Preconditioning for Accelerated Iteratively Reweighted Least Squares in Structured Sparsity Reconstruction

Chen Chen[1], Junzhou Huang[1]*, Lei He[2] and Hongsheng Li[3]
[1]University of Texas at Arlington
[2]Library of Congress
[3]University of Electronic Science and Technology of China

Abstract

In this paper, we propose a novel algorithm for structured sparsity reconstruction. This algorithm is based on the iterative reweighted least squares (IRLS) framework, and accelerated by the preconditioned conjugate gradient method. The convergence rate of the proposed algorithm is almost the same as that of the traditional IRLS algorithms, that is, exponentially fast. Moreover, with the devised preconditioner, the computational cost for each iteration is significantly less than that of traditional IRLS algorithms, which makes it feasible for large scale problems. Besides the fast convergence, this algorithm can be flexibly applied to standard sparsity, group sparsity, and overlapping group sparsity problems. Experiments are conducted on a practical application compressive sensing magnetic resonance imaging. Results demonstrate that the proposed algorithm achieves superior performance over 9 state-of-the-art algorithms in terms of both accuracy and computational cost.

1. Introduction

Compressive sensing (CS) [5] provides the theoretical support for signal reconstruction from undersampled measurements and has been a rather active topic in recent years. If the original data is sparse or compressible, it can be recovered precisely from a small number of measurements. ℓ_1 norm regularization is widely used to induce sparsity and gains great success in many real applications. A general formulation can be written as:

$$\min_x \{F(x) = \frac{1}{2}||Ax - b||_2^2 + \lambda||x||_1\} \tag{1}$$

where $A \in \mathbb{R}^{M\times N}$ is the measurement matrix and $b \in \mathbb{R}^M$ is the vector of measurements; $x \in \mathbb{R}^N$ is the data to be recovered; λ is a positive parameter.

According to structured sparsity theories[2][14], more benefits can be achieved if we could utilize more prior information about the sparsity patterns. For example, the components of the data may be clustered in groups, which is then called group sparse data. Components within the same group tend to be zeros or non-zeros. Sometimes one component may appear in several groups simultaneously, which corresponds to the overlapping group sparsity. A favorable method would be replacing the ℓ_1 norm with $\ell_{2,1}$ norm to model the group sparsity [30][16]:

$$||x||_{2,1} = \sum ||x_{g_i}||_2 \quad i = 1, 2, ..., s \tag{2}$$

where x_{g_i} denotes the components in i-th group and s is the total number of groups. It has been proved that, less measurements are required for structured sparsity recovery, or more precise solution can be obtained with the same number of measurements [2][14][1]. Non-overlapping group sparsity problem is a special case of overlapping group sparsity problems. When the size of every group is one, the problem becomes standard sparsity recovery with ℓ_1 norm regularization. If the generalized problem with overlapping group sparsity is solved, the other two special cases can also be solved.

In literature, many algorithms can be or have already been extended to the non-overlapping group sparsity cases, such as FISTA [3], SPGL1 [4], SpaRSA [29], FOCUSS [10]. However, due to the non-smoothness and non-separableness of the overlapped $\ell_{2,1}$ penalty, there are relatively much fewer algorithms for overlapping group sparsity. SLEP [17], GLO-pridu [21] solve the overlapping sparsity problem by identifying active groups and YALL1 [27] solves it based on the alternating direction method (ADM). However, both SLEP and GLO-pridu are based on proximal methods (e.g. [3]), which cannot achieve a convergence rate better than $F(x^k) - F(x^*) \sim \mathcal{O}(1/k^2)$, where x^* denotes the optimal solution and k is the iteration number. YAL-

*Corresponding author. Email: jzhuang@uta.edu.

L1 relaxes the original problem with augmented Lagrangian and iteratively minimizes the subproblems based on the variable splitting method. Generally, the convergence rate of ADM is no better than $\mathcal{O}(1/k)$ in sparse recovery problems. Although they are very efficient in each iteration, a large number of iterations may be required due to the relatively slow convergence rate. On the other side, the iterative reweighted least squares (IRLS) algorithms have been proved that they converge exponentially fast [9] [6]. Unfortunately, conventional IRLS algorithms contain a large scale inverse operation in each step, which makes them still much more computationally expensive than the fastest proximal methods such as FISTA [1]. In addition, it is unknown how to extend them to the problem with overlapping group sparsity.

In this paper, we propose a novel method for structured sparsity reconstruction based on the IRLS framework. It preserves the fast convergence performance of traditional IRLS, which only needs a few reweighted iterations to achieve an accurate solution. Moreover, we propose a new "pseudo-diagonal" type preconditioner to significantly accelerate the inverse subproblem with preconditioned conjugate gradient (PCG) method. This preconditioner is much more precise than conventional Jacobi diagonal preconditioner. In addition, the proposed preconditioner can be applied even when A is an operator, which is not feasible for most existing preconditioners in PCG methods. Besides the efficiency and fast convergence rate, the proposed algorithm can be flexibly applied to different types of group structures with either overlapping or non-overlapping. For different group configurations, the only change in our scheme is on the group configuration matrix. Extensive experiments are conducted to validate the proposed algorithm on compressive sensing magnetic resonance imaging (CS-MRI) for structured sparsity reconstruction. All experimental results demonstrate that the proposed algorithm outperforms the state-of-the-art methods in terms of accuracy and computational speed.

