\frac{\lambda}{2} \|x^{t+1} - x^t\|^2
\end{bmatrix}
$$

with

$$
(6.12) \qquad
\begin{aligned}
V_1^t &:= \nabla P_1(x^t) - \zeta^t + \bar{l}_t(x^{t+1} - x^t), \\
V_2^t &:= \nabla^2 P_1(x^t)(x^{t+1} - x^t) - \bar{l}_t(x^{t+1} - x^t),
\end{aligned}
$$

where (a) uses the definition of $\widetilde{q}$, [34, Exercise 8.8(c)], [34, Proposition 10.5], and [34, Proposition 8.12] (so that $\partial P_2^*(\zeta^t) = \widehat{\partial} P_2^*(\zeta^t)$), and (b) uses (2.1) and the fact that $\zeta^t \in \partial P_2(x^t)$. On the other hand, we have from [34, Theorem 8.6] that

$$
\partial \widetilde{F}(x^{t+1}, x^t, \zeta^t, \bar{l}_t) \supseteq \widehat{\partial} \widetilde{F}(x^{t+1}, x^t, \zeta^t, \bar{l}_t)
$$

$$
(6.13) \qquad \overset{\text{(a)}}{\supseteq}
\begin{bmatrix}
\frac{1}{\|x^{t+1}\|} \partial \|x^{t+1}\|_1 - \frac{\|x^{t+1}\|_1}{\|x^{t+1}\|^3} x^{t+1} \\
0 \\
0 \\
0
\end{bmatrix}
+ \widehat{N}_{[\widetilde{q} \le 0]}(x^{t+1}, x^t, \zeta^t, \bar{l}_t),
$$

where (a) uses [34, Corollary 10.9], (2.3), and [34, Corollary 8.11], and the facts that $\widehat{\partial} \delta_{[\widetilde{q} \le 0]}(x^{t+1}, x^t, \zeta^t, \bar{l}_t) = \widehat{N}_{[\widetilde{q} \le 0]}(x^{t+1}, x^t, \zeta^t, \bar{l}_t)$ and $\widehat{N}_{\|\cdot\| \ge \rho}(x^{t+1}) = \{0\}$.

Let $\lambda_t \ge 0$ be a Lagrange multiplier of (6.1) with $l_t = \bar{l}_t$, which exists thanks to Lemma 6.3. In view of the inequality and the last equality in (6.10) and using (6.4)

with $(\widetilde{x}, l_t) = (x^{t+1}, \bar{l}_t)$, we deduce that $\lambda_t \in N_{\mathbb{R}_-}(\widetilde{q}(x^{t+1}, x^t, \zeta^t, \bar{l}_t))$, which in turn implies that $\frac{\lambda_t}{\|x^{t+1}\|} \in N_{\mathbb{R}_-}(\widetilde{q}(x^{t+1}, x^t, \zeta^t, \bar{l}_t))$. We can hence let $\lambda = \frac{\lambda_t}{\|x^{t+1}\|}$ in (6.11) to obtain an element in $\widehat{N}_{[\widetilde{q} \leq 0]}(x^{t+1}, x^t, \zeta^t, \bar{l}_t)$. Plugging this particular element into (6.13) yields

$$(6.14) \qquad \partial\widetilde{F}(x^{t+1}, x^t, \zeta^t, \bar{l}_t) \supseteq \begin{bmatrix} \frac{1}{\|x^{t+1}\|} \partial\|x^{t+1}\|_1 - \frac{\|x^{t+1}\|_1}{\|x^{t+1}\|^3} x^{t+1} + \frac{\lambda_t}{\|x^{t+1}\|} V_1^t \\ \frac{\lambda_t}{\|x^{t+1}\|} V_2^t \\ \frac{\lambda_t}{\|x^{t+1}\|}(x^t - x^{t+1}) \\ \frac{\lambda_t}{2\|x^{t+1}\|}\|x^{t+1} - x^t\|^2 \end{bmatrix},$$

where $V_1^t$ and $V_2^t$ are given in (6.12). On the other hand, applying (6.5) with $(\widetilde{x}, l_t) = (x^{t+1}, \bar{l}_t)$ and recalling that $\omega_t = \|x^t\|_1/\|x^t\|$, we obtain

$$(6.15) \qquad \begin{aligned} \partial\|x^{t+1}\|_1 &\ni \frac{\|x^t\|_1}{\|x^t\|^2} x^t - \lambda_t(\nabla P_1(x^t) - \zeta^t) - (\alpha + \lambda_t \bar{l}_t)(x^{t+1} - x^t) \\ &= \frac{\|x^t\|_1}{\|x^t\|^2} x^t - \lambda_t V_1^t - \alpha(x^{t+1} - x^t). \end{aligned}$$

