Iterative Reweighted ℓ_2/ℓ_1 Recovery Algorithms for Compressed Sensing of Block Sparse Signals

Zeinab Zeinalkhani and Amir H. Banihashemi, *Senior Member, IEEE*

***Abstract*—In many applications of compressed sensing the signal is block sparse, i.e., the non-zero elements of the sparse signal are clustered in blocks. Here, we propose a family of iterative algorithms for the recovery of block sparse signals. These algorithms, referred to as *iterative reweighted ℓ_2/ℓ_1 minimization* algorithms (IR-ℓ_2/ℓ_1), solve a weighted ℓ_2/ℓ_1 minimization in each iteration. Our simulation and analytical results on the recovery of both ideally and approximately block sparse signals show that the proposed iterative algorithms have significant advantages in terms of accuracy and the number of required measurements over non-iterative approaches as well as existing iterative methods. In particular, we demonstrate that, by increasing the block length, the performance of the proposed algorithms approaches the Wu-Verdú theoretical limit. The improvement in performance comes at a rather small cost in complexity increase. Further improvement in performance is achieved by using a priori information about the location of non-zero blocks, even if such a priori information is not perfectly reliable.**

***Index Terms*—Block sparsity, compressed sensing, iterative recovery algorithms, iterative reweighted ℓ_2/ℓ_1 minimization, ℓ_2/ℓ_1 minimization.**

I. Introduction

COMPRESSED sensing (CS) is a signal processing technique that can recover an unknown signal with high probability from a small set of linear projections, called *measurements*, conditioned on the sparsity of the signal in some domain (see [1]–[3], for details).

Let $\bar{\mathbf{r}} \in \mathbb{C}^N$ denote an N dimensional complex signal, which is sparse in domain $\mathbf{\Psi}$, where $\mathbf{\Psi}$ is an $N \times N$ *basis matrix*. In particular, we assume that representation $\bar{\mathbf{x}}$ of $\bar{\mathbf{r}}$ in domain $\mathbf{\Psi}$ is a K-*sparse signal*, i.e., $\bar{\mathbf{x}} \in \mathbb{C}^N$ has only K non-zero elements $(K \ll N)$. Further assume that $\mathbf{\Phi}$ is an $M \times N$ random *measurement matrix* with complex elements in general. For instance, let the entries of $\mathbf{\Phi}$ be independent and identically distributed (i.i.d.) real random variables from a zero-mean Gaussian distribution with variance $1/N$. Mathematically, the M-dimensional measurement vector $\mathbf{y}$ is given by $\mathbf{y} = \mathbf{\Phi}\bar{\mathbf{r}} = \mathbf{\Phi}\mathbf{\Psi}\bar{\mathbf{x}}$. In compressed sensing, the problem is to find the sparsest vector $\hat{\mathbf{x}}$ of dimension N given $\mathbf{y}$, i.e.,

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \quad \|\mathbf{x}\|_0 \quad \text{s.t. } \mathbf{y} = \mathbf{A}\mathbf{x}, \tag{1}$$

where $\mathbf{A} = \mathbf{\Phi}\mathbf{\Psi}$ is called the *sensing matrix*.[1] Note that $\bar{\mathbf{x}}$ denotes the true signal, while $\hat{\mathbf{x}}$ and $\mathbf{x}$ denote the estimated signal, which is the solution of the optimization problem, and the variable in the optimization problem, respectively. It is well-known that in the setting of Problem (1), the K-sparse signal $\bar{\mathbf{x}}$ can be recovered with high probability, using only $M = K + 1$ measurements if $\mathbf{\Phi}$ is randomly generated independently of $\bar{\mathbf{x}}$ [4]. In other words, the solution of ℓ_0-norm minimization (ℓ_0 minimization, in brief) is, with high probability, the true signal, i.e., $\hat{\mathbf{x}} = \bar{\mathbf{x}}$. The problem in (1), however, is NP-hard [5].

Fortunately, the ℓ_0 minimization problem in (1) can be relaxed to an ℓ_1 minimization problem:

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

that can be solved in polynomial time using linear programming techniques and yet can provide the same solution as in (1) under certain conditions [1]–[3], [5].

In the presence of noise, Problem (2) can be written as

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \ \|\mathbf{x}\|_1 \quad \text{s.t. } \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \le \alpha, \tag{3}$$

where α bounds the ℓ_2-norm of the noise in the measurements. If the measurements are corrupted by a zero-mean Gaussian noise with variance σ_n^2, a likely upper bound on the ℓ_2-norm of the noise is $\alpha = \sigma_n\sqrt{M + 2\sqrt{2M}}$ (details can be found in [7], Section 3.3). Problem (3) is referred to as LASSO [8] and is a convex problem (a second order cone programming problem [9]) that can be solved efficiently in polynomial time, using convex programming tools [10], [11].

It is proved that when matrix $\mathbf{A}$ is a Gaussian random matrix, the solution of Problem (2), $\hat{\mathbf{x}}$, coincides with the true signal $\bar{\mathbf{x}}$ (solution of (1)), if $M = O(K\log(N/K))$ [1]–[3], [5], [12], which is much smaller than N, but can still be a large number compared to K.

There are a number of iterative methods such as *iterative reweighted least squares* (IRLS) [13], [14], *iterative reweighted ℓ_1 minimization* (IRℓ_1) [7], [15] and *iterative support detection*

Manuscript received July 16, 2014; revised November 25, 2014 and April 24, 2015; accepted May 26, 2015. Date of publication June 02, 2015; date of current version July 21, 2015. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Wenwu Wang. This work appeared in part in the *Proceedings of the IEEE International Conference on Communications* (ICC), Ottawa, ON, Canada, June 10–15, 2012.

The authors are with the Department of Systems and Computer Engineering, Carleton University, Ottawa, ON K1S 5B6 Canada (e-mail: zzeinab@sce.carleton.ca; ahashemi@sce.carleton.ca).

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.2015.2441032

[1] In (1) and in the rest of the paper, we use the ℓ_p-norm with the definition

$$\|\mathbf{x}\|_p = \begin{cases} |\{i : x_i \neq 0, i = 1, \ldots, N\}| & p = 0 \\ \left(\sum_{i=1}^N |x_i|^p\right)^{1/p} & 0 < p < \infty \\ \max_{i=1,\ldots,N} |x_i| & p = \infty. \end{cases}$$

(ISD) [16] that in many cases, can recover the sparse signal with far fewer measurements than the standard ℓ_1 minimization for both noisy and noiseless cases.

In a reconstruction procedure, one can also reduce the number of required measurements by employing a priori information on the support region of the sparse signal [17]. In [14], it is shown that a priori information can be used by the IRLS algorithm to improve the reconstruction performance in terms of the number of required measurements, the number of iterations and the convergence time. Zhang and Qiu [18] incorporated a priori information in the orthogonal matching pursuit algorithm in a spectrum sensing application to improve the sensing performance.

In this paper, we consider the reconstruction of *block sparse signals* (i.e., signals whose non-zero elements occur in clusters). Such signals appear in various applications such as spectrum sensing [19], equalization of sparse communication channels [20], and in general when dealing with multi-band signals [21]. Compressed sensing of block sparse signals has been considered in a number of recent publications [22]–[30], where different assumptions about the sensing matrix were made and different techniques including convex and non-convex relaxations were used for the recovery of the signal. In [27], ℓ_2/ℓ_1-*norm minimization* (ℓ_2/ℓ_1 minimization, in brief) was proposed as a technique to reconstruct block sparse signals. The main result of [27] on the recovery of block sparse signals using the ℓ_2/ℓ_1 relaxation was derived based on the null-space characterization of the sensing matrix $\mathbf{A}$. In particular, it was shown in [27] that in the noiseless scenario, perfect recovery using a Gaussian sensing matrix is possible with high probability as the signal length tends to infinity. In [28], also, ℓ_2/ℓ_1 minimization was proposed for the block sparse signal recovery. The notion of block restricted isometry property (RIP) was then used to derive the conditions for perfect recovery. Eldar *et al.* [29] used yet another condition, based on block-coherence measure, to guarantee successful recovery of block sparse signals through ℓ_2/ℓ_1 minimization. In [30], ℓ_q/ℓ_1 minimization with $q \geq 2$, and an alternate convex program, denoted by P'_{ℓ_q/ℓ_1}, for the recovery of block sparse signals were considered. In the same work, conditions for the recovery of the signal based on the notions of mutual and cumulative subspace coherence were also derived.[2]

A popular category of recovery algorithms for compressed sensing are those based on *message-passing algorithms*, see, e.g., [31]–[42]. In particular, the approximate message-passing (AMP) algorithm of [36] has attracted much attention due to its remarkable performance/complexity trade-off [43]. AMP has also been applied to the recovery of block sparse signals [38], [44]–[46]. In this context, it is demonstrated in [44] that AMP with James-Stein's shrinkage estimator can outperform ℓ_2/ℓ_1 minimization for large block sizes. In particular, it is proved in [44] that in the asymptotic regime of infinite block length, this variant of AMP performs close to the information theoretic limit. This result is complemented with the result of [45], which indicates that for the regime of a fixed block size and very small undersampling ratios, AMP cannot provide any improvement over ℓ_2/ℓ_1 minimization. Complexity-wise, AMP has the same order of computational complexity as ℓ_2/ℓ_1 minimization [40].

Here, we propose a family of iterative algorithms to recover block sparse signals. In each iteration, the proposed algorithms solve a weighted ℓ_2/ℓ_1 minimization with the weights being updated in each iteration based on the solution of the previous iteration, hence the name "iterative reweighted (IR) ℓ_2/ℓ_1 minimization." Different algorithms within the family differ in the way that the weights are updated from one iteration to the next. In this paper, we mainly focus on two algorithms from this family, referred to as IR$^{(1)}$-ℓ_2/ℓ_1 and IR$^{(2)}$-ℓ_2/ℓ_1. In the former algorithm, the weights are updated inversely proportional to the ℓ_2-norm of the corresponding blocks in the solution of the previous iteration. In the latter algorithm, however, some weights are set to zero, if the ℓ_2-norm of the corresponding blocks in the previous iteration is larger than a certain threshold. Otherwise, the weights are set to one. In general, all the IR-ℓ_2/ℓ_1 algorithms perform the standard ℓ_2/ℓ_1 minimization in the first iteration and then improve the solution in subsequent iterations. Our results show that the iterative improvement to standard ℓ_2/ℓ_1 minimization is attainable for noiseless scenarios and for cases where measurements are corrupted with noise whose power relative to the signal is below a certain threshold. For such cases, we demonstrate through simulation results that significant improvements in terms of the number of required measurements for recovery can be obtained compared to the standard ℓ_2/ℓ_1 minimization in both noiseless and noisy cases. Particularly, as the length of the blocks increases, the required number of measurements for the reconstruction of the block sparse signal using IR-ℓ_2/ℓ_1 algorithms approaches the theoretical limit of K [47]. In addition, we demonstrate by a number of examples that the proposed algorithms can outperform James-Stein AMP [44] by a rather large margin in noiseless scenarios and in noisy scenarios with high signal-to-noise ratio (SNR) and small number of measurements (small under-sampling ratios). In general, our numerical results show that the proposed approaches can reconstruct block sparse signals better than existing AMP variants [38], [40], [44]. That said, current AMP variants might be using denoisers that are not well suited to block sparse signals.

Moreover, we analyze IR-ℓ_2/ℓ_1 algorithms and give theoretical guarantees on the success of these algorithms in recovering block sparse signals. We also show that at the presence of a priori information on the support region of the block sparse signal, the performance of both IR$^{(1)}$-ℓ_2/ℓ_1 and IR$^{(2)}$-ℓ_2/ℓ_1 algorithms can be improved, even if such information is not perfectly reliable. In general, our theoretical and simulation results demonstrate the superior performance of the proposed iterative recovery algorithms compared to the existing schemes. This often comes at a rather small cost in computational complexity.

In many practical scenarios, $\bar{\mathbf{x}}$ is not ideally block sparse (i.e., $\bar{\mathbf{x}}$ has very few or no zero elements but has many close to zero elements in clusters). Such signals are called *approximately block sparse signals,* which will be defined more precisely later. In such cases, it is often impossible to recover the signal $\bar{\mathbf{x}}$ accurately from a relatively small number of measurements M. In such scenarios, we establish upper bounds on the signal reconstruction error of the proposed iterative algorithms and show

[2]It was demonstrated in [30] that for some cases, P'_{ℓ_2/ℓ_1} can outperform ℓ_2/ℓ_1 minimization. In the context of the iterative algorithms proposed here, where we use a weighted version of ℓ_2/ℓ_1 minimization in each iteration, however, we observed no notable difference in the performance of iterative algorithms when ℓ_2/ℓ_1 minimization was replaced with P'_{ℓ_2/ℓ_1}.

improvements over the bounds derived for the standard ℓ_2/ℓ_1 minimization.

This paper is organized as follows. In Section II, we present the notations and definitions that will be used throughout the paper. In Section III, we provide the background on the compressed sensing of block sparse signals. In Section IV, weighted ℓ_2/ℓ_1 as the core of iterative algorithms for reconstruction of block sparse signals is proposed. We then present the family of IR-ℓ_2/ℓ_1 algorithms in the absence and presence of a priori information about the location of non-zero blocks. We also analyze the iterative algorithms in the context of the recovery of both ideally and approximately block sparse signals, with focusing mainly on IR$^{(1)}$-ℓ_2/ℓ_1 and IR$^{(2)}$-ℓ_2/ℓ_1. In Section V, simulation results are presented. We conclude the paper in Section VI.

II. Basic Notations and Definitions

In this paper, we use capital bold letters to refer to matrices and lowercase bold letters for vectors. In addition, the bold symbols $\mathbf{1}$ and $\mathbf{0}$ refer to the vectors of all ones and all zeros, respectively. All vectors will be column vectors and the superscript T denotes the transpose operation.

Let $\mathbb{C}$ and $\mathbb{R}$ be the fields of complex and real numbers, respectively. For a matrix $\mathbf{A}$ with N columns, we denote its null space by $\mathcal{N}(\mathbf{A}) = \{\boldsymbol{\eta} \in \mathbb{C}^N | \mathbf{A}\boldsymbol{\eta} = \mathbf{0}\}$.

Definition 1: Suppose signal $\bar{\mathbf{x}}$ consists of L sub-vectors, $\{\bar{\mathbf{u}}_1, \bar{\mathbf{u}}_2, \ldots, \bar{\mathbf{u}}_L\}$, where the sub-vectors can be either zero or non-zero (a non-zero vector can have both zero and non-zero elements). The signal $\bar{\mathbf{x}} = [\bar{\mathbf{u}}_1^T, \bar{\mathbf{u}}_2^T, \ldots, \bar{\mathbf{u}}_L^T]^T$ is called *ideally P-block sparse* (or *P-block sparse*, in brief) if at most $P \ll L$ out of L sub-vectors are non-zero.

We use the notations $f_0 < \cdots < f_L$ for the boundaries of the blocks, i.e., the sub-vector $\bar{\mathbf{u}}_i$, $i \in \{1, \ldots, L\}$, which is the ith segment of $\bar{\mathbf{x}}$, corresponds to the elements whose indices are in the range $f_{i-1}, \ldots, f_i - 1$. In this paper, we use the terms *block*, *sub-vector* and *cluster*, interchangeably. We also assume that the block boundaries are known for the recovery. Similar assumption is made in other works including [27].