2. Related Work

2.1. IRLS for Standard Sparsity

The conventional IRLS algorithms solve the standard sparse problem in constrained form:

$$\min_x ||x||_1, \text{subject to } Ax = b \tag{3}$$

In these methods, the ℓ_1 norm is replaced by a reweighted ℓ_2 norm [6]:

$$\min_x x^T W x, \text{subject to } Ax = b \tag{4}$$

The diagonal weight matrix W in the k-th iteration is computed from the solution of the current iteration x^k, that is, the diagonal elements $W_i^k = |x_i^k|^{-1}$. With current weights W^k, it has a closed form solution for x^{k+1}:

$$x^{k+1} = (W^k)^{-1}A^T(A(W^k)^{-1}A^T)^{-1}b \tag{5}$$

The algorithm can be summarized in Algorithm 1. It has been proved that the IRLS algorithm converges exponentially fast under mild conditions [9]:

$$||x^k - x^*||_1 \leq \mu||x^{k-1} - x^*||_1 \leq \mu^k||x^0 - x^*||_1 \tag{6}$$

where μ is a fixed constant with $\mu < 1$. However, this algorithm is rarely used in compressive sensing applications especially for large scale problems. That is because the inverse of $A(W^k)^{-1}A^T$ takes $\mathcal{O}(M^3)$ if A is a $M \times N$ sampling matrix. Even with higher convergence rate, traditional IRLS still cannot compete with the fastest first-order algorithms such as FISTA [3] (some results are shown in [1]). Moreover, none of previous IRLS methods [6][9][10] could solve the overlapping group sparsity problems, which significantly limits the usage.

Algorithm 1 IRLS

Input: A,b,x^1,$k = 1$
while not meet the stopping criterion **do**
  Update W: $W_i^k = |x_i^k|^{-1} \quad \forall\, W_i^k$
  Update x: $x^{k+1} = (W^k)^{-1}A^T(A(W^k)^{-1}A^T)^{-1}b$
  Update $k = k + 1$
end while

2.2. Majorization Minimization Method

Both the traditional IRLS algorithms and the proposed algorithm can be derived from majorization minimization (MM) method. MM method is a general scheme to construct an optimization algorithm. We want to minimize a function $f(x)$ and x^k denote the solution of the k-th iteration. It may be difficult to minimize the function $f(x)$ directly. Alternatively, we could minimize its majorizer $g(x|x^k)$ by MM theory. $g(x|x^k)$ should satisfy that:

$$g(x|x^k) \geq f(x) \quad \forall x \tag{7}$$
$$g(x^k|x^k) = f(x^k) \tag{8}$$

It indicates that $g(x|x^k)$ is always greater than or equal to $f(x)$ and is tangent to $f(x)$ at x^k. One can select any majorizer function that satisfy the above conditions. A brief review of MM method is introduced in [15]. Minimizing the majorizer function $g(x|x^k)$ also decreases the actual function value of $f(x)$ due to the properties (7) (8).

3. Algorithm

3.1. General model

Consider a general problem for overlapping group sparsity [30][16]:

$$\min_x \{F(x) = \frac{1}{2}||Ax-b||_2^2 + \lambda||G\Phi x||_{2,1}\} \tag{9}$$

where Φ is the sparse basis and is optional. A good choice of Φ for natural images/signals would be an orthogonal wavelet transform. G is a binary matrix for group configuration, which is constructed by rows of the identity matrix. By different settings of G, this model can handle overlapping group, non-overlaping group and standard sparse problems. Simple examples of G for different types of group sparse problems are shown in Figure 1. Although G may have large scales, it can be efficiently represented by a sparse matrix. This kind of indexing matrix has been used in the previous work YALL1 [27].

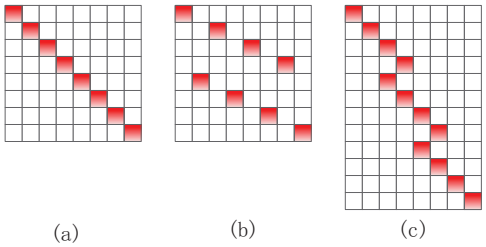


Figure 1. Examples of group configuration matrix G for a signal of size 8. The red elements denote 1s and white elements denote 0s. (a) standard sparsity case where G is the identical matrix. (b) non-overlapping groups of [1,3,5,7] and [2,4,6,8]. (c) overlapping groups of [1,2,3,4], [3,4,5,6] and [5,6,7,8]. Their group sizes are 1,4 and 4, respectively.