Combining (6.14) and (6.15), we see further that

$$(6.16) \qquad \partial\widetilde{F}(x^{t+1}, x^t, \zeta^t, \bar{l}_t) \ni \begin{bmatrix} J_1^t \\ \frac{\lambda_t}{\|x^{t+1}\|} V_2^t \\ \frac{\lambda_t}{\|x^{t+1}\|}(x^t - x^{t+1}) \\ \frac{\lambda_t}{2\|x^{t+1}\|}\|x^{t+1} - x^t\|^2 \end{bmatrix},$$

where

$$J_1^t := \frac{1}{\|x^{t+1}\|}\left(\frac{\|x^t\|_1}{\|x^t\|^2} x^t - \frac{\|x^{t+1}\|_1}{\|x^{t+1}\|^2} x^{t+1}\right) - \frac{\alpha}{\|x^{t+1}\|}(x^{t+1} - x^t).$$

Next, recall from Lemma 6.1(iii) that

$$(6.17) \qquad \|x^{t+1}\| \geq \delta \quad \text{for all } t \in \mathbb{N}_+.$$

Using this together with our assumption that $\{x^t\}$ is bounded, we see that there exists $L_1 > 0$ such that

$$\left\|\frac{\|x^t\|_1}{\|x^t\|^2} x^t - \frac{\|x^{t+1}\|_1}{\|x^{t+1}\|^2} x^{t+1}\right\| \leq L_1\|x^{t+1} - x^t\| \quad \text{for all } t.$$

Combining the above three displays, we deduce that

$$(6.18) \qquad \|J_1^t\| \leq \frac{L_1 + \alpha}{\delta}\|x^{t+1} - x^t\|.$$

On the other hand, one can see from (6.17), the definition of $V_2^t$ (see (6.12)), the boundedness of $\{\lambda_t, \bar{l}_t\}$ (see Theorem 6.4(ii) and Lemma 6.1(ii)), the continuity of $\nabla^2 P_1$, and the boundedness of $\{x^t\}$ that there exist $L_2 > 0$ and $L_3 > 0$ such that

$$(6.19) \qquad \frac{\lambda_t}{\|x^{t+1}\|} \leq \frac{L_2}{\delta} \quad \text{and} \quad \left\|\frac{\lambda_t}{\|x^{t+1}\|} V_2^t\right\| \leq L_3\|x^{t+1} - x^t\|.$$

Moreover, we can see from Theorem 6.4(i) that there exists $\underline{t} \in \mathbb{N}_+$ such that

$$\|x^{t+1} - x^t\|^2 \leq \|x^{t+1} - x^t\|$$

whenever $t \geq \underline{t}$. Now we can conclude from (6.16), (6.18), (6.19), and the above display that there exists $\kappa > 0$ such that

$$\text{dist}(0, \partial \widetilde{F}(x^{t+1}, x^t, \zeta^t, \bar{l}_t)) \leq \kappa \|x^{t+1} - x^t\|$$

for all $t \geq \underline{t}$. This completes the proof.                                        □

When the sequence $\{x^t\}$ generated by $\text{MBA}_{\ell_1/\ell_2}$ is bounded, one can show that the set of accumulation points $\Omega$ of $\{(x^{t+1}, x^t, \zeta^t, \bar{l}_t)\}$ is compact. This together with Lemma 6.1(iii) and the continuity of $\widetilde{F}$ on its domain shows that $\widetilde{F}$ is constant on $\Omega \subseteq \text{dom}\,\partial\widetilde{F}$. Using this together with Proposition 6.5 and Lemma 6.1(iii), one can prove the following convergence result by imposing additional KL assumptions on $\widetilde{F}$. The proof is standard and follows the line of arguments as in [1, 2, 3, 9, 26, 40]. We omit the proof here for brevity.

THEOREM 6.6 (global convergence and convergence rate of $\text{MBA}_{\ell_1/\ell_2}$). *Consider* (1.2) *and assume that $P_1$ is twice continuously differentiable. Suppose that Assumption* 6.2 *holds. Let $\{x^t\}$ be the sequence generated by $MBA_{\ell_1/\ell_2}$ and assume that $\{x^t\}$ is bounded. If $\widetilde{F}$ in (6.8) is a KL function, then $\{x^t\}$ converges to a Clarke critical point $x^*$ of* (1.2) *($x^*$ is stationary if $q$ is in addition regular at $x^*$). Moreover, if $\widetilde{F}$ is a KL function with exponent $\theta \in [0, 1)$, then the following statements hold:*
(i) *If $\theta = 0$, then $\{x^t\}$ converges finitely.*
(ii) *If $\theta \in (0, \frac{1}{2}]$, then there exist $c_0 > 0$, $Q_1 \in (0, 1)$ and $\underline{t} \in \mathbb{N}_+$ such that*

$$\|x^t - x^*\| \leq c_0 Q_1^t \ \text{ for } \ t > \underline{t}.$$

(iii) *If $\theta \in (\frac{1}{2}, 1)$, then there exist $c_0 > 0$ and $\underline{t} \in \mathbb{N}_+$ such that*

$$\|x^t - x^*\| \leq c_0 t^{-\frac{1-\theta}{2\theta-1}} \ \text{ for } \ t > \underline{t}.$$

