A Null-Space-Based Weighted ℓ_1 Minimization Approach to Compressed Sensing*

Shenglong Zhou[†], Naihua Xiu[‡], Yingnan Wang[‡], Lingchen Kong[‡] and Hou-Duo Qi[§]

June 6, 2015; Revised December 27, 2015

Abstract

It has become an established fact that the constrained ℓ_1 minimization is capable of recovering the sparse solution from a small number of linear observations and the reweighted version can significantly improve its numerical performance. The recoverability is closely related to the Restricted Isometry Constant (RIC) of order s (s is an integer), often denoted as δ_s. A class of sufficient conditions for successful k-sparse signal recovery often take the form $\delta_{tk} < c$, for some $t \geq 1$ and c being a constant. When $t > 1$, such a bound is often called RIC bound of high order. There exist a number of such bounds of RICs, high order or not. For example, a high order bound is recently given by Cai and Zhang [10]: $\delta_{tk} < \sqrt{(t-1)/t}$, and this bound is known sharp for $t \geq 4/3$. In this paper, we propose a new weighted ℓ_1 minimization which only requires the following RIC bound that is more relaxed (i.e., bigger) than the above mentioned bound:

$$\delta_{tk} < \sqrt{\frac{t-1}{t-(1-\omega^2)}},$$

where $t > 1$ and $0 < \omega \leq 1$ is determined by two optimizations of a similar type over the null space of the linear observation operator. In tackling the combinatorial nature of the two optimization problems, we develop a reweighted ℓ_1 minimization that yields a sequence of approximate solutions, which enjoy strong convergence properties. Moreover, the numerical performance of the proposed method is very satisfactory when compared to some of the state-of-the-art methods in compressed sensing.

Keywords: Compressed sensing, weighted ℓ_1 minimization, restricted isometry constant, null space property.

1 Introduction

Over the past decade, compressed sensing has seen revolutionary advances both in theory and algorithms. A significant number of publications have appeared in various disciplines and we only list a few of ground-breaking papers that pioneered the advances [16, 13, 12]. For more references, see the survey papers [5, 31] and the monographs [18, 20, 29]. Among many beautiful theoretical results is the well-established fact that the constrained ℓ_1 minimization is capable of recovering the sparse information based only on a small number of observations. Moreover, the

*This work is supported by the National Natural Science Foundation of China (11431002, 11171018, 71271021).
[†]School of Mathematics, University of Southampton, Southampton SO17 1BJ, UK (sz3g14@soton.ac.uk).
[‡]Department of Applied Mathematics, Beijing Jiaotong University, Beijing, China (nhxiu@bjtu.edu.cn).
[‡]Department of Applied Mathematics, Beijing Jiaotong University, Beijing, China (wyn1982@hotmail.com).
[‡]Department of Applied Mathematics, Beijing Jiaotong University, Beijing, China (konglchen@126.com).
[§]School of Mathematics, University of Southampton, Southampton SO17 1BJ, UK (hdqi@soton.ac.uk).

reweighted ℓ_1 minimization provides a more robust numerical performance. This has been well argued by Candès, Wakin and Boyd [14]. The recoverability of ℓ_1 minimization is closely related to the Restricted Isometry Constant (RIC) of the observation operator and a large number of RIC bounds have been derived, see e.g., [11, 6, 2, 3, 27, 8, 9, 40, 10], which also motivated our research in this paper. In particular, some RIC bounds have been known to be sharp (i.e., tight). The purpose of this paper is to show that one of the recently discovered sharp bounds for the constrained ℓ_1 minimization can actually be relaxed when the weighted ℓ_1 minimization is used. Our result prompts us to design a corresponding reweighted ℓ_1 minimization, whose convergence and numerical performance are demonstrated to be satisfactory when compared to other popular methods. In the following, we explain how we have derived our weighted scheme.

(a) ℓ_1 minimization and RIC bounds. Suppose we have a known measurement matrix $\Phi \in \mathbb{R}^{m\times n}$ with $m \ll n$ and the actual observation vector $b \in \mathbb{R}^m$. The purpose is to recover the sparsest vector $x \in \mathbb{R}^n$ such that $\Phi x = b$. The underlying optimization model is the ℓ_0 minimization:

$$\min \ \|x\|_0, \qquad \text{s.t.} \qquad \Phi x = b, \tag{1}$$

where $\|x\|_0$ is ℓ_0-norm of x, i.e., the number of nonzero entries in x. It goes without needing to explain further that under certain conditions this combinatorial optimization problem (1) can be satisfactorily solved by its convex relaxation: the ℓ_1 minimization:

$$\min \ \|x\|_1, \qquad \text{s.t.} \qquad \Phi x = b, \tag{2}$$

where $\|x\|_1 = \sum_{i=1}^n |x_i|$ is the ℓ_1 norm of x. As commented in [14], "the use of ℓ_1 relaxation has become so widespread that it could arguably be considered the modern least squares", see, e.g., [5, 6, 8, 9, 10, 22, 27, 34, 38, 40].

One of the key theory behind the ℓ_1 minimization is about the RIC of the measurement matrix Φ introduced by Candès and Tao [13]. Suppose $1 \le k \le n$ is an integer, the RIC of order k is defined as the smallest number δ_k such that for all k-sparse vectors $x \in \mathbb{R}^n$ (i.e., $\|x\|_0 \le k$), it satisfies

$$(1-\delta_k)\|x\|^2 \le \|\Phi x\|^2 \le (1+\delta_k)\|x\|^2,$$

where $\|x\|$ is the Euclidean norm of x (also known as the ℓ_2 norm). The RIC of order k is denoted as δ_k. In general, δ_{tk} for $t > 1$ is often referred as the RIC of a high order. There certainly exists a research race in deriving better (i.e., higher values) RIC bounds [1, 6, 8, 9, 10, 27, 40]. For example, $\delta_{2k} < 0.5746$ jointly with $\delta_{8k} < 1$ [40], an improved bound $\delta_{2k} < 4/\sqrt{41}$ [1], sharp bounds $\delta_{2k} < \sqrt{2}/2$ [10] and $\delta_k < 1/3$ [8]. In particular, Cai and Zhang [10, Theorem 1.1] proved an RIC bound of high order in the form:

$$\delta_{tk} < \sqrt{\frac{t-1}{t}}, \tag{3}$$

and this bound is sharp for $t \ge 4/3$. Note that (3) implies sharp bound $\delta_{2k} < \sqrt{2}/2$ by choosing $t = 2$. Under each of those bounds, the constrained ℓ_1 minimization (2) is able to exactly recover the k sparse solution of $\Phi x = b$ provided that it exists. One interesting question is whether those sharp bounds can be improved in the weighted scheme. The answer is positive under certain circumstances.

Before moving on, we would like to clarify the usage of two terminological terms, namely RIC bound and Bound of high order. In compresed sensing, a class of sufficient conditions for successful sparse signal recovery through the ℓ_1 minimization (2) often take the following form:

$$\delta_{tk} < c, \tag{4}$$

for some $t \geq 1$ and c being a constant. In this paper, we refer (4) as the RIC bound or simply the bound. This follows the routine taken up by Cai et al. [6]. The term of RIC bound is often used when the focus of research was to improve it (i.e., to improve c). Otherwise, referring (4) as a sufficient condition would be adequate. We further distinguish the situation $t = 1$ from the remaining case $t > 1$. If $t > 1$, we refer it as the RIC bound of high order. For example, the bound in (3) is one of high order tk. The usage of high order was also adopted in Cai and Zhang [10].

(b) Weighted ℓ_1 minimization and RIC bounds. Let $\mathrm{w} \in \mathbb{R}^n$ be nonnegative ($\mathrm{w}_i \geq 0$ for $i = 1, \ldots, n$). The w-weighted norm of $x \in \mathbb{R}^n$ is defined as

$$\|x\|_{1,\mathrm{w}} := \sum_{i=1}^{n} \mathrm{w}_i |x_i|.$$

The w-weighted ℓ_1 minimization is

$$\min \ \|x\|_{1,\mathrm{w}}, \qquad \text{s.t.} \qquad \Phi x = b. \tag{5}$$

At this point, it is convenient to introduce another notation for the weighted ℓ_1: $\mathrm{w} \circ x$, which is the vector of the componentwise product $(\mathrm{w}_i x_i)$ between w and x. The operation $\circ$ in literature is known as the Hadamard product of vectors of same size. Therefore $\|x\|_{1,\mathrm{w}} = \|\mathrm{w} \circ x\|_1$. This new notation will be very convenient later on for our convergence analysis. In numerical implementation, w is often updated each step, resulting in the reweighted ℓ_1 minimization scheme in compressed sensing. This scheme has proved to be much more efficient than the unweighted version [14]. Ever since, there have appeared a large number of reweighted ℓ_1 minimization with a variety of updating rules, see [32, 21, 39, 38, 36] and the references therein.

A particular weighting scheme that has attracted much attention is when partial information is available for the actual support of the optimal sparse solution [32]. Suppose that x^* is the k-sparse solution and we let T_* to denote its support (i.e., the set of indices of nonzero elements in x^*). Suppose T_0 is another index set that approximates T_*. Friedlander et al. [21] proposed the following weighting scheme to define $\mathrm{w} \in \mathbb{R}^n$:

$$\mathrm{w}_i := \begin{cases} \omega & \text{if } i \in T_0 \\ 1 & \text{if } i \in T_0^c, \end{cases} \tag{6}$$

where T_0^c denotes the complementary set of T_0 in $\{1, \ldots, n\}$ and $0 \leq \omega \leq 1$. What makes this weighting scheme more valuable than many others is that it gives rise to RIC bounds of high orders (see [21, Theorem 3]). In particular, when $|T_0| = k$ (i.e., there are k indices in T_0 and this implies $\rho = 1$ in [21, Eq.(15)]), the sufficient condition in [21, Eq.(15)] that guarantees successful signal recovery becomes

$$\delta_{2k} < \frac{1}{\sqrt{2}(\omega + (1-\omega)\sqrt{2(1-\alpha)}) + 1}, \tag{7}$$

where $\alpha = |T_0 \cap T_*|/k$ (i.e., the percentage of true indices in T). An interesting property of this sufficient condition is that $\delta_{2k} \to 1$ as $\alpha \to 1$ and $\omega = 0$. In other words, the high order bound of δ_{2k} gets close to 1 if the more and more accurate support set is used in the weighted scheme. The corresponding (sharp) bound for the unweighted scheme is $\sqrt{2}/2$. A significant improvement is achieved through this particular weighting scheme, though it remains an issue as how to accurately approximate the true support T_*.

In this paper, we propose a novel choice of T_0 based on the null space of Φ. We will derive a RIC bound of high order δ_{tk} for any $t > 1$, following the spirit of Cai and Zhang [10]. We explain our bound below.

(c) A new weighting scheme. Let $\mathcal{N}$ be the null space of Φ. We consider the ℓ_1-unit sphere in the null space defined by

$$\mathcal{B}_1 := \{h \in \mathbb{R}^n \mid h \in \mathcal{N}, \quad \|h\|_1 = 1\}.$$

Let $\mathbb{I}(k)$ denote the collection of all index sets of size not greater than k:

$$\mathbb{I}(k) := \{S \subset \{1, \ldots, n\} \ \mid \ |S| \le k\}.$$

For an index set T and $x \in \mathbb{R}^n$, we let x_T be the vector of x with its elements not indexed by T being set 0. Our selection of T_0 for the weighting scheme (6) is obtained from the following (combinatorial) optimization problem:

$$(T_0, \ \widehat{h}) := \mathop{\arg\max}_{T \in \mathbb{I}(k), \ h \in \mathcal{B}_1} \ \|h_T\|_1. \tag{8}$$

Our main result (see Theorem 3.2) states that there exists $0 \le \omega = \gamma_\Phi \le 1$ (see (13) for the definition of γ_Φ) such that the condition

$$\delta_{tk} < \sqrt{\frac{t-1}{t-(1-\omega^2)}}, \tag{9}$$

for some $t > 1$ is sufficient for the weighted ℓ_1 minimization (5) to recover a k sparse solution of (2). Here, if tk is not an integer, we replace it by $\lceil tk \rceil$, where $\lceil a \rceil$ stands for the smallest integer that is no less than a. It is easy to see that the bound in (9) is bigger than the Cai-Zhang bound in (3) [10] when $\omega < 1$. An interesting property is that this bound also approaches 1 if ω approaches 0. This property is similar to that enjoyed by (7). The difference is that our bound in (9) is for any high order $t > 1$, while (7) is just for $t = 2$. We also note that in the extreme case where $\alpha = 1$ (i.e., when the true support is used for T_0 in (6)) and $t = 2$, the bound in (9) is bigger than that in (7). In general, they belong to different types of bounds.

Before we finish this part, we would like to point out that problem (8) is not easy to solve. In the numerical part, we will propose an update mechanism (see (27)) that approximates the solution of this problem. This results in a new iteratively re-weighted ℓ_1 minimization. We will show that, under reasonable assumptions, our update formula will correctly identify the actual support set (see Theorem 4.4). Moreover, the numerical results show that it performs satisfactorily comparing to some of the state-of-the-art algorithms in compressed sensing.

