Support Recovery From Noisy Random Measurements via Weighted ℓ_1 Minimization

Jun Zhang, *Member, IEEE*, Urbashi Mitra, *Fellow, IEEE*, Kuan-Wen Huang, and Nicolò Michelusi, *Senior Member, IEEE*

***Abstract*—Herein, the sample complexity of general weighted ℓ_1 minimization in terms of support recovery from noisy underdetermined measurements is analyzed. This analysis generalizes prior work on ℓ_1 minimization by considering arbitrary weighting. The explicit relationship between the weights and the sample complexity is stated such that for random matrices with i.i.d. Gaussian entries, the weighted ℓ_1 minimization recovers the support of the underlying signal with high probability as the problem dimension increases. This result provides a measure that is predictive of relative performance of different algorithms. Motivated by the analysis, a new iterative reweighted strategy is proposed for binary signal recovery. In the binary sparsity-Promoting Reweighted ℓ_1 minimization (bPRL1) algorithm, a sequence of weighted ℓ_1 minimization problems are solved where partial support recovery is used to prune the optimization; furthermore, the weights used for the next iteration are updated by the current estimate. bPRL1 is compared to other weighted algorithms through the proposed measure and numerical results are shown to provide superior performance for a spectrum occupancy estimation problem motivated by cognitive radio.**

***Index Terms*—Weighted ℓ_1 minimization, support recovery, sample complexity, partial support recovery, cognitive radio.**

I. Introduction

CONSIDER support recovery of a high-dimensional sparse signal from a small number of noisy measurements. More specifically, suppose that $\boldsymbol{x}^* \in \mathbb{R}^n$ is an unknown k-sparse signal, i.e., n-dimensional vector having at most k non-zero elements, and $\mathbf{A} \in \mathbb{R}^{m\times n}$ is a known measurement matrix satisfying $k < m < n$. The measurement process can be modeled as

$$\boldsymbol{y} = \mathbf{A}\boldsymbol{x}^* + \boldsymbol{z} \tag{1}$$

where $\boldsymbol{z}$ is a noise vector. Our goal is to recover the support of $\boldsymbol{x}^*$; exact support recovery means that $\mathrm{supp}(\boldsymbol{x}^*) = \{i : x_i^* \neq 0\}$ and its signs are properly identified, i.e., support and sign consistency.

This is a fundamental problem in many fields, exploiting tools from compressed sensing (CS) [2], [3], machine learning [4] and statistics [5]. Given the support, one can straightforwardly recover the signal via least-squares on the received signal projected onto the support set. In several key applications, the support is physically more significant than the component values. For example, in a cognitive radio network scenario [6], further investigated in Sec. III, the secondary users are allowed to utilize the frequency bands assigned to the primary users when the bands are not in current use. To detect the spectrum opportunities, we are concerned with estimating the number of active frequency bands and their locations exactly. However, this might be a difficult task, due to the complexity and energy cost of sensing, in real time, all the frequency bands to detect available spectrum opportunities. In contrast, one may leverage the fact that the spectrum occupancy varies slowly over time, as a result of users leaving and joining the network at random times, and thus develop CS techniques to measure the spectrum and detect sparse variations in spectrum occupancy, as demonstrated in [7]. In brain-computer interfaces [8], electroencephalography (EEG) is a popular noninvasive technology to probe human brain activities with excellent spatial-temporal resolution. A relatively fine spatial resolution is required to localize the neural electrical activities from a huge number of potential locations, where neural activity is represented by non-zero potential locations [9].

There has been a great deal of work [10]–[13] on the analysis of classical CS methods, such as ℓ_1 minimization [14] and Orthogonal Matching Pursuit (OMP) [15], for exact support recovery. Among them, [10] shows that in a high-dimensional setting, the ℓ_1 minimization

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \frac{1}{2}\left\|\mathbf{A}\boldsymbol{x} - \boldsymbol{y}\right\|_2^2 + h\|\boldsymbol{x}\|_1, \tag{2}$$

can, with high probability, exactly recover the support of $\boldsymbol{x}^*$ from $m > 2k\log(n-k)$ random measurements.

Manuscript received September 12, 2017; revised March 5, 2018 and April 25, 2018; accepted April 26, 2018. Date of publication May 23, 2018; date of current version July 31, 2018. The associate editor coordinating the review of this manuscript and approving if for publication was Prof. Mark A. Davenport. This work was supported in part by the following Grants and organizations: AFOSR FA9550-12-1-0215, NSF CNS-1213128, NSF CCF-1410009, NSF CPS-1446901, ONR N00014-09-1-0700, ONR N00014-15-1-2550, and in part by the Royal Academy of Engineering, the Fulbright Foundation and the Leverhulme Trust. The work of J. Zhang was supported by NSFC under Grant 61403085 and by the science and technology program of Guangzhou under Grant 201804010293. The work of N. Michelusi was supported by NSF CNS-1642982 and by DARPA #108818. This paper was presented in part at the IEEE International Symposium on Information Theory, Barcelona, Spain, July 2016. *(Corresponding author: Jun Zhang.)*

J. Zhang is with the School of Information Engineering, Guangdong University of Technology, Guangzhou 510006, China (e-mail: jzhang@gdut.edu.cn).

U. Mitra and K. W. Huang are with the Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089 USA (e-mail: ubli@usc.edu; kuanwenh@usc.edu).

N. Michelusi is with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907 USA (e-mail: michelus@purdue.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2018.2838553

The ℓ_1 minimization assumes no prior information other than the sparsity prior on $\boldsymbol{x}^*$. However, significant performance gains can be obtained by exploiting not only standard sparsity, but also structural priors of $\boldsymbol{x}^*$, e.g., partially known support [16]–[18]. A unified method by which to incorporate both the sparsity and the structural priors is a weighted ℓ_1 minimization [7], [16], [17], [19];

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \frac{1}{2} \|\mathbf{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + h \sum_{i=1}^{n} w_i |x_i|, \tag{3}$$

where $w_i \geqslant 0$ denotes the weight for component i, which can play a key role in the incorporation of structural priors by penalizing different components of $\hat{\boldsymbol{x}}$ with nonuniform weights. There are numerous CS algorithms that employ (3), but with different weighting assumptions. A partial list includes: modified-BPDN [16], compact belief incorporation-based recovery (CBIR) [7], support driven reweighted ℓ_1 minimization (SDRL1) [20], iterative reweighted ℓ_1 minimization (IRL1) [19] and threshold iterative support detection (threshold-ISD) [17]. In this paper, we are interested in developing a performance guarantee for the weighted ℓ_1 minimization for determining exact support in the noisy setting.

A. Prior work on Weighted ℓ_1 Minimization

In the noiseless version of weighted ℓ_1 minimization, we can pursue the recovery of $\boldsymbol{x}^*$ by solving

$$\min_{\boldsymbol{x}} \sum_{i=1}^{n} w_i |x_i| \text{ subject to } \boldsymbol{y} = \mathbf{A}\boldsymbol{x}. \tag{4}$$

Several works have studied the exact recovery of $\boldsymbol{x}^*$ using prior support information to determine the weights in (4), e.g., [18], [21]–[23]. There has been a focus on the "two weights" version of (4), a special case of the weighted ℓ_1 minimization. In particular, a modified-CS method was proposed in [18], where zero weights are assigned to the known support and unit weights are assigned to the others. This work showed that when a large proportion of the support is known, the restricted-isometry-based recovery conditions of the modified-CS are weaker than those for traditional ℓ_1 minimization. A mutual-coherence-based recovery condition for the modified-CS was proposed in [21], which also relaxes the analogous ℓ_1 minimization condition ensuring the recovery of any k-sparse signal. More generally, [22] analyzed the performance of (4) over a non-uniform sparsity model where the entries of the unknown vector fall into two sets, with entries of each set having a specific probability of being non-zero. A Grassman angle approach was used to establish a recovery condition from the angle of sample complexity for (4), which focuses primarily on the case that there are only two different probabilities, and hence only two weights. In [23], given a support estimate $\widetilde{T} \subset \{1, \ldots, n\}$, one solves (4) where $w_i = \omega \in [0, 1]$ whenever $i \in \widetilde{T}$, and $w_i = 1$ otherwise. Further, sufficient and necessary conditions for exact recovery using this method are derived. Bounds on the number of Gaussian measurements for these conditions to be satisfied are determined.

There is a substantial body of work [24]–[30] examining the noisy setting and the use of weighted ℓ_1 minimization or its variants for sparse signal recovery. These works investigate the robust and stable recovery problem, *i.e.*, how accurate is $\hat{\boldsymbol{x}}$ when compared to $\boldsymbol{x}^*$ in terms of ℓ_2 norm of error. Specifically, [24] studied a weighted ℓ_1 minimization, named *innovative basis pursuit de-noising* (iBPDN), which incorporates the known support by penalizing it with zero weights, and shows that iBPDN has similar stability behavior to that of traditional ℓ_1 minimization. In [25], a sufficient condition for weighted ℓ_1 minimization for the robust and stable recovery is analyzed, where $w_i = \omega \in [0, 1]$ is applied to the support estimate and $w_i = 1$ otherwise. Relying on the accuracy and size of the support estimate, it is proved in [25] that weighted ℓ_1 minimization is stable and robust under weaker sufficient conditions than the analogous conditions for ℓ_1 minimization. Further, [26] extended the analysis of [25] to multiple sets. Recently, sample complexity analyses [28]–[30] have been carried out for the robust and stable recovery of weighted ℓ_1 minimization. For example, [28] provided an unifying theory to compute sample complexity for general regularized linear regression when the measurement matrix belongs to the Gaussian ensemble and [30] derived another unified sample complexity for weighted ℓ_1 minimization in terms of how well a given prior model for the sparsity support aligns with the true underlying support.

Our work has two major differences compared to the existing results. Firstly, we focus on the *support* recovery problem in the noisy setting, while the existing results in the noisy setting are concerned with the robust and stable recovery problem. Since support recovery is a stronger theoretical criterion than robust and stable recovery, the existing results offer no guarantee that the support of $\boldsymbol{x}^*$ can be recovered exactly in this case. Secondly, unlike most of the previous studies that restricted the weighted ℓ_1 minimization to "two weights" models with $w_i \in [0, 1]$, we study weighted ℓ_1 minimization where $w_i \in [0, +\infty), \forall 1 \leqslant i \leqslant n$, namely, the general "multiple weights" model.

The current work is an extension of [1], which considered only strictly positive weights ($w_i > 0$). Several popular algorithms allow for zero-valued weights [16]–[18], thus the analysis of $w_i \geq 0$ is relevant. While this change in assumptions may seem small, the analysis is quite different. For example, in some cases, the problem in Equation (3) cannot recover the support exactly with probability one. This fact necessitates our new result which characterizes when we can recover the support and which additional constraints are needed for the cases of $w_i \in [0, +\infty)$.

B. Contributions

This paper provides a sample complexity analysis for the general weighted ℓ_1 minimization for exact support recovery when random matrices with i.i.d. Gaussian random entries are used. The main contributions of this work are two-fold.

1) We prove that, in the noisy scenario, if there exists at least one zero weight outside of the support of $\boldsymbol{x}^*$, the solution, $\hat{\boldsymbol{x}}$ in (3) will not match the true support of $\boldsymbol{x}^*$ with probability one. This drawback can be overcome by introducing thresholding. With thresholding, we establish a set of sufficient conditions under which $\hat{\boldsymbol{x}}$ can exactly recover the support of $\boldsymbol{x}^*$ with high probability. These conditions reveal the explicit relationship between the weights and the sample complexity. For example,

in the case of $w_i > 0$, we show that if the magnitude of the non-zero entries of $\boldsymbol{x}^*$ are large enough, $\hat{\boldsymbol{x}}$ has the true support of $\boldsymbol{x}^*$ with high probability, provided the number of measurements is sufficiently large, i.e., $m > 2\eta k \log(n-k)$, where η is an explicit function of the weights. This scaling law extends the work in [10] by incorporating arbitrary weighting. Interestingly, η provides a key measure, dependent on the weights, which is predictive of performance.

2) Motivated by the analysis, an iterative weighted strategy, named the *binary sparsity-Promoting Reweighted ℓ_1 minimization* (bPRL1) algorithm[1], is proposed to recover a binary sparse signal, which utilizes partial support recovery to prune the optimization and updates the weights iteratively. bPRL1 is compared to other weighted algorithms through the proposed measure and numerical results and is shown to provide superior performance for a spectrum sensing problem motivated by cognitive radio. We observe that in some cases where other methods achieve a 50% rate of exact support recovery, bPRL1 can achieve a 87% rate of exact support recovery.

C. Organization and Notation

The remainder of this paper is organized as follows. In Section II, we detail the optimality conditions of (3) and establish sufficient conditions for exact support recovery. Then, the main results of this paper are derived. Further, motivated by the spectrum sensing problem, we propose the bPRL1 algorithm for binary signal recovery in Section III. In Section IV, some numerical simulations demonstrate the performance gains of the bPRL1 algorithm. Conclusions are provided in Section V.

We briefly introduce the notation used in this paper. Scalars are denoted by lowercase letters (e.g., m), vectors by lowercase bold italicized letters (e.g., $\boldsymbol{x}$), matrices by uppercase boldface letters (e.g., $\mathbf{A}$) and sets by uppercase calligraphic letters (e.g., $\mathcal{S}$). We denote the i-th entry of a vector $\boldsymbol{x}$ by x_i, the i-th row of a matrix $\mathbf{A}$ by $\mathbf{A}^i$ and the i-th column of a matrix $\mathbf{A}$ by $\mathbf{A}_i$. We denote by $\mathbf{A}_{\mathcal{S}}$ (or $\boldsymbol{x}_{\mathcal{S}}$) the reduced dimension matrix (or vector) constructed by the columns (or entries) of $\mathbf{A}$ (or $\boldsymbol{x}$) whose indices are in the set $\mathcal{S}$. If $\mathbf{A}$ is a diagonal matrix, $\mathbf{A}_{\mathcal{S}}$ is a diagonal matrix constructed from the diagonal elements of $\mathbf{A}$ whose indices are in the set $\mathcal{S}$. The support set of a vector $\boldsymbol{x}$ is denoted by supp($\boldsymbol{x}$), the cardinality of a set $\mathcal{S}$ by $|\mathcal{S}|$ and $[n] \triangleq \{1, \ldots, n\}$. Given $\mathcal{S} \subseteq [n]$, its complement is denoted by $\mathcal{S}^c \triangleq [n] \setminus \mathcal{S}$. The notation $\|\boldsymbol{x}\|_p$ denotes the ℓ_p norm of $\boldsymbol{x}$. The notation $\mathbf{A}^T$ denotes the transpose of $\mathbf{A}$. Throughout this paper, $\boldsymbol{x}^*$ is the original k-sparse signal that satisfies $k < m \ll n$. Without loss of generality, we assume that the elements of $\mathbf{A}$ are drawn randomly and independently from the standard Gaussian distribution, i.e., $a_{i,j} \sim \mathcal{N}(0,1)$, and the entries of $\boldsymbol{z}$, $z_i \sim \mathcal{N}(0, \sigma_z^2)$.

