Medical Image Analysis 49 (2018) 141–152

Contents lists available at ScienceDirect

Medical Image Analysis

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

Fast iteratively reweighted least squares algorithms for analysis-based sparse reconstruction

Chen Chen[a], Lei He[b], Hongsheng Li[c], Junzhou Huang[d,*]

[a] Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, USA
[b] Office of Strategic Initiatives, Library of Congress, USA
[c] Department of Electronic Engineering, the Chinese University of Hong Kong, Hong Kong
[d] Department of Computer Science and Engineering, University of Texas at Arlington, USA

ARTICLE INFO

Article history:
Received 5 December 2017
Revised 4 August 2018
Accepted 6 August 2018
Available online 7 August 2018

Keywords:
Structured sparsity
Total variation
Overlapping group sparsity
Image reconstruction
Preconditioned conjugate gradient descent

ABSTRACT

In this paper, we propose a novel algorithm for analysis-based sparsity reconstruction. It can solve the generalized problem by structured sparsity regularization with an orthogonal basis and total variation (TV) regularization. The proposed algorithm is based on the iterative reweighted least squares (IRLS) framework, and is accelerated by the preconditioned conjugate gradient method. The proposed method is motivated by that, the Hessian matrix for many applications is diagonally dominant. The convergence rate of the proposed algorithm is empirically shown to be almost the same as that of the traditional IRLS algorithms, that is, linear convergence. Moreover, with the specifically devised preconditioner, the computational cost for the subproblem is significantly less than that of traditional IRLS algorithms, which enables our approach to handle large scale problems. In addition to the fast convergence, it is straightforward to apply our method to standard sparsity, group sparsity, overlapping group sparsity and TV based problems. Experiments are conducted on practical applications of compressive sensing magnetic resonance imaging. Extensive results demonstrate that the proposed algorithm achieves superior performance over 14 state-of-the-art algorithms in terms of both accuracy and computational cost.

1. Introduction

Ill-posed problems widely exist in medical imaging and computer vision. In order to seek a meaningful solution, regularization is often used if we have certain prior knowledge. With the emerging of compressive sensing (CS) (Candes et al., 2006; Donoho, 2006), sparsity regularization has been an active topic in recent years. If the original data is sparse or compressible, it can be recovered precisely from a small number of measurements. The ℓ_1 norm is usually used to induce sparsity and gains great success in many real applications. The optimization problems can be written as:

$$\min_{\mathbf{x}}\{F(\mathbf{x}) = \frac{1}{2}||\mathbf{A}\mathbf{x}-\mathbf{b}||_2^2 + \lambda||\mathbf{x}||_1\}, \tag{1}$$

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

According to structured sparsity theories (Baraniuk et al., 2010; Huang et al., 2011b), 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 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 (Jacob et al., 2009). A favorable method would be replacing the ℓ_1 norm with $\ell_{2,1}$ norm to model the group sparsity (Yuan and Lin, 2005):

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

where $\mathbf{x}_{g_i}$ denotes the components in ith group and s is the total number of groups. It has been proven that, fewer measurements are required for structured sparsity recovery, or more accurate solution can be obtained with the same number of measurements (Baraniuk et al., 2010; Huang et al., 2011b; Bach et al., 2011).

In many real-world applications, the data itself is not sparse, but it can be sparsely represented in some transformation domains. This leads to the analysis-based sparsity regularization problem:

$$\min_{\mathbf{x}}\{F(\mathbf{x}) = \frac{1}{2}||\mathbf{A}\mathbf{x}-\mathbf{b}||_2^2 + \lambda||\Psi\mathbf{x}||_{2,1}\}, \tag{3}$$

* Corresponding author.
E-mail address: jzhuang@uta.edu (J. Huang).

https://doi.org/10.1016/j.media.2018.08.002
1361-8415/

where Ψ denotes some sparifying operator, e.g., the wavelet transform or finite difference operator. In addition to the analysis-based sparsity model, the synthesis sparsity model is also widely used in signal processing (Cleju et al., 2012). An empirical study shows that analysis-based recovery works better for a large number of signals, but is less robust for approximately sparse signals or when fewer measurements are available (Cleju et al., 2012). In this article, we are interested in the analysis-based sparsity model, where the sparsifying operator Ψ is a composite of the finite difference matrices or a group configuration matrix combined with an orthogonal sparsifying basis (See Section 3 and dummyTXdummy- 4 for more details). The standard sparsity and non-overlapping group sparsity minimization problem are special cases of problem (3). In this work, we focus on the image reconstruction applications where A is an undersampling matrix/operator.

When Ψ is an orthogonal basis, many efficient algorithms can be used to solve the standard sparsity and non-overlapping group sparsity minimization, such as FISTA (Beck and Teboulle, 2009b), SPGL1 (Berg and Friedlander, 2008), SpaRSA (Wright et al., 2009), FOCUSS (Gorodnitsky and Rao, 1997). However, there are relatively fewer algorithms for overlapping group sparsity, due to the difficulty of dealing with the non-smoothness and non-separability of the overlapping $\ell_{2,1}$ penalty. SLEP (Liu et al., 2009; Yuan et al., 2013), GLO-pridu (Mosci et al., 2010) solve the overlapping group sparsity problem by identifying active groups, and YALL1 (Deng et al., 2011) solves it with the alternating direction method (ADM). Both SLEP and GLO-pridu are based on the proximal gradient descent method (e.g., FISTA Beck and Teboulle, 2009b), which cannot achieve a convergence rate better than $F(\mathbf{x}^k) - F(\mathbf{x}^*) \sim \mathcal{O}(1/k^2)$, where $\mathbf{x}^*$ denotes an optimal solution and k is the iteration number. YALL1 relaxes the original problem and iteratively minimizes the corresponding 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 hand, the iterative reweighted least squares (IRLS) algorithms have been proven to converge exponentially fast (Daubechies et al., 2010) if $\mathbf{A}$ satisfies the restricted isometry property (RIP). 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. Some other algorithms can solve the sparsity or group sparsity based denoising problems efficiently (Chen and Selesnick, 2014b; 2014a), but they cannot solve the general linear inverse problem (3) directly.