Consider the Young's inequality holding for a general function $g(\cdot) : \mathbb{R} \longmapsto \mathbb{R}$:

$$\sqrt{g(x)} \leq \frac{\sqrt{g(y)}}{2} + \frac{g(x)}{2\sqrt{g(y)}} \tag{10}$$

with $g(y) > 0$ and $g(x) \geq 0$. The equality holds only when $g(x) = g(y)$. Based on this, we have

$$\begin{aligned} ||G\Phi x||_{2,1} &= \sum_{i=1}^{s} ||(G\Phi x)_{g_i}||_2 \\ &\leq \sum_{i=1}^{s} [\frac{||(G\Phi x^k)_{g_i}||_2}{2} + \frac{||(G\Phi x)_{g_i}||_2^2}{2||(G\Phi x^k)_{g_i}||_2}] \end{aligned} \tag{11}$$

Writing it in matrix form and we can majorize $F(x)$ by MM method:

$$\begin{aligned} Q(x|x^k) = \frac{1}{2}||Ax-b||_2^2 + \frac{\lambda}{2}x^T\Phi^T G^T W^k G\Phi x \\ + \frac{\lambda}{2}\sum_{i=1}^{s}\frac{1}{W_{g_i}^k} \end{aligned} \tag{12}$$

where Φ^T denotes the inverse transform of Φ; W^k is the group-wise weights. The weight of i-th group $W_{g_i}^k$ can be obtained by:

$$W_{g_i}^k = (||(G\Phi x^k)_{g_i}||_2^2 + \epsilon)^{-1/2} \tag{13}$$

ϵ is very small number to avoid the weight becoming infinity. Supposing the signal x to be recovered is of length N and G is a N'-by-N matrix, then W^k is a N'-by-N' diagonal matrix and has the following form:

$$W^k = \begin{Bmatrix} W_{g_1}^k & & & & & \\ & \ddots & & & & \\ & & W_{g_1}^k & & & \\ & & & \ddots & & \\ & & & & W_{g_s}^k & \\ & & & & & W_{g_s}^k \end{Bmatrix} \tag{14}$$

where each group-wise weight $W_{g_i}^k$ is duplicated $|g_i|$ times and $|g_i|$ denotes the size of the i-th group. One can find that the group-wise weights are all related to G. With different settings of G, the group-wise weights are directly acquired. Variant-size group sparsity problems also can be flexibly handled in this model. An interesting case would be the standard sparse problem, where each group contains only one element and the group-wise weight matrix W is the same as that in IRLS algorithm [9] [6].

Now the problem becomes:

$$x^{k+1} = \arg\min_x Q(x|x^k) \tag{15}$$

Note that $W_{g_i}^k$ is independent on x and can be considered as a constant. We iteratively update W^k with x^k and solve x^{k+1} based on current W^k. Our algorithm is also a IRLS type algorithm with exponentially fast convergence rate.

3.2. Accelerating with PCG

In each iteration, W^k is easy to update with (14) (13). To solve (15), a simple way is to let the first order gradient of $Q(x|x^k)$ be zero as it is a quadratic convex function:

$$(A^TA + \lambda\Phi^T G^T W^k G\Phi)x - A^Tb = 0 \tag{16}$$

The way to solve (16) determines the efficiency of the whole algorithm. The exact inverse of the system matrix $S = A^TA + \lambda\Phi^T G^T W^k G\Phi$ takes $\mathcal{O}(N^3)$ time. It is impractical to compute S^{-1} for many cases especially when the size of S is huge. An alternative way is to approximate the solution of (16) with classical conjugate gradient (CG) decent method. It is much faster than computing the exact solution. Besides CG, a better way is the preconditioned conjugate gradient (PCG) method. The design of preconditioner is problem-dependent, which should be

as close as possible to the system matrix S and can be inversed efficiently. Therefore, it is not an easy task to design a good preconditioner in general due to this tradeoff. For signal/image reconstruction, such preconditioner has not been found in existing IRLS algorithms [6][9][10].

By observing that S is usually diagonally dominant in reconstruction problems, we define a new preconditioner for best approximation in Frobenius norm $||\cdot||_F$:

$$P = \arg\min_{X \in \mathcal{D}} ||S - X||_F \tag{17}$$

where $\mathcal{D}$ denotes the class of diagonal or "pseudo-diagonal" matrices. Here, the pseudo-diagonal matrix means a non-diagonal matrix whose inverse can be obtained efficiently like a diagonal matrix with $\mathcal{O}(N)$ time. Please note that the $G^T W^k G$ is always diagonal for any kind of G in Figure 1. Due to the strong constraint, the possible diagonal or "pseudo-diagonal" candidates for (17) are enumerable. When Φ is the wavelet transform, it is not hard to find $P = (\overline{A^T A} I + \lambda \Phi^T G^T W^k G \Phi)$ where $\overline{A^T A}$ is the mean of diagonal elements of $A^T A$ and I denotes the identity matrix. As A is known for the application, $\overline{A^T A}$ can be pre-estimated before the first iteration and is fixed for each iteration. Therefore in each iteration, $P^{-1} = \Phi^T (\overline{A^T A} I + \lambda G^T W^k G)^{-1} \Phi$ can be obtained with linear time.