II. Sample Complexity Analysis for Weighted ℓ_1 Minimization

A. Support Recovery Conditions

In this section, we detail the optimality conditions of the weighted ℓ_1 minimization, then investigate the conditions under which it is possible to exactly recover the support of $\boldsymbol{x}^*$ (or the support of its reduced dimensional vector). For this section (Section II-A), we mainly consider the case that $\mathbf{A}$ and $\boldsymbol{z}$ are deterministic and fixed. In the sequel (Section II-B), we will consider statistical models on $\mathbf{A}$ and $\boldsymbol{z}$ and leverage the results in the current section which serve to provide conditions for support recovery, conditioned on $\mathbf{A}$ and $\boldsymbol{z}$.

Without loss of generality, the weighted ℓ_1 minimization can be rewritten as follows:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \frac{1}{2m} \|\mathbf{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + h\|\mathbf{W}\boldsymbol{x}\|_1, \tag{5}$$

where $\mathbf{W} \in \mathbb{R}^{n \times n}$ is a diagonal matrix, whose diagonal vector is $\boldsymbol{w}$ with $w_i \in [0, +\infty)$, $h > 0$ is a tuning parameter to achieve a tradeoff between the data fidelity and its sparsity and m is the number of the measurements. We denote the subdifferential of $\|\mathbf{W}\boldsymbol{x}\|_1$ on $\boldsymbol{x}$ as

$$\begin{aligned}\partial\|\mathbf{W}\boldsymbol{x}\|_1 &= \{\mathbf{W}\boldsymbol{u} | \boldsymbol{u}^T\mathbf{W}\boldsymbol{x} = \|\mathbf{W}\boldsymbol{x}\|_1, \ \|\boldsymbol{u}\|_\infty \leq 1\} \\ &= \{\mathbf{W}\boldsymbol{u} | u_i = \text{sign}(x_i), \text{ if } w_i x_i \neq 0 \\ &\qquad \text{and } u_i \in [-1, 1], \text{ otherwise}\},\end{aligned} \tag{6}$$

where $\mathbf{W}$ is independent on $\boldsymbol{u}$, $\text{sign}(x_i) = 1$ when $x_i > 0$ and $\text{sign}(x_i) = -1$ when $x_i < 0$. In the paper, we also denote $\text{sign}(x_i) = 0$ when $x_i = 0$. From convex analysis, we have the following Lemma.

Lemma 1: (**Global Minimum and Uniqueness**): Assume that $\mathbf{A}$, $\boldsymbol{z}$ and the weights w_i are deterministic and fixed. (a) A vector $\hat{\boldsymbol{x}} \in \mathbb{R}^n$ is a global minimum of (5) if and only if $\exists\, \mathbf{W}\hat{\boldsymbol{u}} \in \partial\|\mathbf{W}\hat{\boldsymbol{x}}\|_1$, such that

$$\frac{1}{m}\mathbf{A}^T(\mathbf{A}\hat{\boldsymbol{x}} - \boldsymbol{y}) + h\mathbf{W}\hat{\boldsymbol{u}} = \mathbf{0}. \tag{7}$$

(b) Suppose $\mathcal{V}$ is any subset of $[n] \setminus \{i : w_i = 0\}$ with $|\mathcal{V}| \geqslant n - m$. If $|\hat{u}_i| < 1$ for all $i \in \mathcal{V}$ and $\mathbf{A}_{\mathcal{V}^c}$ is full column rank [2], then $\hat{\boldsymbol{x}}$ is the unique minimum.

The proof of this Lemma is given in Appendix A.

According to the proof of Lemma 1, if (7) can be split into

$$\begin{aligned}&\frac{1}{m}\mathbf{A}_{\mathcal{V}^c}^T(\boldsymbol{y} - \mathbf{A}_{\mathcal{V}^c}\hat{\boldsymbol{x}}_{\mathcal{V}^c}) = h \cdot \mathbf{W}_{\mathcal{V}^c} \cdot \hat{\boldsymbol{u}}_{\mathcal{V}^c} \\ &\left|\frac{\mathbf{A}_i^T}{m}(\boldsymbol{y} - \mathbf{A}_{\mathcal{V}^c}\hat{\boldsymbol{x}}_{\mathcal{V}^c})\right| < hw_i \quad \text{for } i \in \mathcal{V},\end{aligned} \tag{8}$$

there exists an $\hat{\boldsymbol{u}}$, which satisfies $|\hat{u}_i| < 1$ for $i \in \mathcal{V}$ such that $\hat{x}_i = 0$ hold for $i \in \mathcal{V}$. According to (b) in Lemma 1, $\hat{\boldsymbol{x}}$ is the unique minimum of (5) provided $|\mathcal{V}| \geqslant n - m$. Let $\mathcal{S} = \text{supp}(\boldsymbol{x}^*)$. We further establish a sufficient condition under which (5) recovers the support of $\boldsymbol{x}^*$ exactly, i.e., $\text{sign}(\hat{\boldsymbol{x}}) = \text{sign}(\boldsymbol{x}^*)$.

Lemma 2: (**Exact Support Recovery**): Assume that $\mathbf{A}$, $\boldsymbol{z}$ and the weights w_i are deterministic and fixed. Then the support of $\boldsymbol{x}^*$ can be recovered exactly from (5), provided the following

[1]This algorithm was presented at the 2016 IEEE International Symposium on Information Theory (ISIT) conference [1].

[2]It follows from $|\mathcal{V}| \geqslant n - m$ that $m \geqslant n - |\mathcal{V}| = |\mathcal{V}^c|$ holds. Hence, the condition that $\mathbf{A}_{\mathcal{V}^c}$ is full column rank always holds.

events are satisfied

$$\begin{cases} \left| \frac{\mathbf{A}_i^T}{m} \left\{ (\mathbf{I} - \mathbf{A}_{\mathcal{S}} \mathbf{A}_{\mathcal{S}}^+) \boldsymbol{z} + mh {\mathbf{A}_{\mathcal{S}}^+}^T \mathbf{W}_{\mathcal{S}} \boldsymbol{u}_{\mathcal{S}} \right\} \right| \\ \qquad\qquad < hw_i, \ \ \forall\, i \in \mathcal{S}^c & \text{(9a)} \\ \operatorname{sign}(\boldsymbol{x}_{\mathcal{S}}^* + \mathbf{A}_{\mathcal{S}}^+ \boldsymbol{z} - mh(\mathbf{A}_{\mathcal{S}}^T \mathbf{A}_{\mathcal{S}})^{-1} \mathbf{W}_{\mathcal{S}} \boldsymbol{u}_{\mathcal{S}}) \\ \qquad\qquad = \operatorname{sign}(\boldsymbol{x}_{\mathcal{S}}^*) & \text{(9b)} \end{cases}$$

where $\mathbf{A}_{\mathcal{S}}^+ = (\mathbf{A}_{\mathcal{S}}^T \mathbf{A}_{\mathcal{S}})^{-1} \mathbf{A}_{\mathcal{S}}^T$ and $\boldsymbol{u}_{\mathcal{S}} = \operatorname{sign}(\boldsymbol{x}_{\mathcal{S}}^*)$.

The proof of this Lemma is given in Appendix B.

Remark 1: The condition in (9a) is a recovery guarantee for the zero entries in $\boldsymbol{x}^*$, which is the key challenge for the exact support recovery of $\boldsymbol{x}^*$, while the one in (9b) guarantees the sign recovery of the nonzero entries in $\boldsymbol{x}^*$.

If there exists at least one zero weight outside of the support of $\boldsymbol{x}^*$, we obtain the following pessimistic conclusion.

Proposition 1: (**Failure Condition**): Suppose that the weights w_i are deterministic and fixed, the entries of both $\mathbf{A}$ and $\boldsymbol{z}$ are drawn randomly from the Gaussian distribution. If there exists at least one zero weight outside of the support of $\boldsymbol{x}^*$, the support of $\boldsymbol{x}^*$ with probability one can not be recovered exactly by solving (5) alone.

The proof of this Proposition is given in Appendix C.

From the proof of Proposition 1, we observe that although $\hat{\boldsymbol{x}}$ can not recover the zero entries in $\boldsymbol{x}^*$ whose indices are included in $\mathcal{T} \setminus \mathcal{S}$, where $\mathcal{T} \triangleq (\operatorname{supp}(\boldsymbol{w}))^c$, it is possible to guarantee that $\operatorname{supp}(\boldsymbol{x}^*) \subset \operatorname{supp}(\hat{\boldsymbol{x}})$ holds. Hence, redefine $\mathcal{V}$ be any subset of $\{\mathcal{S} \cup \mathcal{T}\}^c$ with $|\mathcal{V}| \geqslant n - m$. We investigate the guarantee for exact recovery of $\operatorname{sign}(\boldsymbol{x}_{\mathcal{S}\cup\mathcal{V}}^*)$ by employing (5) and obtain the following sufficient condition.

Lemma 3: (**Partial Support Recovery**): Assume that $\mathbf{A}$, $\boldsymbol{z}$ and the weights w_i are deterministic and fixed. Then the support of $\boldsymbol{x}_{\mathcal{S}\cup\mathcal{V}}^*$ can be recovered exactly from (5), provided the following events are satisfied

$$\begin{cases} \left| \frac{\mathbf{A}_i^T}{m} \left\{ (\mathbf{I} - \mathbf{A}_{\mathcal{V}^c} \mathbf{A}_{\mathcal{V}^c}^+) \boldsymbol{z} + mh {\mathbf{A}_{\mathcal{V}^c}^+}^T \mathbf{W}_{\mathcal{V}^c} \hat{\boldsymbol{u}}_{\mathcal{V}^c} \right\} \right| \\ \qquad\qquad < hw_i, \quad \forall\, i \in \mathcal{V} & \text{(10a)} \\ \Lambda > \left\| \mathbf{A}_{\mathcal{V}^c}^+ \boldsymbol{z} - mh(\mathbf{A}_{\mathcal{V}^c}^T \mathbf{A}_{\mathcal{V}^c})^{-1} \mathbf{W}_{\mathcal{V}^c} \hat{\boldsymbol{u}}_{\mathcal{V}^c}) \right\|_\infty & \text{(10b)} \end{cases}$$

where Λ is the smallest absolute nonzero entry in $\boldsymbol{x}^*$ and $\mathbf{A}_{\mathcal{V}^c}^+ = (\mathbf{A}_{\mathcal{V}^c}^T \mathbf{A}_{\mathcal{V}^c})^{-1} \mathbf{A}_{\mathcal{V}^c}^T$.

The proof of this Lemma is given in Appendix D.

Remark 2: (i) When $\mathcal{T} \setminus \mathcal{S}$ is empty and $\mathcal{V} = \mathcal{S}^c$, Lemma 3 reduces to Lemma 2. (ii) Lemma 3 provides an opportunity to combine a threshold method to recover the signs of the remaining components from $\hat{\boldsymbol{x}}$, so that, eventually, exact support recovery of $\boldsymbol{x}^*$ can be achieved.

B. Sample Complexity Analysis

To elaborate the precise requirements on the number of measurements needed, we analyze the sample complexity for the general weighted ℓ_1 minimization to exactly recover the support of $\boldsymbol{x}^*$ (or the support of its reduced dimensional vector). We will provide high probability results under key assumptions for the distributions of $\mathbf{A}$ and $\boldsymbol{z}$. As before, $\mathbf{W}$ is deterministic and fixed. The main results are stated as follows.

Theorem 1: Assume that $\mathcal{V}$ is any subset of $\{\mathcal{S} \cup \mathcal{T}\}^c$ with $|\mathcal{V}| \geqslant n - m$, the elements of $\mathbf{A}$, $a_{i,j} \sim \mathcal{N}(0, 1)$, the entries of $\boldsymbol{z}$, $z_i \sim \mathcal{N}(0, \sigma_z^2)$ and $\mathbf{W}$ is deterministic and fixed. One solves (5) to recover $\boldsymbol{x}^*$ from the measurements $\boldsymbol{y} = \mathbf{A}\boldsymbol{x}^* + \boldsymbol{z}$. Define

$$g(h) = c_1 \tau h + 20 \sqrt{\frac{\sigma_z^2 \log(n - |\mathcal{V}|)}{m}} \tag{11}$$

where $\tau = \max\{w_i : i \in \mathcal{V}^c\}$ and c_1 is a positive constant. Suppose

$$m > \max\left(|\mathcal{V}^c|, 2\eta k \log(|\mathcal{V}|)(1 + \epsilon') \left(1 + \frac{\sigma_z^2}{h^2 \xi k}\right) \right), \tag{12}$$

where $\xi = \sum_{i \in \mathcal{V}^c} w_i^2 / k$, $\eta = \max_{i \in \mathcal{V}} \{\frac{\xi}{w_i^2}\}$ and $\epsilon' = \max\{\epsilon, 8\sqrt{\frac{n - |\mathcal{V}|}{m}}\}$ for some fixed $\epsilon > 0$.

(a) If $\Lambda > g(h)$, then $\operatorname{sign}(\hat{\boldsymbol{x}}_{\mathcal{S}\cup\mathcal{V}}) = \operatorname{sign}(\boldsymbol{x}_{\mathcal{S}\cup\mathcal{V}}^*)$ holds, with probability greater than

$$1 - c_2 \exp(-c_3 \min\{m\epsilon^2, n - |\mathcal{V}|, \log|\mathcal{V}|\}) \tag{13}$$

for some positive constants c_2 and c_3.

(b) Compute vector $\widetilde{\boldsymbol{x}}$ by

$$\widetilde{x}_i = \begin{cases} \hat{x}_i + g(h) & \hat{x}_i \leq -g(h) \\ 0 & |\hat{x}_i| \leq g(h) \\ \hat{x}_i - g(h) & \hat{x}_i \geq g(h) \end{cases}, \tag{14}$$

i.e., the soft-thresholding nonlinearity. If $\Lambda > 2g(h)$, then $\widetilde{\boldsymbol{x}}$ recovers the support of $\boldsymbol{x}^*$ exactly, with probability greater than

$$1 - c_2 \exp(-c_3 \min\{m\epsilon^2, n - |\mathcal{V}|, \log|\mathcal{V}|\}) \tag{15}$$

for some positive constants c_2 and c_3.

The proof of this Theorem is given in Appendix E.

C. Discussion and Numerical Validation of Theorem 1

We first discuss the case that $w_i > 0$, $\forall 1 \leqslant i \leqslant n$, and then extend the discussion to the case that allows some weights to be zero, i.e., $w_i \geqslant 0$, $\forall 1 \leqslant i \leqslant n$. For w_i strictly positive, we can tighten the results of Theorem 1.

1) $w_i > 0$ case: As the problem dimension increases, we assume that m is sufficiently large as specified in Theorem 1. Hence, let $\mathcal{V} = \mathcal{S}^c$, inequality $8\sqrt{\frac{k}{m}} < \epsilon$ holds for any fix $\epsilon > 0$ [10]. We obtain the following Corollary.