The paper is organized as follows. In Section 2, we introduce the known weighted null space property in terms of the ℓ_1-unit sphere $\mathcal{B}_1$. We prove our new bound (9) in Section 3. In Section 4, we describe our reweighted ℓ_1 scheme and study its convergence analysis. We report some numerical results in Section 5, where extensive comparison with a number of popular methods shows the satisfactory performance of our proposed method. We conclude the paper with a few remarks in Section 6.

2 Weighted Null Space Property

In this section, we introduce the Weighted Null Space Property (WNSP) [25], which is a generalization of the famous Null Space Property (NSP) in compressed sensing. We rephrase the WNSP in a slightly different way to suit our need later in deriving the new RIC bound. We also include an example to highlight the key difference between WNSP and NSP.

It is well known that the NSP is a necessary and sufficient condition for the ℓ_1 minimization (2) to produce the k-sparse solution of $\Phi x = b$ [17, 22, 23]. We say that the matrix Φ satisfies the NSP of order k if it holds

$$\|h_S\|_1 < \|h_{S^c}\|_1, \qquad \forall \ S \in \mathbb{I}(k), \ 0 \ne h \in \mathcal{N}.$$

It can also be defined by restricting h to be in the ℓ_1 unit sphere in $\mathcal{N}$ and we state this below.

$$\|h_S\|_1 < \|h_{S^c}\|_1, \qquad \forall\ S \in \mathbb{I}(k),\ h \in \mathcal{B}_1. \tag{10}$$

Our definition of WNSP is to replace the ℓ_1 norm in (10) by the weighted ℓ_1 norm and we formally describe it below.

Definition 2.1 (WNSP) *For a given weight* $\mathrm{w} \in \mathbb{R}^n$ *with* $\mathrm{w} \geq 0$ *the matrix* Φ *is to said to have the* $\mathrm{w}-$*weighted null space property of order* k *if it holds*

$$\|h_S\|_{1,\mathrm{w}} < \|h_{S^c}\|_{1,\mathrm{w}}, \qquad \forall\ S \in \mathbb{I}(k),\ h \in \mathcal{B}_1. \tag{11}$$

We note that this definition is equivalent to the one defined in [25], where $0 \neq h$ runs over the null space $\mathcal{N}$. It is known that the WNSP of order k is a necessary and sufficient condition for the weighted ℓ_1 minimization (2) to recover the k-sparse solution of $\Phi x = b$ (assuming it has one). The next example shows a scenario in which ℓ_1 minimization (2) fails to recover the $1-$sparse solution, while we have a plenty of choices for the weighted ℓ_1 minimization (5) to recover it. The problem will be used again to explain our main result.

Example 2.2 *Let the measurement matrix* Φ *and observation vector* b *be given as*

$$\Phi = \begin{pmatrix} 4/5 & 0 & 3/10 \\ 0 & 4/5 & 3/10 \end{pmatrix}, \quad b = \begin{pmatrix} 3/5 \\ 3/5 \end{pmatrix}.$$

It is easy to verify the unique solution of ℓ_0 and ℓ_1 minimizations are $x^{(0)} = (0, 0, 2)^\top$ and $x^{(1)} = (\frac{3}{4}, \frac{3}{4}, 0)^\top$ respectively. However, if we set the weight $\mathrm{w} \in \mathbb{R}^3$ to satisfy: $\mathrm{w}_2 = \mathrm{w}_1$, $\mathrm{w}_3 < \frac{3}{4}\mathrm{w}_1$ and $0 < \mathrm{w}_1 \leq 1$, we can verify that the weighted ℓ_1 minimization (5) is able to recover the $1-$sparse solution x^0. This is because the w-weighted NSP (11) is satisfied by verifying it by directly using

$$\mathcal{B}_1 = \left\{ h = (3h_3/8,\ 3h_3/8,\ -h_3)^\top \ \middle|\ h_3 = \pm\frac{4}{7} \right\}.$$

It is also easy to verify that NSP (10) is not satisfied by choosing $S = \{3\}$, which yields

$$\|h_S\|_1 = |h_3| > |\frac{6}{8}h_3| = \|h_{S^c}\|_1, \qquad \text{for } h \in \mathcal{B}_1$$

violating the NSP inequality (10).

In next section, we will derive our main RIC bound that generalizes the Cai-Zhang bound (3) to the weighted ℓ_1 minimization. Some of our proof techniques are taken from [10]. In particular, we will use the main technical lemma about sparse representation of a special polytope in [10, Lemma 1.1], which is stated below for easy reference.

Lemma 2.3 *For a positive number* α *and a positive integer* s*, define the polytope* $T(\alpha, s) \subset \mathbb{R}^n$ *by*

$$T(\alpha, s) = \{v \in \mathbb{R}^n \mid \|v\|_\infty \leq \alpha, \|v\|_1 \leq s\alpha\},$$

where $\|\cdot\|_\infty$ *is the infinity norm. For any* $v \in \mathbb{R}^n$*, define the set* $U(\alpha, s, v) \subset \mathbb{R}^n$ *of sparse vectors by*

$$U(\alpha, s, v) = \{u \in \mathbb{R}^n \mid supp(u) \subseteq supp(v), \|u\|_0 \leq s, \|u\|_1 = \|v\|_1, \|u\|_\infty \leq \alpha\}.$$

Then $v \in T(\alpha, s)$ *if and only if* v *is in the convex hull of* $U(\alpha, s, v)$*. In particular, any* $v \in T(\alpha, s)$ *can be expressed as*

$$v = \sum_{i=1}^{N} \lambda_i u_i, \quad and \quad 0 \leq \lambda_i \leq 1, \quad \sum_{i=1}^{N} \lambda_i = 1, \quad u_i \in U(\alpha, s, v).$$

3 New RIC Bound

Our main purpose in this section is to prove that the use of the weight defined by (6) and (8) will lead to an improved RIC bound (9). We first address the choice of ω in (6). We consider the following problem:

$$(T_1,\ \widetilde{h}) := \mathop{\arg\max}_{T\in\mathbb{I}(k)\backslash\{T_0\},\ h\in\mathcal{B}_1} \ \|h_T\|_1. \tag{12}$$

We define a quantity that is the ratio of the optimal objectives from (8) and (12):

$$\gamma_\Phi := \frac{\|\widetilde{h}_{T_1}\|_1}{\|\widehat{h}_{T_0}\|_1}. \tag{13}$$

Obviously, $0 \le \gamma_\Phi \le 1$. We note that both problems (8) and (12) may have multiple solutions. Hence, for technical correctness, $(T_0, \widehat{h})$ should be one of the optimal solutions. We use $=$ (instead of $\in$) in (8) for the sake of simplicity (this comments also applies to problem (12)). We also note that the quantity γ_Φ is well defined because only the optimal objectives from the two optimization problems are involved. Furthermore, γ_Φ would depend on the choice of T_0 in (8) in the case there are multiple solutions. Different choices of T_0 may lead to different numerators in (13), but with the same denominator. However, this dependence does not enter into our analysis. What actually matters to us is the optimality condition of (12).

We have the following technical results. The first result says that $\widehat{h}$ is also the w-weighted ℓ_1 solution over the ℓ_1-unit sphere $\mathcal{B}_1$. The second result says that if ω is chosen bigger than γ_Φ in (6), then $(T_0, \widehat{h})$ is also an optimal solution of the w-weighted problem of (8). The third result studies when it further satisfies the weighted NSP.

Lemma 3.1 *Let the weight vector* w *be defined as in (6) in which T_0 is a solution of (8), and γ_Φ be defined as (13) and (12) corresponding to such T_0, the following results hold.*

(i) For any $0 \le \omega \le 1$ in (6), $\widehat{h}$ has the following minimality property:

$$\|\widehat{h}\|_{1,\mathrm{w}} = \min_{h\in\mathcal{B}_1} \ \|h\|_{1,\mathrm{w}}. \tag{14}$$

(ii) If ω in (6) is chosen to satisfy $\gamma_\Phi \le \omega \le 1$, we must have

$$\|\widehat{h}_{T_0}\|_{1,\mathrm{w}} = \max_{T\in\mathbb{I}(k),\ h\in\mathcal{B}_1} \ \|h_T\|_{1,\mathrm{w}}. \tag{15}$$

(iii) For any weight vector w *that has the property of (15), the* w*-weighted NSP of order k holds provided that*

$$\|\widehat{h}_{T_0^c}\|_1 > \omega\|\widehat{h}_{T_0}\|_1. \tag{16}$$

Proof. (i) For any $0 \le \omega \le 1$ and for $h \in \mathcal{B}_1$, we have

$$\begin{aligned}
\|\widehat{h}\|_{1,\mathrm{w}} &= \omega\|\widehat{h}_{T_0}\|_1 + \|\widehat{h}_{T_0^C}\|_1 = \omega\|\widehat{h}_{T_0}\|_1 + 1 - \|\widehat{h}_{T_0}\|_1 \\
&\le (\omega-1)\|h_{T_0}\|_1 + 1 = (\omega-1)\|h_{T_0}\|_1 + \|h_{T_0}\|_1 + \|h_{T_0^c}\|_1 \\
&= \omega\|h_{T_0}\|_1 + \|h_{T_0^c}\|_1 = \|h\|_{1,\mathrm{w}},
\end{aligned}$$

where the first inequality used the optimality property in (8). This establishes (14).

(ii) There are two cases to consider. Case 1: $(T_0, \overline{h})$ is an optimal solution of the optimization problem in (15) for some $\overline{h}$. Case 2: T_0 is not an optimal set of (15). For Case 1, we have for any $h \in \mathcal{B}_1$,

$$\begin{aligned}
\|h_{T_0}\|_{1,\mathrm{w}} \leq \|\overline{h}_{T_0}\|_{1,\mathrm{w}} &= \max_{T\in\mathbb{I}(k),\ h\in\mathcal{B}_1} \|h_T\|_{1,\mathrm{w}} \\
&= \max_{h\in B_1} \|h_{T_0}\|_{1,\mathrm{w}} && \text{(because } T_0 \text{ is optimal)} \\
&= \omega \max_{h\in\mathcal{B}_1} \|h_{T_0}\|_1 && \text{(because of the definition of w)} \\
&= \omega\|\widehat{h}_{T_0}\|_1 && \text{(because of (8))} \\
&= \|\widehat{h}_{T_0}\|_{1,\mathrm{w}}.
\end{aligned}$$

For Case 2, for any $h \in \mathcal{B}_1$ and any $T \in \mathbb{I}(k) \setminus \{T_0\}$, one has

$$\begin{aligned}
\|\widehat{h}_{T_0}\|_{1,\mathrm{w}} &= \omega\|\widehat{h}_{T_0}\|_1 && \text{(by definition of w)} \\
&\geq \|\widetilde{h}_{T_1}\|_1 && \text{(because of } \omega \geq \gamma_\Phi) \\
&\geq \|h_T\|_1 && \text{(because of (12))} \\
&\geq \|h_T\|_{1,\mathrm{w}} && \text{(because of w} \leq 1).
\end{aligned}$$

For both cases, we established that $\|\widehat{h}_{T_0}\|_{1,\mathrm{w}}$ is an upper bound of the problem (15). Since $(T_0, \widehat{h})$ is feasible with respect to (15), $\|\widehat{h}_{T_0}\|_{1,\mathrm{w}}$ is actually the optimal objective of (15).

(iii) It follows from (16) that we have

$$\|\widehat{h}_{T_0^c}\|_{1,\mathrm{w}} = \|\widehat{h}_{T_0^c}\|_1 > \omega\|\widehat{h}_{T_0}\|_1 = \|\widehat{h}_{T_0}\|_{1,\mathrm{w}}. \tag{17}$$

For any $h \in \mathcal{B}_1$ any $T \in \mathbb{I}(k)$, we have the following chain of inequalities:

$$\begin{aligned}
\|h_{T^c}\|_{1,\mathrm{w}} &= \|h\|_{1,\mathrm{w}} - \|h_T\|_{1,\mathrm{w}} \\
&\geq \|\widehat{h}\|_{1,\mathrm{w}} - \|h_T\|_{1,\mathrm{w}} && \text{(because of (14))} \\
&= \|\widehat{h}_{T_0}\|_{1,\mathrm{w}} + \|\widehat{h}_{T_0^c}\|_{1,\mathrm{w}} - \|h_T\|_{1,\mathrm{w}} \\
&> 2\|\widehat{h}_{T_0}\|_{1,\mathrm{w}} - \|h_T\|_{1,\mathrm{w}} && \text{(because of (17))} \\
&\geq \|h_T\|_{1,\mathrm{w}}, && \text{(because of (15))}
\end{aligned}$$

This establishes the w-weighted NSP. □

We are ready to present our first main result, which is a sufficient condition in terms of δ_{tk} to ensure the sufficient condition (16) for the w-weighted NSP.

Theorem 3.2 *For any given ω in defining the* w*-weight (6) that has the property (15), in which T_0 is a solution of (8), if the matrix Φ satisfies*

$$\delta_{tk} < \delta_{\mathcal{N}}(t) := \sqrt{\frac{t-1}{t-(1-\omega^2)}} \tag{18}$$

for some $t > 1$, then each k-sparse solution $\hat{x}$ of the weighted ℓ_1 minimization (5) is the solution of (1)*.*

Proof. Since it is known that the w-weighted NSP of order k is necessary and sufficient for the w-weighted ℓ_1 minimization to produce the k-sparse solution of (1) [25], it is sufficient for us to prove that the condition (18) is sufficient for the condition (16) to hold in Lemma 3.1(iii).

