Iteratively Reweighted Least Squares for Block-sparse Recovery

Shuang Li, Qiuwei Li, Gang Li, Xiongxiong He and Liping Chang
Zhejiang Key Lab. for Signal Processing
College of Information Engineering
Zhejiang University of Technology
310023 Hangzhou, Zhejiang, P.R. China
Email: lishuang0809@gmail.com

Abstract—The compressive sensing (CS) theory has shown that sparse signals can be reconstructed exactly from much fewer measurements than traditionally believed. What's more, using $\ell_p$-norm minimization with $p < 1$ can do so with much fewer measurements than with $p = 1$. In this paper, a novel algorithm is proposed for computing local minima of the nonconvex problem in the block-sparse system. A series of experiments are presented to show the remarkable performance of our proposed algorithm in block sparse signal recovery, and compare the recovery ability of this algorithm with the IRLS and BOMP algorithm.

Index Terms—Compressive sensing, underdetermined systems of linear equations, nonconvex optimization, block sparse signal reconstruction, BIRLS.

I. INTRODUCTION

CS is an emerging framework which states that sparse signals, that is, signals that have a concise linear representation on an appropriate dictionary, can be exactly recovered from a number of linear projections of dimension considerably lower than the number of samples traditionally required [1]. Excellent reviews on the basics of CS and new directions of research on CS can be found from [2], [3]. There are three basic elements in the CS system: the sensing matrix, the sparse representation dictionary and the reconstruction algorithm.

Consider a signal $x \in \mathbb{R}^N$ can often be expressed as a linear combination of a small number of signals taken from a “resource” database, which is called the dictionary. Elements of the dictionary are typically unit norm functions called atoms. Let us denote the dictionary as $\Psi$, and the atoms as $\psi_k, k = 1, \cdots, L$, that is, $\Psi = \left[ \psi_1 \ \psi_2 \ \cdots \ \psi_L \right] \in \mathbb{R}^{N \times L}$, where $L$ is the size of the dictionary. The dictionary is over-complete ($L > N$) when it spans the signal space and its atoms are linearly dependent. In that case, every signal can be represented as a linear combination of atoms in the dictionary:

$$x = \sum_{k=1}^{L} s_k \psi_k \triangleq \Psi s \tag{1}$$

where $s \triangleq \left[ s_1 \ s_2 \ \cdots \ s_L \right]^T$ is a coefficient vector that represents $x$ in dictionary $\Psi$. We say that a signal $x$ is sparse in some dictionary $\Psi$, that is to say, $x$ has at least one sparse representing vector $s$ that has few non-zero coefficients.

Such class of signals which are sparse in some dictionary can be accurately recovered from a measurement vector $y \in \mathbb{R}^M$ with $M << N$, where the measurement paradigm consists of linear projections of the signal vector via a carefully chosen projection matrix $\Phi \in \mathbb{R}^{M \times N}$:

$$y = \Phi x \tag{2}$$

Where $x$ is the original signal needed to be recover. The projection matrix $\Phi$, alternatively called measurement matrix or sensing matrix, acts as a multitude of probes on the information contained in the original signal $y$.

By substituting $x$ in (2) with (1), $z$ can be rewritten as

$$y = \Phi \Psi s \triangleq As \tag{3}$$

where the matrix $A = \Phi \Psi$ is sometimes referred to as equivalent dictionary of the CS system, which is often encountered in engineering environment. As these systems have more unknowns than equations, there exists an infinite number of solutions. It is apparently impossible to identify which of these candidate solutions is indeed the proper one without some additional information. In many instances, the object we wish to recover is known to be sparse (or approximately so) when expressed in certain basis which significantly makes the search for solutions feasible since the simplest solution now tends to be the proper one. This is where the sparsity constraint comes into play.

The spark of a given matrix $A$, denoted as $spark(A)$, is defined as the smallest number of columns in $A$ that are linearly dependent [4]. If $spark(A) \geq 2K$, then for each measurement vector $y \in \mathbb{R}^M$ there exists at most one $K$ sparse signal $\theta$ such that $y = A\theta$. It can be seen that the larger the $spark$ of $A$, the bigger the signal space among which the CS systems can guarantee exact recovery. This property of matrices for the study of the uniqueness of sparse solutions is of great importance.