Another special case of (3) is the total variation (TV) reconstruction problem, where Ψ denotes the first-order finite difference matrices and is non-orthogonal. There are efficient algorithms specially designed for TV reconstruction such as RecPF (Yang et al., 2010) and SALSA (Afonso et al., 2010). Both of them are relaxed by ADM. The efficient transformation in RecPF (Yang et al., 2010) requires that $\mathbf{A}^T\mathbf{A}$ can be diagonalized by the Fourier transform, while SALSA (Afonso et al., 2010) requires $\mathbf{A}\mathbf{A}^T = \mathbf{I}$. Due to these restrictions, these two methods cannot be applied to certain reconstruction applications, e.g., CS imaging (Xiao and Yang, 2010). Moreover, it is unknown how to extend them to solve the joint total variation (JTV) problems (Bresson and Chan, 2008; Huang et al., 2012). The ADM-based methods often have slower convergence rate. Generalized minimization methods can be used, such as the split Bregman method (Goldstein and Osher, 2009), FISTA (Beck and Teboulle, 2009a) and IRN (Rodríguez and Wohlberg, 2009), but they do not consider the special structure of undersampling matrix $\mathbf{A}$ in reconstruction.

In this article, we propose a novel method for the analysis-based sparsity reconstruction (3) in the IRLS framework. It preserves the fast convergence performance of traditional IRLS, which only requires a few reweighted iterations to achieve an accurate solution. An incomplete Cholesky preconditioner is proposed to significantly accelerate the inverse subproblem with preconditioned conjugate gradient (PCG) method. We call our method fast iterative reweighted least squares (FIRLS). This preconditioner is based on the observation that $\mathbf{A}^T\mathbf{A}$ is often diagonally dominant in the image reconstruction problems, such as compressed sensing, image inpainting and CS-MRI. With the same computation complexity, the proposed preconditioner provides more precise results than conventional Jacobi diagonal preconditioner. In addition, the proposed preconditioner can be applied even when $\mathbf{A}$ is an operator, e.g., the Fourier or wavelet transform, which is not feasible for most existing preconditioners of the PCG methods. Besides the efficiency and fast convergence rate, the proposed method can be easily applied to different sparsity patterns, e.g., overlapping group sparsity, TV and JTV. We validate the proposed method on CS-MRI for tree sparsity, joint sparsity, TV and JTV based reconstruction. Extensive experimental results demonstrate that the proposed algorithm outperforms the state-of-the-art methods in terms of both accuracy and computational speed. Primary results in this work has been presented in Chen et al. (2014).

Our major contributions in this article include: (a) We introduce a novel incomplete Cholesky preconditioner, which is an extension of the conventional diagonal preconditioner and can be potentially used in a wide range of applications; (b) For overlapping group sparsity minimization, we propose an efficient preconditioner (16) to solve the large scale inverse problem by utilizing the diagonally dominant structure of $\mathbf{A}^T\mathbf{A}$; (c) We exploit the inexact LU decomposition of the finite difference operators and propose a novel preconditioning (36) for TV regularization; (d) Extensive experiments are conducted to evaluate the performance of our method on different reconstruction tasks. The results demonstrate that our preconditioning strategy achieves significant improvement over state-of-the-arts in terms of both accuracy and efficiency.

2. Related work: IRLS

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

$$\min_{\mathbf{x}} ||\mathbf{x}||_1, \text{ subject to } \mathbf{A}\mathbf{x} = \mathbf{b}. \tag{4}$$

In practice, the ℓ_1 norm is replaced by a reweighted ℓ_2 norm (Chartrand and Yin, 2008):

$$\min_{x} \mathbf{x}^T\mathbf{W}\mathbf{x}, \text{ subject to } \mathbf{A}\mathbf{x} = \mathbf{b}. \tag{5}$$

The diagonal weight matrix $\mathbf{W}$ in the kth iteration is computed from the solution of the current iteration $\mathbf{x}^k$, i.e., $\mathbf{W}_i^k = |\mathbf{x}_i^k|^{-1}$. With current weights $\mathbf{W}^k$, we can derive the closed form solution for $\mathbf{x}^{k+1}$:

$$\mathbf{x}^{k+1} = (\mathbf{W}^k)^{-1}\mathbf{A}^T(\mathbf{A}(\mathbf{W}^k)^{-1}\mathbf{A}^T)^{-1}\mathbf{b}. \tag{6}$$

It has been proven that the IRLS algorithm converges exponentially fast (linear convergence in the terminology of numerical optimization) if the measurement matrix $\mathbf{A}$ satisfies the RIP (Daubechies et al., 2010):

$$||\mathbf{x}^k - \mathbf{x}^*||_1 \le \mu||\mathbf{x}^{k-1} - \mathbf{x}^*||_1 \le \mu^k||\mathbf{x}^0 - \mathbf{x}^*||_1, \tag{7}$$

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 $\mathbf{A}(\mathbf{W}^k)^{-1}\mathbf{A}^T$ takes $\mathcal{O}(M^3)$ if A is a $M \times N$ sampling matrix. Even with the higher convergence rate, traditional IRLS still cannot compete with the fastest first-order algorithms such as FISTA (Beck and Teboulle, 2009b) (some results have been shown in Bach et al., 2011). Most existing IRLS methods like Chartrand and Yin (2008);

Daubechies et al. (2010); Gorodnitsky and Rao (1997) cannot solve the overlapping group sparsity problems, which is a strong limitation. In Bach et al. (2011), the IRLS is extended to solve the overlapping group sparsity problems. However, the inverse subproblem is solved in the closed form, which makes it impossible for large scale problems. As we will show later, the proposed preconditioning method can significantly accelerate the speed of conventional IRLS algorithms.