*Remark* 6.7 (KL property of $\widetilde{F}$ corresponding to (1.3), (1.4), and (1.5)).
(i) In both (1.3) and (1.4), we have $q = P_1$ being analytic and $P_2^* = \delta_{\{0\}}$. Hence $\widetilde{F}$ becomes $\widetilde{F}(x, y, \zeta, w) = \frac{\|x\|_1}{\|x\|} + \delta_\Delta(x, y, \zeta, w)$ with $\Delta = \{(x, y, \zeta, w) : P_1(y) + \langle \nabla P_1(y), x - y \rangle + \frac{w}{2}\|x - y\|^2 \leq 0, \zeta = 0, \|x\| \geq \rho\}$. Hence, the graph of $\widetilde{F}$ is

$$\left\{ (x, y, \zeta, w, z) : \begin{array}{ll} \|x\|_1 = z\|x\|, & \|x\| \geq \rho, \quad \zeta = 0, \\ P_1(y) + \langle \nabla P_1(y), x - y \rangle + \frac{w}{2}\|x - y\|^2 \leq 0 \end{array} \right\},$$

which is semianalytic [24, p. 596]. This means that $\widetilde{F}$ is subanalytic [24, Definition 6.6.1]. Moreover, the domain of $\widetilde{F}$ is closed and $\widetilde{F}|_{\text{dom}\,\widetilde{F}}$ is continuous. Therefore, $\widetilde{F}$ satisfies the KL property according to [7, Theorem 3.1].
(ii) For (1.5), first note that $P_2$ is a convex piecewise linear-quadratic function (see, for example, the proof of [26, Theorem 5.1]). Then $P_2^*$ is also a piecewise linear-quadratic function thanks to [34, Theorem 11.14]. Thus, one can see that $\widetilde{q}$ corresponding to (1.5) is semialgebraic and so is the set $\Theta = \{(x, y, \zeta, w) : \widetilde{q}(x, y, \zeta, w) \leq 0\}$. Therefore $\widetilde{F}$ is semialgebraic as the sum of the semialgebraic functions $x \mapsto \frac{\|x\|_1}{\|x\|} + \delta_{\|\cdot\| \geq \rho}(x)$ and $\delta_\Theta$ and is hence a KL function [2].

Using Theorem 6.6, Remark 6.7, Proposition 5.4, and the discussions preceding it, and recalling that continuously differentiable functions are regular, we have the following immediately corollary.

COROLLARY 6.8 (global convergence of $\text{MBA}_{\ell_1/\ell_2}$ for (1.3), (1.4), and (1.5)).
(i) *Suppose that $MBA_{\ell_1/\ell_2}$ is applied to (1.3) or (1.4). Then the sequence generated converges to a stationary point of the problem if the sequence is bounded.*
(ii) *Suppose that $MBA_{\ell_1/\ell_2}$ is applied to (1.5). Then the sequence generated converges to a Clarke critical point of the problem if the sequence is bounded.*

**7. Numerical simulations.** In this section, we perform numerical experiments on solving random instances of (1.3), (1.4), and (1.5) by $\text{MBA}_{\ell_1/\ell_2}$. All numerical experiments are performed in MATLAB 2019b on a 64-bit PC with an Intel Core i7-6700 CPU (3.40GHz) and 32GB of RAM.

We set $l_{\min} = 10^{-8}$, $l_{\max} = 10^8$, and $\alpha = 1$ in $\text{MBA}_{\ell_1/\ell_2}$. We let $l_0^0 = 1$ and compute, for each $t \geq 1$,

$$
l_t^0 = \begin{cases} \max\left\{l_{\min}, \min\left\{\frac{\langle d_x^t, d_g^t \rangle}{\|d_x^t\|^2}, l_{\max}\right\}\right\} & \text{if } \langle d_x^t, d_g^t \rangle \geq 10^{-12}, \\ \max\left\{l_{\min}, \min\left\{\frac{l_{t-1}}{2}, l_{\max}\right\}\right\} & \text{otherwise,} \end{cases}
$$

where $d_x^t = x^t - x^{t-1}$ and $d_g^t = \xi^t - \xi^{t-1}$ with $\xi^t = \nabla P_1(x^t) - \zeta^t$: specifically, $\zeta^t = 0$ when solving (1.3) and (1.4), while for (1.5), we pick any $\zeta^t \in \text{Proj}_S(Ax^t - b)$, which can be obtained by finding the largest $r$ entries of $Ax^t - b$.

We initialize $\text{MBA}_{\ell_1/\ell_2}$ at some feasible point $x_{\text{feas}}$ and terminate $\text{MBA}_{\ell_1/\ell_2}$ when

$$
(7.1) \qquad \|x^t - x^{t-1}\| \leq tol \cdot \max\{\|x^t\|, 1\};
$$

we will specify the choices of $x_{\text{feas}}$ and *tol* in each of the subsections below.