For simplicity, we denote $h := \widehat{h}$ and it will not cause any confusion in this proof. Then, the condition (16) becomes

$$\|h_{T_0^c}\|_1 > \omega \|h_{T_0}\|_1.$$

Suppose on the contrary that we have

$$\|h_{T_0^c}\|_1 \leq \omega \|h_{T_0}\|_1. \tag{19}$$

We will derive a contradiction.

By setting $\beta := \|h_{T_0}\|_1 / k$, (19) becomes

$$\|h_{T_0^c}\|_1 \leq \omega k \beta.$$

The following proof technique is taken from [10]. We now decompose $h_{T_0^c}$ into two parts, $h_{T_0^c} = h^{(1)} + h^{(2)}$, where

$$h_i^{(1)} = \begin{cases} (h_{T_0^c})_i, & |(h_{T_0^c})_i| > \beta/a, \\ 0, & \text{otherwise}, \end{cases} \qquad h_i^{(2)} = \begin{cases} (h_{T_0^c})_i, & |(h_{T_0^c})_i| \leq \beta/a, \\ 0, & \text{otherwise}, \end{cases}$$

and $a > 0$ satisfies $\omega k a$ being an integer. Therefore $h^{(1)}$ is at most $\omega k a$-sparse as a result of the facts that $\|h^{(1)}\|_1 \leq \|h_{T_0^c}\|_1 \leq \omega k \beta$ and all non-zero entries of $h^{(1)}$ have magnitude larger than β/a. By letting $\|h^{(1)}\|_0 = m$, we have

$$\begin{aligned} &\|h^{(2)}\|_1 = \|h_{T_0^c}\|_1 - \|h^{(1)}\|_1 \leq [\omega k a - m]\, \beta/a, \\ &\|h^{(2)}\|_\infty \leq \beta/a. \end{aligned}$$

Applying Lemma 2.3 with $s := \omega k a - m$, $h^{(2)}$ can be expressed as a convex combination of sparse vectors, i.e.,

$$h^{(2)} = \sum_{i=1}^{N} \lambda_i u_i,$$

where u_i is s-sparse, $\|u_i\|_1 = \|h^{(2)}\|_1, \|u_i\|_\infty \leq \beta/a,\ i = 1, 2, \ldots, N$. Therefore, we have

$$\|u_i\|^2 \leq (\omega k a - m)\|u_i\|_\infty^2 \leq \frac{\omega k}{a}\beta^2 \leq \frac{\omega}{a}\|h_{T_0}\|^2 \leq \frac{\omega}{a}\|h_{T_0} + h^{(1)}\|^2, \tag{20}$$

where the third and last inequalities are respectively the consequences of $\|h_{T_0}\|_1 \leq \sqrt{k}\|h_{T_0}\|_2$ and the disjoint supports of h_{T_0} and $h^{(1)}$.

For any $\mu \geq 0$, denoting $\eta_i := h_{T_0} + h^{(1)} + \mu u_i$, we obtain

$$\sum_{j=1}^{N} \lambda_j \eta_j - \eta_i/2 = h_{T_0} + h^{(1)} + \mu h^{(2)} - \eta_i/2 = (1/2 - \mu)\left(h_{T_0} + h^{(1)}\right) - \mu u_i/2 + \mu h, \tag{21}$$

where $\eta_i, \sum_{i=1}^{N} \lambda_j \eta_j - \frac{1}{2}\eta_i - \mu h$ are all $(\omega k a + k)$-sparse vectors due to the facts: $\|h_{T_0}\|_0 \leq k$, $\|h^{(1)}\|_0 = m$ and $\|u_i\|_0 \leq \omega k a - m$. Since $\Phi h = 0$, together with (21), we have

$$\Phi\Big(\sum_{j=1}^{N} \lambda_j \eta_j - \frac{1}{2}\eta_i\Big) = \Phi\Big((\frac{1}{2} - \mu)(h_{T_0} + h^{(1)}) - \frac{1}{2}\mu u_i\Big).$$

It follows from [10, Eq.(25)] (setting $c = 1/2$ there), we have

$$\sum_{i=1}^{N} \lambda_i \Big\|\Phi(\sum_{j=1}^{N} \lambda_j \eta_j - \frac{1}{2}\eta_i)\Big\|^2 = \frac{1}{4}\sum_{i=1}^{N} \lambda_i \|\Phi \eta_i\|^2. \tag{22}$$

Let $\mu := \frac{\sqrt{(a+\omega)a}-a}{\omega} > 0$. Suppose it holds:

$$\delta := \delta_{\omega ka+k} < \sqrt{\frac{a}{a+\omega}}. \tag{23}$$

We now derive a contradiction from (22) with (23). We omit some computation in the following chain of inequalities.

$$\begin{aligned}
0 &= \sum_{i=1}^{N} \lambda_i \Big\| \Phi\Big((\frac{1}{2}-\mu)(h_{T_0}+h^{(1)}) - \frac{1}{2}\mu u_i \Big) \Big\|^2 - \frac{1}{4}\sum_{i=1}^{N} \lambda_i \|\Phi\eta_i\|^2 \\
&\leq (1+\delta)\sum_{i=1}^{N} \lambda_i \left[(\frac{1}{2}-\mu)^2 \|h_{T_0}+h^{(1)}\|^2 + \frac{\mu^2}{4}\|u_i\|^2 \right] - \frac{1-\delta}{4}\sum_{i=1}^{N} \lambda_i \left(\|h_{T_0}+h^{(1)}\|^2 + \mu^2\|u_i\|^2 \right) \\
&= \sum_{i=1}^{N} \lambda_i \left[\Big((1+\delta)(\frac{1}{2}-\mu)^2 - \frac{1-\delta}{4} \Big) \|h_{T_0}+h^{(1)}\|^2 + \frac{1}{2}\delta\mu^2\|u_i\|^2 \right] \\
&\leq \sum_{i=1}^{N} \lambda_i \|h_{T_0}+h^{(1)}\|^2 \left[\mu^2 - \mu + \delta\Big(\frac{1}{2} - \mu + (1+\frac{\omega}{2a})\mu^2 \Big) \right] \\
&= \|h_{T_0}+h^{(1)}\|^2 \left[\mu^2 - \mu + \delta\Big(\frac{1}{2} - \mu + (1+\frac{\omega}{2a})\mu^2 \Big) \right] \\
&= \|h_{T_0}+h^{(1)}\|^2 \Big(\frac{1}{2} - \mu + (1+\frac{\omega}{2a})\mu^2 \Big) \left[\delta - \sqrt{\frac{a}{a+\omega}} \right] \\
&< 0,
\end{aligned}$$

which is the contradiction we have desired. Here, the second inequality is derived from (20) and the last inequality holds because of (23) and the fact:

$$\frac{1}{2} - \mu + (1+\frac{\omega}{2a})\mu^2 = (1+\frac{\omega}{2a})(\mu - \frac{1}{2(1+\omega/2a)})^2 + \frac{1}{2} - \frac{1}{4(1+\omega/2a)} \geq \frac{1}{2} - \frac{1}{4(1+\omega/2a)} > 0.$$

By setting $a = \frac{t-1}{\omega}$ in condition (23), we obtain condition (18). This proves that condition (18) is a sufficient condition for the w-weighted NSP to hold. Hence the proof is completed. □

We note that the special choice $\omega = 1$, which is always feasible for (18), recovers the Cai-Zhang bound (3). The chance is that ω can take a much smaller value than 1 and hence the bound in Theorem 3.2 significantly improves its original version (3). We illustrate this point by revisiting Example 2.2. We already showed that there exist weights w such that the weighted ℓ_1 minimization can recover the ℓ_0 minimization solution. It is straightforward to calculate that

$$\begin{array}{ll} T_0 = \{3\}, & \widehat{h} = (3/14,\ 3/14,\ 4/7)^\top, \\ T_1 = \{1\} \text{ or } \{2\}, & \widetilde{h} = (3/14,\ 3/14,\ 4/7)^\top. \end{array}$$

Therefore, $\gamma_\Phi = 3/8$. Setting the weight w in (6) by choosing $\omega \geq \gamma_\Phi$ will ensure (15). The quantity $\delta_{\mathcal{N}}(2) = 0.9363$.

On the other hand, we can use the formula (see [27, 40])

$$\delta_{2k} = \max_{S \in \mathbb{I}(2k)} \|\Phi_S^\top \Phi_S - I_{2k}\|,$$

to calculate δ_2, where $\|\cdot\|$ denotes the spectral norm of a matrix. By setting $n = 3$ and $k = 1$, we obtain $\delta_2 = 0.9224$. Hence, we have $\delta_2 < \delta_{\mathcal{N}}(2)$. Theorem 3.2 guarantees that the weighted

ℓ_1 minimization will be able to recover the ℓ_0 solution of the problem in Example 2.2. As a matter of fact, any choice of $\omega \geq 3/8$ will be enough.

As highlighted in Introduction, the high order bound in (18) can go near 1 if ω is allowed to approach 0. The following table indicates this trend and can be used to compare to existing bounds of various high orders.

ω	δ_{2k}	δ_{3k}	δ_{4k}
1	$\sqrt{2}/2$	$\sqrt{6}/3$	$\sqrt{3}/2$
3/4	0.800	0.883	0.917
1/2	0.894	0.942	0.960
1/4	0.970	0.984	0.989

Table 1: Bounds on δ_{tk} of high order $t = 2, 3, 4$ base on formula (18)

4 A Reweighted ℓ_1 Minimization

The main remaining task is to identify T_0 in (8), which is as hard to solve as the original ℓ_0 minimization (1). Therefore, it has to be iteratively and numerically approximated and this naturally leads to the reweighted ℓ_1 minimization approach. In this section, we will propose a novel updating rule for the weight vector w. Our ultimate result is that the new rule will allow us eventually identify T_0 to be the true support of an optimal solution of the ℓ_1 minimization problem. To accomplish this task, we adopt the popular reweighted ℓ_1 approach to solve (2):

$$x^\ell := \arg\min_{x \in \mathbb{R}^n} \quad f(x; \mu_\ell, \mathrm{w}^\ell) := \frac{1}{2}\|\Phi x - b\|^2 + \mu_\ell \|x\|_{1,\mathrm{w}^\ell}, \tag{24}$$

where $\mu_\ell > 0$ is the penalty parameter and w^ℓ is the current weight vector at the ℓth iteration. One advantage of the approach in (24) is that the powerful `YALL1` solver (available at `http://yall1.blogs.rice.edu/`) can be called to solve it at each iteration.

We describe the rule and the resulting algorithm in the first subsection below. Convergence analysis of reweighted ℓ_1 algorithms is never going to be trivial and there seems to be no standard scheme to follow, as well demonstrated in Zhao and Li [39]. Our main convergence result relies on a few very technical lemmas concerning the new updating rule. We report those results in the second subsection.

4.1 The Rule for Updating the Weight Vector and the Algorithm

Let $x^\ell, x^{\ell-1}$ be the two iterates already computed. Let k_ℓ be the size of the support of x^ℓ (i.e., $k_\ell := |\mathrm{supp}(x^\ell)|$). Let $h^\ell := x^\ell - x^{\ell-1}$ and sort h^ℓ in decreasing order in terms of absolute values of its elements:

$$|h^\ell_{j^\ell_1}| \geq |h^\ell_{j^\ell_2}| \geq \cdots \geq |h^\ell_{j^\ell_{k_\ell}}| \geq |h^\ell_{j^\ell_{k_\ell+1}}| \geq \cdots \geq |h^\ell_{j_n}|. \tag{25}$$

We further define two index sets that respectively corresponds to T_0 in (8) and T_1 in (12):

$$T_0^\ell := \{j_1^\ell, j_2^\ell, \cdots, j_{k_\ell}^\ell\}, \qquad T_1^\ell := \{j_2^\ell, j_3^\ell, \cdots, j_{k_\ell+1}^\ell\}. \tag{26}$$

We define γ_Φ^ℓ below to mimic the quantity γ_Φ in (13), and γ_ℓ in case that γ_Φ^ℓ is too small:

$$\gamma_\Phi^\ell := \frac{\|h^\ell_{T_1^\ell}\|_1}{\|h^\ell_{T_0^\ell}\|_1} \qquad \text{and} \qquad \gamma_\ell := \epsilon_1 + \theta_\ell \gamma_\Phi^\ell,$$

where $\epsilon_1 < 1$ is a small positive number and $\theta_\ell = \alpha_1\theta_{\ell-1}$ with $\alpha_1 > 1$ and $\theta_0 > 0$ being given. The weight vector w is updated according to the following rule:

$$\mathrm{w}_i^{\ell+1} = \begin{cases} \gamma_\ell, & \text{if } i \in T_0^\ell, \\ \epsilon_1 + \theta_\ell + \frac{1}{|x_i^\ell|+\epsilon_2}, & \text{if } i \notin T_0^\ell \end{cases} \tag{27}$$

where $0 < \epsilon_2 \le 1$. We have two brief comments on (27).