Corollary 1: Assume that the elements of $\mathbf{A}$, $a_{i,j} \sim \mathcal{N}(0, 1)$, the entries of $\boldsymbol{z}$, $z_i \sim \mathcal{N}(0, \sigma_z^2)$ and $\mathbf{W}$ is deterministic and fixed. Consider the optimization in (5) to recover $\boldsymbol{x}^*$ from the measurements $\boldsymbol{y} = \mathbf{A}\boldsymbol{x}^* + \boldsymbol{z}$. Define

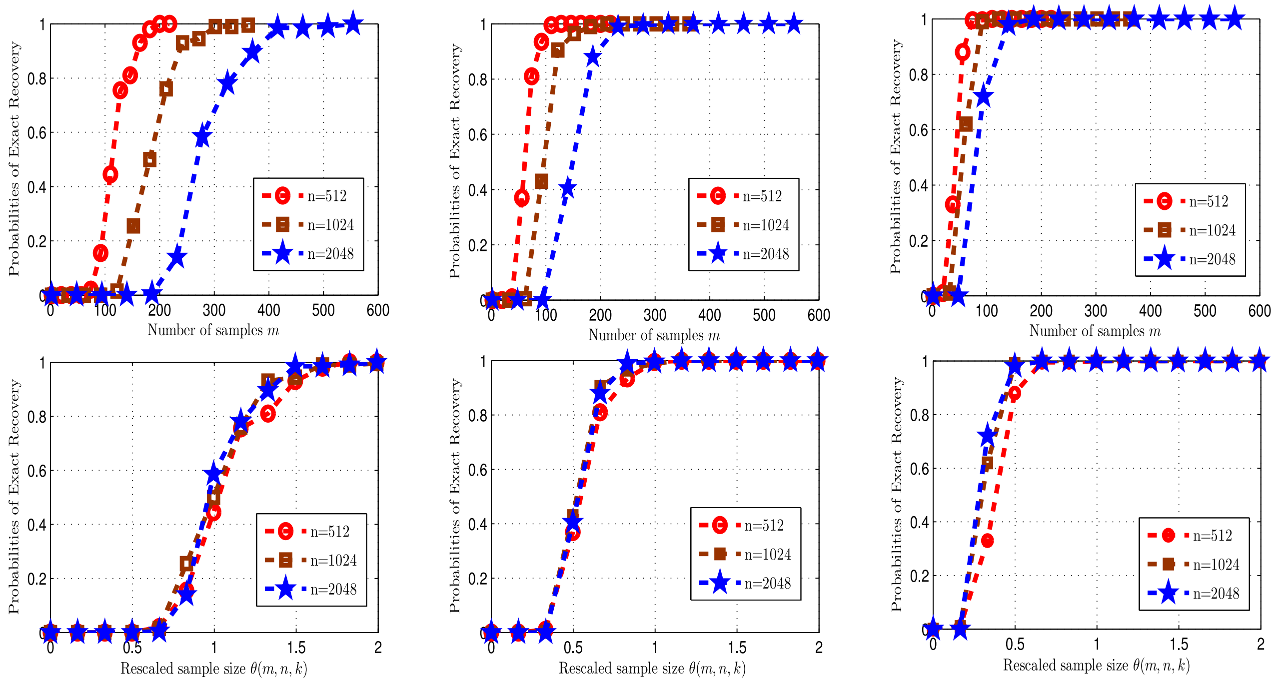
$$g(h) = c_1 \tau h + 20 \sqrt{\frac{\sigma_z^2 \log(k)}{m}}, \tag{16}$$

where $\tau = \max\{w_i : i \in \mathcal{S}\}$ and c_1 is a positive constant. If $\Lambda > g(h)$ and for some fixed $\epsilon > 0$, if

$$m > \max\left(k, 2\eta k \log(n - k)(1 + \epsilon) \left(1 + \frac{\sigma_z^2}{h^2 \xi k}\right) \right), \tag{17}$$

where $\xi = \sum_{i \in \mathcal{S}} w_i^2 / k$ and $\eta = \max_{i \in \mathcal{S}^c} \{\frac{\xi}{w_i^2}\}$, then $\hat{\boldsymbol{x}}$ recovers the support of $\boldsymbol{x}^*$ exactly, with probability greater than

$$1 - c_2 \exp(-c_3 \min\{k, \log(n - k)\}) \tag{18}$$



(a) $\eta = 1$. (b) $\eta = 0.5$. (c) $\eta = 0.25$.

Fig. 1. Simulation results of optimization problem (5) with $\eta = 1$, $\eta = 0.5$ and $\eta = 0.25$, respectively. The top of each sub figure shows the probabilities of exact recovery versus the sample size m for three different problem size n. The bottom of each sub figure shows the probabilities of exact recovery versus the rescaled sample size size $\theta(m,n,k) = m/[2k\log(n-k)]$. In all cases, sparsity $k = \lceil 0.4n^{0.5} \rceil$.

for some positive constants c_2 and c_3.

Remark 3: Corollary 1 indicates that if $m > 2\eta k\log(n-k)$ holds and the smallest magnitude of the non-zero entries of $\boldsymbol{x}^*$ is large enough, the optimization problem (5) can, with high probability, recover the support of $\boldsymbol{x}^*$ exactly where the important parameter η is directly related to the weights. In practice, we can reduce the sample complexity to some extent by selecting weights such that η is minimized.

Remark 4: The sample complexity analysis [10] for standard ℓ_1 minimization is a special case of Corollary 1 where $\mathbf{W}$ is an identity matrix. Similar to the analysis in [10], if we set

$$h = \sqrt{\frac{\phi_n}{\varsigma}\frac{2\sigma_z^2\log(n-k)}{m}} \tag{19}$$

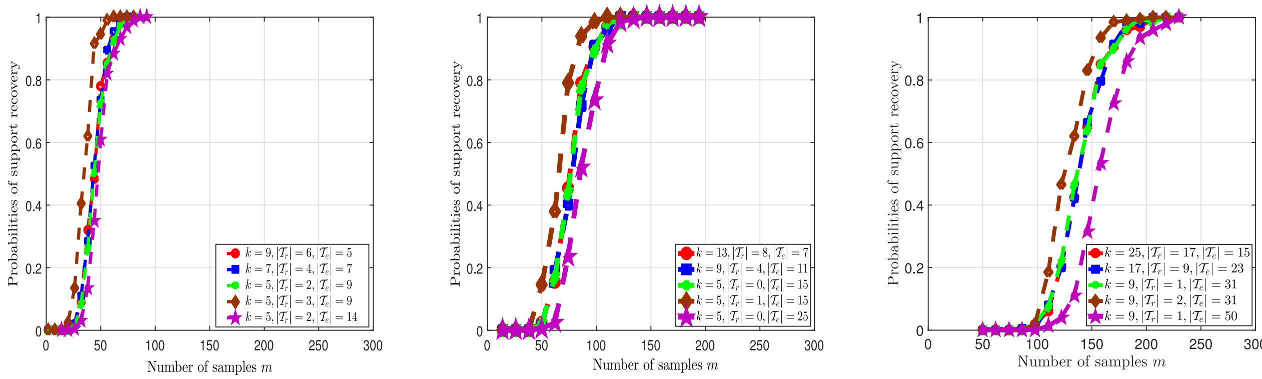
for some $\phi_n \geq 2$, where $\varsigma = \min\left\{w_i^2 : i \in \mathcal{V}\right\}$, then it suffices to have

$$m > \max\left(k, 2\eta k\log(n-k)\left((1+\epsilon)^{-1} - \frac{1}{\phi_n}\right)^{-1}\right) \tag{20}$$

for some $\epsilon > 0$. Further, if we choose the standard Gaussian matrix and an h with $\phi_n \to +\infty$, Theorem 1, with high probability, guarantees the exact support recovery of $\boldsymbol{x}^*$ with approximately $m = \max(k, 2\eta k\log(n-k))$ samples.

Theorem 1 provides a scaling law for general weighted ℓ_1 minimization in terms of exact support recovery, which is related to the system parameters $(n, k, \mathbf{W})$. Moreover, according to the scaling law, the parameter η, which is an explicit function of the weights, can predict the relative performance of weighed ℓ_1 minimization with different weights. We have conducted simulations to validate the scaling law and the η. In the experiments, the non-zero elements of the k-sparse signal is ± 1 uniformly at random. The components of $\mathbf{A}$ are drawn randomly from the standard Gaussian distribution and noise level $\sigma_z = 0.5$. Based on Remark 4, the choice of h follows equation (19) with $\phi_n = 9$ in our experiments. At first, BPDN [14] is employed to recover $\boldsymbol{x}^*$. According to Theorem 1, BPDN, as a special case of the weighted ℓ_1 minimization, has $\eta = 1$. In the top portion of Fig. 1(a), we plot the probabilities of exact support recovery versus the sample size m for three different problem sizes $n \in \{512, 1024, 2048\}$, and $k = \lceil 0.4\sqrt{n} \rceil$ in each case. We repeat the experiment 200 times at each point. The probabilities of exact support recovery vary from zero to one; the larger the problem size, the more samples required. However, according to the scaling predicted by Theorem 1, i.e., $m >= 2\eta\zeta k\log(n-k)$, ζ is a constant. Thus, the bottom portion of Fig. 1(a) plots the same experimental results, but the probabilities of exact support recovery are now plotted versus an "appropriately rescaled" version of the sample size, i.e., $\theta(m,n,k) = m/[2k\log(n-k)]$. In the bottom portion of Fig. 1(a), all of the curves now line up with one another, even though the problem sizes and sparsity levels vary dramatically. All of the cases have probabilities of exact support recovery that equal to one at $\theta(m,n,k) = \zeta \approx 2$. Thus, the experimental result matches the theoretical prediction of the scaling law very well.

To further validate the predictive properties of η, the same experiments are performed with weighted ℓ_1 minimization with non-uniform weights. Two classes of weights are tested where one weights the non-zero elements of k-sparse signals with



(a) $n = 512$. (b) $n = 1024$. (c) $n = 2048$.

Fig. 2. The probabilities of identifying $\text{sign}(\boldsymbol{x}^*_{\mathcal{S}\cup\mathcal{V}})$ versus the sample size m for three different problem size $n \in \{512, 1024, 2048\}$.

$w_i = \sqrt{2}/2$ and another with $w_i = 1/2$. Both strategies weight the zero elements with $w_i = 1$. We used the YALL1 toolbox [31] to implement the weighted ℓ_1 minimization. The experimental results are plotted in Fig. 1(b) and Fig. 1(c), respectively. According to Theorem 1, the weighted ℓ_1 minimization has $\eta = 0.5$ and $\eta = 0.25$ for the two classes of weights. As shown in the bottom portions of both Fig. 1(b) and Fig. 1(c), the curves achieve the probability of exact support recovery equals to one at $\theta(m,n,k) = \frac{1}{2}\zeta \approx 1$ and $\theta(m,n,k) = \frac{1}{4}\zeta \approx 0.5$, respectively. We observed that the simulation results match the theoretical predictions of η very well, which indicates that η provides a good prediction of the relative performance of weighted ℓ_1 minimization with different weights.

2) $w_i \geqslant 0$ case: There is a large body of work [16]–[18], such as modified-BPDN [16], [18], which weights the partially known support with zero weight. However, errors in the partially known support will lead to zero weights outside of the support [32]. In this scenario, the support of $\boldsymbol{x}^*$ may be impossible to be recovered exactly by solving (5) alone. Theorem 1 indicates that even for this situation, (5) can, under certain conditions, exactly recover the support of $\boldsymbol{x}^*$ by means of a thresholding method.

Remark 5: A special case of weighted ℓ_1 minimization is modified-CS [18] which weights the partially known support as zero. According to condition (12), if the partially known support is accurate, i.e., $\mathcal{V} = \mathcal{S}^c$, this weighting strategy ensures that $\eta < 1$ holds. Comparing with the classical result $m > 2k\log(n-k)$ [10] required by the BPDN where $\eta = 1$, modified-CS achieves a reduced sample complexity for exact support recovery by exploiting the support information as predicted by the respective values of η.

Remark 6: The sample complexity of the weighted ℓ_1 minimization in terms of exact support recovery of $\boldsymbol{x}^*_{\mathcal{S}\cup\mathcal{V}}$ is dominated by the system parameters $(n, |\mathcal{V}|, \eta k)$. More specifically, let $\mathcal{V} = \{\mathcal{S}\cup\mathcal{T}\}^c$ with $|\mathcal{V}| \geqslant n - m$, where $\mathcal{T}$ can be divided into $\mathcal{T}_r \triangleq \mathcal{T}\cap\mathcal{S}$ and $\mathcal{T}_e \triangleq \mathcal{T}\setminus\mathcal{S}$. Then, (i) an increase in $|\mathcal{T}_r|$ yields a decrease of η, which reduces the sample complexity; (ii) an increase in $|\mathcal{T}_e|$ results in the increase of the term $\log(|\mathcal{V}|)(1 + 8\sqrt{\frac{n-|\mathcal{V}|}{m}})$ so that more samples are required to exactly recover the support; (iii) If the triples $(n, k + |\mathcal{T}_e|, \eta k)$ are the same, the same sample complexity is required for exact support recovery of $\boldsymbol{x}^*_{\mathcal{S}\cup\mathcal{V}}$.

We conduct further numerical simulations to validate Theorem 1 and Remark 6. In the experiments, the non-zero elements of the k-sparse signal are ± 10 uniformly at random. The components of $\mathbf{A}$ are drawn randomly from the standard Gaussian distribution and noise standard deviation is $\sigma_z = 0.05$. The weights for weighted ℓ_1 minimization are restricted to $w_i \in \{0, 1\}$, i.e., a modified-CS model. At first, we vary k, $|\mathcal{T}_e|$ and $|\mathcal{T}_r|$, but keep $k + |\mathcal{T}_e|$ and ηk constant. According to Theorem 1 and (iii) in Remark 6, the model should exactly recover $\text{sign}(\boldsymbol{x}^*_{\mathcal{S}\cup\mathcal{V}})$ with the same sample complexity. In Fig. 2, we plot the probabilities of exactly recovering $\text{sign}(\boldsymbol{x}^*_{\mathcal{S}\cup\mathcal{V}})$ versus the sample size m for three different problem sizes $n \in \{512, 1024, 2048\}$. The experiment was repeated 200 times at each point. In all of the cases, the red, blue and green curves line up with one another, which indicate the experimental result match (iii) in Remark 6 very well. Further, we make $|\mathcal{T}_r|$ or $|\mathcal{T}_e|$ increase in all of the cases. It can be inferred from (i) and (ii) in Remark 6 that the sample complexity will decrease or increase, respectively. As shown in Fig. 2, all of the experimental results match these theoretical predictions very well. On the other hand, although not shown in Fig. 2, we observe that in all of the experiments, $\text{sign}(\boldsymbol{x}^*)$ can not be recovered exactly by solving (5) alone, which validates the conclusion in Proposition 1. However, if we compute vector $\widetilde{\boldsymbol{x}}$ by $\widetilde{x}_i = \text{sign}(\hat{x}_i)(|\hat{x}_i| - 5)_+$, then $\text{sign}(\tilde{\boldsymbol{x}}) = \text{sign}(\boldsymbol{x}^*)$ holds.

Remark 7: Even though some zero entries in $\boldsymbol{x}^*$ can not be recovered by solving (5) alone, since zero weights are assigned to these indices, it is possible to recover them by means of a thresholding method. This fact is important for real applications because the prior information is often inaccurate, which leads to the result that a fraction of the weights are biased. Fortunately, utilizing a thresholding method, the weighted ℓ_1 minimization is robust to the inappropriate weights.