Definition 2: Signal $\bar{\mathbf{x}}$ is called *approximately P-block sparse with parameter* δ, if it consists of L sub-vectors $\{\bar{\mathbf{u}}_1, \ldots, \bar{\mathbf{u}}_L\}$ such that $\sum_{i=P+1}^{L} \|\bar{\mathbf{u}}_{\pi(i)}\|_2 \leq \delta$, for some $\delta \geq 0$, where $P \ll L$ and π is a permutation of the set $\{1, \ldots, L\}$ such that $\|\bar{\mathbf{u}}_{\pi(1)}\|_2 \geq \|\bar{\mathbf{u}}_{\pi(2)}\|_2 \geq \cdots \geq \|\bar{\mathbf{u}}_{\pi(L)}\|_2$. We call π a *proper permutation* with respect to the set $\{\|\bar{\mathbf{u}}_i\|_2\}_{1 \leq i \leq L}$.[3]

For a given subset Γ of $\{1, \ldots, L\}$, $\mathbf{x}_\Gamma$ denotes the sub-vector of $\mathbf{x}$ consisting of $\mathbf{u}_i$ vectors with $i \in \Gamma$. For an index set $S \subset \{1, 2, \ldots, L\}$, S^c and $|S|$ denote the complement of S with respect to $\{1, 2, \ldots, L\}$ and the size of set S, respectively.

III. Block Sparse Compressed Sensing Background

To exploit the block structure of a block sparse signal $\bar{\mathbf{x}}$ in the recovery process, the following ℓ_2/ℓ_1 minimization was proposed in [27]

$$\begin{aligned}\hat{\mathbf{x}} &= [\hat{\mathbf{u}}_1^T, \hat{\mathbf{u}}_2^T, \ldots, \hat{\mathbf{u}}_L^T]^T \\ &= \arg\min_{\mathbf{u}_1, \ldots, \mathbf{u}_L} \sum_{i=1}^{L} \|\mathbf{u}_i\|_2 \quad \text{s.t. } \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq \alpha, \end{aligned} \tag{4}$$

[3]To apply this definition to random signals, we require the signal to satisfy this condition with probability (close to) 1.

where α bounds the ℓ_2-norm of the noise in measurements. The problem in (4) is a relaxation of the following ℓ_2/ℓ_0 minimization problem

$$\begin{aligned}\bar{\mathbf{x}} &= [\bar{\mathbf{u}}_1^T, \bar{\mathbf{u}}_2^T, \ldots, \bar{\mathbf{u}}_L^T]^T \\ &= \arg\min_{\mathbf{u}_1, \ldots, \mathbf{u}_L} |\{i : \|\mathbf{u}_i\|_2 \neq 0, i = 1, \ldots, L\}| \\ &\quad \text{s.t.} \| \mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq \alpha. \end{aligned}$$

In the rest of the paper, we use the terms "true block sparse signal" and "solution of ℓ_2/ℓ_0 minimization," interchangeably, to refer to $\bar{\mathbf{x}}$.

The simulation results in [27] demonstrated that ℓ_2/ℓ_1 minimization outperformed the ℓ_1 minimization in the case of block sparse signals. The following theorem introduces a necessary and sufficient condition on the sensing matrix $\mathbf{A}$ so that $\hat{\mathbf{x}}$ and $\bar{\mathbf{x}}$ coincide for any ideally P-block sparse signal $\bar{\mathbf{x}}$.

Theorem 1: [27] Let $\bar{\mathbf{x}} = [\bar{\mathbf{u}}_1^T, \ldots, \bar{\mathbf{u}}_L^T]^T$ be the true P-block sparse signal and $\hat{\mathbf{x}} = [\hat{\mathbf{u}}_1^T, \ldots, \hat{\mathbf{u}}_L^T]^T$ be a minimizer of (4) for $\alpha = 0$. Further, assume that $\mathbf{A}$ is an $M \times N$ sensing matrix. Then, signals $\hat{\mathbf{x}}$ and $\bar{\mathbf{x}}$ coincide for any P-block sparse $\bar{\mathbf{x}}$ *if and only if* for all index sets $S \subset \{1, 2, \ldots, L\}$ with $|S| \leq P$ and for all $\boldsymbol{\eta} \in \mathcal{N}(\mathbf{A}) \setminus \mathbf{0}$, we have

$$\sum_{i \in S} \|\boldsymbol{\eta}_i\|_2 < \gamma \sum_{i \in S^c} \|\boldsymbol{\eta}_i\|_2, \tag{5}$$

for some $0 < \gamma < 1$.

Here, $\boldsymbol{\eta}_i$ is the ith block of $\boldsymbol{\eta}$ (the block boundaries of $\boldsymbol{\eta}$ are the same as those of $\bar{\mathbf{x}}$). In the following, we use the notation $\bar{\gamma}$ as the infimum of all values of γ that satisfy (5).

In practice, the signals are not ideally block sparse. In [48], it was proved that when $\bar{\mathbf{x}}$ is approximately P-block sparse, the difference between the solution of ℓ_2/ℓ_1 minimization with $\alpha = 0$, $\hat{\mathbf{x}}$, and the true signal, $\bar{\mathbf{x}}$, in ℓ_2/ℓ_1 norm (i.e., $\sum_{i=1}^{L} \|\hat{\mathbf{u}}_i - \bar{\mathbf{u}}_i\|_2$), is bounded from above.

Theorem 2: [48] Let $\bar{\mathbf{x}}$ be the true approximately P-block sparse signal with parameter δ, and $\hat{\mathbf{x}}$ be a minimizer of (4) for $\alpha = 0$. Further, assume that $\mathbf{A}$ is an $M \times N$ sensing matrix. Then, for $0 < \bar{\gamma} < 1$, we have

$$\sum_{i=1}^{L} \|\hat{\mathbf{u}}_i - \bar{\mathbf{u}}_i\|_2 \leq 2 \left(\frac{1 + \bar{\gamma}}{1 - \bar{\gamma}} \right) \delta, \tag{6}$$

if and only if for all index sets $S \subset \{1, 2, \ldots, L\}$ with $|S| \leq P$ and for all $\boldsymbol{\eta} \in \mathcal{N}(\mathbf{A}) \setminus \mathbf{0}$, we have

$$\sum_{i \in S} \|\boldsymbol{\eta}_i\|_2 < \bar{\gamma} \sum_{i \in S^c} \|\boldsymbol{\eta}_i\|_2, \tag{7}$$

where $\boldsymbol{\eta}_i$ is the ith block of $\boldsymbol{\eta}$ with the block boundaries the same as those of $\bar{\mathbf{x}}$.

IV. Iterative Reweighted (IR-) ℓ_2/ℓ_1 Minimization

Although ℓ_1 minimization can solve ℓ_0 minimization problem under certain conditions, it requires a larger number of measurements for the signal reconstruction due to its dependency on the magnitude of the signal. If we can also reduce (eliminate) the dependency of the ℓ_2/ℓ_1 minimization (4) on the ℓ_2-norm of the blocks by exploiting a weighting strategy, the reconstruction performance can be improved. In the following, we first present weighted ℓ_2/ℓ_1 as the core of iterative algorithms for the reconstruction of block sparse signals. We then propose the iterative

algorithms when no a priori information is available. Furthermore, we modify the reconstruction process to use the a priori information about the support region of the block sparse signal. We also analyze the success of weighted ℓ_2/ℓ_1 minimization in the recovery of ideally block sparse signals where the measurements are noiseless. We then generalize the analysis to the case of approximately block sparse signals.

A. Weighted ℓ_2/ℓ_1 Minimization

Consider the following weighted ℓ_2/ℓ_1 minimization problem:

$$\begin{aligned}\hat{\mathbf{x}} &= [\hat{\mathbf{u}}_1^T, \hat{\mathbf{u}}_2^T, \ldots, \hat{\mathbf{u}}_L^T]^T \\ &= \arg\min_{\mathbf{u}_1,\ldots,\mathbf{u}_L} \sum_{i=1}^{L} \omega_i \|\mathbf{u}_i\|_2 \quad \text{s.t. } \|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2 \le \alpha, \end{aligned} \tag{8}$$

where ω_i's are the weights that can be considered as free parameters.

The signal reconstruction can be improved if the values of the weights are properly set. In particular, in this paper, we use the formulation in (8) iteratively, where the weights in the current iteration are determined based on the estimate of the signal obtained in the previous iteration. In each iteration, the convex problem (8) is solved by using convex programming tools [10], [11], [49]. The computational complexity of solving (8) is essentially the same as that of unweighted (standard) ℓ_2/ℓ_1 minimization. Therefore, for an algorithm that solves (8) iteratively, the computational complexity is roughly $j_{\max}$ times that of the standard ℓ_2/ℓ_1 minimization, where $j_{\max}$ is the maximum number of iterations.

Our focus in this paper will be mainly on two incarnations of the iterative algorithm: IR$^{(1)}$-ℓ_2/ℓ_1 and IR$^{(2)}$-ℓ_2/ℓ_1.

1) IR$^{(1)}$-ℓ_2/ℓ_1*:* In this algorithm, the weights are updated inversely proportional to the ℓ_2-norm of their corresponding blocks in the previous iteration.

2) IR$^{(2)}$-ℓ_2/ℓ_1*:* In this algorithm, the weights only take values 0 and 1. A weight will be zero if the ℓ_2-norm of the corresponding block in the previous iteration is larger than a certain threshold. Otherwise, the weight will be set at one. The terms with zero weight disappear from the summation in the formulation of (8). The summation is thus performed over an index set Γ, which is a subset of $\{1,\ldots,L\}$. We use the term *truncated* ℓ_2/ℓ_1 *minimization* to refer to (8) in this scenario.

In the rest of the paper, we often use the following version of (8), in which all the weights appearing in the summation are non-zero:

$$\begin{aligned}\hat{\mathbf{x}} &= [\hat{\mathbf{u}}_1^T, \hat{\mathbf{u}}_2^T, \ldots, \hat{\mathbf{u}}_L^T]^T \\ &= \arg\min_{\mathbf{u}_1,\ldots,\mathbf{u}_L} \sum_{i\in\Gamma} \omega_i \|\mathbf{u}_i\|_2 \quad \text{s.t. } \|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2 \le \alpha, \end{aligned} \tag{9}$$

where $\Gamma \subseteq \{1,\ldots,L\}$ is a given index set.

B. Iterative Reweighted ℓ_2/ℓ_1 Minimization With no A Priori Information

Consider the weighted ℓ_2/ℓ_1 minimization problem in (8). In the absence of any a priori information, we propose an iterative reconstruction scheme, IR-ℓ_2/ℓ_1, where the ℓ_2-norm of estimated blocks in one iteration is used to select the corresponding weights in the next iteration. Each iteration of IR-ℓ_2/ℓ_1 thus consists of two steps: (i) updating the weights, and (ii) signal reconstruction. In Step (i), some weights may be selected to be zero, effectively reducing the support set from $\{1,\ldots,L\}$ to a subset of it, Γ. In the following, we refer to this as "support detection" and limit the operations of Step (i) to support detection and two specific update rules for non-zero weights. We initialize the proposed iterative algorithms with the solution to the standard ℓ_2/ℓ_1 minimization and demonstrate that by properly choosing the update rules for the index set Γ and non-zero weights, the initial solution can be much improved iteratively.

The algorithmic framework of IR-ℓ_2/ℓ_1 is given in Algorithm 1:

Algorithm 1: Iterative reweighted ℓ_2/ℓ_1 minimization (IR-ℓ_2/ℓ_1)
Inputs:$\mathbf{\Phi}$ (or $\mathbf{A}$), $\mathbf{y}$, ρ_0, β, α, $j_{\max}$ and $f_0, f_1 \ldots, f_L$
Output: $\hat{\mathbf{x}} = [\hat{\mathbf{u}}_1^T, \hat{\mathbf{u}}_2^T, \ldots, \hat{\mathbf{u}}_L^T]^T$

---

1) Initialization step:
   (a) Set the iteration number $j \leftarrow 1$ and $\rho \leftarrow \rho_0$;
   (b) $\Gamma^{(1)} \leftarrow \{1,\ldots,L\}$;
   (c) $\omega_i^{(1)} = 1, i \in \{1,\ldots,L\}$;
   (d) Obtain $\hat{\mathbf{x}}^{(1)}$ by solving **(9)** for $\Gamma = \Gamma^{(1)}$ and $\omega_i = \omega_i^{(1)}$;
2) **While** $j < j_{\max}$, **do** :
   (a) Support detection:
   Determine $\Gamma^{(j+1)}$ using $\hat{\mathbf{x}}^{(j)}$;
   (b) Update the non-zero weights:
   Determine $\omega_i^{(j+1)}$, for $i \in \Gamma^{(j+1)}$ using $\hat{\mathbf{x}}^{(j)}$;
   (c) Signal reconstruction:
   Obtain $\hat{\mathbf{x}}^{(j+1)}$ by solving **(9)** for $\Gamma = \Gamma^{(j+1)}$ and $\omega_i = \omega_i^{(j+1)}$;
   (d) Update ρ:
   if $\frac{\|\hat{\mathbf{x}}^{(j+1)} - \hat{\mathbf{x}}^{(j)}\|_2}{\|\hat{\mathbf{x}}^{(j)}\|_2} < \sqrt{\frac{\rho}{100}}$ **then** $\rho \leftarrow \rho/10$;
   (e) $j \leftarrow j+1$.

In Algorithm 1, the superscript (j) is used to refer to a variable, a vector or a set in iteration j.

In support detection, we update $\Gamma^{(j+1)}$ such that $\Gamma^{(j+1)} \leftarrow \{i : \|\hat{\mathbf{u}}_i^{(j)}\|_2 < \epsilon^{(j)}\}$. We set the threshold $\epsilon^{(j)}$ for the support detection in iteration j at $\epsilon^{(j)} = \max_i \|\hat{\mathbf{u}}_i^{(j)}\|_2/\beta^j$ for $i = 1,\ldots,L$, with $\beta > 1$. Note that this selection reduces the threshold as the iterations proceed. The choice of β affects the efficiency of the algorithm. If β is selected to be too large, $\epsilon^{(j)}$ values will be too small, resulting in the estimated index set Γ to be too small, making the reconstruction step ineffective due to excessive truncation. If, on the other hand, β is selected to be too small, then too many iterations will be required to accurately reconstruct the block sparse signal, decreasing the speed of recovery.

For updating the non-zero weights, we choose the weights as

$$\omega_i^{(j+1)} = \frac{1}{\|\hat{\mathbf{u}}_i^{(j)}\|_2 + \rho}, \quad i \in \Gamma^{(j+1)}. \tag{10}$$

To justify the choice for non-zero weights in (10), we note that in (9), large weights will discourage non-zero blocks whereas small weights will encourage such blocks. Therefore, the weights could be chosen inversely proportional to the true ℓ_2-norm of the blocks. With the lack of knowledge about true

ℓ_2-norms, here, we use the updating rule of (10), in which ℓ_2-norms of estimated blocks in the previous iteration are used instead of true ℓ_2-norms, and ρ is a regularization constant to prevent instability. We will prove that these weights guarantee $\hat{\mathbf{x}} = \bar{\mathbf{x}}$ under certain conditions.

One may also note that if the ℓ_2-norm of the ith block in the jth iteration, $\|\hat{\mathbf{u}}_i^{(j)}\|_2$, is small, the index of this block would be in the set $\Gamma^{(j+1)}$ and $\omega_i^{(j+1)}$ would be large. So the blocks with small ℓ_2-norm are pushed towards becoming zero. This thus yields a block sparse solution.

At each iteration, the set of block indices, Γ, is updated in Step 2(a). The size of this set however may not necessarily decrease with iterations. After updating Γ, all the components of $\mathbf{x}$ are updated in Step 2(c), including both the components corresponding to Γ and those that correspond to Γ^c.

Since, the overall optimization problem is non-convex, the proposed algorithm would produce local minima. To avoid being trapped in an undesirable local minimum, choosing a suitable starting point is important. As a good starting point, in the first iteration (i.e., $j = 1$), the algorithm finds the solution of the standard ℓ_2/ℓ_1 minimization (i.e., in the initialization step, we set $\Gamma^{(1)} \leftarrow \{1, \ldots, L\}$ and $\omega_i^{(1)} = 1, \forall i \in \{1, 2, \ldots, L\}$). In this iteration, non-zero and zero blocks are weighted the same since there is no a priori information about the zero or non-zero blocks available.

