Signal Processing xxx (xxxx) xxx

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

A novel property to modify weighted l_1 minimization for improved compressed sensing

Yudong He *, Baeck Hyun Woo, Fauzan Abdurrahim, Richard H.Y. So

Department of Industry Engineering and Decision Analytics, The Hong Kong University of Science and Technology, Hong Kong

ARTICLE INFO

Keywords:
Sparsity
Sparsity-inducing property
Weighted l_1 minimization
Compressed sensing
Underdetermined linear system

ABSTRACT

Weighted l_1 minimization schemes are common methods to achieve compressed sensing (CS). However, they fail in the presence of inaccurate prior knowledge or improper scaling of weights due to inappropriately assigned large weights causing large and destructive errors in signal recovery. This paper proposes a theory-based algorithm to identify and correct such destructive weights for each signal entry. The enhancement is achieved through a novel sparsity-inducing property (SIP) which establishes a necessary condition for successful signal recovery. SIP outperforms existing properties such as coherence, restricted isometry property, and nullspace property by indicating which signal entries fail to be recovered. This unique advantage enables us to correct destructive weights that do not satisfy the SIP condition, making signal recovery successful where it previously failed. Results from many numerical experiments demonstrate that our proposed method can improve the signal recovery capability, robustness, and stability of the weighted l_1 minimization for a wide range of applications, including sparse and compressive signal recovery, noise-aware recovery, sparse error correction, fast image acquisition, and sub-Nyquist sampling.

1. Introduction

1.1. Compressed sensing and weighted l_1 minimization

Compressed sensing (CS) refers to the recovery of signal with n coefficients (e.g., $\mathbf{x}_o \in \mathbb{R}^n$) from m linear measurements where m is less than n [1,2]. This requires $\mathbf{x}_o$ to be sparse (i.e., most of its coefficients are zero) or compressible (not strictly sparse but most of its coefficients are close to zero) [3–5]. CS is widely applied in fields such as medical imaging [6–8], radar imaging [9–11], seismology [12,13], astronomy [14,15], and facial recognition with severe occlusion [16]. CS is achieved through an optimization framework called l_1 minimization [3]. For simplicity, in the absence of noise, l_1 minimization can be formulated as

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1, \quad \text{s.t. } \mathbf{y} = \mathbf{A}\mathbf{x}, \tag{1}$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is the solution vector (signal), $\|\mathbf{x}\|_1 := \sum_i |x_i|$, called l_1 norm, and $\mathbf{y} \in \mathbb{R}^m$ contains known linear measurements with $\mathbf{y} = \mathbf{A}\mathbf{x}_o$ for some known sensing matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ (also called measurement matrix in literature). However, in practice, l_1 minimization cannot provide accurate results due to insufficient sparsity levels of $\mathbf{x}_o$. To increase recovery accuracy, weights have been assigned on entries of $\mathbf{x}_o$. Such methods have been referred to as weighted l_1 minimization schemes:

$$\min_{\mathbf{x}} \|\mathbf{W}\mathbf{x}\|_1, \quad \text{s.t. } \mathbf{y} = \mathbf{A}\mathbf{x}, \tag{2}$$

where $\mathbf{W} = \text{diag}(w_1, \ldots, w_n)$, called weight matrix, is a diagonal matrix with positive main diagonal entries, i.e., $w_i > 0, \forall i \in \{1, \ldots, n\}$. A typical weighting principle is to put large weights on small entries of $\mathbf{x}_o$, and small weights on large entries so as to ensure the reconstructed signal has the same sparsity pattern of $\mathbf{x}_o$ [17].

1.2. Prior works on weighted l_1 minimization and their shortcomings

Assigning appropriate weights has been the subject of many studies. These studies can be classified into two categories. Studies belonging to the first category use signal knowledge as input to construct weights. Specifically, Vaswani and Lu [18] proposed a weighted l_1 minimization called modified-CS, which assigns zero weight to the support of the true signal and unit weight to others. Friedlander et al. [19] proposed a more general weighting scheme by putting a small but nonzero weight on the support. Mansour and Yilmaz [20] and Needell et al. [21] extended the two-weight setting to a multiple-weight setting by assuming that multiple support estimates are available. Other subsequent

* Corresponding author.
E-mail address: yhebh@connect.ust.hk (Y. He).

https://doi.org/10.1016/j.sigpro.2024.109828
Received 14 August 2024; Received in revised form 6 November 2024; Accepted 26 November 2024
Available online 4 December 2024
0165-1684/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

studies on support-based weighting schemes can be found in [22–24]. Besides utilizing the support information, Khajehnejad et al. [25] used uncertainty information where coefficients of the true signal were partitioned into two sets according to their probabilities of being nonzero. Similarly, Zhang et al. [26] proposed a binary signal recovery algorithm utilizing the prior probability of each entry being nonzero. These studies suffer a common problem, namely, when the required prior knowledge is inaccurate, these weighting schemes generate overlarge weights on nonzero entries leading to huge and destructive errors in signal recovery (i.e., destructive weights). Existing analyses, which utilize concepts such as restricted isometry property (RIP) [27], nullspace property (NSP) [28], weighted robust NSP [29], and coherence [30], can quantitatively assess how the accuracy of required knowledge affect recovery performance. However, all of them fail to provide methods to assign more appropriate weights in the presence of imperfect prior knowledge.

The second category of weighting schemes does not assume any prior knowledge. Instead, these schemes involve solving a sequence of weighted l_1 minimization problems where the weights of the current weighted l_1 minimization are set to the reciprocal of the solution of the previous weighted l_1 minimization. These weighting schemes are known as reweighting schemes. Following the first proposal of Candes et al. [31], various reweighting schemes have been developed for different applications. For example, Peng et al. [32] proposed a reweighting scheme for image restoration, Zhang et al. [33] proposed a reweighting scheme for phase retrieval from magnitude-only measurements, and Wang et al. [34] proposed another for bearing fault diagnosis. These schemes applied a single scaling factor to scale all weights through empirical trial-and-error methods. The purpose was to control the rate of change of the weights with respect to the signal coefficients and prevent the weights from becoming infinite. Unfortunately, these schemes did not provide a theoretical explanation for their empirical choices of scaling factors. Worst, their solutions needed to be tailored to specific characteristics of the signal, the measurement matrix, and other factors [31–36]. Any changes in signal characteristics can render their schemes ineffective [37–39]. In summary, there are two deficiencies in existing weighted l_1 minimization schemes: (1) inaccurate prior knowledge and (2) improper scaling of weights. Both these points lead to the assignment of destructive (i.e., inappropriate) weights. Unfortunately, we currently lack a method to prevent it.

1.3. Research gap

In many applications, especially those involving dynamic situations such as transmission or sampling of non-stationary signals, prior knowledge of the signal will be inaccurate causing failure in existing weighted l_1 minimization schemes. While customized empirical simulation can determine a scaling factor for the weights, a theory-based analytical method to determine the range of destructive weight per signal entry that can prevent the assignment of a single destructive weight could not be found.

1.4. Contribution

In this paper, we propose a novel theory-based method to identify and correct destructive weights in weighted l_1 minimization. We begin by defining a new property of sensing matrices and deriving a necessary condition for successful signal recovery by weighted l_1 minimization equipped with arbitrary weights. This property, which we refer to as the sparsity-inducing property (SIP), differs fundamentally from existing properties such as coherence, RIP, and NSP. The latter can only assert whether recovery succeeded or failed, while SIP provides analytical details and determines signal coefficients that are failed to be recovered. For the signal coefficients that are not recovered, our algorithm will determine the maximum weight values beyond which destructive errors will occur. These calculations are conducted for each signal entry of $\mathbf{x}_o$ and a weight-clipping algorithm will correct (i.e., reduce) the value of the destructive weights. The benefits of applying the weight-clipping method include but are not limited to enhancing the overall quality of weights to improve the signal recovery capabilities of weighted l_1 minimization; increasing the robustness of weighting schemes, enabling them to maintain satisfactory performance despite inaccurate prior knowledge or inappropriate weight scaling; and improving the stability of constructing high-quality weights. We demonstrate these benefits and the versatility of the proposed method by applying it to representative applications that utilize different l_1-based optimization frameworks. These frameworks include equality-constrained l_1 minimization, inequality-constrained l_1 minimization, constraint-free l_1 minimization, l_1-analysis minimization, and total variation minimization.

Our contribution can be summarized in three key points: (1) We derive a new property, SIP, providing requirements of successful signal recovery weighted l_1 minimization methods. (2) We introduce weight-clipping inspired by SIP to correct excessive weights, enhancing signal recovery. (3) We integrate weight-clipping into Candes' reweighting scheme and evaluate its improvement through extensive simulations across representative applications of CS.

1.5. Organization

The rest of the paper is organized as follows. In Section 2, we introduce SIP and present our theoretical results. In Section 3, we introduce the proposed weight-clipping method. Section 4 describes experiments that demonstrate the effectiveness and broad applicability of the proposed method in terms of sparse and compressible signal recovery, noise-aware recovery, sparse error correction, image processing, and sub-Nyquist sampling. Finally, we conclude and discuss future research directions in Section 5.

2. Theoretical results

2.1. SIP in unweighted l_1 minimization

l_1 minimization appears to possess an oracular ability, as it seemingly discerns which signal coefficients are nonzero. However, it occasionally behaves blindly, wrongly estimating a nonzero signal coefficient as zero. We note that this phenomenon is associated with a set of scalar values linked to the sensing matrix, which we refer to as 'sparsity-inducing constants (SICs)' in this paper. As our findings can be extended to the complex domain, not limited to the real domain, we use $\mathbb{F}$ to denote both the real domain $\mathbb{R}$ and complex domain $\mathbb{C}$.

Definition 1 (*Sparsity-Inducing Constant*). Given a matrix $\mathbf{A} = [\mathbf{a}_1|\mathbf{a}_2|\cdots|\mathbf{a}_n] \in \mathbb{F}^{m\times n}$ where $m < n$, let $\mathbb{B}_i$ denote the set containing all the $m \times m$ full rank sub-matrices of $\mathbf{A}$ except the ith column. For each integer $i = 1, 2, \ldots, n$, define the ith sparsity-inducing constant r_i of the matrix $\mathbf{A}$ as

$$r_i = \min_{\mathbf{B}\in\mathbb{B}_i} \|\mathbf{B}^{-1}\mathbf{a}_i\|_1. \tag{3}$$

Particularly, if $\mathbb{B}_i$ is a null set, then define $r_i = +\infty$.

The SIC possesses the following two types of invariance.

Proposition 1 (*Permutation Invariance*). *r_i remains unchanged after permuting any column of $\mathbf{A}$ except the ith column.*

Proof. Denote $\mathbf{P} \in \mathbb{R}^{m\times m}$ any column-wise permutation matrix. Then $\|(\mathbf{BP})^{-1}\mathbf{a}_i\|_1 = \|\mathbf{PB}^{-1}\mathbf{a}_i\|_1 = \|\mathbf{B}^{-1}\mathbf{a}_i\|_1$. ■

Proposition 2 (*Invertible Linear Transformation Invariance*). *r_i remains unchanged after left multiplying $\mathbf{A}$ by any invertible matrix.*

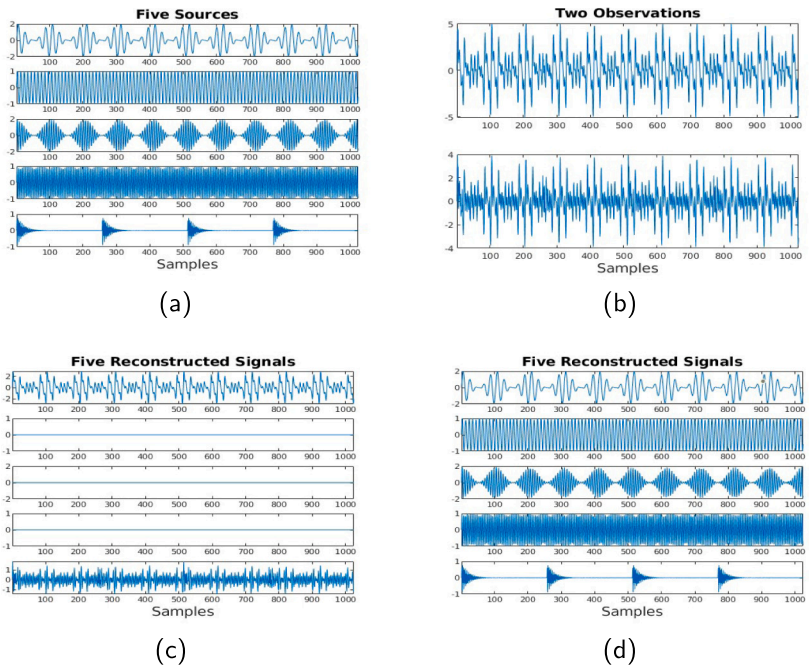

Fig. 1. Example of a failed signal separation due to the mixing matrix $\mathbf{A}$ having SICs less than one. (a) Five source signals. (b) Two mixed signals. (c) Failed l_1 minimization reconstruction with destructive errors. (d) Successful reconstruction (31 dB signal-to-noise ratio) by the proposed weight-clipping method deduced by SIP theory (see Section 3).