(i) The quantity $\gamma_\Phi^\ell \le 1$ is well defined. Its denominator $\|h_{T_0^\ell}^\ell\|$ can only be *zero* when $h^\ell = 0$. Our algorithm would have already terminated if $\|h^\ell\|$ is small enough. The weighting scheme in (27) is actually a scaled version of what we used in (6). In fact, the sequence $\{\theta_\ell\}$ goes to ∞ and hence we have

$$\frac{\mathrm{w}^{\ell+1}}{\theta_\ell} = \begin{cases} \gamma_\Phi^\ell + \frac{\epsilon_1}{\theta_\ell}, & \text{if } i \in T_0^\ell, \\ 1 + \frac{\epsilon_1}{\theta_\ell} + \frac{1}{\theta_\ell(|x_i^\ell|+\epsilon_2)}, & \text{if } i \notin T_0^\ell \end{cases} \approx \begin{cases} \gamma_\Phi^\ell, & \text{if } i \in T_0^\ell, \\ 1, & \text{if } i \notin T_0^\ell. \end{cases}$$

T_0^ℓ will be proved to be unchanged for all sufficiently large ℓ. In addition, the weight sequence $\{\mathrm{w}^\ell\}$ are bounded from below by ϵ_1.

(ii) The updating rule for the indices not in T_0^ℓ drew experiences from [14], where $\mathrm{w}_i^{\ell+1} = 1/(|x_i^\ell| + \epsilon)$ ($\epsilon > 0$ is small) is used and from [19], where $\mathrm{w}_i^{\ell+1} = 1/(|x_i^\ell| + \epsilon)^{1-q}$, $q \in (0, 1)$ is used. The key feature of ours is that we update w based on the two consecutive iterates rather than just using the current iterate. We like to point it out that we chose the difference vector between the two consecutive iterates to approximate the null-space. This is the cheapest approximation as the iterates are already computed. It remains to be explored if there are other heuristic (better and cheaper) approximation to the null space.

The overall algorithm is stated in Table 2. We note that the refinement step in Step 3 will be justified by our main result Thm. 4.4(ii). To distinguish the method from the many existing ones, we call it the Modified Iterative Reweighted ℓ_1 Minimization `MIRL`$_1$ and we state its convergence analysis in next subsection.

Table 2: The framework of MIRL$_1$.

Modified Iterative Reweighted ℓ_1 Minimization (MIRL$_1$)
Step 0 Initialize $x^0, \omega^1, \mu_1, \theta_0 > 0, 1 > \alpha_0 > 0, \alpha_1 > 1$. `tol` is the given tolerance. Set $t \Leftarrow 1$;
Step 1 Compute $$x^\ell \in \mathrm{argmin}_x\ f(x;\ \mu_\ell,\ \mathrm{w}^\ell);$$
Step 2 If $\frac{\|x^\ell - x^{\ell-1}\|}{\max\{1, \|x^{\ell-1}\|\}} \le$ `tol`, go to **Step 3**. Otherwise, update $\mathrm{w}^{\ell+1}$ from $x^\ell, x^{\ell-1}$ by (27) and $\mu_{\ell+1}$ by $$\mu_{\ell+1} = \begin{cases} \alpha_0\mu_\ell, & \text{if } \beta_\ell > 1 \\ \beta_\ell\mu_\ell, & \text{if } \beta_\ell \le 1 \end{cases} \qquad \text{with} \quad \beta_\ell = \frac{\|x^\ell\|_{1,\mathrm{w}^\ell}}{\|x^\ell\|_{1,\mathrm{w}^{\ell+1}}}$$ set $\ell \Leftarrow \ell + 1$, go to **Step 1**.
Step 3 Refine x^ℓ to output $\hat{x}$ by $\hat{x}_\Gamma \in \arg\min_{y\in\mathbb{R}^{\lvert\Gamma\rvert}} \|\Phi_\Gamma y - b\|, \hat{x}_{\Gamma^c} = 0$, where $\Gamma = \mathrm{supp}(x^\ell)$.

4.2 Convergence Analysis

In this part, we will first establish a few technical results and then state our main result. Let $\{x^\ell\}$ be the sequence generated by our $\texttt{MIRL}_1$ method. Since $x^{\ell+1} = \arg\min_x\ f(x, \mu_{\ell+1}, \mathrm{w}^{\ell+1})$, we have the first-order optimality condition involving the ℓ_1 norm

$$\Phi^\top(\Phi x^{\ell+1} - b) + \mu_{\ell+1}\Big(\mathrm{w}^{\ell+1} \circ \mathrm{v}^{\ell+1}\Big) = 0 \ \text{ and } \ \mathrm{v}_i^{\ell+1} = \begin{cases} 1 & x_i^{\ell+1} > 0 \\ -1 & x_i^{\ell+1} < 0 \\ \beta & x_i^{\ell+1} = 0,\ \beta \in [-1, 1]. \end{cases} \tag{28}$$

In fact, the vector $\mathrm{v}^{\ell+1}$ is a subgradient of the ℓ_1 norm $\|x\|_1$ at point $x^{\ell+1}$. In all, we have the following sequences to consider: $\{x^\ell,\ \mu_\ell,\ \mathrm{w}^\ell,\ \mathrm{v}^\ell\}$. Our first result shows that the function values $\{f(x^\ell;\ \mu_\ell,\ \mathrm{w}^\ell)\}$ are decreasing.

Proposition 4.1 *Let the sequence $\{x^\ell,\ \mu_\ell,\ \mathrm{w}^\ell\}$ be generated by $\texttt{MIRL}_1$. The sequence enjoys the following properties.*

(i) It holds

$$\mu_\ell\|x^\ell\|_{1,\mathrm{w}^\ell} \geq \mu_{\ell+1}\|x^\ell\|_{1,\mathrm{w}^{\ell+1}}$$

(ii) Denote $f(x^\ell,\ \mu_\ell,\ \mathrm{w}^\ell)$ by f_ℓ. We then have

$$f_\ell - f_{\ell+1} \geq \frac{1}{2}\|\Phi x^{\ell+1} - \Phi x^\ell\|^2 + \mu_\ell\|x^\ell\|_{1,\mathrm{w}^\ell} - \mu_{\ell+1}\|x^\ell\|_{1,\mathrm{w}^{\ell+1}}.$$

Therefore, by the property in (i), the sequence $\{f_\ell\}$ is decreasing.

Proof. (i) It follows from the definition of $\mu_{\ell+1}$ and β_ℓ in Step 2 of $\texttt{MIRL}_1$ that

$$\begin{aligned}
\mu_\ell\|x^\ell\|_{1,\mathrm{w}^\ell} - \mu_{\ell+1}\|x^\ell\|_{1,\mathrm{w}^{\ell+1}} &= \begin{cases} \mu_\ell\|x^\ell\|_{1,\mathrm{w}^\ell} - \alpha_0\mu_\ell\|x^\ell\|_{1,\mathrm{w}^{\ell+1}}, & \text{if } \ \beta_\ell > 1 \\ \mu_\ell\|x^\ell\|_{1,\mathrm{w}^\ell} - \beta_\ell\mu_\ell\|x^\ell\|_{1,\mathrm{w}^{\ell+1}}, & \text{if } \ \beta_\ell \leq 1 \end{cases} \\
&\geq \begin{cases} \mu_\ell\|x^\ell\|_{1,\mathrm{w}^\ell} - \alpha_0\mu_\ell\|x^\ell\|_{1,\mathrm{w}^\ell}, & \text{if } \ \beta_\ell > 1 \\ \mu_\ell\|x^\ell\|_{1,\mathrm{w}^\ell} - \mu_\ell\|x^\ell\|_{1,\mathrm{w}^\ell}, & \text{if } \ \beta_\ell \leq 1 \end{cases} \\
&\geq 0.
\end{aligned}$$

(ii) Direct calculation yields

$$\begin{aligned}
& f_\ell - f_{\ell+1} \\
= \ & \frac{1}{2}\|\Phi x^{\ell+1} - \Phi x^\ell\|^2 + \langle x^\ell - x^{\ell+1}, \Phi^\top(\Phi x^{\ell+1} - b)\rangle + \mu_\ell\|x^\ell\|_{1,\mathrm{w}^\ell} - \mu_{\ell+1}\|x^{\ell+1}\|_{1,\mathrm{w}^{\ell+1}} \\
= \ & \frac{1}{2}\|\Phi x^{\ell+1} - \Phi x^\ell\|^2 + \langle x^{\ell+1} - x^\ell, \mu_{\ell+1}\mathrm{w}^{\ell+1} \circ \mathrm{v}^{\ell+1}\rangle + \mu_\ell\|x^\ell\|_{1,\mathrm{w}^\ell} - \mu_{\ell+1}\|x^{\ell+1}\|_{1,\mathrm{w}^{\ell+1}} \qquad (29) \\
\geq \ & \frac{1}{2}\|\Phi x^{\ell+1} - \Phi x^\ell\|^2 + \mu_{\ell+1}\|x^{\ell+1}\|_{1,\mathrm{w}^{\ell+1}} - \mu_{\ell+1}\|x^\ell\|_{1,\mathrm{w}^{\ell+1}} + \mu_\ell\|x^\ell\|_{1,\mathrm{w}^\ell} - \mu_{\ell+1}\|x^{\ell+1}\|_{1,\mathrm{w}^{\ell+1}} \\
= \ & \frac{1}{2}\|\Phi x^{\ell+1} - \Phi x^\ell\|^2 + \mu_\ell\|x^\ell\|_{1,\mathrm{w}^\ell} - \mu_{\ell+1}\|x^\ell\|_{1,\mathrm{w}^{\ell+1}}, \qquad (30)
\end{aligned}$$

where the second equality is from (28), and the first inequality is due to $\mathrm{v}_i^{\ell+1}x_i^{\ell+1} = |x_i^{\ell+1}|$ and $|\mathrm{v}_i^{\ell+1}| \leq 1$. $\square$

By the updating rule for μ_ℓ in Step 2 of $\texttt{MIRL}_1$ algorithm, it is easy to see that the sequence $\{\mu_\ell\}$ is decreasing and is bounded from below by 0. Hence it converges. We let

$$\underline{\mu} := \lim_{\ell\to\infty} \mu_\ell \geq 0$$

The case that $\underline{\mu} = 0$ is not desired as it essentially implies that the reweighted ℓ_1 optimization (24) will eventually lead to the least-square solution without the ℓ_1 regularization term (i.e., this term eventually vanishes). Consequently, no sparse solutions would be guaranteed. Unfortunately, we were unable to ensure this case not happening through the update scheme of μ_ℓ. Therefore, we assume $\underline{\mu} > 0$. We also assume each subproblem (24) always finds a sparse iterate that satisfies

$$\tau := \inf_{\ell\geq 1}\left\{ \min_{i\in \mathrm{supp}(x^\ell)} |x_i^\ell| \right\} > 0. \tag{31}$$

This technical assumption will greatly simplify the techniques used to derive our main result. We note that some of the results are still true without this assumption. However, the technical gain without it is not significant.

The following technical result first establishes the boundedness of the sequence $\{x^\ell\}$. It then establishes the useful fact that for any two consecutive iterates of the subgradient vectors $\{\mathrm{v}^\ell,\ \mathrm{v}^{\ell+1}\}$ and the weight vectors $\{\mathrm{w}^\ell,\ \mathrm{w}^{\ell+1}\}$, the change over the support set of the iterate x^ℓ will be vanishing as ℓ goes to ∞.

Lemma 4.2 *We have the following technical results:*

(i) The sequence $\{x^\ell\}$ is bounded.

(ii) For any $i^\ell \in supp(x^\ell)$, we have

$$\lim_{\ell\to\infty} |\mathrm{v}_{i^\ell}^\ell - \mathrm{v}_{i^\ell}^{\ell+1}| = 0 \qquad \textit{and} \qquad \lim_{\ell\to\infty} |\mathrm{w}_{i^\ell}^\ell - \mathrm{w}_{i^\ell}^{\ell+1}| = 0.$$

Proof. (i) Since $\underline{\mu} > 0$, and $\mu_\ell \geq \underline{\mu}$, $\mathrm{w}^\ell \geq \epsilon_1$ for all $\ell = 1, 2, \ldots$, we have

$$\underline{\mu}\epsilon_1 \|x^\ell\|_1 \leq \mu_\ell \|x^\ell\|_{1,\mathrm{w}^\ell} \leq \frac{1}{2}\|\Phi x^\ell - b\|^2 + \mu_\ell\|x^\ell\|_{1,\mathrm{w}^\ell} = f(x^\ell, \mu_\ell, \mathrm{w}^\ell) \leq f(0, \mu_\ell, \mathrm{w}^\ell) = \frac{1}{2}\|b\|^2.$$

Therefore, the sequence $\{x^\ell\}$ is bounded.

