Weak Stability of ℓ_1-minimization Methods in Sparse Data Reconstruction

YUN-BIN ZHAO*, HOUYUAN JIANG †, and ZHI-QUAN LUO‡

(1st version December 2015, 2nd version May 2016, Revised April 2017)

Abstract. As one of the most plausible convex optimization methods for sparse data reconstruction, ℓ_1-minimization plays a fundamental role in the development of sparse optimization theory. The stability of this method has been addressed in the literature under various assumptions such as restricted isometry property (RIP), null space property (NSP), and mutual coherence. In this paper, we propose a unified means to develop the so-called weak stability theory for ℓ_1-minimization methods under the condition called weak range space property of a transposed design matrix, which turns out to be a necessary and sufficient condition for the standard ℓ_1-minimization method to be weakly stable in sparse data reconstruction. The reconstruction error bounds established in this paper are measured by the so-called Robinson's constant. We also provide a unified weak stability result for standard ℓ_1-minimization under several existing compressed-sensing matrix properties. In particular, the weak stability of ℓ_1-minimization under the constant-free range space property of order k of the transposed design matrix is established for the first time in this paper. Different from the existing analysis, we utilize the classic Hoffman's Lemma concerning the error bound of linear systems as well as the Dudley's theorem concerning the polytope approximation of the unit ℓ_2-ball to show that ℓ_1-minimization is robustly and weakly stable in recovering sparse data from inaccurate measurements.

Key words. Sparsity optimization, ℓ_1-minimization, convex optimization, linear program, weak stability, weak range space property

AMS subject classifications: 90C05, 90C25, 90C31, 94A12, 15A29.

*School of Mathematics, University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom (zhaoyy@maths.bham.ac.uk). The research of this author was partially supported by the Engineering and Physical Sciences Research Council (EPSRC) under grant #EP/K00946X/1.

†Judge Business School, University of Cambridge, Trumpington Street, Cambridge CB2 1AG, United Kingdom (h.jiang@jbs.cam.ac.uk).

‡Shenzhen Research Institute of Big Data, Chinese University of Hong Kong, Shenzhen, Guangdong Province, China (luozq@cuhk.edu.cn).

1 Introduction

Data might be contaminated by some form of random noise and the measurements of data are subject to quantization error. Thus a huge effort in sparse data reconstruction is made to ensure the reconstruction algorithms stable in the sense that reconstruction errors stay under control when the measurements are slightly inaccurate and when the data is not exactly sparse (see, e.g., [2, 21, 22, 25]). One of the widely used reconstruction models is the ℓ_1-minimization

$$\min_x \{\|x\|_1 : \ \|Ax - y\|_p \leq \varepsilon\}, \tag{1}$$

where $\|\cdot\|_p$ is the ℓ_p-norm with $p \geq 1$ ($p = 1, 2, \infty$ will be considered in this paper). In the above model, $A \in \mathbb{R}^{m\times n}$ ($m < n$) is a full-row-rank matrix called a design or sensing matrix which is a collection of known or learned dictionaries, $y = A\widehat{x} + u$ is the acquired measurement vector for the data $\widehat{x}$ to be reconstructed, and u represents the measurement error with level $\|u\|_p \leq \varepsilon$. The size of ε is closely tied with the noise power. In this paper, the given data (A, y, ε) is referred to as the problem data of (1). When $\varepsilon = 0$, (1) is reduced to the so-called standard ℓ_1-minimization, i.e., $\min\{\|x\|_1 : \ Ax = y\}$. The use of ℓ_1-norm to promote sparsity in data processing has actually a long history (see, e.g., [32, 39, 31, 19, 33, 14, 34]), but a significant development of theory and algorithms for sparse data reconstruction has been made only recently in the framework of compressed sensing (see, e.g. [16, 10, 9, 17, 8, 21, 22]).

Assume that an unknown vector, denoted by $\widehat{x}$, satisfies $\|A\widehat{x} - y\|_p \leq \varepsilon$. In traditional compressed sensing setting, it is generally assumed that problem (1) admits a unique optimal solution, in which case it is interesting to know how close the unique solution of (1) to $\widehat{x}$. This leads to the traditional stability analysis for ℓ_1-minimization methods. The major results in this aspect have been achieved by Donoho, Candès, Romberg, Tao, and others (e.g., [17, 10, 9, 8]). However, from a mathematical point of view, we still need to understand the general stability (which is referred to as the weak stability in this paper) of a reconstruction model by taking into account the settings where the problem might possess multiple optimal solutions or the sensing matrix A might admit a certain less restrictive property than existing assumptions. Moreover, the study of weak stability will also provide a novel stability result under existing stability conditions. Let us first recall the notation of best k-term approximation before we introduce the weak stability. Let k be an integer number and define

$$\sigma_k(x)_1 := \inf_z \{\|x - z\|_1 : \|z\|_0 \leq k\},$$

where $x \in \mathbb{R}^n$ and $\|z\|_0$ denotes the number of nonzero entries of $z \in \mathbb{R}^n$. $\sigma_k(x)_1$ is called the ℓ_1-error of best k-term approximation. Let x^* be an optimal solution of (1) with given problem data (A, y, ε). Problem (1) is said to be weakly stable for noise-free reconstruction ($\varepsilon = 0$) if for any feasible vector x of the problem, there is a solution x^* of (1) such that

$$\|x - x^*\| \leq C\sigma_k(x)_1, \tag{2}$$

where $\|\cdot\|$ is a norm and C is a constant depending on the problem data (A, y). Problem (1) is said to be robustly and weakly stable for noisy reconstruction ($\varepsilon > 0$) if for any feasible vector x of the problem, there is a solution x^* of (1) such that

$$\|x - x^*\| \leq C_1\sigma_k(x)_1 + C_2\varepsilon, \tag{3}$$

where C_1 and C_2 are constants determined by the problem data (A, y, ε).

When the solution x^* of (1) is unique (for instance, when $\varepsilon = 0$ and when the matrix A admits the restricted isometry property (RIP) or null space property (NSP), see Definition 2.1), the weak stability can be reduced to the normal stability if constants C, C_1 and C_2 are often measured in terms of RIP or NSP constants. Candès and Tao [10, 11] introduced the notion of the RIP with constant δ_K, where K is a certain integer number, and they proved in [11] that if $\delta_{2k} + \delta_{3k} < 1$, all k-sparse vectors can be exactly reconstructed via standard ℓ_1-minimization. Furthermore, Candès, Romberg and Tao [9] have shown that the stability of problem (1) with $p = 2$ can be guaranteed if $\delta_{3k} + 3\delta_{4k} < 2$. This result was improved to $\delta_{2k} < \sqrt{2} - 1$ in [8], and was further improved by several researchers (see, e.g., [24, 4, 35, 5, 25, 1]). Finally, Cai and Zhang [6] has improved this bound to $\delta_{2k} < 1/\sqrt{2}$.

The NSP of order k (see Definition 2.1) is a necessary and sufficient condition for every k-sparse vector to be exactly reconstructed with standard ℓ_1-minimization. This NSP property appeared in [18, 16, 28] and was formally called the null space property by Cohen et al. [15]. The NSP is strictly weaker than the RIP (see, e.g., [23, 3]). It was shown [15, 38, 22, 25, 3] that the stable NSP or robust NSP (which is a strengthened version of the NSP of order k) guarantees the stability of ℓ_1-minimization. A typical feature of RIP- and NSP-based stability results for ℓ_1-minimization methods is that the coefficients C, C_1 and C_2 in (2) and (3) are measured by the RIP constant, stable NSP constant or the robust NSP constant.

The range space property (RSP) of order k of A^T (see Definition 2.1) was introduced in [45]. This property is also a necessary and sufficient condition for recovering every k-sparse vector with standard ℓ_1-minimization. So this property is equivalent to the NSP of order k. If the RSP is only defined locally at a specific vector x^*, it is called the individual RSP of A^T at x^*, which is a nonuniform recovery condition for a specific vector [45]. A stability analysis at a specific vector for ℓ_1-minimization has been carried out in [43], under an assumption equivalent to the individual RSP. Note that RSP of order k of A^T and NSP of order k are constant-free conditions in the sense that their definitions do not involve any constant, unlike the stable or robust NSP of order k. Although the stability of ℓ_1-minimization methods has been extensively studied under various conditions in the literature, the weak stability of these methods has not been properly established at present. In this paper, we consider a more relaxed constant-free condition than RSP of order k of A^T. We ask whether the weak stability of ℓ_1-minimization methods can be developed under less restrictive constant-free matrix properties than the existing ones.

We note that the optimal solution x^* of (1) is not determined by the problem data A only. Clearly, x^* is jointly determined by all problem data (A, y, ε) of (1). Different measurement vector y and noise level ε together with different choice of the norm in (1) will affect the optimal solution of (1) as well. In other words, in addition to A, the problem data (y, ε) will also directly or indirectly affect the reconstruction ability and stability of ℓ_1-minimization methods. Exploiting adequate problem data will levitate the dependence on the matrix property, and might yield a weak stability result under less restrictive assumptions than existing conditions.

The purpose of this paper is to establish such weak stability results for ℓ_1-minimization methods under a constant-free and mild matrix property. We prove that the so-called *weak range space property of* A^T (see Definition 2.2) is a desired sufficient condition for many ℓ_1-minimization methods to be weakly stable in sparse data reconstruction. We show that this condition is also necessary for standard ℓ_1-minimization to be weakly stable for any given measurement vector $y \in \{Ax : \|x\|_0 \leq k\}$. This property is directly tied to and originated naturally from the fundamental Karush-Kuhn-Tucker (KKK) optimality conditions for linear optimization. It is well known that the optimality conditions completely characterize the optimal solutions x^* of ℓ_1-minimization through problem data no matter whether the optimal solution of the problem is unique or not. We will demonstrate that the weak RSP of order k of A^T, together with a classic error bound of linear systems developed by Hoffman [30] and Robinson [37], provides an efficient way to develop the weak stability theory for ℓ_1-minimization. Existing RIP, NSP, mutual coherence conditions and their variants imply the weak RSP of A^T, and we show that each of these existing conditions implies the same reconstruction error bounds in terms of the so-called Robinson's constants depending on the problem data. Moreover, the weak stability of ℓ_1-minimization under the RSP of order k of A^T or NSP of order k is immediately obtained for the first time, as a special case of the general weak stability results established in this paper.

This paper is organized as follows. In section 2, we give the definitions of several key matrix properties and recall the Robinson's constant and Hoffman's lemma. We also prove that the weak RSP of order k of A^T is a necessary condition for standard ℓ_1-minimization with measurements $y \in \{Ax : \|x\|_0 \leq k\}$ to be weakly stable in sparse data reconstruction. In section 3, we characterize the weak stability of standard ℓ_1-minimization under the weak RSP. In section 4, we show the robust weak stability of the ℓ_1-minimization problem with linearly representable constraints, i.e., $p = 1$ and $p = \infty$ in (1). In section 5, we prove the robust weak stability of quadratically constrained ℓ_1-minimization.

Notation. Unless otherwise stated, the identity matrix of any order will be denoted by I and a vector of ones will be denoted by e. The nonnegative orthant in $\mathbb{R}^n$ will be denoted by $\mathbb{R}^n_+$. The set of $m \times n$ matrices is denoted by $\mathbb{R}^{m\times n}$. The p-norm of a vector is defined as $\|x\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$, where $p \geq 1$. In particular, when $p = \infty$, the p-norm is reduced to $\|x\|_\infty = \max_{1\leq i\leq n} |x_i|$. The induced matrix norm of A is defined as $\|A\|_{p\to q} = \max_{\|x\|_p\leq 1} \|Ax\|_q$. For a vector $x \in \mathbb{R}^n$, $|x|$, $(x)^+$ and $(x)^-$ denote the vectors in $\mathbb{R}^n$ with components $|x|_i := |x_i|$, $[(x)^+]_i := \max\{x_i, 0\}$ and $[(x)^-]_i := \min\{x_i, 0\}, i = 1, \ldots, n$, respectively. Given a subset $S \subseteq \{1, \ldots, n\}$ and a vector $x \in \mathbb{R}^n$, we use $|S|$ to denote the cardinality of S, $\overline{S}$ to denote the complement of S, i.e., $\overline{S} = \{1, \ldots, n\}\backslash S$, and x_S to denote the subvector of x by deleting the components x_i with $i \notin S$. For matrix A, A^T denotes the transpose of A, $\mathcal{R}(A^T)$ the range space of A^T, and $\mathcal{N}(A)$ the null space of A. For any vectors $x, y \in \mathbb{R}^n$, $x \leq y$ means $x_i \leq y_i$ for all $i = 1, \ldots, n$. A vector x is said to be k-sparse if it admits at most k nonzero entries, i.e., $\|x\|_0 \leq k$.

2 Weak RSP of order k of A^T and Robinson's constant

In this section, we provide some notions and facts that will be used throughout the remainder of the paper. Let us first recall some important matrix properties that have been widely used in sparse recovery framework.

Definition 2.1. (a) (RIP of order $2k$) [10, 8] The matrix A is said to satisfy the restricted isometry property of order $2k$ with constant $\delta_{2k} \in (0, 1)$ if $(1-\delta_{2k})\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta_{2k})\|x\|_2^2$ holds for all k-sparse vector $x \in \mathbb{R}^n$.

(b) (NSP of order k) [15, 44, 25] The matrix A is said to satisfy the null space property of order k if $\|v_S\|_1 < \|v_{\overline{S}}\|_1$ holds for any $v \in \mathcal{N}(A)$ and any $S \subseteq \{1, \ldots, n\}$ with $|S| \leq k$.

(c) (Stable NSP of order k) [15, 44, 25] The matrix A is said to satisfy the stable null space property of order k with constant $\rho \in (0, 1)$ if $\|v_S\|_1 \leq \rho\|v_{\overline{S}}\|_1$ holds for any $v \in \mathcal{N}(A)$ and any

$S \subseteq \{1, \ldots, n\}$ with $|S| \leq k$.

(d) (Robust NSP of order k) [15, 25] The matrix A is said to satisfy the robust null space property of order k with constants $\rho \in (0, 1)$ and $\tau > 0$ if $\|v_S\|_1 \leq \rho \|v_{\overline{S}}\|_1 + \tau \|Av\|$ holds for any $v \in \mathbb{R}^n$ and any $S \subseteq \{1, \ldots, n\}$ with $|S| \leq k$.

(e) (RSP of order k of A^T) [45] The matrix A^T is said to satisfy the range space property of order k if for any disjoint subsets S_1, S_2 of $\{1, \ldots, n\}$ with $|S_1| + |S_2| \leq k$, there is a vector $\eta \in \mathcal{R}(A^T)$ satisfying that $\eta_i = 1$ for $i \in S_1$, $\eta_i = -1$ for $i \in S_2$, $|\eta_i| < 1$ for $i \notin S_1 \cup S_2$.

The notion (e) above arises from the uniqueness analysis for the solution of linear ℓ_1-minimization. In fact, for any given $\widehat{x}$, it is known that $\widehat{x}$ is the unique solution to the problem $\min\{\|z\|_1 : Az = A\widehat{x}\}$ if and only if $A_{supp(\widehat{x})}$ (the submatrix of A formed by deleting the columns corresponding to the indices not in $supp(\widehat{x}) = \{i : \widehat{x}_i \neq 0\}$) has full column rank and the following property holds: there is a vector $\eta \in \mathcal{R}(A^T)$ such that $\eta_i = 1$ for $\widehat{x}_i > 0$, $\eta_i = -1$ for $\widehat{x}_i < 0$, and $|\eta_i| < 1$ for $\widehat{x}_i = 0$. The sufficiency of the above statement was shown in [26], and the necessity of the above statement was first shown in [36]. This fact was also rediscovered and proved independently in [27, 45, 25, 42]. However, this uniqueness property depends on the individual vector $\widehat{x}$, and thus it is insufficient for the uniform reconstruction of all k-sparse vectors via ℓ_1-minimization. To exactly reconstruct every k-sparse vector with ℓ_1-minimization, this individual property is strengthened to the RSP of order k of A^T in [45] so that it is independent of any individual vector. Given a matrix $A \in \mathbb{R}^{m \times n}$, it is shown in [45] that every k-sparse vector $\widehat{x} \in \mathbb{R}^n$ can be exactly reconstructed by the ℓ_1-minimization method

$$\min\{\|z\|_1 : Az = y := A\widehat{x}\} \tag{4}$$

if and only if A^T admits the RSP of order k. So the RSP of order k of A^T is a necessary and sufficient condition for the uniform recovery of all k-sparse vectors, and hence it is equivalent to the NSP of order k. An advantage of the RSP concept is that it can be easily extended to sparse data reconstruction with more complex structure than (4) (see, e.g., [46, 49]). We now introduce the weak RSP of order k which is a relaxation of the RSP of order k.