Algorithm 2 FIRLS

Input: A,b,x^1, G, λ, $k = 1$
while not meet the stopping criterion **do**
  Update W^k by (14) (13)
  Update $S = A^T A + \lambda \Phi^T G^T W^k G \Phi$
  Update $P = \Phi^T (\overline{A^T A} I + \lambda G^T W^k G) \Phi$
  while not meet the PCG stopping criterion **do**
    Update x^{k+1} by PCG for $Sx = A^T b$ with preconditioner P
  end while
  Update $k = k + 1$
end while

Several advantages of the proposed preconditioner can be found when compared with existing ones [22] [25][23]. To get the inverse, fast Fourier transforms are involved in recent circulant preconditioners that used in image deblurring [22] [25], while our model only requires linear time to obtain P^{-1}. Compared with conventional Jacobi preconditioner that is used in total variation minimizing [23], we do not discard all non-diagonal information and therefore the preconditioner is more accurate. Moreover, because it is required to calculate the exact values of S for existing preconditioners [22] [25][23], there is no known way to extend them to the case when A or Φ is an operator. Interestingly, the conventional Jacobi preconditioner can be derived by (17), when the original data is sparse (i.e. $\Phi = 1$) and A is a numerical matrix.

Now, our method can be summarized in Algorithm 2. We called it Fast Iterative Reweighted Least Squares (FIRLS). Although our algorithm has double loops, we observe that only 10 to 30 PCG iterations are sufficient to obtain a solution very close to the optimal one for problem (16). In each inner PCG iteration, the dominated cost is by applying S and P^{-1}, which is denoted by $\mathcal{O}(\mathcal{C}_S + \mathcal{C}_P)$. When A and Φ are dense matrices, $\mathcal{O}(\mathcal{C}_S + \mathcal{C}_P) = \mathcal{O}(N^2)$. When A and Φ are the partial Fourier transform and wavelet transform in CS-MRI [18], it is $\mathcal{O}(N \log N)$.

4. Application: Compressive Sensing MRI

Compressive sensing MRI (CS-MRI) [18] is one of the most successful application of compressive sensing and sparsity regularization. Partial but not full k-space data is acquired and the final MR image can be reconstructed by exploiting the sparsity of the image. With little information loss, this scheme could significant accelerate MRI acquisition. In CS-MRI, $A = RF$ is an undersampled Fourier operator, where F is the Fourier transform and $R \in \mathbb{R}^{M \times N}$ is a selection matrix containing M rows of the identity matrix. Therefore, $A^T A = F^T R^T R F$ is diagonally dominant as $R^T R$ is diagonal. Based on (17), $\overline{A^T A}$ is just the sampling ratio (a fixed scalar).

The MR images are often assumed to be sparse under the wavelet basis [18][19][13][12]. Furthermore, the wavelet coefficients of a natural image yield a quadtree. If a coefficient is zero or nonzero, its parent coefficient also tends to be zero or nonzero. This wavelet tree structure has already been successfully utilized in MR image reconstruction [7, 8]. This problem is approximated by overlapping group sparsity, where each coefficient and its parent coefficient are assigned into one group [7]. Due to the difficulty of overlapping group sparsity regularization, conventional wavelet reconstruction methods [18][19][13] cannot solve it. Compared with the variable splitting strategy used in [7], small changes in the group configuration matrix G will make our algorithm applicable for the tree-based overlapping group sparsity reconstruction.

To assist clinic diagnose, multiple MR images with different contrast are often acquired simultaneously from the same anatomical cross section. For example, T1 and T2 weighted MR images could distinguish fat and edema better, respectively. Different from the CS-MRI for individual MR imaging, multi-contrast reconstruction for weighted MR images means the simultaneous reconstruction of multiple T1/T2-weighted MR images. Joint sparsity across different contrasts is widely used in recent multi-contrast reconstruction methods [20][11]. Following their group settings, our method also could be used to accelerate the reconstruction.

5. Experiments

5.1. Experiment Setup

The experiments are conducted using Matlab on a desktop with 3.4GHz Intel core i7 3770 CPU. First, we validate the efficiency of the proposed preconditioner and convergence speed of our method. The proposed method is then compared with state-of-the-art algorithms on the application of tree structured MRI and multi-contrast MRI. To avoid confusion, we denote the standard sparse version of our algorithm as FIRLS_L1 and the overlapping group sparse version as FIRLS_OG. The non-overlapping group version FIRLS_MT is compared in multi-contrast MRI reconstruction. For fair comparisons, all codes are downloaded from the authors' websites and we carefully follow their experiment setup.

Note that some algorithms need a very small number of iterations to converge (higher convergence rate), while they cost more time in each iteration (higher complexity). The others take less time in each iteration, however, more iterations are required. As we are interested in fast reconstruction, an algorithm is said to be better if it can achieve higher reconstruction accuracy with less computational time.