However, it is not an easy task as computing the spark of a matrix has combinatorial complexity. It is therefore preferable to use some alternative properties of $A$ that can be easily manipulated to provide recovery guarantees such as the mutual coherence [4].
In principle, the problem of finding the sparsest representation of a given signal can be cast as the following optimization problem:

\[ P_{\ell_0} : \min_{\theta} ||\theta||_0 \quad \text{s.t.} \quad y = A\theta \quad (4) \]

where \( ||\theta||_0 \) counts the number of nonzero elements in \( \theta \).

The work of Donoho [5] and Candès [6] has shown that under appropriate conditions, the problem \( P_{\ell_0} \) can be solved efficiently by the following problem \( P_{\ell_1} \), since finding the sparse representation of a given signal using \( P_{\ell_0} \) problem is NP-hard [7]. In previous work, there are many research papers addressing conditions under which the problem \( P_{\ell_1} \) is efficiently by the following problem [8]-[10].

\[ P_{\ell_1} : \min_{\theta} ||\theta||_1 \quad \text{s.t.} \quad y = A\theta \quad (5) \]

Recently, there has been growing interest in recovering sparse representations of signals which have nonzero entries appearing in blocks rather than arbitrarily spreading throughout the signals. They are the so-called block-sparse signals. In this paper, we are interested in these block-sparse signals which arise naturally in dealing with the multi-band signals [11], face recognition [12], and the measurements of gene expression levels [13], etc.

The main objective and contribution of this paper are given as follows, respectively.

- **Objective**: We consider the problem of computing the local minima of the nonconvex problem in the block-sparse system, that is to find the sparse representation of a given signal which is known to be block sparse when expressed in the proper basis.

- **Contribution**: We propose a new algorithm called Block Iteratively Reweighted Least Squares (BIRLS) based on the IRLS [14]-[15] to solve the problem of computing the local minima of the nonconvex problem in the block-sparse system. Experiments are given to show the remarkable performance of this algorithm compared with the IRLS and BOMP algorithm.

The outline of this paper is arranged as follows. In Section II, some basic knowledge about the block sparse reconstruction is briefly reviewed. In Section III, we present our objective for block sparse recovery and propose a new algorithm — BIRLS to solve it. Simulation results are presented in Section IV to show the outperformance of our proposed method in improving signal reconstruction accuracy. Some concluding remarks are given in Section V to end this paper.

II. PRELIMINARIES

In this section, we briefly introduce the definition of block-sparsity and some basic work on the recovery of block-sparse signals.

A. Block-Sparsity

In this subsection, we consider the problem of representing a vector \( y \in \mathbb{R}^M \) in a given equivalent dictionary \( A \in \mathbb{R}^{M \times L} \) with \( M \leq L \), i.e., \( y = A\theta \) with \( \theta \in \mathbb{R}^L \). As the system of equation \( y = A\theta \) is underdetermined, there exists an infinite number of solutions. In this paper, we consider the case of sparse vector \( \theta \) with nonzero entries appearing in blocks rather than arbitrarily spreading throughout the vector. Specific examples include signals which lie in unions of subspaces [16].

For a given \( \theta \), we re-arrange it into a new vector, denoted as \( \tilde{\theta} \), by swapping the positions of the elements in \( \theta \) such that \( \tilde{\theta} \) is in the following block structure:

\[ \theta \rightarrow \tilde{\theta} = \left[ \tilde{\theta}^T_1 \ldots \tilde{\theta}^T_m \right]^T \quad (6) \]

where \( \tilde{\theta}^T_m \in \mathbb{R}^{m \times 1} \) denotes the \( m \)th block of the coefficient vector \( \tilde{\theta} \) with \( \sum_{m=1}^B l_m = L \), and \( B \) is the number of blocks. The representation \( \tilde{\theta} \) of \( \theta \) is based on the Sparse Agglomerative Clustering (SAC) algorithm [17].

Accordingly, the equivalent dictionary \( A \), corresponding to \( \theta \), can be re-arranged partitioned into \( \tilde{A} \):

\[ A \rightarrow \tilde{A} = \left[ \tilde{A}^T_1 \ldots \tilde{A}^T_B \right] \quad (7) \]

Define \( I(z) \) as \( I(z) = 1, \forall z \neq 0 \), and otherwise \( I(z) = 0 \). A vector \( \theta \) is said to be \( K \)-block sparse if

\[ ||\theta||_0 \psi = \sum_{m=1}^B I(||\tilde{\theta}^T|m||_2) \leq K \]

which denotes the number of nonzero blocks in \( \theta \) or \( \tilde{\theta} \).