Definition 2.2. (Weak RSP of order k of A^T) The matrix A^T is said to satisfy the weak range space property of order k if for any disjoint subsets S_1, S_2 of $\{1, \ldots, n\}$ with $|S_1|+|S_2| \leq k$, there is a vector $\eta \in \mathcal{R}(A^T)$ satisfying that

$$\eta_i = 1 \text{ for } i \in S_1, \ \eta_i = -1 \text{ for } i \in S_2, \ |\eta_i| \leq 1 \text{ for } i \notin S_1 \cup S_2. \tag{5}$$

Different from the RSP of order k, the inequality "$|\eta_i| \leq 1$ for $i \notin S_1 \cup S_2$" in Definition 2.2 is not required to hold strictly. The weak RSP of order k of A^T is a strengthened optimality condition for the individual problem (4). In fact, by the KKT optimality condition, $\widehat{x}$ is an optimal solution of (4) if and only if there is a vector $\eta \in \mathcal{R}(A^T)$ satisfying $\eta_i = 1$ for $\widehat{x}_i > 0$, $\eta_i = -1$ for $\widehat{x}_i < 0$, and $|\eta_i| \leq 1$ otherwise. Define the specific pair of (S_1, S_2) with $S_1 = \{i : \widehat{x}_i > 0\}$ and $S_2 = \{i : \widehat{x}_i < 0\}$. The KKT optimality condition implies that the condition (5) holds for such a specific pair (S_1, S_2). This can be called the individual weak RSP of A^T at $\widehat{x}$. If we expect that every k-sparse vector $\widehat{x}$ is an optimal solution to the ℓ_1-minimization problem with measurements $y = A\widehat{x}$, then condition (5) must hold for any disjoint subsets (S_1, S_2) with $|S_1 \cup S_2| \leq k$ in order to cover all possible cases of k-sparse vectors. This naturally yields the matrix property described in Definition 2.2.

The RIP of order $2k$ with $\delta_{2k} \leq 1/\sqrt{2}$ implies that every k sparse vector can be exactly recovered by ℓ_1-minimization (e.g., [6]). Thus it implies the RSP of order k of A^T which is equivalent to the NSP of order k. We see that the recovery condition $\mu_1(k) + \mu_1(k-1) < 1$ presented in [41] also implies the NSP of order k (see, e.g., Theorem 5.15 in [25]), where $\mu_1(k)$ is the so-called accumulative coherence defined as

$$\mu_1(k) = \max_{i \in \{1,\ldots,n\}} \max \left\{ \sum_{j \in S} |a_i^T a_j| : \ S \subseteq \{1,\ldots,n\}, \ |S| = k, \ i \notin S \right\},$$

where $a_i, i = 1,\ldots,n$ are the ℓ_2-normalized columns of A. Thus we have the following relation:

$$\left.\begin{array}{r} \text{RIP of order } 2k \Rightarrow \\ \text{Stable NSP of order } k \Rightarrow \\ \text{Robust NSP of order } k \Rightarrow \\ \mu_1(k) + \mu_1(k-1) < 1 \Rightarrow \end{array}\right\} \ \text{NSP of order } k \ \Leftrightarrow \ \begin{array}{l} \text{RSP of order } k \text{ of } A^T \Rightarrow \\ \text{weak RSP of order } k \text{ of } A^T. \end{array}$$

The weak RSP is the mildest one among the above-mentioned matrix properties. To see how mild such a condition is, let us first prove that the weak RSP of order k of A^T is a necessary condition for standard ℓ_1-minimization with any given measurement vector $y \in \{Ax : \|x\|_0 \leq k\}$ to be weakly stable in sparse data reconstruction.

Theorem 2.3. *Let A be a given $m \times n$ $(m < n)$ matrix with rank$(A) = m$. Suppose that for any given measurement vector $y \in \{Ax : \|x\|_0 \leq k\}$, the following holds: For any $x \in \mathbb{R}^n$ satisfying $Ax = y$, there is a solution x^* of the problem* $\min\{\|z\|_1 : Az = y\}$ *such that $\|x - x^*\| \leq C\sigma_k(x)_1$, where $\|\cdot\|$ is a norm and C is a constant dependent on the problem data (A, y). Then A^T must satisfy the weak RSP of order k.*

Proof. Assume that (S_1, S_2) is an arbitrary pair of disjoint subsets of $\{1,\ldots,n\}$ with $|S_1| + |S_2| \leq k$. Under the assumption of the theorem, we now prove that there exists a vector $\eta \in \mathcal{R}(A^T)$ satisfying (5). Then, by Definition 2.2, A^T must admit the weak RSP of order k. Indeed, let $\widehat{x}$ be a k-sparse vector in $\mathbb{R}^n$ such that

$$\{i : \widehat{x}_i > 0\} = S_1, \ \{i : \widehat{x}_i < 0\} = S_2. \tag{6}$$

Consider the problem (4), i.e., $\min\{\|z\|_1 : Az = y := A\widehat{x}\}$. By the assumption, there is an optimal solution x^* to this problem such that $\|\widehat{x} - x^*\| \leq C\sigma_k(\widehat{x})_1$, where C depends on the problem data (A, y). Since $\widehat{x}$ is k-sparse, the right-hand side of the inequality above is equal to zero, and hence $\widehat{x} = x^*$. This, together with (6), implies that

$$\{i : x_i^* > 0\} = S_1, \ \{i : x_i^* < 0\} = S_2, \ x_i^* = 0 \text{ for all } i \notin S_1 \cup S_2. \tag{7}$$

Note that x^* is an optimal solution to the convex problem (4). x^* must satisfy the optimality condition, i.e., there exists a vector $u \in \mathbb{R}^m$ such that $A^T u \in \partial\|x^*\|_1$, where $\partial\|x^*\|_1$ is the subgradient of the ℓ_1-norm at x^*, i.e.,

$$\partial\|x^*\|_1 = \{v \in \mathbb{R}^n : \ v_i = 1 \text{ for } x_i^* > 0, \ v_i = -1 \text{ for } x_i^* < 0, \ |v_i| \leq 1 \text{ otherwise}\}.$$

By setting $\eta = A^T u \in \partial\|x^*\|_1$, we immediately see that $\eta_i = 1$ for $x_i^* > 0$, $\eta_i = -1$ for $x_i^* < 0$, and $|\eta_i| \leq 1$ for $x_i^* = 0$. This, together with (7), implies that the vector $\eta = A^T u$ satisfies (5). Since S_1 and S_2 are arbitrary disjoint subsets of $\{1,\ldots,n\}$ with $|S_1| + |S_2| \leq k$. Thus A^T must satisfy the weak RSP of order k. $\square$

In the next section, we show that the converse of the above result is also valid (see Theorem 3.2 and Corollary 3.3 for details). We will use a classic error bound for linear systems established by Hoffman [30]. Let us first recall a constant introduced by Robinson [37]. Let $P \in \mathbb{R}^{n_1 \times q}$ and $Q \in \mathbb{R}^{n_2 \times q}$ be two real matrices. Define a set $F \subseteq \mathbb{R}^{n_1+n_2}$ by

$$F = \{(b,d) : \text{ for some } z \in \mathbb{R}^q \text{ such that } Pz \leq b \text{ and } Qz = d\}.$$

Let $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$ be norms on $\mathbb{R}^q$ and $\mathbb{R}^{n_1+n_2}$, respectively. Robinson [37] has shown that the quantity

$$\mu_{\alpha,\beta}(P,Q) := \max_{\|(b,d)\|_\beta \leq 1, (b,d) \in F} \min_{z \in \mathbb{R}^q} \{\|z\|_\alpha : \ Pz \leq b, \ Qz = d\} \tag{8}$$

is a finite real number. It has also been shown in [37] that the extreme value above is attained. In this paper, we use $\alpha = \infty$, in which case $\|x\|_\infty$ is a polyhedral norm in the sense that the closed unit ball $\{x : \|x\|_\infty \leq 1\}$ is a polyhedron. Define the optimal value of the internal minimization in (8) as

$$g(b,d) = \min\{\|z\|_\alpha : \ Pz \leq b, \ Qz = d\}, \quad (b,d) \in F.$$

Then

$$\mu_{\alpha,\beta}(P,Q) = \max_{(b,d) \in \widetilde{B} \cap F} g(b,d),$$

where $\widetilde{B} = \{(b,d) : \ \|(b,d)\|_\beta \leq 1\}$ is the unit ball in $\mathbb{R}^{n_1 \times n_2}$. As pointed out in [37], the function $g(b,d)$ is convex over F if $\|\cdot\|_\alpha$ is a polyhedral norm. In this case, $\mu_{\alpha,\beta}(P,Q)$ is the maximum of a convex function over the bounded set $\widetilde{B} \cap F$.

Let $M' \in \mathbb{R}^{m \times q}$ and $M'' \in \mathbb{R}^{\ell \times q}$ be two given matrices. Consider (P,Q) of the form

$$P = \begin{bmatrix} I_N & 0 \\ -I & 0 \end{bmatrix} \in \mathbb{R}^{(|N|+m) \times (m+\ell)}, \ Q = \begin{bmatrix} M' \\ M'' \end{bmatrix}^T \in \mathbb{R}^{q \times (m+\ell)},$$

where N is a subset of $\{1, \ldots, m\}$ and I_N is obtained from the $m \times m$ identity matrix I by deleting the rows corresponding to indices not in N. Robinson [37] defined the following constant:

$$\sigma_{\alpha,\beta}(M', M'') := \max_{N \subseteq \{1,\ldots,m\}} \mu_{\alpha,\beta}\left(\begin{bmatrix} I_N & 0 \\ -I & 0 \end{bmatrix}, \begin{bmatrix} M' \\ M'' \end{bmatrix}^T \right). \tag{9}$$

As shown in [37], the well known Hoffman's Lemma [30] in terms of constant (9) with $(\alpha, \beta) = (\infty, 2)$ is stated as follows.

Lemma 2.4. (Hoffman) *Let $M' \in \mathbb{R}^{m \times q}$ and $M'' \in \mathbb{R}^{\ell \times q}$ be given matrices and $\mathcal{F} = \{x \in \mathbb{R}^q : M'x \leq b, M''x = d\}$. For any vector x in $\mathbb{R}^q$, there is a point $x^* \in \mathcal{F}$ with*

$$\|x - x^*\|_2 \leq \sigma_{\infty,2}(M', M'') \left\| \begin{bmatrix} (M'x - b)^+ \\ M''x - d \end{bmatrix} \right\|_1.$$

The constant $\sigma_{\alpha,\beta}(M', M'')$, defined in (9), is referred to as the *Robinson's constant* determined by (M', M''). Given the solution set $\mathcal{F}$ of a linear system, Hoffman's error bound claims that the distance from a point in space to $\mathcal{F}$ can be measured in terms of the Robinson's constant and the quantity of the linear system being violated at this point.

In the remainder of the paper, we use Lemma 2.4 to develop a weak-stability theory for ℓ_1-minimization problems. The purpose of this study is to estimate the distance between an unknown vector (which is the target data to reconstruct) and the solution of the ℓ_1-minimization problem. Note that the solution set of a linear optimization problem is a polyhedron which can be represented as the solution set of a certain linear system by using the KKT optimality condition. From this observation, a recovery error bound via ℓ_1-minimization is similar to the Hoffman's error bound, although they are not completely the same since sparsity is also involved in sparse data reconstruction. However, this similarity or connection motivates one to use Hoffman's error bound combined with sparsity assumption to form a new analytic method for studying stability issues in sparse data reconstruction. This is different from the standard analytic methods in this area.

Our analysis not only provides a new tool to the study of stability issues of ℓ_1-minimization, but also makes it possible to go beyond the standard framework of methods (such as RIP and NSP based ones) in order to develop stability results under mild conditions or in general settings. As we have pointed out, most existing conditions can be relaxed to the assumption made in this paper. Traditional recovery error bounds are often established in terms of RIP constant, stable or robust stable NSP constant or their variants. Our assumption is a constant-free condition in the sense that the definition of this condition does not involve any constant that is difficult to certify. Under the constant-free weak RSP of A^T discussed in this paper, we use Robinson's constant to express stability coefficients in reconstruction error bounds. The error bound established under this assumption can apply to a wide range of matrix conditions, leading to a somewhat unified version of error bounds for sparse data reconstruction (see, e.g., Corollary 3.5). This is different from a standard analysis, which often requires an assumption-to-assumption analysis and the resulting error bounds often depends on an assumed individual assumption. Hoffman's Lemma and Robinson's constant provide a new perspective and an efficient way to interpret the sparse-signal-recovery behavior of ℓ_1-minimization methods.

3 Weak stability of ℓ_1-minimization in noise-free settings

In this section, we consider the case where the nonadaptive measurements $y \in \mathbb{R}^m$ are accurate, i.e., $y = A\widehat{x}$, where $\widehat{x} \in \mathbb{R}^m$ is the sparse data to reconstruct. The situation with inaccurate measurements will be discussed in later sections. Given a matrix A and the noiseless measurements y, the compressed sensing theory indicates that if A admits some strong property, the standard ℓ_1-minimization

$$\min\{\|x\|_1 : Ax = y\} \tag{10}$$

can exactly reconstruct the sparse data $\widehat{x}$ in the sense that the unique solution x^* of (10) coincides with $\widehat{x}$. In many situations, however, the data $\widehat{x}$ is not exactly sparse and it can only be claimed that $\widehat{x}$ is close to a sparse vector. In these situations, it is important to know whether the reconstruction is weakly stable. In section 2, we have shown that the weak RSP of order k of A^T is a necessary condition for standard ℓ_1-minimization with any given measurements $y \in \{Ax : \|x\|_0 \leq k\}$ to be weakly stable. In this section, we further show that this condition is also sufficient for the problem to be weakly stable. Note that the problem (10) can be written as the linear program

$$\min_{(x,t)}\{e^T t : \ Ax = y, \ -x + t \geq 0, \ x + t \geq 0, \ t \geq 0\}, \tag{11}$$

to which the dual problem is given as

$$\max_{(w,u,v)} \left\{ y^T w : \ A^T w - u + v = 0, \ u + v \leq e, \ (u,v) \geq 0 \right\}. \tag{12}$$

Thus, by the optimality conditions, the solution of (10) can be characterized as follows.

Lemma 3.1. *x^* is an optimal solution of (10) if and only if there exist vectors $t^*, u^*, v^* \in \mathbb{R}^n_+$ and $w^* \in \mathbb{R}^m$ such that $(x^*, t^*, u^*, v^*, w^*) \in D$, where*

$$\begin{aligned} D = \{(x,t,u,v,w) : \ & Ax = y, \ x \leq t, \ -x \leq t, \ A^T w - u + v = 0, \ u + v \leq e, \\ & y^T w = e^T t, \ (u,v,t) \geq 0\}. \end{aligned} \tag{13}$$

Moreover, any $(x,t,u,v,w) \in D$ satisfies that $t = |x|$.

The first assertion follows directly from the optimality conditions of (11) and (12). The second assertion is implied from (13) and can be directly seen from (11) as well. In fact, x^* is an optimal solution of (10) if and only if x^*, together with $t^* = |x^*|$, is an optimal solution of (11). Note that (13) is of the form

$$D = \{z = (x,t,u,v,w) : \ M'z \leq b, \ M''z = d\}, \tag{14}$$

where $b = (0,0,e,0,0,0)$ and $d = (y,0,0)$ and

$$M' = \begin{pmatrix} I & -I & 0 & 0 & 0 \\ -I & -I & 0 & 0 & 0 \\ 0 & 0 & I & I & 0 \\ 0 & 0 & -I & 0 & 0 \\ 0 & 0 & 0 & -I & 0 \\ 0 & -I & 0 & 0 & 0 \end{pmatrix}, \ M'' = \begin{pmatrix} A & 0 & 0 & 0 & 0 \\ 0 & 0 & -I & I & A^T \\ 0 & -e^T & 0 & 0 & y^T \end{pmatrix}. \tag{15}$$

In the remainder of the paper, we use c, c_1, c_2 to denote the following constants:

$$c = \|(AA^T)^{-1}A\|_{\infty\to\infty}, \ c_1 = \|(AA^T)^{-1}A\|_{\infty\to 1}, \ c_2 = \|(AA^T)^{-1}A\|_{\infty\to 2}. \tag{16}$$

We now prove the main result of this section.