Since an all-zero sub-vector $\hat{\mathbf{u}}_i^{(j)}, i \in \{1, 2, \ldots, L\}$ can cause instability if the parameter ρ is equal to zero, we choose ρ to be strictly positive. Similar to [13] and [14], to avoid getting trapped in a local minimum, we start by setting ρ to be a relatively large value (e.g., ρ_0 can be selected close to the standard deviation of ℓ_2-norm of non-zero blocks) and reduce its value by a factor of 10,[4] in the $(j+1)$th iteration if

$$\frac{\|\hat{\mathbf{x}}^{(j+1)} - \hat{\mathbf{x}}^{(j)}\|_2}{\|\hat{\mathbf{x}}^{(j)}\|_2} < \sqrt{\frac{\rho}{100}}. \tag{11}$$

Two special cases:

1) IR$^{(1)}$-ℓ_2/ℓ_1*:* In this case, Algorithm 1 is executed without the support detection at Step 2(a), i.e., in all iterations, we set $\Gamma^{(j)} \leftarrow \{1, \ldots, L\}$.[5]

By choosing the non-zero weights based on (10), this iterative algorithm attempts to find a local minimum of the objective function $\sum_{i=1}^{L} \|\mathbf{u}_i\|_2/(\|\mathbf{u}_i\|_2+\rho)$ that more closely than the objective function of the ℓ_2/ℓ_1 minimization resembles the ℓ_0 minimization. In early iterations, the signal estimate is inaccurate, however, the locations of the active blocks with high ℓ_2-norm are likely to be identified. In the following iterations, the influence of those high-energy blocks is down-weighted. This implies that the active blocks with smaller levels of energy would be up-weighted and can thus be detected easier.

2) IR$^{(2)}$-ℓ_2/ℓ_1*:* In this case, Algorithm 1 is executed without the weight-updating step 2(b), i.e., in all iterations, we set $\omega_i^{(j)} = 1, \forall i \in \Gamma^{(j)}$.[6] In other words, from the inaccurate estimate of the signal in the early iterations, the weights corresponding to the high-energy blocks are likely to be set to 0. Otherwise, they will be set to 1. In fact, it is highly probable that Γ^c includes the indices of the high-energy blocks and, hence, $\mathbf{x}_\Gamma$ would be sparser than $\mathbf{x}$. This implies that the low-energy blocks can be detected easier.

By choosing a suitable value for β, this algorithm often converges very fast (in less than 4 iterations). Unlike the case for IR$^{(2)}$-ℓ_2/ℓ_1, the convergence speed of IR$^{(1)}$-ℓ_2/ℓ_1 cannot be adjusted by any parameter in the algorithm. Therefore, in some applications where the computational time is important, with setting $j_{\max}$ to a small number, IR$^{(2)}$-ℓ_2/ℓ_1 would outperform IR$^{(1)}$-ℓ_2/ℓ_1. This will be verified in Section V.

C. Iterative Reweighted ℓ_2/ℓ_1 Minimization With a Priori Information About the Support Region of the Sparse Signal

In practical applications such as magnetic resonance imaging (MRI) and wideband spectrum sensing, the partial information about the support region of the sparse signal may be available a priori. In general, in the applications whose sparsity patterns are varying slowly with time, the support estimate can be used from the previous time slot. In [50], the authors showed that in MRI, the maximum support changes for two subsequent sequences are less than 2%. In wideband spectrum sensing also the support of the block sparse signal changes slowly over time. Moreover, in the context of spectrum sensing, some of the frequency bands are heavily used by the users such as local radio stations, local TV stations, etc. [18], and this information can be available at the recovery terminal.

Here, we assume that the index set of non-zero blocks is partially known, possibly with some error. We use such side-information to improve the recovery performance via changing the initial step of IR-ℓ_2/ℓ_1 algorithms.

Let $\mathcal{Q}$ be a subset of $\{1, 2, \ldots, L\}$ containing the indices of the blocks that are known to be non-zero, i.e., $\|\bar{\mathbf{u}}_p\|_2 \neq 0, \forall p \in \mathcal{Q}$. This prior information, however, may not be perfectly reliable. To measure the reliability of the side information, we define *reliability*, $\mathcal{R}$, as the ratio of the expected number of correctly identified components of $\mathcal{Q}$ (i.e., the components that represent the true non-zero blocks) to the size of $\mathcal{Q}$. In some applications such as spectrum sensing, when the knowledge about the statistical behavior of the signal occupancy or the changing rate of the support is available, the value of $\mathcal{R}$ would also be available. For instance, if the average changes of the support from a time slot to the next is 5%, set $\mathcal{Q}$ can include the indices of occupied blocks in the previous time slot and $\mathcal{R}$ would be 0.95. It is clear that in the applications in which the support changes slowly over time, $\mathcal{R}$ is close to 1.

Based on the a priori information, we modify Algorithm 1 by initializing the weights as follows

$$\omega_i^{(1)} = \begin{cases} 1, & \text{if } i \notin \mathcal{Q} \\ 1 - \mathcal{R} + \tau\xi(\mathcal{R} - 1), & \text{otherwise,} \end{cases} \tag{12}$$

where $0 \leq \tau \ll 1$ is a non-negative small constant and $\xi(.)$ is defined by $\xi(x) = 1$ if $x = 0$; and $= 0$, otherwise. If $\mathcal{R} = 1$, it means that all the blocks with indices in $\mathcal{Q}$ are non-zero. The corresponding weights are thus assigned the value τ. We experimentally found that when τ is a small positive constant (e.g., 10^{-2}), IR-ℓ_2/ℓ_1 algorithms have better performance than

[4]The value 10 is selected empirically. Our experiments show that the performance of the proposed algorithms is not very sensitive to the small changes in the value of the decreasing rate of ρ per iteration.

[5]IR$^{(1)}$-ℓ_2/ℓ_1 can be considered as the adaptation of IRℓ_1 applied to ℓ_2/ℓ_1 minimization rather than ℓ_1 minimization.

[6]IR$^{(2)}$-ℓ_2/ℓ_1 can be considered as the adaptation of ISD applied to ℓ_2/ℓ_1 minimization rather than ℓ_1 minimization.

when $\tau = 0.7$ The weighting strategy in the next iterations is the same as the strategy presented in Algorithm 1.

Our results show that the required number of measurements for the reconstruction decreases by increasing the value of $\mathcal{R}$, and is always smaller than what is needed at the absence of any a priori information.

D. Sufficient Conditions for the Recovery of Sparse Signals by Weighted ℓ_2/ℓ_1 Minimization: Noiseless Scenario ($\alpha = 0$)

In this part, we introduce a sufficient exact recovery condition on the sensing matrix $\mathbf{A}$ so that the solution of (9) coincides with the true signal $\bar{\mathbf{x}}$, if for a given $\Gamma \subseteq \{1, \ldots, L\}$, $\bar{\mathbf{x}}_\Gamma$ is a P-block sparse signal.

Theorem 3: Let $\bar{\mathbf{x}}$ be the true block sparse signal (solution of ℓ_2/ℓ_0 minimization) so that for a given set $\Gamma \subseteq \{1, \ldots, L\}$, $\bar{\mathbf{x}}_\Gamma$ is a P-block sparse sub-vector of $\bar{\mathbf{x}}$ and $\hat{\mathbf{x}}$ is a minimizer of (9) for $\alpha = 0$. Further assume that $\mathbf{A}$ is an $M \times N$ sensing matrix and $\omega_i = 1/(\rho + \|\bar{\mathbf{u}}_i\|_2), \forall i \in \Gamma$, where $\rho > 0$ is a regularization constant. If for all index sets $S \subset \Gamma$ with $|S| \leq P$ and for all $\boldsymbol{\eta} \in \mathcal{N}(\mathbf{A}) \setminus \mathbf{0}$, we have

$$\sum_{i \in S} \|\boldsymbol{\eta}_i\|_2 < \bar{\gamma} \sum_{i \in \Gamma \cap S^c} \|\boldsymbol{\eta}_i\|_2, \tag{13}$$

for some $0 < \bar{\gamma} < 1$, where $\boldsymbol{\eta}_i$ is the ith block of $\boldsymbol{\eta}$ with the block boundaries the same as those of $\bar{\mathbf{x}}$, then $\hat{\mathbf{x}}$ will be the unique minimizer of (9) and $\hat{\mathbf{x}} = \bar{\mathbf{x}}$.[8]

Proof: The true signal $\bar{\mathbf{x}} = [\bar{\mathbf{u}}_1^T, \ldots, \bar{\mathbf{u}}_L^T]^T$ uniquely solves (9) for a given index set Γ if and only if

$$\sum_{i \in \Gamma} \omega_i \|\bar{\mathbf{u}}_i\|_2 < \sum_{i \in \Gamma} \omega_i \|\bar{\mathbf{u}}_i + \boldsymbol{\eta}_i\|_2, \quad \forall \boldsymbol{\eta} \in \mathcal{N}(\mathbf{A}),\ \boldsymbol{\eta} \neq \mathbf{0}. \tag{14}$$

Let $S = \Gamma \cap J$, where J is the index set of non-zero blocks of $\bar{\mathbf{x}}$. Thus, we can write

$$\begin{aligned}
\sum_{i \in \Gamma} \omega_i \|\bar{\mathbf{u}}_i + \boldsymbol{\eta}_i\|_2 &= \sum_{i \in S} \omega_i \|\bar{\mathbf{u}}_i + \boldsymbol{\eta}_i\|_2 + \sum_{i \in \Gamma \cap S^c} \omega_i \|\mathbf{0} + \boldsymbol{\eta}_i\|_2 \\
&= \sum_{i \in S} \omega_i \|\bar{\mathbf{u}}_i + \boldsymbol{\eta}_i\|_2 - \sum_{i \in S} \omega_i \|\bar{\mathbf{u}}_i\|_2 \\
&\quad + \sum_{i \in S} \omega_i \|\boldsymbol{\eta}_i\|_2 + \sum_{i \in S} \omega_i \|\bar{\mathbf{u}}_i\|_2 \\
&\quad - \sum_{i \in S} \omega_i \|\boldsymbol{\eta}_i\|_2 + \sum_{i \in \Gamma \cap S^c} \omega_i \|\boldsymbol{\eta}_i\|_2 \\
&= \sum_{i \in S} \omega_i \underbrace{\left(\|\bar{\mathbf{u}}_i + \boldsymbol{\eta}_i\|_2 - \|\bar{\mathbf{u}}_i\|_2 + \|\boldsymbol{\eta}_i\|_2\right)}_{\geq 0} \\
&\quad + \sum_{i \in S} \omega_i \|\bar{\mathbf{u}}_i\|_2 \\
&\quad + \left(\sum_{i \in \Gamma \cap S^c} \omega_i \|\boldsymbol{\eta}_i\|_2 - \sum_{i \in S} \omega_i \|\boldsymbol{\eta}_i\|_2 \right).
\end{aligned} \tag{15}$$

Since $\sum_{i \in S} \omega_i \|\bar{\mathbf{u}}_i\|_2 = \sum_{i \in \Gamma} \omega_i \|\bar{\mathbf{u}}_i\|_2$ (by the definition of S), it is sufficient to have $\sum_{i \in \Gamma \cap S^c} \omega_i \|\boldsymbol{\eta}_i\|_2 > \sum_{i \in S} \omega_i \|\boldsymbol{\eta}_i\|_2$ for (14) to hold true. To show that this inequality is valid, we can write

$$\sum_{i \in S} \left(\frac{\rho}{\rho + \|\bar{\mathbf{u}}_i\|_2} \right) \|\boldsymbol{\eta}_i\|_2 \overset{(a)}{<} \sum_{i \in S} \|\boldsymbol{\eta}_i\|_2 \overset{(b)}{<} \sum_{i \in \Gamma \cap S^c} \|\boldsymbol{\eta}_i\|_2, \tag{16}$$

where (16a) holds true because $\rho/(\rho + \|\bar{\mathbf{u}}_i\|_2) < 1$ for $i \in S$ and (16b) follows from (13) as $0 < \bar{\gamma} < 1$. By dividing both sides of (16) by $\rho > 0$, and also using the fact that $\|\bar{\mathbf{u}}_i\|_2 = \mathbf{0}$ for $i \in \Gamma \cap S^c$, we can write

$$\sum_{i \in S} \frac{1}{\rho + \|\bar{\mathbf{u}}_i\|_2} \|\boldsymbol{\eta}_i\|_2 < \sum_{i \in \Gamma \cap S^c} \frac{1}{\rho + \|\bar{\mathbf{u}}_i\|_2} \|\boldsymbol{\eta}_i\|_2. \tag{17}$$

Therefore,

$$\sum_{i \in \Gamma \cap S^c} \omega_i \|\boldsymbol{\eta}_i\|_2 > \sum_{i \in S} \omega_i \|\boldsymbol{\eta}_i\|_2. \tag{18}$$

■

One should note that the statement of Theorem 3 is for a fixed set of weights ω_i and a fixed index set Γ. It therefore does not take into account the evolution of weights and the set Γ throughout iterations.

In the following, we discuss the result of Theorem 3 for two special cases of IR$^{(1)}$-ℓ_2/ℓ_1 and IR$^{(2)}$-ℓ_2/ℓ_1.

1) IR$^{(1)}$-ℓ_2/ℓ_1*:* Since in IR$^{(1)}$-ℓ_2/ℓ_1, an (untruncated) weighted ℓ_2/ℓ_1 minimization is solved in each iteration (i.e., $\Gamma = \{1, \ldots, L\}$), the null space property in (13) can be rewritten as

$$\sum_{i \in S} \|\boldsymbol{\eta}_i\|_2 < \bar{\gamma} \sum_{i \in S^c} \|\boldsymbol{\eta}_i\|_2. \tag{19}$$

Comparing the null space property of the sensing matrix in (19) and Inequality (5) for the exact recovery of block sparse signals by ℓ_2/ℓ_1 minimization, we can see that the sufficient conditions on matrix $\mathbf{A}$ for exact reconstruction are the same for both untruncated weighted ℓ_2/ℓ_1 minimization and ℓ_2/ℓ_1 minimization. One should note that Condition (5) is also a necessary condition on matrix $\mathbf{A}$ for exact recovery by ℓ_2/ℓ_1 minimization (see the proof in [27]). This is, however, not the case for the untruncated weighted ℓ_2/ℓ_1 minimization, i.e., Condition (19) is sufficient but not necessary for the success of untruncated weighted ℓ_2/ℓ_1 minimization. From the proof, we can see that Inequality (18) is also another sufficient condition on matrix $\mathbf{A}$ for the exact reconstruction and it is not difficult to see that Inequality (18) with $\Gamma = \{1, \ldots, L\}$ may hold true depending on the values of ω_i, even if (19) is not satisfied. In other words, if $\mathbf{A}$ does not satisfy (19), ℓ_2/ℓ_1 minimization will fail, yet, weighted ℓ_2/ℓ_1 minimization may still succeed. This implies that IR$^{(1)}$-ℓ_2/ℓ_1 minimization would outperform the ℓ_2/ℓ_1 minimization in the recovery of block sparse signals. In fact, our simulation results, presented in Section V, demonstrate that for a given sparsity level, the number of required measurements for an accurate recovery by IR$^{(1)}$-ℓ_2/ℓ_1 is significantly smaller than that of the ℓ_2/ℓ_1 minimization.

[7]This is particularly the case when the blocks with indices in $\mathcal{Q}$ have relatively small ℓ_2 norms compared to non-zero blocks whose indices are not in $\mathcal{Q}$. In such circumstances, not having the blocks with indices in $\mathcal{Q}$ as part of the objective function will result in over-estimating the ℓ_2 norm of those blocks in the first iteration and will consequently deteriorate the performance of the algorithm.

[8]Note that (13) implies the linear independence of all the columns of $\mathbf{A}$ in blocks whose indices are in Γ^c.