(ii) It follows from Prop. 4.1(ii) that $\{f_\ell\}$ is monotonically decreasing and bounded below ($f_\ell \geq 0$). Thus it converges and we have

$$\lim_{\ell\to\infty} \quad \|\Phi x^{\ell+1} - \Phi x^\ell\| = 0, \tag{32}$$

$$\lim_{\ell\to\infty} \quad \mu_\ell\|\mathrm{w}^\ell \circ x^\ell\|_1 - \mu_{\ell+1}\|\mathrm{w}^{\ell+1}\circ x^\ell\|_1 = 0. \tag{33}$$

It then follows from (28) and (32) that $\lim_{\ell\to\infty}(\mu_\tau \mathrm{w}^\ell \circ \mathrm{v}^\ell - \mu_{\tau+1}\mathrm{w}^{\ell+1}\circ \mathrm{v}^{\ell+1}) = \lim_{\ell\to\infty}\Phi^\top(\Phi x^{\ell+1} - \Phi x^\ell) = 0$, which combining $\lim_{\ell\to\infty}\mu_{\ell+1} = \lim_{\ell\to\infty}\mu_\ell = \underline{\mu} > 0$ yields

$$\lim_{\ell\to\infty} \mathrm{w}^\ell \circ \mathrm{v}^\ell = \lim_{\ell\to\infty} \mathrm{w}^{\ell+1}\circ \mathrm{v}^{\ell+1}. \tag{34}$$

By (29) we have

$$\begin{aligned} f_\ell - f_{\ell+1} \quad = \quad & \frac{1}{2}\|\Phi x^{\ell+1} - \Phi x^\ell\|^2 + \mu_\ell\|\mathrm{w}^\ell\circ x^\ell\|_1 - \mu_{\ell+1}\|\mathrm{w}^{\ell+1}\circ x^\ell\|_1 + \\ & \underbrace{\left\langle x^{\ell+1} - x^\ell, \mu_{\ell+1}\mathrm{w}^{\ell+1}\circ \mathrm{v}^{\ell+1}\right\rangle + \mu_{\ell+1}\|\mathrm{w}^{\ell+1}\circ x^\ell\|_1 - \mu_{\ell+1}\|\mathrm{w}^{\ell+1}\circ x^{\ell+1}\|_1}_{:=\Theta}. \end{aligned} \tag{35}$$

which further implies

$$\begin{aligned}\Theta = \mu_{\ell+1}\sum_{i=1}^n \mathrm{w}_i^{\ell+1}\left(|x_i^\ell| - |x_i^{\ell+1}| - \mathrm{v}_i^{\ell+1}x_i^\ell + \mathrm{v}_i^{\ell+1}x_i^{\ell+1}\right) &= \mu_{\ell+1}\sum_{i=1}^n \mathrm{w}_i^{\ell+1}\left(|x_i^\ell| - \mathrm{v}_i^{\ell+1}x_i^\ell\right)\\ &\geq \epsilon_1\underline{\mu}\sum_{i=1}^n\left(|x_i^\ell| - \mathrm{v}_i^{\ell+1}x_i^\ell\right) \geq 0,\end{aligned}$$

where the last inequality is due to $|\mathrm{v}_i^{\ell+1}| \leq 1, i = 1,\ldots,n$. Now taking into account of the limits in (32)-(35) and $\lim_{\ell\to\infty} f_\ell - f_{\ell+1} = 0$, we derive that

$$\lim_{\ell\to\infty} |x_i^\ell| - \mathrm{v}_i^{\ell+1}x_i^\ell = 0, \quad i = 1, 2, \ldots, n,$$

which indicates that

$$\lim_{\ell\to\infty} |\mathrm{v}_{i^\ell}^\ell - \mathrm{v}_{i^\ell}^{\ell+1}||x_{i^\ell}^\ell| = \lim_{\ell\to\infty} ||x_{i^\ell}^\ell| - \mathrm{v}_{i^\ell}^{\ell+1}x_{i^\ell}^\ell| = 0, \quad \forall\ i^\ell \in \mathrm{supp}(x^\ell), \tag{36}$$

Since Assumption (31) holds, we must have $\lim_{\ell\to\infty}|\mathrm{v}_{i^\ell}^\ell - \mathrm{v}_{i^\ell}^{\ell+1}| = 0,\ \forall\ i^\ell \in \mathrm{supp}(x^\ell)$, which combining (34) yields $\lim_{\ell\to\infty}|\mathrm{w}_{i^\ell}^\ell - \mathrm{w}_{i^\ell}^{\ell+1}| = 0,\ \forall\ i^\ell \in \mathrm{supp}(x^\ell)$. This completes our proof. □

The next result says that the support set at x^ℓ is identified by our index set T_0^ℓ for sufficiently large ℓ.

Lemma 4.3 *The following results hold.*

(i) For all sufficiently large index ℓ, we have supp$(x^\ell) = T_0^\ell$.

(ii) $\lim_{\ell\to\infty} \|x^\ell - x^{\ell+1}\| = 0$.

Proof. (i) We are going to establish three facts in order to prove the claim in (i). They are: For sufficiently large ℓ, we must have

$$\mathrm{supp}(x^\ell) \subseteq T_0^{\ell-1} \qquad \text{and} \qquad \mathrm{supp}(x^\ell) \subseteq T_0^\ell,$$

and the whole sequence $\{\gamma_\ell\}$ is bounded. We proved one by one.

Suppose there exists a subsequence $\{\ell_j\}$ such that $\mathrm{supp}(x^{\ell_j}) \not\subseteq T_0^{\ell_j-1}$. There exists a sub-subsequence of $\{\ell_j\}$ (for simplicity, we still denote it as $\{\ell_j\}$) such that there exists an infinite index sequence $\{i^{\ell_j}\}$ such that $i^{\ell_j} \in \mathrm{supp}(x^{\ell_j})$, but $i^{\ell_j} \notin T_0^{\ell_j-1}$. For notational simplicity, denote $i := i^{\ell_j}$. According to the update rule (27), we have

$$\mathrm{w}_i^{\ell_j} = \epsilon_1 + \theta_{\ell_j-1} + \frac{1}{|x_i^{\ell_j-1}| + \epsilon_2} \ \to\ \infty \qquad (\text{because } \theta_{\ell_j-1} \to \infty).$$

This leads to the following contradiction:

$$\infty = \lim_{\ell_j\to\infty} \underline{\mu}\tau\mathrm{w}_i^{\ell_j} \leq \lim_{\ell_j\to\infty} \mu_{\ell_j}\|\mathrm{w}^{\ell_j} \circ x^{\ell_j}\| \leq \lim_{\ell_j\to\infty} f(x^{\ell_j}, \mu_{\ell_j}, \mathrm{w}^{\ell_j}) \leq \frac{1}{2}\|b\|^2. \tag{37}$$

Therefore, we must have

$$\mathrm{supp}(x^\ell) \subseteq T_0^{\ell-1} \tag{38}$$

for sufficiently large ℓ.

We now prove the sequence $\{\gamma_\ell\}$ is bounded. Because of (38) and by the update rule (27), we have for $i^\ell \in \mathrm{supp}(x^\ell)$, $\mathrm{w}_{i^\ell}^\ell = \gamma_{\ell-1}$ for all ℓ sufficiently large. Same reasoning as in (37), we can claim that the sequence $\{\gamma_{\ell-1}\}$ is bounded, which also means the boundedness of $\{\gamma_\ell\}$.

Finally, we suppose that there exits a subsequence $\{\ell_j\}$ such that $\mathrm{supp}(x^{\ell_j}) \not\subseteq T_0^{\ell_j}$. As reasoning in the first case, there exists an infinite index sequence $\{i^{\ell_j}\}$ such that $i^{\ell_j} \in \mathrm{supp}(x^{\ell_j})$, but not in $T_0^{\ell_j}$. Similarly, denote $i := i^{\ell_j}$. According to the update rule (27), we have

$$\mathrm{w}_i^{\ell_j+1} = \epsilon_1 + \theta_{\ell_j} + \frac{1}{|x_i^{\ell_j}| + \epsilon_2} \ \to \ \infty \qquad (\text{because } \theta_{\ell_j} \to \infty).$$

In the meantime, because of the established fact (38), we have $i \in T_0^{\ell_j-1}$, hence $\mathrm{w}_i^{\ell_j} = \gamma_{\ell_j-1}$, which is bounded. Therefore

$$\left(\mathrm{w}_i^{\ell_j+1} - \mathrm{w}_i^{\ell_j}\right) \ \to \ \infty,$$

which contradicts Lemma 4.2(ii) on the index set $\mathrm{supp}(x^\ell)$. Hence, we must have

$$\mathrm{supp}(x^\ell) \subseteq T_0^\ell \tag{39}$$

for sufficiently large ℓ.

By the definition of T_0^ℓ, its size equals the size of $\mathrm{supp}(x^\ell)$. It follows from (39), we must have $\mathrm{supp}(x^\ell) = T_0^\ell$ for sufficiently large ℓ.

(ii) We note that $\gamma_\ell = \epsilon_1 + \theta_\ell \gamma_\Phi^\ell$. Since $\{\gamma_\ell\}$ is bounded and $\theta_\ell \to \infty$, we must have $\gamma_\Phi^\ell \to 0$ as $\ell \to \infty$. By the definition of γ_Φ^ℓ, we can only have the following cases. Case 1: $h^\ell \to 0$; and Case 2:

$$h_0 := \liminf_{\ell\to\infty} |h_{j_1^\ell}^\ell| > 0 \quad \text{and} \quad \lim_{\ell\to\infty} h_i^\ell = 0, \quad \forall\, i \neq j_1^\ell,$$

where we use j_1 to denote the index of the first largest absolute value of the elements in h^ℓ (Note: j_1 should depend on ℓ, for simplicity we drop the dependence). Case 1 is what the result claims. Now we prove that Case 2 is impossible.

Performing the inner product on both sides of (28) with $x^{\ell+1}$ and using the property of $\mathrm{v}^{\ell+1}$, we have

$$\langle \Phi x^{\ell+1},\ \Phi x^{\ell+1} - b\rangle + \mu_{\ell+1} \|\mathrm{w}^{\ell+1} \circ x^{\ell+1}\|_1 = 0.$$

We then have

$$\begin{aligned} & \mu_\ell \|\mathrm{w}^\ell \circ x^\ell\|_1 - \mu_{\ell-1}\|\mathrm{w}^{\ell-1} \circ x^{\ell-1}\|_1 \\ = & -\langle \Phi x^\ell,\ \Phi x^\ell - b\rangle + \langle \Phi x^{\ell-1},\ \Phi x^{\ell-1} - b\rangle \\ = & \langle \Phi x^{\ell-1} - \Phi x^\ell, \Phi x^{\ell-1} + \Phi x^\ell\rangle + \langle \Phi x^\ell - \Phi x^{\ell-1},\ b\rangle. \end{aligned}$$

The limit in (32), the boundedness of $\{x^\ell\}$ (proved in Lemma 4.2(i)), and $\lim_{\ell\to\infty} \mu_\ell = \underline{\mu}$ lead to

$$0 = \lim_{\ell\to\infty} \left(\mu_\ell \|\mathrm{w}^\ell \circ x^\ell\|_1 - \mu_{\ell-1}\|\mathrm{w}^{\ell-1} \circ x^{\ell-1}\|_1\right) = \underline{\mu} \lim_{\ell\to\infty} \left(\|\mathrm{w}^\ell \circ x^\ell\|_1 - \|\mathrm{w}^{\ell-1} \circ x^{\ell-1}\|_1\right).$$

Since $\underline{\mu} > 0$,

$$\lim_{\ell\to\infty} \ \|\mathrm{w}^\ell \circ x^\ell\|_1 - \|\mathrm{w}^{\ell-1} \circ x^{\ell-1}\|_1 = 0. \tag{40}$$

We note that we have established in (i) that $\mathrm{supp}(x^\ell) = T_0^\ell$ and in (39) that $\mathrm{supp}(x^\ell) \subseteq T_0^{\ell-1}$ for all sufficiently large ℓ. We therefore have

$$\mathrm{supp}(x^\ell) = T_0^\ell \subseteq \mathrm{supp}(x^{\ell-1}) = T_0^{\ell-1} \subseteq \mathrm{supp}(x^{\ell-2}) = T_0^{\ell-2}$$

for all sufficiently large ℓ. Since j_1^ℓ is the first index in T_0^ℓ we must have

$$j_1^\ell \in T_0^{\ell-1} \subseteq T_0^{\ell-2}$$

for all sufficiently large ℓ. Those facts will be used in the following proof, where for simplicity we denote $j_1 := j_1^\ell$ and $T := \mathrm{supp}(x^{\ell-1})$. It follows that

$$\begin{aligned}
& \lim_{\ell\to\infty} \left| \|\mathrm{w}^\ell \circ x^\ell\|_1 - \|\mathrm{w}^{\ell-1} \circ x^{\ell-1}\|_1 \right| \\
= \;& \lim_{\ell\to\infty} \left| \mathrm{w}_{j_1}^\ell |x_{j_1}^\ell| - \mathrm{w}_{j_1}^{\ell-1}|x_{j_1}^{\ell-1}| + \sum_{i\neq j_1} (\mathrm{w}_i^\ell |x_i^\ell| - \mathrm{w}_i^{\ell-1}|x_i^{\ell-1}|) \right| \\
\geq \;& \lim_{\ell\to\infty} \left(\left| \mathrm{w}_{j_1}^\ell |x_{j_1}^\ell| - \mathrm{w}_{j_1}^{\ell-1}|x_{j_1}^{\ell-1}| \right| - \sum_{i(\neq j_1)\in T} \left| \mathrm{w}_i^\ell |x_i^\ell| - \mathrm{w}_i^{\ell-1}|x_i^{\ell-1}| \right| - \sum_{i\notin T} \mathrm{w}_i^\ell |x_i^\ell| \right) \\
= \;& \lim_{\ell\to\infty} \left(|\gamma_{\ell-1} x_{j_1}^\ell - \gamma_{\ell-2} x_{j_1}^{\ell-1}| - \sum_{i(\neq j_1)\in T} |\gamma_{\ell-1} x_i^\ell - \gamma_{\ell-2} x_i^{\ell-1}| - \sum_{i\notin T} \mathrm{w}_i^\ell |x_i^\ell| \right) \\
\geq \;& \lim_{\ell\to\infty} \left(|\gamma_{\ell-1} h_{j_1}^\ell| - |(\gamma_{\ell-1} - \gamma_{\ell-2}) x_{j_1}^{\ell-1}| - \sum_{i(\neq j_1)\in T} |\gamma_{\ell-1} x_i^\ell - \gamma_{\ell-2} x_i^{\ell-1}| - \sum_{i\notin T} \mathrm{w}_i^\ell |x_i^\ell - x_i^{\ell-1}| \right) \\
= \;& \lim_{\ell\to\infty} \left| \gamma_{\ell-1} h_{j_1}^\ell \right| \geq \epsilon_1 |h_0| > 0,
\end{aligned}$$

which contradicts the established fact (40). Here, the second equality follows by the facts that for sufficiently large ℓ and all $i \in T$, it holds $i \in T = T_0^{\ell-1}$, $\mathrm{w}_i^\ell = \gamma_{\ell-1}$, and $i \in T \subseteq T_0^{\ell-2}$, $\mathrm{w}_i^{\ell-1} = \gamma_{\ell-2}$ from the definition of w^ℓ, and $\lim_{\ell\to\infty} |\mathrm{v}_i^{\ell-1} - \mathrm{v}_i^\ell| = 0$; the last equality follows from $\lim_{\ell\to\infty} \gamma_{\ell-1} - \gamma_{\ell-2} = 0$ and $\lim_{\ell\to\infty} x_i^\ell - x_i^{\ell-1} = \lim_{\ell\to\infty} h_i^\ell = 0$, for $i \neq j_1^\ell$. This contradiction means that Case 1 holds, which is the claim in the result. $\square$