Theorem 3.2. *Let $A \in \mathbb{R}^{m\times n}$ $(m < n)$ be a given matrix with rank$(A) = m$, and let y be any given vector in $\mathbb{R}^m$. If A^T satisfies the weak RSP of order k, then, for any $x \in \mathbb{R}^n$, there is an optimal solution x^* of (10) such that*

$$\|x - x^*\|_2 \leq \gamma \left\{2\sigma_k(x)_1 + (1+c)\|Ax - y\|_1\right\}, \tag{17}$$

where c is a constant given in (16), and $\gamma = \sigma_{\infty,2}(M', M'')$ is the Robinson's constant with (M', M'') given as (15). In particular, if x satisfies $Ax = y$, then there is an optimal solution x^* of (10) such that

$$\|x - x^*\|_2 \leq 2\gamma\sigma_k(x)_1. \tag{18}$$

Proof. Let x be any given vector in $\mathbb{R}^n$ and let $t = |x|$. Let S denote the support set of the k-largest components of $|x|$. Let $S = S_+ \cup S_-$, where $S_+ = \{i \in S : x_i > 0\}$ and $S_- = \{i \in S : x_i < 0\}$. We now construct a vector $(\widetilde{u}, \widetilde{v}, \widetilde{w})$ such that it is a feasible point to the problem (12). Since A^T has the weak RSP of order k, there exists a vector $\eta \in \mathcal{R}(A^T)$ such that $A^T\widetilde{w} = \eta$ for some $\widetilde{w} \in \mathbb{R}^m$ and η satisfies that

$$\eta_i = 1 \text{ for } i \in S_+, \ \eta_i = -1 \text{ for } i \in S_-, \ |\eta_i| \leq 1 \text{ for } i \notin S = S_+ \cup S_-,$$

from which we see that $(A^T\widetilde{w})_S = \eta_S = \mathrm{sign}(x_S)$. We construct $(\widetilde{u}, \widetilde{v})$ as follows: $\widetilde{u}_i = 1$ and $\widetilde{v}_i = 0$ for $i \in S_+$; $\widetilde{u}_i = 0$ and $\widetilde{v}_i = 1$ for $i \in S_-$; $\widetilde{u}_i = (|\eta_i| + \eta_i)/2$ and $\widetilde{v}_i = (|\eta_i| - \eta_i)/2$ for all $i \notin S$. From this construction, $(\widetilde{u}, \widetilde{v})$ satisfies that $(\widetilde{u}, \widetilde{v}) \geq 0$, $\widetilde{u} + \widetilde{v} \leq e$ and $A^T\widetilde{w} = \eta = \widetilde{u} - \widetilde{v}$. Thus $(\widetilde{u}, \widetilde{v}, \widetilde{w})$ is a feasible vector to the problem (12). We now estimate the distance of $(x, t, \widetilde{u}, \widetilde{v}, \widetilde{w})$ to the set D given by (13) which can be written as (14). By applying Lemma 2.4 to (14), for the point $(x, t, \widetilde{u}, \widetilde{v}, \widetilde{w})$, where $t = |x|$, there exists a point $(x^*, t^*, u^*, v^*, w^*) \in D$ such that

$$\left\| \begin{bmatrix} x \\ t \\ \widetilde{u} \\ \widetilde{v} \\ \widetilde{w} \end{bmatrix} - \begin{bmatrix} x^* \\ t^* \\ u^* \\ v^* \\ w^* \end{bmatrix} \right\|_2 \leq \gamma \left\| \begin{bmatrix} (x-t)^+ \\ (-x-t)^+ \\ (\widetilde{u} + \widetilde{v} - e)^+ \\ Ax - y \\ A^T\widetilde{w} - \widetilde{u} + \widetilde{v} \\ e^T t - y^T \widetilde{w} \\ (\vartheta)^- \end{bmatrix} \right\|_1, \qquad (19)$$

where $(\vartheta)^-$ denotes the vector $((\widetilde{u})^-, (\widetilde{v})^-, (t)^-)$, and $\gamma = \sigma_{\infty,2}(M', M'')$ is the Robinson's constant determined by (M', M'') given as (15). By the choice of $(\widetilde{u}, \widetilde{v}, \widetilde{w})$ and the fact $t = |x|$, we have

$$(x-t)^+ = (-x-t)^+ = 0, \ (\widetilde{u} + \widetilde{v} - e)^+ = 0, \ A^T\widetilde{w} - \widetilde{u} + \widetilde{v} = 0, \ (\vartheta)^- = 0.$$

Thus the inequality (19) is reduced to

$$\|(x, t, \widetilde{u}, \widetilde{v}, \widetilde{w}) - (x^*, t^*, u^*, v^*, w^*)\|_2 \quad \leq \quad \gamma \left\| \begin{bmatrix} Ax - y \\ e^T t - y^T \widetilde{w} \end{bmatrix} \right\|_1. \qquad (20)$$

Denote by $h = Ax - y$. By the choice of $(t, \widetilde{u}, \widetilde{v}, \widetilde{w})$, we see that

$$e^T t - y^T \widetilde{w} = e^T |x| - (Ax - h)^T \widetilde{w} = \|x\|_1 - x^T(A^T\widetilde{w}) + h^T\widetilde{w}.$$

Substituting this into (20) and noting that

$$\|x - x^*\|_2 \leq \|(x, t, \widetilde{u}, \widetilde{v}, \widetilde{w}) - (x^*, t^*, u^*, v^*, w^*)\|_2,$$

we obtain

$$\|x - x^*\|_2 \leq \gamma \left\{ \|Ax - y\|_1 + \left| \|x\|_1 - x^T(A^T\widetilde{w}) + h^T\widetilde{w} \right| \right\}. \qquad (21)$$

Note that A has full row rank and $\|\eta\|_\infty \leq 1$. From $A^T\widetilde{w} = \eta$, we see that

$$\|\widetilde{w}\|_\infty = \|(AA^T)^{-1}A\eta\|_\infty \leq \|(AA^T)^{-1}A\|_{\infty\to\infty}\|\eta\|_\infty \leq c, \qquad (22)$$

where c is a constant given in (16). Note that

$$(x_S)^T(A^T\widetilde{w})_S = (x_S)^T\eta_S = (x_S)^T\mathrm{sign}(x_S) = \|x_S\|_1.$$

Therefore,

$$
\begin{aligned}
\left| \|x\|_1 - x^T(A^T\widetilde{w}) + h^T\widetilde{w} \right| &= \left| \|x\|_1 - (x_S)^T(A^T\widetilde{w})_S - (x_{\overline{S}})^T(A^T\widetilde{w})_{\overline{S}} + h^T\widetilde{w} \right| \\
&= \left| \|x\|_1 - \|x_S\|_1 - (x_{\overline{S}})^T(A^T\widetilde{w})_{\overline{S}} + h^T\widetilde{w} \right| \\
&= \left| \sigma_k(x)_1 - (x_{\overline{S}})^T(A^T\widetilde{w})_{\overline{S}} + h^T\widetilde{w} \right| \\
&\leq \sigma_k(x)_1 + \left| (x_{\overline{S}})^T(A^T\widetilde{w})_{\overline{S}} \right| + |h^T\widetilde{w}| \\
&\leq 2\sigma_k(x)_1 + \|h\|_1 \cdot \|\widetilde{w}\|_\infty \\
&\leq 2\sigma_k(x)_1 + c\|Ax - y\|_1, \qquad (23)
\end{aligned}
$$

where the second inequality follows from the fact

$$
\left| (x_{\overline{S}})^T(A^T\widetilde{w})_{\overline{S}} \right| \leq \|x_{\overline{S}}\|_1 \left\| (A^T\widetilde{w})_{\overline{S}} \right\|_\infty = \|x_{\overline{S}}\|_1 \|\eta_{\overline{S}}\|_\infty \leq \|x_{\overline{S}}\|_1 = \sigma_k(x)_1,
$$

and the final inequality follows from (22). Substituting (23) into (21) yields the estimate (17), as desired. In particular, if x is a solution to the underdetermined linear system $Az = y$, then (17) is reduced to (18). $\square$

Under the weak RSP of order k of A^T, Theorem 3.2 indicates that the standard ℓ_1-minimization problem, i.e., problem (1) with $\varepsilon = 0$, is weakly stable for any given $y \in \mathbb{R}^m (= \{Ax : x \in \mathbb{R}^n\}$ since A is underdetermined with full row rank). In particular, it is weakly stable for any given $y \in \{Ax : \|x\|_0 \leq k\} \subseteq \mathbb{R}^m$. Theorem 2.3 indicates that if the standard ℓ_1-minimization problem is weakly stable for any given $y \in \{Ax : \|x\|_0 \leq k\}$, then A^T must satisfy the weak RSP of order k. Merging Theorems 2.3 and 3.2 immediately yields the following statement.

Corollary 3.3. *Let $A \in \mathbb{R}^{m\times n} (m < n)$ be a matrix with rank$(A) = m$. Then the standard ℓ_1-minimization problem $\min\{\|x\|_1 : Ax = y\}$ is weakly stable in sparse data reconstruction for any given measurements $y \in \{Ax : \|x\|_0 \leq k\}$ if and only if A^T satisfies the weak RSP of order k.*

Thus the weak RSP of A^T is the mildest condition, which cannot be relaxed without damaging the weak stability of ℓ_1-minimization problems.

Remark 3.4. Uniform recovery requires that every k-sparse vector can be reconstructed by ℓ_1-minimization. This means that every k-sparse vector is an optimal solution of ℓ_1-minimization. Then the classic KKT optimality condition naturally yields the matrix property of weak RSP of A^T. Therefore, no matter what (deterministic or random) design matrix A is used, the weak RSP of order k of A^T is a fundamental property required for achieving the uniform recovery with ℓ_1-minimization as a decoding method. The existence of a matrix with such a property follows directly from that of RIP matrices. We recall the following fact: (Candés, Tao, etc.) *Let A be an $m \times n$ Gaussian or Bernoulli random matrix. Then there exists a universal constant $C > 0$ such that the RIP constant of $A/\sqrt{m}$ satisfies $\delta_{2k} \leq \xi$ (where $0 < \xi < 1$) with probability at least $1 - \epsilon$ provided*

$$
m \geq C\xi^{-2}\left(k(1 + \ln(n/k)) + \ln(2\epsilon^{-1})\right).
$$

This fact was first shown by Candès and Tao [10] and it was improved later to the above statement by Candès and other researchers. Taking $\xi = 1/\sqrt{2}$, Cai et al. [6] have shown that the RIP of order $2k$ with constant $\delta_{2k} < 1/\sqrt{2}$ guarantees the uniform recovery of k-parse vectors via ℓ_1-minimization method. Note that the uniform recovery of k-parse vectors via ℓ_1-minimization is equivalent to that A^T satisfies the RSP of order k (see [45] for details), and hence A^T satisfies the weak RSP of order k. Combining these facts and taking $\xi = 1/\sqrt{2}$, we immediately obtain the following statement: *Let A be an $m \times n$ Gaussian or Bernoulli random matrix. Then there exists a universal constant $C > 0$ such that $A^T/\sqrt{m}$ satisfies the weak RSP of order k with probability at least $1 - \epsilon$ provided*

$$m \geq 2C\left(k(1 + \ln(n/k)) + \ln(2\epsilon^{-1})\right). \tag{24}$$

By Theorem 3.2, when A^T satisfies the weak RSP of order k, the error bound (18) always holds. Combining Theorem 3.2 and the above statements yields the following fact: *Let A be an $m \times n$ $(m < n)$ Gaussian or Bernoulli random matrix with full row rank, and let y be a given vector in $\mathbb{R}^m$. Then there exists a universal constant $C > 0$ such that with probability at least $1 - \epsilon$, the standard ℓ_1-minimization problem with matrix $A/\sqrt{m}$ is weakly stable, provided that (24) is satisfied.*

From Theorem 3.2, we obtain a unified stability result for several existing matrix properties.

Corollary 3.5. *Let (A, y) be given, where $y \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m \times n}$ $(m < n)$ with rank$(A) = m$. Suppose that A admits one of the following properties:*

(p1) *RIP of order $2k$ with constant $\delta_{2k} < 1/\sqrt{2}$.*

(p2) *A is a matrix with ℓ_2-normalized columns and $\mu_1(k) + \mu_1(k-1) < 1$, where $\mu_1(k)$ is the accumulated mutual coherence.*

(p3) *The stable NSP of order k with constant $0 < \rho < 1$.*

(p4) *The robust NSP of order k with constant $0 < \rho < 1$ and $\tau > 0$.*

(p5) *The NSP of order k.*

(p6) *The RSP of order k of A^T.*

Then, for any $x \in \mathbb{R}^n$, the optimal solution x^* of (10) approximates x with error

$$\|x - x^*\|_2 \leq 2\gamma\sigma_k(x)_1 + \gamma(1 + c)\|Ax - y\|_1,$$

where c is a constant given in (16) and $\gamma = \sigma_{\infty,2}(M', M'')$ is the Robinson's constant determined by (15). In particular, for any x with $Ax = y$, the optimal solution x^* of (10) approximates x with error $\|x - x^*\|_2 \leq 2\gamma\sigma_k(x)_1$.

The above corollary follows directly from Theorem 3.2, since each of the properties (p1)–(p6) implies the weak RSP of order k of A^T as well as the uniqueness of the optimal solution x^* of (10). Corollary 3.5 is a unified weak stability result in the sense that every matrix property of (p1)–(p6) implies the same error bound in terms of the Robinson's constant. The weak stability result of this type is new and established in this paper for the first time.

4 Robust weak stability of linearly constrained models

In more realistic situations, the measurements y for the unknown sparse data $\widehat{x} \in \mathbb{R}^n$ are inaccurate, and thus $y = A\widehat{x} + u$, where u denotes the measurement error satisfying $\|u\| \leq \varepsilon$ for some norm $\|\cdot\|$ and noise level $\varepsilon > 0$. Thus we consider the robust weak stability of (1) with a known level $\varepsilon > 0$. In this section, we focus on the following problems:

$$\min\{\|x\|_1 : \quad \|Ax - y\|_\infty \leq \varepsilon\}, \tag{25}$$

$$\min\{\|x\|_1 : \quad \|Ax - y\|_1 \leq \varepsilon\}, \tag{26}$$

corresponding to $p = \infty$ and $p = 1$ in (1), respectively. The case $p = 2$ in (1) will be treated separately in section 5. Problems (25) and (26) are referred to as the ℓ_1-minimization with ℓ_∞-norm and ℓ_1-norm constraints, respectively. A common feature of (25) and (26) is that their constraints can be linearly represented. This structure makes it possible to extend the approach in section 3 to establish the robust weak stability of (25) and (26).

4.1 ℓ_1-minimization with ℓ_∞-norm constraint

We first consider the problem (25), which can be written as

$$\min_{(x,t)} \left\{ e^T t : \; -x + t \geq 0, \; x + t \geq 0, \; t \geq 0, \; -\varepsilon e \leq Ax - y \leq \varepsilon e \right\} \tag{27}$$

to which the dual problem is given as

$$\max_{(u,v,w,w')} \left\{ (y - \varepsilon e)^T w - (y + \varepsilon e)^T w' : \; A^T(w - w') = u - v, \; u + v \leq e, \; (u, v, w, w') \geq 0 \right\}. \tag{28}$$

Clearly, x^* is an optimal solution of (25) if and only if (x^*, t^*) with $t^* = |x^*|$ is an optimal solution of (27). By the optimality condition of a linear program, we can immediately characterize the solution set of (25) as follows.

Lemma 4.1. *x^* is an optimal solution of (25) if and only if there exist vectors t^*, u^*, v^* in $\mathbb{R}^n_+$ and w^*, w'^* in $\mathbb{R}^m_+$ such that $(x^*, t^*, u^*, v^*, w^*, w'^*) \in D^{(\infty)}$ where*

$$\begin{aligned} D^{(\infty)} = \{(x, t, u, v, w, w') : \quad & -x + t \geq 0, \; x + t \geq 0, \; -\varepsilon e \leq Ax - y \leq \varepsilon e, \\ & A^T(w - w') = u - v, \; u + v \leq e, \\ & e^T t = (y - \varepsilon e)^T w - (y + \varepsilon e)^T w', \\ & (t, u, v, w, w') \geq 0\}. \end{aligned} \tag{29}$$

Moreover, for any $(x, t, u, v, w, w') \in D^{(\infty)}$, it must hold that $t = |x|$.