III. Binary Signal Recovery: Promoting Sparsity and Minimizing η via bPRL1 Algorithm

In this section, a new iterative reweighted algorithm is proposed for the estimation of a sparse binary vector. Due to the binary property, binary vector estimation is equivalent to support recovery. Our algorithm is motivated by the spectrum occupancy estimation problem in [7]. Therein, primary users (PUs) occupy a portion of the spectrum band; secondary users (SUs) attempt to estimate the spectrum to detect available spectrum opportunities to use for their own data transmissions. We denote the spectrum occupancy vector (SOV) as $\boldsymbol{x}^* \in \{0,1\}^n$, where $x_i^* = 1$ if the i-th spectrum band is occupied by a PU, and $x_i^* = 0$ if it is not occupied (hence available for use by SUs). The goal for SUs is to detect $\boldsymbol{x}^*$ with high accuracy, in order to make use of all unused spectrum opportunities, while minimizing the interference generated to PUs. To this end, SUs collect m compressed noisy measurements as

$$y_i = \mathbf{A}^i \boldsymbol{x}^* + z_i, \forall i = 1, 2, \ldots, m, \tag{21}$$

based on which they attempt to recover $\boldsymbol{x}^*$, where $z_i \sim \mathcal{N}(0, \sigma_z^2)$ is Gaussian noise, i.i.d. over time and across the SUs, and $\mathbf{A}^i$ is the measurement vector at SU i. This measurement model is the result of filtering operations performed at each SU across the spectrum bins, thus $\mathbf{A}^i$ denotes the filtering coefficient vector, which also includes the signal attenuation between the primary user (PU) and the SU. We assume that the entries of $\mathbf{A}^i$ are Gaussian with zero mean and unit variance. Equation (21) can be expressed in matrix notation as

$$\boldsymbol{y} = \mathbf{A}\boldsymbol{x}^* + \boldsymbol{z}. \tag{22}$$

Because PUs join and leave the network at random times, there is temporal correlation in the SOV. Thus, past spectrum measurements provide information so as to estimate $\boldsymbol{x}^*$ at the current time. We denote the availability of prior information by the vector $\boldsymbol{\beta}$, where β_i denotes the prior probability that $x_i = 1$, which is a result of all past observations. Note that if $\beta_i \approx \frac{1}{2}$, the prior is not informative about the current state of the frequency bin i.

Utilizing $\boldsymbol{\beta}$, the *maximum-a-posteriori* (MAP) estimate $\hat{\boldsymbol{x}}^{(\mathrm{MAP})}$ is given by [7]

$$\begin{aligned}\hat{\boldsymbol{x}}^{(\mathrm{MAP})} &= \arg\max_{\boldsymbol{x} \in \{0,1\}^n} \mathbb{P}(\boldsymbol{x}|\boldsymbol{\beta}, \mathbf{A}, \boldsymbol{y}) \\ &= \arg\min_{\boldsymbol{x} \in \{0,1\}^n} \frac{1}{2} \|\boldsymbol{y} - \mathbf{A}\boldsymbol{x}\|_2^2 + \sigma_z^2 \sum_i w_i x_i,\end{aligned} \tag{23}$$

where $w_i = \ln(\frac{1-\beta_i}{\beta_i})$. On the other hand, an initial estimate of $\boldsymbol{x}^*$ based on $\boldsymbol{\beta}$ alone is $\hat{\boldsymbol{x}} = \chi(\boldsymbol{\beta} \geq 0.5)$ where $\chi(\cdot)$ is the entry-wise indicator function. The *residual uncertainty vector* $\boldsymbol{e}^*$ can be computed via $\boldsymbol{e}^* = \hat{\boldsymbol{x}} \oplus \boldsymbol{x}^*$ where $\oplus$ is the XOR operator. Therefore, incorporating $\hat{\boldsymbol{x}}$ into (23), the MAP estimation of $\boldsymbol{x}^*$ can be transformed into the MAP estimation of $\boldsymbol{e}^*$, which satisfies

$$\hat{\boldsymbol{e}}^{(\mathrm{MAP})} = \hat{\boldsymbol{x}} \oplus \hat{\boldsymbol{x}}^{(\mathrm{MAP})}. \tag{24}$$

Combining (23) and (24), one has [7]

$$\hat{\boldsymbol{e}}^{(\mathrm{MAP})} = \arg\min_{\boldsymbol{e} \in \{0,1\}^n} \frac{1}{2} \left\|\hat{\boldsymbol{y}} - \hat{\mathbf{A}}\boldsymbol{e}\right\|_2^2 + \sigma_z^2 \|\mathbf{W}\boldsymbol{e}\|_1, \tag{25}$$

where $\hat{\boldsymbol{y}} = \boldsymbol{y} - \mathbf{A}\hat{\boldsymbol{x}}$, $\hat{\mathbf{A}} = \mathbf{A}(\mathbf{I} - 2\mathrm{diag}(\hat{\boldsymbol{x}}))$, $\mathbf{W} = \mathrm{diag}(\boldsymbol{w})$ and

$$w_i = (1 - 2\hat{x}_i) \ln\left(\frac{1-\beta_i}{\beta_i}\right). \tag{26}$$

Remark 8: Due to the discrete set $\{0,1\}^n$, the optimization specified by Equation (25) has exponential complexity. We convert to a convex problem by relaxing $\{0,1\}^n$ to the convex set $[0,1]^n$ enabling the use of convex optimization techniques. Suppose $\tilde{\boldsymbol{e}}^{(\mathrm{MAP})}$ is the solution of the corresponding convex problem. Then $\hat{\boldsymbol{e}}^{(\mathrm{MAP})}$ can be approximately obtained by computing $\chi(\tilde{\boldsymbol{e}}^{(\mathrm{MAP})} \geq 0.5)$.

If $\boldsymbol{\beta}$ is informative, $\boldsymbol{e}^*$ is well-modeled as sparse even though $\boldsymbol{x}^*$ may not be. In other words, the prior estimate $\hat{\boldsymbol{x}}$ may differ from $\boldsymbol{x}^*$ only in few entries, due to an informative prior. Furthermore, a more effective sparsity measure $\|\mathbf{W}\boldsymbol{e}\|_1$ can be obtained via the non-uniform weighting which is a function of $\boldsymbol{\beta}$. By exploiting sparsity, we can better estimate $\boldsymbol{e}^*$ than $\boldsymbol{x}^*$. However, a lack of informative conditional priors can have a serious impact on our ability to estimate $\boldsymbol{e}^*$. Consider the following example, let $\mathcal{V}^c = \mathcal{S}$, the support of $\boldsymbol{e}^*$. And suppose that $e_1^* = 0$ (i.e., $i = 1 \in \mathcal{V}$) and $\beta_1 = 0.49$. Using (26), the weight w_1 assigned to e_1 is approximately zero. According to our theoretical result in Theorem 1, it incurs a relatively large η such that the sample complexity for exact support recovery increases. Hence, if $\boldsymbol{\beta}$ is uninformative (e.g., close to 1/2), the sparsity assumption of $\boldsymbol{e}^*$ can be violated with high probability. At the same time, the uninformative priors can lead to weights that incur biases that further degrade performance. It motivates us to develop a new algorithm to mitigate the impact of uninformative conditional belief.

Our strategy is to use $\hat{\boldsymbol{e}}^{(\mathrm{MAP})}$ and $\tilde{\boldsymbol{e}}^{(\mathrm{MAP})}$ to update the initial estimate $\hat{\boldsymbol{x}}$ and the weights $\mathbf{W}$, respectively. Then a new optimization problem can be constructed, which has better recoverability. Firstly, we make an assumption that even if inexact recovery occurs, $\hat{\boldsymbol{e}}^{(\mathrm{MAP})}$ may lead to good *partial* support recovery. In fact, this assumption is supported by recent results [33], which show the requirement that *most of the support of current estimate belong to the support of the true signal* is far less stringent than the requirement for exact recovery. Further, we obtain the following Proposition.

Proposition 2: If most of the support of $\hat{\boldsymbol{e}}^{(\mathrm{MAP})}$ as defined in Equation (25) belongs to the support of $\boldsymbol{e}^*$, i.e.,

$$\frac{\mathrm{supp}(\hat{\boldsymbol{e}}^{(\mathrm{MAP})}) \cap \mathrm{supp}(\boldsymbol{e}^*)}{\mathrm{supp}(\hat{\boldsymbol{e}}^{(\mathrm{MAP})})} > \frac{1}{2}, \tag{27}$$

Then $\left\|\hat{\boldsymbol{e}}^{(\mathrm{MAP})}) \oplus \hat{\boldsymbol{x}} \oplus \boldsymbol{x}^*\right\|_0 < \|\hat{\boldsymbol{x}} \oplus \boldsymbol{x}^*\|_0 = \|\boldsymbol{e}^*\|_0$ holds.

The proof of this Proposition is given in Appendix F.

For ease of description, we redefine $\hat{\boldsymbol{x}}^{(0)} = \hat{\boldsymbol{x}}$, $\boldsymbol{e}^{(1)} = \boldsymbol{e}^*$, $\hat{\boldsymbol{e}}^{(1)} = \hat{\boldsymbol{e}}^{(\mathrm{MAP})}$, $\hat{\boldsymbol{x}}^{(1)} = \hat{\boldsymbol{e}}^{(1)} \oplus \hat{\boldsymbol{x}}^{(0)}$ and $\boldsymbol{e}^{(2)} = \hat{\boldsymbol{x}}^{(1)} \oplus \boldsymbol{x}^*$. Proposition 2 motivates us to update $\hat{\boldsymbol{x}}$ to $\hat{\boldsymbol{x}}^{(1)}$ because the corresponding *residual uncertainty vector*, i.e., $\boldsymbol{e}^{(2)}$, is more sparse than $\boldsymbol{e}^*$. Secondly, we use $\tilde{\boldsymbol{e}}^{(\mathrm{MAP})}$ to update $\mathbf{W}$. Consider the specific situations that $\beta_1 = 0$ (or 1), then we have the initial estimate $\hat{x}_1^{(0)} = 0$ (or 1). If $\tilde{e}_1^{(\mathrm{MAP})} = 1$, then x_1^* is more likely to be one (or zero) and we hope that w_1 tends to zero (or $+\infty$). Conversely, if $\tilde{e}_1^{(\mathrm{MAP})} = 0$, then x_1^* is more likely to be zero (or one) and $w_1 \to +\infty$ (or $\to 0$) is preferred. Hence, to assign the appropriate weights, we use $\tilde{\boldsymbol{e}}^{(\mathrm{MAP})}$ to update the weights by adapting the approximated conditional belief $\boldsymbol{\beta}^{(1)} = \frac{\boldsymbol{\beta} + \tilde{\boldsymbol{e}}^{(\mathrm{MAP})}}{2}$. This update rule provides a simple tradeoff between the conditional belief of $\boldsymbol{x}^*$ and the current estimate. Finally, similar to

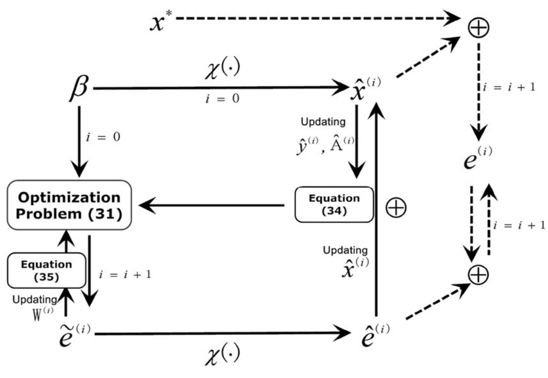


Fig. 3. A sketch of bPRL1 Algorithm and an illustration for promoting sparsity in $\boldsymbol{e}^{(i)}$.

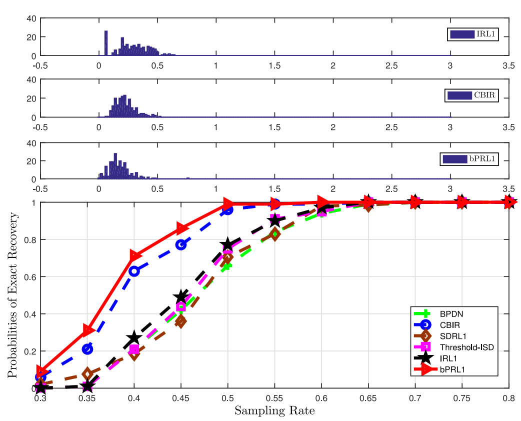


Fig. 4. The histograms of η and the probability of exact recovery for different algorithms when $\tau = 0.1$.

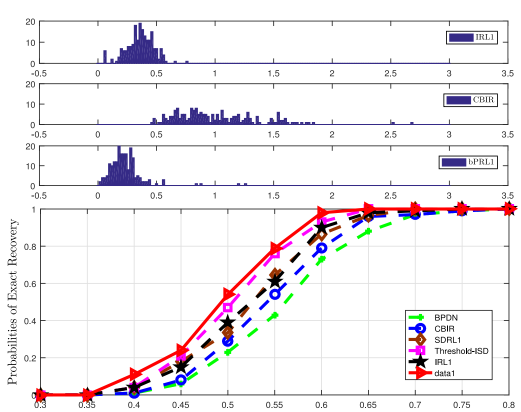


Fig. 5. The histograms of η and the probability of exact recovery for different algorithms when $\tau = 0.2$.

(25), a new optimization problem with the updated estimate $\hat{\boldsymbol{x}}^{(1)}$ and the approximated conditional belief $\boldsymbol{\beta}^{(1)}$ can be constructed.

Clearly, the above argument can be extended to $\hat{\boldsymbol{x}}^{(i)}$ with $i \geq 1$. Denote $\hat{\boldsymbol{x}}^{(i)} = \hat{\boldsymbol{x}}^{(i-1)} \oplus \hat{\boldsymbol{e}}^{(i)}$, for $i \geq 1$. A sequence of optimization problems with better recoverability (due $\boldsymbol{e}$ being more sparse and $\boldsymbol{\beta}^{(i)}$ being more informative) are constructed as follows,

$$\tilde{\boldsymbol{e}}^{(i)} = \arg\min_{\boldsymbol{e}\in[0,1]^n} \frac{1}{2}\left\|\hat{\boldsymbol{y}}^{(i-1)} - \hat{\mathbf{A}}^{(i-1)}\boldsymbol{e}\right\|_2^2 + \sigma_z^2 \left\|\mathbf{W}^{(i-1)}\boldsymbol{e}\right\|_1, \tag{28}$$

where $\hat{\boldsymbol{y}}^{(i-1)} = \boldsymbol{y} - \mathbf{A}\hat{\boldsymbol{x}}^{(i-1)}$, $\hat{\mathbf{A}}^{(i-1)} = \mathbf{A}(\mathbf{I} - 2\mathrm{diag}(\hat{\boldsymbol{x}}^{(i-1)}))$ and

$$w_j^{(i-1)} = (1 - 2\hat{x}_j^{(i-1)})\ln\left(\frac{1-\beta_j^{(i-1)}}{\beta_j^{(i-1)}}\right) \tag{29}$$

with $\boldsymbol{\beta}^{(0)} = \boldsymbol{\beta}$. Further, $\hat{\boldsymbol{e}}^{(i)} \triangleq \chi(\tilde{\boldsymbol{e}}^{(i)} \geq 0.5)$ and $\boldsymbol{\beta}^{(i)}$ can be updated as

$$\boldsymbol{\beta}^{(i)} = \frac{\boldsymbol{\beta}^{(i-1)} + \tilde{\boldsymbol{e}}^{(i)}}{2}, \quad i \geq 1. \tag{30}$$

We summarize the bPRL1 algorithm in Algorithm 1.

Algorithm 1: bPRL1 Algorithm.