**7.1. Robust compressed sensing problems (1.5).** We generate a sensing matrix $A \in \mathbb{R}^{(p+\iota)\times n}$ with i.i.d. standard Gaussian entries and then normalize each column of $A$. Next, we generate the original signal $x_{\text{orig}} \in \mathbb{R}^n$ as a $k$-sparse vector with $k$ i.i.d. standard Gaussian entries at random (uniformly chosen) positions. We then generate a vector $z_\iota \in \mathbb{R}^\iota$ with i.i.d. standard Gaussian entries and set $z \in \mathbb{R}^{p+\iota}$ to be a vector with the first $p$ entries being zero and the last $\iota$ entries being $2\,\text{sign}(z_\iota)$. The vector $b$ in (1.5) is then generated as $b = Ax_{\text{orig}} - z + 0.01\varepsilon$, where $\varepsilon \in \mathbb{R}^{p+\iota}$ has i.i.d. standard Gaussian entries. Finally, we set $\sigma = 1.2\|0.01\varepsilon\|$ and $r = 2\iota$. In $\text{MBA}_{\ell_1/\ell_2}$, we set $x_{\text{feas}} = A^\dagger b$,[2] and $tol = 10^{-6}$ in (7.1).

In our numerical tests, we consider $(n, p, k, \iota) = (2560i, 720i, 80i, 10i)$ with $i \in \{2, 4, 6, 8, 10\}$. For each $i$, we generate 20 random instances as described above. The computational results are shown in Table 1. We present the time $t_{\text{qr}}$ for the (reduced) QR decomposition when generating $x_{\text{feas}}$, the CPU times $t_{\text{mba}}$ and $t_{\text{sum}}$,[3] the recovery error $\text{RecErr} = \frac{\|x_{\text{out}} - x_{\text{orig}}\|}{\max\{1, \|x_{\text{orig}}\|\}}$, and the Residual $= \text{dist}^2(Ax_{\text{out}} - b, S) - \sigma^2$, averaged over the 20 random instances, where $x_{\text{out}}$ is the approximate solution returned by $\text{MBA}_{\ell_1/\ell_2}$. We see that $x_{\text{orig}}$ are approximately recovered in a reasonable period of time.

---

[2] We compute $A^\dagger b$ via the MATLAB commands [Q,R] = qr(A',0); xfeas = Q*(R'\b).

[3] $t_{\text{mba}}$ is the run time of $\text{MBA}_{\ell_1/\ell_2}$, while $t_{\text{sum}}$ includes the run time of $\text{MBA}_{\ell_1/\ell_2}$, the time for performing (reduced) QR factorization on $A^T$, and the time for computing $Q(R^{-1})^T b$.

TABLE 1
*Random tests on robust CS.*

| $i$ | $t_{\mathrm{qr}}$ | $t_{\mathrm{mba}}(t_{\mathrm{sum}})$ | RecErr | Residual |
|---|---|---|---|---|
| 2 | 0.5 | 1.2 ( 1.7) | 3.3e-02 | -3e-11 |
| 4 | 3.1 | 4.1 ( 7.2) | 3.3e-02 | -5e-11 |
| 6 | 9.8 | 8.3 ( 18.1) | 3.3e-02 | -9e-11 |
| 8 | 24.0 | 14.3 ( 38.4) | 3.3e-02 | -1e-10 |
| 10 | 43.6 | 21.5 ( 65.3) | 3.3e-02 | -2e-10 |

TABLE 2
*Random tests on CS problems with Cauchy noise (tol = $10^{-6}$ for $\mathrm{SCP}_{\mathrm{ls}}$).*

| $i$ | $t_{\mathrm{qr}}$ | CPU | | RecErr | | Residual | |
|---|---|---|---|---|---|---|---|
| | | $\mathrm{SCP}_{\mathrm{ls}}$ | $\mathrm{MBA}_{\ell_1/\ell_2}$ | $\mathrm{SCP}_{\mathrm{ls}}$ | $\mathrm{MBA}_{\ell_1/\ell_2}$ | $\mathrm{SCP}_{\mathrm{ls}}$ | $\mathrm{MBA}_{\ell_1/\ell_2}$ |
| 2 | 0.5 | 10.0 | 0.6 ( 11.1) | 1.3e-01 | 6.5e-02 | -2e-07 | -8e-08 |
| 4 | 3.0 | 52.4 | 2.0 ( 57.5) | 1.3e-01 | 6.6e-02 | -6e-07 | -2e-07 |
| 6 | 9.4 | 87.3 | 4.1 ( 100.9) | 1.3e-01 | 6.6e-02 | -9e-07 | -2e-07 |
| 8 | 23.4 | 281.6 | 7.0 ( 312.1) | 1.3e-01 | 6.5e-02 | -1e-06 | -3e-07 |
| 10 | 42.4 | 285.5 | 11.4 ( 339.5) | 1.3e-01 | 6.5e-02 | -2e-06 | -4e-07 |