We can even make a stronger claim based on the result in Lemma 4.3(ii). Assume that x^∞ is an isolated limit of $\{x^\ell\}$. It then follows from Moré and Sorensen [28] (also see [24, Prop. 7]) that the whole sequence $\{x^\ell\}$ converges to x^∞. The following result, which would be much easier to prove for x^∞ being isolated, concerns any accumulation point of the whole sequence.

Theorem 4.4 *Suppose the condition (31) holds and $\underline{\mu} > 0$. Then we have the following results.*

(i) There exists a support set $\Gamma \subset \{1, 2, \cdots, n\}$ such that

$$\mathrm{supp}(x^\ell) = T_0^\ell = \Gamma \qquad \textit{for all sufficiently large } \ell.$$

(ii) Let $\{x^{\ell_j}\}$ be any converging subsequence with its limit denoted as x^∞. Then $x^\infty = (x_\Gamma^\infty;\ x_{\Gamma^c}^\infty)$ with

$$x_\Gamma^\infty = \operatorname*{argmin}_{y\in\mathbb{R}^{|\Gamma|}} \frac{1}{2}\|\Phi_\Gamma y - b\|^2 + \gamma^\infty \underline{\mu} \|y\|_1 \qquad \textit{and} \qquad x_{\Gamma^c}^\infty = 0,$$

where Φ_Γ consists of columns in Φ indexed by Γ, and

$$\gamma^\infty = \lim_{\ell_j\to\infty} \gamma_{\ell_j} = \frac{\|\Phi_\Gamma^\top (\Phi x^\infty - b)\|_1}{\underline{\mu}|\Gamma|}.$$

Proof. (i) It is sufficient to prove that $\mathrm{supp}(x^\ell) = \mathrm{supp}(x^{\ell+1})$ for all sufficiently large ℓ. Suppose it is not true, then there exists a subsequence $\{\ell_j\}$ such that $\mathrm{supp}(x^{\ell_j}) \neq \mathrm{supp}(x^{\ell_j+1})$.

Without loss of generality, we assume that there exists an infinite index sequence $\{i^{\ell_j}\}$ such that $i^{\ell_j} \in \mathrm{supp}(x^{\ell_j})$, but not in $\mathrm{supp}(x^{\ell_j+1})$. With Lemma 4.3(ii), we obtain that

$$0 = \lim_{\ell_j \to \infty} \|x^{\ell_j} - x^{\ell_j+1}\| \geq \lim_{\ell_j \to \infty} |x^{\ell_j}_{i^{\ell_j}} - x^{\ell_j+1}_{i^{\ell_j}}| = \lim_{\ell_j \to \infty} |x^{\ell_j}_{i^{\ell_j}}| \geq \tau > 0.$$

This contradiction establishes (i), noting the fact $\mathrm{supp}(x^\ell) = T_0^\ell$ that has been established in Lemma 4.3(i).

(ii) In view of the result in (i), the weight vector w in (27) is updated by (when ℓ is sufficiently large)

$$\mathrm{w}_i^{\ell+1} = \begin{cases} \gamma_\ell, & i \in \Gamma \\ \epsilon_1 + \theta_\ell + \frac{1}{\epsilon_2}, & i \in \Gamma^c. \end{cases} \tag{41}$$

The first order optimality condition (28) at $x^{\ell+1}$ with the additional information that $\mathrm{supp}(x^{\ell+1}) = \Gamma$ implies

$$\begin{aligned} 0 &= \Phi_\Gamma^\top(\Phi_\Gamma x_\Gamma^{\ell+1} - b) + \mu_{\ell+1}\gamma_\ell \mathrm{v}_\Gamma^{\ell+1} \\ 0 &= \Phi_{\Gamma^c}^\top(\Phi_\Gamma x_\Gamma^{\ell+1} - b) + \mu_{\ell+1}\mathrm{w}_{\Gamma^c}^{\ell+1} \circ \mathrm{v}_{\Gamma^c}^{\ell+1} \\ 0 &= x_{\Gamma^c}^{\ell+1}, \end{aligned} \tag{42}$$

where (and hereafter) $x_\Gamma^{\ell+1}$ denotes the subvector of $x^{\ell+1}$ indexed by Γ. By (41), $\mathrm{w}_i^{\ell+1} \to \infty$, for $\forall\, i \in \Gamma^c$, which implies $\mathrm{v}_i^{\ell+1} \to 0$, $\forall\, i \in \Gamma^c$.

Now we compute the convergence on the subsequence $\{x^{\ell_j}\}$. From the first equation in (42) for sufficient large ℓ_j, we have

$$\gamma_{\ell_j} = |\gamma_{\ell_j}| = \frac{\|\Phi_\Gamma^\top(\Phi_\Gamma x_\Gamma^{\ell_j+1} - b)\|_1}{\mu_{\ell_j+1}\|\mathrm{v}_\Gamma^{\ell_j+1}\|_1} = \frac{\|\Phi_\Gamma^\top(\Phi_\Gamma x_\Gamma^{\ell_j+1} - b)\|_1}{\mu_{\ell_j+1}|\Gamma|} \longrightarrow \frac{\|\Phi_\Gamma^\top(\Phi x^\infty - b)\|_1}{\underline{\mu}|\Gamma|} := \gamma^\infty,$$

Taking the limit of the first equation in (42), we get

$$0 = \Phi_\Gamma^\top(\Phi_\Gamma x_\Gamma^\infty - b) + \underline{\mu}\gamma^\infty \mathrm{v}_\Gamma^\infty, \tag{43}$$

where v^∞ is the limit of $\mathrm{v}_i^{\ell_j+1}$ on the support $i \in \Gamma$ (the limit exists because of (28)). Dividing the both sides of the second equation in (42) by $\|\mathrm{w}_\Gamma^{\ell+1}\|$ (which goes to ∞), using the fact that $\mathrm{v}_i^{\ell_j+1} \to 0$ for $i \in \Gamma^c$, and taking the limit, we have

$$0 = \mathbf{0}_{\Gamma^c}^\top(\Phi_\Gamma x_\Gamma^\infty - b) = \lim_{\ell_j \to \infty} \left\{ \frac{\Phi_{\Gamma^c}^\top}{\|\mathrm{w}_{\Gamma^c}^{\ell_j+1}\|}(\Phi_\Gamma x_\Gamma^{\ell_j+1} - b) + \mu_{\ell_j+1}\frac{\mathrm{w}_{\Gamma^c}^{\ell_j+1}}{\|\mathrm{w}_{\Gamma^c}^{\ell_j+1}\|} \circ \mathrm{v}_{\Gamma^c}^{\ell_j+1} \right\}. \tag{44}$$

where $\mathbf{0}$ is the zero matrix. The equations in (43) and (44) are actually the optimality conditions for the following optimization problem

$$x_\Gamma^\infty \in \operatorname*{argmin}_{y \in \mathbb{R}^{|\Gamma|}} \frac{1}{2}\|\Phi_\Gamma y - b\|^2 + \gamma^\infty \underline{\mu}\|y\|_1, \qquad x_{\Gamma^c}^\infty = 0.$$

Thus, we complete the whole proof. □

5 Numerical Experiments

In this section, we will report some numerical comparison between our proposed $\texttt{MIRL}_1$ algorithm and some of the latest algorithms in compressed sensing. We will test the exact recovery $y = \Phi x$ by utilizing the sensing matrix Φ which is chosen as in [35]. More specifically, two types of sensing matrices of $\Phi = [\Phi_1, \cdots, \Phi_n] \in \mathbb{R}^{m\times n}$ will be generated, namely, random Gaussian matrix and random partial discrete cosine transform (DCT) matrix:

$$\begin{array}{lll} \text{Gaussian:} & \Phi_j \overset{i.i.d.}{\sim} \mathcal{N}(0, m^{-1}I_m), & j = 1,2,\ldots,n, \\ \text{DCT:} & \Phi_{ij} = m^{-1/2}\cos(2\pi(j-1)\psi_i), & i = 1,2,\cdots,m, \quad j = 1,2,\ldots,n, \end{array}$$

where $\psi \in \mathbb{R}^m \overset{i.i.d.}{\sim} \mathcal{U}([0,1]^m)$ whose components are uniformly and independently sampled from $[0,1]$. We generate the 'true' original signal and the measurement vector b by the following pseudo Matlab codes:

$$x_{\rm orig} = \texttt{zeros}(n,1), \quad \Gamma = \texttt{randperm}(n), \quad x_{\rm orig}(\Gamma(1:k)) = \texttt{randn}(k,1),$$

where k is the sparsity of the original signal $x_{\rm orig}$. To remove the possibility of $x_{\rm orig}$ containing too many elements of small magnitude, we add 0.01 to each none zero entry:

$$x_{\rm orig} = x_{\rm orig} + 0.01 * \text{sign}(x_{\rm orig}), \quad b = \Phi x_{\rm orig}.$$

For the convenience of comparison, we denote by x the signal which is to be recovered by our algorithms. All the experiments are carried out on a CPU 3.2GHz RMA 4.0GB desktop.

5.1 Interpretation for Parameters

The stopping criterion used in Step 2 of $\texttt{MIRL}_1$ algorithm is

$$\frac{\|x^\ell - x^{\ell+1}\|}{\max\{1, \|x^\ell\|\}} \le 10^{-2}.$$

The stopping criterion for `YALL1` solver is $\texttt{tol} = 10^{-4}$. We will utilize the default Matlab package `linsolve` in Step 3. For initialization, we always start with $x^0 = 0$ and the initialized weight $\mathrm{w}^1 = (1,1,\cdots,1)^\top$, and the remaining parameters are set as $\epsilon_1 = 10^{-10}$, $\theta_0 = 0.1\mu_1 m/n$, $\alpha_1 = 1.005$, $\alpha_0 = 0.2$, $\mu_1 = 0.01\|\Phi^\top b\|_\infty$. For x^ℓ, we sort its components in decreasing magnitude: $|x^\ell_{j_1}| \ge |x^\ell_{j_2}| \ge \cdots \ge |x^\ell_{j_n}|$. Instead of fixing ϵ_2 in updating w in (27), we use ϵ_2^ℓ as suggested by [14]:

$$\epsilon_2^\ell = \max\{10^{-3}, |x^\ell_{j_0}|\}, \qquad \text{and} \qquad j_0 = \lceil \frac{m}{4\log(n/m)} \rceil. \tag{45}$$

A crucial quantity in defining T_0^ℓ is the number of nonzero entries in x^ℓ. In order to avoid including too many such elements because x^ℓ tends to have many elements of small magnitude, we propose to use

$$k_\ell(r) := \min \Big\{ s : \sum_{i=1}^{s} |x^\ell_{j_i}| \ge r\|x^\ell\|_1 \Big\}, \tag{46}$$

where $r \in (0,1)$ controls the total percentage of the contribution of the first s largest elements (in terms of the magnitude) in x^ℓ to its ℓ_1 norm $\|x^\ell\|_1$. Throughout the computation, r is chosen by (we also tried other values and this value seemed to give a balanced results in all tests).

$$r_0 = \begin{cases} \frac{1}{\ln(n/m)}, & \ln(n/m) > 1, \\ 0.7, & \text{otherwise} . \end{cases}$$