Input: $\boldsymbol{y}$, $\mathbf{A}$, $\boldsymbol{\beta}$ and δ.
Output: $\hat{\boldsymbol{x}}^{(i)}$
Initialization

$i = 1$,
$\hat{\boldsymbol{x}} = \chi(\boldsymbol{\beta} \geq 0.5)$,
$\hat{\boldsymbol{y}}^{(0)} = \boldsymbol{y} - \mathbf{A}\hat{\boldsymbol{x}}$,
$\hat{\mathbf{A}}^{(0)} = \mathbf{A}(\mathbf{I} - 2\mathrm{diag}(\hat{\boldsymbol{x}}))$ and
$w_j^{(0)} = (1 - 2\hat{x}_j)\ln\left(\frac{1-\beta_j}{\beta_j}\right)$, for $1 \leq j \leq n$.

Repeat

1) Solve the optimization problem:

$$\tilde{\boldsymbol{e}}^{(i)} = \arg\min_{\boldsymbol{e}\in[0,1]^n} \frac{1}{2}\left\|\hat{\boldsymbol{y}}^{(i-1)} - \hat{\mathbf{A}}^{(i-1)}\boldsymbol{e}\right\|_2^2 + \sigma_z^2 \left\|\mathbf{W}^{(i-1)}\boldsymbol{e}\right\|_1 \tag{31}$$

$$\hat{\boldsymbol{e}}^{(i)} = \chi(\tilde{\boldsymbol{e}}^{(i)} \geq 0.5) \tag{32}$$

2) Sparsity-promotion:

$$\hat{\boldsymbol{x}}^{(i)} = \hat{\boldsymbol{x}}^{(i-1)} \oplus \hat{\boldsymbol{e}}^{(i)} \tag{33}$$

$$\hat{\boldsymbol{y}}^{(i)} = \boldsymbol{y} - \mathbf{A}\hat{\boldsymbol{x}}^{(i)}$$

$$\hat{\mathbf{A}}^{(i)} = \mathbf{A}(\mathbf{I} - 2\mathrm{diag}(\hat{\boldsymbol{x}}^{(i)})) \tag{34}$$

3) Weight update:

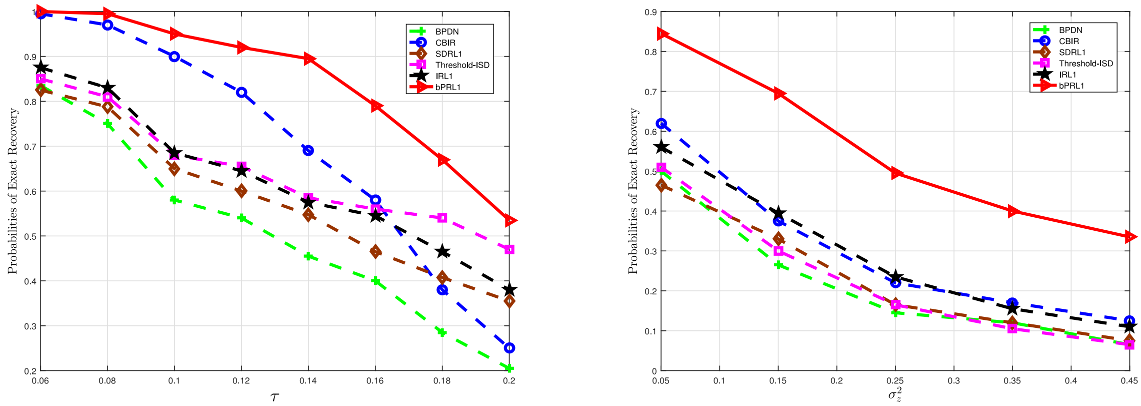
$$\boldsymbol{\beta}^{(i)} = \frac{\boldsymbol{\beta}^{(i-1)} + \tilde{\boldsymbol{e}}^{(i)}}{2}$$

$$w_j^{(i)} = (1 - 2\hat{x}_j^{(i)})\ln\left(\frac{1-\beta_j^{(i)}}{\beta_j^{(i)}}\right) \tag{35}$$

4) $i = i + 1$;

Until $(\hat{\boldsymbol{x}}^{(i)} = \hat{\boldsymbol{x}}^{(i-1)})$ or $(i > \delta)$

To summarize the essence of the bPRL1 algorithm, we provide an algorithm sketch, as well as an illustration for promoting the sparsity in $\boldsymbol{e}^{(i)}$ in Fig. 3 where the solid lines represent the algorithm sketch and the dashed lines describe the promotion of sparsity in $\boldsymbol{e}^{(i)}$, respectively. The core of the bPRL1 algorithm is a dynamic updating of the optimization problem (28) at each iteration. The updating modifies not only the weights, but also the *residual uncertainty vector*. For the i-th iteration of the bPRL1 algorithm, $\boldsymbol{e}^{(i)}$ is the *residual uncertainty vector* of optimization problem (28), which is more sparse than $\boldsymbol{e}^{(i-1)}$



(a) τ versus probability of exact recovery.

(b) Noise variance σ_z^2 versus probability of exact recovery.

Fig. 6. Comparisons between six methods in terms of τ and noise variance σ_z^2 versus probability of exact recovery.

provided the condition in Proposition 2 is satisfied. At the same time, the current output $\tilde{\boldsymbol{e}}^{(i)}$ is also used to update the weights, so that η is minimized. These two modifications work together to improve recovery.

Remark 9: In our experiments, the iterative bPRL1 algorithm generally converged in 2–3 iterations. Thus, the added computational cost for improved support recovery appears to be moderate.

IV. Simulation Results

In this section, we compare bPRL1 to BPDN [14], CBIR [7], SDRL1 [20], IRL1 [19] and threshold-ISD [17]. For BPDN and CBIR, fixed weights are employed where BPDN uses $\mathbf{W} = \mathbf{I}$ and CBIR determines w_i via (26). We observe that bPRL1, SDRL1, IRL1 and threshold-ISD are iterative reweighted algorithms. SDRL1 identifies two support estimates that are updated in every iteration and applies constant weights (i.e., 0 and 1/2) on these estimates. At the s-th iteration, IRL1 computes

$$w_i = (|x_i^{(s-1)} + \rho|)^{p-1}, \tag{36}$$

where $x_i^{(s-1)}$ is the i-th element of the solution at the previous iteration, $p \in [0, 1]$ and ρ is a regularization parameter. Threshold-ISD estimates the support set by employing the current solution. In the subsequent iteration, a weight of zero is applied to the elements whose indices are included in the set. Note that for CBIR, IRL1 and bPRL1, the weights $w_i \in (0, +\infty)$; we have $w_i \in \{0, 1\}$ and $w_i \in \{0, 1/2, 1\}$ for threshold-ISD and SDRL1, respectively. Both IRL1 and threshold-ISD are state-of-the-art in terms of the number of measurements required. However, for the particular problem considered herein, bPRL1 offers improved performance. We underscore that our bPRL1 actively exploits the quality of the prior information, whereas the other algorithms do not. Thus, the performance improvement is to be expected. In our experiments, we used the CVX toolbox [34] for BPDN, CBIR, SDRL1 and bPRL1 implementations; for IRL1 and threshold-ISD implementations, we employed the open source code [35] provided in [17]. However, to constrain the optimization domain to $[0, 1]^n$, we replaced the YALL1 toolbox [31] in the open source code with the CVX toolbox [34]. All test code was tested on a Thinkpad X220i with dual Intel(R) Core(TM) i3 CPUs 2.30GHz and 4GB of memory.

We first generate a random prior, $\pi_i \sim U[0, 1]$ where

$$\begin{cases} \pi_i \triangleq \mathbb{P}(x_i^* = 1), \ \forall i \in [1, n] \\ \sum_{i=1}^{n} \mathbb{P}(x_i^* = 1) = \gamma n \end{cases}. \tag{37}$$

Then, the n-dimensional SOV $\boldsymbol{x}^*$ is generated using this prior. In the experiment, we set $n = 60$ and the sparsity rate $\gamma = 0.2$. The measurement $\boldsymbol{y} = \mathbf{A}\boldsymbol{x}^* + \boldsymbol{z}$ is constructed where the elements of $\mathbf{A}$, $a_{i,j} \sim \mathcal{N}(0, 1)$ and the entries of $\boldsymbol{z}$, $z_i \sim \mathcal{N}(0, 0.05)$. To emulate the conditional prior in [7], we generate conditional priors randomly to consider the cases of informative/uninformative priors. In particular, we select β_i as follows: if $\boldsymbol{x}_i^* = 0$, then $\boldsymbol{\beta_i} \sim U[\tau, \tau + 0.6]$ and if $\boldsymbol{x}_i^* = 1$, then $\boldsymbol{\beta_i} \sim U[0.4 - \tau, 1 - \tau]$. Herein, we provide results for $\tau = 0.1$ and $\tau = 0.2$. According to the model, when $\tau = 0.1$, $\boldsymbol{\beta}$ provides an informative prior such that we can expect that both CBIR and bPRL1 will achieve better performance than the others. However, when $\tau = 0.2$, we have

$$\boldsymbol{\beta_i} \sim \begin{cases} U[0.2, 0.8], & \boldsymbol{x}_i^* = 0 \\ U[0.2, 0.8], & \boldsymbol{x}_i^* = 1 \end{cases}. \tag{38}$$

The conditional prior $\boldsymbol{\beta}$ in this case does not provide any information for both CBIR and bPRL1. Note that we have used $\boldsymbol{\beta}$ to construct an initial support estimate $\hat{\boldsymbol{x}}$, then all of the algorithms in the experiments recover the corresponding *residual uncertainty vector* $\mathbf{e}^*$ rather than $\boldsymbol{x}^*$ directly.

Fig. 4 and Fig. 5 depict the sampling rate (i.e., $\frac{m}{n}$) versus the probability of exact recovery when $\tau = 0.1$ and $\tau = 0.2$, respectively. The experiments are repeated 200 times for each sampling rate. Although not shown here, if $\tau = 0$, CBIR and bPRL1 achieve the same, near unity probability of exact recovery and the other schemes are worse, but have essentially coincident behavior. The top portion of Fig. 4 and Fig. 5 show the histograms of η for three classes of weighted ℓ_1 minimization based recovery algorithms: the IRL1 [19], the CBIR [7] and the proposed bPRL1 algorithm. Note that as all of the algorithms examined couple weighted ℓ_1 minimization with hard-thresholding, small valued elements will be clipped to zero. Thus we compute the average η for the last iteration of each estimation by eliminating the smallest 35% of the weights in the complement of the

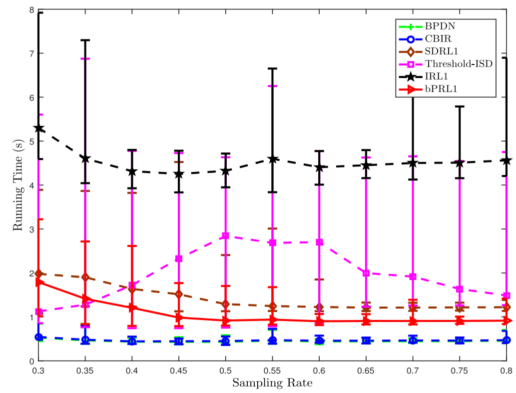


Fig. 7. Comparisons between BPDN, CBIR, IRL1, Threshold-ISD and bPRL1 in terms of τ and noise variance σ_z^2 versus probability of exact recovery.

support set. If we compare the modes of these histograms, we see that in Fig. 4, $\bar{\eta}_{\text{IRL1}} > \bar{\eta}_{\text{CBIR}} > \bar{\eta}_{\text{bPRL1}}$, suggesting that for a fixed sampling rate, we expect the probability of exact recovery to be the best for bPRL1; indeed, this is reflected in the numerical results in the lower portion of the figure. Interestingly, for $\tau = 0.2$, $\bar{\eta}_{\text{CBIR}} > \bar{\eta}_{\text{IRL1}} > \bar{\eta}_{\text{bPRL1}}$, suggesting that CBIR will provide worse performance than IRL1, this is consistent with what we see in the bottom portion of Fig. 5. The bPRL1 still attains the best performance in this case.

Further, Fig. 6(a) presents the probability of exact recovery versus τ for the six methods where the sampling rate is set to 0.5. We observe that the performance of CBIR degrades rapidly as τ increases. When $\tau > 0.18$, CBIR is inferior to SDRL1, IRL1 and threshold-ISD. However, bPRL1 is far superior to the other algorithms even if the conditional prior is uninformative. Fig. 6(b) depicts the probability of exact recovery versus noise variance σ_z^2 where τ=0.15 and the sampling rate is fixed to 0.5. These curves show while varying the noise levels, bPRL1 still maintains a significant improvement on recoverability. We have also compared the computational complexity of the six methods. The results are depicted in Fig. 7. As BPDN and CBIR are not iterative, they are the least complex. However, bPRL1 has lower complexity than SDRL1, threshold-ISD and IRL1 as it converges in very few iterations. Thus, for our problem of interest, exact support recovery in the presence of potentially uninformative priors, bPRL1 provides the best performance/complexity tradeoff.

V. Conclusions

We provide a sample complexity analysis for the general weighted ℓ_1 minimization problem for support recovery which generalizes prior work for standard ℓ_1 minimization by considering arbitrary weights. This result provides a measure which allows one to predict the performance of weighted ℓ_1 minimization based approaches. Motivated by the theory, a new iterative reweighting algorithm, named bPRL1, is proposed which exploits prior information and promotes sparsity through the pruning of the signal within an iteration. Signal pruning capitalizes on partial support recovery and the binary nature of the signal of interest. Our previously computed measure proves to be predictive of simulation results wherein bPRL1 offers superior performance over other reweighting methods that also exploit prior information. We observe that in some cases where other methods achieve a 50% rate of exact recovery, bPRL1 can achieve a 87% rate. At present, we are working on extending the bPRL1 to the non-binary case.

Appendix A
Proof of Lemma 1

Proof: It is known that (5) can be transformed into an equivalent constrained problem with a continuous objective function over a compact set [36]. Therefore, its minimum is always achieved. Based on the first order optimality condition, $\hat{\boldsymbol{x}}$ is a global minimum of (5) if and only if $\exists\, \mathbf{W}\hat{\boldsymbol{u}} \in \partial\|\mathbf{W}\hat{\boldsymbol{x}}\|_1$, such that $\frac{1}{m}\mathbf{A}^T(\mathbf{A}\hat{\boldsymbol{x}} - \boldsymbol{y}) + h\mathbf{W}\hat{\boldsymbol{u}} = \mathbf{0}$. Thus (a) is established. According to standard duality theory [37], given the subgradient $\mathbf{W}\hat{\boldsymbol{u}} \in \mathbb{R}^n$, any optimum $\tilde{\boldsymbol{x}} \in \mathbb{R}^n$ of (5) must satisfy the complementary slackness condition $\hat{\boldsymbol{u}}^T\mathbf{W}\tilde{\boldsymbol{x}} = \|\mathbf{W}\tilde{\boldsymbol{x}}\|_1$. For all i such that $|\hat{u}_i| < 1$ and $w_i \neq 0$, this condition holds if and only if $\tilde{x}_i = 0$. Further, if $|\hat{u}_i| < 1$ for all $i \in \mathcal{V}$ and $\mathbf{A}_{\mathcal{V}^c}$ is full column rank, then (5) is strictly convex because its solution is restricted to the form $(\hat{\boldsymbol{x}}_{\mathcal{V}^c}, \mathbf{0})$, and so its optimum is unique. Thus (b) applies.