Proof. Denote $\mathbf{C} \in \mathbb{F}^{m\times m}$ any invertible matrix. Then $\|(\mathbf{CB})^{-1}\mathbf{Ca}_i\|_1 = \|\mathbf{B}^{-1}\mathbf{C}^{-1}\mathbf{Ca_i}\|_1 = \|\mathbf{B}^{-1}\mathbf{a}_i\|_1$. ■

Permutation Invariance can assist in narrowing the search space for computing $\{r_i\}_{i=1}^n$, and Invertible Linear Transformation Invariance highlights that any invertible linear transformation of the measurements or sensing matrix will not affect any properties solely determined by $\{r_i\}_{i=1}^n$. We find that the blind behavior of l_1 minimization mentioned earlier is related to SIC. Note that, we assume that both optimization problems (4) and (2) have a unique optimal solution, and those problems without a unique optimal solution are out of the scope of our theorem.

Theorem 1 (*Compressed Loss Theorem*). *Given a matrix $\mathbf{A} \in \mathbb{F}^{m\times n}$ where $m < n$, let r_i denote its ith sparsity-inducing constant. If $r_i \le 1$, then $x_i^* = 0$, where x_i^* is the ith entry of the unique optimal solution to the problem*

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1, \quad s.t.\ \mathbf{y} = \mathbf{Ax}. \tag{4}$$

Proof. See Appendix A where Proposition 1 is used. ■

Based on Theorem 1, when $r_i \le 1$, the ith signal coefficient estimate becomes zero, irrespective of its true value. This implies that the compressive-recovery framework employing l_1 minimization results in the loss of certain information. Consequently, the concept of the proposed SIP emerges, characterized by the generation of zero estimates without considering the true values when recovering the signal from the compressed data. This aligns with the intuition that if a column vector $\mathbf{a}_i$ is relatively small (indicating a small r_i), the associated signal coefficient x_i contributes less to satisfying the equality constraint. In other words, the value of x_i is less significant. Therefore, optimizing $\|\mathbf{x}\|_1$ will result in a zero estimate for x_i.

For the sake of illustration, let us say:

$$\mathbf{A} = \begin{bmatrix} 100 & 0 & 1 \\ 0 & 100 & 1 \end{bmatrix}.$$

For any $\mathbf{x}_o = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}^\mathsf{T}$, we have $\mathbf{x}^* = \begin{bmatrix} x_1^* & x_2^* & 0 \end{bmatrix}^\mathsf{T}$ with $x_1^* = x_1 + 0.01x_3$ and $x_2^* = x_2 + 0.01x_3$ such that $\mathbf{Ax}_o = \mathbf{Ax}^*$ and $\|\mathbf{x}_o\|_1 \ge \|\mathbf{x}^*\|_1$ (the equality holds only if $x_3 = 0$). Hence, solving the problem (4) fails to recover x_3 if $x_3 \ne 0$. This failure can be identified and corrected by the proposed SIP.

The necessary condition for the perfect recovery of signals based on SIP is as follows:

Theorem 2 (*Necessary Condition under SIP*). *Define $\mathbb{I}$ as the support of $\mathbf{x}_o$, i.e., $\mathbb{I} := supp(x_o) \subseteq \{1, \dots, n\}$ and $\mathbb{S}$ as the set containing the column index of $\mathbf{A}$ whose corresponding SIC is less or equal to one, i.e., $\mathbb{S} = \{i \mid r_i \le 1\}$. Given $\mathbf{y} = \mathbf{Ax}_o$, the solution to the l_1 minimization (4) is exactly $\mathbf{x}_o$ only if $\mathbb{I} \cap \mathbb{S} = \varnothing$.*

Proof. This is a straightforward implication of Theorem 1. ■

SIP differs from existing properties characterizing a sensing matrix such as coherence, RIP, and NSP. Next, we present a case where coherence, RIP, and NSP become uninformative, yet SIP provides valuable information. We also show how destructive the error is when the necessary condition under SIP is not satisfied. We applied l_1 minimization to achieve signal separation, a technique widely used in audio signal processing [40–42]. We generated five static signals (illustrated in Fig. 1(a)) incorporating two modulated signals, two periodic signals, and one pulse signal, and mixed them into two-channel signals (depicted in Fig. 1(b)) by linearly combining them using a matrix encoding their amplitude amplification and attenuation

$$\mathbf{A} = \begin{bmatrix} 1.4374 & 1.1547 & 0.9639 & 0.8141 & 0.7148 \\ 0.7043 & 0.7559 & 0.8381 & 1.0044 & 1.3572 \end{bmatrix}.$$

The five static signals are not sparse in the time domain but possess non-overlapping narrow frequency bands. Therefore, we performed the l_1 minimization in the Short-time Fourier transform (STFT) domain, and then obtained time-domain signals via inverse STFT. The mixing matrix $\mathbf{A}$ does not respect coherence, RIP, or NSP. Except that the recovery will not be perfect, no more details can be obtained from these properties. Nonetheless, given the SICs of the matrix

$$\mathbf{r} = \begin{bmatrix} 1.6757 & 0.8982 & 0.8532 & 0.8687 & 2.1076 \end{bmatrix},$$

SIP asserts that the recovery will not only fail (according to Theorem 2) but also the middle three channel signals will be erroneously reconstructed as zero (according to Theorem 1). The numerical result

displayed in Fig. 1(c) corroborates this assertion. This additional information provided by SIP is valuable as it opens up a way to modify l_1 minimization. Fig. 1(d) exhibits the successful recovery outcome after applying the proposed weight-clipping method, an application of SIP theory to improve the recovery ability of l_1 minimization, which will be elaborated in Section 3. In summary, the major difference between SIP and prior works, especially with regard to coherence, RIP, and NSP, is that SIP pinpoints specific signal recovery failures; that is, it reveals which signal coefficients on the support set fail to be recovered. Prior works cannot do so.

2.2. SIP in weighted l_1 minimization

Weighted l_1 minimization is more general than unweighted l_1 minimization, as the latter can be viewed as a special case of the former. Therefore, the theorem presented in the previous section is extended below to generalize SIP for weighted l_1 minimization:

Theorem 3 (*Generalized Necessary Condition under SIP*). *Define $\mathbb{I}$ as the support of $\mathbf{x}_o$ and $\mathbb{S}_W$ as the set containing the column index of $\mathbf{AW}^{-1}$ whose corresponding SIC is less or equal to one. Given $\mathbf{y} = \mathbf{Ax}_o$, the solution to the weighted l_1 minimization (2) is exactly $\mathbf{x}_o$ only if $\mathbb{I} \cap \mathbb{S}_W = \varnothing$.*

Proof. Suppose $\mathbf{x}^*$ is the unique optimal solution to a weighted l_1 minimization with a sensing matrix $\mathbf{A}$ and a weight matrix $\mathbf{W}$, and $\mathbf{z}^*$ is the unique optimal solution to an unweighted l_1 minimization with a sensing matrix $\mathbf{AW}^{-1}$. These two optimization problems are equivalent in the sense that $\mathbf{x}^* = \mathbf{W}^{-1}\mathbf{z}^*$. Hence, $\mathbf{x}^*$ and $\mathbf{z}^*$ share the same support. If $\mathbb{I} \cap \mathbb{S}_W \neq \varnothing$, according to Theorem 1, there are zero values at entries of $\mathbf{z}^*$ indexed by $\mathbb{I}$. Therefore, $\mathbf{x}^*$ does not share the same support of $\mathbf{x}_o$, which means $\mathbf{x}^* \neq \mathbf{x}_o$. ■

If $\mathbb{S}_W$ is an empty set, the requirement of SIP is satisfied. However, if $\mathbb{S}_W$ is not an empty set and has an intersection with the support set $\mathbb{I}$, the corresponding weights (one-to-one mapping by the index contained in $\mathbb{I} \cap \mathbb{S}_W$) are destructive because, according to Theorem 3, the corresponding nonzero coefficients in the support of $\mathbf{x}_o$ will be erroneously estimated as zero, leading to failed signal recovery. We use this new insight to quantitatively explain the limitation of using a single scaling factor for scaling weights in the iterative reweighting scheme. For each trial, we generated a sparse signal of length $n = 15$ with k nonzero entries, i.e., $k = \|\mathbf{x}_o\|_0$ where $\|\cdot\|_0$ is the l_0 norm. The locations of the nonzero entries were randomly selected and their values were chosen independently from a standard Gaussian distribution. We set $m = 8$ and sampled a $m \times n$ random matrix $\mathbf{A}$ with independent and identically distributed (i.i.d.) standard Gaussian entries. Given the data $\mathbf{y} = \mathbf{Ax}_o$, weights were constructed following the classic iterative reweighting scheme [31] with four iterations. The reweighting process involved a sequence of iterations using the solution from the previous iteration to calculate the weights for the current iteration: for each $i = 1, \ldots, n$,

$$w_i^{(l+1)} = \frac{1}{|x_i^{(l)}| + \epsilon}, \tag{5}$$

where $x_i^{(l)}$ is the optimal solution to the weighted l_1 minimization at the lth iteration, $w_i^{(l+1)}$ is the weight for the $(l+1)$th iteration, and ϵ is a scaling factor used to scale the weights for preventing infinity in the case of $x_i^{(l)} = 0$. We conducted 500 trials, recording $|\mathbb{S}_W|$ (the size of $\mathbb{S}_W$) for each trial. The results are shown in Fig. 2(a). As ϵ decreased, the frequency of the occurrence of a large $|\mathbb{S}_W|$ increased, implying the reweighting process more easily generated destructive weights. $\mathbb{S}_W$ was not an empty set in more than 95% of trials when $\epsilon = 0.01$. This percentage was much higher than the 7% observed in the unweighted l_1 minimization. This finding explains the lack of efficacy of the iterative reweighting scheme when ϵ is too small. Although it is easier for $\mathbb{S}_W$ to be an empty set at larger ϵ values, as shown in Fig. 2(b), the empirical probability of perfect recovery[1] did not increase beyond 0.1 and even decreased at $\epsilon = 10$ because the weights became almost identical. The maximum improvement of the probability of perfect recovery was less than 0.1 through an exhaustive search of ϵ. This numerical experiment demonstrates the limitation of using a single scaling factor (i.e., ϵ) to uniformly scale all weights. This limitation can be overcome by the proposed weight-clipping method that independently regulates the weight of each entry according to its SIC value. This ability will be introduced in the next section.

3. Weight-Clipping

3.1. Clipping overly large weights

According to the SIP theory, an effective weight matrix should make $\mathbb{S}_W$ an empty set. If $\mathbb{S}_W$ is not empty, an appropriate modification to the weight matrix should be made to ensure that $\mathbb{S}_W$ becomes empty. Our idea is that if r_i, the ith SIC of $\mathbf{AW}^{-1}$, is less than or equal to one, then w_i is a destructive weight identified by Theorem 3. We correct w_i by reducing its value to make r_i greater than one. If the r_i is already greater than one, we keep the value of w_i. This operation essentially clips overly large weights, with the following mathematical expression: for each $i = 1, \ldots, n$,

$$w_i' = \begin{cases} r_i w_i/\alpha, & \text{if } r_i < \alpha, \\ w_i, & \text{otherwise,} \end{cases} \quad \forall i = 1, \ldots, n, \tag{6}$$

where w_i is the original weight, w_i' is the modified/clipped weight, r_i is the ith SIC of the matrix $\mathbf{AW}^{-1}$, and α is a parameter introduced to allow a flexible clipping range. After clipping, the value of weights gets smaller because clipping occurs when $r_i/\alpha < 1$. To ensure r_i is greater than one after clipping, the parameter α should be chosen greater than one. For brevity, we refer to this weight modification method (6) as weight-clipping. Three numerical experiments were implemented to demonstrate the effectiveness of the proposed method.

Firstly, the experiment depicted in Fig. 1 was replicated. Unweighted l_1 minimization can be viewed as a weighted l_1 minimization with the identity matrix as the weight matrix. Examining the SIC values shows that assigning unit weights to the middle three channel signals is inappropriate. We applied the weight-clipping with $\alpha = 1.1$ to the identity matrix and used the clipped weight matrix to reconstruct the signal. As evidenced in Fig. 1(d), compared to the unweighted l_1 minimization shown in Fig. 1(c), the reconstruction with the modified weight matrix showed significant improvement.

In the second experiment, we repeated the experiment whose results are shown in Fig. 2 with the application of the proposed weight-clipping at each reweighting iteration. As demonstrated in Fig. 3, the perfect recovery probability noticeably increased (by 0.3 when $k = 4$) compared to the original non-clipping reweighted l_1 minimization. Additionally, it was found that a larger value of α (e.g., $\alpha = 3$) led to better results than a smaller value (e.g., $\alpha = 1.1$). This is because when $\alpha = 1.1$, certain nonzero signal coefficients were still inaccurately estimated as zero despite their corresponding r_i values exceeding (but close to) one. By Increasing α, r_i values rose, leading to successful recovery. Consequently, it is hypothesized that stronger necessary conditions than that in Theorem 3 may exist. Identifying such conditions will be a valuable focus for future research.