2) $\mathrm{IR}^{(2)}$-ℓ_2/ℓ_1*:* In Theorem 3, with setting $\omega_i = 1, \forall i \in \Gamma$, the exact recovery condition on $\mathbf{A}$ for the truncated ℓ_2/ℓ_1 minimization is also (13). This theorem implies that depending on the set Γ, truncated ℓ_2/ℓ_1 minimization, compared to ℓ_2/ℓ_1 minimization, can reconstruct the signals with a larger number of non-zero blocks. In the exact reconstruction of block sparse signals by the standard ℓ_2/ℓ_1 minimization, the total number of non-zero blocks in the signal can not be greater than P (i.e., exact recovery is possible only when $\bar{\mathbf{x}}$ is a P-block sparse signal). In the truncated ℓ_2/ℓ_1 minimization, however, it is sufficient for $\bar{\mathbf{x}}_\Gamma$ to be a P-block sparse signal, which means $\bar{\mathbf{x}}$ itself can have more than P non-zero blocks.

In the $\mathrm{IR}^{(2)}$-ℓ_2/ℓ_1 algorithm, the set Γ is updated in each iteration so that the indices of the blocks with the largest ℓ_2-norm are most likely to be in the set Γ^c. This implies that, as the iterations progress, the signal constrained to the support set Γ is likely to be sparser than the original signal, thus improving the performance compared to the standard ℓ_2/ℓ_1 minimization. This will be verified with our simulation results, presented in Section V.

E. Error Bound on the Recovery of Approximately Block Sparse Signals With Weighted ℓ_2/ℓ_1 Minimization: Noiseless Scenario ($\alpha = 0$)

In this section, we will consider the case of so-called "approximately" block sparse signals and prove that when $\bar{\mathbf{x}}_\Gamma$ is approximately P-block sparse, for a given set Γ, the difference between the solution of weighted ℓ_2/ℓ_1 minimization with $\alpha = 0$, $\hat{\mathbf{x}}$, and the true signal, $\bar{\mathbf{x}}$, in ℓ_2/ℓ_1 norm (i.e., $\sum_{i=1}^{L} \|\hat{\mathbf{u}}_i - \bar{\mathbf{u}}_i\|_2$), is bounded from above.

Theorem 4: Let $\bar{\mathbf{x}}$ be the true signal (solution of ℓ_2/ℓ_0 minimization) so that for a given set $\Gamma \subseteq \{1, \ldots, L\}$, $\bar{\mathbf{x}}_\Gamma$ is an approximately P-block sparse sub-vector of $\bar{\mathbf{x}}$ with parameter δ_Γ. Further, assume that $\mathbf{A}$ is an $M \times N$ sensing matrix such that for all sets $\Upsilon \subseteq \{1, \ldots, L\}$ with $|\Upsilon| = |\Gamma|$, all sets $S \subset \Upsilon$ with $|S| = P$ and all $\boldsymbol{\eta} \in \mathcal{N}(\mathbf{A}) \setminus \mathbf{0}$, we have

$$\sum_{i \in S} \|\boldsymbol{\eta}_i\|_2 < \bar{\gamma} \sum_{i \in \Upsilon \cap S^c} \|\boldsymbol{\eta}_i\|_2, \tag{20}$$

for some $0 < \bar{\gamma} < 1$, where $\boldsymbol{\eta}_i$ is the ith block of $\boldsymbol{\eta}$ with the block boundaries the same as those of $\bar{\mathbf{x}}$. Let π be the proper permutation with respect to the set $\{\|\bar{\mathbf{u}}_i\|_2\}_{i \in \Gamma}$. If $\hat{\mathbf{x}}$ is a minimizer of (9) for $\alpha = 0$ with $\omega_i = 1/(\rho + \|\bar{\mathbf{u}}_i\|_2), \forall i \in \Gamma$, where $\rho > 0$ is a regularization constant, then

$$\sum_{i=1}^{L} \|\hat{\mathbf{u}}_i - \bar{\mathbf{u}}_i\|_2 \leq 2C\delta_\Gamma, \tag{21}$$

with

$$C = \frac{\rho + \kappa}{\rho(1 - \Omega)} \left((1 + \bar{\gamma}) + \bar{\gamma} \max\{1, \frac{|\Gamma^c|}{P}\} \right), \tag{22}$$

$\Omega = \bar{\gamma}(\rho + \kappa)/(\rho + \mu)$, $\mu = \|\bar{\mathbf{u}}_{\pi(P)}\|_2$ and $\kappa = \|\bar{\mathbf{u}}_{\pi(P+1)}\|_2$.

Proof: Outline: Since $\Gamma \subseteq \{1, \ldots, L\}$, we can write $\sum_{i=1}^{L} \|\hat{\mathbf{u}}_i - \bar{\mathbf{u}}_i\|_2 = \sum_{i \in \Gamma} \|\hat{\mathbf{u}}_i - \bar{\mathbf{u}}_i\|_2 + \sum_{i \in \Gamma^c} \|\hat{\mathbf{u}}_i - \bar{\mathbf{u}}_i\|_2$. We first find upper bounds on $\sum_{i \in \Gamma} \|\hat{\mathbf{u}}_i - \bar{\mathbf{u}}_i\|_2$ and $\sum_{i \in \Gamma^c} \|\hat{\mathbf{u}}_i - \bar{\mathbf{u}}_i\|_2$, respectively. We then combine them to obtain an upper bound on $\sum_{i=1}^{L} \|\hat{\mathbf{u}}_i - \bar{\mathbf{u}}_i\|_2$.

Upper bound on $\sum_{i \in \Gamma} \|\hat{\mathbf{u}}_i - \bar{\mathbf{u}}_i\|_2$: Since $\hat{\mathbf{x}}$ is the solution of (9), we have $\sum_{i \in \Gamma} \omega_i \|\hat{\mathbf{u}}_i\|_2 \leq \sum_{i \in \Gamma} \omega_i \|\bar{\mathbf{u}}_i\|_2$. In addition, let $\boldsymbol{\eta} = \hat{\mathbf{x}} - \bar{\mathbf{x}}$, which means $\boldsymbol{\eta} \in \mathcal{N}(\mathbf{A})$. Therefore we have $\sum_{i \in \Gamma} \omega_i \|\bar{\mathbf{u}}_i + \boldsymbol{\eta}_i\|_2 \leq \sum_{i \in \Gamma} \omega_i \|\bar{\mathbf{u}}_i\|_2$. Further, for the set $S \subset \Gamma$, we can write

$$\begin{aligned}
\sum_{i \in \Gamma} \omega_i \|\bar{\mathbf{u}}_i\|_2 &\geq \sum_{i \in \Gamma} \omega_i \|\bar{\mathbf{u}}_i + \boldsymbol{\eta}_i\|_2 \\
&= \sum_{i \in S} \omega_i \|\bar{\mathbf{u}}_i + \boldsymbol{\eta}_i\|_2 + \sum_{i \in \Gamma \cap S^c} \omega_i \|\bar{\mathbf{u}}_i + \boldsymbol{\eta}_i\|_2 \\
&\geq \sum_{i \in S} \omega_i \|\bar{\mathbf{u}}_i\|_2 - \sum_{i \in \Gamma \cap S^c} \omega_i \|\bar{\mathbf{u}}_i\|_2 + \sum_{i \in \Gamma \cap S^c} \omega_i \|\boldsymbol{\eta}_i\|_2 \\
&\quad - \sum_{i \in S} \omega_i \|\boldsymbol{\eta}_i\|_2,
\end{aligned} \tag{23}$$

where the last inequality follows from the triangle inequality. Since $\sum_{i \in \Gamma} \omega_i \|\bar{\mathbf{u}}_i\|_2 = \sum_{i \in S} \omega_i \|\bar{\mathbf{u}}_i\|_2 + \sum_{i \in \Gamma \cap S^c} \omega_i \|\bar{\mathbf{u}}_i\|_2$ and similarly, $\sum_{i \in \Gamma} \omega_i \|\boldsymbol{\eta}_i\|_2 = \sum_{i \in S} \omega_i \|\boldsymbol{\eta}_i\|_2 + \sum_{i \in \Gamma \cap S^c} \omega_i \|\boldsymbol{\eta}_i\|_2$, Inequality (23) can be rewritten as follows

$$2 \sum_{i \in \Gamma \cap S^c} \omega_i \|\bar{\mathbf{u}}_i\|_2 \geq \sum_{i \in \Gamma} \omega_i \|\boldsymbol{\eta}_i\|_2 - 2 \sum_{i \in S} \omega_i \|\boldsymbol{\eta}_i\|_2. \tag{24}$$

Now, let S be the set $\{\pi(i)\}_{1 \leq i \leq P}$ (i.e., $\Gamma \cap S^c = \{\pi(i)\}_{P+1 \leq i \leq |\Gamma|}$), then we have

$$\sum_{i \in S} \omega_i \|\boldsymbol{\eta}_i\|_2 = \sum_{i \in S} \frac{1}{\rho + \|\bar{\mathbf{u}}_i\|_2} \|\boldsymbol{\eta}_i\|_2 \leq \frac{1}{\rho + \mu} \sum_{i \in S} \|\boldsymbol{\eta}_i\|_2, \tag{25}$$

since for all $i \in S$, $\|\bar{\mathbf{u}}_i\|_2 \geq \mu$ (by the definition of μ). On the other hand, for all $i \in S^c$, $\|\bar{\mathbf{u}}_i\|_2 \leq \kappa$ (where $\kappa = \|\bar{\mathbf{u}}_{\pi(P+1)}\|_2$, by definition). Thus, we have

$$\sum_{i \in \Gamma \cap S^c} \omega_i \|\boldsymbol{\eta}_i\|_2 = \sum_{i \in \Gamma \cap S^c} \frac{1}{\rho + \|\bar{\mathbf{u}}_i\|_2} \|\boldsymbol{\eta}_i\|_2 \geq \frac{1}{\rho + \kappa} \sum_{i \in \Gamma \cap S^c} \|\boldsymbol{\eta}_i\|_2. \tag{26}$$

Let $\Upsilon = \Gamma$ in (20). By combining (20), (25) and (26), it can be shown that

$$\sum_{i \in S} \omega_i \|\boldsymbol{\eta}_i\|_2 < \Omega \sum_{i \in \Gamma \cap S^c} \omega_i \|\boldsymbol{\eta}_i\|_2 , \tag{27}$$

where $\Omega = \bar{\gamma}(\rho + \kappa)/(\rho + \mu)$. Again by rewriting $\sum_{i \in \Gamma \cap S^c} \omega_i \|\boldsymbol{\eta}_i\|_2$ as $\sum_{i \in \Gamma} \omega_i \|\boldsymbol{\eta}_i\|_2 - \sum_{i \in S} \omega_i \|\boldsymbol{\eta}_i\|_2$ in the right hand side of (27), it can be easily shown that

$$\frac{1 - \Omega}{\Omega} \sum_{i \in S} \omega_i \|\boldsymbol{\eta}_i\|_2 < \sum_{i \in \Gamma} \omega_i \|\boldsymbol{\eta}_i\|_2 - 2 \sum_{i \in S} \omega_i \|\boldsymbol{\eta}_i\|_2, \tag{28}$$

since $\kappa \leq \mu$ and $0 < \bar{\gamma} < 1$ and thus $0 < \Omega < 1$. Now, Inequality (28) can be written as

$$\begin{aligned}
\sum_{i \in S} \omega_i \|\boldsymbol{\eta}_i\|_2 &< \frac{\Omega}{1 - \Omega} \left(\sum_{i \in \Gamma} \omega_i \|\boldsymbol{\eta}_i\|_2 - 2 \sum_{i \in S} \omega_i \|\boldsymbol{\eta}_i\|_2 \right) \\
&\leq \frac{2\Omega}{1 - \Omega} \sum_{i \in \Gamma \cap S^c} \omega_i \|\bar{\mathbf{u}}_i\|_2,
\end{aligned} \tag{29}$$

where the last inequality follows from (24). Therefore, we can write

$$\begin{aligned}\sum_{i\in\Gamma\cap S^c}\|\boldsymbol{\eta}_i\|_2 &\overset{(a)}{\leq}(\rho+\kappa)\sum_{i\in\Gamma\cap S^c}\omega_i\|\boldsymbol{\eta}_i\|_2\\ &\overset{(b)}{\leq}(\rho+\kappa)\left(\sum_{i\in S}\omega_i\|\boldsymbol{\eta}_i\|_2+2\sum_{i\in\Gamma\cap S^c}\omega_i\|\bar{\mathbf{u}}_i\|_2\right)\\ &\overset{(c)}{<}(\rho+\kappa)\left(\frac{2\Omega}{1-\Omega}\sum_{i\in\Gamma\cap S^c}\omega_i\|\bar{\mathbf{u}}_i\|_2+2\sum_{i\in\Gamma\cap S^c}\omega_i\|\bar{\mathbf{u}}_i\|_2\right)\\ &\overset{(d)}{\leq}\frac{2(\rho+\kappa)}{1-\Omega}\frac{1}{\rho}\sum_{i\in\Gamma\cap S^c}\|\bar{\mathbf{u}}_i\|_2, \end{aligned}\tag{30}$$

where (30a), (30b), (30c) and (30d), respectively, follow from (26), (23), (29) and the fact that $\omega_i\leq 1/\rho,\forall i$. By setting $\Upsilon=\Gamma$ in (20) and using (30), we have

$$\begin{aligned}\sum_{i\in\Gamma}\|\boldsymbol{\eta}_i\|_2 &<(1+\bar{\gamma})\sum_{i\in\Gamma\cap S^c}\|\boldsymbol{\eta}_i\|_2\\ &<\frac{2(1+\bar{\gamma})(\rho+\kappa)}{\rho(1-\Omega)}\sum_{i\in\Gamma\cap S^c}\|\bar{\mathbf{u}}_i\|_2. \end{aligned}\tag{31}$$

Finally, by setting $\boldsymbol{\eta}_i=\hat{\mathbf{u}}_i-\bar{\mathbf{u}}_i$, in (31), the upper bound on $\sum_{i\in\Gamma}\|\hat{\mathbf{u}}_i-\bar{\mathbf{u}}_i\|_2$ is derived as

$$\sum_{i\in\Gamma}\|\hat{\mathbf{u}}_i-\bar{\mathbf{u}}_i\|_2<\frac{2(1+\bar{\gamma})(\rho+\kappa)}{\rho(1-\Omega)}\sum_{i\in\Gamma\cap S^c}\|\bar{\mathbf{u}}_i\|_2.\tag{32}$$

Upper bound on $\sum_{i\in\Gamma^c}\|\hat{\mathbf{u}}_i-\bar{\mathbf{u}}_i\|_2$: We study two cases:

Case 1: $|\Gamma^c|\leq P$. We can find $S_1\subset S$ such that $|S_1\cup\Gamma^c|=P$. Therefore, we can write:

$$\sum_{i\in\Gamma^c}\|\hat{\mathbf{u}}_i-\bar{\mathbf{u}}_i\|_2\leq\sum_{i\in S_1\cup\Gamma^c}\|\hat{\mathbf{u}}_i-\bar{\mathbf{u}}_i\|_2\overset{(a)}{<}\bar{\gamma}\sum_{i\in\Gamma\cap S^c}\|\hat{\mathbf{u}}_i-\bar{\mathbf{u}}_i\|_2,\tag{33}$$

where (33a) follows from the null space property for $\mathbf{A}$ in (20). Clearly, $\Upsilon=S_1\cup\Gamma^c\cup(\Gamma\cap S^c)$ and $|\Upsilon|=|\Gamma|$.