The set $D^{(\infty)}$ can be written as

$$D^{(\infty)} = \{z = (x, t, u, v, w, w') : \; M^{(1)} z \leq b^{(1)}, \; M^{(2)} z = b^{(2)}\}, \tag{30}$$

where $b^{(2)} = 0$ and

$$M^{(1)} = \begin{pmatrix} I & -I & 0 & 0 & 0 & 0 \\ -I & -I & 0 & 0 & 0 & 0 \\ A & 0 & 0 & 0 & 0 & 0 \\ -A & 0 & 0 & 0 & 0 & 0 \\ 0 & -I & 0 & 0 & 0 & 0 \\ 0 & 0 & I & I & 0 & 0 \\ 0 & 0 & 0 & 0 & -I_m & 0 \\ 0 & 0 & 0 & 0 & 0 & -I_m \\ 0 & 0 & -I & 0 & 0 & 0 \\ 0 & 0 & & -I & 0 & 0 \end{pmatrix}, \quad b^{(1)} = \begin{pmatrix} 0 \\ 0 \\ y + \varepsilon e \\ \varepsilon e - y \\ 0 \\ e \\ 0 \\ 0 \\ 0 \end{pmatrix}, \tag{31}$$

$$M^{(2)} = \begin{pmatrix} 0 & 0 & -I & I & A^T & -A^T \\ 0 & e^T & 0 & 0 & -(y-\varepsilon e)^T & (y+\varepsilon e)^T \end{pmatrix}, \tag{32}$$

where I and I_m are n- and m-dimensional identity matrices, respectively. We now show that the robust weak stability of (25) is guaranteed under the weak RSP of order k of A^T.

Theorem 4.2. *Let the problem data (A, y, ε) of (25) be given, where $\varepsilon > 0$, $y \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m\times n}$ $(m < n)$ with rank$(A) = m$. Let A^T satisfy the weak RSP of order k. Then for any $x \in \mathbb{R}^n$, there is an optimal solution x^* of (25) such that*

$$\|x - x^*\|_2 \leq \gamma_1 \left\{ \|(Ax - y - \varepsilon e)^+\|_1 + \|(Ax - y + \varepsilon e)^-\|_1 + 2\sigma_k(x)_1 + c_1\varepsilon + c_1\|Ax - y\|_\infty \right\},$$

where c_1 is the constant given in (16) and $\gamma_1 = \sigma_{\infty,2}(M^{(1)}, M^{(2)})$ is the Robinson's constant determined by $(M^{(1)}, M^{(2)})$ given in (31) and (32). In particular, for any x with $\|Ax - y\|_\infty \leq \varepsilon$, there is a solution x^* of (25) such that

$$\|x - x^*\|_2 \leq 2\gamma_1 \{\sigma_k(x)_1 + c_1\varepsilon\}.$$

Proof. For any given $x \in \mathbb{R}^n$, we consider a vector (t, u, v, w, w') satisfying the following properties: $t = |x|$ and (u, v, w, w') satisfies $A^T(w - w') = u - v$, $u + v \leq e$ and $(u, v, w, w') \geq 0$, i.e., (u, v, w, w') is a feasible vector to problem (28). Note that the set (29) can be written as (30). For such a vector (x, t, u, v, w, w'), applying Lemma 2.4 with $(M', M'') = (M^{(1)}, M^{(2)})$ being given in (31) and (32), there must exist a vector $(x^*, t^*, u^*, v^*, w^*, w'^*) \in D^{(\infty)}$ such that

$$\left\| \begin{bmatrix} x \\ t \\ u \\ v \\ w \\ w' \end{bmatrix} - \begin{bmatrix} x^* \\ t^* \\ u^* \\ v^* \\ w^* \\ w'^* \end{bmatrix} \right\|_2 \leq \gamma_1 \left\| \begin{bmatrix} (x - t)^+ \\ (-x - t)^+ \\ (Ax - y - \varepsilon e)^+ \\ (Ax - y + \varepsilon e)^- \\ A^T(w - w') - u + v \\ (u + v - e)^+ \\ e^T t - (y - \varepsilon e)^T w + (y + \varepsilon e)^T w' \\ (\widehat{\vartheta})^- \end{bmatrix} \right\|_1, \tag{33}$$

where $(\widehat{\vartheta})^-$ is short for the vector $((u)^-, (v)^-, (t)^-, (w)^-, (w')^-)$, and $\gamma_1 = \sigma_{\infty,2}(M^{(1)}, M^{(2)})$ is the Robinson's constant with $(M^{(1)}, M^{(2)})$ being given by (31) and (32). By the nonnegativity of (u, v, t, w, w'), we see that $(\widehat{\vartheta})^- = 0$. Since $t = |x|$ and (u, v, w, w') is feasible to problem (28), we see that

$$(x - t)^+ = (-x - t)^+ = 0,\ A^T(w - w') - u + v = 0,\ (u + v - e)^+ = 0.$$

Thus the system (33) is reduced to

$$\|(x, t, u, v, w, w') - (x^*, t^*, u^*, v^*, w^*, w'^*)\|_2 \leq \gamma_1 \left\| \begin{bmatrix} (Ax - y - \varepsilon e)^+ \\ (Ax - y + \varepsilon e)^- \\ e^T t - (y - \varepsilon e)^T w + (y + \varepsilon e)^T w' \end{bmatrix} \right\|_1.$$

Let $\phi = y - Ax$. We see that

$$\begin{aligned} e^T t - (y - \varepsilon e)^T w + (y + \varepsilon e)^T w' &= e^T t - y^T(w - w') + \varepsilon e^T(w + w') \\ &= e^T|x| - (Ax + \phi)^T(w - w') + \varepsilon e^T(w + w') \\ &= e^T|x| - x^T A^T(w - w') - \phi^T(w - w') + \varepsilon e^T(w + w'). \end{aligned}$$

Merging the above two relations leads to

$$\begin{aligned}
&\|(x,t,u,v,w,w') - (x^*,t^*,u^*,v^*,w^*,w'^*)\|_2 \\
&\le \gamma_1\{\|(Ax-y-\varepsilon e)^+\|_1 + \|(Ax-y+\varepsilon e)^-\|_1 \\
&+ \left|e^T|x| - x^T A^T(w-w') - \phi^T(w-w') + \varepsilon e^T(w+w')\right|\}.
\end{aligned} \tag{34}$$

By the weak RSP of order k of A^T, we now construct a specific vector $(\widetilde{u},\widetilde{v},\widetilde{w},\widetilde{w}')$ which is feasible to problem (28). To this goal, let S denote the support set of the k-largest components of $|x|$. Let $S_+ = \{i \in S : x_i > 0\}$ and $S_- = \{i \in S : x_i < 0\}$. Clearly, $S = S_+ \cup S_-$. Since A^T satisfies the weak RSP of order k, there exists a vector $\eta \in \mathcal{R}(A^T)$ such that $\eta = A^T g$ for some $g \in \mathbb{R}^m$ and η satisfies the following conditions:

$$\eta_i = 1 \text{ for } i \in S_+,\ \eta_i = -1 \text{ for all } i \in S_-, \ \text{ and } |\eta_1| \le 1 \text{ for } i \in \overline{S} = \{1,\dots,n\} \setminus S.$$

Construct $(\widetilde{u},\widetilde{v})$ as follows: $\widetilde{u}_i = 1$ and $\widetilde{v}_i = 0$ for $i \in S_+$; $\widetilde{u}_i = 0$ and $\widetilde{v}_i = 1$ for $i \in S_-$; $\widetilde{u}_i = (1+\eta_i)/2$ and $\widetilde{v}_i = (1-\eta_i)/2$ for all $i \in \overline{S}$. By this construction, we see that $\widetilde{u} - \widetilde{v} = \eta$. Moreover, by setting $\widetilde{w} = (g)^+$ and $\widetilde{w}' = -(g)^-$, we see that $\widetilde{w} \ge 0$, $\widetilde{w}' \ge 0$, $\widetilde{w} - \widetilde{w}' = g$. It is easy to see that the vector $(\widetilde{u},\widetilde{v},\widetilde{w},\widetilde{w}')$ specified as above satisfies the conditions

$$\widetilde{u} + \widetilde{v} \le e,\ A^T(\widetilde{w} - \widetilde{w}') = \widetilde{u} - \widetilde{v},\ (\widetilde{u},\widetilde{v},\widetilde{w},\widetilde{w}') \ge 0$$

which indicates that $(\widetilde{u},\widetilde{v},\widetilde{w},\widetilde{w}')$ is a feasible vector to problem (28). Thus it follows from (34) that for the vector $(x, t = |x|, \widetilde{u},\widetilde{v},\widetilde{w},\widetilde{w}')$, there is a point in $D^{(\infty)}$, denoted still by $(x^*,t^*,u^*,v^*,w^*,w'^*)$, such that

$$\begin{aligned}
&\|(x,t,\widetilde{u},\widetilde{v},\widetilde{w},\widetilde{w}') - (x^*,t^*,u^*,v^*,w^*,w'^*)\|_2 \\
&\le \gamma_1\{\|(Ax-y-\varepsilon e)^+\|_1 + \|(Ax-y+\varepsilon e)^-\|_1 \\
&+ \left|e^T|x| - x^T A^T(\widetilde{w}-\widetilde{w}') - \phi^T(\widetilde{w}-\widetilde{w}') + \varepsilon e^T(\widetilde{w}+\widetilde{w}')\right|\}.
\end{aligned} \tag{35}$$

By the construction of $(\widetilde{u},\widetilde{v},\widetilde{w},\widetilde{w}')$, we see that $[A^T(\widetilde{w}-\widetilde{w}')]_S = (\widetilde{u}-\widetilde{v})_S = \eta_S = \text{sign}(x_S)$. Thus

$$\begin{aligned}
&\left|e^T|x| - x^T A^T(\widetilde{w}-\widetilde{w}') - \phi^T(\widetilde{w}-\widetilde{w}') + \varepsilon e^T(\widetilde{w}+\widetilde{w}')\right| \\
&= \left|\|x\|_1 - (x_S)^T[A^T(\widetilde{w}-\widetilde{w}')]_S - (x_{\overline{S}})^T[A^T(\widetilde{w}-\widetilde{w}')]_{\overline{S}} - \phi^T(\widetilde{w}-\widetilde{w}') + \varepsilon e^T(\widetilde{w}+\widetilde{w}')\right| \\
&= \left|\|x\|_1 - \|x_S\|_1 - (x_{\overline{S}})^T[A^T(\widetilde{w}-\widetilde{w}')]_{\overline{S}} - \phi^T(\widetilde{w}-\widetilde{w}') + \varepsilon e^T(\widetilde{w}+\widetilde{w}')\right| \\
&\le \sigma_k(x)_1 + |(x_{\overline{S}})^T[A^T(\widetilde{w}-\widetilde{w}')]_{\overline{S}}| + |\phi^T(\widetilde{w}-\widetilde{w}')| + \varepsilon|e^T(\widetilde{w}+\widetilde{w}')| \\
&\le \sigma_k(x)_1 + \|(x_{\overline{S}})\|_1\|[A^T(\widetilde{w}-\widetilde{w}')]_{\overline{S}}\|_\infty + |\phi^T g| + \varepsilon|e^T((g)^+ - (g)^-)| \\
&\le 2\sigma_k(x)_1 + \|g\|_1\|\phi\|_\infty + \varepsilon\|g\|_1,
\end{aligned} \tag{36}$$

where the last inequality follows from the fact $\|[A^T(\widetilde{w}-\widetilde{w}')]_{\overline{S}}\|_\infty = \|\eta_{\overline{S}}\|_\infty \le 1$. Since A^T has full column rank, it follows from $A^T g = \eta$ that $g = (AA^T)^{-1}A\eta$, and hence

$$\|g\|_1 = \|(AA^T)^{-1}A\eta\|_1 \le \|(AA^T)^{-1}A\|_{\infty\to 1}\|\eta\|_\infty \le \|(AA^T)^{-1}A\|_{\infty\to 1} = c_1. \tag{37}$$

Merging (35), (36) and (37) yields the bound

$$\begin{aligned}
\|x-x^*\|_2 &\le \|(x,t,\widetilde{u},\widetilde{v},\widetilde{w},\widetilde{w}') - (x^*,t^*,u^*,v^*,w^*,w'^*)\|_2 \\
&\le \gamma_1\left\{\|(Ax-y-\varepsilon e)^+\|_1 + \|(Ax-y+\varepsilon e)^-\|_1 + 2\sigma_k(x)_1 + \|g\|_1(\|\phi\|_\infty + \varepsilon)\right\} \\
&\le \gamma_1\left\{\|(Ax-y-\varepsilon e)^+\|_1 + \|(Ax-y+\varepsilon e)^-\|_1 + 2\sigma_k(x)_1 + c_1\|y-Ax\|_\infty + c_1\varepsilon\right\},
\end{aligned}$$

as desired. In particular, when x satisfies the constraint of (25), i.e., $\|y-Ax\|_\infty \le \varepsilon$, the above estimate reduces to $\|x-x^*\|_2 \le 2\gamma_1\{\sigma_k(x)_1 + c_1\varepsilon\}$. $\square$

4.2 ℓ_1-minimization with ℓ_1-norm constraint

We now show the robust weak stability of problem (26). Note that (26) is equivalent to

$$\min_{(x,r)} \left\{ \|x\|_1 : \ |Ax-y| \le r, \ e^T r \le \varepsilon, \ r \in \mathbb{R}^m_+ \right\}. \tag{38}$$

It is evident that x^* is an optimal solution of (26) if and only if there is a vector r^* such that (x^*, r^*) is an optimal solution of (38). We may further write (38) as the linear program

$$\min_{(x,t,r)} \left\{ e^T t : \ x \le t, \ -x \le t, \ t \ge 0, \ Ax - y \le r, -Ax + y \le r, \ e^T r \le \varepsilon, \ r \ge 0 \right\}. \tag{39}$$

The dual problem of (39) is given by

$$\begin{array}{ll} \max & -y^T(v_3 - v_4) - \varepsilon v_5 \\ \text{s.t} & A^T(v_3 - v_4) + v_1 - v_2 = 0, \ v_3 + v_4 \le v_5 e, \ v_1 + v_2 \le e, \\ & v_i \ge 0, \ i = 1, \ldots, 5, \end{array} \tag{40}$$

where $v_1, v_2 \in \mathbb{R}^n_+, v_3, v_4 \in \mathbb{R}^m_+$, and $v_5 \in \mathbb{R}_+$. By the optimality condition of a linear program, the solution set of (26) can be characterized as follows.

Lemma 4.3. *x^* is an optimal solution of (26) if and only if there exist vectors $t^*, v_1^*, v_2^* \in \mathbb{R}^n_+, v_3^*, v_4^*, r^* \in \mathbb{R}^m_+$ and $v_5^* \in \mathbb{R}_+$ such that $(x^*, t^*, r^*, v_1^*, \ldots, v_5^*) \in D^{(1)}$, where*

$$D^{(1)} = \{(x,t,r,v_1,\ldots,v_5) : \begin{array}{l} x \le t, \ \ -x \le t, \ Ax - r \le y, \ \ -Ax - r \le -y, \\ e^T r \le \varepsilon, \ (r,t) \ge 0, \ A^T(v_3 - v_4) + v_1 - v_2 = 0, \\ v_3 + v_4 - v_5 e \le 0, \ v_1 + v_2 \le e, \\ e^T t = -y^T(v_3 - v_4) - v_5\varepsilon, \ v_i \ge 0, \ i = 1, \ldots, 5\}. \end{array} \tag{41}$$

Moreover, for any $(x,t,r,v_1,\ldots,v_5) \in D^{(1)}$, it must hold that $t = |x|$.

In order to apply Lemma 2.4 in the proof of the next theorem, we rewrite $D^{(1)}$ as

$$D^{(1)} = \{z = (x,t,r,v_1,\ldots,v_5) : \ M^* z \le b^*, \ M^{**} z = b^{**}\}, \tag{42}$$

where $b^{**} = 0, b^*$ is a vector consisting of $0, y, -y, e$ and ε, The matrix M^* captures all coefficients of the inequalities in (41), and M^{**} is the matrix capturing all coefficients of the equalities in (41). The entries of M^* and M^{**} are given by the problem data (A, y, ε). M^* and M^{**} are omitted here. We have the following stability result.