The third experiment is an application of the proposed weight-clipping to a weighting scheme that requires prior support knowledge. Suppose $\mathbb{T}_o \subset \{1, \ldots, n\}$ is the support of the original signal $\mathbf{x}_o$, $\mathbb{T}$ is an estimate of $\mathbb{T}_o$. For simplicity, let $\mathbb{T}$ be a subset of $\mathbb{T}_o$, and $|\mathbb{T}| = [acc \cdot |\mathbb{T}_o|]$, where $acc \in (0, 1)$ is the estimation accuracy and $[\cdot]$ means

[1] A perfect recovery is defined if the maximum absolute difference between $\mathbf{x}_o$ and its estimate $\mathbf{x}_{\text{est}}$ is less than $1\text{e}-3$, i.e., $\|\mathbf{x}_o - \mathbf{x}_{\text{est}}\|_\infty < 1\text{e}-3$ where $\|\cdot\|_\infty$ is the infinity norm.

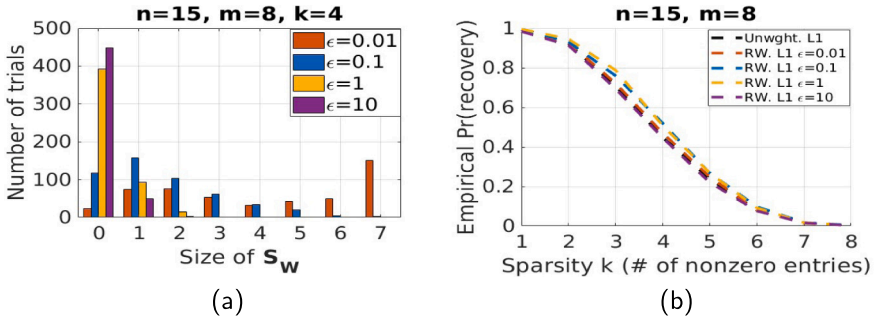

Fig. 2. (a) Histogram of the size of $\mathbb{S}_W$ defined in Theorem 3. Weights are constructed by the 'iterative reweighting scheme [31]' with different values of ϵ. (b) The empirical probability of perfect recovery[1], as a function of sparsity level $k = \|\mathbf{x}_o\|_0$.

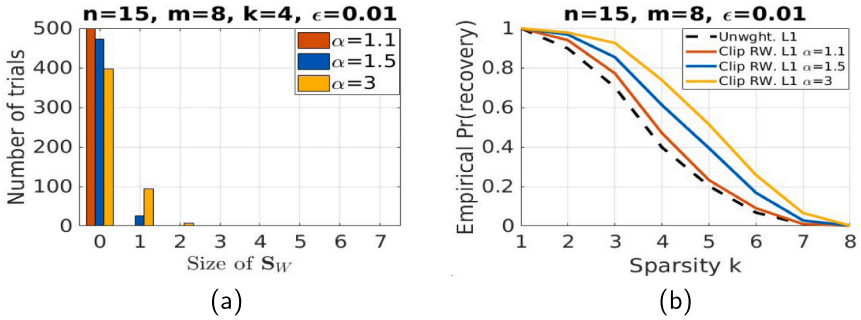

Fig. 3. (a) Histogram of the size of $\mathbb{S}_W$ defined in Theorem 3. Weights are constructed by the 'iterative reweighting scheme [31]' with a fixed $\epsilon = 0.01$, and at each reweighting iteration clipped by the proposed weight-clipping (6) with different values of α. (b) The empirical probability of perfect recovery (declared when $\|\mathbf{x}_o - \mathbf{x}_{\text{est}}\|_\infty < 1\text{e}-3$).

taking the closest integer. We try to solve the following optimization problem

$$\min_{\mathbf{x}} \sum_{i=1}^{n} |w_i x_i|, \text{s.t. } \mathbf{y} = \mathbf{A}\mathbf{x}, w_i = \begin{cases} 0, & i \in \mathbb{T} \\ 1, & \text{otherwise.} \end{cases} \tag{7}$$

This problem is initially introduced in [18]. Its limitation is that the recovery ability is sensitive to the accuracy of the support estimation as a large weight can be erroneously assigned to a nonzero entry leading to zero estimation. Herein, we apply weight-clipping to control the size of the weights, improving the quality of the weights when the support estimation is inaccurate, i.e., $acc < 1$. A total of 500 recovery trials were conducted, with each trial generating a sparse signal $\mathbf{x}_o$ of length $n = 15$, containing k nonzero Gaussian entries. The sparse signal was then measured by a randomly generated Gaussian matrix, yielding $m = 8$ measurements. The support-based weighted l_1 minimization (7) was applied to reconstruct $\mathbf{x}_o$ from these measurements, serving as a benchmark for recovery performance comparison after implementing the proposed weight-clipping (6). We tried two estimation accuracy levels, i.e., $acc = 0.5$ and $acc = 0.8$. The results, presented in Fig. 4, show a notable improvement after applying weight-clipping, particularly when $acc = 0.8$. This experiment demonstrates that the proposed weight-clipping can improve weight quality in the presence of imperfect prior knowledge.

3.2. Numerical upper bounds of SIC

Weight-clipping utilizing the upper bounds of SIC is as effective as utilizing exact SIC values if the parameter α is chosen wisely but the former is much more computationally efficient. In this regard, we introduce two algorithms capable of computing the upper bounds of SIC in polynomial time. Recall the definition of SIC (see Definition 1), the value of SIC is the minimal l_1 norm of $\mathbf{B}^{-1}\mathbf{a}$ where $\mathbf{B}$ is a submatrix of $\mathbf{A}$. The idea of computing an upper bound of SIC is to find a submatrix (say $\hat{\mathbf{B}}$) that gives a reasonably small l_1 norm. Since $\|\mathbf{B}^{-1}\mathbf{a}\|_1$ is inversely proportional to the absolute determinant of $\mathbf{B}$ ($\|\mathbf{B}^{-1}\mathbf{a}\|_1 = \|\text{adj}(\mathbf{B})\mathbf{a}\|_1 / |\det(\mathbf{B})|$ where $\det(\mathbf{B})$ is the determinant of $\mathbf{B}$ and $\text{adj}(\mathbf{B})$ is the adjugate matrix of $\mathbf{B}$), $\hat{\mathbf{B}}$ could be the submatrix with the largest possible absolute determinant. Unfortunately, finding the submatrix with the largest absolute determinant is NP-hard [43]. Thus, we use the heuristic algorithm, Khachiyan's algorithm (re-expressed in Algorithm 1 for convenience and consistency) proposed in [44] to find a submatrix with a reasonably large absolute determinant. In Algorithm 1, $\|\cdot\|_2$ denotes the l_2 norm, and $(\cdot)^{\mathsf{T}}$ denotes the transpose for a real-valued vector or conjugate transpose for a complex-valued vector. $\mathbf{A}_{\mathbb{S}}$ is a column submatrix ($\mathbb{S}$ contains the column indices) of $\mathbf{A}$. Then, the upper bound of SIC is given by $\|\hat{\mathbf{B}}^{-1}\mathbf{a}\|_1$. The specific procedure is presented in Algorithm 2.

Algorithm 1: Khachiyan's Algorithm to Find the Largest Absolute Determinant Submatrix [44].

Input: $\mathbf{A} \in \mathbb{F}^{m\times n}$
Output: $\mathbf{B} \in \mathbb{F}^{m\times m}$
$\mathbf{V} \leftarrow \mathbf{A}$;
$\mathbb{S} \leftarrow \emptyset$;
while $|\mathbb{S}| < m$ **do**
  Search for the largest l_2 norm column vector $\mathbf{v}_j \in \mathbf{V}$;
  $\mathbb{S} \leftarrow \mathbb{S} \cup j$;
  $\mathbf{V} \leftarrow \mathbf{V} - \mathbf{v}_j\mathbf{v}_j^{\mathsf{T}}\mathbf{V}/\|\mathbf{v}_j\|_2^2$;
end
$\mathbf{B} \leftarrow \mathbf{A}_{\mathbb{S}}$;

Inspection of Algorithm 2 indicates that Algorithm 1 needs to be executed n times, some of which may have duplicate outputs. Therefore, Algorithm 2 has certain redundancy in the calculation. In some applications such as MRI image processing, n can be large. If one would like to shorten the computation time, we can run a slightly modified Algorithm 1 once and reuse its output. This faster procedure which gives the second (perhaps larger) upper bound of SIC is presented in Algorithm 3.

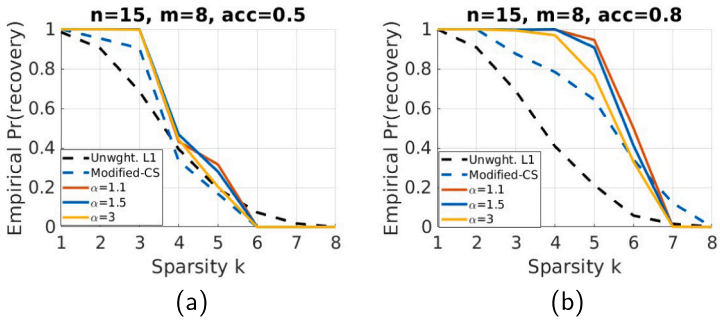

Fig. 4. The empirical probability of perfect recovery as a function of sparsity level $k = \|\mathbf{x}_o\|_0$. The black dash curve represents the unweighted l_1 minimization. The blue dash curve represents the support-based weighted l_1 minimization [18]. The solid curves represent support-based weighted l_1 minimization incorporating the proposed weight-clipping (6) with different values of α. (a) The support estimation accuracy is 0.5. (b) The support estimation accuracy is 0.8.

Algorithm 2: Algorithm to Calculate an Upper Bound of SIC.

Input: Matrix $\mathbf{A} \in \mathbb{F}^{m\times n}$ with $m < n$
Output: $\{\hat{r}_i\}_{i=1}^n$, the upper bound of SICs of $\mathbf{A}$
$i \leftarrow 1$;
while $i \le n$ **do**
  $\mathbf{a}_i$ represents the i-th column vector of $\mathbf{A}$;
  $\mathbf{A}_i$ represents the matrix containing all the column vectors of $\mathbf{A}$ except $\mathbf{a}_i$;
  if $\mathbf{A}_i$ *is not full row-rank* **then**
    $\hat{r}_i \leftarrow \infty$;
  else
    Execute Algorithm 1 with the input of $\mathbf{A}_i$, and denote the output as $\mathbf{B}_i$;
    $\hat{r}_i \leftarrow \|\mathbf{B}_i^{-1}\mathbf{a}_i\|_1$;
  end
  $i \leftarrow i + 1$;
end

Algorithm 3: Faster Algorithm to Calculate an Upper Bound of SIC.

Input: Matrix $\mathbf{A} \in \mathbb{F}^{m\times n}$ with $m < n$
Output: $\{\hat{r}_i\}_{i=1}^n$, the upper bound of SICs of $\mathbf{A}$
$\mathbf{V} \leftarrow \mathbf{A}$;
Initialize a sequence $(s_k)_{k=1}^{m+1}$;
$k \leftarrow 1$;
while $k \le m + 1$ **do**
  Search for the largest l_2 norm column vector $\mathbf{v}_j \in \mathbf{V}$;
  $s_k \leftarrow j$;
  $\mathbf{V} \leftarrow \mathbf{V} - \mathbf{v}_j\mathbf{v}_j^{\mathrm{T}}\mathbf{V}/\|\mathbf{v}_j\|_2^2$;
  $k \leftarrow k + 1$;
end
$i \leftarrow 1$;
while $i \le n$ **do**
  $\mathbf{a}_i$ represents the vector of the i-th column of $\mathbf{A}$;
  $\mathbb{S}_{m,i}$ represents a set which contains the first m elements of $(s_k)_{k=1}^{m+1}$ except the element i;
  $\mathbf{B}_i \leftarrow \mathbf{A}_{\mathbb{S}_{m,i}}$;
  $\hat{r}_i \leftarrow \|\mathbf{B}_i^{-1}\mathbf{a}_i\|_1$;
  $i \leftarrow i + 1$;
end

4. Numerical experiments

In order to verify and establish the validity and benefits of the proposed weight-clipping, seven numerical experiments were conducted to examine the effect of three factors: signal distribution, noise, and sparse model. The design of these experiments follows the test pipeline of the well-known iterative reweighting scheme [31]. Five types of l_1 minimization were involved. Specifically, the use of equality-constrained l_1 minimization to investigate Gaussian, Bernoulli, and compressible signal recovery is reported in Sections 4.1, 4.2, and 4.3; Inequality-constrained l_1 minimization to examine the impact of noise is reported in Section 4.4; Unconstrained l_1 minimization, total variation minimization, and l_1-analysis minimization to examine three common sparse models are reported in Sections 4.5, 4.6, and 4.7, respectively. The aim of these experiments is to demonstrate that the proposed weight-clipping can help weighted l_1 minimization achieve better, more robust, and more stable CS in various situations.