3. FIRLS for overlapping group sparsity

3.1. An alternative formulation for overlapping group sparsity

We consider the overlapping group sparsity regularization problem first (Yuan and Lin, 2005; Jacob et al., 2009). The mixed $\ell_{2,1}$ norm in (3) may contain overlapping groups. It can be rewritten in the analysis-based sparsity form:

$$\min_{\mathbf{x}}\{F(\mathbf{x}) = \frac{1}{2}||\mathbf{A}\mathbf{x}-\mathbf{b}||_2^2 + \lambda||\mathbf{G}\Phi\mathbf{x}||_{2,1}\}, \tag{8}$$

where Φ denotes an orthogonal sparse basis and is optional. A good choice of Φ for natural images/signals would be the orthogonal wavelet transform, which is used in this article. Our method can also be applied if Φ is not orthogonal but the inverse of preconditioner can be obtained efficiently. For some applications, orthogonal transforms may lead to "blocking" artifacts. $\mathbf{G}$ is a binary matrix for group configuration, which is constructed by the rows of the identity matrix. With different settings of $\mathbf{G}$, this model can handle overlapping group, non-overlaping group and standard sparsity problems. Tree sparsity can also be approximated by this model (Kim and Xing, 2010; Liu and Ye, 2010; Jenatton et al., 2011). Although $\mathbf{G}$ may have large scales, it can be efficiently implemented by a sparse matrix. This kind of indexing matrix has been used in the previous work YALL1 (Deng et al., 2011). With this reformulation, $\Psi = \mathbf{G}\Phi$ and the $\ell_{2,1}$ norm in (8) is now non-overlapping.

We relax the mixed $\ell_{2,1}$ norm based on Young's inequality:

$$\begin{aligned}||\mathbf{G}\Phi\mathbf{x}||_{2,1} &= \sum_{i=1}^{s}||(\mathbf{G}\Phi\mathbf{x})_{g_i}||_2 \le \sum_{i=1}^{s}(||(\mathbf{G}\Phi\mathbf{x})_{g_i}||_2^2+\varepsilon)^{1/2}\\ &\le \sum_{i=1}^{s}\Big[\frac{(||(\mathbf{G}\Phi\mathbf{x}^k)_{g_i}||_2^2+\varepsilon)^{1/2}}{2} + \frac{||(\mathbf{G}\Phi\mathbf{x})_{g_i}||_2^2+\varepsilon}{2(||(\mathbf{G}\Phi\mathbf{x}^k)_{g_i}||_2^2+\varepsilon)^{1/2}}\Big].\end{aligned} \tag{9}$$

where ε is a very small constant to avoid the denominator going to zero. In Chen and Selesnick (2014b), a similar way is used to majorize the original function. Writing it in matrix form and we can majorize $F(\mathbf{x})$ by the majorization minimization (MM) method (Hunter and Lange, 2004):

$$\begin{aligned}Q(\mathbf{x},\mathbf{W}^k) = \frac{1}{2}||\mathbf{A}\mathbf{x}-\mathbf{b}||_2^2 + \frac{\lambda}{2}\mathbf{x}^T\Phi^T\mathbf{G}^T\mathbf{W}^k\mathbf{G}\Phi\mathbf{x}\\ +\frac{\lambda}{2}\sum_{i=1}^{s}\frac{1}{\mathbf{W}_{g_i}^k}+c,\end{aligned} \tag{10}$$

where Φ^T denotes the inverse transform of Φ; $\mathbf{W}^k$ is the group-wise weights; c is a constant with respect to $\mathbf{x}$ containing the remaining terms in (9). Based on our majorization in (10), we are actually minimizing an upper bound $\sum_{i=1}^{s}(||(\mathbf{G}\Phi\mathbf{x})_{g_i}||_2^2+\varepsilon)^{1/2}$ of the original function $||\mathbf{G}\Phi\mathbf{x}||_{2,1}$. Our solution is guaranteed to converge to a local minimum by the MM theory. The weight of ith group $\mathbf{W}_{g_i}^k$ can be obtained by:

$$\mathbf{W}_{g_i}^k = (||(\mathbf{G}\Phi\mathbf{x}^k)_{g_i}||_2^2+\varepsilon)^{-1/2}. \tag{11}$$

Suppose that the signal $\mathbf{x}$ to be recovered is of length N and $\mathbf{G}$ is a N'-by-N matrix, then $\mathbf{W}^k$ is a N'-by-N' diagonal matrix and has the following form:

$$\mathbf{W}^k = \begin{Bmatrix}\mathbf{W}_{g_1}^k & & & & & \\ & \cdots & & & & \\ & & \mathbf{W}_{g_1}^k & & & \\ & & & \cdots & & \\ & & & & \mathbf{W}_{g_s}^k & \\ & & & & & \mathbf{W}_{g_s}^k\end{Bmatrix}, \tag{12}$$

where each group-wise weight $\mathbf{W}_{g_i}^k$ 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 $\mathbf{G}$. With different settings of $\mathbf{G}$, the group-wise weights are directly derived. 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 $\mathbf{W}$ is the same as in IRLS algorithm (Daubechies et al., 2010; Chartrand and Yin, 2008).

The next the problem is to solve:

$$\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}} Q(\mathbf{x},\mathbf{W}^k). \tag{13}$$

Note that $\mathbf{W}_{g_i}^k$ is independent of $\mathbf{x}$ and can be considered as a constant. We iteratively update $\mathbf{W}^k$ with $\mathbf{x}^k$ and solve $\mathbf{x}^{k+1}$ based on current $\mathbf{W}^k$. Our algorithm is also an IRLS typed algorithm.