5.2. The Accuracy of the Proposed Preconditioner

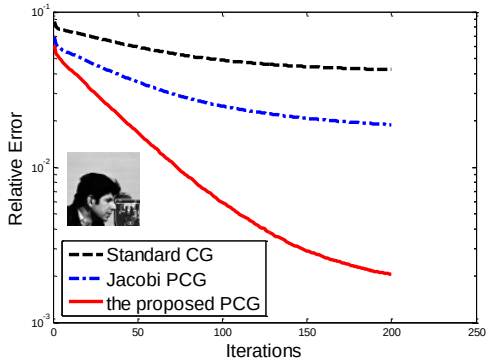


Figure 2. Convergence rate comparison among standard CG, Jacobi PCG and the proposed PCG in terms of relative errors.

One of our contributions is the proposed pseudo-diagonal preconditioner for sparse recovery. First, we conduct an experiment to validate its effectiveness. Without loss of generality, a patch (64×64) cropped from the cameraman image is used for reconstruction, which is feasible to obtain the closed form solution of S^{-1} for evaluation. As most existing preconditioners cannot support the inverse of operators, the sampling matrix is set as the random projection and Φ is a dense matrix for wavelet basis here. Figure 2 demonstrates the performance of the proposed PCG compared with Jacobi PCG and standard CG for the problem (16). The performance of the proposed PCG with less than 50 iterations is better than that of CG and Jacobi PCG with 200 iterations. Although Jacobi preconditioner is diagonal, it remove all the non-diagonal elements which makes the preconditioner less precise. This experiment demonstrates that the inner loop subproblem in our algorithm is solved efficiently due to the advantage of the proposed preconditioner.

5.3. Convergence Rate and Computational Complexity

The superiority of the proposed method comes from its fast convergence rate, that only a small number of iterations can achieve high reconstruction accuracy. In addition, each iteration has lower computational cost. To validate its fast convergence rate, we compare it with three existing algorithms with known convergence rate. They are the IST algorithm SpaRSA [29], FISTA [3] and IRLS algorithm FOCUSS [10], with $\mathcal{O}(1/k)$, $\mathcal{O}(1/k^2)$ and exponential convergence rates, respectively. Mean squared error (MSE) is used as the evaluation metric.

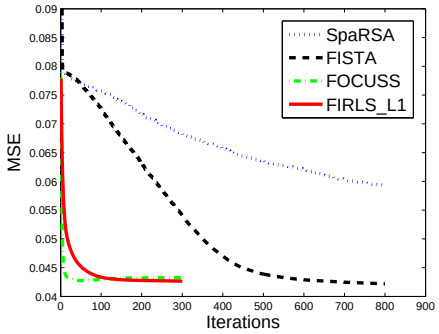


Figure 3. Convergence Rate Comparison among FOCUSS, FISTA and SpaRSA.

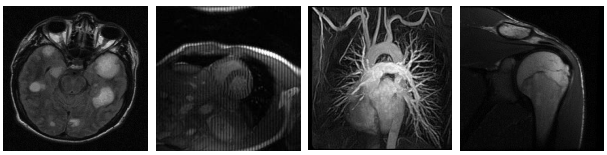
The test data is a random 1D signal of length 4000, with 10% elements being non-zeros. The number of measurements are 800. Figure 3 demonstrates the comparison. In each iteration, FOCUSS needs to compute the inverse of a large-scale matrix, and the proposed method uses 30 PCG iterations to approximate the inverse. Both FOCUSS and the proposed method converge within 200 iterations. FISTA tends to converge at about 800 iterations. However, SpaRSA requires much more than 800 iterations to converge. Table 1 lists the reconstruction results at different CPU time between FOCUSS and the proposed method. The proposed algorithm always achieves more accurate result in much less time. After convergence, the 0.0005 difference in terms of MSE may be caused by approximation or rounding errors. With the size of the data becomes larger, the time cost of FOCUSS will increase at a cubic speed. More importantly, it is unknown how to solve the overlapping group sparse problem with FOCUSS.

Table 1. Computational cost comparison between FOCUSS [10] and the proposed method

	FOCUSS [10]			FIRLS_L1		
TIME (SECONDS)	64.8	110.8	727.7	10.5	29.8	120.2
MSE	0.0485	0.0442	0.0432	0.0481	0.0440	0.0427

5.4. CS-MRI with Wavelet Tree Sparsity

A practical application CS-MRI is used to validate the performance of our method for overlapping group sparsity recovery. The sampling matrix is the partial Fourier transform. We follow the sampling strategy of previous works [19][13][7], which randomly choose more Fourier coefficients from low frequency and less on high frequency. The sampling ratio is defined as the number of sampled measurements divided by the total size of the signal/image. Similar as previous works, Signal-to-Noise Ratio (SNR) is used for result evaluation: $SNR = 10\log_{10}(V_s/V_n)$ where V_n is the MSE between the original signal x_0 and the solution x; $V_s = var(x_0)$ denotes the variance of the values in x_0. The test images used in previous works [19][13][7] are shown in Figure 4, with the same size 256×256.