During the experiment, weight-clipping was added to the iterative reweighting scheme, and the modified algorithm was called clipping-reweighted l_1 minimization, whose equality-constrained version is summarized as follows:

1. Set the iteration count l to zero and $w_i^{(0)} = 1$, $i = 1, \dots, n$.
2. Solve the weighted l_1 minimization problem
$$\mathbf{x}^{(l)} = \arg\min_{\mathbf{x}} \|\mathbf{W}^{(l)}\mathbf{x}\|_1, \quad \text{s.t. } \mathbf{y} = \mathbf{A}\mathbf{x}. \tag{8}$$
3. Update the weight: for each $i = 1, \dots, n$,
$$w_i^{(l+1)} = \frac{1}{|x_i^{(l)}| + \epsilon}. \tag{9}$$
4. Calculate the SIC upper bound of $\mathbf{A}(\mathbf{W}^{(l+1)})^{-1}$ using Algorithm 3. Denote them as $\{\hat{r}_i^{(l+1)}\}_{i=1}^n$.
5. Clip the weight: for each $i = 1, \dots, n$,
$$w_i^{(l+1)} \leftarrow \begin{cases} \hat{r}_i^{(l+1)} w_i^{(l+1)}/\alpha, & \text{if } \hat{r}_i^{(l+1)} < \alpha, \\ w_i^{(l+1)}, & \text{otherwise.} \end{cases} \tag{10}$$
6. Terminate on convergence or when l reaches the maximum number of iterations l_{max}. Otherwise, $l \leftarrow l + 1$ and go to step 2.

For the choice of α in Step 5, empirically, for larger values of m and n, a larger value α should be set. An adaptive strategy is provided in Section 4.3 and tested in Sections 4.3, 4.4, 4.5, and 4.6 for automatically selecting α when it cannot be predetermined. Kindly note that the modified algorithm does not change the stopping criteria of the original iterative reweighting scheme. Our SIP solution provides a better initial point for the iterative reweighting scheme and helps to prevent the original reweighting process from converging to a wrong (local) optimum.

To solve the optimization problem (8), one can first solve
$$\mathbf{z}^{(l)} = \arg\min_{\mathbf{z}} \|\mathbf{z}\|_1, \quad \text{s.t. } \mathbf{y} = \mathbf{A}(\mathbf{W}^{(l)})^{-1}\mathbf{z}, \tag{11}$$
which can be fed directly into the existing l_1 minimization solver; then $\mathbf{x}^{(l)}$ can be obtained by $\mathbf{x}^{(l)} = (\mathbf{W}^{(l)})^{-1}\mathbf{z}^{(l)}$. The l_1 minimization

solver we used in this section comes from the software package l_1-*MAGIC* (available at https://candes.su.domains/software/l1magic/). A cycle from steps 1 to 6 is referred to as a trial.

The above six steps provide a framework that combines weight-clipping and iterative reweighting. Variations can be made in Steps 2–4 for specific situations. For example, when noise is present in the measurements, one can solve an unconstrained l_1 minimization instead of the constrained l_1 minimization (8) in Step 2. This paper discusses five individual cases in Sections 4.3–4.7. These are: noise-free measurements; noisy measurements; sparse error in the measurements; non-sparse signals but with sparse gradients; and non-sparse signals in the time domain but sparse in the frequency domain.

4.1. Sparse and compressible signal recovery

An experiment was conducted to investigate the ability of the proposed clipping-reweighted l_1 minimization to recover sparse signals. For each trial, we sampled a signal $\mathbf{x}_o$ of length $n = 256$ from one of the three types of distributions: (1) a k-sparse[2] Gaussian signal with i.i.d. standard Gaussian coefficients randomly placed on k positions, (2) a k-sparse Bernoulli signal with i.i.d. symmetric Bernoulli ± 1 coefficients, or (3) a compressible signal that was the sum of a k-sparse Gaussian signal with $k = 45$ and a randomly permuted sequence $\{i^{-1/p}\}_{i=1}^{n}$ for a fixed $p = 0.7$. Before summation, the sequence was randomly sign-flipped and normalized so that the maximum absolute coefficient was 1. From the signal, we collected $m = 100$ measurements using $m \times n$ matrix $\mathbf{A}$ with i.i.d. standard Gaussian entries. We set the maximum number of iterations $l_{\max} = 4$ (more iterations hardly improved the result). Parameters ϵ and α were fixed during the reweighting iterations. We conducted 500 trials for each combination of α, ϵ, and k.

The empirical probabilities of perfect recovery of reweighted l_1 minimization with fixed $\epsilon = 0.1$ (labeled as RW. L1 in the figure) are compared with the clipping-reweighted l_1 minimization with fixed $\epsilon = 0.1$ and various α values (labeled as Clip RW. L1) in Fig. 5(a). A perfect recovery is declared when the maximum absolute difference between $\mathbf{x}_o$ and its estimate $\mathbf{x}_{\text{est}}$ is less than $1\text{e}-3$ (i.e., $\|\mathbf{x}_o - \mathbf{x}_{\text{est}}\|_\infty < 1\text{e}-3$). We see a clear improvement over the original reweighted l_1 minimization. With the appropriate value of α, the oversampling factor m/k (the smaller, the better) required to achieve perfect recovery in all 500 trials decreased from approximately $100/33 \approx 3$ to approximately $100/38 \approx 2.6$; and when $k = 45$, the probability of perfect recovery increased from 0.45 to 0.8. These improvements by weight-clipping were quite robust for a wide range of α values. Fig. 5(b) illustrates the comparison among the performance of unweighted l_1 minimization (labeled as Unweight. L1 in the figure); reweighted l_1 minimization with various ϵ; and clipping-reweighted l_1 minimization with a fixed $\alpha = 30$. As seen in the figure, the clipping-reweighted l_1 minimization with a suboptimal choice of $\epsilon = 0.01$ still possessed a comparable performance to the original reweighted l_1 minimization with an optimal $\epsilon = 0.1$.

The performance for Bernoulli signal recovery is shown in Fig. 5(c). The limited decay of Bernoulli coefficients results in less variation in weights, which in turn reduces the impact of weight-clipping. Nonetheless, our results demonstrate that weight-clipping can effectively improve the reweighted l_1 minimization for better recovery of sparse signals with Bernoulli coefficients, albeit to a lesser degree than for other signal types.

The result for compressible signals is shown in Fig. 5(d). The performance metric is l_2 reconstruction error ratio (the smaller, the better), which is defined as $\|\mathbf{x}_o - \mathbf{x}_{\text{crw}}^{(4)}\|_2 / \|\mathbf{x}_o - \mathbf{x}_{\text{rw}}^{(4)}\|_2$, where $\mathbf{x}_{\text{rw}}^{(4)}$ and $\mathbf{x}_{\text{crw}}^{(4)}$ are the estimate of $\mathbf{x}_o$ given by the original and clipping reweighted l_1 minimization with four iterations, respectively. The average l_2 reconstruction error ratio in this experiment was 0.59, indicating that the average reconstruction error was reduced by 41% through weight-clipping.

[2] A signal is said to be k-sparse if it has k nonzero coefficients.

Fig. 5(e) illustrates an example of signal reconstruction with $n = 50$, $m = 20$, and $k = 8$, demonstrating how weight-clipping facilitates successful signal recovery. The reconstruction achieved by clipping-reweighted l_1 minimization (in red) was perfect, while the reconstruction by the original reweighted l_1 minimization (in blue) failed. The failure of the latter can be attributed to the weight on the 39th signal coefficient being destructive weight (as suggested by $\hat{r}_{39}^{(4)} = 1.05$, which is too close to 1), causing the 39th signal coefficient to be incorrectly estimated as zero. In contrast, the weight clipped according to the SIP theory was more appropriate (as suggested by $\hat{r}_{39}^{(4)} = 9.40$), resulting in a successful signal recovery.

4.2. Recovering non-stationary signals with changing variances

In this experiment, we investigated the recovery performance of non-stationary signals with changing variances. For each trial, we generated a signal $\mathbf{x}_o$ of length $n = 256$ containing k nonzero entries. The positions of these nonzero entries were randomly chosen, and their values were drawn from an i.i.d Gaussian distribution with a variance of σ^2. We collected $m = 100$ measurements from the signal using $m \times n$ matrix $\mathbf{A}$ with i.i.d. standard Gaussian entries. The maximum number of reweighting iterations was set to four ($l_{\max} = 4$). We conducted 500 trials for each of the three selected values of σ^2: 0.1, 1, and 10. We implemented the original reweighted l_1 minimization with different ϵ values while fixing $\epsilon = 0.1$ for the proposed clipping-reweighted l_1 minimization and setting $\alpha = 30$. The results are displayed in Fig. 6. We can observe that for the original reweighted l_1 minimization, the optimal choice of ϵ varied with different σ^2 values. For instance, $\epsilon = 0.1$ was optimal when $\sigma^2 = 0.1$, while $\epsilon = 1$ was optimal when $\sigma^2 = 10$. However, with weight-clipping, $\epsilon = 0.1$ consistently achieved recovery performance comparable to, or even better than, the best benchmarks with finely tuned ϵ values. This experiment demonstrates that weight-clipping improves the robustness of the iterative reweighting scheme, enabling effective non-stationary signal recovery with a fixed parameter setting.

4.3. Adaptive choice of α

The second numerical experiment was conducted to investigate the situation where the value of α cannot be predetermined but must be estimated during the reconstruction process. Our approach is to set α in proportion to the median of $\{\hat{r}_i\}_{i=1}^{n}$ as this can increase the value of a small r_i after weight-clipping, thus avoiding false zero estimation due to SIP. The parameter α of (10) can be calculated by

$$\alpha = 2 \cdot \text{med}(\{\hat{r}_i^{(l)}\}_{i=1}^{n}). \tag{12}$$

We repeated the experiment whose result was shown in Fig. 5(b) with the adaptively selected α. The result is shown in Fig. 7. The performance of the clipping-reweighted l_1 minimization using adaptively selected α was comparable to that of using carefully tuned α. The same conclusion can be drawn when $n = 10, 50,$ and 100. The use of 2 as the factor in (12) was an arbitrary choice from test runs and it worked well.

4.4. Recovery from noisy measurements

In Sections 4.1 and 4.2, we report that the proposed weight-clipping gives the benefits of recovering sparse signals in the noise-free situation using equality-constrained l_1 minimization. In this section, we report an experiment conducted to examine the benefits of weight-clipping for recovering sparse signals in the presence of noise. The aim of the experiment was to recover sparse signals from noisy data $\mathbf{y} = \mathbf{A}\mathbf{x}_o + \mathbf{e}$, where $\mathbf{e}$ was an unknown deterministic or stochastic noise. This problem was formulated by a quadratic inequality constraint [4,45]:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1, \quad \text{s.t. } \|\mathbf{y} - \mathbf{A}\mathbf{x}\| \le \delta, \tag{13}$$