Theorem 4.4. *Let the problem data (A, y, ε) of (26) be given, where $\varepsilon > 0$, $y \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m \times n}$ $(m < n)$ with rank$(A) = m$. Let A^T satisfy the weak RSP of order k. Then for any $x \in \mathbb{R}^n$, there is an optimal solution x^* of (26) such that*

$$\|x - x^*\|_2 \le \gamma_2 \left\{ 2\sigma_k(x)_1 + (\|Ax - y\|_1 - \varepsilon)^+ + c(\varepsilon + \|y - Ax\|_1) \right\},$$

where c is the constant given in (16), and $\gamma_2 = \sigma_{\infty,2}(M^*, M^{**})$ is the Robinson's constant determined by (M^*, M^{**}) in (42). In particular, for any x with $\|Ax - y\|_1 \le \varepsilon$, there is an optimal solution x^* of (26) such that

$$\|x - x^*\|_2 \le 2\gamma_2 \{\sigma_k(x)_1 + c\varepsilon\}. \tag{43}$$

Proof. Let x be any vector in $\mathbb{R}^n$, and let $(t, r, v_1, \ldots, v_5)$ satisfy the following properties: $t = |x|$, $r = |Ax - y|$, and $(v_1, \ldots, v_5)$ is feasible to (40), i.e.,

$$A^T(v_3 - v_4) + v_1 - v_2 = 0, \; v_1 + v_2 \leq e, \; v_3 + v_4 \leq v_5 e, \; (v_1, \ldots, v_5) \geq 0.$$

For such a vector $(x, t, r, v_1, \ldots, v_5)$, applying Lemma 2.4 with $(M', M'') = (M^*, M^{**})$ where M^* and M^{**} are the matrices in (42), there exists a point $(x^*, t^*, r^*, v_1^*, \ldots, v_5^*)$ in $D^{(1)}$ defined by (41) such that

$$\left\| \begin{bmatrix} x \\ t \\ r \\ v_1 \\ \vdots \\ v_5 \end{bmatrix} - \begin{bmatrix} x^* \\ t^* \\ r^* \\ v_1^* \\ \vdots \\ v_5^* \end{bmatrix} \right\|_2 \leq \gamma_2 \left\| \begin{bmatrix} (x-t)^+ \\ (-x-t)^+ \\ (Ax-y-r)^+ \\ (Ax-y+r)^- \\ (e^T r - \varepsilon)^+ \\ A^T(v_3 - v_4) + v_1 - v_2 \\ (v_1 + v_2 - e)^+ \\ (v_3 + v_4 - v_5 e)^+ \\ e^T t + y^T(v_3 - v_4) + v_5 \varepsilon \\ (\vartheta^*)^- \end{bmatrix} \right\|_1, \tag{44}$$

where $(\vartheta^*)^-$ is the short for the vector $((t)^-, (r)^-, (v_1)^-, \ldots, (v_5)^-)$, and $\gamma_2 = \sigma_{\infty,2}(M^*, M^{**})$ is the Robinson's constant determined by the matrices (M^*, M^{**}) in (42). By the nonnegativity of $(t, r, v_1, ..., v_5)$, we see that $(\vartheta^*)^- = 0$. Since $(v_1, \ldots, v_5)$ is feasible to (40), we also have

$$(x-t)^+ = (-x-t)^+ = 0, \; A^T(v_3 - v_4) + v_1 - v_2 = 0, \; (v_1 + v_2 - e)^+ = 0,$$

$$(v_3 + v_4 - v_5 e)^+ = 0, \; (Ax - y - r)^+ = (Ax - y + r)^- = 0.$$

Thus the inequality (44) is reduced to

$$\|(x, t, r, v_1, \ldots, v_5) - (x^*, t^*, r^*, v_1^*, \ldots, v_5^*)\|_2 \leq \gamma_2 \left\| \begin{bmatrix} (e^T r - \varepsilon)^+ \\ e^T t + y^T(v_3 - v_4) + v_5 \varepsilon \end{bmatrix} \right\|_1. \tag{45}$$

Furthermore, letting $\phi = y - Ax$, we see that

$$\begin{aligned} e^T t + y^T(v_3 - v_4) + v_5 \varepsilon &= e^T|x| + (Ax + \phi)^T(v_3 - v_4) + v_5 \varepsilon \\ &= e^T|x| + x^T A^T(v_3 - v_4) + \phi^T(v_3 - v_4) + v_5 \varepsilon. \end{aligned} \tag{46}$$

Merging (45) and (46) leads to

$$\begin{aligned} &\|(x, t, r, v_1, \ldots, v_5) - (x^*, t^*, r^*, v_1^*, \ldots, v_5^*)\|_2 \\ &\leq \gamma_2 \left\{ (e^T r - \varepsilon)^+ + \left| e^T|x| + x^T A^T(v_3 - v_4) + \phi^T(v_3 - v_4) + v_5 \varepsilon \right| \right\}. \end{aligned} \tag{47}$$

We now construct a specific vector $(\widetilde{v}_1, \ldots, \widetilde{v}_5)$ which is feasible to problem (40). To this goal, we still let S be the support set of the k-largest components of $|x|$, and we still decompose S as $S = S_+ \cup S_-$, where $S_+ = \{i \in S : x_i > 0\}$ and $S_- = \{i \in S : x_i < 0\}$. Let $\overline{S} = \{1, \ldots, n\} \setminus S$. Since A^T satisfies the weak RSP of order k, there exists a vector $\eta = A^T g$ for some $g \in \mathbb{R}^m$ satisfying that $\eta_i = 1$ for $i \in S_+$, $\eta_i = -1$ for $i \in S_-$, and $|\eta_i| \leq 1$ for $i \in \overline{S}$. Define the vectors $\widetilde{v}_1$ and $\widetilde{v}_2$ as follows: $(\widetilde{v}_1)_i = 1$ and $(\widetilde{v}_2)_i = 0$ for $i \in S_+$; $(\widetilde{v}_1)_i = 0$ and $(\widetilde{v}_2)_i = 1$ for $i \in S_-$; $(\widetilde{v}_1)_i = (|\eta_i| + \eta_i)/2$ and $(\widetilde{v}_2)_i = (|\eta_i| - \eta_i)/2$ for all $i \in \overline{S}$. This construction ensures that $(\widetilde{v}_1, \widetilde{v}_2) \geq 0$, $\widetilde{v}_1 + \widetilde{v}_2 \leq e$, and $\widetilde{v}_1 - \widetilde{v}_2 = \eta$. Moreover, by setting

$$\widetilde{v}_3 = |(g)^-| = -(g)^-, \; \widetilde{v}_4 = (g)^+, \;\; \widetilde{v}_5 = \|g\|_\infty,$$

we see that $\widetilde{v}_3 \geq 0, \widetilde{v}_4 \geq 0$, and

$$\widetilde{v}_3 + \widetilde{v}_4 = -(g)^- + (g)^+ = |g| \leq \|g\|_\infty e = v_5 e, \;\; \widetilde{v}_3 - \widetilde{v}_4 = -(g)^- - (g)^+ = -g.$$

Note that $A^T(\widetilde{v}_3 - \widetilde{v}_4) = -A^T g = -\eta = \widetilde{v}_2 - \widetilde{v}_1$. Therefore, the vector $(\widetilde{v}_1, \ldots, \widetilde{v}_5)$ constructed as above is feasible to the problem (40). We also note that $\eta_S = \text{sign}(x_S)$ and $\|\eta_{\overline{S}}\|_\infty \leq 1$. Then it follows from (47) that for the vector $(x, t = |x|, r = |Ax - y|, \widetilde{v}_1, \ldots, \widetilde{v}_5)$, there exists a point in $D^{(1)}$, denoted still by $(x^*, t^*, r^*, v_1^*, \ldots, v_5^*)$, such that

$$\begin{aligned}
&\|(x, t, r, \widetilde{v}_1, \ldots, \widetilde{v}_5) - (x^*, t^*, r^*, v_1^*, \ldots, v_5^*)\|_2 \\
&\leq \gamma_2 \left[(e^T r - \varepsilon)^+ + \left|e^T |x| + x^T A^T(\widetilde{v}_3 - \widetilde{v}_4) + \phi^T(\widetilde{v}_3 - \widetilde{v}_4) + \widetilde{v}_5 \varepsilon\right|\right]. \\
&= \gamma_2 \left[(e^T r - \varepsilon)^+ + \left|\|x\|_1 + (x_S)^T [A^T(\widetilde{v}_3 - \widetilde{v}_4)]_S + (x_{\overline{S}})^T [A^T(\widetilde{v}_3 - \widetilde{v}_4)]_{\overline{S}} - \phi^T g + \widetilde{v}_5 \varepsilon\right|\right] \\
&= \gamma_2 \left[(e^T r - \varepsilon)^+ + \left|\|x\|_1 - \|x_S\|_1 - (x_{\overline{S}})^T \eta_{\overline{S}} - \phi^T g + \widetilde{v}_5 \varepsilon\right|\right] \\
&\leq \gamma_2 \left[(e^T r - \varepsilon)^+ + \sigma_k(x)_1 + \|x_{\overline{S}}\|_1 \|\eta_{\overline{S}}\|_\infty + |\phi^T g| + \widetilde{v}_5 \varepsilon\right] \\
&\leq \gamma_2 \left[(e^T r - \varepsilon)^+ + 2\sigma_k(x)_1 + \|\phi\|_1 \|g\|_\infty + \|g\|_\infty \varepsilon\right]. \qquad (48)
\end{aligned}$$

As $\|\eta\|_\infty = 1$ and $g = (AA^T)^{-1} A\eta$, we have $\|g\|_\infty \leq \|(AA^T)^{-1} A\|_{\infty \to \infty} = c$. We also note that $r = |Ax - y| = |\phi|$, which indicates that $e^T r = \|Ax - y\|_1 = \|\phi\|_1$. Thus it follows from (48) that

$$\begin{aligned}
\|x - x^*\|_2 &\leq \|(x, t, r, \widetilde{v}_1, ..., \widetilde{v}_5) - (x^*, t^*, r^*, v_1^*, ..., v_5^*)\|_2 \\
&\leq \gamma_2 \left[(\|Ax - y\|_1 - \varepsilon)^+ + 2\sigma_k(x)_1 + c(\|y - Ax\|_1 + \varepsilon)\right].
\end{aligned}$$

In particular, when x satisfies the constraint of (26), i.e., $\|y - Ax\|_1 \leq \varepsilon$, the above estimate is reduced to (43). The proof is complete. □

Similar to Corollary 3.5, we immediately have the following result.

Corollary 4.5. *Let the problem data (A, y, ε) be given, where $\varepsilon > 0$, $y \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m \times n}$ $(m < n)$ with rank$(A) = m$. Let c and c_1 be the constants given in (16), and let γ_1 and γ_2 be the Robinson's constants given in Theorems 4.2 and 4.4, respectively. Suppose that the solutions to (25) and (26) are unique. If A satisfies one of the conditions (p1)–(p6) in Corollary 3.5, then the following statements hold:*

(i) *For any x satisfying $\|Ax - y\|_\infty \leq \varepsilon$, the solution x^* of (25) approximates x with error*

$$\|x - x^*\|_2 \leq 2\gamma_1 \{\sigma_k(x)_1 + c_1 \varepsilon\}.$$

(ii) *For any x satisfying $\|Ax - y\|_1 \leq \varepsilon$, the solution $x^\#$ of (26) approximates x with error*

$$\|x - x^\#\|_2 \leq 2\gamma_2 \{\sigma_k(x)_1 + c\varepsilon\}.$$

A difference between Corollary 4.5 and existing results is in that the constants γ_1 and γ_2 in Corollary 4.5 are Robinson's constants instead of RIP or NSP constants. Each of the matrix properties (p1)–(p6) in Corollary 3.5 implies an identical error bound.

5 Robust weak stability of quadratically constrained models

We now consider the robust weak stability of the quadratically constrained ℓ_1-minimization

$$\gamma^* := \min_x \{\|x\|_1 : \; \|Ax - y\|_2 \leq \varepsilon\}, \qquad (49)$$

where $\varepsilon > 0$, and γ^* denotes the optimal value of the problem. Let S^* denote the set of optimal solutions of (49), which can be represented as

$$S^* = \{x \in \mathbb{R}^n : \ \|x\|_1 \le \gamma^*, \ \|Ax - y\|_2 \le \varepsilon\}.$$

Let $B = \{z \in \mathbb{R}^m : \|z\|_2 \le 1\}$ be the unit ℓ_2-ball. Then problem (49) can be written as

$$\gamma^* = \min_x \{\|x\|_1 : \ u = (Ax - y)/\varepsilon, \ u \in B\}. \tag{50}$$

Since the constraint of (49) is nonlinear, Lemma 2.4 does not apply to this situation directly. We need to establish several auxiliary results in order to show the robust weak stability of (49). The main idea is to approximate B with a polytope. We recall that B is the intersection of half spaces $a^T z \le 1$ tangent to its surface, i.e.,

$$B = \bigcap_{\|a\|_2 = 1} \left\{z \in \mathbb{R}^m : \ a^T z \le 1\right\}. \tag{51}$$

We also recall the Hausdorff metric of two sets $S_1, S_2 \subseteq \mathbb{R}^m$:

$$\delta^{\mathcal{H}}(S_1, S_2) = \max\left\{\sup_{z' \in S_1} \inf_{z \in S_2} \|z' - z\|_2, \ \sup_{z \in S_2} \inf_{z' \in S_1} \|z' - z\|_2\right\}. \tag{52}$$

By taking a finite number of half-spaces in (51) to approximate B, Dudley [20] established the following result. (A more discussion on the polytope approximation of B can be found, for instance, in [7].)

Lemma 5.1. (Dudley [20]) *There exists a constant τ such that for every integer number $K > m$ there is a polytope*

$$\mathcal{P}_K = \bigcap_{\|a^i\|_2 = 1, 1 \le i \le K} \left\{z \in \mathbb{R}^m : (a^i)^T z \le 1\right\}, \tag{53}$$

achieving

$$\delta^{\mathcal{H}}(B, \mathcal{P}_K) \le \frac{\tau}{K^{2/(m-1)}}, \tag{54}$$

where $\delta^{\mathcal{H}}(\cdot, \cdot)$ is the Hausdorff metric defined by (52).

From the above lemma, we see that $\mathcal{P}_K$ can approximate B to any level of accuracy provided that K is sufficiently large. For $\mathcal{P}_K$ given by (53), we use $M_{\mathcal{P}_K} := [a^1, \ldots, a^K]$ to denote the matrix with $a^i \in \mathbb{R}^m, \ i = 1, \ldots, K$ as its columns. We also use the symbol $\mathrm{Col}(M_{\mathcal{P}_K}) = \{a^1, a^2, \ldots, a^K\}$ to denote the set of columns of $M_{\mathcal{P}_K}$. Thus $\mathcal{P}_K$ can be written as

$$\mathcal{P}_K = \{z \in \mathbb{R}^m : (M_{\mathcal{P}_K})^T z \le e\},$$

where e is the vector of ones in $\mathbb{R}^K$. Let $\{\mathcal{P}_K\}_{K>m}$ be any sequence of the polytopes given as (53) and satisfying (54). Consider the sequence of polytopes $\{\widetilde{\mathcal{P}}_J\}_{J>m}$, where

$$\widetilde{\mathcal{P}}_J = \bigcap_{m < K \le J} \mathcal{P}_K. \tag{55}$$

Thus $\widetilde{\mathcal{P}}_J$ is still a polytope formed by a finite number of half space $(a^i)^T z \le 1$ where $\|a^i\|_2 = 1$. We still use $M_{\widetilde{\mathcal{P}}_J}$ to denote the matrix with these vectors a^i's as columns, so

$$\widetilde{\mathcal{P}}_J = \{z \in \mathbb{R}^m : (M_{\widetilde{\mathcal{P}}_J})^T z \le e\}.$$

We still use $\mathrm{Col}(M_{\widetilde{\mathcal{P}}_J})$ to denote the collection of column vectors of $M_{\widetilde{\mathcal{P}}_J}$.

In what follows, for a given compact convex set $T \subseteq \mathbb{R}^n$, we denote the projection of x into T by $\pi_T(x) := \mathrm{argmin}\{\|x-w\|_2 : \ w \in T\}$. We first prove the following lemma.