Appendix B
Proof of Lemma 2

Proof: Define an n-dimensional vector $\boldsymbol{x}^\dagger$ as

$$\begin{cases} \boldsymbol{x}^\dagger_{\mathcal{S}} = \boldsymbol{x}^*_{\mathcal{S}} + \mathbf{A}^+_{\mathcal{S}}\boldsymbol{z} - mh(\mathbf{A}^T_{\mathcal{S}}\mathbf{A}_{\mathcal{S}})^{-1}\mathbf{W}_{\mathcal{S}}\boldsymbol{u}_{\mathcal{S}} \\ \boldsymbol{x}^\dagger_{\mathcal{S}^c} = \mathbf{0} \end{cases}. \quad (39)$$

If the conditions in (9) are satisfied, we will prove that vector $\boldsymbol{x}^\dagger$ is the unique minimum of (5). According to (9b), we have

$$\text{sign}(\boldsymbol{x}^\dagger_{\mathcal{S}}) = \text{sign}(\boldsymbol{x}^*_{\mathcal{S}}). \quad (40)$$

Further, utilizing the equality $\mathbf{W}_{\mathcal{S}}\boldsymbol{u}_{\mathcal{S}} = \mathbf{W}_{\mathcal{S}} \times \text{sign}(\boldsymbol{x}^*_{\mathcal{S}})$, it follows

$$\boldsymbol{x}^\dagger_{\mathcal{S}} = \boldsymbol{x}^*_{\mathcal{S}} + \mathbf{A}^+_{\mathcal{S}}\boldsymbol{z} - mh(\mathbf{A}^T_{\mathcal{S}}\mathbf{A}_{\mathcal{S}})^{-1}\mathbf{W}_{\mathcal{S}} \times \text{sign}(\boldsymbol{x}^\dagger_{\mathcal{S}}). \quad (41)$$

Thus letting $\mathcal{V} = \mathcal{S}^c$ and $\mathcal{V}^c = \mathcal{S}$, we have that $\boldsymbol{x}^\dagger_{\mathcal{S}}$ satisfies the first condition in (8) through multiplying both sides in (41) by $\mathbf{A}^T_{\mathcal{S}}\mathbf{A}_{\mathcal{S}}$. Further, substituting (41) into the term $(\boldsymbol{y} - \mathbf{A}_{\mathcal{S}}\boldsymbol{x}^\dagger_{\mathcal{S}})$, we have that $\forall i \in \mathcal{V}$

$$\left|\frac{\mathbf{A}^T_i}{m}(\boldsymbol{y} - \mathbf{A}_{\mathcal{S}}\boldsymbol{x}^\dagger_{\mathcal{S}})\right| < hw_i, \quad (42)$$

where the inequality in (42) utilizes the fact that $\mathbf{W}_{\mathcal{S}} \times \text{sign}(\boldsymbol{x}^\dagger_{\mathcal{S}}) = \mathbf{W}_{\mathcal{S}}\boldsymbol{u}_{\mathcal{S}}$ and follows from the condition in (9a). Hence, according to the sufficient conditions in (8), $\boldsymbol{x}^\dagger$ is the unique minimum of (5), i.e., $\hat{\boldsymbol{x}} = \boldsymbol{x}^\dagger$. Based on (39) and (40), $\text{sign}(\hat{\boldsymbol{x}}) = \text{sign}(\boldsymbol{x}^*)$ holds.

Appendix C
Proof of Proposition 1

Proof: Recall that $\mathbf{W}\hat{\boldsymbol{u}}$ is the subdifferential of $\|\mathbf{W}\hat{\boldsymbol{x}}\|_1$ as defined in Equation (6). If there is a nonempty set $\mathcal{T}_e = \{i : w_i = 0 \text{ and } i \notin \mathcal{S}\}$, we have that $\mathbf{W}_{\mathcal{T}_e}\hat{\boldsymbol{u}}_{\mathcal{T}_e} = \mathbf{0}_{|\mathcal{T}_e|\times 1}$ holds, where $\mathbf{0}_{|\mathcal{T}_e|\times 1}$ represents the vector of zeros with the size

$|\mathcal{T}_e| \times 1$. According to Lemma 1, if the support of $\boldsymbol{x}^*$ can be recovered exactly by solving (5), i.e., $\text{sign}(\hat{\boldsymbol{x}}) = \text{sign}(\boldsymbol{x}^*)$, then it is necessary that $\exists\, \mathbf{W}\hat{\boldsymbol{u}} \in \partial\|\mathbf{W}\hat{\boldsymbol{x}}\|_1$, such that

$$[\mathbf{A}_\mathcal{S}, \mathbf{A}_{\mathcal{T}_e}]^T \left(\boldsymbol{y} - [\mathbf{A}_\mathcal{S}, \mathbf{A}_{\mathcal{T}_e}]\begin{bmatrix}\hat{\boldsymbol{x}}_\mathcal{S}\\ \mathbf{0}_{|\mathcal{T}_e|\times 1}\end{bmatrix}\right) = mh\begin{bmatrix}\mathbf{W}_\mathcal{S}\hat{\boldsymbol{u}}_\mathcal{S}\\ \mathbf{0}_{|\mathcal{T}_e|\times 1}\end{bmatrix}. \quad (43)$$

Defining $\mathbf{B} \triangleq [\mathbf{A}_\mathcal{S}, \mathbf{A}_{\mathcal{T}_e}]^T[\mathbf{A}_\mathcal{S}, \mathbf{A}_{\mathcal{T}_e}]$ and $\boldsymbol{p} \triangleq ([\mathbf{A}_\mathcal{S}, \mathbf{A}_{\mathcal{T}_e}]^T \boldsymbol{z} - mh[\begin{smallmatrix}\mathbf{W}_\mathcal{S}\hat{\boldsymbol{u}}_\mathcal{S}\\ \mathbf{0}_{|\mathcal{T}_e|\times 1}\end{smallmatrix}])$, (43) can be rewritten as

$$\mathbf{B}\begin{bmatrix}\hat{\boldsymbol{x}}_\mathcal{S} - \boldsymbol{x}^*_\mathcal{S}\\ \mathbf{0}_{|\mathcal{T}_e|\times 1}\end{bmatrix} = \boldsymbol{p}. \quad (44)$$

We first consider the case that $|\mathcal{T}_e| \leq m - k$. Since $\mathbf{A}$ is a Gaussian random matrix, its submatrix $[\mathbf{A}_\mathcal{S}, \mathbf{A}_{\mathcal{T}_e}]$ is full column rank with probability one. Thus $\mathbf{B} \in R^{(k+|\mathcal{T}_e|)\times(k+|\mathcal{T}_e|)}$ is full rank with probability one [38] and its columns are linearly independent. Moreover, given that the noise vector $\boldsymbol{z}$ is selected randomly from a Gaussian distribution, $\boldsymbol{p}$ is also randomly distributed, which is linearly independent with any $k + |\mathcal{T}_e| - 1$ columns of $\mathbf{B}$ with a probability of one [39]. Therefore, the probability that (44) holds is zero.

Further, we consider the case that $|\mathcal{T}_e| > m - k$, where $[\mathbf{A}_\mathcal{S}, \mathbf{A}_{\mathcal{T}_e}]$ is full row rank with probability one. According to matrix theory [38], we have that

$$\text{rank}(\mathbf{B}) = \text{rank}\left([\mathbf{A}_\mathcal{S}, \mathbf{A}_{\mathcal{T}_e}]\right) = m, \quad (45)$$

where $\text{rank}(\mathbf{B})$ represents the rank of matrix $\mathbf{B}$. Because of the random property of $\boldsymbol{p}$, there are m columns of $\mathbf{B}$ that are linearly independent with $\boldsymbol{p}$. Therefore, it follows that

$$\text{rank}([\mathbf{B}, \boldsymbol{p}]) = m + 1 \neq \text{rank}(\mathbf{B}) \quad (46)$$

According to matrix theory [38], we have that

$$\mathbf{B}\boldsymbol{x} = \boldsymbol{p} \quad (47)$$

is an inconsistent equation. Hence, (44) holds with probability zero. Based on the above analysis, the conclusion in Proposition 1 holds.

Appendix D
Proof of Lemma 3

Proof: There always exists a vector $\hat{\boldsymbol{x}}$ that is the global minimum of (5) and satisfies (7). Recall that $\mathcal{S} = \text{supp}(\boldsymbol{x}^*)$ and $\mathcal{T} = \{i : w_i = 0\}$. Suppose $\mathcal{V}$ be any subset of $\{\mathcal{S} \cup \mathcal{T}\}^c$ with $|\mathcal{V}| \geq n - m$. According to Lemma 1, $\hat{\boldsymbol{x}}$ is the unique minimum of (5) if

$$\left|\frac{\mathbf{A}_i^T}{m}(\boldsymbol{y} - \mathbf{A}_{\mathcal{V}^c}\hat{\boldsymbol{x}}_{\mathcal{V}^c})\right| < hw_i \quad \textit{for } i \in \mathcal{V}. \quad (48)$$

Moreover, according to Equation (8) and Footnote 1, $\hat{\boldsymbol{x}}$ can be determined as

$$\begin{cases}\hat{\boldsymbol{x}}_{\mathcal{V}^c} = \boldsymbol{x}^*_{\mathcal{V}^c} + \mathbf{A}^+_{\mathcal{V}^c}\boldsymbol{z} - mh(\mathbf{A}^T_{\mathcal{V}^c}\mathbf{A}_{\mathcal{V}^c})^{-1}\mathbf{W}_{\mathcal{V}^c}\hat{\boldsymbol{u}}_{\mathcal{V}^c}\\ \hat{\boldsymbol{x}}_\mathcal{V} = \mathbf{0}\end{cases}. \quad (49)$$

Substituting (49) into (48), we obtain the condition in (10a). Obviously, we have that $\mathcal{S} \subseteq \mathcal{V}^c$. Therefore, if the condition in (10b) is satisfied, according to Equation (49), it follows that

$$\text{sign}(\hat{\boldsymbol{x}}_\mathcal{S}) = \text{sign}(\boldsymbol{x}^*_\mathcal{S}). \quad (50)$$

Based on Equations (49) and (50), we have that $\text{sign}(\hat{\boldsymbol{x}}_{\mathcal{S}\cup\mathcal{V}}) = \text{sign}({\boldsymbol{x}^*}_{\mathcal{S}\cup\mathcal{V}})$ so that the conclusion in Lemma 3 holds.

Appendix E
Proof of Theorem 1

In this section, the proof of Theorem 1 uses the techniques in [10], with appropriate modification to account for the weighted ℓ_1 norm that replaces the ℓ_1 norm.

Proof: Based on Lemma 3, we conclude that optimization problem (5) can recover the support of $\boldsymbol{x}^*_{\mathcal{S}\cup\mathcal{V}}$ exactly, provided the events in (10) are satisfied. Therefore, we first derive a precise condition under which (10a) is satisfied with high probability. Further, by bounding the quantity $(\mathbf{A}^+_{\mathcal{V}^c}\boldsymbol{z} - mh(\mathbf{A}^T_{\mathcal{V}^c}\mathbf{A}_{\mathcal{V}^c})^{-1}\mathbf{W}_{\mathcal{V}^c}\hat{\boldsymbol{u}}_{\mathcal{V}^c})$, another condition can be obtained to guarantee $\text{sign}(\hat{\boldsymbol{x}}_\mathcal{S}) = \text{sign}(\boldsymbol{x}^*_\mathcal{S})$ holds with high probability. Then, according to Lemma 3, the support of $\boldsymbol{x}^*_{\mathcal{S}\cup\mathcal{V}}$ is, with high probability, recovered exactly from the solution of (5).

For (10a), conditioned on $\mathbf{A}_{\mathcal{V}^c}$ and noise $\boldsymbol{z}$, we have that

$$\Gamma_i \triangleq \frac{\mathbf{A}_i^T}{mh}\left[(\mathbf{I} - \mathbf{A}_{\mathcal{V}^c}\mathbf{A}^+_{\mathcal{V}^c})\boldsymbol{z} + mh\mathbf{A}_{\mathcal{V}^c}(\mathbf{A}^T_{\mathcal{V}^c}\mathbf{A}_{\mathcal{V}^c})^{-1}\mathbf{W}_{\mathcal{V}^c}\hat{\boldsymbol{u}}_{\mathcal{V}^c}\right] \quad (51)$$

is zero-mean Gaussian with variance at most

$$\text{var}(\Gamma_i|\mathbf{A}_{\mathcal{V}^c},\ \boldsymbol{z}) \leq \left\|\mathbf{A}_{\mathcal{V}^c}\left(\mathbf{A}^T_{\mathcal{V}^c}\mathbf{A}_{\mathcal{V}^c}\right)^{-1}\mathbf{W}_{\mathcal{V}^c}\hat{\boldsymbol{u}}_{\mathcal{V}^c} + (\mathbf{I} - \mathbf{A}_{\mathcal{V}^c}\mathbf{A}^+_{\mathcal{V}^c})\frac{\boldsymbol{z}}{mh}\right\|_2^2. \quad (52)$$

Further, because

$$\left\langle \mathbf{A}_{\mathcal{V}^c}(\mathbf{A}^T_{\mathcal{V}^c}\mathbf{A}_{\mathcal{V}^c})^{-1}\boldsymbol{a},\ (\mathbf{I} - \mathbf{A}_{\mathcal{V}^c}\mathbf{A}^+_{\mathcal{V}^c})\boldsymbol{b}\right\rangle = 0 \quad (53)$$

holds for all vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, we have

$$\left\langle \mathbf{A}_{\mathcal{V}^c}(\mathbf{A}^T_{\mathcal{V}^c}\mathbf{A}_{\mathcal{V}^c})^{-1}\mathbf{W}_{\mathcal{V}^c}\hat{\boldsymbol{u}}_{\mathcal{V}^c},\ (\mathbf{I} - \mathbf{A}_{\mathcal{V}^c}\mathbf{A}^+_{\mathcal{V}^c})\frac{\boldsymbol{z}}{mh}\right\rangle = 0. \quad (54)$$

It follows that

$$\text{var}(\Gamma_i|\mathbf{A}_{\mathcal{V}^c},\ \boldsymbol{z}) \leq \left\|\mathbf{A}_{\mathcal{V}^c}\left(\mathbf{A}^T_{\mathcal{V}^c}\mathbf{A}_{\mathcal{V}^c}\right)^{-1}\mathbf{W}_{\mathcal{V}^c}\hat{\boldsymbol{u}}_{\mathcal{V}^c}\right\|_2^2 + \left\|(\mathbf{I} - \mathbf{A}_{\mathcal{V}^c}\mathbf{A}^+_{\mathcal{V}^c})\frac{\boldsymbol{z}}{mh}\right\|_2^2. \quad (55)$$

For the first term in equation (55), we have

$$\begin{aligned}&\left\|\mathbf{A}_{\mathcal{V}^c}\left(\mathbf{A}^T_{\mathcal{V}^c}\mathbf{A}_{\mathcal{V}^c}\right)^{-1}\mathbf{W}_{\mathcal{V}^c}\hat{\boldsymbol{u}}_{\mathcal{V}^c}\right\|_2^2\\ &\quad= \frac{1}{m}\hat{\boldsymbol{u}}^T_{\mathcal{V}^c}\mathbf{W}_{\mathcal{V}^c}\left(\frac{\mathbf{A}^T_{\mathcal{V}^c}\mathbf{A}_{\mathcal{V}^c}}{m}\right)^{-1}\mathbf{W}_{\mathcal{V}^c}\hat{\boldsymbol{u}}_{\mathcal{V}^c}\\ &\quad\leq \frac{1}{m}\left\|\hat{\boldsymbol{u}}^T_{\mathcal{V}^c}\mathbf{W}_{\mathcal{V}^c}\right\|_2\left\|\left(\frac{\mathbf{A}^T_{\mathcal{V}^c}\mathbf{A}_{\mathcal{V}^c}}{m}\right)^{-1}\mathbf{W}_{\mathcal{V}^c}\hat{\boldsymbol{u}}_{\mathcal{V}^c}\right\|_2\\ &\quad\leq \frac{1}{m}\left\|\hat{\boldsymbol{u}}^T_{\mathcal{V}^c}\mathbf{W}_{\mathcal{V}^c}\right\|_2^2\left|\left|\left|\left(\frac{\mathbf{A}^T_{\mathcal{V}^c}\mathbf{A}_{\mathcal{V}^c}}{m}\right)^{-1}\right|\right|\right|_2,\end{aligned} \quad (56)$$

where the first inequality follows from the Cauchy-Schwartz inequality, $|||\cdot|||_2$ represents the spectral norm and the second inequality follows from the definition of matrix norm.