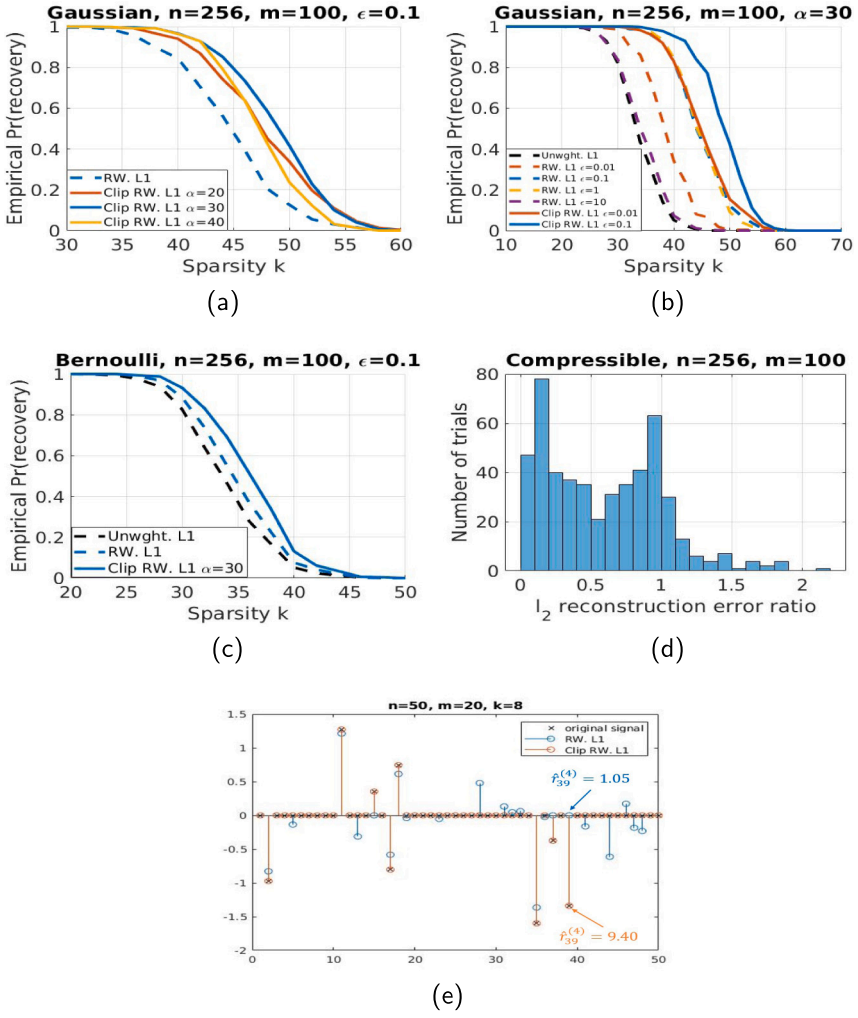

Fig. 5. Results of sparse signal recovery from noise-free measurements. Gaussian signal recovery is shown in (a) with $\epsilon = 0.1$ and various α and in (b) with $\alpha = 30$ and various ϵ. The signal to be recovered has a length of $n = 256$ with k Gaussian coefficients. The number of random measurements is $m = 100$. Sparse signal recovery with Bernoulli ± 1 coefficients is shown in (c). The statistics of the ratio (the smaller, the better) between the l_2 reconstruction error of clipping-reweighted l_1 minimization and the l_2 reconstruction error of original reweighted l_1 minimization for compressible signals are shown by the histogram in (d). An example of recovering a sparse signal with Gaussian coefficients demonstrating the effectiveness of weight-clipping is shown in (e).

where δ is set to ensure that the original solution $\mathbf{x}_o$ is feasible in the case of deterministic noise or feasible with high probability in the case of stochastic noise. Accordingly, we adapted Step 2 of the equality-constrained clipping-reweighted algorithm as:

$$\mathbf{x}^{(l)} = \arg\min_{\mathbf{x}} \|\mathbf{W}^{(l)}\mathbf{x}\|_1, \quad \text{s.t. } \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \le \delta. \tag{14}$$

For each trial, we sampled a k-sparse Gaussian signal of length $n = 256$ with $k = 45$. The locations of the nonzero coefficients were chosen randomly, and their values were independently sampled from a standard Gaussian distribution. We set $m = 100$ and sampled $m \times n$ measurement matrix $\mathbf{A}$ with i.i.d. standard Gaussian entries. We further normalized $\mathbf{A}$ such that each column vector had unit l_2 norm magnitude. We sampled noise $\mathbf{z}$ of length $m = 100$ from standard Gaussian distribution, and subsequently, multiplied it by the factor $\sigma = \frac{1}{\sqrt{m}}\beta\|\mathbf{A}\mathbf{x}_o\|_2$. We set $\delta^2 = \sigma^2(m + 2\sqrt{2m})$ as a likely upper bound of $\|\mathbf{z}\|_2^2$. Specifically, each entry of $\mathbf{z}$ followed a Gaussian distribution ($z_i/\sigma \sim N(0,1)$). The mean and variance of z_i^2/σ^2 were 1 and 2, respectively. By the central limit theorem, when m was not too small, $\|\mathbf{z}\|_2^2/\sigma^2$ approximately followed a Gaussian distribution with mean m and variance $2m$. Thus, the probability that $\|\mathbf{z}\|_2^2$ exceeded δ^2 was the probability that a Gaussian variable exceeded its mean by at least two standard deviations, which was about 2.5%.

We set $\epsilon = 0.1$ and chose α according to our adaptive strategy (12). The number of reweighting iterations was set to four. We still used the l_2 reconstruction error ratio as a performance metric to highlight the improvement afforded by the clipping-reweighted l_1 minimization when compared to the original reweighted l_1 minimization.

We carried out 500 trials for each of the two values of β. The results are presented in Fig. 8. When the noise level was small (e.g., $\beta = 0.05$), the clipping-reweighted l_1 minimization outperformed the original reweighted l_1 minimization, with an average l_2 reconstruction error ratio of 0.81. The superiority persisted until $\beta = 0.2$, where the average l_2 reconstruction error ratio was 0.99. In the experiment of $\beta = 0.2$, some nonzero coefficients that were incorrectly estimated as zero remained estimated as zero even after their weights were clipped. Therefore, we conclude that the lack of effectiveness of the proposed weight-clipping when $\beta \ge 0.2$ is likely due to the weights not being clipped aggressively enough. Overall, weight-clipping facilitates sparse signal recovery from measurements with a certain degree of noise.

4.5. Sparse and non-stationary error correction

Suppose a transmitter sends a block of arbitrary signal $\mathbf{x}_o \in \mathbb{R}^n$ (not necessarily sparse) to a remote receiver. The problem is that an unknown error occurs during transmission, so what is received is a

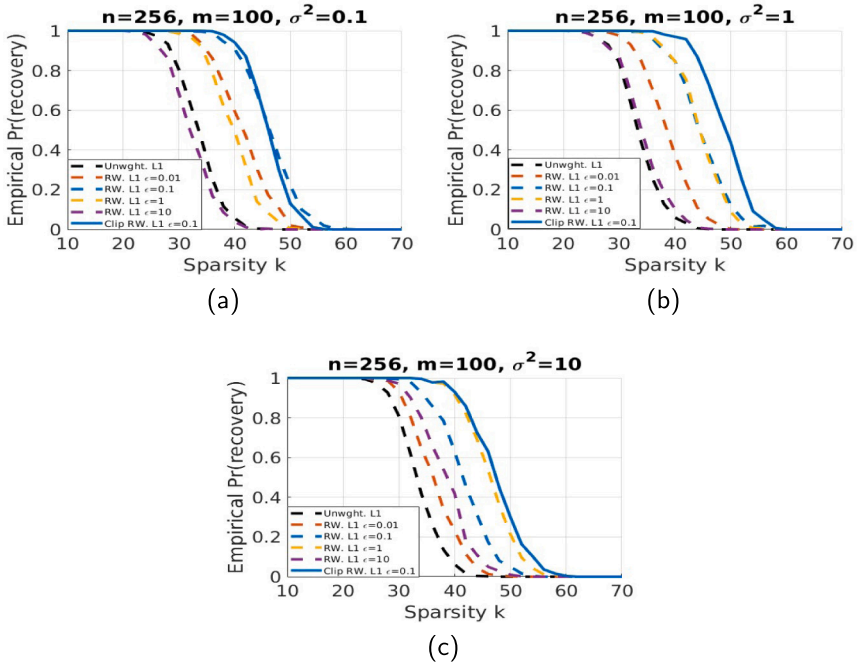

Fig. 6. Results of sparse signal recovery from noise-free measurement. The length of measurement is $m = 100$. The desired signal $\mathbf{x}_o$ has a length of $n = 256$ and k nonzero entries following i.i.d. Gaussian distribution with a variance of σ^2. (a) $\sigma^2 = 0.1$. (b) $\sigma^2 = 1$. (c) $\sigma^2 = 10$.

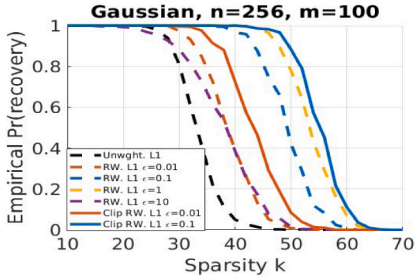

Fig. 7. Sparse signal recovery from noise-free measurements using adaptively selected α (see (12)). The black dash curve represents the unweighted l_1 minimization. The colorful dash curves represent the original reweighted l_1 minimization with various ϵ values. The solid curves represent the proposed clipping-reweighted l_1 minimization with adaptive α.

block of the corrupted signal. If the error occurs only intermittently, it is proposed in [46] to send over-complete linear measurements $\mathbf{A}\mathbf{x}_o$ instead of $\mathbf{x}_o$, where $\mathbf{A} \in \mathbb{R}^{m\times n}$ with $m > n$. The desired signal $\mathbf{x}_o$ can be recovered by solving an unconstrained l_1 minimization formulated as below

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_1, \tag{15}$$

where $\mathbf{y} = \mathbf{A}\mathbf{x}_o + \mathbf{e}_o$ is the received corrupted signal, and $\mathbf{e}_o$ is an arbitrary and unknown error vector. The theorem posits that if the proportion of the corrupted entries is within a certain range, or in other words, if $\mathbf{e}_o$ is sufficiently sparse, the solution to (15) exactly recovers $\mathbf{x}_o$. Candes et al. [31] empirically demonstrate that the recovery ability of reweighted l_1 minimization surpasses that of unweighted l_1 minimization. However, the performance of their reweighted l_1 minimization is sensitive to the characteristics of the error, as shown in this experiment. We prove that weight-clipping can enhance the reweighted l_1 minimization, helping to maintain or even improve recovery performance in the presence of non-stationary errors.

Different from the previous cases, in this task, we pursue the sparsity of $\mathbf{e} = \mathbf{y} - \mathbf{A}\mathbf{x}$ instead of $\mathbf{x}$. Therefore, it is no longer $\mathbf{A}$ but $\mathbf{A}^\dagger = (\mathbf{A}^\mathsf{T}\mathbf{A})^{-1}\mathbf{A}^\mathsf{T}$, the Moore–Penrose inverse matrix of $\mathbf{A}$, plays the role of the sensing matrix. Accordingly, we adapted Steps 2–4 of the clipping-reweighted algorithm for this error-correction problem:

2. Solve the weighted l_1 minimization problem
$$\mathbf{x}^{(l)} = \arg\min_{\mathbf{x}} \|\mathbf{W}^{(l)}(\mathbf{y} - \mathbf{A}\mathbf{x})\|_1. \tag{16}$$

3. Set $\mathbf{e}^{(l)} = \mathbf{y} - \mathbf{A}\mathbf{x}^{(l)}$, and update the weight: for each $i = 1, \ldots, n$,
$$w_i^{(l+1)} = \frac{1}{|e_i^{(l)}| + \epsilon}. \tag{17}$$

4. Calculate the upper bound of SIC of $\mathbf{A}^\dagger(\mathbf{W}^{(l+1)})^{-1}$ using Algorithm 3. Denote them as $\{\hat{r}_i^{(l+1)}\}_{i=1}^n$.

For each trial, we sampled a signal $\mathbf{x}_o$ with $n = 128$ i.i.d. standard Gaussian entries. We set $m = 4n = 512$, and sampled an $m \times n$ matrix $\mathbf{A}$ with i.i.d. standard Gaussian entries, producing the up-sampled signal $\mathbf{A}\mathbf{x}_o$ for transmission. To simulate the intermittent errors during the transmission process, we randomly selected k entries of $\mathbf{A}\mathbf{x}_o$ and replaced each entry with (1) its sign flip, simulating random changes in signal polarity; (2) its product with a randomly generated factor drawn from a standard Gaussian distribution, simulating multiplicative interference; (3) its sum with a randomly generated number from a standard Gaussian distribution, simulating additive interference; or (4) an i.i.d. standard Gaussian sample, simulating signal loss during transmission (replacing signal samples with Gaussian noise samples). We ran 500 trials for each type of error. We implemented the original reweighted l_1 minimization with various values of ϵ, while for the clipping-reweighted l_1 minimization, we fixed ϵ at 0.1, with α adaptively determined by (12). The probability of perfect recovery (declared when $\|\mathbf{x}_o - \mathbf{x}_{\text{est}}\|_\infty < 1\text{e}-3$) after four reweighted iterations is shown in Fig. 9. We observed that for the original reweighted l_1 minimization, the optimal value of ϵ varied with the error type. For example, the parameter $\epsilon = 0.1$ was optimal for the third type of error but performed poorly with the other error types. In comparison, the clipping-reweighted l_1 minimization, with a constant $\epsilon = 0.1$, either matched or exceeded the best benchmarks across all four error

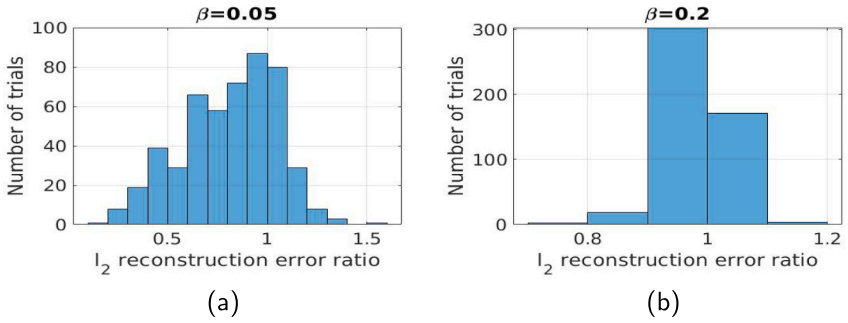

Fig. 8. Results of sparse signal recovery from noisy measurements. Gaussian signal recovery is shown in (a) with $\beta = 0.05$ and (b) with $\beta = 0.2$ from $m = 100$ random measurements of length $n = 256$ signals with $k = 45$ nonzero Gaussian coefficients. Histogram shows the statistic of the ratio of l_2 reconstruction error between the proposed clipping-reweighted l_1 minimization and the original reweighted l_1 minimization.