Case 2: $|\Gamma^c|>P$. Let $S_2\subset\Gamma^c$ be the index set of the P largest $\|\hat{\mathbf{u}}_i-\bar{\mathbf{u}}_i\|_2$ for $i\in\Gamma^c$. Therefore, we have

$$\frac{1}{|\Gamma^c|}\sum_{i\in\Gamma^c}\|\hat{\mathbf{u}}_i-\bar{\mathbf{u}}_i\|_2<\frac{1}{P}\sum_{i\in S_2}\|\hat{\mathbf{u}}_i-\bar{\mathbf{u}}_i\|_2.\tag{34}$$

On the other hand, from the null space property (20), by setting $\Upsilon=S_2\cup(\Gamma\cap S^c)$, we have

$$\sum_{i\in S_2}\|\hat{\mathbf{u}}_i-\bar{\mathbf{u}}_i\|_2<\bar{\gamma}\sum_{i\in\Gamma\cap S^c}\|\hat{\mathbf{u}}_i-\bar{\mathbf{u}}_i\|_2.\tag{35}$$

Therefore, from (34) and (35), we obtain

$$\sum_{i\in\Gamma^c}\|\hat{\mathbf{u}}_i-\bar{\mathbf{u}}_i\|_2<\bar{\gamma}\frac{|\Gamma^c|}{P}\sum_{i\in\Gamma\cap S^c}\|\hat{\mathbf{u}}_i-\bar{\mathbf{u}}_i\|_2.\tag{36}$$

Combining Case 1, i.e., (33) and Case 2, i.e., (36), we have

$$\sum_{i\in\Gamma^c}\|\hat{\mathbf{u}}_i-\bar{\mathbf{u}}_i\|_2<\max\{1,\frac{|\Gamma^c|}{P}\}\bar{\gamma}\sum_{i\in\Gamma\cap S^c}\|\hat{\mathbf{u}}_i-\bar{\mathbf{u}}_i\|_2.\tag{37}$$

By setting $\boldsymbol{\eta}_i=\hat{\mathbf{u}}_i-\bar{\mathbf{u}}_i$ in (30) and using (37), we obtain the following upper bound on $\sum_{i\in\Gamma^c}\|\hat{\mathbf{u}}_i-\bar{\mathbf{u}}_i\|_2$:

$$\sum_{i\in\Gamma^c}\|\hat{\mathbf{u}}_i-\bar{\mathbf{u}}_i\|_2<\frac{2(\rho+\kappa)\bar{\gamma}}{\rho(1-\Omega)}\max\{1,\frac{|\Gamma^c|}{P}\}\sum_{i\in\Gamma\cap S^c}\|\bar{\mathbf{u}}_i\|_2.\tag{38}$$

Finally, by using the two upper bounds of (32) and (38), we can write

$$\begin{aligned}\sum_{i=1}^{L}\|\hat{\mathbf{u}}_i-\bar{\mathbf{u}}_i\|_2&=\sum_{i\in\Gamma}\|\hat{\mathbf{u}}_i-\bar{\mathbf{u}}_i\|_2+\sum_{i\in\Gamma^c}\|\hat{\mathbf{u}}_i-\bar{\mathbf{u}}_i\|_2\\ &\leq 2C\sum_{i\in\Gamma\cap S^c}\|\bar{\mathbf{u}}_i\|_2,\end{aligned}\tag{39}$$

where

$$C=\frac{\rho+\kappa}{\rho(1-\Omega)}\left(1+\bar{\gamma}+\bar{\gamma}\max\{1,\frac{|\Gamma^c|}{P}\}\right).\tag{40}$$

Since S is assumed to be the index set of the P largest elements of $\{\|\bar{\mathbf{u}}_i\|_2\}_{i\in\Gamma}$, based on the definition of $\bar{\mathbf{x}}_\Gamma$, we have $\sum_{i\in\Gamma\cap S^c}\|\bar{\mathbf{u}}_i\|_2\leq\delta_\Gamma$. This together with (39) results in (21). ■

Similar to Theorem 3, Theorem 4 is limited to fixed values of ω_i and a fixed set Γ. The evolution of these entities with iterations is not considered.

In the following, we discuss the result of Theorem 4 for the special cases of $\mathrm{IR}^{(1)}$-ℓ_2/ℓ_1 and $\mathrm{IR}^{(2)}$-ℓ_2/ℓ_1.

1) $\mathrm{IR}^{(1)}$-ℓ_2/ℓ_1*:* When $\Gamma=\{1,\ldots,L\}$, the null space property for weighted ℓ_2/ℓ_1 in (20) is reduced to that of ℓ_2/ℓ_1 in (7) and the upper bound in (32) on $\sum_{i\in\Gamma}\|\hat{\mathbf{u}}_i-\bar{\mathbf{u}}_i\|_2$ is reduced to

$$\sum_{i=1}^{L}\|\hat{\mathbf{u}}_i-\bar{\mathbf{u}}_i\|_2\leq 2C^{(1)}\sum_{i\in S^c}\|\bar{\mathbf{u}}_i\|_2\leq 2C^{(1)}\delta,\tag{41}$$

where $C^{(1)}=(1+\bar{\gamma})(\rho+\kappa)/(\rho(1-\Omega))$ and the last inequality follows from the assumption that S is the index set of the P largest elements of $\{\|\bar{\mathbf{u}}_i\|_2\}_{1\leq i\leq L}$. It can be easily shown that, for sufficiently large $\bar{\gamma}$, i.e., $\bar{\gamma}>(\kappa(\rho+\mu))/(\mu(\rho+\kappa))$, the upper bound in (6) is larger than that of (41). As can be seen in (7), for a given $\mathbf{A}$, $\bar{\gamma}$ is monotonically increasing in P. Therefore, larger values of $\bar{\gamma}$ are of interest. This implies that the proposed algorithm would be superior to the ℓ_2/ℓ_1 minimization in the reconstruction of approximately block sparse signals. This is also verified by our simulation results presented in Section V.

2) $\mathrm{IR}^{(2)}$-ℓ_2/ℓ_1*:* In truncated ℓ_2/ℓ_1 minimization, since $\omega_i=1,\forall i\in\Gamma$, we have

$$\begin{aligned}\sum_{i=1}^{L}\|\boldsymbol{\eta}_i\|_2&=\sum_{i\in S}\|\boldsymbol{\eta}_i\|_2+\sum_{i\in\Gamma\cap S^c}\|\boldsymbol{\eta}_i\|_2+\sum_{i\in\Gamma^c}\|\boldsymbol{\eta}_i\|_2\\ &\overset{(a)}{\leq}\left(1+\bar{\gamma}+\bar{\gamma}\max\{1,\frac{|\Gamma^c|}{P}\}\right)\sum_{i\in\Gamma\cap S^c}\|\boldsymbol{\eta}_i\|_2\\ &\overset{(b)}{<}\frac{2}{1-\bar{\gamma}}\left(1+\bar{\gamma}+\bar{\gamma}\max\{1,\frac{|\Gamma^c|}{P}\}\right)\sum_{i\in\Gamma\cap S^c}\|\bar{\mathbf{u}}_i\|_2\\ &\overset{(c)}{\leq}2C^{(2)}\delta_\Gamma,\end{aligned}\tag{42}$$

where $C^{(2)}=(1+\bar{\gamma}+\bar{\gamma}\max\{1,|\Gamma^c|/P\})/(1-\bar{\gamma})$. Inequality (42a) follows from (37) and the null space property given in (20) as $\Upsilon=\Gamma$, Inequality (42b) follows from (24) when $\omega_i=$

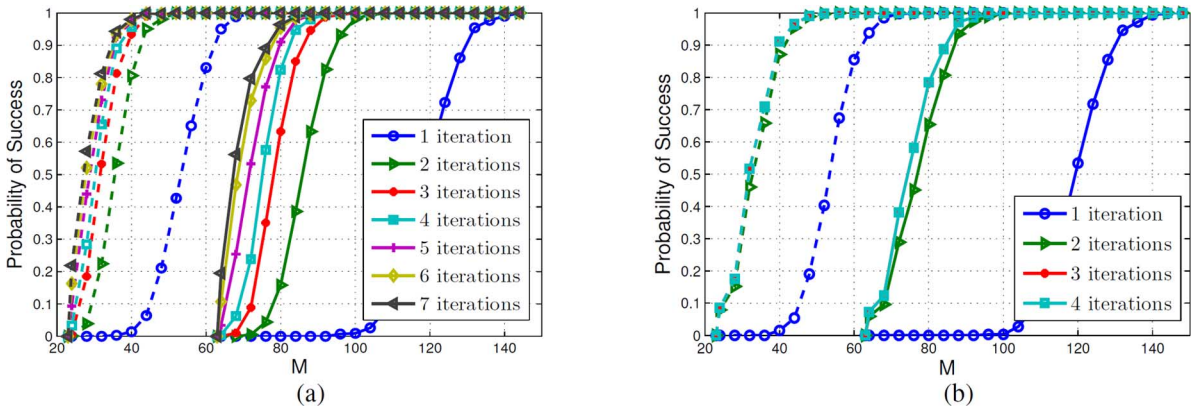

Fig. 1. Probability of success of (a) IR$^{(1)}$-ℓ_2/ℓ_1, and (b) IR$^{(2)}$-ℓ_2/ℓ_1 for different number of iterations versus M for $N = 256$, $L = 32$ and Signal Model I. Solid and dashed curves correspond to $P = 8$ and $P = 3$, respectively.

$1, \forall i \in \Gamma$, and the null space property of (20) as $\Upsilon = \Gamma$, and finally, Inequality (42c) follows from the assumption that set $S \subset \Gamma$ is the index set of the P largest elements of $\{\|\bar{\mathbf{u}}_i\|_2\}_{i\in\Gamma}$.

Unlike the case for the upper bound of (41), it is not easy to compare the error bound of (42) with that of (6). Although, $C^{(2)}$ is clearly larger than $(1+\bar{\gamma})/(1-\bar{\gamma})$, it is not as easy to compare δ with δ_Γ. Since $\bar{\mathbf{x}}_\Gamma$ is approximately P-block sparse, $\bar{\mathbf{x}}$ would be an approximately P'-block sparse with $P' > P$. Therefore, according to the definition of δ and δ_Γ, we can not easily compare them. In IR$^{(2)}$-ℓ_2/ℓ_1 algorithm, Γ is iteratively updated so that the blocks with larger ℓ_2-norm are likely to be in Γ^c. Therefore, we expect that δ_Γ would be much smaller than δ such that the effect of $C^{(2)}$ would be dominated by the effect of δ_Γ. Particularly, when the ℓ_2-norm of non-zero blocks decays fast enough if sorted in descending order, $\delta_\Gamma \ll \delta$. This would imply that the upper bound in (42) can be smaller than (6). Thus, IR$^{(2)}$-ℓ_2/ℓ_1 may outperform the ℓ_2/ℓ_1 minimization. Our simulation results in Section V, in fact, verify the superior performance of IR$^{(2)}$-ℓ_2/ℓ_1 minimization in the reconstruction of approximately block sparse signals compared to the standard ℓ_2/ℓ_1 minimization.

V. Simulation Results

In this section, we present some simulation results to demonstrate the effectiveness of IR-ℓ_2/ℓ_1 algorithms with focusing on IR$^{(1)}$-ℓ_2/ℓ_1 and IR$^{(2)}$-ℓ_2/ℓ_1 as recovery algorithms for block sparse signals in both noiseless and noisy scenarios. We also provide comparisons with other recovery algorithms. Subsections V-A and V-B of this section are concerned with the recovery of signals with real elements in noiseless and noisy scenarios, respectively. The application of the proposed algorithms to the problem of wideband spectrum sensing, in which the signal is complex, can be found in [51]. In Subsection V-A, we first discuss the recovery of ideally block sparse signals in Part V-A1, followed by that of approximately block sparse signals (Part V-A2) and recovery at the presence of a priori information about the location of non-zero blocks (Part V-A3). In Part V-A1 itself, the performance of the proposed algorithms, the role of block length on the performance, the effect of signal model on recovery, comparison with other recovery algorithms, and other reweighted iterative algorithms are discussed, respectively.

We consider a signal with dimension N, which consists of L non-overlapping, consecutive and equal-sized blocks. The measurement matrix is randomly constructed with i.i.d. elements, where each element is a zero-mean Gaussian random variable with variance $1/N$. We also assume that the basis matrix $\mathbf{\Psi}$ is the $N \times N$ identity matrix $\mathbf{I}_N$. Therefore, $\mathbf{A} = \Phi$. For IR$^{(2)}$-ℓ_2/ℓ_1, we use $\beta = 4$ for noiseless scenarios (Subsection V-A) and $\beta = 2$ for noisy scenarios (Subsection V-B).

Each experiment is repeated 1000 times and each trial for an experiment consists of (a) randomly creating a signal $\bar{\mathbf{x}}$, (b) constructing a random $M \times N$ measurement matrix, (c) creating M measurements accordingly (and adding noise to each measurement in noisy scenarios), and (d) obtaining $\hat{\mathbf{x}}$ using the reconstruction algorithm.

A. Noiseless Scenarios

In noiseless scenarios, we declare "success" in the recovery of a signal when $\sum_{i=1}^{L} \|\hat{\mathbf{u}}_i - \bar{\mathbf{u}}_i\|_2 / \sum_{i=1}^{L} \|\bar{\mathbf{u}}_i\|_2 \leq 10^{-3}$.

1) Ideally Block Sparse Signals: To model a P-block sparse signal, we select P blocks (out of L) randomly and assign the entries of these blocks using a zero-mean Gaussian distribution with variance σ_s^2 in an i.i.d. fashion. The entries of the other blocks are set to zero. We set $\sigma_s^2 = 1$, and then normalize the signal such that $\sum_{i=1}^{L} \|\bar{\mathbf{u}}_i\|_2 = 1$. This model is referred to as Model I.

Performance of IR-ℓ_2/ℓ_1 Algorithms: In this part, we consider signals with $N = 256$ and $L = 32$. Fig. 1 shows the probability of success of IR$^{(1)}$-ℓ_2/ℓ_1 and IR$^{(2)}$-ℓ_2/ℓ_1 versus M for different number of iterations and two sparsity ratios of 25% ($K = 64$ and $P = 8$) and $\approx$ 10% ($K = 24$ and $P = 3$). We observe that most of the benefit comes from the first few iterations. Therefore, the added computational cost compared to the standard ℓ_2/ℓ_1 minimization is rather small. Furthermore, as seen in these figures, the convergence speed of IR$^{(2)}$-ℓ_2/ℓ_1 is, in general, higher than that of IR$^{(1)}$-ℓ_2/ℓ_1. While IR$^{(2)}$-ℓ_2/ℓ_1 converges in 2–3 iterations, this number for IR$^{(1)}$-ℓ_2/ℓ_1 is about 5–6. We also see that for a fixed N and a fixed ratio of M/K, the probability of success increases with M (or K) in both iterative algorithms. For example, in Fig. 1(a), when $M = 30$ and $P = 3$, the probability of success is 0.6 after 5 iterations while when $M = 80$ and $P = 8$, the probability of success is more than 0.9 after 5 iterations. In Fig. 1(b), also, the probability of

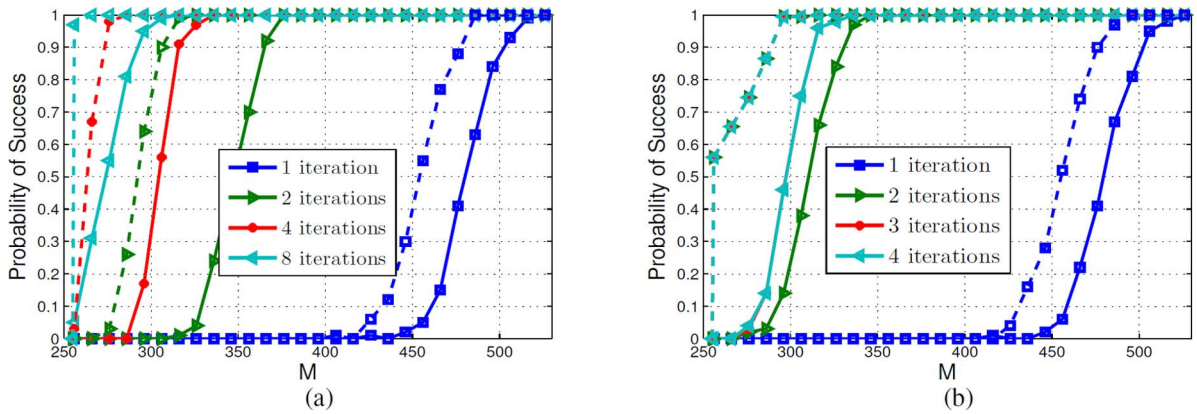

Fig. 2. Probability of success of (a) IR$^{(1)}$-ℓ_2/ℓ_1, and (b) IR$^{(2)}$-ℓ_2/ℓ_1 for different number of iterations versus M for two different block lengths of 8 and 32 ($N = 1024$, $K = 256$ and Signal Model I). Solid and dashed curves correspond to $L = 128$ (block length of $N/L = 8$) and $L = 32$ (block length of $N/L = 32$), respectively.

success is 0.3 after 2 iterations with $M = 30$ and $P = 3$, while the probability of success is 0.65 after 2 iterations with $M = 80$ and $P = 8$.