At the same time, we have

$$\begin{aligned}&\left|\left|\left|\left(\frac{\mathbf{A}_{\mathcal{V}^c}^T\mathbf{A}_{\mathcal{V}^c}}{m}\right)^{-1}\right|\right|\right|_2\\&\leq |||\mathbf{I}^{-1}|||_2+\left|\left|\left|\left(\frac{\mathbf{A}_{\mathcal{V}^c}^T\mathbf{A}_{\mathcal{V}^c}}{m}\right)^{-1}-\mathbf{I}^{-1}\right|\right|\right|_2. \end{aligned}\tag{57}$$

Applying Lemma 9 in [10], it follows that the event

$$\left|\left|\left|\left(\frac{\mathbf{A}_{\mathcal{V}^c}^T\mathbf{A}_{\mathcal{V}^c}}{m}\right)^{-1}\right|\right|\right|_2\leq 1+8\sqrt{\frac{n-|\mathcal{V}|}{m}}\tag{58}$$

is satisfied with probability greater than $1-2\exp(-\frac{n-|\mathcal{V}|}{2})$.

Recall that the definition of vector $\hat{\boldsymbol{u}}_{\mathcal{V}^c}$. We have

$$\left\|\hat{\boldsymbol{u}}_{\mathcal{V}^c}^T\mathbf{W}_{\mathcal{V}^c}\right\|_2^2\leq\sum_{i\in\mathcal{V}^c}W_{\mathcal{V}^c,i}^2=k\xi,\tag{59}$$

where $\xi=\frac{\sum_{i\in\mathcal{V}^c}W_{\mathcal{V}^c,i}^2}{k}$ and $W_{\mathcal{V}^c,i}$ represents the i-th diagonal element in the matrix $\mathbf{W}_{\mathcal{V}^c}$. Consequently, combining Equations (56), (58) and (59), we obtain that event

$$\left\|\mathbf{A}_{\mathcal{V}^c}(\mathbf{A}_{\mathcal{V}^c}^T\mathbf{A}_{\mathcal{V}^c})^{-1}\mathbf{W}_{\mathcal{V}^c}\hat{\boldsymbol{u}}_{\mathcal{V}^c}\right\|_2^2\leq\left(1+8\sqrt{\frac{n-|\mathcal{V}|}{m}}\right)\frac{\xi k}{m}\tag{60}$$

is satisfied with probability greater than $1-2\exp(-\frac{n-|\mathcal{V}|}{2})$.

Turning to the second term in (55), we have

$$\left\|(\mathbf{I}-\mathbf{A}_{\mathcal{V}^c}\mathbf{A}_{\mathcal{V}^c}^+)\frac{\boldsymbol{z}}{mh}\right\|_2^2\leq\frac{1}{mh^2}\frac{\|\boldsymbol{z}\|_2^2}{m}\tag{61}$$

since $(\mathbf{I}-\mathbf{A}_{\mathcal{V}^c}\mathbf{A}_{\mathcal{V}^c}^+)$ is an orthogonal projection matrix. On the other hand, $\|\boldsymbol{z}\|_2^2/\sigma_z^2$ is a χ^2 random variable with m degrees of freedom. Thus, applying the tail bounds for χ^2 random variable (see Appendix J in [10]), we have that for all $\epsilon\in(0,1/2)$,

$$\mathbb{P}\left[\left\|(\mathbf{I}-\mathbf{A}_{\mathcal{V}^c}\mathbf{A}_{\mathcal{V}^c}^+)\frac{\boldsymbol{z}}{mh}\right\|_2^2\geq(1+\epsilon)\frac{\sigma_z^2}{mh^2}\right]\leq\exp\left(-\frac{3m\epsilon^2}{16}\right).\tag{62}$$

Combining (55), (58) and (62), we have that event

$$\begin{aligned}&\mathrm{var}(\Gamma_i)\geq\Pi\\&\triangleq\left(1+\max\left\{\epsilon,\,8\sqrt{\frac{n-|\mathcal{V}|}{m}}\right\}\right)\left(\frac{\xi k}{m}+\frac{\sigma_z^2}{mh^2}\right)\end{aligned}\tag{63}$$

is satisfied with probability less than $4\exp(-c_1\min\{m\epsilon^2,n-|\mathcal{V}|\})$ for some $c_1>0$.

Consequently, applying the standard Gaussian tail bounds (see Appendix A in [10]), we have

$$\begin{aligned}&\mathbb{P}\left[\max_{i\in\mathcal{V}}|\Gamma_i|\geq w_i\right]\\&=\mathbb{P}\left[\max_{i\in\mathcal{V}}|\Gamma_i|\geq w_i|\mathrm{var}(\Gamma_i)<\Pi\right]\mathbb{P}\left[\mathrm{var}(\Gamma_i)<\Pi\right]\\&\quad+\mathbb{P}\left[\max_{i\in\mathcal{V}}|\Gamma_i|\geq w_i|\mathrm{var}(\Gamma_i)\geq\Pi\right]\mathbb{P}[\mathrm{var}(\Gamma_i)\geq\Pi]\\&\leq\mathbb{P}\left[\max_{i\in\mathcal{V}}|\Gamma_i|\geq w_i|\mathrm{var}(\Gamma_i)<\Pi\right]+\mathbb{P}[\mathrm{var}(\Gamma_i)\geq\Pi]\\&\leq 2|\mathcal{V}|\exp\left(-\frac{w_i^2}{2\Pi}\right)+4\exp(-c_1\min\{m\epsilon^2,n-|\mathcal{V}|\}).\end{aligned}\tag{64}$$

Further, the exponential term in (64) is decaying, provided

$$m>2\eta k\log|\mathcal{V}|(1+\epsilon')\left(1+\frac{\sigma_z^2}{h^2\xi k}\right),\tag{65}$$

where $\eta=\max_{i\in\mathcal{V}}\{\frac{\xi}{w_i^2}\}$ and $\epsilon'=\max\{\epsilon,8\sqrt{\frac{n-|\mathcal{V}|}{m}}\}$. Combining with the rank condition on $\mathbf{A}_{\mathcal{V}^c}$, i.e., $m\geqslant|\mathcal{V}^c|$, it follows that the condition (12) holds.

For (10b), we establish a bound on $(\mathbf{A}_{\mathcal{V}^c}^+\boldsymbol{z}-mh(\mathbf{A}_{\mathcal{V}^c}^T\mathbf{A}_{\mathcal{V}^c})^{-1}\mathbf{W}_{\mathcal{V}^c}\hat{\boldsymbol{u}}_{\mathcal{V}^c})$ and applying the triangle inequality, we have

$$\begin{aligned}&\left\|(\mathbf{A}_{\mathcal{V}^c}^+\boldsymbol{z}-mh(\mathbf{A}_{\mathcal{V}^c}^T\mathbf{A}_{\mathcal{V}^c})^{-1}\mathbf{W}_{\mathcal{V}^c}\hat{\boldsymbol{u}}_{\mathcal{V}^c})\right\|_\infty\\&\leqslant\left\|\mathbf{A}_{\mathcal{V}^c}^+\boldsymbol{z}\right\|_\infty+mh\left\|(\mathbf{A}_{\mathcal{V}^c}^T\mathbf{A}_{\mathcal{V}^c})^{-1}\mathbf{W}_{\mathcal{V}^c}\hat{\boldsymbol{u}}_{\mathcal{V}^c}\right\|_\infty.\end{aligned}\tag{66}$$

For the first term in (66), conditioned on $\mathbf{A}_{\mathcal{V}^c}$, random vector $\mathbf{A}_{\mathcal{V}^c}^+\boldsymbol{z}$ is zero-mean Gaussian with variance at most

$$\frac{\sigma_z^2}{m}\left|\left|\left|\left(\frac{\mathbf{A}_{\mathcal{V}^c}^T\mathbf{A}_{\mathcal{V}^c}}{m}\right)^{-1}\right|\right|\right|_2\tag{67}$$

As analyzed in [10],

$$\mathbb{P}\left[\Upsilon\triangleq\frac{\sigma_z^2}{m}\left|\left|\left|\left(\frac{\mathbf{A}_{\mathcal{V}^c}^T\mathbf{A}_{\mathcal{V}^c}}{m}\right)^{-1}\right|\right|\right|_2\geq\frac{9\sigma_z^2}{m}\right]\leq 2\exp\left(-\frac{m}{2}\right)\tag{68}$$

By the total probability rule, it follows

$$\begin{aligned}&\mathbb{P}\left[\left\|\mathbf{A}_{\mathcal{V}^c}^+\boldsymbol{z}\right\|_\infty>t\right]\\&\leq\mathbb{P}\left[\left\|\mathbf{A}_{\mathcal{V}^c}^+\boldsymbol{z}\right\|_\infty>t|\Upsilon<\frac{9\sigma_z^2}{m}\right]+\mathbb{P}\left(\Upsilon\geq\frac{9\sigma_z^2}{m}\right)\end{aligned}\tag{69}$$

Using the standard Gaussian tail bounds and the conclusion in (42) of [10], it follows

$$\mathbb{P}\left[\left\|\mathbf{A}_{\mathcal{V}^c}^+\boldsymbol{z}\right\|_\infty\geq 20\sqrt{\frac{\sigma_z^2\log(n-|\mathcal{V}|)}{m}}\right]\leq 4\exp(-c_1 m)\tag{70}$$

holds for some $c_1>0$.

For the second term in (66), we have

$$m\left\|\left(\mathbf{A}_{\mathcal{V}^c}^T\mathbf{A}_{\mathcal{V}^c}\right)^{-1}\mathbf{W}_{\mathcal{V}^c}\hat{\boldsymbol{u}}_{\mathcal{V}^c}\right\|_\infty \leq \left\|\left[\left(\frac{\mathbf{A}_{\mathcal{V}^c}^T\mathbf{A}_{\mathcal{V}^c}}{m}\right)^{-1}-\mathbf{I}^{-1}\right]\mathbf{W}_{\mathcal{V}^c}\hat{\boldsymbol{u}}_{\mathcal{V}^c}\right\|_\infty+\left\|\mathbf{I}^{-1}\mathbf{W}_{\mathcal{V}^c}\hat{\boldsymbol{u}}_{\mathcal{V}^c}\right\|_\infty. \tag{71}$$

According to Lemma 5 in [10], we have

$$\mathbb{P}\left\{\left\|\left[\left(\frac{\mathbf{A}_{\mathcal{V}^c}^T\mathbf{A}_{\mathcal{V}^c}}{m}\right)^{-1}-\mathbf{I}^{-1}\right]\mathbf{W}_{\mathcal{V}^c}\hat{\boldsymbol{u}}_{\mathcal{V}^c}\right\|_\infty \geq c_2\|\mathbf{W}_{\mathcal{V}^c}\hat{\boldsymbol{u}}_{\mathcal{V}^c}\|_\infty\right\} \leq 4\exp(-c_3\min\{n-|\mathcal{V}|,\log(|\mathcal{V}|)\}) \tag{72}$$

holds for some c_2, $c_3 > 0$. Therefore, it follows

$$\mathbb{P}\left\{mh\left\|\left(\mathbf{A}_{\mathcal{V}^c}^T\mathbf{A}_{\mathcal{V}^c}\right)^{-1}\mathbf{W}_{\mathcal{V}^c}\hat{\boldsymbol{u}}_{\mathcal{V}^c}\right\|_\infty \geqslant c_4h\|\mathbf{W}_{\mathcal{V}^c}\hat{\boldsymbol{u}}_{\mathcal{V}^c}\|_\infty\right\} \leq 4\exp(-c_3\min\{n-|\mathcal{V}|,\log(|\mathcal{V}|)\}) \tag{73}$$

holds for some $c_3, c_4 > 0$. Combining (66), (70) and (73), we have that event

$$\left\|(\mathbf{A}_{\mathcal{V}^c}^+\boldsymbol{z}-mh(\mathbf{A}_{\mathcal{V}^c}^T\mathbf{A}_{\mathcal{V}^c})^{-1}\mathbf{W}_{\mathcal{V}^c}\hat{\boldsymbol{u}}_{\mathcal{V}^c})\right\|_\infty \leq c_4h\|\mathbf{W}_{\mathcal{V}^c}\hat{\boldsymbol{u}}_{\mathcal{V}^c}\|_\infty+20\sqrt{\frac{\sigma_z^2\log(n-|\mathcal{V}|)}{m}} \triangleq g(h) \tag{74}$$

is satisfied with probability greater than $1-c_4^{'}\exp(-c_3^{'}\min\{n-|\mathcal{V}|,\log(|\mathcal{V}|)\})$. Therefore, if $\forall i\in\mathcal{S}\ \ |x_i^*|>g(h)$ holds, we have that (10b) holds with high probability. Combining the probabilities that the two events in (10) are satisfied, it follows from Lemma 3 that (a) in Theorem 1 holds.

Further, if $|x_i^*|>2g(h)$, for $i\in\mathcal{S}$, it follows from (49) and (74) that for $i\in\mathcal{S}$,

$$|\hat{x}_i| \geq |x_i^*| - \left\|(\mathbf{A}_{\mathcal{V}^c}^+\boldsymbol{z}-mh(\mathbf{A}_{\mathcal{V}^c}^T\mathbf{A}_{\mathcal{V}^c})^{-1}\mathbf{W}_{\mathcal{V}^c}\hat{\boldsymbol{u}}_{\mathcal{V}^c})\right\|_\infty > g(h) \tag{75}$$

and for $i\in\mathcal{V}^c\setminus\mathcal{S}$,

$$\begin{aligned}|\hat{x}_i| &\leq |x_i^*| + \left\|(\mathbf{A}_{\mathcal{V}^c}^+\boldsymbol{z}-mh(\mathbf{A}_{\mathcal{V}^c}^T\mathbf{A}_{\mathcal{V}^c})^{-1}\mathbf{W}_{\mathcal{V}^c}\hat{\boldsymbol{u}}_{\mathcal{V}^c})\right\|_\infty \\ &= \left\|(\mathbf{A}_{\mathcal{V}^c}^+\boldsymbol{z}-mh(\mathbf{A}_{\mathcal{V}^c}^T\mathbf{A}_{\mathcal{V}^c})^{-1}\mathbf{W}_{\mathcal{V}^c}\hat{\boldsymbol{u}}_{\mathcal{V}^c})\right\|_\infty \\ &\leq g(h)\end{aligned} \tag{76}$$

hold with probability greater than 1-$c_2\exp(-c_3\ \min\{m\epsilon^2, n-|\mathcal{V}|,\log(|\mathcal{V}|)\})$ for some positive constants c_2 and c_3. Therefore, applying Equation (14), the zero entries of $\boldsymbol{x}^*$ in indices $\mathcal{V}^c\setminus\mathcal{S}$ can, with high probability, be recovered exactly. Then (b) in Theorem 1 holds.