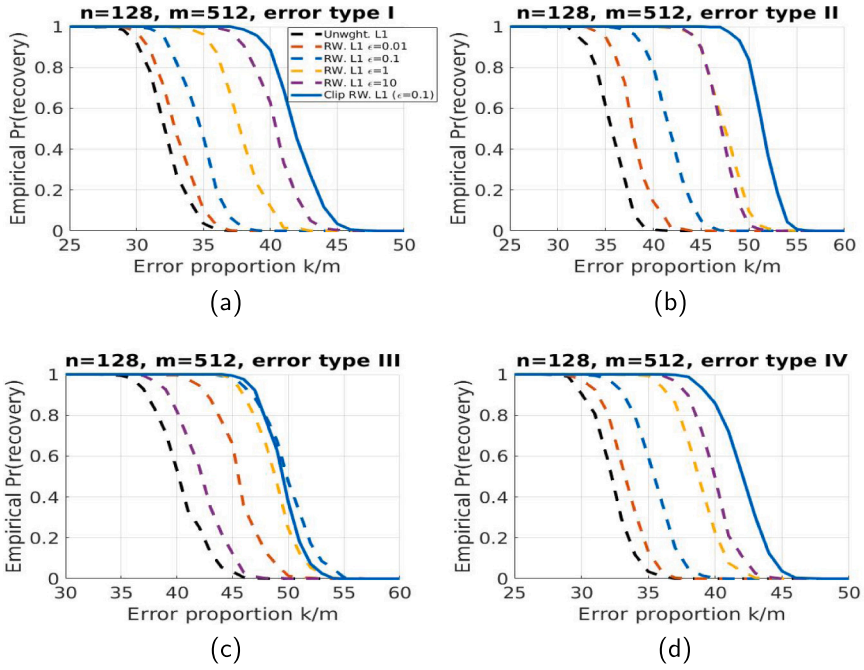

Fig. 9. Results of signal recovery from measurements with intermittent error. The desired signal $\mathbf{x}_o$ has length $n = 128$, and the corrupted measurement vector $\mathbf{y}$ has length $m = 4n = 512$ with k corrupted entries. The empirical probability of successful recovery depends on the error proportion k/m. The black dashed curve represents unweighted l_1 minimization, the colorful dashed curves represent original reweighted l_1 minimization with various ϵ values, and the solid curve represents clipping-reweighted l_1 minimization with $\epsilon = 0.1$ and adaptively chosen α. Subfigures (a), (b), (c), and (d) share the same legend and illustrate the performance of signal recovery subject to different error types.

types. These results suggest that our proposed algorithm offers a more reliable solution in scenarios where errors vary dynamically and timely adjustment of ϵ is not feasible. In conclusion, our proposed algorithm provides a more robust approach for sparse error correction.

4.6. Total variation minimization for fast image acquisition

This section reports an experiment conducted to investigate the effectiveness of applying weight-clipping to recover graph signals with sparse variation. The total variation (TV) minimization that minimizes TV norm is developed to reduce the sampling requirement for MRI to achieve faster image acquisition [6]. Weighted TV minimization by minimizing a weighted TV norm can increase the sharpness of the reconstructed image. However, destructive weights are assigned to some pixels, leading to blurred regions. Our proposed weight-clipping can correct destructive weights identified by our SIP theory, thus resulting in clearer images.

The TV norm of a two-dimensional array $(x_{i,j})$, $1 \le i, j \le n$, is defined as

$$\|(x_{i,j})\|_{\mathrm{TV}} := \sum_{1\le i,j\le n} \|(Dx)_{i,j}\|_2, \tag{18}$$

where $(Dx)_{i,j}$ is a two-dimensional array containing discrete gradients along vertical and horizontal directions. In detail, $(Dx)_{i,j} = (x_{i+1,j} - x_{i,j}, x_{i,j+1} - x_{i,j})$ when $1 \le i, j \le n-1$ while at the boundary $(Dx)_{n,j} = (0, x_{n,j+1} - x_{n,j})$ when $1 \le j \le n-1$, $(Dx)_{i,n} = (x_{i+1,n} - x_{i,n}, 0)$ when $1 \le i \le n-1$, and $(Dx)_{n,n} = (0, 0)$. For the convenience of subsequent mathematical processing, we rewrite the TV norm into a matrix form

$$\|\mathbf{x}\|_{\mathrm{TV}} = \sum_{1\le k\le n^2} \sqrt{(\mathbf{D_v x})_k^2 + (\mathbf{D_h x})_k^2}, \tag{19}$$

where $\mathbf{x} \in \mathbb{R}^{n^2}$ represents an image of $n \times n$ pixels stretched along one-dimension, $\mathbf{D_v}$ and $\mathbf{D_h}$ are $n^2 \times n^2$ vertical and horizontal forward differencing matrices, respectively, and $(\cdot)_k$ is the kth entry of a vector. It has been proved by Candes, Romber, and Tao [3] that if an image has sufficiently sparse gradients (i.e., $(Dx)_{i,j}$ is nonzero for only a small number of indices i, j), equality-constrained TV minimization can recover the image using incomplete measurements. The equality-constrained TV minimization is formulated as

$$\min_{\mathbf{x}} \|\mathbf{x}\|_{\mathrm{TV}}, \quad \text{s.t. } \mathbf{y} = \mathbf{A}\mathbf{x}. \tag{20}$$

A weighted TV norm, which is more general, is defined as

$$\|\mathbf{x}\|_{\mathbf{w},\mathrm{TV}} = \sum_{1\le k\le n^2} w_k \sqrt{(\mathbf{D_v x})_k^2 + (\mathbf{D_h x})_k^2}. \tag{21}$$

In this experiment, the algorithm to simulate the clipping-reweighted TV minimization is given in the following steps:

1. Set the iteration count l to zero and $w_k^{(0)} = 1$, $k = 1, \dots, n^2$.
2. Solve the weighted TV minimization problem
$$\mathbf{x}^{(l)} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_{\mathbf{w}^{(l)},\mathrm{TV}}, \quad \text{s.t. } \mathbf{y} = \mathbf{A}\mathbf{x}. \tag{22}$$
3. Update the weight: for each $k = 1, \dots, n^2$,
$$w_k^{(l+1)} = \frac{1}{\sqrt{(\mathbf{D_v x})_k^2 + (\mathbf{D_h x})_k^2 + \epsilon}}. \tag{23}$$
4. Calculate the upper bounds of SIC of $\mathbf{A}(\mathbf{W}^{(l+1)}\mathbf{D_v} + \gamma\mathbf{I}_{n^2})^{-1}$ and $\mathbf{A}(\mathbf{W}^{(l+1)}\mathbf{D_h} + \gamma\mathbf{I}_{n^2})^{-1}$ using Algorithm 3. Denote them as $\{\hat{r}_{v,k}^{(l+1)}\}_{k=1}^{n^2}$ and $\{\hat{r}_{h,k}^{(l+1)}\}_{k=1}^{n^2}$, respectively. Define $\hat{r}_k^{(l+1)} := \min(\hat{r}_{v,k}^{(l+1)}, \hat{r}_{h,k}^{(l+1)})$, $\forall k$.
5. Clip the weight: for each $k = 1, \dots, n^2$,
$$w_k^{(l+1)} \leftarrow \begin{cases} \hat{r}_k^{(l+1)} w_k^{(l+1)}/\alpha, & \text{if } \hat{r}_k^{(l+1)} < \alpha, \\ w_k^{(l+1)}, & \text{otherwise.} \end{cases} \tag{24}$$
6. Terminate on convergence or when l reaches the maximum number of iterations l_{max}. Otherwise, $l \leftarrow l + 1$ and go to step 2.

In Step 4, $\gamma\mathbf{I}_{n^2}$ is used to prevent singularity, where $\mathbf{I}_{n^2}$ is an $n^2 \times n^2$ identity matrix, and γ is set to $1\mathrm{e}-3$.

We conducted a reconstruction test on the Shepp–Logan phantom image [47] to evaluate the above-proposed 6-step clipping-reweighted algorithm. Results are shown in Fig. 10. The test image used (as shown in Fig. 10(a)) consisted of 100×100 pixels, 831 of which had nonzero gradients. We measured the image by sampling its discrete Fourier transform (DFT) coefficients along 10 pseudo-radial lines. The sampling pattern is illustrated in Fig. 10(b); it resulted in a total of 961 real-value measurements. Fig. 10(c) highlights the reconstruction by minimizing energy (also called back-projection) which solves

$$\min_{\mathbf{x}} \|\mathbf{x}\|_2, \quad \text{s.t. } \mathbf{y} = \mathbf{A}\mathbf{x}. \tag{25}$$

Fig. 10(d) displays the result of unweighted TV minimization, and Fig. 10(e) shows the result of reweighted TV minimization with $\epsilon = 0.1$ after four iterations. We experimented with different values of ϵ and found that 0.1 was the optimal choice. Fig. 10(f) presents the reconstructed image of the proposed clipping-reweighted TV minimization after four iterations with $\epsilon = 0.1$, and the parameter α was computed according to the adaptive strategy (12). Inspections of Fig. 10 indicate that the reconstructed image obtained from the clipping-reweighted TV minimization contained more details and less blur on the edges. Our result also outperformed the benchmarks in terms of peak signal-to-noise ratio (PSNR), a standard objective evaluation metric for assessing image quality, where a higher PSNR indicates better image quality. The proposed algorithm is robust to variations in the parameter ϵ. Specifically, when the clipping-reweighted TV minimization was performed with a suboptimal $\epsilon = 0.01$, the PSNR of the reconstructed image reached 24.11 dB. This was higher than what PSNR achieved by the original reweighted TV minimization using ϵ values of 0.01, 0.1, and 1, which recorded PSNRs of 20.89 dB, 22.92 dB, and 21.85 dB, respectively. These findings suggest that weight-clipping can improve the performance of weighted TV minimization, leading to more accurate and robust fast MRI techniques.

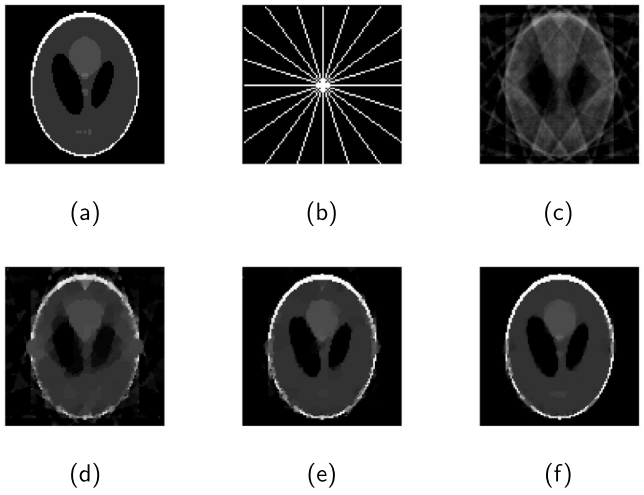

Fig. 10. Results of recovering an image by TV minimization: (a) Original 100×100 Shepp–Logan Phantom image [47]. (b) 2D Fourier-domain sampling pattern. We sampled the white regions' DFT coefficients to reconstruct the image. (c) Reconstruction by back-projection, PSNR = 16.21 dB. (d) Minimum TV reconstruction, PSNR = 19.36 dB. (e) Reweighted TV reconstruction, PSNR = 22.92 dB. (f) Clipping-reweighted TV reconstruction, PSNR = 25.78 dB.

4.7. l_1-Analysis minimization and Sub-Nyquist sampling

One of the well-known applications of l_1-analysis minimization is sub-Nyquist sampling, which aims to sample a signal at a lower rate than the Nyquist rate (i.e., twice the highest frequency of the analog signal). The traditional analog-to-digital converters (ADCs) uniformly sample an analog signal at or above the Nyquist rate. Analog-to-information conversion (AIC) based on CS theory is an alternative sampling technique. An AIC system consists of multiple random modulation pre-integration (RMPI) units to rapidly change the polarity of the analog signal, which is then integrated and quantized. These RMPIs work in parallel to generate a highly compressed information vector, which can be used to reconstruct the discretized signal. This allows signal sampling at rates much lower than the Nyquist rate. More details can be found in [2].