Lemma 5.2. *Let $\{\mathcal{P}_K\}_{K>m}$ be any sequence of the polytopes defined by (53) and satisfying (54). For any $J>m$, let $\widetilde{\mathcal{P}}_J$ be given as (55). Then for any point $z \in \mathbb{R}^m$ with $\|z\|_2 = 1$, there exists a column vector a^i of $M_{\widetilde{\mathcal{P}}_J}$, i.e., $a^i \in Col(M_{\widetilde{\mathcal{P}}_J})$, such that*

$$\|z - a^i\|_2 \leq \sqrt{\frac{2\tau}{J^{2/(m-1)}+\tau}}.$$

Proof. Let z be any given point on the unit sphere, i.e., $\|z\|_2 = 1$. Since $B \subseteq \widetilde{\mathcal{P}}_J$, where $J>m$, the straight line passing through z and the center of B crosses a point, denoted by z', on the surface of polytope $\widetilde{\mathcal{P}}_J$. Clearly, $z = z'/\|z'\|_2$, i.e., z is the projection of z' onto B. Note that $B \subseteq \widetilde{\mathcal{P}}_J \subseteq \mathcal{P}_J$ for any $J>m$. By the definition of Hausdorff metric and Lemma 5.1, we obtain

$$\|z - z'\|_2 \leq \delta^{\mathcal{H}}(B, \widetilde{\mathcal{P}}_J) \leq \delta^{\mathcal{H}}(B, \mathcal{P}_J) \leq \frac{\tau}{J^{2/(m-1)}}. \tag{56}$$

Since z' is on the surface of $\widetilde{\mathcal{P}}_J$, there is a vector $a^{i_0} \in \mathrm{Col}(M_{\widetilde{\mathcal{P}}_J})$ such that $(a^{i_0})^T z' = 1$. Note that $\|z'-z\|_2 = \|z' - \frac{z'}{\|z'\|_2}\|_2 = \|z'\|_2 - 1$, $\|a^{i_0}\|_2 = \|z\|_2 = 1$ and $(a^{i_0})^T z' = 1$. We immediately have

$$\begin{aligned} \|z - a^{i_0}\|_2^2 &= 2(1-(a^{i_0})^T z) = 2(1 - \frac{(a^{i_0})^T z'}{\|z'\|_2}) = 2(1-\frac{1}{\|z'\|_2}) \\ &= \frac{2\|z'-z\|_2}{\|z'-z\|_2+1} \leq \frac{2(\tau/J^{2/(m-1)})}{(\tau/J^{2/(m-1)})+1} = \frac{2\tau}{\tau + J^{2/(m-1)}}, \end{aligned}$$

where the inequality follows from (56). $\square$

Recall that S^* is the set of optimal solutions of (49). We now prove the next lemma.

Lemma 5.3. *Let $\{\mathcal{P}_K\}_{K>m}$ and $\widetilde{\mathcal{P}}_J$ be given as Lemma 5.2. Let $S_{\widetilde{\mathcal{P}}_J}$ be the set*

$$S_{\widetilde{\mathcal{P}}_J} = \{x \in \mathbb{R}^n : \ \|x\|_1 \leq \gamma^*, \ u = (Ax-y)/\varepsilon, \ u \in \widetilde{\mathcal{P}}_J\}, \tag{57}$$

where γ^* is the optimal value of (49). Then $\delta^{\mathcal{H}}(S^*, S_{\widetilde{\mathcal{P}}_J}) \to 0$ as $J \to \infty$.

Proof. Note that $B \subseteq \widetilde{\mathcal{P}}_J \subseteq \mathcal{P}_J$ for every $J>m$. By the definition of Hausdorff metric and Lemma 5.1, we see that

$$\delta^{\mathcal{H}}(B, \widetilde{\mathcal{P}}_J) \leq \delta^{\mathcal{H}}(B, \mathcal{P}_J) \leq \frac{\tau}{J^{2/(m-1)}}, \ J>m. \tag{58}$$

Note that $S_{\widetilde{\mathcal{P}}_J}$, given by (57), can be rewritten as

$$S_{\widetilde{\mathcal{P}}_J} = \{x \in \mathbb{R}^n : \ \|x\|_1 \leq \gamma^*, \ (M_{\widetilde{\mathcal{P}}_J})^T(Ax-y) \leq \varepsilon e\},$$

where γ^* is the optimal value of (49). Clearly, $S^* \subseteq S_{\widetilde{\mathcal{P}}_J}$ due to the fact $B \subseteq \widetilde{\mathcal{P}}_J$. We now prove that $\delta^{\mathcal{H}}(S^*, S_{\widetilde{\mathcal{P}}_J}) \to 0$ as $J \to \infty$. Since S^* is a subset of $S_{\widetilde{\mathcal{P}}_J}$, by the definition of Hausdorff metric, we see that

$$\delta^{\mathcal{H}}(S^*, S_{\widetilde{\mathcal{P}}_J}) = \sup_{w \in S_{\widetilde{\mathcal{P}}_J}} \inf_{z \in S^*} \|w - z\|_2 = \sup_{w \in S_{\widetilde{\mathcal{P}}_J}} \|w - \pi_{S^*}(w)\|_2, \tag{59}$$

where $\pi_{S^*}(w) \in S^*$ is the projection of w into S^*. The projection operator $\pi_{S^*}(w)$ is continuous in w and $S_{\widetilde{\mathcal{P}}_J}$ is compact convex set for any $\widetilde{\mathcal{P}}_J$. Thus for every polytope $\widetilde{\mathcal{P}}_J$, the superimum in (59) can be attained, i.e., there exists a point, denoted by $w^*_{\widetilde{\mathcal{P}}_J} \in S_{\widetilde{\mathcal{P}}_J}$, such that

$$\delta^{\mathcal{H}}(S^*, S_{\widetilde{\mathcal{P}}_J}) = \left\| w^*_{\widetilde{\mathcal{P}}_J} - \pi_{S^*}(w^*_{\widetilde{\mathcal{P}}_J}) \right\|_2. \tag{60}$$

We also note that $S^* \subseteq S_{\widetilde{\mathcal{P}}_{J+1}} \subseteq S_{\widetilde{\mathcal{P}}_J}$ for any $J > m$, which implies that $\delta^{\mathcal{H}}(S^*, S_{\widetilde{\mathcal{P}}_{J+1}}) \leq \delta^{\mathcal{H}}(S^*, S_{\widetilde{\mathcal{P}}_J})$. Thus $\{\delta^{\mathcal{H}}(S^*, S_{\widetilde{\mathcal{P}}_J})\}_{J>m}$ is a non-increasing nonnegative sequence. There must exist a number $\delta \geq 0$ such that

$$\lim_{J\to\infty} \delta^{\mathcal{H}}(S^*, S_{\widetilde{\mathcal{P}}_J}) = \delta \geq 0.$$

We now further prove that $\delta = 0$. Note that $w^*_{\widetilde{\mathcal{P}}_J} \in S_{\widetilde{\mathcal{P}}_J}$ for any $J > m$. Thus

$$\left\| w^*_{\widetilde{\mathcal{P}}_J} \right\|_1 \leq \gamma^*, \ (M_{\widetilde{\mathcal{P}}_J})^T (A w^*_{\widetilde{\mathcal{P}}_J} - y) \leq \varepsilon e \text{ for any } J > m. \tag{61}$$

The inequality (61) implies that the sequence $\{w^*_{\widetilde{\mathcal{P}}_J}\}_{J>m}$ is bounded and satisfies that

$$\left[\sup_{a^i \in \mathrm{Col}(M_{\widetilde{\mathcal{P}}_J})} (a^i)^T (A w^*_{\widetilde{\mathcal{P}}_J} - y) \right] \leq \varepsilon \ \text{ for any } J > m.$$

Note that for any $m < J' \leq J$, we have $\mathrm{Col}(M_{\widetilde{\mathcal{P}}_{J'}}) \subseteq \mathrm{Col}(M_{\widetilde{\mathcal{P}}_J})$. Thus the inequality above implies that for any fixed integer number $J' > m$,

$$\left[\sup_{a^i \in \mathrm{Col}(M_{\widetilde{\mathcal{P}}_{J'}})} (a^i)^T (A w^*_{\widetilde{\mathcal{P}}_J} - y) \right] \leq \varepsilon \ \text{ for any } J \geq J'.$$

Note that the sequence $\{w^*_{\widetilde{\mathcal{P}}_J}\}_{J \geq J'}$ is bounded. Pasting through to a subsequence if necessary, we may assume that $w^*_{\widetilde{\mathcal{P}}_J} \to w^*$ with $\|w^*\|_1 \leq \gamma^*$. Thus it follows from the above inequality that

$$\sup_{a^i \in \mathrm{Col}(M_{\widetilde{\mathcal{P}}_{J'}})} (a^i)^T (A w^* - y) \leq \varepsilon, \tag{62}$$

which holds for any given $J' > m$. We now prove that (62) implies that $\|Aw^* - y\|_2 \leq \varepsilon$. We show this by contradiction. Assume that $\|Aw^* - y\|_2 > \varepsilon$, which by the definition of the ℓ_2-norm implies that

$$\max_{\|a\|_2=1} a^T (Aw^* - y) = \|Aw^* - y\|_2 > \varepsilon.$$

The maximum above attains at $a^* = (Aw^* - y)/\|Aw^* - y\|_2$. By continuity, there exists a neighborhood of a^*, namely, $U = a^* + \delta^* B$, where $\delta^* > 0$ is a small number, such that any point $w \in U \cap \{z \in \mathbb{R}^m : \|z\|_2 = 1\}$ satisfies that

$$w^T (Aw^* - y) \geq \frac{1}{2} \left(\|Aw^* - y\|_2 + \varepsilon \right). \tag{63}$$

Note that $\widetilde{\mathcal{P}}_J$ achieves (58). Let J' be an integer number such that $\sqrt{\frac{2\tau}{(J')^{2/(m-1)}+\tau}} \leq \delta^*$. Applying Lemma 5.2 to $\widetilde{\mathcal{P}}_{J'}$, we conclude that for the vector a^*, there is a vector $a^i \in \mathrm{Col}(M_{\widetilde{\mathcal{P}}_{J'}})$ such that

$$\|a^i - a^*\|_2 \leq \sqrt{\frac{2\tau}{(J')^{2/(m-1)} + \tau}} \leq \delta^*,$$

which, together with the fact $\|a^i\|_2 = 1$, implies that $a^i \in U \cap \{z \in \mathbb{R}^m : \|z\|_2 = 1\}$. Thus it follows from (63) that

$$(a^i)^T(Aw^* - y) \geq \frac{1}{2}(\|Aw^* - y\|_2 + \varepsilon) > \varepsilon.$$

This contradicts (62). Thus w^* must satisfy that $\|Aw^* - y\|_2 \leq \varepsilon$. This together with the fact $\|w^*\|_1 \leq \gamma^*$ implies that $w^* \in S^*$. As a result, $\pi_{S^*}(w^*) = w^*$. It follows from (60) and the continuity of $\pi_{S^*}(\cdot)$ that

$$\delta = \lim_{J\to\infty} \delta^{\mathcal{H}}(S^*, S_{\widetilde{\mathcal{P}}_J}) = \lim_{J\to\infty} \|w^*_{\widetilde{\mathcal{P}}_J} - \pi_{S^*}(w^*_{\widetilde{\mathcal{P}}_J})\|_2 = \|w^* - \pi_{S^*}(w^*)\|_2 = 0,$$

as desired. $\square$

We will also make use of the following property of a projection operator.

Lemma 5.4. *Let S' and S'' be compact convex sets in $\mathbb{R}^n$. Then for any $x \in \mathbb{R}^n$,*

$$\|\pi_{S'}(x) - \pi_{S''}(x)\|_2^2 \leq \delta^{\mathcal{H}}(S', S'')(\|x - \pi_{S'}(x)\|_2 + \|x - \pi_{S''}(x)\|_2).$$

Proof. By the property of projection operators, we have

$$(x - \pi_{S'}(x))^T(v - \pi_{S'}(x)) \leq 0 \text{ for all } v \in S', \tag{64}$$

$$(x - \pi_{S''}(x))^T(u - \pi_{S''}(x)) \leq 0 \text{ for all } u \in S''. \tag{65}$$

We project $\pi_{S''}(x) \in S''$ into S' to get the point $\widehat{v} = \pi_{S'}(\pi_{S''}(x)) \in S'$ and we project $\pi_{S'}(x) \in S'$ into S'' to get the point $\widehat{u} = \pi_{S''}(\pi_{S'}(x)) \in S''$. By the definition of Hausdorff metric, we have

$$\|\widehat{v} - \pi_{S''}(x)\|_2 \leq \delta^{\mathcal{H}}(S', S''), \;\; \|\widehat{u} - \pi_{S'}(x)\|_2 \leq \delta^{\mathcal{H}}(S', S''). \tag{66}$$

Substituting $\widehat{v}$ into (64) and $\widehat{u}$ into (65) yields

$$(x - \pi_{S'}(x))^T(\widehat{v} - \pi_{S'}(x)) \leq 0, \;\; (x - \pi_{S''}(x))^T(\widehat{u} - \pi_{S''}(x)) \leq 0,$$

which implies the first inequality below

$$\begin{aligned}
\|\pi_{S'}(x) - \pi_{S''}(x)\|_2^2 &= (\pi_{S'}(x) - x + x - \pi_{S''}(x))^T(\pi_{S'}(x) - \pi_{S''}(x)) \\
&= -(x - \pi_{S'}(x))^T(\pi_{S'}(x) - \pi_{S''}(x)) + (x - \pi_{S''}(x))^T(\pi_{S'}(x) - \pi_{S''}(x)) \\
&\leq -(x - \pi_{S'}(x))^T(\widehat{v} - \pi_{S''}(x)) + (x - \pi_{S''}(x))^T(\pi_{S'}(x) - \widehat{u}) \\
&\leq \|x - \pi_{S'}(x)\|_2\|\pi_{S''}(x) - \widehat{v}\|_2 + \|x - \pi_{S''}(x)\|_2\|\pi_{S'}(x) - \widehat{u}\|_2 \\
&\leq \delta^{\mathcal{H}}(S', S'')(\|x - \pi_{S'}(x)\|_2 + \|x - \pi_{S''}(x)\|_2),
\end{aligned}$$

where the final inequality follows from (66). $\square$

For each $K > 2m$, by Lemma 5.1, there is a polytope $\mathcal{P}_K$ of the form (53) achieving (54), and $\mathcal{P}_K$ can be represented as $\mathcal{P}_K = \{z \in \mathbb{R}^m : (M_{\mathcal{P}_K})^T z \leq e\}$. We now add the following $2m$ half spaces

$$(\pm\varrho_i)^T z \leq 1, \;\; i = 1, \ldots, m$$

to $\mathcal{P}_K$, where ϱ_i $(i = 1, \ldots, m)$ denotes the i-th column vector of the $m \times m$ identity matrix. Let $\widehat{K}$ denote the cardinality of the set $\mathrm{Col}(M_{\mathcal{P}_K}) \cup \{\pm\varrho_i : i = 1, \ldots, m\}$. This yields the polytope

$$\mathcal{P}_{\widehat{K}} := \mathcal{P}_K \cap \{z \in \mathbb{R}^m : \;\varrho_i^T z \leq 1, \;\; -\varrho_i^T z \leq 1, \;\; i = 1, \ldots, m\}, \tag{67}$$

Therefore,

$$\mathrm{Col}(M_{\mathcal{P}_{\widehat{K}}}) = \mathrm{Col}(M_{\mathcal{P}_K}) \cup \{\pm\varrho_i : i = 1, \ldots, m\} \tag{68}$$

and $\widehat{K} = |\mathrm{Col}(M_{\mathcal{P}_{\widehat{K}}})|$. Clearly, $K \le \widehat{K} \le K + 2m$ which together with $K > 2m$ implies that $1 \le \widehat{K}/K \le 2$. Let τ be the constant in Lemma 5.1 and let $\tau' = 4^{1/(m-1)}\tau$. By the definition of Hausdorff metric and Lemma 5.1, we see that the polytope $\mathcal{P}_{\widehat{K}}$ constructed as (67) satisfies

$$\delta^{\mathcal{H}}(B, \mathcal{P}_{\widehat{K}}) \le \delta^{\mathcal{H}}(B, \mathcal{P}_K) \le \frac{\tau}{K^{2/(m-1)}} = \frac{\tau}{\widehat{K}^{2/(m-1)}} \left(\frac{\widehat{K}}{K}\right)^{2/(m-1)} \le \frac{\tau'}{\widehat{K}^{2/(m-1)}}. \tag{69}$$

We use the set $\mathcal{P}_{\widehat{K}}$ defined as (67), which achieves (69), to construct the sequence of polytopes $\{\widetilde{\mathcal{P}}_J\}$ as follows:

$$\widetilde{\mathcal{P}}_J = \bigcap_{m < \widehat{K} \le J} \mathcal{P}_{\widehat{K}}. \tag{70}$$

Let $S_{\widetilde{\mathcal{P}}_J}$ denote the set (57) with $\widetilde{\mathcal{P}}_J$ being given by (70). Then Lemma 5.3 remains valid for the sequence of polytopes given by (70). So $\delta^{\mathcal{H}}(S^*, S_{\widetilde{\mathcal{P}}_J}) \to 0$ as $J \to \infty$.