Next, we highlight the benefit of Step 3 in $\texttt{MIRL}_1$, which is justified by our main result Thm. 4.4(ii). We call this step the refinement step. In the test reported in Table 3, we run 100

n	$k = \lceil 0.01n \rceil$		$k = \lceil 0.05n \rceil$	
	$\|x - x_{\text{orig}}\|$	cpu Time (s)	$\|x - x_{\text{orig}}\|$	cpu Time (s)
1000	1.94e-03 | 1.73e-15	0.040 | 0.034	6.16e-03 | 5.06e-15	0.049 | 0.047
2000	1.73e-03 | 3.46e-15	0.241 | 0.232	8.39e-03 | 1.00e-14	0.300 | 0.311
3000	2.15e-03 | 4.97e-15	0.716 | 0.714	9.19e-03 | 1.33e-14	0.967 | 0.961
4000	2.35e-03 | 5.97e-15	1.167 | 1.169	1.08e-02 | 1.81e-14	1.556 | 1.560
5000	2.31e-03 | 7.37e-15	1.812 | 1.813	1.33e-02 | 2.22e-14	2.341 | 2.357
6000	3.68e-03 | 8.42e-15	2.495 | 2.500	1.26e-02 | 2.62e-14	3.443 | 3.476
7000	3.24e-03 | 1.04e-14	3.456 | 3.459	1.58e-02 | 3.06e-14	4.926 | 4.991
8000	3.46e-03 | 1.29e-14	5.088 | 5.114	1.52e-02 | 3.39e-14	6.355 | 6.482
9000	4.08e-03 | 1.44e-14	5.853 | 5.848	1.62e-02 | 4.02e-14	8.360 | 8.665
10000	2.78e-03 | 1.56e-14	7.660 | 7.696	1.59e-02 | 4.33e-14	10.77 | 11.36

Table 3: Results of MIRL_1 with and without the refinement step. The first column under each measurement column of $\|x - x_{\text{orig}}\|$ and cpu Time is without the refinement step and the second column is with the refinement step.

independent trials from $n = 1000$ to $n = 10000$ with $m = n/4$, and $k = \lceil 0.01n \rceil$ and $\lceil 0.05n \rceil$, respectively. We report the results by $\texttt{MIRL}_1$ with and without the refinement step. The left column underneath $\|x - x_{\text{orig}}\|$ and Time in Table 3 is for without the refinement step and the right column is after the refinement step. One can easily observe that $\texttt{MIRL}_1$ with the refinement returns far more accurate solutions than it does without the refinement step. Moreover, the cpu time (in seconds) is almost the same.

To finish this subsection, we would like to emphasize that the refinement step is also useful for all other methods tested in this paper, but only in the situation when they were able to obtain the correct support set. In fact, if one method gets a wrong support set, this refinement step would lead to a far worse solution. We did test the idea of including this refinement step in all other methods. Our finding is that this refinement step failed to make a difference in terms of the success rate of recovery explained in the next subsection, over 100 random runs for each test problem. Because of this, we will not include this refinement step in the other methods in the subsequent comparisons.

5.2 Comparison of Different Algorithms

We will compare our method $\texttt{MIRL}_1$ with $\texttt{ADMML}_1$ [4], $\texttt{IRL}_1$ [14], $\texttt{IRLSL}_q$ [26], $\texttt{DCAL}_{1-2}$ [35] and two greedy algorithms CoSaMP [30] and SP [15], which try to solve the following problems respectively,

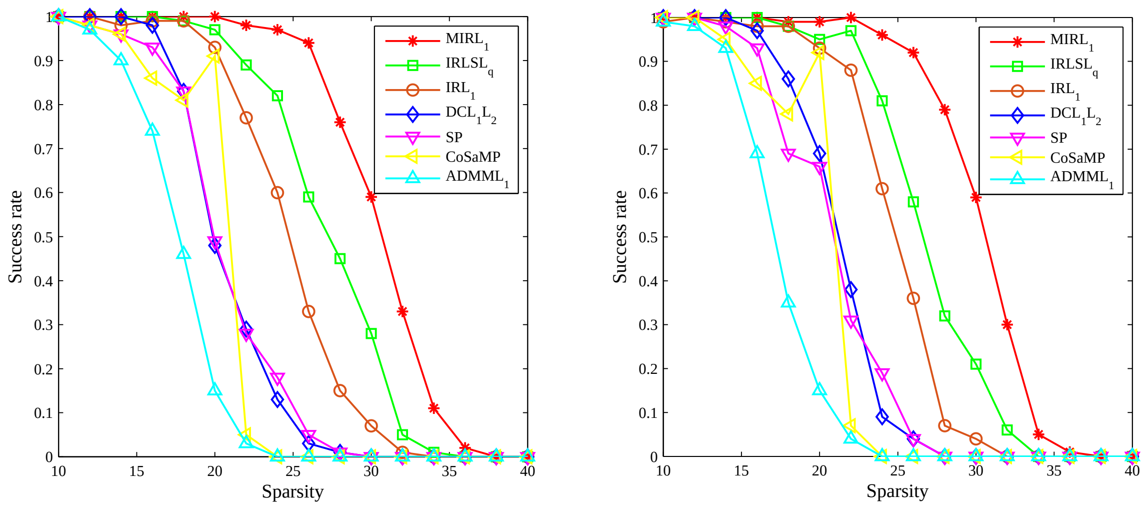
$$
\begin{aligned}
&\texttt{ADMML}_1: && \min_x \frac{1}{2}\|\Phi x - b\|^2 + \mu\|x\|_1;\\
&\texttt{IRL}_1: && \min_x \sum_{i=1}^{n} \log(|x_i| + \epsilon), \quad \text{s.t.} \quad \Phi x = b;\\
&\texttt{IRLSL}_q: && \min_x \frac{1}{2}\|\Phi x - b\|^2 + \mu\|x\|_q^q;\\
&\texttt{DCAL}_{1-2}: && \min_x \frac{1}{2}\|\Phi x - b\|^2 + \mu(\|x\|_1 - \|x\|).
\end{aligned}
$$

The MATLAB codes for $\texttt{ADMML}_1$, $\texttt{IRLSL}_q$, $\texttt{DCAL}_{1-2}$, CoSaMP and SP can be downloaded from the respective authors' websites. Similar to [35], we also take advantage of YALL1 solver to solve the weighted ℓ_1 minimization subproblems of $\texttt{IRL}_1$ instead of the default ℓ_1-MAGIC. Moreover in $\texttt{IRL}_1$, the smoothing parameter ϵ is adaptively updated as introduced in [14], i.e., (45), and the outer stopping criterion is the same as that of $\texttt{MIRL}_1$. In particular, we used the following values in Table 4 for those methods and the remaining parameters are set to their default values.

$\texttt{IRL}_1$	$tol = 10^{-5}$	$\texttt{CoSaMP}$	$tol = 10^{-5}, maxit = 1000$
$\texttt{IRLSL}_q$	$maxit = 1000, q = 0.5, tol = 10^{-5}$	$\texttt{SP}$	$tol = 10^{-5}, maxit = 1000$
$\texttt{ADMML}_1$	$\lambda = 10^{-5}, \rho = 10^{-5}, \alpha = 1,$ $\epsilon^{abs} = 10^{-7}, \epsilon^{rel} = 10^{-5}$		

Table 4: Values of some parameters in different algorithms.

(a) Gaussian and Partial DCT matrix. We begin with running 100 independent trials and recording the corresponding success rates at sparsity levels from 10 to 40. It is well known that greedy algorithms $\texttt{CoSaMP}$ and $\texttt{SP}$ would perform relatively well when the true sparsity $k = \|x_{\text{orig}}\|_0$ is known in advance. From Figure 1, one can observe that $\texttt{MIRL}_1$ behaves the best for both sensing matrices, followed by $\texttt{IRLSL}_q$ and $\texttt{IRL}_1$ which both are better than $\texttt{DCAL}_{1-2}$, $\texttt{SP}$ and $\texttt{CoSaMP}$. It appears that $\texttt{ADMML}_1$ came last.



(a) Gaussian Matrix (b) Partial DCT Matrix

Figure 1: Success rates using sensing matrices. $m = 64, n = 256, 10 \leq k \leq 40$.

When the true signal is relatively sparse, for instance $k = \lceil 0.01n \rceil$, it is known that the greedy algorithms $\texttt{CoSaMP}$ and $\texttt{SP}$ have their exceptional advantages compared with regularized algorithms. However, as the signal becoming less sparse, they do not perform as well as, even much worse than the others. Therefore in the remaining test, we will not include these two methods any more. Now let $m = n/4$, $k = \lceil 0.01n \rceil$, $\lceil 0.05n \rceil$, and $\lceil 0.08n \rceil$ for different dimension n. We run 100 independent trials for Gaussian matrices and record the average errors $\|\Phi x - b\|$, $\|x - x_{\text{orig}}\|$ and $\texttt{cpu}$ time. As indicated in Figure 2, $\texttt{MIRL}_1$ tends to outperform others in terms of the smallest recovered errors and computing time. When $k = \lceil 0.01n \rceil$, $\texttt{DCAL}_{1-2}$ and $\texttt{ADMML}_1$ behave much better than $\texttt{IRL}_1$ and $\texttt{IRLSL}_q$. However, when $k = \lceil 0.08n \rceil$, $\texttt{DCAL}_{1-2}$ and $\texttt{ADMML}_1$ fail to recover the signal because the $\|x - x_{\text{orig}}\|$ are larger than 1 when $n \geq 1500$, see Table 5.

(b) Toeplitz Correlation matrix. Next, we test examples related to the Toeplitz Correlation matrix which is widely used in statistics, and recently sparse property of variables under such matrices has drawn large amounts of attention, see, i.e., [33]. In this simulation, each row $\phi_i (i = 1, \cdots, m)$ of the design matrix Φ is generated from $\mathcal{N}(0, \Sigma)$ distribution with Toeplitz

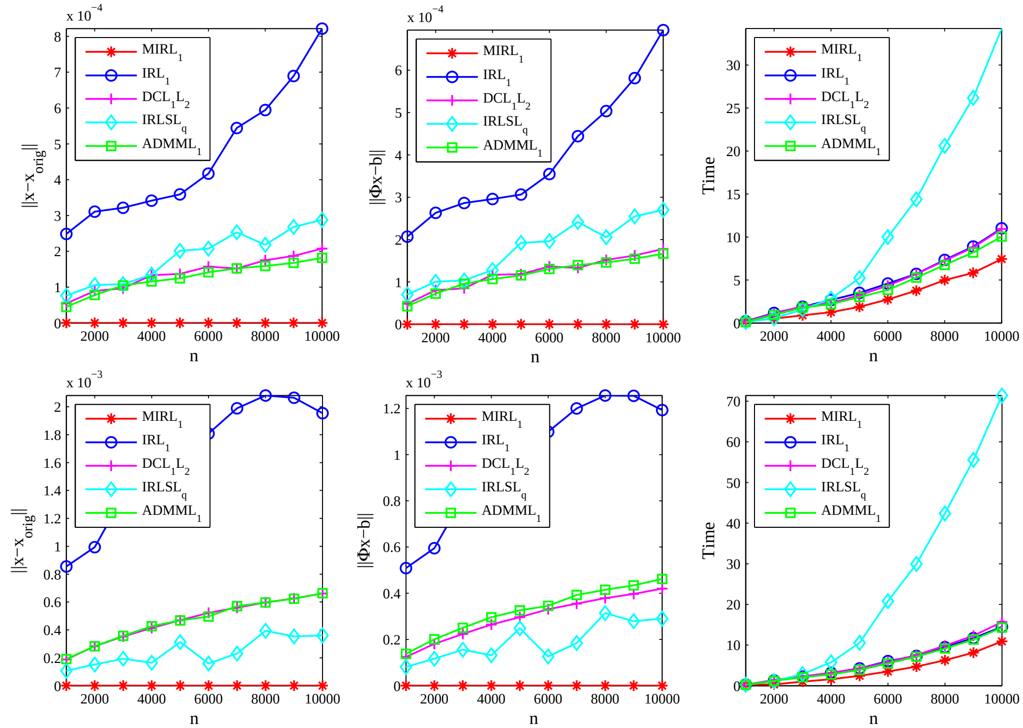


Figure 2: Recovered errors and cpu time. Data shown in the first and second row is generated with $k = \lceil 0.01n \rceil$ and $\lceil 0.05n \rceil$ respectively.

Correlation matrix $\Sigma_{ij} = 2^{-|i-j|}$. Then, each row of Φ possesses the following format

$$\phi_i \overset{i.i.d.}{\sim} \Sigma^{1/2}\mathcal{N}(0, I_n), \qquad i = 1, 2, \ldots, m.$$

We still run 100 independent trials for this test, and record the data (in left part of Figure 3) generated by the 5 algorithms in terms of their success rate for the instances whose sparsity ranges from 10 to 40. Clearly, $\texttt{MIRL}_1$ behaves the best in such circumstance, followed by $\texttt{IRLSL}_q$, $\texttt{IRL}_1$ and $\texttt{DCAL}_{1-2}$ which all perform better than $\texttt{ADMML}_1$.