This section reports the application of the proposed weight-clipping to sub-Nyquist sampling. Weight-clipping was applied in a commonly used sparse model, where the signal was assumed to be not sparse but had a sparse representation. Under this assumption, the signal could be recovered by the l_1-analysis minimization [48–50], which has the form

$$\min_{\mathbf{x}} \|\mathbf{\Psi}^*\mathbf{x}\|_1, \quad \text{s.t. } \mathbf{y} = \mathbf{A}\mathbf{x}. \tag{26}$$

That is, we directly search for a signal $\mathbf{x}$ whose representation $\mathbf{z} = \mathbf{\Psi}^*\mathbf{x}$ is sparse, where $\mathbf{\Psi}$ is a dictionary composed of common bases of the objects of interest, and $\mathbf{\Psi}^*$ is the adjoint operator of $\mathbf{\Psi}$.

The last experiment applied weight-clipping to the iterative reweighting scheme. The proposed clipping-reweighted l_1-analysis minimization algorithm used is summarized as follows:

1. Set the iteration count l to zero and $w_j^{(0)} = 1$, $\forall j \in \mathbb{J}$ ($\mathbb{J}$ contains the column indexes of the dictionary $\mathbf{\Psi}$).
2. Solve the weighted l_1-analysis minimization problem
$$\mathbf{x}^{(l)} = \arg\min_{\mathbf{x}} \|\mathbf{W}^{(l)}\mathbf{\Psi}^*\mathbf{x}\|_1, \quad \text{s.t. } \mathbf{y} = \mathbf{A}\mathbf{x}. \tag{27}$$
3. Put $\mathbf{z}^{(l)} = \mathbf{\Psi}^*\mathbf{x}$, and update the weight: $\forall j \in \mathbb{J}$,
$$w_j^{(l+1)} = \frac{1}{|z_j^{(l)}| + \epsilon}. \tag{28}$$

Table 1
Performance (measured by SNR) of different algorithms in sub-Nyquist sampling tasks, where m is the number of RMPI units in AIC. Bold indicates statistical significance (paired-sample t-test with a significance level of 0.05).

algo \ m	30	40	50
Unweight. L1	4.35 ± 5.84 dB	4.84 ± 6.84 dB	6.91 ± 6.91 dB
RW. L1	12.28 ± 7.00 dB	14.38 ± 5.68 dB	17.47 ± 5.88 dB
Clip RW. L1	$\mathbf{13.41 \pm 5.91}$ dB	$\mathbf{15.72 \pm 5.06}$ dB	$\mathbf{18.39 \pm 5.55}$ dB

4. Calculate the SIC upper bound of $\mathbf{A}\boldsymbol{\Psi}(\mathbf{W}^{(l+1)})^{-1}$ using Algorithm 3. Denote them as $\{\hat{r}_j^{(l+1)}\}_{j\in\mathbb{J}}$.
5. Clip the weight: $\forall j \in \mathbb{J}$,
$$w_j^{(l+1)} \leftarrow \begin{cases} \hat{r}_j^{(l+1)} w_j^{(l+1)}/\alpha, & \text{if } \hat{r}_j^{(l+1)} < \alpha, \\ w_j^{(l+1)}, & \text{otherwise.} \end{cases} \tag{29}$$
6. Terminate on convergence or l reaching the maximum number of iteration l_{max}. Otherwise, $l \leftarrow l+1$ and go to step 2.

The last experiment was an AIC simulation conducted to sample amplitude modulation (AM) signals. For each trial, an AM signal was generated:

$$x_o(t) = [1 + 0.5\cos(2\pi f_m t)]\sin(2\pi f_c t), \tag{30}$$

where the carrier signal was

$$c(t) = \sin(2\pi f_c t), \tag{31}$$

and the message signal was a cosine wave

$$m(t) = 0.5\cos(2\pi f_m t). \tag{32}$$

The simulation code randomly chose the carrier frequency f_c between 1 GHz and 3.33 GHz, following a uniform distribution, and the message signal frequency f_m between 10 MHz and 50 MHz. To simulate an AIC system that consisted of m RMPI units and quantized the integration after every n addition, we sampled an $m \times n$ matrix $\mathbf{A}$ with i.i.d. Bernoulli ± 1 entries and used $\mathbf{A}$ to obtain m measurements from the AM signal. We set $n = 256$ and the polarity change rate of RMPI was set to 10 GHz. Successful reconstruction implied that a signal with a maximum frequency of 5 GHz theoretically could be sampled at a rate of approximately $10^{10}/256 \approx 40$ MHz. This greatly reduced the sub-Nyquist sampling rate, and as such could reduce cost and save resources. In this simulation, $\boldsymbol{\Psi}^*$ was defined as a set of STFT using Gaussian windows with lengths of 32, 64, 128, and 256. This led to the Gabor dictionary with approximately 17 times overcomplete (i.e., $|\mathbb{J}|/n - 1 \approx 17$). We set $l_{\max} = 4$, $\epsilon = 0.1$, and $\alpha = 30$. We simulated three different scales of AIC systems, namely with 30, 40, and 50 RMPI units (i.e., $m = 30, 40, 50$), and repeated the tests 100 times for each scale. The accuracy of the reconstructed signal was measured in terms of signal-to-noise ratio (SNR), and the results are shown in Table 1. Fig. 11 depicts one of the results. Our findings indicate that after applying weight-clipping, there was an increase in both the accuracy and the stability of the reconstruction, as evidenced by an increase in the means of SNRs and a decrease in the variance.

5. Conclusion

In this paper, we have discovered a novel property of weighted l_1 minimization and introduced a new necessary condition for successful signal recovery. This property allows us to identify and correct destructive weights that cause large errors in signal recovery. Extensive simulation experiments have confirmed the effectiveness of our proposed method across representative applications of compressed sensing, signal types, and l_1 minimization variants. Specifically, we have shown that our method can make weight values more appropriate when prior knowledge used in weighting schemes is inaccurate or when weights are improperly scaled. Furthermore, we demonstrate that incorporating our method into Candes's reweighting scheme enhances its ability to recover signals, improves its robustness against non-stationary signals, and increases its stability in generating high-quality weights.

CRediT authorship contribution statement

Yudong He: Writing – original draft, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Baeck Hyun Woo:** Writing – review & editing. **Fauzan Abdurrahim:** Writing – review & editing. **Richard H.Y. So:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study is partially supported by the Research Grants Council (Hong Kong) (RGC) through projects 16209823 and C5052-23G.

Appendix A. The proof of Theorem 1

We need the following two lemmas.

Lemma 1. *For any $\mathbf{a} \in \mathbb{F}^{n\times 1}$ and $\mathbf{b} \in \mathbb{F}^{n\times 1}$ where $\mathbb{F}$ is either the real number set $\mathbb{R}$ or complex number set $\mathbb{C}$, the reverse triangle inequality holds: $\|\mathbf{a}-\mathbf{b}\|_1 \geq \|\mathbf{a}\|_1 - \|\mathbf{b}\|_1$.*

Proof. We omit the proof since it is a widely accepted inequality. ■

Lemma 2. *Given a matrix $\mathbf{A} = [\mathbf{a}_1|\mathbf{a}_2|\cdots|\mathbf{a}_n] \in \mathbb{F}^{m\times n}$ where $\mathbb{F}$ is either the real number set $\mathbb{R}$ or the complex number set $\mathbb{C}$, if $\forall i \in \{1,\dots,n\}$, $\|\mathbf{a}_i\|_1 \leq 1$, then $\|\mathbf{v}\|_1 \geq \|\mathbf{A}\mathbf{v}\|_1$ for any $\mathbf{v} \in \mathbb{F}^{n\times 1}$.*

Proof. For any $\mathbf{v} \in \mathbb{F}^{n\times 1}$, we have

$$\|\mathbf{A}\mathbf{v}\|_1 = \|\sum_{i=1}^{n} v_i \mathbf{a}_i\|_1 \leq \sum_{i=1}^{n} |v_i| \|\mathbf{a}_i\|_1. \tag{A.1}$$

Because $\forall i \in \{1,\dots,n\}$, $\|\mathbf{a}_i\|_1 \leq 1$, it follows that

$$\|\mathbf{v}\|_1 = \sum_{i=1}^{n} |v_i| \geq \sum_{i=1}^{n} |v_i| \|\mathbf{a}_i\|_1. \tag{A.2}$$

Now we can conclude this proof from (A.1) and (A.2). ■

We introduce new notations here. We partition the matrix $\mathbf{A} \in \mathbb{F}^{m\times n}$ into three submatrices $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ where $\mathbf{B} \in \mathbb{F}^{m\times m}$ is assumed to be invertible, and $\mathbf{C}$ and $\mathbf{D}$ are arbitrary submatrices containing the remaining columns. Note that, following Proposition 1, permuting columns in $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$ does not affect the conclusion of this theorem. Thus, we use T_B, T_C, and T_D, three disjoint subsets of $\{1,\dots,n\}$, to denote the column indices of these three submatrices, without concern about the order of the indices. For any linear equations $\mathbf{y} = \mathbf{A}\mathbf{x}$, we partition $\mathbf{x}$ into three sub-vectors $\mathbf{x}_{T_B} \in \mathbb{F}^{m\times 1}$, $\mathbf{x}_{T_C}$, and $\mathbf{x}_{T_D}$ accordingly, so $\mathbf{y} = \mathbf{B}\mathbf{x}_{T_B} + \mathbf{C}\mathbf{x}_{T_C} + \mathbf{D}\mathbf{x}_{T_D}$. Theorem 1 can be derived immediately from the following Theorem 4.

Theorem 4. *If the l_1 norm of all columns of $\mathbf{B}^{-1}\mathbf{C}$ is smaller or equal to 1, then $\mathbf{x}^*_{T_C} = 0$, where $\mathbf{x}^*_{T_C}$ is the sub-vector of $\mathbf{x}^*$ corresponding to the subset T_C, and $\mathbf{x}^*$ is the unique optimal solution to the problem*

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1, \quad s.t.\ \mathbf{y} = \mathbf{A}\mathbf{x}. \tag{A.3}$$

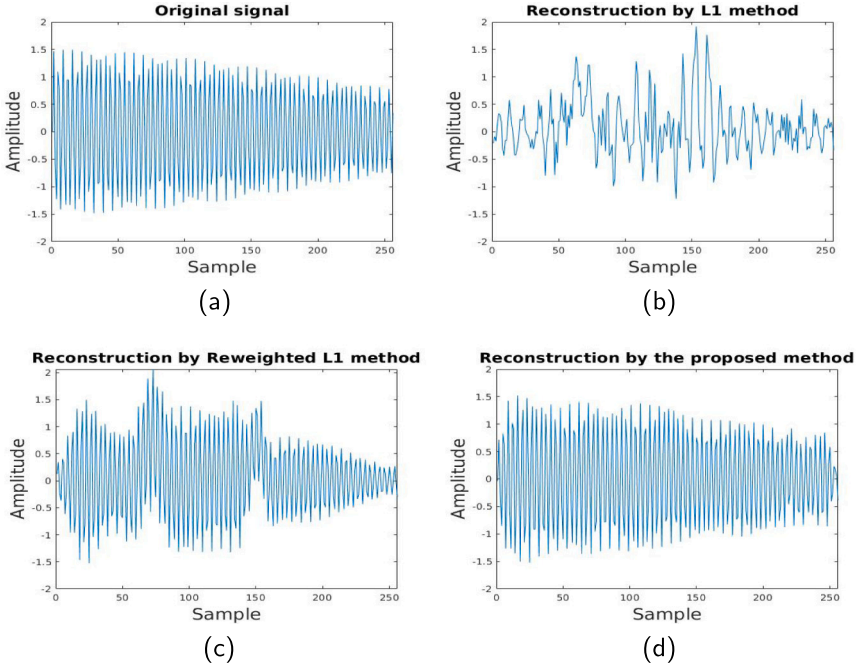

Fig. 11. Clipping-reweighted l_1-analysis minimization (our proposed method) can result in better signal reconstruction in the sub-Nyquist sampling task. (a) Original signal sampled at 10 GHz. (b) l_1-analysis reconstruction, SNR = −0.34 dB (c) Reweighted l_1-analysis reconstruction, SNR = 7.48 dB. (d) Clipping-reweighted l_1-analysis reconstruction, SNR = 16.11 dB.

Proof. Construct a solution $\mathbf{x}'$ such that $\mathbf{x}'_{T_C} = 0$, $\mathbf{x}'_{T_D} = \mathbf{x}^*_{T_D}$, where $\mathbf{x}^*_{T_D}$ is already optimal, and $\mathbf{x}'_{T_B} = \mathbf{B}^{-1}(\mathbf{y} - \mathbf{D}\mathbf{x}^*_{T_D})$ for the sake of feasibility. We shall prove that $\mathbf{x}'$ is an optimal solution. Denote any change of $\mathbf{x}'$ by $\Delta\mathbf{x}$, and the corresponding changes on the three sub-vectors by $\Delta\mathbf{x}_{T_B}$, $\Delta\mathbf{x}_{T_C}$, and $\Delta\mathbf{x}_{T_D}$. Because $\mathbf{x}'_{T_D}$ is already optimal, it is only necessary to prove that for any feasible change $\Delta\mathbf{x}$ where $\Delta\mathbf{x}_{T_D} = 0$, $\|\mathbf{x}' + \Delta\mathbf{x}\|_1 \geq \|\mathbf{x}'\|_1$.