**7.2. CS problems with Cauchy noise (1.4).** Similar to the previous subsection, we generate the sensing matrix $A \in \mathbb{R}^{m \times n}$ with i.i.d. standard Gaussian entries and then normalize each column of $A$. We then generate the original signal $x_{\mathrm{orig}} \in \mathbb{R}^n$ as a $k$-sparse vector with $k$ i.i.d. standard Gaussian entries at random (uniformly chosen) positions. However, we generate $b$ as $b = Ax_{\mathrm{orig}} + 0.01\varepsilon$ with $\varepsilon_i \sim \mathrm{Cauchy}(0,1)$, i.e., $\varepsilon_i = \tan(\pi(\widetilde{\varepsilon}_i - 1/2))$ for some random vector $\widetilde{\varepsilon} \in \mathbb{R}^m$ with i.i.d. entries uniformly chosen in $[0,1]$. Finally, we set $\gamma = 0.02$ and $\sigma = 1.2\|0.01\varepsilon\|_{LL_2,\gamma}$.

We compare the $\ell_1$ minimization model (which minimizes $\ell_1$ norm in place of $\ell_1/\ell_2$ in (1.4); see [43, equation (5.8)] with $\mu = 0$) with our $\ell_1/\ell_2$ model. We use $\mathrm{SCP}_{\mathrm{ls}}$ in [43] for solving the $\ell_1$ minimization model. We use the same parameter settings for $\mathrm{SCP}_{\mathrm{ls}}$ as in [43, section 5], except that we terminate $\mathrm{SCP}_{\mathrm{ls}}$ when (7.1) is satisfied with $tol = 10^{-6}$ in Table 2. We initialize $\mathrm{MBA}_{\ell_1/\ell_2}$ at the approximate solution $x_{\mathrm{scp}}$ given by $\mathrm{SCP}_{\mathrm{ls}}$ and terminate $\mathrm{MBA}_{\ell_1/\ell_2}$ when (7.1) is satisfied with $tol = 10^{-6}$.

In our numerical experiments, we consider $(n,m,k) = (2560i, 720i, 80i)$ with $i \in \{2,4,6,8,10\}$. For each $i$, we generate 20 random instances as described above. Our computational results are presented in Table 2, which are averaged over the 20 random instances. Here we show the CPU time $t_{\mathrm{qr}}$ for performing (reduced) QR decomposition on $A^T$, the CPU time,[4] the recovery error $\mathrm{RecErr} = \frac{\|x_{\mathrm{out}} - x_{\mathrm{orig}}\|}{\max\{1, \|x_{\mathrm{orig}}\|\}}$, and the residual $\mathrm{Residual} = \|Ax_{\mathrm{out}} - b\|_{LL_2,\gamma} - \sigma$ of both $\mathrm{SCP}_{\mathrm{ls}}$ and $\mathrm{MBA}_{\ell_1/\ell_2}$, where $x_{\mathrm{out}}$ is the approximate solution returned by the respective algorithm. We see that the recovery error is significantly improved by solving the nonconvex model.

Finally, as suggested by one reviewer, we investigate the effect of initialization on the performance of $\mathrm{MBA}_{\ell_1/\ell_2}$. Specifically, we test $\mathrm{SCP}_{\mathrm{ls}}$ and $\mathrm{MBA}_{\ell_1/\ell_2}$ on the same set of instances used in Table 2 but terminate $\mathrm{SCP}_{\mathrm{ls}}$ when (7.1) is satisfied with $tol = 10^{-3}$. We then initialize $\mathrm{MBA}_{\ell_1/\ell_2}$ at the approximate solution returned by $\mathrm{SCP}_{\mathrm{ls}}$ and terminate $\mathrm{MBA}_{\ell_1/\ell_2}$ when (7.1) is satisfied with $tol = 10^{-6}$. The

---

[4]For $\mathrm{MBA}_{\ell_1/\ell_2}$, the time in parentheses is the total run time including the time for computing the initial point $A^\dagger b$ for $\mathrm{SCP}_{\mathrm{ls}}$ and the run times of $\mathrm{SCP}_{\mathrm{ls}}$ and $\mathrm{MBA}_{\ell_1/\ell_2}$, and the time without parentheses is the actual run time of $\mathrm{MBA}_{\ell_1/\ell_2}$ starting from $x_{\mathrm{feas}} = x_{\mathrm{scp}}$.