B. Recovery of Block-Sparse Signals

The problem of finding a representation of a signal \( y \) that uses the minimum number of blocks of \( \tilde{A} \) can be formulated as follows [18].

\[ P_{p/\ell_0} : \min_{\tilde{\theta}} \sum_{m=1}^B I(||\tilde{\theta}^T|m||_p) \quad \text{s.t.} \quad y = \tilde{A}\tilde{\theta} \quad (8) \]

where \( ||\tilde{\theta}^T|m||_p \) denotes the \( \ell_p \)-norm of \( \tilde{\theta}^T|m| \) and the objective counts the number of nonzero blocks of \( \tilde{\theta} \).

It is obvious that solving (8) is an NP-hard problem since it requires searching exhaustively over all choices of a few blocks of \( \tilde{A} \) and checking whether they span the observed signal. Then, the \( \ell_1 \) relaxation of it is proposed [18]:

\[ P_{\ell_1/\ell_1} : \min_{\tilde{\theta}} \sum_{m=1}^B ||\tilde{\theta}^T|m||_1 \quad \text{s.t.} \quad y = \tilde{A}\tilde{\theta} \quad (9) \]

For \( p \geq 1 \), this is a convex problem, and can be solved efficiently by using the convex programming tools [19].
III. Iteratively Reweighted Least Squares for Block-sparse Minimization

In this section, we consider the problem of block-sparse recovery in a union of subspaces. We assume that the equivalent dictionary $\mathbf{A}$ consists of $B$ blocks and that the maximum number of atoms in each block $\mathbf{A}[m] \in \mathbb{R}^{M \times \ell_m}$ is $s$, i.e., $l_m \leq s$, $\forall m \in [1, B]$. Our goal is to find the sparse representation of a measurement signal $y$, under the condition of an overcomplete equivalent dictionary with a block structure given.

As already mentioned, a popular way to render the problem $P_{\ell_p/\ell_0}$ more tractable is to relax the highly discontinuous $\ell_0$-norm, replacing it by a continuous or even smooth approximation. Examples of such relaxation options include replacing it with $\ell_1$-norm or even by smooth functions such as $\sum \log(1 + \alpha \|m\|^2)$ [21], etc. In this work, we mainly deal with the problem $P_{\ell_2/\ell_1}$ and propose a new method to solve it.

For the special case of $p = 2$, we have

$$\|\hat{\theta}[m]\|_p = \|\hat{\theta}[m]\|_2 = \frac{1}{\|\hat{\theta}[m]\|_2} \|\hat{\theta}[m]\|_2^2$$

where $\omega_m \triangleq \frac{1}{\|\hat{\theta}[m]\|_2}$, $W[m] \triangleq diag(\omega_m, \ldots, \omega_m) \in \mathbb{R}^\ell \times m$, $m \in [1, B]$, and $W[m]^{\frac{1}{2}}$ denotes the square root of $W[m]$. In order to provide stability and to avoid the situation that $\omega_m \to \infty$ when $\|\hat{\theta}[m]\|_2 = 0$, a small parameter $\varepsilon \in (0, 1)$ is introduced:

$$\omega_m \triangleq \frac{1}{\|\hat{\theta}[m]\|_2 + \varepsilon}$$

Plugging these results into problem $P_{\ell_2/\ell_1}$ (9), then, the problem $P_{\ell_2/\ell_1}$ changes to

$$P_{\ell_2/\ell_1} : \min_{\hat{\theta}} \sum_{m=1}^B \omega_m^2 \|\hat{\theta}[m]\|_2 \quad s.t. \quad y = A\hat{\theta}$$

$$\phantom{P_{\ell_2/\ell_1}} = \min_{\hat{\theta}} \sum_{m=1}^B \|W[m]^{\frac{1}{2}} \hat{\theta}[m]\|_2^2 = \|W^{\frac{1}{2}} \hat{\theta}\|_2^2 \quad s.t. \quad y = A\hat{\theta}$$

where $W = diag(W[1], \ldots, W[B]) \in \mathbb{R}^{\ell \times L}$ is a diagonal positive-definite matrix, named as the weighted matrix.

Based on the above idea, we replace the penalty function in (9) by a weighted $\ell_2$-norm, and propose the following block-sparse minimization problem which can be efficiently solved by our proposed algorithm, named the Block Iteratively Reweighted Least Squares (BIRLS):

$$WP_{\ell_p/\ell_1} : \min_{\hat{\theta}} \sum_{m=1}^B \omega_m^p \|\hat{\theta}[m]\|_2^2 \quad s.t. \quad y = A\hat{\theta}$$

where

$$\omega_m \triangleq \frac{1}{\|\hat{\theta}[m]\|_2 + \varepsilon}^{1-\frac{p}{2}}, \quad 0 \leq p \leq 1$$

As to the problem $WP_{\ell_p/\ell_1}$ defined in (10), our proposed algorithm to solve it is outlined as follows.