Effect of Block Length on Recovery Performance: Fig. 2 demonstrates the performance of IR$^{(1)}$-ℓ_2/ℓ_1 and IR$^{(2)}$-ℓ_2/ℓ_1 when $N = 1024$, $K = 256$ and thus the sparsity ratio $K/N = P/L$ is 0.25. Two different block lengths N/L of 8 and 32 are considered in Fig. 2. As demonstrated in [27], increasing the block length leads to an improvement in the recovery performance of ℓ_2/ℓ_1 minimization. Since IR-ℓ_2/ℓ_1 algorithms execute ℓ_2/ℓ_1 minimization in the first iteration, a better performance of the latter would imply a better performance for the former. In addition, this general trend is somehow expected since between the two signals with the same dimension and sparsity, the one with smaller blocks on average contains more uncertainty. In fact, Fig. 2 shows that the performance of both iterative algorithms improves with increasing the block length. For instance, for the larger block length of 32, in the 8th iteration, IR$^{(1)}$-ℓ_2/ℓ_1 succeeds with overwhelming probability in reconstruction of the block sparse signals when the number of measurements is only 4% more than the theoretical limit of K, i.e., 266 (Fig. 2(a)). In addition, with $M = K = 256$, the probability of success of 0.97 can be achieved when the block length is 32 ($P = 8$).

Moreover, Fig. 2(b) shows that IR$^{(2)}$-ℓ_2/ℓ_1 converges even faster as the length of the blocks increases. This is due to the fact that by increasing the block length, the difference between the ℓ_2-norm of non-zero blocks decreases on average.

Effect of Signal Model on IR-ℓ_2/ℓ_1 Algorithms: To investigate the effect of signal model on the performance of the recovery algorithms, we use a second signal model, referred to as Model II. Similar to Model I, P out of L blocks are chosen randomly as non-zero blocks. The elements of the pth non-zero block, however, follow a zero mean Gaussian distribution with variance σ_p^2 (i.e., $\mathcal{N}(0, \sigma_p^2)$), where σ_p is the pth element of a random permutation of the sequence $\{i^{-1/\theta}\}_{i=1}^{P}$ for a fixed θ. Clearly, increasing θ results in decreasing the decaying rate of the ℓ_2-norm of non-zero blocks. We set $\theta = 0.7$, and normalize $\bar{\mathbf{x}}$ so that $\sum_{i=1}^{L} \|\bar{\mathbf{u}}_i\|_2 = 1$. In Figs. 3(a) and 3(b), we have shown ℓ_2-norm of blocks of a realization of a block sparse signal with $N = 256$, $L = 32$ and $P = 8$ that follows Model I and II, respectively. We observe that in Model II, ℓ_2-norm of the non-zero blocks decays faster than Model I, if sorted in descending order. We found that when the ℓ_2-norm of non-zero blocks decays faster, the convergence speed of IR$^{(2)}$-ℓ_2/ℓ_1 would be slower while the convergence speed of IR$^{(1)}$-ℓ_2/ℓ_1 does not change much. Therefore, we set $j_{\max}$ to 5 for both IR$^{(1)}$-ℓ_2/ℓ_1 and IR$^{(2)}$-ℓ_2/ℓ_1 for Signal Model II. Fig. 4 shows the probability of success of iterative reweighted algorithms and the standard ℓ_2/ℓ_1 minimization for the two signal models. Unlike the ℓ_2/ℓ_1 minimization, the performance of the proposed algorithms depends on the signal model. In particular,

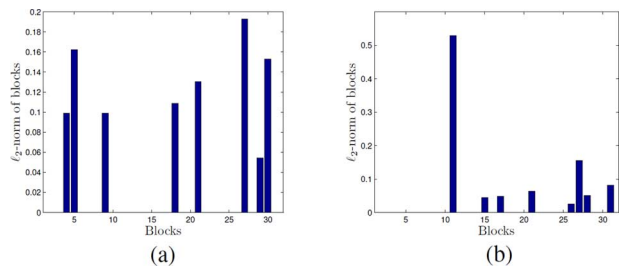

Fig. 3. ℓ_2-norm of the blocks in one realization of the signal for Signal Models I and II ($N = 256$, $K = 64$, $L = 32$ and $P = 8$). (a) Model I (b) Model II.

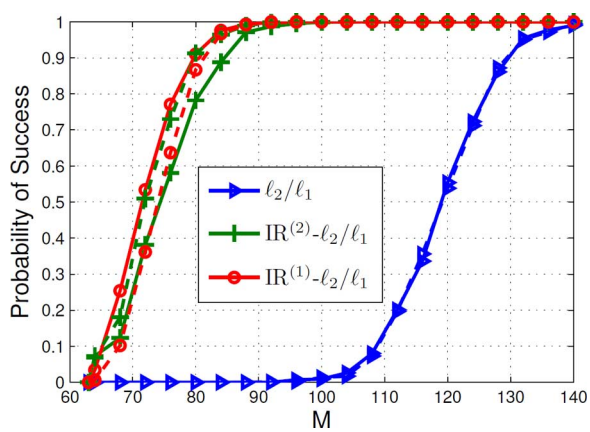

Fig. 4. Probability of success of different algorithms for Signal Models I and II versus M ($N = 256$, $K = 64$, $L = 32$ and $P = 8$). Solid and dashed lines correspond to Model I and Model II, respectively.

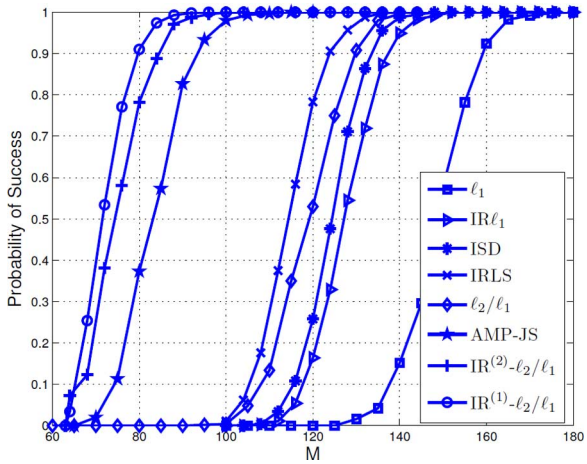

Fig. 5. Probability of success of different algorithms versus M ($N = 256$, $K = 64$, $L = 32$, $P = 8$ and Signal Model I).

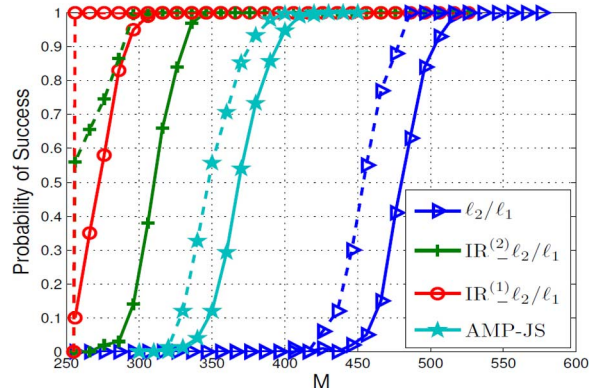

Fig. 6. Probability of success of different recovery algorithms versus M for block lengths of $N/L = 8$ (solid line) and 32 (dashed line) ($N = 1024$, $K = 256$ and Signal Model I). For IR$^{(1)}$-ℓ_2/ℓ_1 and IR$^{(2)}$-ℓ_2/ℓ_1, $j_{\max}$ is selected to be 8 and 3, respectively.

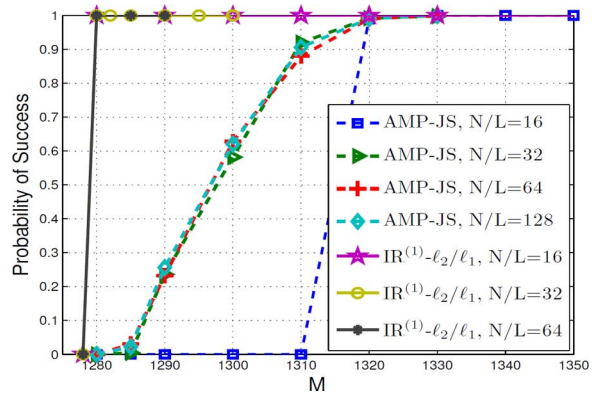

Fig. 7. Probability of success of AMP-JS and IR$^{(1)}$-ℓ_2/ℓ_1 for different block lengths versus M ($N = 5120$, $K = 1280$ and Signal Model I).

IR$^{(2)}$-ℓ_2/ℓ_1 performs better on Model-II signals, while the trend for IR$^{(1)}$-ℓ_2/ℓ_1 is opposite. In fact, the results show that for signals with Model II, IR$^{(2)}$-ℓ_2/ℓ_1 outperforms IR$^{(1)}$-ℓ_2/ℓ_1.

In the IR$^{(2)}$-ℓ_2/ℓ_1 algorithm, the threshold $\epsilon^{(j)}$ decreases as the iterations proceed. Therefore, for Signal Model I, after a very small number of iterations, there would not be any non-zero blocks left in $\bar{\mathbf{x}}_\Gamma$. This thus implies that the performance will not be improved in the subsequent iterations. For Signal Model II, however, the performance would still be improved by increasing the number of iterations since it is more likely to have some non-zero blocks left in $\bar{\mathbf{x}}_\Gamma$. On the other hand, a slight deterioration in the recovery performance of IR$^{(1)}$-ℓ_2/ℓ_1 is expected with Signal Model II. This is due to the zero blocks that are likely to be estimated as non-zero blocks with small ℓ_2-norm, in the early iterations. In the subsequent iterations, those zero blocks and non-zero blocks with small ℓ_2-norm are weighted the same. However, for the recovery of Model-I signals, non-zero blocks are likely to have much smaller weights than zero blocks.

Comparison With Other Recovery Algorithms: In Fig. 5, we compare the performance of iterative reweighted ℓ_2/ℓ_1 minimization algorithms to that of ℓ_1 minimization [6], IRℓ_1 [7], ISD [16], IRLS [13], standard ℓ_2/ℓ_1 minimization [27] and James-Stein AMP (AMP-JS) [44] for $N = 256$, $L = 32$, and $P = 8$. Based on the results of Fig. 1, we select the maximum number of iterations (i.e., $j_{\max}$) for IR$^{(1)}$-ℓ_2/ℓ_1 to be 5 and for IR$^{(2)}$-ℓ_2/ℓ_1 to be 3. We found that the selected parameters (the number of iterations and β) for IR$^{(1)}$-ℓ_2/ℓ_1 and IR$^{(2)}$-ℓ_2/ℓ_1 are also good choices for IRℓ_1 and ISD, respectively. We set the parameter ρ in IRℓ_1 the same as that in IR$^{(1)}$-ℓ_2/ℓ_1, i.e., we set ρ_0 to the standard deviation of nonzero signal components and reduce its value by a factor of 10 in each iteration if Inequality (11) holds. The parameters of IRLS are set to those recommended in [13]. AMP-JS is performed with the maximum number of iterations equal to 300. The algorithm stops if the maximum number of iterations is reached or if the normalized ℓ_2-norm of the difference between successive estimates, i.e., $\|\hat{\mathbf{x}}^{(i)} - \hat{\mathbf{x}}^{(i-1)}\|_2/\|\hat{\mathbf{x}}^{(i)}\|_2$, is less than 10^{-5}. Fig. 5 shows that the two iterative algorithms outperform all the other algorithms, including standard ℓ_2/ℓ_1 and AMP-JS, by a large margin. In particular, we note that the smallest value of M that results in the probability of success of almost one is 92 and 96 for IR$^{(1)}$-ℓ_2/ℓ_1 and IR$^{(2)}$-ℓ_2/ℓ_1, respectively. This value, however, is greater than 110 for all the other algorithms, including ℓ_2/ℓ_1 minimization and AMP-JS.

To have a more clear picture of the comparison between the proposed algorithms and AMP-JS, we have shown the probability of success curves of the three algorithms for Signal Model I with $N = 1024$ and $K = 256$ and two block lengths of $N/L = 8$ and 32 in Fig. 6. The curves for standard ℓ_2/ℓ_1 minimization are also included for reference. Fig. 6 shows that while the performance of AMP-JS improves by increasing the block length, it is still far from the performance of the proposed algorithms. More results for $N = 5120$ and $K = 1280$ are presented in Fig. 7, where IR$^{(1)}$-ℓ_2/ℓ_1 and AMP-JS are compared at different block lengths. The maximum number of iterations for AMP-JS in Fig. 7 is selected to be 1000, and the stopping criterion for the normalized ℓ_2-norm of the difference between successive estimates is set at 10^{-6}.[9] For IR$^{(1)}$-ℓ_2/ℓ_1, we use $j_{\max} = 10$ for block lengths 16 and 32, and $j_{\max} = 8$ for block length 64. Fig. 7 shows that the performance of

[9]One should note that for the AMP-JS results presented in Fig. 7, we declare success if the normalized error is less than 0.01 rather than 0.001 used for IR$^{(1)}$-ℓ_2/ℓ_1. If such an advantage is not given to AMP-JS, its performance measured by the probability of success will deteriorate noticeably. For example, for the scenario depicted in Fig. 7, with the normalized error threshold of 0.001 for success, AMP-JS with $N/L \geq 32$ requires about M = 1340 to achieve a success probability of 0.6 instead of M = 1300 as shown in Fig. 7.

TABLE I
AVERAGE RUNTIME OF THE RECOVERY ALGORITHMS OF FIG. 5 FOR $M = 170$.

Algorithm	Average Runtime (sec.)
ℓ_1 minimization	0.028
IRℓ_1	0.074
ISD	0.059
IRLS	0.127
AMP-JS	0.021
ℓ_2/ℓ_1 minimization	0.033
IR$^{(2)}$-ℓ_2/ℓ_1	0.081
IR$^{(1)}$-ℓ_2/ℓ_1	0.095

TABLE II
AVERAGE RUNTIME OF THE RECOVERY ALGORITHMS OF FIG. 6 FOR $M = 520$.

Algorithm	Average Runtime (sec.)	
	$N/L = 32$	$N/L = 8$
AMP-JS	0.064	0.131
ℓ_2/ℓ_1 minimization	0.182	0.246
IR$^{(2)}$-ℓ_2/ℓ_1	0.245	0.485
IR$^{(1)}$-ℓ_2/ℓ_1	0.325	0.713

TABLE III
AVERAGE RUNTIME OF THE RECOVERY ALGORITHMS OF FIG. 7 FOR $M = 1330$.

Algorithm	Average Runtime (sec.)		
	$N/L = 64$	$N/L = 32$	$N/L = 16$
AMP-JS	8.066	8.876	12.329
IR$^{(1)}$-ℓ_2/ℓ_1	69.948	87.288	108.589

AMP-JS improves with increasing the block length but saturates at about $N/L = 32$. In all block lengths, IR$^{(1)}$-ℓ_2/ℓ_1 outperforms AMP-JS, and the performance of the former is practically the same as the theoretical limit even for the block length of 16.