(a) (b) (c) (d)

Figure 4. The original images: (a) Brain; (b) Cardiac; (c) Chest; (d) Shoulder.

Tree-structured CS-MRI method [7] has been shown to be superior to standard CS-MRI methods [18][19][13]. Therefore, we compare our algorithm with two latest and fastest algorithms, turbo AMP [26] and WaTMRI [7]. In addition, overlapping group sparsity algorithms SLEP [17] and YALL1 [27] are also compared. The total number of iterations is 100 except that turbo AMP only runs 10 iterations due to its higher time complexity.

A visual comparison on the Brain image is shown in Figure 5, with a 25% sampling ratio. Artifacts can be found on the results by SLEP [17] and YALL1 [27]. The image reconstructed by the AMP [26] tends to be blurry when compared with the original. Our method and WaTMRI [7] produce the most accurate results. Note that total variation regularization is incorporated in WaTMRI [7], but our method still achieves higher SNR. Besides SNR, we also compare the mean structural similarity [28] (MSSIM) of different images, which mimics the human visual system. The MSSIM for the images recovered by AMP [26], WaTMRI [7], SLEP [17], YALL1 [27] and the proposed method are 0.8890, 0.8654, 0.8561, 0.7857 and 0.9009. In terms of MSSIM, our method still has the best performance, which

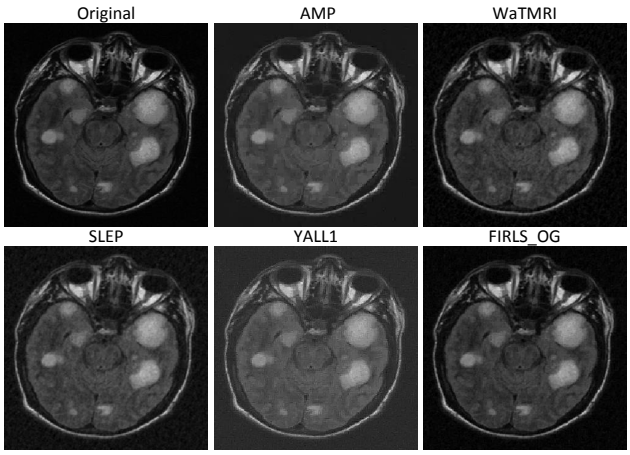


Figure 5. Visual comparison on the Brain image with 25% sampling. The SNRs of AMP [26], WaTMRI [7], SLEP [17], YALL1 [27] and the proposed method are 15.91, 16.72, 16.49, 12.86 and 18.39, respectively.

is consistent with the observation in terms of SNR.

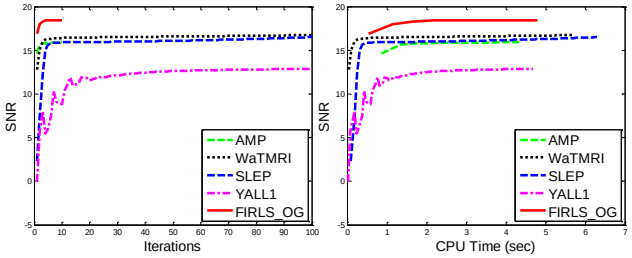


Figure 6. Convergence speed comparison on the Brain image with 25% sampling. Left: SNR vs outer loop iterations. Right: SNR vs CPU time. The SNRs of reconstructed images with these algorithms are 15.91, 16.72, 16.49, 12.86 and 18.39 respectively. The time costs are 4.34 seconds, 5.73 seconds, 6.28 seconds, 4.71 seconds and 4.80 seconds, respectively.

The corresponding convergence speed of the this experiment is presented in Figure 6. From SNR versus outer loop iterations, the proposed algorithm far exceeds that of all other algorithms, which is due to the fast convergence rate of IRLS. However, there is no known convergence rate better than $\mathcal{O}(1/k^2)$ for WaTMRI and SLEP, and $\mathcal{O}(1/k)$ for YALL1, respectively. These results are consistent with that in previous work [7]. For the same number of total iterations, the computational cost of our method is comparable to the fastest one YALL1, and it significantly outperforms

YALL1 in terms of reconstruction accuracy. SLEP has the same formulation as ours. To reach our result in this experiment, it requires around 500 iterations and cost about 43 seconds. Similar results can be obtained on the other testing images. The results on the four images with different sampling ratios are listed in Table 2.

Table 2. Average SNR (dB) comparisons on the four MR images with various sampling ratios.