Appendix F
Proof of Proposition 2

Proof: If most of the support of $\hat{\boldsymbol{e}}^{(\mathrm{MAP})}$ belong to the support of $\boldsymbol{e}^*$, i.e.,

$$\frac{\left|\mathrm{supp}(\hat{\boldsymbol{e}}^{(\mathrm{MAP})})\cap\mathrm{supp}(\boldsymbol{e}^*)\right|}{\left|\mathrm{supp}(\hat{\boldsymbol{e}}^{(\mathrm{MAP})})\right|} > \frac{1}{2}, \tag{77}$$

we have

$$\left\|\hat{\boldsymbol{e}}^{(\mathrm{MAP})}\oplus\hat{\boldsymbol{x}}\oplus\boldsymbol{x}^*\right\|_0 = \left\|\hat{\boldsymbol{e}}^{(\mathrm{MAP})}\oplus\boldsymbol{e}^*\right\|_0 = |\mathrm{supp}(\hat{\boldsymbol{e}}^{(\mathrm{MAP})})\cap\mathrm{supp}(\boldsymbol{e}^*)^c| + |\mathrm{supp}(\boldsymbol{e}^*)\cap\mathrm{supp}(\hat{\boldsymbol{e}}^{(\mathrm{MAP})})^c|. \tag{78}$$

Because we have $|\mathrm{supp}(\hat{\boldsymbol{e}}^{(\mathrm{MAP})})\cap\mathrm{supp}(\boldsymbol{e}^*)^c| + |\mathrm{supp}(\hat{\boldsymbol{e}}^{(\mathrm{MAP})})\cap\mathrm{supp}(\boldsymbol{e}^*)| = |\mathrm{supp}(\hat{\boldsymbol{e}}^{(\mathrm{MAP})})|$, it follows that

$$|\mathrm{supp}(\hat{\boldsymbol{e}}^{(\mathrm{MAP})})\cap\mathrm{supp}(\boldsymbol{e}^*)^c| < |\mathrm{supp}(\hat{\boldsymbol{e}}^{(\mathrm{MAP})})\cap\mathrm{supp}(\boldsymbol{e}^*)| \tag{79}$$

by applying (77).

Hence, combining (78) and (79), we have

$$\begin{aligned}&\left\|\hat{\boldsymbol{e}}^{(\mathrm{MAP})}\oplus\hat{\boldsymbol{x}}\oplus\boldsymbol{x}^*\right\|_0 \\ &< |\mathrm{supp}(\hat{\boldsymbol{e}}^{(\mathrm{MAP})})\cap\mathrm{supp}(\boldsymbol{e}^*)| + |\mathrm{supp}(\boldsymbol{e}^*)\cap\mathrm{supp}(\hat{\boldsymbol{e}}^{(\mathrm{MAP})})^c| \\ &= |\mathrm{supp}(\boldsymbol{e}^*)| = \|\boldsymbol{e}^*\|_0\,.\end{aligned} \tag{80}$$

References

[1] J. Zhang, U. Mitra, K.-W. Huang, and N. Michelusi, "Support recovery from noisy random measurements via weighted ℓ_1 minimization," in *Proc. IEEE Int. Symp. Inf. Theory*, 2016, pp. 1426–1430.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[3] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[4] W. Wu, S. Nagarajan, and Z. Chen, "Bayesian machine learning for EEG/MEG signal processing," *IEEE Signal Process. Mag.*, vol. 33, no. 1, pp. 14–36, Jan. 2016.

[5] J. Fan and R. Li, "Variable selection via non-concave penalized likelihood and its oracle properties," *J. Amer. Statist. Assoc.*, vol. 96, no. 456, pp. 1348–1360, 2001.

[6] M. Levorato, U. Mitra, and M. Zorzi, "Cognitive interference management in retransmission-based wireless networks," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 3023–3046, May 2012.

[7] N. Michelusi and U. Mitra, "Cross-layer estimation and control for cognitive radio: Exploiting sparse network dynamics," *IEEE Trans. Cogn. Commun. Netw.*, vol. 1, no. 1, pp. 128–145, Mar. 2015.

[8] Y. Li *et al.*, "Multimodal BCIs: Target detection, multi-dimensional control, and awareness evaluation in patients with disorder of consciousness," *Proc. IEEE*, vol. 104, no. 2, pp. 332–352, Feb. 2016.

[9] K. Liu, Z. L. Yu, W. Wu, Z. Gu, Y. Li, and S. Nagarajan, "Bayesian electromagnetic spatio-temporal imaging of extended sources with Markov random field and temporal basis expansion," *NeuroImage*, vol. 139, pp. 385–404, 2016.

[10] M. J. Wainwright, "Sharp thresholds for high-dimensional and noisy sparsity recovery using-constrained quadratic programming (Lasso)," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2183–2202, May 2009.

[11] C. Dossal, M.-L. Chabanol, G. Peyré, and J. Fadili, "Sharp support recovery from noisy random measurements by ℓ_1-minimization," *Appl. Comput. Harmon. Anal.*, vol. 33, no. 1, pp. 24–43, 2012.

[12] J. Wang, "Support recovery with orthogonal matching pursuit in the presence of noise," *IEEE Trans. Signal Process.*, vol. 63, no. 21, pp. 5868–5877, Nov. 2015.

[13] J. Wen, Z. Zhou, J. Wang, X. Tang, and Q. Mo, "A sharp condition for exact support recovery with orthogonal matching pursuit," *IEEE Trans. Signal Process.*, vol. 65, no. 6, pp. 1370–1382, Mar. 2017.

[14] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, 2001.

[15] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[16] W. Lu and N. Vaswani, "Modified basis pursuit denoising (modified-BPDN) for noisy compressive sensing with partially known support," in *Proc. 35th Int. Conf. Acoust., Speech, Signal Process.*, 2010, pp. 3926–3929.

[17] Y. Wang and W. Yin, "Sparse signal reconstruction via iterative support detection," *SIAM J Imag. Sci.*, vol. 3, no. 3, pp. 462–491, 2010.

[18] N. Vaswani and W. Lu, "Modified-CS: Modifying compressive sensing for problems with partially known support," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4595–4607, Sep. 2010.

[19] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Anal. Appl.*, vol. 14, no. 5–6, pp. 877–905, 2008.

[20] H. Mansour and Ö. Yilmaz, "Support driven reweighted ℓ_1 minimization," in *Proc. 37th Int. Conf. Acoust., Speech, Signal Process.*, 2012, pp. 3309–3312.

[21] C. Herzet, C. Soussen, J. Idier, and R. Gribonval, "Exact recovery conditions for sparse representations with partial support information," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7509–7524, Nov. 2013.

[22] M. Khajehnejad, W. Xu, S. Avestimehr, and B. Hassibi, "Analyzing weighted ℓ_1 minimization for sparse recovery with nonuniform sparse models," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 1985–2001, May 2011.

[23] H. Mansour and R. Saab, "Recovery analysis for weighted ℓ_1-minimization using the null space property," *Appl. Comput. Harmon. Anal.*, vol. 43, no. 1, pp. 23–38, 2017.

[24] L. Jacques, "A short note on compressed sensing with partially known signal support," *Signal Process.*, vol. 90, no. 12, pp. 3308–3312, 2010.

[25] M. Friedlander, H. Mansour, R. Saab, and Ö. Yilmaz, "Recovering compressively sampled signals using partial support information," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 1122–1134, Feb. 2012.

[26] H. Mansour and Ö. Yilmaz, "Weighted ℓ_1 minimization with multiple weighting sets," *Proc. SPIE*, vol. 8138, 2011, Art. no. 813 809.

[27] N. Ghadermarzy and H. Mansour, "Non-convex compressed sensing using partial support information," *Sampling Theory Signal Image Process.*, vol. 13, no. 3, pp. 249–270, 2014.

[28] C. Thrampoulidis, S. Oymak, and B. Hassibi, "Regularized linear regression: A precise analysis of the estimation error," in *Proc. 28th Conf. Learn. Theory*, 2015, pp. 1683–1709.

[29] H. Rauhut and R. Ward, "Interpolation via weighted ℓ_1 minimization," *Appl. Comput. Harmonic Anal.*, vol. 40, no. 2, pp. 321–351, 2016.

[30] B. Bah and R. Ward, "The sample complexity of weighted sparse approximation," *IEEE Trans. Signal Process.*, vol. 64, no. 12, pp. 3145–3155, Jun. 2016.

[31] YALL1 Toolbox, 2011. [Online]. Available: http://yall1.blogs.rice.edu/

[32] J. Zhang, Y. Li, Z. Gu, and Z. Yu, "Effect of errors in partially known support on the recoverability of modified compressive sensing," *EURASIP J. Signal Process.*, vol. 2012, no. 1, pp. 1–7, 2012.

[33] G. Reeves and M. C. Gastpar, "Approximate sparsity pattern recovery: Information-theoretic lower bounds," *IEEE Trans. Inf. Theory*, vol. 59, no. 6, pp. 3451–3465, Jun. 2013.

[34] CVX Toolbox, 2017. [Online]. Available: http://cvxr.com/cvx/

[35] ISD, 2009. [Online]. Available:http://www.caam.rice.edu/~optimization/L1/ISD/

[36] J.-J. Fuchs, "Recovery of exact sparse representations in the presence of bounded noise," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3601–3608, Oct. 2005.

[37] D. G. Luenberger, *Introduction to Linear and Nonlinear Programming*. Reading, MA, USA: Addison-Wesley, 1973.

[38] X. Zhang, *Matrix Analysis and Applications*. Beijing, China: Tsinghua Univ. Press, 2004.

[39] Y. Li, A. Cichocki, and S.-i. Amari, "Analysis of sparse representation and blind source separation," *Neural Comput.*, vol. 16, pp. 1193–1234, 2004.

Jun Zhang (M'14) received the Bachelor's and Master's degrees in computer science from Xiangtan University, Xiangtan, China, in 2002 and 2005, respectively, and the Ph.D. degree in pattern recognition and intelligence system from the South China University of Technology, Guangzhou, China, in 2012.

Jun Zhang was a Postdoctoral Research Fellow in electronic engineering with the University of Southern California, Los Angeles, CA, USA, from 2015 to 2016. He is currently an Associate Professor with the School of Information Engineering, Guangdong University of Technology, Guangzhou, where he has been the Head of the Electrical Engineering Department since 2016. His current research interests include the fields of compressive sensing, machine learning, and biomedical signal processing.

Urbashi Mitra (F'07) received the B.S. and M.S. degrees from the University of California at Berkeley, Berkeley, CA, USA, and the Ph.D. degree from Princeton University, Princeton, NJ, USA. After a six-year stint with the Ohio State University, Columbus, OH, USA, she joined the Department of Electrical Engineering, University of Southern California, Los Angeles, CA, USA, where she is currently the Gordon S. Marshall Chair in Engineering. Her research interests include wireless communications, biological communication, underwater acoustic communications, communication and sensor networks, detection and estimation and the interface of communication, sensing and control. She is the inaugural Editor-in-Chief of the IEEE TRANSACTIONS ON MOLECULAR, BIOLOGICAL, AND MULTI-SCALE COMMUNICATIONS. She has served on the IEEE Fourier Award for Signal Processing, the IEEE James H. Mulligan, Jr. Education Medal and the IEEE Paper Prize committees. She has been an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING (2012–2015), the IEEE TRANSACTIONS ON INFORMATION THEORY (2007–2011), the IEEE JOURNAL OF OCEANIC ENGINEERING (2006–2011), and the IEEE TRANSACTIONS ON COMMUNICATIONS (1996–2001). She has Co-Chaired: (technical program) 2018 IEEE International Workshop on Signal Processing Advances in Wireless Communications in Kalamata, Greece, 2014 IEEE International Symposium on Information Theory in Honolulu, HI, USA, 2014 IEEE Information Theory Workshop in Hobart, Tasmania, the IEEE 2012 International Conference on Signal Processing and Communications, Bangalore India, and the IEEE Communication Theory Symposium at ICC 2003 in Anchorage, AK, USA; and was the General Co-Chair for the first ACM Workshop on Underwater Networks at Mobicom 2006, Los Angeles, CA. She was the Tutorials Chair for IEEE ISIT 2007 in Nice, France and the Finance Chair for the IEEE ICASSP 2008 in Las Vegas, NV, USA. She served as the Co-Director of the Communication Sciences Institute, University of Southern California, from 2004 to 2007. She was the recipient of 1996 National Science Foundation (NSF) CAREER Award and 1997 OSU College of Engineering MacQuigg Award for Teaching, 2000 OSU College of Engineering Lumley Award for Research, 2001 Okawa Foundation Award, the Texas Instruments Visiting Professorship (Fall 2002, Rice University), the Texas Instruments Visiting Professorship (Fall 2002, Rice University), the 2009 DCOSS Applications and Systems Best Paper Award, the 2012 Globecom Signal Processing for Communications Symposium Best Paper Award, the 2012 NAE Lillian Gilbreth Lectureship, the USC Center for Excellence in Research Fellowship (2010–2013), the 2014–2015 IEEE Communications Society Distinguished Lecturer, the 2017 IEEE Communications Society Women in Communications Engineering Technical Achievement Award, a 2016 UK Royal Academy of Engineering Distinguished Visiting Professorship, a 2016 US Fulbright Scholar Award, and a 2016–2017 UK Leverhulme Trust Visiting Professorship.

Kuan-Wen Huang received the B.S. degree in electrical engineering from National Taiwan University, Taipei, Taiwan in 2012. He is currently working toward the Ph.D. degree in electrical engineering with the University of Southern California, Los Angeles, CA, USA. His current research focuses primarily on sparse signal processing, machine learning and optimization.

Nicolò Michelusi (S'09– M'13– SM'18) received the B.Sc. (with hons.), M.Sc. (with hons.) and Ph.D. degrees from the University of Padova, Padova, Italy, in 2006, 2009, and 2013, respectively, and the M.Sc. degree in telecommunications engineering from the Technical University of Denmark, Lyngby, Denmark, in 2009, as part of the T.I.M.E. double degree program. He was a Postdoctoral Research Fellow with the Ming-Hsieh Department of Electrical Engineering, University of Southern California, USA, in 2013–2015. He is currently an Assistant Professor with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA. His research interests lie in the areas of 5G wireless networks, millimeter-wave communications, stochastic optimization, and distributed optimization. He is an Associate Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and serves as a reviewer for several IEEE Transactions.