To compare the computational complexity of different recovery algorithms, in Tables I–III, we have reported the average runtime of the algorithms for scenarios corresponding to Figs. 5–7, respectively. All algorithms that involve solving a convex optimization problem (ℓ_1, IRℓ_1, ISD, ℓ_2/ℓ_1, IR$^{(1)}$-ℓ_2/ℓ_1, IR$^{(2)}$-ℓ_2/ℓ_1) are implemented using SpaRSA [49], [52]. In such implementations, the maximum number of iterations and the stopping criterion for the normalized ℓ_2-norm of the difference between successive estimates are set at the same values as those of AMP-JS in each scenario, i.e., 300 and 10^{-5}, respectively, for the scenarios in Figs. 5 and 6, and 1000 and 10^{-6}, respectively, for the scenario of Fig. 7. The reported runtimes are based on running the algorithms in MATLAB 7.11.0 environment on a desktop computer with an Intel Core i7 2.8 GHz CPU and 8 GB of RAM, and under the Microsoft Windows 7 operating system. The number of measurements for the results of Tables I– III are selected to be $M = 170$, 520 and 1330, respectively. These are the smallest values that result in success probability of about one for all the algorithms involved in each scenario. The results of Table I show that among all recovery algorithms, the one that performs the best, i.e., IR$^{(1)}$-ℓ_2/ℓ_1, has also the largest runtime. This runtime, however, is not much larger than those of the other algorithms. In particular, the average runtime of IR$^{(1)}$-ℓ_2/ℓ_1 is only about 3 and 5 times those of ℓ_2/ℓ_1 minimization and AMP-JS, respectively. The results of Tables II and III show that the runtime of all algorithms reduces with increasing the block length. These results also demonstrate that, for $N = 1024$, the average runtime of IR$^{(1)}$-ℓ_2/ℓ_1 is about 5 times that of AMP-JS for both block lengths 8 and 32. This ratio increases to about 9 to 10 for $N = 5120$ at different block lengths.

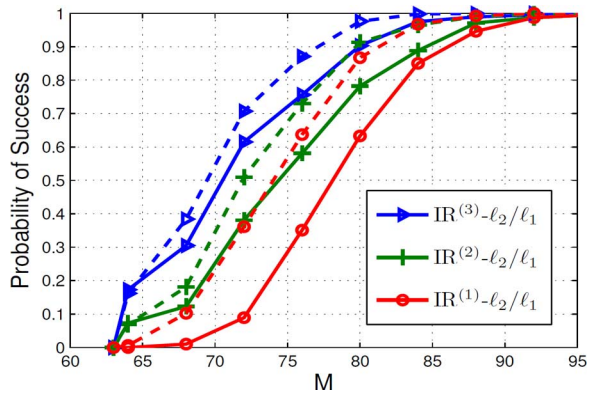

Fig. 8. Probability of success of different iterative reweighted ℓ_2/ℓ_1 recovery algorithms versus M ($N = 256$, $L = 32$ and $P = 8$). Solid lines correspond to Signal Model I with $j_{\max} = 3$ and dashed lines correspond to Signal Model II with $j_{\max} = 5$.

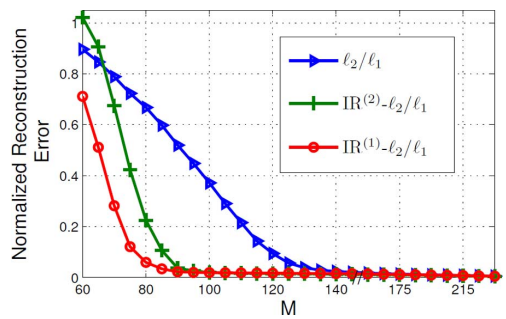

Fig. 9. Normalized reconstruction error for IR$^{(1)}$-ℓ_2/ℓ_1, IR$^{(2)}$-ℓ_2/ℓ_1 and the standard ℓ_2/ℓ_1 minimization for approximately block sparse signals versus M ($N = 256$, $L = 32$ and $P = 8$).

Other IR-ℓ_2/ℓ_1 Algorithms: Clearly, in IR-ℓ_2/ℓ_1 algorithms, the reweighting strategy plays an important role in the reconstruction performance. So far in the paper, we focused on two special cases of these algorithms. In general, however, there may be other reweighting strategies that result in better reconstruction performance depending for example on the signal distribution or the desirable trade-off between performance and the speed of convergence. For instance, one can execute Algorithm 1 with both support detection and weight-updating steps (i.e., Steps 2(a) and 2(b)). We refer to this option as IR$^{(3)}$-ℓ_2/ℓ_1, and consider the case with $\beta = 3$. The probability of success of this algorithm applied to block sparse signals with $N = 256$, $L = 32$ and $P = 8$ is depicted in Fig. 8 versus M for Signal Models I and II when the maximum number of iterations is set to 3 and 5, respectively. For comparison, we have also presented the results for IR$^{(1)}$-ℓ_2/ℓ_1 and IR$^{(2)}$-ℓ_2/ℓ_1. To have a fair comparison, for each signal model, we use the same number of iterations for all algorithms. Fig. 8 demonstrates that for both signal models, IR$^{(3)}$-ℓ_2/ℓ_1 outperforms the other two algorithms. Note

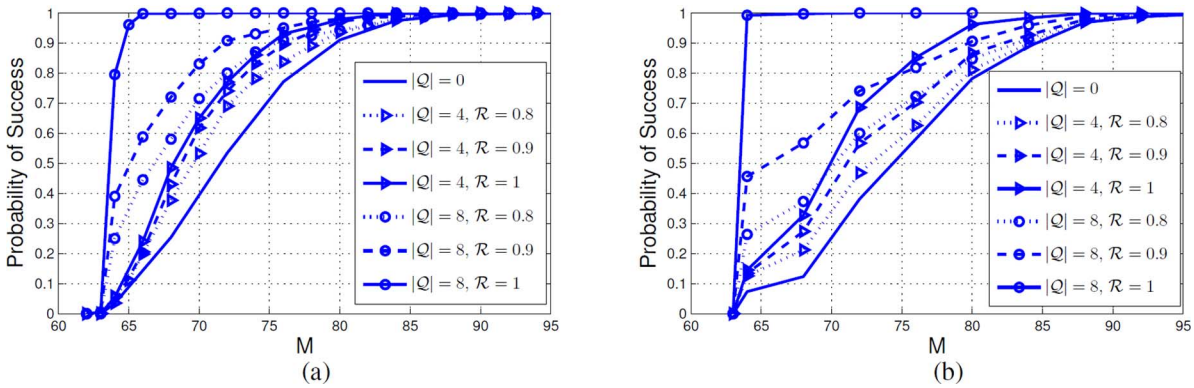

Fig. 10. Probability of success of (a) IR$^{(1)}$-ℓ_2/ℓ_1, and (b) IR$^{(2)}$-ℓ_2/ℓ_1 versus M at the presence of a priori information on the non-zero block locations ($N = 256$, $L = 32$, $P = 8$ and Signal Model I).

that since IR$^{(1)}$-ℓ_2/ℓ_1 needs more iterations to converge, by setting $j_{\max}$ to a small number, its performance would be inferior to IR$^{(2)}$-ℓ_2/ℓ_1 and IR$^{(3)}$-ℓ_2/ℓ_1. However, if $j_{\max}$ is selected to be a larger number, IR$^{(1)}$-ℓ_2/ℓ_1 would outperform the other two iterative algorithms.

2) Approximately Block Sparse Signals: An approximately block sparse signal is generated by two Gaussian distributions as in [53]. In this model, we select P blocks randomly and assign the entries of these blocks using a zero-mean Gaussian distribution with variance σ_s and the entries of the rest of the blocks are assigned using another zero-mean Gaussian distribution with variance σ'_s, where $\sigma'_s < \sigma_s$. For each set of blocks, the entries are selected in an i.i.d. fashion, and independent of the entries of the other set's blocks. Here, we set $\sigma_s = 1$ and $\sigma'_s = 0.01$, then normalize the signal such that $\sum_{i=1}^{L} \|\bar{\mathbf{u}}_i\|_2 = 1$.

In Fig. 9, we have provided the normalized reconstruction error (i.e., $\sum_{i=1}^{L} \|\hat{\mathbf{u}}_i - \bar{\mathbf{u}}_i\|_2 / \sum_{i=1}^{L} \|\bar{\mathbf{u}}_i\|_2$) of IR$^{(1)}$-$\ell_2/\ell_1$, IR$^{(2)}$-$\ell_2/\ell_1$ and the standard ℓ_2/ℓ_1 minimization when the signal is approximately block sparse with $N = 256$, $L = 32$ and $P = 8$. The parameter $j_{\max}$ for IR$^{(1)}$-ℓ_2/ℓ_1 and IR$^{(2)}$-ℓ_2/ℓ_1 is set to 5 and 3, respectively. Fig. 9 demonstrates that for smaller values of M, the reconstruction error of the iterative methods is, in general, much smaller than that of the standard method. The anomaly is for values of M less than about 70, where IR$^{(2)}$-ℓ_2/ℓ_1 performs inferior to the standard ℓ_2/ℓ_1. This can be attributed to the low reliability of support detection in the first iteration that consequently deteriorates the performance of IR$^{(2)}$-ℓ_2/ℓ_1 in subsequent iterations.

3) Block Sparse Signal Recovery With A Priori Information: In the following experiment, we study the effect of a priori information about non-zero blocks on the performance of the proposed algorithms. We use Signal Model I with $N = 256$, $L = 32$, $P = 8$ and set $\tau = 0.01$ to initialize the weights. Figs. 10(a) and 10(b) show the probability of success of IR$^{(1)}$-ℓ_2/ℓ_1 and IR$^{(2)}$-ℓ_2/ℓ_1 for different combinations of $|\mathcal{Q}|$ and $\mathcal{R}$, respectively. The parameter $j_{\max}$ for the two algorithms are selected to be 5 and 3, respectively. Note that the solid curves belong to the cases where the a priori information is perfectly reliable ($\mathcal{R} = 1$). In both figures, for a given value of $\mathcal{R}(|\mathcal{Q}|)$, increasing $|\mathcal{Q}|(\mathcal{R})$ improves the success probability. (In particular, for both algorithms, if $|\mathcal{Q}| = P = 8$ and $\mathcal{R} = 1$, the success probability tends to one for $M > K = 64$.) This is expected as increasing $|\mathcal{Q}|(\mathcal{R})$ increases the total amount of reliable a priori information. The figure also shows that any a priori information improves the performance compared to the case with no a priori information ($|\mathcal{Q}| = 0$).

B. Noisy Scenarios

In this subsection, we consider the case where measurements are contaminated with noise. For the measurement matrix and the generation of results based on random experiments, we follow the same assumptions as those in Section V-A. The only difference is that, we add i.i.d. zero-mean Gaussian noise with variance σ_n^2 to measurements. For the signal, we use Model I but without any normalization.

Based on the results presented in section V-A, in the absence of noise, increasing the number of iterations improves the performance of the proposed iterative algorithms. In the presence of noise, however, this is not generally the case. In particular, for a given SNR, there is a certain threshold I_{th} on the number of iterations beyond which the performance starts to deteriorate. This implies that there is a threshold SNR$_{th}$ on the SNR values for which the proposed algorithms provide an improvement over the standard ℓ_2/ℓ_1 minimization, i.e., for SNR values less than SNR$_{th}$, $I_{th} = 1$. In the following, for each SNR value, we select $j_{\max} = I_{th}$ for the iterative algorithms.

We consider a block sparse signal with $N = 256$, $L = 32$ and $P = 8$. In Figs. 11(b)–11(d), we have reported the performance of the proposed algorithms in comparison with the standard ℓ_2/ℓ_1 minimization and AMP-JS for $\sigma_n = 0.01$, 0.05 and 0.1, respectively. In Fig. 11(a), curves for noiseless scenario are presented for reference. The figures show that although, IR$^{(2)}$-ℓ_2/ℓ_1 performs inferior to IR$^{(1)}$-ℓ_2/ℓ_1 for smaller values of M, it performs practically the same as IR$^{(1)}$-ℓ_2/ℓ_1 for larger values of M. Moreover, the figures demonstrate that the proposed algorithms consistently outperform the standard ℓ_2/ℓ_1 minimization, although the amount of improvement decreases with decreasing SNR. As expected, $j_{\max} = I_{th}$ has also a decreasing trend with SNR for both iterative algorithms. Compared to AMP-JS, although the proposed algorithms are superior in the high SNR region and smaller number of measurements (small under-sampling ratios), they are outperformed by AMP-JS at lower SNR values and larger number of measurements.

To study the effect of block length on the relative performance of different algorithms in the presence of noise, we also con-

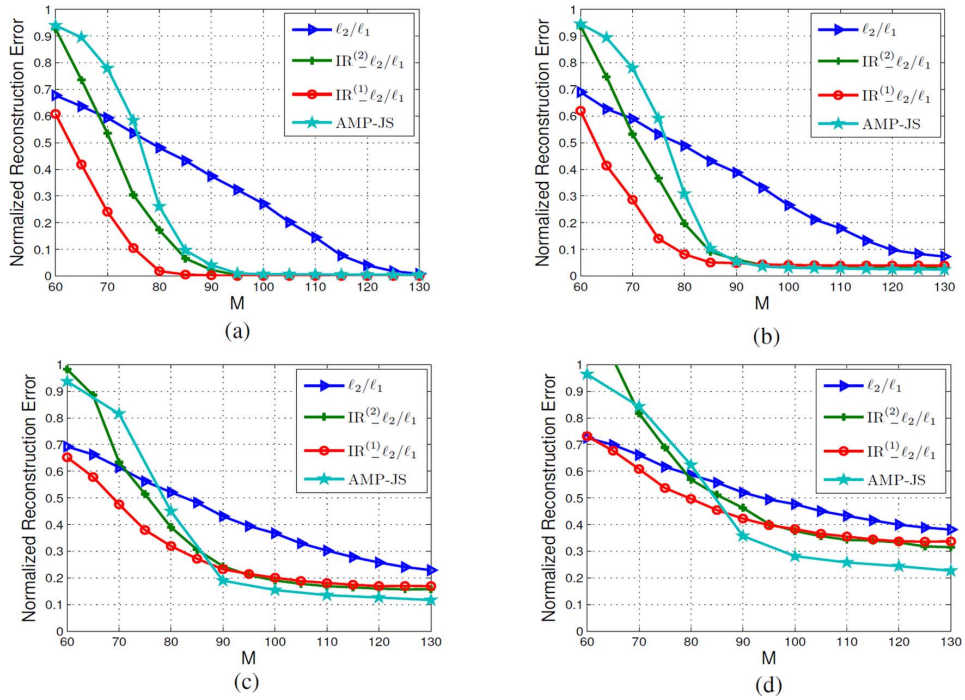

Fig. 11. Normalized reconstruction error of different recovery algorithms versus M for various SNR values ($N = 256$, $L = 32$, $P = 8$ and Signal Model I). For IR$^{(1)}$-ℓ_2/ℓ_1, $j_{\max}$ is selected to be 5, 5, 3, 3 in (a), (b), (c) and (d), respectively. For IR$^{(2)}$-ℓ_2/ℓ_1, $j_{\max}$ is selected to be 3, 3, 2, 2 in (a), (b), (c) and (d), respectively. (a) Noiseless (b) Noisy with $\sigma_n = 0.01(\text{SNR} = 33.98\ \text{dB})$ (c) Noisy with $\sigma_n = 0.05(\text{SNR} = 20.0\ \text{dB})$ (d) Noisy with $\sigma_n = 0.1(\text{SNR} = 13.98\ \text{dB})$.