3.2. Accelerating with PCG

In each iteration, W^k can be easily updated with (12) and (11). To solve (13), a simple way is to let the first order derivative of $Q(\mathbf{x}|\mathbf{x}^k)$ be zero as it is a quadratic convex function:

$$(\mathbf{A}^T\mathbf{A}+\lambda\Phi^T\mathbf{G}^T\mathbf{W}^k\mathbf{G}\Phi)\mathbf{x} - \mathbf{A}^T\mathbf{b} = 0. \tag{14}$$

The way to solve (14) determines the efficiency of the whole algorithm. The exact inverse of the Hessian matrix $\mathbf{S} = \mathbf{A}^T\mathbf{A} + \lambda\Phi^T\mathbf{G}^T\mathbf{W}^k\mathbf{G}\Phi$ takes $\mathcal{O}(N^3)$ time. It is impractical to compute $\mathbf{S}^{-1}$ for many cases especially when the size of $\mathbf{S}$ is large. An alternative way is to approximate the solution of (14) with the classical conjugate gradient (CG) decent method. It is much faster than computing the exact solution. In addition to CG, a better way is the preconditioned conjugate gradient (PCG) method (Saad, 2003). The design of preconditioner is problem-dependent, which should be as close as possible to the Hessian matrix $\mathbf{S}$ and can be inversed efficiently. Therefore, it is not an easy task to design a good preconditioner in general due to this tradeoff. In signal/image reconstruction, such preconditioner has not been found in existing IRLS algorithms (Chartrand and Yin, 2008; Daubechies et al., 2010; Gorodnitsky and Rao, 1997).

We define a new preconditioner for the best approximation in Frobenius norm $||\cdot||_F$:

$$\mathbf{P}^* = \arg\min_{\mathbf{X}\in\mathcal{D}}||\mathbf{S}-\mathbf{X}||_F, \tag{15}$$

where $\mathcal{D}$ denotes a class of incomplete Cholesky preconditioner (Kershaw, 1978), whose inverse can be obtained efficiently with $\mathcal{O}(N)$ time. Instead of discarding all the non-diagonal entries like the conventional diagonal preconditioner, such incomplete Cholesky preconditioner keeps more information so that the algorithm converges much faster.

For the subproblem (14), it is difficult to exactly optimize the incomplete Cholesky preconditioners. Fortunately, we could find a set of possible candidates due to the strong constraint, such as $diag(\mathbf{A}^T\mathbf{A}+\lambda\Phi^T\mathbf{G}^T\mathbf{W}^k\mathbf{G}\Phi)$, $(a\mathbf{I}+\lambda\Phi^T\mathbf{G}^T\mathbf{W}^k\mathbf{G}\Phi)$ or $diag(\mathbf{A}^T\mathbf{A}+\lambda\Phi^T w\Phi)$ etc. Here, a, w denote the mean of diagonal elements of the matrix $\mathbf{A}^T\mathbf{A}$, $\mathbf{G}^T\mathbf{W}^k\mathbf{G}$, respectively, which are scalars; $diag()$ means setting all non-diagonal elements to zero and $\mathbf{I}$ denotes the identity matrix. Note that the $\mathbf{G}^T\mathbf{W}^k\mathbf{G}$ is always diagonal for

any kind of $\mathbf{G}$. In addition, we observe that $\mathbf{A}^T\mathbf{A}$ is often diagonally dominant in the image reconstruction problems. For example, in CS-MRI, $\mathbf{A} = \mathbf{RF}$ where $\mathbf{F}$ denotes the Fourier transform and $\mathbf{R} \in \mathbb{R}^{M\times N}$ $(M < N)$ is a selection matrix containing M rows of the identity matrix. Therefore, $\mathbf{A}^T\mathbf{A} = \mathbf{F}^T\mathbf{R}^T\mathbf{RF}$ is diagonally dominant as $\mathbf{R}^T\mathbf{R}$ is diagonal. For the image inpainting problem, $\mathbf{A}^T\mathbf{A} = \mathbf{R}^T\mathbf{R}$ is diagonal. This structure also holds when $\mathbf{A}$ is a random projection matrix. Based on this property, we proposed a preconditioner by discarding the non-diagonal elements of $\mathbf{A}^T\mathbf{A}$:

$$\mathbf{P} = (a\mathbf{I} + \lambda\Phi^T\mathbf{G}^T\mathbf{W}^k\mathbf{G}\Phi). \tag{16}$$

The preconditioning error in Frobenius norm $||\mathbf{S} - \mathbf{P}||_F$ is very small, due to diagonally dominant structure of $\mathbf{A}^T\mathbf{A}$. As $\mathbf{A}$ is known for many applications, a can be pre-estimated and is fixed for each iteration. Therefore in each iteration, $\mathbf{P}^{-1} = \Phi^T(a\mathbf{I} + \lambda\mathbf{G}^T\mathbf{W}^k\mathbf{G})^{-1}\Phi$ can be efficiently obtained.

Several advantages of the proposed preconditioner can be found when compared with existing ones (Papandreou and Yuille, 2011; Lefkimmiatis et al., 2012). To get the inverse, fast Fourier transforms are used in recent circulant preconditioners for image deblurring (Papandreou and Yuille, 2011; Lefkimmiatis et al., 2012), while our model only requires linear time to obtain $\mathbf{P}^{-1}$. We do not discard all non-diagonal information like the Jacobi preconditioner and therefore the preconditioner is more accurate. Moreover, our model can also handle the case when A or Φ is an operator, while other preconditioners (Papandreou and Yuille, 2011; Lefkimmiatis et al., 2012; Rodríguez and Wohlberg, 2009) cannot because they require the exact values of $\mathbf{S}$.