Sampling Ratio	20%	23%	25%	28%	30%
AMP [26]	11.64	15.7	16.43	17.08	17.44
WaTMRI [7]	15.56	17.43	18.23	19.22	20.45
SLEP [17]	11.59	16.51	17.36	18.51	20.07
YALL1 [27]	12.13	13.29	14.12	15.29	16.07
FIRLS_OG	**15.67**	**18.78**	**19.43**	**20.53**	**21.52**

5.5. Multi-contrast MRI with Joint Sparsity

The multi-contrast MR images are extracted from the SRI24 Multi-Channel Brain Atlas Data [24]. An example of the test images is shown in Figure 7. We compare our method with the fastest multi-contrast MRI methods [20][11], which use the algorithms SPGL1_MMV [4] and FCSA_MT to solve the corresponding problems, respectively. The experiment setup is the similar as in the previous experiments, except group setting is constructed for joint sparsity (non-overlapping) case.

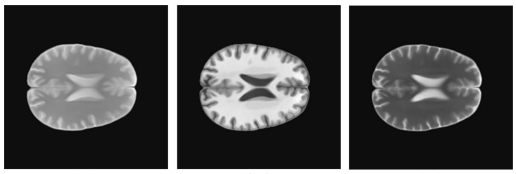
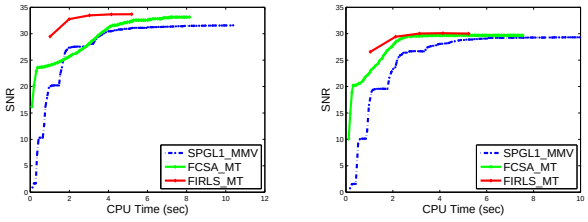


Figure 7. The original images for multi-contrast MRI.

Figure 8 shows the performance comparisons among SPGL1_MMV [4], FCSA_MT [11] and the proposed method FIRLS_MT on the example images shown in Figure 7. Each algorithm runs 100 iterations. After convergence, three algorithms achieve similar accuracy for 20% sampling and SPGL1 is only slightly worse than others for 25% sampling. From the curves, our method is always better than SPGL1_MMV and FCSA_MT, i.e., higher accuracy for the same reconstruction time.

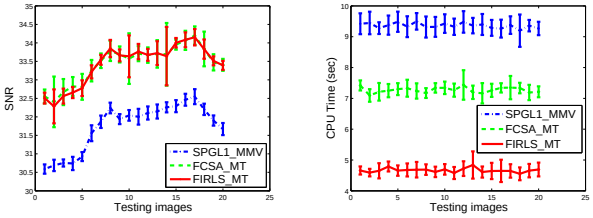
To demonstrate the fast convergence property of our method, we then conduct experiments on 20 set images (i.e. total 60 images) that extracted from SRI24. Different from the tree-based CS-MRI, each algorithm for non-overlapping group sparsity converges much faster. Therefore, we terminate these algorithms when a fixed tolerance reached (e.g. 10^{-3} of relative change on x). To reduce randomness, all



(a) (b)

Figure 8. (a) Performance comparison for multi-contrast MRI with 25% sampling. The average time costs of SPGL1_MMV, FCSA_MT, and the proposed method are 10.38 seconds, 8.15 seconds, 5.19 seconds. Their average SNRs are 31.58, 33.12 and 33.69. (b) Performance comparison for multi-contrast MRI with 20% sampling. Their average time costs are 9.98 seconds, 7.54 seconds, 5.23 seconds. Their average SNRs are 29.31, 29.69 and 30.01.

algorithms run 100 times and the reconstruction results are shown in Figure 9. With 25% sampling, the accuracy of our method is almost the same as FCSA_MT, and always better than SPGL1. To reach the convergence, our method is consistently faster than the other two algorithms. These results demonstrate the efficiency of proposed method.



(a) (b)

Figure 9. Performance comparison on 60 images from SRI24 dataset with 25% sampling. (a) SNR comparison. (b) CPU time comparison. The average convergence time for SPGL1, FCSA_MT and the proposed FIRLS_MT is 9.3 seconds, 7.2 seconds, 4.6 seconds, respectively.

5.6. Discussion

The first and second experiments validate the fast convergence speed of our method due to the proposed preconditioner. The advantages over the state-of-the-arts are further validated on practical application CS-MRI on two group settings, overlapping groups with tree sparsity and non-overlapping groups with joint sparsity. Although results on both problems are promising, some difference can be found. The non-overlapping group sparsity problem is often easier to solve. For example, the subproblem in FISTA has closed form solution for joint sparsity but not for overlapping group sparsity. However, our method has similar difficulty for non-overlapping and overlapping group sparsity.

That is why our method outperforms the fastest methods on joint sparsity reconstruction, and significantly outperforms those for tree-sparsity reconstruction.

The superior performance of the proposed preconditioner also attributes to the structure of the system matrix S, which is often diagonally dominant in reconstruction problems (e.g. A is random projection or partial Fourier transform). It can be applied to other applications where S is not diagonally dominant (e.g. image blurring), and still be more accurate than Jacobi preconditioner as it keeps more non-diagonal information.