Proposed algorithm:

**Step I:** Firstly, initialize the iteration count $n = 1$, and $\omega_m^{(0)} = 1$, $m = 1, \ldots, B$. Then, we perform the following three steps at each iteration $n$.

**Step II:** For the given weighted matrix $W^{(n-1)}$, find the block-sparse representation $\hat{\theta}^{(n)}$ by solving the following minimization problem:

$$\hat{\theta}^{(n)} = \arg \min_{\hat{\theta}} \sum_{m=1}^B \omega_m^{(n-1)} \|\hat{\theta}[m]\|_2^2$$

$$\phantom{\hat{\theta}^{(n)}} = \arg \min_{\hat{\theta}} \|W^{(n-1)}^{\frac{1}{2}} \hat{\theta}\|_2^2 \quad s.t. \quad y = A\hat{\theta} \quad (11)$$

**Step III:** With the obtained $\hat{\theta}^{(n)}$, update the weighted matrix $W^{(n)}$:

$$\omega_m^{(n)} = \frac{1}{\|\hat{\theta}^{(n)}[m]\|_2^2 + \varepsilon}^{1-\frac{p}{2}}$$

$$\phantom{W^{(n)}} = diag(\omega_m^{(n)}, \ldots, \omega_m^{(n)})$$

$$W^{(n)} = diag(W^{(n)}[1], \ldots, W^{(n)}[B])$$

**Step IV:** Terminate on convergence or when $n$ attains a specified maximum number of iterations $n_{max}$. Otherwise, increment $n$ and go to **Step II**.

The minimization problem defined in (11) can be attacked by solving the Euler-Lagrange equation of (11). Introducing the Lagrange multiplier $\lambda$, we define the Lagrangian as

$$L(\hat{\theta}) = \|W^{(n-1)}^{\frac{1}{2}} \hat{\theta}\|_2^2 + \lambda^T (A\hat{\theta} - y).$$

Taking a derivative of $L(\hat{\theta})$ with respect to $\hat{\theta}$, we have the requirement

$$\frac{\partial L(\hat{\theta})}{\partial \hat{\theta}} = 2W^{(n-1)}^{\frac{1}{2}} \hat{\theta} + A^T \lambda.$$

Thus the solution is obtained as

$$\hat{\theta}_{opt} = -\frac{1}{2} W^{(n-1)-1} A^T \lambda, \quad (12)$$

where $W^{(n-1)-1}$ denotes the inverse of $W^{(n-1)}$.

Plugging this solution into the constraint $y = A\hat{\theta}$, we get the Lagrange multiplier

$$\lambda = -2[AW^{(n-1)}A^T]^{-1} y.$$

Assigning this to the equation (12) generates the optimal solution

$$\hat{\theta}^{(n)} = \hat{\theta}_{opt} = W^{(n-1)-1} A^T [AW^{(n-1)}A^T]^{-1} y.$$

In this work, we relax the nonconvex optimization problem (9) by our proposed algorithm BIRLS, which made the recovery problem more tractable and efficient.
IV. SIMULATION RESULTS

In this section, we evaluate the contribution of the proposed BIRLS algorithm, and compare the sparse recovery ability of this algorithm with the IRLS algorithm [14] and the BOMP algorithm [20] via the following experiments.

We randomly select entries of a $N \times L$ matrix $A$ from a mean-zero Gaussian distribution, then scale the columns to have unit $\ell_2$-norm. For convenience, all blocks are assumed to have a same size $s$, i.e., $\ell_m = s, \forall m \in [1, B]$. The iteration is run until the change in relative $\ell_2$-norm from the previous iterate is less than $\sqrt{\varepsilon}/100$, at which point $\varepsilon$ (initialized to 1) is reduced by a factor of 10, then the iteration repeated beginning with the previous solution.

We use the normalized representation error

$$e = \frac{\|\hat{\theta}_0 - \hat{\theta}_r\|_F^2}{\|\hat{\theta}_0\|_F^2}$$

to evaluate the performance of each algorithm, where $\hat{\theta}_0$ denotes the original block sparse signal and $\hat{\theta}_r$ denotes the reconstruction signal.