Our method can be summarized in Algorithm 1. We denote this overlapping group sparsity version as FIRLS_OG. n is the iteration counter for the inner PCG loop. In each inner PCG iteration, the dominated cost is by applying $\mathbf{S}$ and $\mathbf{P}^{-1}$, which is denoted by $\mathcal{O}(\mathcal{C}_S + \mathcal{C}_P)$. When $\mathbf{A}$ and Φ are dense matrices, $\mathcal{O}(\mathcal{C}_S + \mathcal{C}_P) = \mathcal{O}(N^2)$. When $\mathbf{A}$ and Φ are the partial Fourier transform and wavelet transform in CS-MRI (Lustig et al., 2007), it is $\mathcal{O}(N \log N)$. The PCG stopping criterion depends on the precision requirements of different problems. As the inner loop is time consuming to be solved exactly, we set a fixed number of iterations (e.g., 10–30 iterations) for the PCG.

Algorithm 1: FIRLS_OG.

Input: $\mathbf{A}, \mathbf{b}, \mathbf{x}^1, \mathbf{G}, \lambda, k = 1$
while stopping criterion not met **do**
  Update $\mathbf{W}^k$ by (11) (12)
  Update $\mathbf{S} = \mathbf{A}^T\mathbf{A} + \lambda\Phi^T\mathbf{G}^T\mathbf{W}^k\mathbf{G}\Phi$
  Update $\mathbf{P}^{-1} = \Phi^T(a\mathbf{I} + \lambda\mathbf{G}^T\mathbf{W}^k\mathbf{G})^{-1}\Phi$
  Initialize PCG: $\mathbf{r}_0 = \mathbf{A}^T\mathbf{b} - \mathbf{S}\mathbf{x}^k$, $\mathbf{z}_0 = \mathbf{p}_0 = \mathbf{P}^{-1}\mathbf{r}_0$, $n = 0$, $\mathbf{y}_0 = \mathbf{x}^k$
  while PCG stopping criterion not met **do**
    Update $\alpha_n = \frac{\mathbf{r}_n^T\mathbf{z}_n}{\mathbf{p}_n^T\mathbf{S}\mathbf{p}_n}$
    Update $\mathbf{y}_{n+1} = \mathbf{y}_n + \alpha_n\mathbf{p}_n$
    Update $\mathbf{r}_{n+1} = \mathbf{r}_n - \alpha_n S\mathbf{p}_n$
    Update $\mathbf{z}_{n+1} = \mathbf{P}^{-1}\mathbf{r}_{n+1}$
    Update $\beta_n = \frac{\mathbf{z}_{n+1}^T\mathbf{r}_{n+1}}{\mathbf{z}_n^T\mathbf{r}_n}$
    Update $\mathbf{p}_{n+1} = \mathbf{z}_{n+1} + \beta_n\mathbf{p}_n$
    Update $n = n + 1$
  end while
  Update $\mathbf{x}^{k+1} = \mathbf{y}^n$
  Update $k = k + 1$
end while

3.3. Convergence analysis

Our algorithm follows the rules of MM method. It is guaranteed to converge to a local minimum based on the MM theory. We provide a concise proof of its convergence property.

Theorem 1. *The global optimal solution* $\mathbf{x}^*$ *of (10) is the global optimal solution of original problem (8).*

Proof. Suppose $\mathbf{x}_1^*$ is the global optimal solution of (10) and $\mathbf{x}_2^*$ is the global optimal solution of (8). $\mathbf{W}_1^*$ and $\mathbf{W}_2^*$ are weights of $\mathbf{x}_1^*$ and $\mathbf{x}_2^*$ based on (11) and (12). Consider Q as a function corresponds to $\mathbf{x}$ and $\mathbf{W}$. We have:

$$Q(\mathbf{x}_1^*, \mathbf{W}_1^*) \le Q(\mathbf{x}_2^*, \mathbf{W}), \quad \forall\ \mathbf{W}; \tag{17}$$

$$F(\mathbf{x}_2^*) \le F(\mathbf{x}_1^*). \tag{18}$$

Based on the inequality (9), we have

$$F(\mathbf{x}) \le Q(\mathbf{x}, \mathbf{W}^k) \quad \forall\ \mathbf{x}; \tag{19}$$

$$F(\mathbf{x}^k) = Q(\mathbf{x}^k, \mathbf{W}^k). \tag{20}$$

Therefore,

$$F(\mathbf{x}_2^*) \le F(\mathbf{x}_1^*) = Q(\mathbf{x}_1^*, \mathbf{W}_1^*) \le Q(\mathbf{x}_2^*, \mathbf{W}_2^*) = F(\mathbf{x}_2^*), \tag{21}$$

which indicates $F(\mathbf{x}_1^*) = F(\mathbf{x}_2^*)$. □

Theorem 2. *$F(\mathbf{x}^k)$ is monotonically decreased by Algorithm 1, i.e.,$F(\mathbf{x}^{k+1}) \le F(\mathbf{x}^k)$. In particular, we have* $\lim_{k\to\infty}(F(\mathbf{x}^k) - F(\mathbf{x}^{k+1})) = 0$.