6. Conclusion

We have proposed a novel method for structured sparsity recovery, which is of the IRLS type and preserves its *fast convergence rate*. The subproblem in our scheme is accelerated by the PCG method with a new pseudo-diagonal preconditioner. Due to the *high accuracy* and *efficiency* of this preconditioner, the subproblem can be solved in very low cost, even when it contains transforming operations. Extensive experimental results have demonstrated the *flexibility*, *effectiveness* and *efficiency* of this method on CS-MRI.

References

[1] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski. Optimization with sparsity-inducing penalties. *Foundations and Trends in Machine Learning*, 4(1):1–106, 2011. 1, 2

[2] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde. Model-based compressive sensing. *IEEE Trans. Information Theory*, 56:1982–2001, 2010. 1

[3] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sciences*, 2(1):183–202, 2009. 1, 2, 5

[4] E. Berg and M. Friedlander. Probing the pareto frontier for basis pursuit solutions. *SIAM J. Scientific Computing*, 31:890–912, 2008. 1, 7

[5] E. Candes, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Information Theory*, 52(2):489–509, 2006. 1

[6] R. Chartrand and W. Yin. Iteratively reweighted algorithms for compressive sensing. In *ICASSP*, 2008. 2, 3, 4

[7] C. Chen and J. Huang. Compressive sensing MRI with wavelet tree sparsity. In *NIPS*, 2012. 4, 6, 7

[8] C. Chen and J. Huang. The benefit of tree sparsity in accelerated MRI. *Medical image analysis*, 2013. 4

[9] I. Daubechies, R. DeVore, M. Fornasier, and S. Gunturk. Iteratively reweighted least squares minimization for sparse recovery. *Communications on Pure and Applied Mathematics*, 63(1):1–38, 2010. 2, 3, 4

[10] I. F. Gorodnitsky and B. D. Rao. Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm. *IEEE Trans. Signal Processing*, 45(3):600–616, 1997. 1, 2, 4, 5, 6

[11] J. Huang, C. Chen, and L. Axel. Fast multi-contrast mri reconstruction. In *MICCAI*, 2012. 4, 7

[12] J. Huang, S. Zhang, H. Li, and D. Metaxas. Composite splitting algorithms for convex optimization. *Comput. Vis. Image Und.*, 115(12):1610–1622, 2011. 4

[13] J. Huang, S. Zhang, and D. Metaxas. Efficient MR Image Reconstruction for Compressed MR Imaging. *Medical Image Analysis*, 15(5):670–679, 2011. 4, 6

[14] J. Huang, T. Zhang, and D. Metaxas. Learning with structured sparsity. In *ICML*, 2009. 1

[15] D. Hunter and K. Lange. A tutorial on MM algorithms. *The American Statistician*, 58:30–37, 2004. 2

[16] L. Jacob, G. Obozinski, and J.-P. Vert. Group lasso with overlap and graph lasso. In *ICML*, 2009. 1, 3

[17] J. Liu, S. Ji, and J. Ye. *SLEP: Sparse Learning with Efficient Projections*. Arizona State University, 2009. 1, 6, 7

[18] M. Lustig, D. Donoho, and J. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, 58:1182–1195, 2007. 4, 6

[19] S. Ma, W. Yin, Y. Zhang, and A. Chakraborty. An efficient algorithm for compressed MR imaging using total variation and wavelets. In *CVPR*, 2008. 4, 6

[20] A. Majumdar and R. Ward. Joint reconstruction of multiecho MR images using correlated sparsity. *Magnetic Resonance Imaging*, 29:899–906, 2011. 4, 7

[21] S. Mosci, S. Villa, A. Verri, and L. Rosasco. A primal-dual algorithm for group sparse regularization with overlapping groups. In *NIPS*, 2010. 1

[22] G. Papandreou and A. Yuille. Efficient variational inference in large-scale bayesian compressed sensing. In *ICCV Workshop*, 2011. 4

[23] P. Rodríguez and B. Wohlberg. Efficient minimization method for a generalized total variation functional. *IEEE Trans. Image Processing*, 18(2):322–332, 2009. 4

[24] T. Rohlfing, N. Z. NM, E. Sullivan, and A. Pfefferbaum. The sri24 multichannel atlas of normal adult human brain structure. *Human Brain Mapping*, 31:798–819, 2010. 7

[25] A. B. S. Lefkimmiatis and M. Unser. Hessian-based norm regularization for image restoration with biomedical applications. *IEEE Trans. Image Processing*, 21(3):983–995, 2012. 4

[26] S. Som and P. Schniter. Compressive imaging using approximate message passing and a markov-tree prior. *IEEE Trans. Signal Processing*, 60:3439–3448, 2012. 6, 7

[27] W. Y. W. Deng and Y. Zhang. Group sparse optimization by alternating direction method. Technical report, Rice University, 2011. 1, 3, 6, 7

[28] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Processing*, 13(4):600–612, 2004. 6

[29] S. Wright, R. Nowak, and M. Figueiredo. Sparse reconstruction by separable approximation. *IEEE Trans. Signal Processing*, 57:2479–2493, 2009. 1, 5

[30] M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Series B Stat. Methodol.*, 68(1):49–67, 2005. 1, 3