In the first experiment, we generate a $K = 6$ block sparse signal $\tilde{\theta}_0$ and recover it through the measurement $y = A\tilde{\theta}_0$. The dimension of each block is set to $s = 10$, i.e., the number of blocks in the equivalent dictionary $\bar{A} \in \mathbb{R}^{100 \times 200}$ is $B = \frac{K}{s} = 60$. We use the IRLS and BIRLS to recover the signal from its measurement $y$ with different $p$. The result is shown in Fig. 1-Fig. 3 and the reconstruction error is given in TABLE I.

![Fig. 1. Reconstruction of a K = 6 block-sparse signal $\tilde{\theta}_0 \in \mathbb{R}^{200 \times 1}$. The blue points indicate the original signal $\tilde{\theta}_0$, the blue crosses indicate the IRLS reconstruction signal and the red circles indicate the BIRLS reconstruction signal with $p = 0$.](image)

![Fig. 2. Reconstruction of a K = 6 block-sparse signal $\tilde{\theta}_0 \in \mathbb{R}^{200 \times 1}$. The blue points indicate the original signal $\tilde{\theta}_0$, the blue crosses indicate the IRLS reconstruction signal and the red circles indicate the BIRLS reconstruction signal with $p = 1/2$.](image)

![Fig. 3. Reconstruction of a K = 6 block-sparse signal $\tilde{\theta}_0 \in \mathbb{R}^{200 \times 1}$. The blue points indicate the original signal $\tilde{\theta}_0$, the blue crosses indicate the IRLS reconstruction signal and the red circles indicate the BIRLS reconstruction signal with $p = 1$.](image)

<table>
<thead>
<tr>
<th>TABLE I</th>
<th>THE RECONSTRUCTION ERROR FOR IRLS AND BIRLS</th>
<th>(M = 100, L = 200, K = 6)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IRLS</td>
<td>BIRLS</td>
</tr>
<tr>
<td>$p = 0$</td>
<td>$5.6167 \times 10^{-4}$</td>
<td>$2.6193 \times 10^{-8}$</td>
</tr>
<tr>
<td>$p = 1/2$</td>
<td>$5.7374 \times 10^{-4}$</td>
<td>$4.1687 \times 10^{-8}$</td>
</tr>
<tr>
<td>$p = 1$</td>
<td>$8.0958 \times 10^{-4}$</td>
<td>$4.4523 \times 10^{-4}$</td>
</tr>
</tbody>
</table>

In the second experiment, we mainly compare the reconstruction performance of the BIRLS and BOMP mentioned above with respect to different $p$. Signals are generated with different block-sparsity $K$ which varies from 1 to 20. num = 40
signals for each $K$ are used to deal with the recovery problem. The result in Fig. 4 shows that the reconstruction error decays as $p$ decreases. Although the BOMP algorithm has a wonderful performance when $K$ is small, its space of recoverable signals is much smaller than our BIRLS algorithm.

![Fig. 4. Comparison of BOMP algorithm and BIRLS algorithm, M=100, L=256, s=8, B=32.](image)

In the third experiment, we mainly compare the reconstruction performance of the IRLS and BIRLS algorithm with respect to different $p$. Signals are also generated with different block-sparsity $K$ which varies from 1 to 20 and $num = 40$ signals for each $K$ are used to deal with the recovery problem. The reconstruction result is shown in Fig. 5. It also indicates that the reconstruction error decays as $p$ decreases both in the IRLS algorithm and the BIRLS algorithm. As to the same $p$, the simulation result shows that our proposed BIRLS algorithm significantly outperforms the IRLS algorithm in a block-sparse system and also has a much larger signals recovery space.

![Fig. 5. Comparison of IRLS algorithm and BIRLS algorithm, M=100, L=256, s=8, B=32. The solid curves represent the BIRLS algorithm that outperforms the traditional IRLS algorithm (the dashed curves).](image)

V. CONCLUSIONS

In this paper, the problem of computing the local minima of the nonconvex problem in the block-sparse system is mainly addressed. By replacing the problem $P_{\ell_p/\ell_1}$ with a weighted $\ell_2$-norm, a new algorithm called Block Iteratively Reweighted Least Squares (BIRLS) based on the IRLS is proposed to solve this problem. Experiments are given to show that the reconstruction performance obtained by using our novel method significantly outperforms the IRLS algorithm and the BOMP algorithm.

VI. ACKNOWLEDGMENT

This work was supported by the NSFC-Grants 61273195, CPSF-Grant 2012M511386, ZSFC-Grant Y13F010050, NSFC 61304124 and ZJNSF LY13F010009.

REFERENCES