Proof. With the property (19), we have

$$F(\mathbf{x}^{k+1}) \le Q(\mathbf{x}^{k+1}, \mathbf{W}^k). \tag{22}$$

To balance the cost and accuracy when solving (14), we apply the PCG method to decrease $Q(\mathbf{x}, \mathbf{W}^k)$ and efficiently obtain the solution $\mathbf{x}^{k+1}$. Because $\mathbf{x}^k$ is the initial guess for $\mathbf{x}^{k+1}$, based on the monotonically decreasing property of PCG (Kaasschieter, 1988), we have:

$$Q(\mathbf{x}^{k+1}, \mathbf{W}^k) \le Q(\mathbf{x}^k, \mathbf{W}^k). \tag{23}$$

And we finally get:

$$F(\mathbf{x}^{k+1}) \le Q(\mathbf{x}^{k+1}, \mathbf{W}^k) \le Q(\mathbf{x}^k, \mathbf{W}^k) = F(\mathbf{x}^k). \tag{24}$$

$F(\mathbf{x})$ is convex and bounded. Due to the monotone convergence theorem, we have:

$$\lim_{k\to\infty}(F(\mathbf{x}^k) - F(\mathbf{x}^{k+1})) = 0. \tag{25}$$

□

4. FIRLS for TV

We have presented an efficient algorithm for overlapping group sparsity under an orthogonal sparse basis Φ. In image reconstruction problems, another widely used sparsity regularizer is TV. Due to the non-orthogonality of the TV semi-norm, the FIRLS_OG algorithm cannot be applied to solve the TV problem. In this section, we will present an efficient algorithm for TV based image reconstruction. For brevity and clarity, we first present the algorithm for single channel image reconstruction and then extended it to multi-channel reconstruction (Bresson and Chan, 2008).

4.1. An alternative formulation for TV

TV minimization exploits the sparsity of the image in the gradient domain. For brevity, we assume the image is n by n with $n \times n = N$. Let $\mathbf{D}_1$, $\mathbf{D}_2$ be two N-by-N first-order finite difference matrices in vertical and horizontal directions.

$$\mathbf{D}_1 = \begin{bmatrix} 1 & & & & & & & \\ -1 & 1 & & & & & & \\ & -1 & 1 & & & & & \\ & & \dots & \dots & & & & \\ & & & \dots & \dots & & & \\ & & & & -1 & 1 & & \\ & & & & & -1 & 1 & \end{bmatrix}, \tag{26}$$

$$\mathbf{D}_2 = \begin{bmatrix} 1 & & & & & & & \\ \dots & \dots & & & & & & \\ -1 & \dots & 1 & & & & & \\ & -1 & \dots & 1 & & & & \\ & & \dots & & \dots & & & \\ & & & -1 & \dots & 1 & & \\ & & & & -1 & \dots & 1 & \end{bmatrix}. \tag{27}$$

The main diagonal elements of $\mathbf{D}_1$ and $\mathbf{D}_2$ are all ones. The first diagonal elements below the main diagonal are all minus ones in $\mathbf{D}_1$, while in $\mathbf{D}_2$ n-th diagonal elements below the main diagonal are all minus ones. With these notations, the ℓ_1 and isotropic TV based image reconstruction can be reformulated as:

$$\min_{\mathbf{x}}\{\frac{1}{2}||\mathbf{Ax}-\mathbf{b}||_2^2 + \lambda||\mathbf{D}_1\mathbf{x}||_1 + \lambda||\mathbf{D}_2\mathbf{x}||_1\}, \tag{28}$$

$$\min_{\mathbf{x}}\{F(\mathbf{x}) = \frac{1}{2}||\mathbf{Ax}-\mathbf{b}||_2^2 + \lambda||[\mathbf{D}_1\mathbf{x}, \mathbf{D}_2\mathbf{x}]||_{2,1}\}. \tag{29}$$

Here, the $\ell_{2,1}$ norm is the summation of the ℓ_2 norm of each row, which is a special case of (2). Here and later, we denote $[,]$ as the concatenating of the matrices horizontally. To avoid repetition, all the following derivations only consider isotropic TV function (29). ℓ_1-based TV function can be derived in the same way.

Considering the Young's inequality, we majorize (29) by the MM method (Hunter and Lange, 2004):

$$\begin{aligned} Q(\mathbf{x}, \mathbf{W}^k) = &\frac{1}{2}||\mathbf{Ax}-\mathbf{b}||_2^2 + \frac{\lambda}{2}\mathbf{x}^T\mathbf{D}_1^T\mathbf{W}^k\mathbf{D}_1\mathbf{x} \\ &+\frac{\lambda}{2}\mathbf{x}^T\mathbf{D}_2^T W^k\mathbf{D}_2\mathbf{x} + \frac{\lambda}{2}\mathrm{Tr}((\mathbf{W}^k)^{-1}), \end{aligned} \tag{30}$$

where Tr() denotes the trace. $\mathbf{W}^k$ is a diagonal weight matrix in the kth iteration:

$$\mathbf{W}^k = \begin{Bmatrix} \mathbf{W}_1^k & & & \\ & \mathbf{W}_2^k & & \\ & & \dots & \\ & & & \mathbf{W}_N^k \end{Bmatrix}, \tag{31}$$

with

$$\mathbf{W}_i^k = 1/\sqrt{(\nabla_1\mathbf{x}_i^k)^2 + (\nabla_2\mathbf{x}_i^k)^2 + \varepsilon}, \ i = 1, 2, \dots, N, \tag{32}$$

where ∇_1 and ∇_2 denote the gradients along the x and y directions. When $\mathbf{D}_1 = \mathbf{D}_2 = \mathbf{I}$, it is identical to the ℓ_1 norm minimization as in the conventional IRLS methods (Chartrand and Yin, 2008; Daubechies et al., 2010; Gorodnitsky and Rao, 1997).