TABLE 3
*Random tests on CS problems with Cauchy noise (tol $= 10^{-3}$ for $\mathrm{SCP_{ls}}$).*

| $i$ | $t_{\mathrm{qr}}$ | CPU | | RecErr | | Residual | |
|---|---|---|---|---|---|---|---|
| | | $\mathrm{SCP_{ls}}$ | $\mathrm{MBA}_{\ell_1/\ell_2}$ | $\mathrm{SCP_{ls}}$ | $\mathrm{MBA}_{\ell_1/\ell_2}$ | $\mathrm{SCP_{ls}}$ | $\mathrm{MBA}_{\ell_1/\ell_2}$ |
| 2 | 0.5 | 3.0 | 50.8 ( 54.3) | 1.8e+00 | 1.6e+00 | -3e+01 | -6e-05 |
| 4 | 3.0 | 11.8 | 457.6 ( 472.5) | 4.3e+00 | 4.2e+00 | -1e+02 | -5e-04 |
| 6 | 9.5 | 30.5 | 4.9 ( 44.9) | 2.1e-01 | 6.6e-02 | -9e-01 | -2e-07 |
| 8 | 22.9 | 37.7 | 78.5 ( 139.2) | 9.7e+00 | 9.6e+00 | -6e+01 | -9e-03 |
| 10 | 41.5 | 71.9 | 3164.0 (3277.6) | 2.1e+00 | 1.7e+00 | -1e+02 | -2e-04 |

computational results are presented in Table 3. Not too surprisingly, we can see that $\mathrm{MBA}_{\ell_1/\ell_2}$ can result in large recovery errors with this initialization, though the recovery errors may still be small sometimes (see $i = 6$). Thus, the performance of $\mathrm{MBA}_{\ell_1/\ell_2}$ is quite sensitive to its initialization.

**7.3. Badly scaled CS problems with Gaussian noise (1.3).** In this section, we generate test instances similar to those in [39]. Specifically, we first generate $A = [a_1, \ldots, a_n] \in \mathbb{R}^{m \times n}$ with

$$a_j = \frac{1}{\sqrt{m}} \cos\left(\frac{2\pi wj}{F}\right), \quad j = 1, \ldots, n,$$

where $w \in \mathbb{R}^m$ have i.i.d. entries uniformly chosen in $[0, 1]$. Next, we generate the original signal $x_{\mathrm{orig}} \in \mathbb{R}^n$ using the following MATLAB command:

```
I = randperm(n); J = I(1:k); xorig = zeros(n,1);
xorig(J) = sign(randn(k,1)).*10.^(D*rand(k,1));
```

We then set $b = Ax_{\mathrm{orig}} + 0.01\varepsilon$, where $\varepsilon \in \mathbb{R}^m$ has i.i.d standard Gaussian entries. Finally, we set $\sigma = 1.2\|0.01\varepsilon\|$.

We compare the $\ell_1$ minimization model (which minimizes $\ell_1$ norm in place of $\ell_1/\ell_2$ in (1.3); see [43, equation (5.5)] with $\mu = 0$) with our $\ell_1/\ell_2$ model. The $\ell_1$ minimization model is solved via SPGL1 [5] (version 2.1) using default settings. The initial point for $\mathrm{MBA}_{\ell_1/\ell_2}$ is generated from the approximate solution $x_{\mathrm{spgl1}}$ of SPGL1 as follows: Specifically, since $x_{\mathrm{spgl1}}$ may violate the constraint slightly, we set the initial point of $\mathrm{MBA}_{\ell_1/\ell_2}$ as

$$x_{\mathrm{feas}} = \begin{cases} A^\dagger b + \sigma \frac{x_{\mathrm{spgl1}} - A^\dagger b}{\|Ax_{\mathrm{spgl1}} - b\|} & \text{if } \|Ax_{\mathrm{spgl1}} - b\| > \sigma, \\ x_{\mathrm{spgl1}} & \text{otherwise.} \end{cases}$$

We terminate $\mathrm{MBA}_{\ell_1/\ell_2}$ when (7.1) is satisfied with $tol = 10^{-8}$.

In our numerical tests, we set $n = 1024$, $m = 64$, and consider $k \in \{8, 12\}$, $F \in \{5, 15\}$, and $D \in \{2, 3\}$. For each $(k, F, D)$, we generate 20 random instances as described above. We present the computational results (averaged over the 20 random instances) in Table 4. Here we show the CPU time,[5] the recovery error RecErr $= \frac{\|x_{\mathrm{out}} - x_{\mathrm{orig}}\|}{\max\{1, \|x_{\mathrm{orig}}\|\}}$, and the Residual $= \|Ax_{\mathrm{out}} - b\|^2 - \sigma^2$ of both SPGL1 and $\mathrm{MBA}_{\ell_1/\ell_2}$, where $x_{\mathrm{out}}$ is the approximate solution returned by the respective algorithm. We again observe that the recovery error is significantly improved (on average) by solving

---

[5] For $\mathrm{MBA}_{\ell_1/\ell_2}$, the time in parenthesis is the total run time including the time for computing the feasible point $A^\dagger b$ and the run times of SPGL1 and $\mathrm{MBA}_{\ell_1/\ell_2}$, the time without parenthesis is the actual run time of $\mathrm{MBA}_{\ell_1/\ell_2}$ starting from $x_{\mathrm{feas}}$.