Thus in the remainder of the paper, let $\varepsilon' > 0$ be any fixed small number. From the above discussion, there exists an integer number $J_0 > 2m$ such that

$$\delta^{\mathcal{H}}(S^*, S_{\widetilde{\mathcal{P}}_{J_0}}) \le \varepsilon'. \tag{71}$$

We consider the fixed polytope $\widetilde{\mathcal{P}}_{J_0}$ constructed as above. This polytope is an approximation of B and achieves (71). We use $\widehat{N}$ to denote the number of columns of $M_{\widetilde{\mathcal{P}}_{J_0}}$ and use $e_{\widehat{N}}$ to denote the vector of ones in $\mathbb{R}^{\widehat{N}}$ to distinguish it from e, the vector of ones in $\mathbb{R}^n$. Replacing B in (50) by $\widetilde{\mathcal{P}}_{J_0}$ leads to the following approximation of (49):

$$\gamma^*_{\widetilde{\mathcal{P}}_{J_0}} := \min_x \{\|x\|_1 : \ u = (Ax - y)/\varepsilon, u \in \widetilde{\mathcal{P}}_{J_0}\} = \min_x \{\|x\|_1 : \ (M_{\widetilde{\mathcal{P}}_{J_0}})^T (Ax - y) \le \varepsilon e_{\widehat{N}}\}, \tag{72}$$

where $\gamma^*_{\widetilde{\mathcal{P}}_{J_0}}$ is the optimal value of the above problem. Let

$$S^*_{\widetilde{\mathcal{P}}_{J_0}} = \{x \in \mathbb{R}^n : \ \|x\|_1 \le \gamma^*_{\widetilde{\mathcal{P}}_{J_0}}, \ u = (Ax - y)/\varepsilon, \ u \in \widetilde{\mathcal{P}}_{J_0}\}$$

be the set of optimal solutions of (72), and let $S_{\widetilde{\mathcal{P}}_{J_0}}$ be the set defined by (57) with $\widetilde{\mathcal{P}}_J$ replaced by $\widetilde{\mathcal{P}}_{J_0}$. Clearly, $S^* \subseteq S_{\widetilde{\mathcal{P}}_{J_0}}$. Note that $\gamma^*_{\widetilde{\mathcal{P}}_{J_0}} \le \gamma^*$ due to the fact $B \subseteq \widetilde{\mathcal{P}}_{J_0}$. We immediately see that $S^*_{\widetilde{\mathcal{P}}_{J_0}} \subseteq S_{\widetilde{\mathcal{P}}_{J_0}}$ The problem (72) can be written as

$$\min_{(x,t)} \{e^T t : \ x \le t, \ -x \le t, \ t \ge 0, \ (M_{\widetilde{\mathcal{P}}_{J_0}})^T (Ax - y) \le \varepsilon e_{\widehat{N}}\},$$

to which the dual problem is given as

$$\begin{aligned} \max \quad & -\left[\varepsilon e_{\widehat{N}} + (M_{\widetilde{\mathcal{P}}_{J_0}})^T y\right]^T v_3 \\ \text{s.t.} \quad & A^T M_{\widetilde{\mathcal{P}}_{J_0}} v_3 + v_1 - v_2 = 0, \ v_1 + v_2 \le e, \ (v_1, v_2, v_3) \ge 0. \end{aligned} \tag{73}$$

The following lemma immediately follows from the optimality condition of the above linear program.

Lemma 5.5. *$x^* \in \mathbb{R}^n$ is an optimal solution of (72) if and only if there exist vectors $t^*, v_1^*, v_2^* \in \mathbb{R}^n_+$ and $v_3^* \in \mathbb{R}^{\widehat{N}}_+$ such that $(x^*, t^*, v_1^*, v_2^*, v_3^*) \in D^{(2)}$ where*

$$D^{(2)} = \{(x, t, v_1, v_2, v_3) : \begin{array}{l} x \le t, \ -x \le t, \ (M_{\widetilde{\mathcal{P}}_{J_0}})^T (Ax - y) \le \varepsilon e_{\widehat{N}}, \\ A^T M_{\widetilde{\mathcal{P}}_{J_0}} v_3 + v_1 - v_2 = 0, \ v_1 + v_2 \le e, \\ e^T t = -\left[\varepsilon e_{\widehat{N}} + (M_{\widetilde{\mathcal{P}}_{J_0}})^T y\right]^T v_3, \ (t, v_1, v_2, v_3) \ge 0\}. \end{array} \tag{74}$$

Moreover, for any $(x, t, v_1, v_2, v_3) \in D^{(2)}$, it must hold that $t = |x|$.

To apply Lemma 2.4, we write (74) in the form

$$D^{(2)} = \{z = (x, t, v_1, v_2, v_3) : \ M^+ z \le b^+, \ M^{++} z = b^{++}\}, \tag{75}$$

where $b^{++} = 0$ and

$$M^+ = \begin{pmatrix} I & -I & 0 & 0 & 0 \\ -I & -I & 0 & 0 & 0 \\ (M_{\widetilde{\mathcal{P}}_{J_0}})^T A & 0 & 0 & 0 & 0 \\ 0 & 0 & I & I & 0 \\ 0 & -I & 0 & 0 & 0 \\ 0 & 0 & -I & 0 & 0 \\ 0 & 0 & 0 & -I & 0 \\ 0 & 0 & 0 & 0 & -I_{\widehat{N}} \end{pmatrix}, \quad b^+ = \begin{pmatrix} 0 \\ 0 \\ (M_{\widetilde{\mathcal{P}}_{J_0}})^T y + \varepsilon e_{\widehat{N}} \\ e \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \tag{76}$$

$$M^{++} = \begin{pmatrix} 0 & 0 & I & -I & A^T M_{\widetilde{\mathcal{P}}_{J_0}} \\ 0 & e^T & 0 & 0 & \varepsilon e_{\widehat{N}}^T + y^T M_{\widetilde{\mathcal{P}}_{J_0}} \end{pmatrix}, \tag{77}$$

where I and $I_{\widehat{N}}$ are the $n \times n$ and $\widehat{N} \times \widehat{N}$ identity matrices, respectively. We now prove the main result in this section.

Theorem 5.6. *Let the problem data (A, y, ε) of (49) be given, where $\varepsilon > 0$, $y \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m \times n}$ $(m < n)$ with rank$(A) = m$. Let ε' be any prescribed small number and let the polytope $\widetilde{\mathcal{P}}_{J_0}$ be constructed as (70) and achieve (71). Suppose that A^T satisfies the weak RSP of order k. Then for any $x \in \mathbb{R}^n$, there is an optimal solution x^* of (49) such that*

$$\|x - x^*\|_2 \le 2\gamma_3 \left\{\widehat{N}(\|Ax - y\|_2 - \varepsilon)^+ + 2\sigma_k(x)_1 + c_1\varepsilon + c_2\|Ax - y\|_2\right\} + 2\varepsilon', \tag{78}$$

where c_1 and c_2 are constants given in (16), $\gamma_3 = \sigma_{\infty,2}(M^+, M^{++})$ is the Robinson's constant determined by (M^+, M^{++}) given in (76) and (77). Moreover, for any x with $\|Ax - y\|_2 \le \varepsilon$, there is an optimal solution x^* of (49) such that

$$\|x - x^*\|_2 \le 4\gamma_3\sigma_k(x)_1 + 2\gamma_3(c_1 + c_2)\varepsilon + 2\varepsilon'.$$

Proof. Let x be any vector in $\mathbb{R}^n$ and let $t = |x|$. We still denote by S the support set of the k-largest entries of $|x|$. Let $S_+ = \{i \in S : x_i > 0\}$ and $S_- = \{i \in S : x_i < 0\}$. Then $S = S_+ \cup S_-$. Since A^T satisfies the weak RSP of order k, there exists a vector $\eta = A^T g$ for some $g \in \mathbb{R}^m$, satisfying that $\eta_i = 1$ for $i \in S_+$, $\eta_i = -1$ for $i \in S_-$, and $|\eta_i| \le 1$ for $i \in \overline{S}$, where $\overline{S} = \{1, \ldots, n\} \backslash S$. For the given problem data (A, y, ε), as shown between (67) and (71), there exists an integer number $J_0 > 2m$ such that the polytope $\widetilde{\mathcal{P}}_{J_0}$, given as (70), can approximate B and achieve the bound (71). We now construct a feasible solution $(\widetilde{v}_1, \widetilde{v}_2, \widetilde{v}_3)$ to problem

(73). Set $(\widetilde{v}_1)_i = 1$ and $(\widetilde{v}_2)_i = 0$ for all $i \in S_+$, $(\widetilde{v}_1)_i = 0$ and $(\widetilde{v}_2)_i = 1$ for all $i \in S_-$, and $(\widetilde{v}_1)_i = (|\eta_i| + \eta_i)/2$ and $(\widetilde{v}_2)_i = (|\eta_i| - \eta_i)/2$ for all $i \in \overline{S}$. This choice of $\widetilde{v}_1$ and $\widetilde{v}_2$ ensures that $(\widetilde{v}_1, \widetilde{v}_2) \geq 0$, $\widetilde{v}_1 + \widetilde{v}_2 \leq e$ and $\widetilde{v}_1 - \widetilde{v}_2 = \eta$. We now construct the vector $\widetilde{v}_3$. By the construction of $\widetilde{\mathcal{P}}_{J_0}$, we see that

$$\{\pm \varrho_i : \ i = 1, \ldots, m\} \subseteq \mathrm{Col}(M_{\widetilde{\mathcal{P}}_{J_0}}).$$

It is not difficult to show that there exists a vector $\widetilde{v}_3 \in \mathbb{R}^{\widehat{N}}_+$ satisfying $M_{\widetilde{\mathcal{P}}_{J_0}} \widetilde{v}_3 = -g$ and $\|\widetilde{v}_3\|_1 = \|g\|_1$. In fact, without loss of generality, we assume that $\{-\varrho_i : \ i = 1, \ldots, m\}$ are arranged as the first m columns and $\{\varrho_i : \ i = 1, \ldots, m\}$ are arranged as the second m columns in $M_{\widetilde{\mathcal{P}}_{J_0}}$. For every $i = 1, \ldots, m$, if $g_i \geq 0$, then we set $(\widetilde{v}_3)_i = g_i$; otherwise, if $g_i < 0$, then we set $(\widetilde{v}_3)_{m+i} = -g_i$. All remaining entries of $\widetilde{v}_3 \in \mathbb{R}^{\widehat{N}}$ are set to be zero. By this choice of $\widetilde{v}_3$, we see that $\widetilde{v}_3 \geq 0$, $M_{\widetilde{\mathcal{P}}_{J_0}} \widetilde{v}_3 = -g$ and

$$\|\widetilde{v}_3\|_1 = \|g\|_1 = \|(AA^T)^{-1} A\eta\|_1 \leq \|(AA^T)^{-1} A\|_{\infty \to 1} \|\eta\|_\infty \leq c_1, \tag{79}$$

where c_1 is the constant given in (16).

Let $D^{(2)}$ be given as in Lemma 5.5. $D^{(2)}$ can be written as (75). For the vector $(x, t, \widetilde{v}_1, \widetilde{v}_2, \widetilde{v}_3)$, applying Lemma 2.4 with $(M', M'') = (M^+, M^{++})$ where M^+ and M^{++} are given as (76) and (77), there exists a point in $D^{(2)}$, denoted by $(\widehat{x}, \widehat{t}, \widehat{v}_1, \widehat{v}_2, \widehat{v}_3)$, such that

$$\left\| \begin{bmatrix} x \\ t \\ \widetilde{v}_1 \\ \widetilde{v}_2 \\ \widetilde{v}_3 \end{bmatrix} - \begin{bmatrix} \widehat{x} \\ \widehat{t} \\ \widehat{v}_1 \\ \widehat{v}_2 \\ \widehat{v}_3 \end{bmatrix} \right\|_2 \leq \gamma_3 \left\| \begin{bmatrix} \left((M_{\widetilde{\mathcal{P}}_{J_0}})^T (Ax - y) - \varepsilon e_{\widehat{N}} \right)^+ \\ (x - t)^+ \\ (-x - t)^+ \\ A^T M_{\widetilde{\mathcal{P}}_{J_0}} \widetilde{v}_3 + \widetilde{v}_1 - \widetilde{v}_2 \\ (\widetilde{v}_1 + \widetilde{v}_2 - e)^+ \\ e^T t + \left(\varepsilon e_{\widehat{N}} + (M_{\widetilde{\mathcal{P}}_{J_0}})^T y \right)^T \widetilde{v}_3 \\ t^- \\ (\widetilde{\vartheta})^- \end{bmatrix} \right\|_1, \tag{80}$$

where $(\widetilde{\vartheta})^-$ denotes the vector $((\widetilde{v}_1)^-, (\widetilde{v}_2)^-, (\widetilde{v}_3)^-)$, and $\gamma_3 = \sigma_{\infty,2}(M^+, M^{++})$ is the Robinson's constant determined by (M^+, M^{++}) given in (76) and (77). Note that $t = |x|$ implies that $(x - t)^+ = (-x - t)^+ = t^- = 0$. Also, since $(\widetilde{v}_1, \widetilde{v}_2, \widetilde{v}_3)$ is feasible to (73), we have $(\widetilde{\vartheta})^- = 0$, $(\widetilde{v}_1 + \widetilde{v}_2 - e)^+ = 0$ and $A^T M_{\widetilde{\mathcal{P}}_{J_0}} \widetilde{v}_3 + \widetilde{v}_1 - \widetilde{v}_2 = 0$. Thus (80) is reduced to

$$\|x - \widehat{x}\|_2 \leq \gamma_3 \left\{ \left\| \left[(M_{\widetilde{\mathcal{P}}_{J_0}})^T (Ax - y) - \varepsilon e_N \right]^+ \right\|_1 + \left| e^T t + \left[\varepsilon e_{\widehat{N}} + (M_{\widetilde{\mathcal{P}}_{J_0}})^T y \right]^T \widetilde{v}_3 \right| \right\}. \tag{81}$$

Note that for every $a^i \in \mathrm{Col}(M_{\widetilde{\mathcal{P}}_{J_0}})$, we have $\|a^i\|_2 = 1$ and thus $(a^i)^T (Ax - y) \leq \|Ax - y\|_2$. This implies that $\left[(a^i)^T (Ax - y) - \varepsilon \right]^+ \leq (\|Ax - y\|_2 - \varepsilon)^+$ and hence

$$\left[(M_{\widetilde{\mathcal{P}}_{J_0}})^T (Ax - y) - \varepsilon e_{\widehat{N}} \right]^+ \leq (\|Ax - y\|_2 - \varepsilon)^+ e_{\widehat{N}},$$

and hence

$$\left\| \left[(M_{\widetilde{\mathcal{P}}_{J_0}})^T (Ax - y) - \varepsilon e_{\widehat{N}} \right]^+ \right\|_1 \leq \widehat{N} (\|Ax - y\|_2 - \varepsilon)^+. \tag{82}$$

By the definition of η, we see that $x^T A^T g = x^T \eta = \|x_S\|_1 + x_{\overline{S}}^T \eta_{\overline{S}}$ and thus

$$\left| e^T |x| - x^T A^T g \right| = \left| \|x\|_1 - \|x_S\|_1 - x_{\overline{S}}^T \eta_{\overline{S}} \right| \leq \|x_{\overline{S}}\|_1 + |x_{\overline{S}}^T \eta_{\overline{S}}| \leq 2 \|x_{\overline{S}}\|_1 = 2\sigma_k(x)_1.$$

We also note that

$$\|g\|_2 = \|(AA^T)^{-1}A\eta\|_2 \le \|(AA^T)^{-1}A\|_{\infty\to 2}\|\eta\|_\infty \le c_2, \tag{83}$$

where c_2 is the constant given in (16). Thus, by letting $\phi = Ax - y$ and noting that $M_{\widetilde{\mathcal{P}}_{J_0}}\widetilde{v}_3 = -g$, we have

$$\begin{aligned}
\left|e^T t + \left[\varepsilon e_{\widehat{N}} + (M_{\widetilde{\mathcal{P}}_{J_0}})^T y\right]^T \widetilde{v}_3\right| &= \left|e^T|x| + x^T A^T M_{\widetilde{\mathcal{P}}_{J_0}}\widetilde{v}_3 - \phi^T M_{\widetilde{\mathcal{P}}_{J_0}}\widetilde{v}_3 + \varepsilon e_{\widehat{N}}^T \widetilde{v}_3\right| \\
&= |e^T|x| - x^T A^T g + \phi^T g + \varepsilon e_{\widehat{N}}^T \widetilde{v}_3| \\
&\le 2\sigma_k(x)_1 + |\phi^T g| + |\varepsilon e_{\widehat{N}}^T \widetilde{v}_3| \\
&\le 2\sigma_k(x)_1 + \|\phi\|_2\|g\|_2 + \varepsilon\|\widetilde{v}_3\|_1 \\
&\le 2\sigma_k(x)_1 + c_2\|Ax - y\|_2 + \varepsilon c_1,
\end{aligned} \tag{84}$$

where the last inequality follows from (79) and (83). Merging (81), (82) and (84) leads to

$$\|x - \widehat{x}\|_2 \le \gamma_3\left[|\widehat{N}|(\|Ax - y\|_2 - \varepsilon)^+ + 2\sigma_k(x)_1 + c_1\varepsilon + c_2\|Ax - y\|_2\right]. \tag{85}$$