4.2. Accelerating with PCG and incomplete LU decomposition

After the weight matrix is updated by (31) and (32), the problem is to update $\mathbf{x}$. With the same rule as in the overlapping group sparsity regularization, we have

$$(\mathbf{A}^T\mathbf{A} + \lambda\mathbf{D}_1^T W^k\mathbf{D}_1 + \lambda\mathbf{D}_2^T W^k\mathbf{D}_2)\mathbf{x} = \mathbf{A}^T\mathbf{b}. \tag{33}$$

Similar to (14), the Hessian matrix here is in large scale. We have discussed that the Hessian matrix is not dense but follows some special structure in image reconstruction. A good solver should consider such special structure of the problem. In TV based image deblurring problems, by observing that A has a circulant structure (under periodic boundary conditions), many efficient algorithms have been proposed to accelerate the minimization (Lefkimmiatis et al., 2012; Yang et al., 2009; Chan et al., 2011). However, these algorithms cannot be applied to the TV reconstruction problems.

Based on the diagonally dominant prior information in image reconstruction, we can obtained an accurate preconditioner like (16).

$$\mathbf{P}' = a\mathbf{I} + \lambda\mathbf{D}_1^T\mathbf{W}^k\mathbf{D}_1 + \lambda\mathbf{D}_2^T\mathbf{W}^k\mathbf{D}_2 \tag{34}$$

However, the inverse cannot be efficiently obtained for this preconditioner, due to the non-orthogonality of $\mathbf{D}_1$ and $\mathbf{D}_2$.

Fortunately, $\mathbf{P}'$ has a sparse structure due to the forms of $\mathbf{D}_1$ and $\mathbf{D}_2$. We observe that preconditioner $\mathbf{P}'$ in (34) is a penta-diagonal matrix (Saad, 2003) and is symmetric, which has the following form:

$$\mathbf{P}' = \begin{bmatrix} a_1 & b_1 & & c_1 & & & \\ b_1 & a_2 & b_2 & & c_2 & & \\ & b_2 & a_3 & \dots & & \dots & \\ & & \dots & \dots & & & c_{N-n} \\ c_1 & & & & & & \\ & c_2 & & & & \dots & \\ & & \dots & & \dots & \dots & b_{N-1} \\ & & & c_{N-n} & & b_{N-1} & a_N \end{bmatrix}. \tag{35}$$

Such penta-diagonal matrix has incomplete LU decomposition. Therefore, we propose a new preconditioner for TV minimization:

$$\mathbf{P} = \mathbf{LU} \approx a\mathbf{I} + \lambda\mathbf{D}_1^T\mathbf{W}^k\mathbf{D}_1 + \lambda\mathbf{D}_2^T\mathbf{W}^k\mathbf{D}_2, \tag{36}$$

where

$$\mathbf{L} = \begin{bmatrix} 1 & & & & & & \\ \frac{b_1}{a_1} & 1 & & & & & \\ & \frac{b_2}{a_2} & 1 & & & & \\ & & \dots & \dots & & & \\ \frac{c_1}{a_1} & & & \frac{b_n}{a_n} & 1 & & \\ & \dots & & & \dots & \dots & \\ & & \frac{c_{N-n}}{a_{N-n}} & \dots & & \frac{b_{N-1}}{a_{N-1}} & 1 \end{bmatrix}, \tag{37}$$

$$\mathbf{U} = \begin{bmatrix} a_1 & b_1 & \dots & c_1 & & & \\ & a_2 & b_2 & \dots & c_2 & & \\ & & \dots & \dots & & \dots & \\ & & & a_{N-n} & b_{N-n} & \dots & c_{N-n} \\ & & & & \dots & \dots & \\ & & & & & a_N & b_{N-1} \\ & & & & & & a_N \end{bmatrix}. \tag{38}$$

We assume λ is small for the TV regularization. The decomposition is very accurate therefore as $\mathbf{P}$ is diagonally dominated with $a_i \gg b_i^2, a_i \gg c_i^2$ and $a_i \gg b_i c_i$ for all i. To the best of our knowledge, this incomplete LU decomposition is first proposed for TV minimization. Due to the special structure of $\mathbf{P}$, the incomplete LU decomposition only takes $\mathcal{O}(N)$ time. Therefore, the total time to obtain $\mathbf{P}^{-1} \approx \mathbf{U}^{-1}\mathbf{L}^{-1}$ is $\mathcal{O}(N)$. We can conclude the proposed method for TV reconstruction in Algorithm 2. The inner PCG loop is similar to that in Algorithm 1, which is omitted to avoid repetition.

4.3. Extension to JTV

In many multiple measurement vector problems (MMV), the image with multiple channels has the joint sparsity property. In

Algorithm 2: FIRLS_TV.

Input: $\mathbf{A}$, $\mathbf{b}$, $\mathbf{x}^1$, λ, $k = 1$
while stopping criterion not met **do**
  Update $\mathbf{W}^k$ by (**??**) and (**??**)
  Update $\mathbf{S} = \mathbf{A}^T\mathbf{A} + \lambda\mathbf{D}_1^T\mathbf{W}^k\mathbf{D}_1 + \lambda\mathbf{D}_2^T\mathbf{W}^k\mathbf{D}_2$
  Update $\mathbf{P} = \mathbf{LU} \approx a\mathbf{I} + \lambda\mathbf{D}_1^T\mathbf{W}^k\mathbf{D}_1 + \lambda\mathbf{D}_2^T\mathbf{W}^k\mathbf{D}_2$, $\mathbf{P}^{-1} = \mathbf{U}^{-1}\mathbf{L}^{-1}$
  while PCG stopping criterion not met **do**
    Update $\mathbf{x}^{k+1}$ by PCG for $\mathbf{Sx} = \mathbf{A}^T\mathbf{b}$ with preconditioner $\mathbf{P}$
  end while
  Update $k = k + 1$
end while

these cases, the TV can be extended to joint total variation (JTV) (Bresson and Chan, 2008; Huang et al., 2012):