TABLE 4
*Random tests on badly scaled CS problems with Gaussian noise.*

| $k$ | $F$ | $D$ | CPU | | RecErr | | Residual | |
|---|---|---|---|---|---|---|---|---|
| | | | SPGL1 | $\text{MBA}_{\ell_1/\ell_2}$ | SPGL1 | $\text{MBA}_{\ell_1/\ell_2}$ | SPGL1 | $\text{MBA}_{\ell_1/\ell_2}$ |
| 8 | 5 | 2 | 0.07 | 0.13 ( 0.20) | 3.2e-02 | 2.3e-03 | -4e-05 | -1e-13 |
| 8 | 5 | 3 | 0.06 | 0.14 ( 0.20) | 3.2e-03 | 6.8e-04 | -4e-05 | -2e-11 |
| 8 | 15 | 2 | 0.08 | 3.92 ( 4.01) | 4.7e-01 | 1.5e-01 | -9e-05 | -7e-13 |
| 8 | 15 | 3 | 0.11 | 31.46 ( 31.58) | 3.8e-01 | 5.3e-02 | 2e-02 | -5e-11 |
| 12 | 5 | 2 | 0.06 | 2.26 ( 2.32) | 1.4e-01 | 3.6e-02 | -3e-04 | -8e-13 |
| 12 | 5 | 3 | 0.08 | 4.05 ( 4.14) | 6.0e-02 | 3.8e-03 | 1e-04 | -7e-11 |
| 12 | 15 | 2 | 0.09 | 8.32 ( 8.41) | 5.2e-01 | 2.0e-01 | -1e-04 | -1e-12 |
| 12 | 15 | 3 | 0.11 | 403.80 (403.91) | 5.2e-01 | 1.5e+00 | 6e-02 | -3e-10 |

the nonconvex model in most instances, except when $(k, F, D) = (12, 15, 3)$. In this case, we see that the $x_{\text{spgl1}}$ can be highly infeasible and thus the starting point $x_{\text{feas}}$ provided to $\text{MBA}_{\ell_1/\ell_2}$ may not be a good starting point. This might explain the relatively poor performance of $\text{MBA}_{\ell_1/\ell_2}$ in this case.

## 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] H. ATTOUCH, J. BOLTE, P. REDONT, AND A. SOUBEYRAN, *Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-Łojasiewicz inequality*, Math. Oper. Res., 35 (2010), pp. 438–457.

[3] H. ATTOUCH, J. BOLTE, AND B. F. SVAITER, *Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods*, Math. Program. 137 (2013), pp. 91–129.

[4] A. AUSLENDER, R. SHEFI, AND M. TEBOULLE, *A moving balls approximation method for a class of smooth constrained minimization problems*, SIAM J. Optim., 20 (2010), pp. 3232–3259.

[5] E. VAN DEN BERG AND M. P. FRIEDLANDER, *Probing the Pareto frontier for basis pursuit solutions*, SIAM J. Sci. Comput., 31 (2008), pp. 890–912.

[6] J. BOLTE, Z. CHEN, AND E. PAUWELS, *The multiproximal linearization method for convex composite problems*, Math. Program., 186 (2020), pp. 1–36, doi:10.1007/s10107-019-01382-3.

[7] J. BOLTE, A. DANIILIDIS, AND A. LEWIS, *The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamic systems*, SIAM. J. Optim., 17 (2007), pp. 1205–1223.

[8] J. BOLTE AND E. PAUWELS, *Majorization-minimization procedures and convergence of SQP methods for semi-algebraic and tame programs*, Math. Oper. Res., 41 (2016), pp. 442–465.

[9] J. BOLTE, S. SABACH, AND M. TEBOULLE, *Proximal alternating linearized minimization for nonconvex and nonsmooth problems*, Math. Program., 146 (2014), pp. 459–494.

[10] J. M. BORWEIN AND A. S. LEWIS, *Convex Analysis and Nonlinear Optimization*, 2nd ed., Springer, New York, 2006.

[11] J. M. BORWEIN AND Q. J. ZHU, *Techniques of Variational Analysis*, Springer, New York, 2004.

[12] R. I. BOŢ AND E. R. CSETNEK, *Proximal-gradient algorithms for fractional programming*, Optimization, 66 (2017), pp. 1383–1396.

[13] R. I. BOŢ, M. N. DAO, AND G. LI, *Extrapolated Proximal Subgradient Algorithms for Nonconvex and Nonsmooth Fractional Programs*, https://arxiv.org/abs/2003.04124, 2020.

[14] E. J. CANDÉS, J. K. ROMBERG, AND T. TAO, *Stable signal recovery from incomplete and inaccurate measurements*, Comm. Pure Appl. Math., 59 (2006), pp. 1207–1223.

[15] E. J. CANDÉS AND T. TAO, *Decoding by linear programming*, IEEE Trans. Inform. Theory, 51 (2005), pp. 4203–4251.