(c) Over-sampled partial DCT matrix. Our final test is on the over-sampled partial DCT matrices whose definition is:

$$\Phi_{ij} = m^{-1/2}\cos(2\pi(j-1)\psi_i/F), \quad i = 1, 2, \ldots, m, \quad j = 1, 2, \ldots, n,$$

where $\psi \in \mathbb{R}^m \overset{i.i.d.}{\sim} \mathcal{U}([0,1]^m)$ and the positive integer F is the refinement factor. Similarly to [35], the 'true' signal x_{orig} is generated to satisfy

$$\min_{i,j \in \text{supp}(x_{\text{orig}})} |i - j| \geq 2F.$$

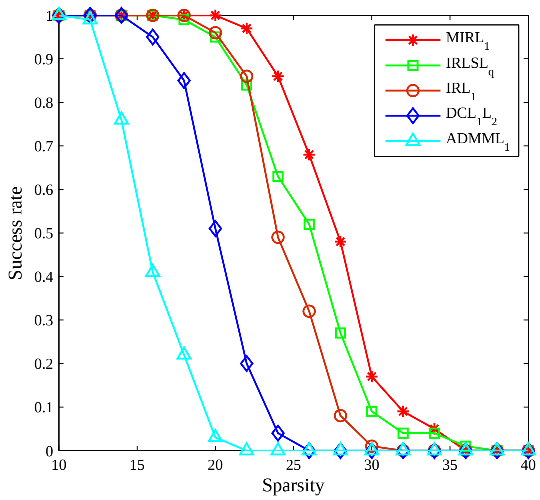
We again run 100 independent trials for this test with $F = 10$, and plot in right part of Figure 3 the performance of the 5 algorithms in terms of their success rate for the instances whose sparsity ranges from 10 to 36. It can be clearly seen that $\texttt{DCAL}_{1-2}$ performs exceptionally well indeed, followed by $\texttt{MIRL}_1$, $\texttt{ADMML}_1$ and $\texttt{IRL}_1$ with almost same performance. Moreover, $\texttt{IRLSL}_q$ behaves not as strongly as others.

6 Conclusion

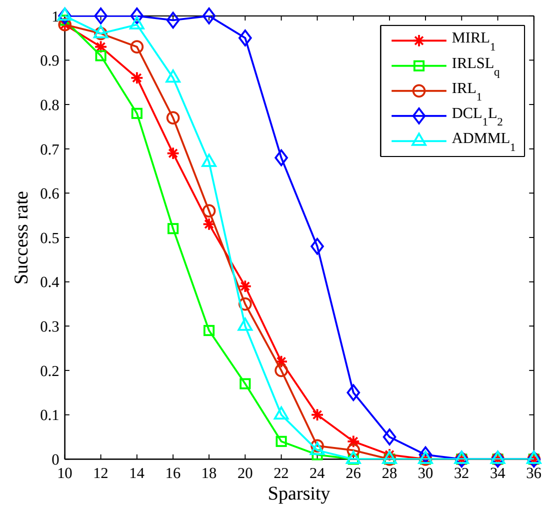
In this paper, we studied a novel weighted ℓ_1-minimization, which makes use of the null space of the observation matrix. Our main result shows that the tight RIC bound on high order

	$\|x - x_{\text{orig}}\|$					cpu Time				
n	MIRL$_1$	IRL$_1$	DCAL$_{1-2}$	IRLSL$_q$	ADMML$_1$	MIRL$_1$	IRL$_1$	DCAL$_{1-2}$	IRLSL$_q$	ADMML$_1$
1000	7.6e-15	5.3e-4	9.3e-1	6.8e-5	2.0e+0	0.10	0.53	2.60	0.39	1.17
1500	1.0e-14	6.0e-4	1.7e+0	1.7e-4	2.8e+0	0.24	1.20	6.33	1.06	1.88
2000	1.4e-14	7.9e-4	2.0e+0	1.8e-4	3.2e+0	0.64	1.78	16.43	1.95	3.33
2500	1.7e-14	8.8e-4	2.7e+0	2.0e-4	3.7e+0	1.35	2.66	28.27	3.46	5.36
3000	2.1e-14	9.3e-4	2.6e+0	2.1e-4	3.6e+0	2.14	3.83	40.98	5.26	7.75
3500	2.4e-14	1.0e-3	3.1e+0	1.4e-4	4.2e+0	2.62	4.34	52.97	7.84	9.54
4000	2.7e-14	1.3e-3	3.6e+0	2.3e-4	4.6e+0	3.62	5.41	75.31	10.93	12.68

Table 5: Comparison of MIRL$_1$, IRL$_1$, DCAL$_{1-2}$ and IRLSL$_q$ when $k = \lceil 0.08n \rceil$.



(a) Toeplitz Correlation Matrix



(b) Over-sampled Partial DCT Matrix

Figure 3: Success rates using other matrices. $m = 100, n = 2000$.

δ_{tk} $(t > 1)$ by Cai and Zhang [10] for the ℓ_1-minimization can be relaxed when the weighted scheme is used. The significance of this result is that there is a way to improve the Cai-Zhang bound via a weighted scheme. We proposed one such scheme. We hope that this research will stimulate alternatives (e.g., less expensive ones) to improve this type of the RIC bound. The main drawback of our scheme is that the optimization problem on the null space is as hard to solve as the original ℓ_0-minimization problem. We overcome this drawback by proposing a novel updating rule for the weight vector, leading to the reweighted ℓ_1-minimization scheme. The new rule is motivated by the constant γ_Φ used in our RIC bound. We proved that under reasonable conditions the reweighted method is capable of identifying the correct support of the sparse solution. This has led to the refinement step, which solves the weighted ℓ_1 minimization on a reduced subspace. Numerical comparison with other leading ℓ_1 solvers demonstrated the effectiveness of our method.

The proof techniques developed for the high order δ_{tk} can actually be extended to study new RIC bounds for δ_k itself. We did not include this part mainly because we intend to focus exclusively on the RIC bounds of the Cai-Zhang type (3). We will explore this in separate research.

Acknowledgement

We sincerely thank the referees as well as the associate editor for their constructive comments on the previous version entitled: "The Weighted ℓ_1 Minimization Approach to Compressed Sensing". Their previous as well as the new comments have significantly improved the quality of the paper.

References

[1] Andersson J. & Strömberg J. (2014) On the Theorem of Uniform Recovery of Random Sampling Matrices. *IEEE Trans. Inform. Theory*, **60**(3), 1700-1710.

[2] Blanchard J.D. & Thompson A. (2010) On Support Sizes Of Restricted Isometry Constants. *Appl. Comput. Harmon. Anal.* **29**, 382-390.

[3] Blanchard J.D., Cartis C. & Tanner J. (2011) Compressed Sensing: How Sharp Is the Restricted Isometry Property? *SIAM Review*, **53**, 105-125.

[4] Boyd S., Parikh N., Chu E., Peleato B. & Eckstein J. (2011) Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Foundations and Trends in Machine Learning*, **3**(1), 1-122.

[5] Bruckstei A.M., Donoho D.L. & Elad A. (2009) From Sparse Solutions of Systems of Equations to Sparse Modeling of Signals and Images. *SIAM Rev.*, 51, 34-81.

[6] Cai T.T., Wang L. & Xu G.W. (2010) New Bounds for Restricted Isometry Constants. *IEEE Trans. Inform. Theory*, **56**, 4388-4394.

[7] Cai T.T., Wang L. & Xu G.W. (2010) Shifting Inequality and Recovery of Sparse Signals. *IEEE Trans. Signal Process.*, **58**, 1300-1308.

[8] Cai T.T. & Zhang A.R. (2013) Sharp RIP Bound for Sparse Signal and Low-rank Matrix Recovery. *Appl. and Comput. Harmon. Anal.*, **35**, 74-93.

[9] Cai T.T. & Zhang A.R. (2013) Compressed Sensing and Affine Rank Minimization under Restricted Isometry. *IEEE Trans. Signal Process.*, **61**, 3279-3290.

[10] Cai T.T. & Zhang A.R. (2014) Sparse Representation of a Polytope and Recovery of Sparse Signals and Low-rank Matrices. *IEEE Trans. Inform. Theory.* **60**, 122-132.

[11] Candès E.J. (2008) The Restricted Isometry Property and its Implications for Compressed Sensing. *Compte Redus Math.* **346**, 589-592.

[12] Candès E.J., Romberg J. & Tao T. (2006) Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information. *IEEE Trans. Inform. Theory*, **52**, 489-509.

[13] Candès E.J. & Tao T. (2005) Decoding by Linear Programming. *IEEE Trans. Inform. Theory*, **51**, 4203-4215.

[14] Candès E.J., Wakin M.B. & Boyd S.P. (2008) Enhancing Sparsity by Reweighted ℓ_1 Minimization. *J. Fourier Anal. Appl.*, **14**, 877-905.

[15] Dai W. & Milenkovic O. (2009) Subspace Pursuit for Compressive Sensing Signal Reconstruction. *IEEE Trans. Inform. Theory*, **55**, 2230-2249.

[16] Donoho D.L. (2006) Compressed Sensing. *IEEE Trans. Inform. Theory*, **52**, 1289-1306.

[17] Donoho D.L. & Elad M. (2003) Optimally Sparse Representation in General (Nonorthogonal) Dictionaries via ℓ_1 Minimization. *PNAS*, **100**, 2197-2202.

[18] Eldar Y.C. & Ku Tyniok G. (2012) Compressed Sensing: Theory and Applications. Cambridge University Press.

[19] Forcart S. & Lai M.J. (2009) Sparsest Solutions of Underdetermined Linear Systems via ℓ_q Minimization for $0 < q \leq 1$. *Appl. Comput. Harmon. Anal.*, **26**, 395-407.

[20] Forcart S. & Rauhut (2013) A Mathematical Introduction to Compressive Sensing. Birkhäuser.

[21] Friedlander M.P., Mansour H., Saab R. & Yilmaz Ö. (2012) Recovering Compressively Sampled Signals Using Partial Support Information. *IEEE Trans. Inform. Theory*, **58**, 1122-1134.

[22] Gribonval R. & Nielsen M. (2003) Sparse Decompositions in Unions of Bases. *IEEE Trans. Inform. Theory*, **49**, 3320-3325.

[23] Feuer A. & Nemirovski A. (2003) On Sparse Representation in Pairs of Bases. *IEEE Trans. Info. Theory*, **49**, 1579-1581.

[24] Kanzow C. & Qi H.-D. (1999) A QP-free Constrained Newton-type Method for Variational Inequality Problems, *Math. Program.* **85**, 81-106.

[25] Khajehnejad M.A., Xu W., Salman A.A. & Hassibi B. (2009) Weighted ℓ_1 Minimization for Sparse Recovery with Prior Information. *IEEE International Symposium on Information Theory*, 483-487.

[26] Lai M.J., Xu Y. & Yin W. (2013) Improved Iteratively Reweighted Least Squares for Unconstrained Smoothed ℓ_q Minimization. *SIAM J. Numer. Anal.*, **51**, 927-957.

[27] Mo Q. & Li S. (2011) New Bounds on the Restricted Isometry Constant δ_{2k}. *Appl. Comput. Harmon. Anal.*, **31**, 460-468.

[28] Moré, J.J. & Sorensen, D.C. (1983) Computing a Trust Region Step. *SIAM J. Sci. Stat. Comput.*, **4**, 553-572

[29] Neal P. & Boyd S. (2013) Proximal Algorithms. *Foundations and Treads in Optimaization*, **1**, 1-96.

[30] Needell D. & Tropp J.A. (2009) CoSaMP: Iterative Signal Recovery from Incomplete and Inaccurate Samples. *Appl. Comput. Harmon. Anal.*, **26**, 301-332.

[31] Ruahut H. (2010) Compressive Sensing and Structured Random Matrices. *Radon Series Comp. Appl. Math.*, **9**, 1-92.

[32] Vaswani N. & Lu, W. (2010) Modified-CS: Modifying Compressive Sensing for Problems with Partially Known Support. *IEEE Trans. Singal Process.,* **58**, 4595-4607.

[33] Wang L. (2013) The L_1 Penalized LAD Estimator for High Dimensional Linear Regression. *Journal of Multivariate Analysis*, **120**, 135-151.

[34] Xu G.W. & Xu Z.Q. (2013) On the ℓ_1 Norm Invariant Convex k-sparse Decomposition of Signals. *J. Oper. Res. Soc. China*, **1**(4), 537-541.

[35] Yin P.H., Luo Y.F., He Q. & Xin J. (2015) Minimization of L1-L2 for Compressed Sensing. To appear in *SIAM J. of Sci. Computing.*

[36] Yu X. & Baek S.J. (2013) Sufficient Conditions on Stable Recovery of Sparse Signals with Partial Support Information. *IEEE Signal Processing Letters*, **20**, 539–542.

[37] Zhang Y. (2013) Theory of Compressive Sensing via ℓ-Minimization: A Non-RIP Analysis and Extensions. *J. Oper. Res. Soc. China*, **1**(1), 79-105.

[38] Zhao Y.B. (2013) RSP-based Analysis for Sparsest and Least ℓ_1-norm Solutions to Underdetermined Linear Systems. *IEEE Trans. Signal Process.*, **61**, 5777-5788.

[39] Zhao Y.B. & Li D. (2012) Reweighted ℓ_1-Minimization for Sparse Solutions to Underdetermined Linear Systems. *SIAM J. Optim.*, **22**, 1065-1088.

[40] Zhou S.L., Kong L.C. & Xiu N.H. (2013) New Bounds for RIC in Compressed Sensing. *J. Oper. Res. Soc. China*, **1**(2), 227-237.