By definition,

$$\begin{aligned}\|\mathbf{x}' + \Delta\mathbf{x}\|_1 =& \|\mathbf{x}'_{T_B} + \Delta\mathbf{x}_{T_B}\|_1 + \|\mathbf{x}'_{T_C} + \Delta\mathbf{x}_{T_C}\|_1 \\ &+ \|\mathbf{x}'_{T_D} + \Delta\mathbf{x}_{T_D}\|_1.\end{aligned} \tag{A.4}$$

Because of the constraint $\mathbf{y} = \mathbf{A}\mathbf{x}$, the following equation must to be satisfied:

$$\Delta\mathbf{x}_{T_B} = \mathbf{B}^{-1}(-\mathbf{C}\Delta\mathbf{x}_{T_C} - \mathbf{D}\Delta\mathbf{x}_{T_D}). \tag{A.5}$$

Combining (A.4) and (A.5), we have

$$\begin{aligned}\|\mathbf{x}' + \Delta\mathbf{x}\|_1 =& \|\mathbf{B}^{-1}(\mathbf{y} - \mathbf{D}\mathbf{x}^*_{T_D} - \mathbf{C}\Delta\mathbf{x}_{T_C})\|_1 \\ &+ \|\Delta\mathbf{x}_{T_C}\|_1 + \|\mathbf{x}^*_{T_D}\|_1.\end{aligned} \tag{A.6}$$

Applying Lemma 1 yields

$$\begin{aligned}\|\mathbf{B}^{-1}(\mathbf{y} - \mathbf{D}\mathbf{x}^*_{T_D} - \mathbf{C}\Delta\mathbf{x}_{T_C})\|_1 \geq& \|\mathbf{B}^{-1}(\mathbf{y} - \mathbf{D}\mathbf{x}^*_{T_D})\|_1 \\ &- \|\mathbf{B}^{-1}\mathbf{C}\Delta\mathbf{x}_{T_C}\|_1.\end{aligned} \tag{A.7}$$

By Lemma 2 and our assumed condition about $\mathbf{B}^{-1}\mathbf{C}$, for any $\Delta\mathbf{x}_{T_C}$, we have

$$\|\Delta\mathbf{x}_{T_C}\|_1 \geq \|\mathbf{B}^{-1}\mathbf{C}\Delta\mathbf{x}_{T_C}\|_1. \tag{A.8}$$

Substituting (A.7) and (A.8) into (A.6) yields

$$\|\mathbf{x}' + \Delta\mathbf{x}\|_1 \geq \|\mathbf{B}^{-1}(\mathbf{y} - \mathbf{D}\mathbf{x}^*_{T_D})\|_1 + \|\mathbf{x}^*_{T_D}\|_1, \tag{A.9}$$

which holds for any feasible $\Delta\mathbf{x}$ where $\Delta\mathbf{x}_{T_D} = 0$.

The l_1 norm of $\mathbf{x}'$ equals the right-hand side of the last inequality, hence, $\mathbf{x}'$ is an optimal solution (also unique according to our assumption). ■

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.sigpro.2024.109828.

Data availability

Data will be made available on request.

References

[1] D.L. Donoho, Compressed sensing, IEEE Trans. Inform. Theory 52 (4) (2006) 1289–1306.
[2] E.J. Candès, M.B. Wakin, An introduction to compressive sampling, IEEE Signal Process. Mag. 25 (2) (2008) 21–30.
[3] E.J. Candès, J. Romberg, T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, IEEE Trans. Inform. Theory 52 (2) (2006) 489–509.
[4] E.J. Candès, J.K. Romberg, T. Tao, Stable signal recovery from incomplete and inaccurate measurements, Comm. Pure Appl. Math. 59 (8) (2006) 1207–1223.
[5] E.J. Candes, T. Tao, Near-optimal signal recovery from random projections: Universal encoding strategies? IEEE Trans. Inf. Theory 52 (12) (2006) 5406–5425.
[6] M. Lustig, D. Donoho, J.M. Pauly, Sparse MRI: The application of compressed sensing for rapid MR imaging, Magn. Reson. Med.: Off. J. Int. Soc. Magn. Reson. Med. 58 (6) (2007) 1182–1195.
[7] G.-H. Chen, J. Tang, S. Leng, Prior image constrained compressed sensing (PICCS): a method to accurately reconstruct dynamic CT images from highly undersampled projection data sets, Med. Phys. 35 (2) (2008) 660–663.
[8] J.C. Ye, Compressed sensing MRI: a review from signal processing perspective, BMC Biomed. Eng. 1 (1) (2019) 1–17.
[9] V.M. Patel, G.R. Easley, D.M. Healy, R. Chellappa, Compressed sensing for synthetic aperture radar imaging, in: 2009 16th IEEE International Conference on Image Processing, ICIP, IEEE, 2009, pp. 2141–2144.
[10] W. Qiu, M. Martorella, J. Zhou, H. Zhao, Q. Fu, Three-dimensional inverse synthetic aperture radar imaging based on compressive sensing, IET Radar Sonar Navig. 9 (4) (2015) 411–420.
[11] J. Yang, T. Jin, C. Xiao, X. Huang, Compressed sensing radar imaging: Fundamentals, challenges, and advances, Sensors 19 (14) (2019) 3100.
[12] F.J. Herrmann, M.P. Friedlander, O. Yilmaz, Fighting the curse of dimensionality: Compressive sensing in exploration seismology, IEEE Signal Process. Mag. 29 (3) (2012) 88–100.

[13] J.B. Muir, Z. Zhan, Seismic wavefield reconstruction using a pre-conditioned wavelet–curvelet compressive sensing approach, Geophys. J. Int. 227 (1) (2021) 303–315.
[14] J. Bobin, J.-L. Starck, R. Ottensamer, Compressed sensing in astronomy, IEEE J. Sel. Top. Sign. Proces. 2 (5) (2008) 718–726.
[15] M.C. Cheung, B. De Pontieu, J. Martínez-Sykora, P. Testa, A.R. Winebarger, A. Daw, V. Hansteen, P. Antolin, T.D. Tarbell, J.-P. Wuelser, et al., Multi-component decomposition of astronomical spectra by compressed sensing, Astrophys. J. 882 (1) (2019) 13.
[16] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2) (2008) 210–227.
[17] R. Chartrand, W. Yin, Iteratively reweighted algorithms for compressive sensing, in: 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, 2008, pp. 3869–3872.
[18] N. Vaswani, W. Lu, Modified-CS: Modifying compressive sensing for problems with partially known support, IEEE Trans. Signal Process. 58 (9) (2010) 4595–4607.
[19] M.P. Friedlander, H. Mansour, R. Saab, Ö. Yilmaz, Recovering compressively sampled signals using partial support information, IEEE Trans. Inform. Theory 58 (2) (2011) 1122–1134.
[20] H. Mansour, Ö. Yilmaz, Weighted-l1 minimization with multiple weighting sets, in: Wavelets and Sparsity XIV, Vol. 8138, SPIE, 2011, pp. 52–64.
[21] D. Needell, R. Saab, T. Woolf, Weighted-minimization for sparse recovery under arbitrary prior information, Inf. Inference: J. IMA 6 (3) (2017) 284–309.
[22] W. Chen, Y. Li, G. Wu, Recovery of signals under the high order RIP condition via prior support information, Signal Process. 153 (2018) 83–94.
[23] W. Chen, Y. Li, Recovery of signals under the condition on RIC and ROC via prior support information, Appl. Comput. Harmon. Anal. 46 (2) (2019) 417–430.
[24] H. Ge, W. Chen, M.K. Ng, On recovery of sparse signals with prior support information via weighted lp-minimization, IEEE Trans. Inform. Theory 67 (11) (2021) 7579–7595.
[25] M.A. Khajehnejad, W. Xu, A.S. Avestimehr, B. Hassibi, Weighted l1 minimization for sparse recovery with prior information, in: 2009 IEEE International Symposium on Information Theory, IEEE, 2009, pp. 483–487.
[26] J. Zhang, U. Mitra, K.-W. Huang, N. Michelusi, Support recovery from noisy random measurements via weighted l_1 minimization, IEEE Trans. Signal Process. 66 (17) (2018) 4527–4540.
[27] E.J. Candès, The restricted isometry property and its implications for compressed sensing, C. R. Math. 346 (9) (2008) 589–592.
[28] H. Mansour, R. Saab, Recovery analysis for weighted l_1 minimization using the null space property, Appl. Comput. Harmon. Anal. 43 (1) (2017) 23–38.
[29] B. Bah, R. Ward, The sample complexity of weighted sparse approximation, IEEE Trans. Signal Process. 64 (12) (2016) 3145–3155.
[30] C. Herzet, C. Soussen, J. Idier, R. Gribonval, Exact recovery conditions for sparse representations with partial support information, IEEE Trans. Inf. Theory 59 (11) (2013) 7509–7524.
[31] E.J. Candes, M.B. Wakin, S.P. Boyd, Enhancing sparsity by reweighted l1 minimization, J. Fourier Anal. Appl. 14 (5) (2008) 877–905.
[32] Y. Peng, J. Suo, Q. Dai, W. Xu, Reweighted low-rank matrix recovery and its application in image restoration, IEEE Trans. Cybern. 44 (12) (2014) 2418–2430.
[33] L. Zhang, G. Wang, G.B. Giannakis, J. Chen, Compressive phase retrieval via reweighted amplitude flow, IEEE Trans. Signal Process. 66 (19) (2018) 5029–5040.
[34] L. Wang, S. Ma, Q. Han, Reweighted dual sparse regularization and convex optimization for bearing fault diagnosis, IEEE Trans. Instrum. Meas. 70 (2020) 1–9.
[35] J.-Y. Wu, L.-C. Huang, M.-H. Yang, C.-H. Liu, Sparse subspace clustering via two-step reweighted l1-minimization: Algorithm and provable neighbor recovery rates, IEEE Trans. Inform. Theory 67 (2) (2020) 1216–1263.
[36] W. Guo, Y. Lou, J. Qin, M. Yan, A novel regularization based on the error function for sparse recovery, J. Sci. Comput. 87 (1) (2021) 31.
[37] S. Arberet, P. Vandergheynst, R.E. Carrillo, J.-P. Thiran, Y. Wiaux, Sparse reverberant audio source separation via reweighted analysis, IEEE Trans. Audio Speech Lang. Process. 21 (7) (2013) 1391–1402.
[38] J. Peng, J. Hampton, A. Doostan, A weighted l_1 minimization approach for sparse polynomial chaos expansions, J. Comput. Phys. 267 (2014) 92–111.
[39] J.-Y. Wu, L.-C. Huang, M.-H. Yang, L.-H. Chang, C.-H. Liu, Enhanced noisy sparse subspace clustering via reweighted L1-minimization, in: 2018 IEEE 28th International Workshop on Machine Learning for Signal Processing, MLSP, IEEE, 2018, pp. 1–6.
[40] P. Bofill, M. Zibulevsky, Underdetermined blind source separation using sparse representations, Signal Process. 81 (11) (2001) 2353–2362.
[41] S. Winter, W. Kellermann, H. Sawada, S. Makino, MAP-based underdetermined blind source separation of convolutive mixtures by Hierarchical clustering and -norm minimization, EURASIP J. Adv. Signal Process. 2007 (2006).
[42] F. Feng, M. Kowalski, Underdetermined reverberant blind source separation: Sparse approaches for multiplicative and convolutive narrowband approximation, IEEE/ACM Trans. Audio Speech Lang. Process. 27 (2) (2019) 442–456, http://dx.doi.org/10.1109/TASLP.2018.2881925.
[43] C.H. Papadimitriou, The largest subdeterminant of a matrix, Δελτίο της Ελληνικής Μαθηματικής Εταιρίας 25 (25) (1984) 95–105.
[44] L. Khachiyan, On the complexity of approximating extremal determinants in matrices, J. Complexity 11 (1) (1995) 138–153.
[45] R. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Stat. Soc. Ser. B Stat. Methodol. 58 (1) (1996) 267–288.
[46] E. Candes, T. Tao, Decoding by linear programming, IEEE Trans. Inform. Theory 51 (12) (2005) 4203–4215.
[47] L.A. Shepp, B.F. Logan, The Fourier reconstruction of a head section, IEEE Trans. Nucl. Sci. 21 (3) (1974) 21–43.
[48] E.J. Candes, Y.C. Eldar, D. Needell, P. Randall, Compressed sensing with coherent and redundant dictionaries, Appl. Comput. Harmon. Anal. 31 (1) (2011) 59–73.
[49] M. Kabanava, H. Rauhut, Analysis l1-recovery with frames and Gaussian measurements, Acta Appl. Math. 140 (1) (2015) 173–195.
[50] M. Genzel, G. Kutyniok, M. März, l1-analysis minimization and generalized (co-) sparsity: when does recovery succeed? Appl. Comput. Harmon. Anal. 52 (2021) 82–140.