[16] R. E. CARRILLO, K. E. BARNER, AND T. C. AYSAL, *Robust sampling and reconstruction methods for sparse signals in the presence of impulsive noise*, IEEE J. Sel. Topics Signal Process., 4 (2010), pp. 392–408.

[17] R. E. CARRILLO, A. B. RAMIREZ, G. R. ARCE, K. E. BARNER, AND B. M. SADLER, *Robust compressive sensing of sparse signals: A review*, EURASIP J. Adv. Signal Process., 108, 2016.

[18] R. Chartrand, *Exact reconstruction of sparse signals via nonconvex minimization*, IEEE Signal Process. Lett., 14 (2007), pp. 707–710.

[19] S. S. Chen, D. L. Donoho, and M. A. Saunders, *Atomic decomposition by basis pursuit*, SIAM. Rev., 43 (2001), pp. 129–159.

[20] X. Chen, Z. Lu, and T. K. Pong, *Penalty methods for a class of non-Lipschitz optimization problems*, SIAM J. Optim., 26 (2016), pp. 1465–1492.

[21] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Classics in Appl. Math. 5, SIAM, Philadelphia, 1990.

[22] J.-P. Crouzeix, J. A. Ferland, and H. V. Nguyen *Revisiting Dinkelbach-type algorithms for generalized fractional programs*, Opsearch, 45 (2008), pp. 97–110.

[23] W. Dinkelbach, *On nonlinear fractional programming*, Manage. Sci., 13 (1967), pp. 492–498.

[24] F. Facchinei and J.-S. Pang, *Finite-dimensional Variational Inequalities and Complementarity Problems, Part* I, Springer, New York, 2003.

[25] 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.

[26] T. Liu, T. K. Pong, and A. Takeda, *A refined convergence analysis of* pDCA$_e$ *with applications to simultaneous sparse recovery and outlier detection*, Comput. Optim. Appl., 73 (2019), pp. 69–100.

[27] Y. Lou, P. Yin, Q. He, and J. Xin, *Computing sparse representation in a highly coherent dictionary based on difference of $L_1$ and $L_2$*, J. Sci. Comput., 64 (2015), pp. 178–196.

[28] B. S. Mordukhovich, *Variational Analysis and Generalized Differentiation* I, Springer, New York, 2006.

[29] B. K. Natarajan, *Sparse approximate solutions to linear systems*, SIAM J. Comput., 24 (1995), pp. 227–234.

[30] H. van Ngai and M. Théra, *Error bounds for systems of lower semicontinuous functions in Asplund spaces*, Math. Program., 116 (2009), pp. 397–427.

[31] L. F. Polania, R. E. Carrillo, M. Blanco-Velasco, and K. E. Barner, *Compressive sensing for ECG signals in the presence of electromyography noise*, in Proceeding of the 38th Annual Northeast Bioengineering Conference, 2012, pp. 295–296.

[32] Y. Rahimi, C. Wang, H. Dong, and Y. Lou, *A scale invariant approach for sparse signal recovery*, SIAM J. Sci. Comput., 41 (2019), pp. A3649–A3672.

[33] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

[34] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer, New York, 1998.

[35] R. Shefi and M. Teboulle, *A dual method for minimizing a nonsmooth objective over one smooth inequality constraint*, Math. Program., 159 (2016), pp. 137–164.

[36] H. Tuy, *Convex Analysis and Global Optimization*, Springer, New York, 1998.

[37] S. A. Vavasis, *Derivation of Compressive Sensing Theorems from the Spherical Section Property*, University of Waterloo, 2009.

[38] C. Wang, M. Tao, J. Nagy, and Y. Lou, *Limited-angle CT reconstruction via the $L_1/L_2$ minimization*, SIAM J. Imaging Sci., 14 (2021), pp. 749–777.

[39] C. Wang, M. Yan, and Y. Lou, *Accelerated schemes for the $L_1/L_2$ minimization*, IEEE Trans. Signal Process., 68 (2020), pp. 2660–2669.

[40] B. Wen, X. Chen, and T. K. Pong, *A proximal difference-of-convex algorithm with extrapolation*, Comput. Optim. Appl., 69 (2018), pp. 297–324.

[41] P. Yin, E. Esser, and J. Xin, *Ratio and difference of $\ell_1$ and $\ell_2$ norms and sparse representation with coherent dictionaries*, Commun. Inf. Syst., 14 (2014), pp. 87–109.

[42] P. Yu, G. Li, and T. K. Pong, *Kurdyka-Łojasiewicz Exponent via inf-Projection*, https://arxiv.org/abs/1902.03635, 2019.

[43] P. Yu, T. K. Pong, and Z. Lu, *Convergence rate analysis of a sequential convex programming method with line search for a class of constrained difference-of-convex optimization problems*, SIAM J. Optim., to appear.

[44] Y. Zhang, *Theory of compressive sensing via L1-minimization: A non-RIP analysis and extensions*, J. Oper. Res. Soc. China, 1 (2013), pp. 79–105.