Note that the set $S_{\widetilde{\mathcal{P}}_{J_0}}$ and S^* are compact convex sets. Let x^* and $\overline{x}$ denote the projection of x onto S^* and $S_{\widetilde{\mathcal{P}}_{J_0}}$ respectively, namely, $x^* = \pi_{S^*}(x) \in S^*$ and $\overline{x} = \pi_{S_{\widetilde{\mathcal{P}}_{J_0}}}(x) \in S_{\widetilde{\mathcal{P}}_{J_0}}$. Since $S^* \subseteq S_{\widetilde{\mathcal{P}}_{J_0}}$, we have $\|x - \overline{x}\|_2 \le \|x - x^*\|_2$. By (71), $\delta^{\mathcal{H}}(S^*, S_{\widetilde{\mathcal{P}}_{J_0}}) \le \varepsilon'$, which together with Lemma 5.4 implies that

$$\|x^* - \overline{x}\|_2^2 \le \delta^{\mathcal{H}}(S^*, S_{\widetilde{\mathcal{P}}_{J_0}})(\|x - x^*\|_2 + \|x - \overline{x}\|_2) \le \varepsilon'(\|x - x^*\|_2 + \|x - \overline{x}\|_2). \tag{86}$$

Note that $\widehat{x} \in S^*_{\widetilde{\mathcal{P}}_{J_0}} \subseteq S_{\widetilde{\mathcal{P}}_{J_0}}$ and $\overline{x}$ is the projection of x into the convex set $S_{\widetilde{\mathcal{P}}_{J_0}}$. Thus $\|x - \overline{x}\|_2 \le \|x - \widehat{x}\|_2$. By triangle inequality and (86), we have

$$\begin{aligned}
\|x - x^*\|_2 &\le \|x - \overline{x}\|_2 + \|\overline{x} - x^*\|_2 \\
&\le \|x - \widehat{x}\|_2 + \|\overline{x} - x^*\|_2 \\
&\le \|x - \widehat{x}\|_2 + \sqrt{\varepsilon'(\|x - x^*\|_2 + \|x - \overline{x}\|_2)}.
\end{aligned} \tag{87}$$

Since $\|x - \overline{x}\|_2 \le \|x - x^*\|_2$, it follows from (87) that

$$\|x - x^*\|_2 \le \|x - \widehat{x}\|_2 + \sqrt{2\varepsilon'\|x - x^*\|_2}, \tag{88}$$

which implies that

$$\|x - x^*\|_2 \le \left(\frac{\sqrt{2\varepsilon'} + \sqrt{2\varepsilon' + 4\|x - \widehat{x}\|_2}}{2}\right)^2 \le 2\varepsilon' + 2\|x - \widehat{x}\|_2,$$

where the last inequality follows from the fact $\left(\frac{a+b}{2}\right)^2 \le \frac{a^2+b^2}{2}$. Combination of the inequality above and (85) immediately yields (78), i.e.,

$$\|x - x^*\|_2 \le 2\varepsilon' + 2\gamma_3\left\{\widehat{N}(\|Ax - y\|_2 - \varepsilon)^+ + 2\sigma_k(x)_1 + c_1\varepsilon + c_2\|Ax - y\|_2\right\}.$$

In particularly, when x satisfies $\|Ax - y\|_2 \le \varepsilon$, the above inequality is reduced to

$$\|x - x^*\|_2 \le 2\varepsilon' + 2\gamma_3\{2\sigma_k(x)_1 + (c_1 + c_2)\varepsilon\} = 4\gamma_3\sigma_k(x)_1 + 2\gamma_3(c_1 + c_2)\varepsilon + 2\varepsilon',$$

as desired. $\square$

We immediately have the following corollary.

Corollary 5.7. *Let the problem data (A, y, ε) be given, where $\varepsilon > 0$, $y \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m\times n}(m < n)$ with $rank(A) = m$. Let ε' be any prescribed small number and the polytope $\widetilde{\mathcal{P}}_{J_0}$ achieve (71). Then under each of the listed conditions in Corollary 3.5, for any $x \in \mathbb{R}^n$ with $\|Ax - y\|_2 \le \varepsilon$ there is an optimal solution x^* of (49) such that*

$$\|x - x^*\|_2 \le 4\gamma_3\sigma_k(x)_1 + 2(\gamma_3 c_1 + \gamma_3 c_2)\varepsilon + 2\varepsilon'.$$

where c_1 and c_2 are given in (16) and γ_3 is the Robinson's constant given in Theorem 5.6.

The weak stability is a more general concept than stability. Any traditional sufficient condition for stability of ℓ_1-minimization problems, by Theorem 2.3, implies the weak RSP of A^T. From a mathematical point of view, we have completely characterized the weak stability of standard ℓ_1-minimization under this assumption (see Corollary 3.3). It is worth emphasizing several important features of the weak RSP of A^T.

(**i**) Uniform recovery of every k-sparse vector is a basic requirement in compressed sensing, and the classic KKT optimality condition is a fundamental tool for understanding the internal mechanism of ℓ_1-minimization methods. The weak RSP of A^T is a natural property capturing both the requirement of uniform recovery and the deepest property of any optimal solution to ℓ_1-minimization. So our assumption is actually a strengthened KKK optimality conditions by taking into account the requirement of uniform recovery. As a result, no matter what (deterministic or random) matrix A is used, the weak RSP of A^T is a fundamental property guaranteeing the success and stableness of ℓ_1-minimization methods in sparse data recovery. As shown by Corollary 3.3, this property cannot be relaxed without damaging the weak stability of ℓ_1-minimization, since it is a necessary and sufficient condition for ℓ_1-minimization to be weakly stable for any measurement $y \in \{Ax : \|x\|_0 \le k\}$.

(**ii**) Our analysis is different from the existing frameworks. It is based on the Hoffman's error bound for linear systems and the polytope approximation of the unit ℓ_2-ball. The weak RSP of A^T is a constant-free matrix property. The coefficients C, C_1 and C_2 in error bounds (2) and (3) are measured by the Robinson's constants, no matter the matrix property is constant-free (such as the weak RSP of A^T, RSP of order k of A^T, or NSP of order k) or is constant-dependent (such as the RIP, stable or robust stable NSP). Thus our analytic method yields a certain unified weak stability result irrespective of an individual assumption on A, provided that the imposed assumption implies the weak RSP of A^T (see Corollaries 3.5, 4.5 and 5.7).

(**iii**) Practical signals are often structured or with some prior information, and typical design matrices in practice are not Gaussian or Bernoulli. This makes the standard analysis and results (based on Gaussian and Bernoulli random matrices) difficult to apply in these situations. Thus the structured sparse data reconstruction recently becomes one of the active research areas in compressed sensing and applied mathematics. The weak RSP concept derived from optimality conditions of convex optimization can be easily adapted to these situations to interpret the behavior of more complex and general recovery problems. For instance, the so-called restricted RSP property of A^T was used to deal with the sign or support recovery of signals in 1-bit compressed sensing problems [49].

It is also worth mentioning that the analytic method in this paper is not difficult to be extended to the study of the weak stability of weighted ℓ_1-minimization problems (e.g., [13, 48, 47]), Dantzig selector [12], and Lasso problems [40, 29].

6 Conclusions

We have shown that the so-called weak range space property of the transposed design matrix is a sufficient constant-free condition for various ℓ_1-minimization problems to be (robustly and) weakly stable in sparse data reconstruction. For noise-free measurements, this matrix property turns out to be a necessary condition for standard ℓ_1-minimization to be weakly stable. All existing stability conditions (such as mutual coherence, RIP, NSP, or their variants) imply our assumption. As a result, certain unified weak stability results have been developed for ℓ_1-minimization under existing matrix properties. In particular, the weak stability under the constant-free null space property of order k and range space property of order k have been established in this paper. Our stability coefficients are measured by the Robinson's constants determined by the problem data. Our study indicates that the reconstruction error bounds via ℓ_1-minimization can be understood from Hoffman's error bounds for linear systems with compressed sensing matrices.

References

[1] J. Andersson and J.O. Strömberg, On the theorem of uniform recovery of structured random matrices, *IEEE Trans. Inform. Theory*, 60 (2014), pp. 1700–1710

[2] A. Bruckstein, D. Donoho and M. Elad, From sparse solutions of systems of equations to sparse modeling of signals and images, *SIAM Rev.*, 51 (2009), pp. 34–81.

[3] J. Cahill, X. Chen and R. Wang, The gap between the null space property and the restricted isometry property, *Linear Algebra Appl.*, 501 (2016), pp. 363–375.

[4] T. Cai, L. Wang, and G. Xu, New bounds for restricted isometry constants, *IEEE Trans. Inform. Theory*, 56 (2010), pp. 4388–4394.

[5] T. Cai and A. Zhang, Sharp RIP bound for sparse signal and low-rank matrix recovery, *Appl. Comput. Harmon. Anal.*, 35 (2013), pp. 74–93.

[6] T. Cai and A. Zhang, Sparse representation of a polytope and recovery of sparse signals and low-rank matrices, *IEEE Trans. Inform Theory*, 60 (2014), pp. 122–132.

[7] G.H.L. Cheang and A. R. Barron, A better approximation for balls, *J. Approx. Theory*, 104 (2000), pp. 183–203.

[8] E. Candès, The restricted isometry property and its implications for compressed sensing, *C.R. Math. Acad. Sci. paris*, 346 (2008), pp. 589–592.

[9] E. Candès, J. Romberg and T. Tao, Stable signal recovery from incomplete and inacurate measurements, *Comm. Pure Appl. Math.*, 59 (2006), pp. 1207–1223.

[10] E. Candès and T. Tao, Decoding by linear programming, *IEEE Trans. Inform. Theory*, 51 (2005), pp. 4203–4215.

[11] E. Candès and T. Tao, Near optimal signal recovery from random projections: Universal encoding stratiges? *IEEE Trans. Inform. Theory,* 52 (2006), pp. 5406–5425.

[12] E. Candes and T. Tao, The Dantzig selector: statistical estimation when p is much larger than n, *Ann. Statist.*, 35 (2007), pp. 2313–2351.

[13] E. Candès, M. Wakin and S. Boyd, Enhancing sparsity by reweighted ℓ_1 minimization, *J. Fourier Anal. Appl.*, 14 (2008), pp. 877–905.

[14] S. Chen, D. Donoho and M. Saunders, Atomic decomposition by ℓ_1-minimization, *SIAM J. Sci. Comput.*, 20 (1998), pp. 33–61.

[15] A. Cohen, W. Dahmen and R. Devore, Compressed sensing and best k-term aproximation, *J. Amer. Math. Soc.*, 22 (2009), pp. 211–231.

[16] D. Donoho and M. Elad, Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization, *Proc. Natl. Acad. Sci.*, 100 (2003), pp. 2197–2202.

[17] D. Donoho, M. Elad and V. Temlyahov, Stable recovery of sparse overcomplete representations in the presenxe of noise, *IEEE Trans. Inform. Theory,* 52 (2006), pp. 6–18.

[18] D. Donoho and X. Huo, Uncertainty principles and ideal atomic decomposition, *IEEE Trans. Inform. Theory,* 47 (2001), pp. 2845–2862.

[19] D. Donoho and B. Logan, Signal recovery and the large sieve, *SIAM J. Appl. Math.*, 52 (1992), pp. 577–591.

[20] R. Dudley, Matric entropy of some classes of sets with differentiable bounaries, *J. Approx. Theory,* 10 (1974), 227-236; Correction, *J. Approx. Theory*, 26 (1979), pp.192–193.

[21] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer, New York, 2010.

[22] Y. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*, Cambridge University Press, 2012.

[23] S. Foucart, Stability and robustness of ℓ_1-minimization with Weibull matrices and redandant distionaries, *Linear Algebra Appl.*, 441 (2014), pp. 4–21.

[24] S. Foucart and M. Lai, Sparsest solutions of undertermined linear systems via ℓ_p-minimization for $0 < q \leq 1$, *Appl. Comput. Harmon. Anal.*, 26 (2009), pp. 395–407.

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

[26] J. Fuchs, On sparse representations in arbitrary redundant bases, *IEEE Trans. Inform. Theory,* 50 (2004), pp. 1341–1344.

[27] M. Grasmair, O. Sherzer and M. Haltmeier, Necessary and sufficient conditions for linear convergence of l1-regularization, *Comm. Pure Appl. Math.*, 64 (2011), pp. 161–182.

[28] R. Gribonval and M. Nielsen, Sparse representation in unions of basis, *IEEE Trans. Inform. Theory*, 49 (2003), pp. 3320–3325.

[29] T. Hastie, R. Tibshirani and M. Wainwright, *Statistical Learning with Sparsity: The Lasso and Generalizations*, Chapman & Hall/CRC, 2015.

[30] A.J. Hoffman, On the approximation solution of systems of linear inequalities, *J. Res. Nat. Bur. Standards*, 49 (1952), pp. 263–265.

[31] S. Levy and P. Fullagar, Reconstruction of a sparse spike train from a portion of its spectrum and application to high-resolution deconvolution, *Geophysics*, 46 (1981), pp. 1235–1243.

[32] B. Logan, Properties of hign-pass signals, PhD thesis, Columbia University, 1965.

[33] O.L. Mangasarian, *Machine learning via polydedral concave minimization*, in Applied Mathematics and Parallel Computing-Festschrift for Klaus Ritter (H. Fischer, B. Riedmueller and S. Schaeffler eds.), Springer, Heidelberg, 1996, pp. 175–188.

[34] O.L. Mangasarian, Minimum-support solutions of polyhedral concave programs, *Optimization*, 45 (1999), pp. 149–162.

[35] Q. Mo and S. Li, New bounds on the restricted isometry constant δ_{2k}, *Appl. Comput. Harmon. Anal.*, 31 (2011), pp. 460–468.

[36] M. Plumbley, On polar polytopes and the recovery of sparse representations, *IEEE Trans. Infom. Theory*, 53 (2007), pp. 3188–3195.

[37] S. M. Robinson, Bounds for error in the solution set of a perturbed linear program, *Linear Algebra Appl.*, 6 (1973), pp. 69–81.

[38] Q. Sun, Recovery of sparsest signals via ℓ_q-minimization, *Appl. Comput. Harmon. Anal.*, 32 (2012), pp.329–341.

[39] H. Taylor, S. Banks and J. McCoy, Deconvolution with the ℓ_1 norm, *Geophysics*, 44 (1979), pp. 39–52.

[40] R. Tibshirani, Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B*, 58 (1996), pp. 267–288.

[41] J.A. Tropp, Greed is good: Algorithmic results for sparse approximation, *IEEE Trans. Inform. Theory*, 50 (2004), pp. 2231–2242.

[42] H. Zhang, W. Yin and L. Cheng, Necessary and sufficient conditions of solution uniqueness in 1-norm minimization, *J. Optim. Theory Appl.*, 164 (2015), pp. 109–122.

[43] H. Zhang, M. Yan and W. Yin, One condition for solution uniqueness and robustness of both l1-synthesis and l1-analysis minimization, *Adv. Comput. Math*, 42 (2016), pp. 1381–1399.

[44] Y. Zhang, Theory of compressive sensing via ℓ_1-mimimization: A non-RIP analysis and extensions, *J. Oper. Res. Soc. China*, 1 (2013), pp. 79–105.

[45] Y.B. Zhao, RSP-based analysis for sparsest and least ℓ_1-norm solutions to underdetermined linear systems, *IEEE Trans. Signal Process.*, 61 (2013), no. 22, pp. 5777–5788.

[46] Y.B. Zhao, Equivalence and strong equivalence between sparsest and least ℓ_1-norm non-negative solutions of linear systems and their applications, *J. Oper. Res. Soc. China*, 2 (2014), pp. 171–193.

[47] Y.B. Zhao and M. Kocvara, A new computational method for the sparsest solutions to systems of linear equations, *SIAM J. Optim.*, 25 (2015), pp. 1110–1134.

[48] Y.B. Zhao and D. Li, Reweighted ℓ_1-minimization for sparse solutions to underdetermined linear systems, *SIAM J. Optim.*, 22 (2012), pp. 1065–1088.

[49] Y.B. Zhao and C. Xu, 1-bit compressive sensing: Reformulation and RRSP-based sign recovery theory, *Sci. China Math.*, 59 (2016), pp. 2049–2074.