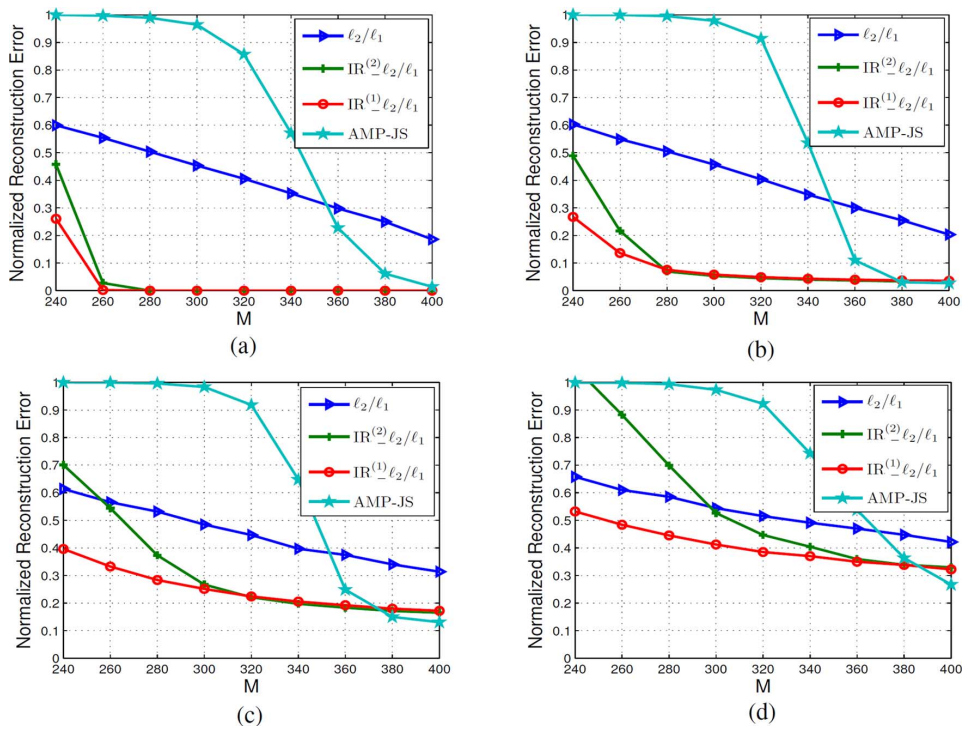

Fig. 12. Normalized reconstruction error of different recovery algorithms versus M for various SNR values ($N = 1024$, $L = 32$, $P = 8$ and Signal Model I). For IR$^{(1)}$-ℓ_2/ℓ_1, $j_{\max}$ is selected to be 5, 3, 3, 2 in (a), (b), (c) and (d), respectively. For IR$^{(2)}$-ℓ_2/ℓ_1, $j_{\max}$ is selected to be 5, 3, 2, 2 in (a), (b), (c) and (d), respectively. (a) Noiseless (b) Noisy with $\sigma_n = 0.01(\text{SNR} = 33.98\ \text{dB})$ (c) Noisy with $\sigma_n = 0.05(\text{SNR} = 20.0\ \text{dB})$ (d) Noisy with $\sigma_n = 0.1(\text{SNR} = 13.98\ \text{dB})$.

sider a block sparse signal with $N = 1024$, $L = 32$ and $P = 8$ (same sparsity ratio of 0.25 as in Fig. 11 but larger block length of 32 instead of 8). Results similar to those of Fig. 11 are presented in Fig. 12 for this case. In comparison with the results of Fig. 11, one can see that increasing the block length works in favor of the proposed algorithms and increases the performance gap between these algorithms, on the one hand, and standard ℓ_2/ℓ_1 minimization and AMP-JS, on the other hand. In particular, the proposed algorithms outperform AMP-JS for all SNR values and almost all values of M simulated in Fig. 12. The exception is for the largest M values at the right hand side of Figs. 12(c) and 12(d).

VI. Conclusion

We proposed a general iterative framework for the recovery of block sparse signals. The family of algorithms defined by this framework are based on iterative application of weighted ℓ_2/ℓ_1 minimization and are thus referred to as iterative reweighted ℓ_2/ℓ_1 minimization (IR-ℓ_2/ℓ_1). Different algorithms within the family differ in the way that the weights are updated from one iteration to the next. We focused particularly on two special cases in this family, i.e., IR$^{(1)}$-ℓ_2/ℓ_1 and IR$^{(2)}$-ℓ_2/ℓ_1. We analyzed both algorithms and discussed their application to block sparse signals in the absence and presence of a priori information about the support of the signal. We derived theoretical guarantees and showed that an accurate reconstruction was possible if the sensing matrix satisfied certain conditions. In the case of approximately block sparse signals, reconstruction error bounds for both algorithms were also derived.

Our results demonstrated that both proposed algorithms can achieve a significant improvement over other recovery algorithms including the standard ℓ_2/ℓ_1 minimization and the recently popular approximate message-passing (AMP) in noiseless scenarios and high SNR regimes. Moreover, our results showed that as the length of the blocks increases, the required number of measurements for the reconstruction of block sparse signals by the proposed algorithms tends to the theoretical limit of the number of non-zero elements. The improvement in performance was at the cost of a modest increase in complexity. Our results also indicated that the relative performance of IR$^{(1)}$-ℓ_2/ℓ_1 and IR$^{(2)}$-ℓ_2/ℓ_1 depended on the signal model as well as the number of iterations. The former algorithm would be desirable for signals with slow decaying distributions for the ℓ_2-norm of the blocks and when a larger number of iterations can be allowed. The latter algorithm, however, is more attractive due to its faster convergence speed particularly for fast decaying signal distributions.

We hope that the results of this paper, particularly those presented at the end of Part V-A1, will motivate more research on other reweighting strategies within the class of IR-ℓ_2/ℓ_1 algorithms.

Acknowledgment

The authors would like to thank the anonymous reviewers whose constructive comments improved the content and the presentation of the paper.

References

[1] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[2] E. Candès and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.
[3] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.
[4] D. Baron, M. B. Wakin, M. Duarte, S. Sarvotham, and R. G. Baraniuk, "Distributed compressed sensing," Jan. 2009, ArXiv:0901.3403v1.
[5] R. G. Baraniuk, "Compressive sensing," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–121, Jul. 2007.
[6] S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comp.*, vol. 20, no. 1, pp. 33–61, 1998.
[7] 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, pp. 877–905, Dec. 2008.
[8] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *J. Roy. Statist. Soc.*, ser. B, vol. 58, no. 1, pp. 267–288, 1996.
[9] F. Alizadeh and D. Goldfarb, "Second-order cone programming," *Math. Program.*, ser. B, vol. 95, no. 1, pp. 3–51, Jan. 2003.
[10] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K: Cambridge Univ. Press, 2004.
[11] M. Grant and S. Boyd, "Cvx User's Guide for cvx Version 1.2," [Online]. Available: http://www.stanford.edu/ boyd/software.html
[12] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[13] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Las Vegas, NV, USA, Apr. 2008, pp. 3869–3872.
[14] C. J. Miosso, R. Borries, M. Argàez, L. Velazquez, C. Quintero, and C. M. Potes, "Compressive sensing reconstruction with prior information by iteratively reweighted least-square," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2424–2431, Jun. 2009.
[15] D. Needell, "Noisy signal recovery via iterative reweighted L1-minimization," in *Proc. Asilomar Conf. Signal, Syst., Comput.*, 2009, pp. 113–117.
[16] Y. Wang and W. Yin, "Sparse signal reconstruction via iterative support detection," *SIAM J. Image Sci.*, vol. 3, no. 3, pp. 462–491, Mar. 2010.
[17] R. Von Borries, C. J. Miosso, and C. Potes, "Compressed sensing using prior information," in *Proc. IEEE Int. Workshop Comp. Adv. Multi-Sensor Adaptive Process. (CAMPSAP)*, Dec. 2007, pp. 121–124.
[18] P. Zhang and R. C. Qiu, *"Modified Orthogonal Matching Pursuit Algorithm for Cognitive Radio Wideband spectrum sensing,"*, Feb. 2011, ArXiv:1102.2881v1.
[19] Z. Tian and G. B. Giannakis, "Compressed sensing for wideband cognitive radios," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Honolulu, HI, USA, Apr. 2007, pp. 1357–1360.
[20] S. Cotter and B. Rao, "Sparse channel estimation via matching pursuit with application to equalization," *IEEE Trans. Commun.*, vol. 50, no. 3, pp. 374–377, Mar. 2002.
[21] M. Mishali and Y. C. Eldar, "Blind multi-band signal reconstruction: Compressed sensing for analog signals," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 993–1009, Mar. 2009.
[22] S. Cotter, B. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2477–2488, Jul. 2005.
[23] J. Tropp, A. C. Gilbert, and M. Strauss, "Algorithms for simultaneous sparse approximation. Part I: greedy pursuit," *IEEE Sel. Topics Signal Process.*, vol. 86, no. 3, pp. 572–588, Apr. 2006.
[24] J. Tropp, "Algorithms for simultaneous sparse approximation. Part II: Convex relaxation," *IEEE J. Sel. Topics Signal Process.*, vol. 86, no. 3, pp. 589–602, Apr. 2006.
[25] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4634–4702, Dec. 2006.
[26] M. Mishali and Y. C. Eldar, "Reduce and boost: recovering arbitrary sets of jointly sparse vectors," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4692–4702, Oct. 2008.
[27] M. Stojnic, F. Parvaresh, and B. Hassibi, "On the reconstruction of block-sparse signals with an optimal number of measuremets," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 3075–3085, Aug. 2009.
[28] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5302–5316, Nov. 2009.
[29] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: uncertainty relations and efficient recovery," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3042–3054, Jun. 2010.
[30] E. Elhamifar and R. Vidal, "Block-sparse recovery via convex optimization," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 4094–4107, Aug. 2012.
[31] T. Tanaka, "A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors," *IEEE Trans. Inf. Theory*, vol. 48, no. 11, pp. 2888–2910, Nov. 2002.
[32] D. Guo and S. Verdu, "Randomly spread CDMA: Asymptotics via statistical physics," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 1983–2010, Jun. 2005.
[33] A. Montanari and D. Tse, "Analysis of belief propagation for nonlinear problems: The example of CDMA (or: How to prove Tanaka's formula)," in *Proc. Inf. Theory Workshop*, Mar. 2006, pp. 160–164.
[34] D. Guo and C.-C. Wang, "Multiuser detection of sparsely spread CDMA," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 3, pp. 421–431, Apr. 2008.

[35] D. Guo, D. Baron, and S. Shamai, "A single-letter characterization of optimal noisy compressed sensing," in *Proc. 47th Allerton Conf. Comm., Contr., Comput.*, Sep. 2009, pp. 52–59.
[36] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci.*, vol. 106, no. 45, pp. 18914–18919, Nov. 2009.
[37] D. Baron, S. Sarvotham, and R. G. Baraniuk, "Bayesian compressive sensing via belief propagation," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 269–280, Jan. 2010.
[38] P. Schniter, "Turbo reconstruction of structured sparse signals," in *Proc. IEEE Conf. Inf. Sci. Sys. (CISS)*, Aug. 2010, pp. 1–6.
[39] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Aug. 2011, pp. 2168–2172.
[40] S. Rangan, A. K. Fletcher, V. K. Goyal, and P. Schniter, "Hybrid generalized approximate message passing with applications to structured sparsity," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2012, pp. 1236–1240.
[41] S. Rangan, A. K. Fletcher, and V. K. Goyal, "Asymptotic analysis of MAP estimation via the replica method and applications to compressed sensing," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1902–1923, Mar. 2012.
[42] Y. Eftekhari, A. Heidarzadeh, A. H. Banihashemi, and I. Lambadaris, "Density evolution analysis of node-based verification-based algorithms in compressed sensing," *IEEE Trans. Inf. Theory*, vol. 58, no. 10, pp. 6616–6645, Oct. 2012.
[43] A. Maleki and A. Montanari, "Analysis of approximate message passing algorithm," in *Proc. IEEE Conf. Inf. Sci. Syst. (CISS)*, Mar. 2010, pp. 1–7.
[44] D. L. Donoho, I. M. Johnstone, and A. Montanari, "Accurate prediction of phase transitions in compressed sensing via a connection to minimax denoising," *IEEE Trans. Inf. Theory*, vol. 59, no. 6, pp. 3396–3433, Jun. 2013.
[45] A. Taeb, A. Maleki, C. Studer, and R. Baraniuk, Maximin Analysis of Message Passing Algorithms for Recovering Block Sparse Signals," Mar. 2013, ArXiv:1303.2389v1, "," .
[46] U. Kamilov, S. Rangan, A. Fletcher, and M. Unser, "Approximate message passing with consistent parameter estimation and applications to sparse learning," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2969–2985, May 2014.
[47] Y. Wu and S. Verdú, "Optimal phase transitions in compressed sensing," *IEEE Trans. Inf. Theory*, vol. 58, no. 10, pp. 6241–6263, Oct. 2010.
[48] M. Stojnic, "ℓ_2/ℓ_1-optimization and its strong thresholds in approximately block-sparse compressed sensing," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Seoul, Korea, Jul. 2009, pp. 473–477.
[49] S. J. Wright, R. D. Nowak, and M. Figueiredo, "Sparse reconstruction by separable approximation," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2479–2493, Jul. 2009.
[50] W. Lu and N. Vaswani, "Regularized modified BPDN for noisy sparse reconstruction with partial erroneous support and signal value knowledge," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 182–196, Jan. 2012.
[51] Z. Zeinalkhani and A. H. Banihashemi, "Iterative recovery algorithms for compressed sensing of wideband block sparse spectrums," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Ottawa, ON, Canada, Jun. 2012, pp. 1655–1659.
[52] S. J. Wright, R. D. Nowak, and M. Figueiredo, "Sparse Reconstruction by Separable Approximation," [Online]. Available: http://www.lx.it.pt/~mtf/SpaRSA/
[53] D. Baron, S. Sarvotham, and R. G. Baraniuk, "Bayesian compressed sensing reconstruction via belief propagation," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 269–280, Jan. 2010.

Zeinab Zeinalkhani received her B.Sc. degrees, both, in electrical engineering (telecommunication and electronics) and her M.Sc. degree in communications engineering from Amirkabir University of Technology (Tehran Polytechnic), in 2006, 2006 and 2009, respectively. She is currently working toward the Ph.D. degree in electrical and computer engineering at Carleton University, Ottawa, ON, Canada.

Her research interests include signal processing, digital and wireless communications, and compressed sensing.

Amir H. Banihashemi (S'90–A'98–M'03–SM'04) received the B.A.Sc. degree in electrical engineering from Isfahan University of Technology, Isfahan, Iran in 1988, and the M.A.Sc. degree in communications engineering from the University of Tehran, Tehran, Iran, in 1991, with the highest academic rank in both classes.

From 1991 to 1994, he was with the Electrical Engineering Research Center and the Department of Electrical and Computer Engineering, Isfahan University of Technology. During 1994–1997, he was with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ontario, Canada, for his Ph.D. degree. In 1997, he joined the Department of Electrical and Computer Engineering, University of Toronto, Toronto, Ontario, Canada, where he worked as a Natural Sciences and Engineering Research Council of Canada (NSERC) Postdoctoral Fellow. He joined the Faculty of Engineering at Carleton University in 1998, where at present he is a Professor in the Department of Systems and Computer Engineering. His research interests include coding and information theory, digital and wireless communications, network coding, theory and implementation of communication algorithms, and compressed sensing and sampling.

Dr. Banihashemi served as an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS from 2003 to 2009. He was the Director of the Broadband Communications and Wireless Systems (BCWS) Centre at Carleton University from 2005 to 2013. He has been involved in many international conferences as chair, member of technical program committee, and member of organizing committee. These include co-chair for the Communication Theory Symposium of Globecom'07, TPC co-chair of the Information Theory Workshop (ITW) 2007, member of the organizing committee for the International Symposium on Information Theory (ISIT) 2008, and co-chair of the Canadian Workshop on Information Theory (CWIT) 2009. Dr. Banihashemi has been a two-time recipient of NSERC Discovery Accelerator Supplement (DAS) in 2008 and 2013.