$$\min_{\mathbf{x}} \left\{ \frac{1}{2}\sum_{t=1}^{T} ||\mathbf{A}_t\mathbf{X}_t - \mathbf{b}_t||_2^2 + \lambda|||[\mathbf{D}_1\mathbf{X}, \mathbf{D}_2\mathbf{X}]||_{2,1} \right\}, \tag{39}$$

where $\mathbf{X} \in \mathbb{R}^{N\times T}$ is a T-channel image with $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_T]$; $\mathbf{A}_t$ is the undersampling matrix for channel t and $\mathbf{b}_t$ is the measurement vector for channel t. Similar as (30), we have:

$$Q(\mathbf{X}, \mathbf{W}^k) = \frac{1}{2}\sum_{t=1}^{T} ||\mathbf{A}_t\mathbf{X}_t - \mathbf{b}_t||_2^2 + \frac{\lambda}{2}\left[\sum_{t=1}^{T} \mathbf{X}_t^T\mathbf{D}_1^T\mathbf{W}^k\mathbf{D}_1\mathbf{X}_t + \sum_{t=1}^{T} \mathbf{X}_t^T\mathbf{D}_2^T\mathbf{W}^k\mathbf{D}_2\mathbf{X}_t + \mathrm{Tr}((\mathbf{W}^k)^{-1}) \right], \tag{40}$$

where

$$\mathbf{W}_i^k = 1/\sqrt{\sum_{t=1}^{T} (\nabla_1\mathbf{X}_{t,i}^k)^2 + (\nabla_2\mathbf{X}_{t,i}^k)^2 + \varepsilon}, \forall\ i, \tag{41}$$

and

$$\mathbf{W}^k = \begin{Bmatrix} \mathbf{W}_1^k & & & \\ & \mathbf{W}_2^k & & \\ & & \cdots & \\ & & & \mathbf{W}_N^k \end{Bmatrix}. \tag{42}$$

It indicates that the weights for $\mathbf{X}_1$ to $\mathbf{X}_T$ are the same. Similar, $\mathbf{X}_t$ can be updated by solving:

$$(\mathbf{A}_t^T\mathbf{A}_t + \lambda\mathbf{D}_1^T\mathbf{W}^k\mathbf{D}_1 + \lambda\mathbf{D}_2^T\mathbf{W}^k\mathbf{D}_2)\mathbf{x} = \mathbf{A}_t^T\mathbf{b}_t. \tag{43}$$

It also can be solved efficiently by the PCG method with the proposed preconditioner. Again to avoid repetition, the algorithm for JTV based reconstruction is not listed.

5. Experiments

5.1. Experiment setup

The experiments are conducted using MATLAB on a desktop computer with 3.4 GHz Intel core i7 3770 CPU. We validate different versions of our method on wavelet tree sparsity based reconstruction, wavelet joint sparsity reconstruction, TV and JTV reconstruction. To avoid confusion, we denote the tree sparsity version as FIRLS_OG and non-overlapping joint sparsity version FIRLS_MT. The version for standard ℓ_1 norm minimization is denoted by FIRLS_L1. FIRLS_TV and FIRLS_JTV denotes the TV and JTV reconstruction, respectively. We also compare the corresponding IRLS algorithms without our preconditioning (i.e., the subproblem is solved by CG), which are named IRLS_L1, IRLS_TV, etc.

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 considered to be better if it can achieve higher reconstruction accuracy with less computational time.

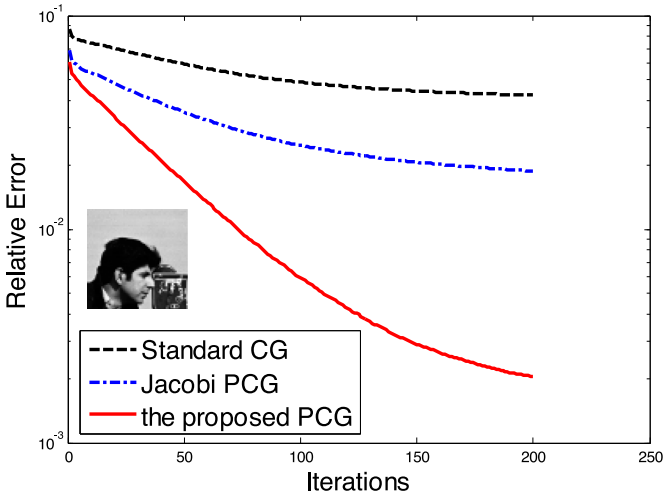


Fig. 1. Convergence rate comparison among standard CG, Jacobi PCG and the proposed PCG for ℓ_1 norm minimization.

5.2. The accuracy of the proposed preconditioner

One of our contributions is the proposed incomplete Cholesky preconditioner for sparse recovery. First, we conduct an experiment to validate its effectiveness with the orthogonal wavelet basis. 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 $\mathbf{S}^{-1}$ 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. Fig. 1 demonstrates the performance of the proposed PCG compared with Jacobi PCG and standard CG for the problem (14). 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 removes all the non-diagonal elements which makes the preconditioner less precise.

To validate the effectiveness of the proposed preconditioner in TV reconstruction, we take experiments on the Shepp-Logan phantom image with 64×64 pixels. The Shepp-Logan phantom image is very smooth and is an ideal example to validate TV reconstruction. The relative errors of CG, PCG Jacobi and the proposed method are shown in Fig. 2. It shows that only 20 iterations of PCG with the proposed preconditioner can outperform conventioanal CG with 200 iterations. Jacobi PCG requires approximately 2 times of iterations to reach the same accuracy as our method, because it discards all non-diagonal information directly and makes the preconditioning less precise. Comparing with the results in Fig. 1, our preconditioner seems less powerful on TV reconstruction. This is expected as we further decompose the preconditioner into two triangle matrices $\mathbf{L}$ and $\mathbf{U}$, which introduces minor approximation error. However, it still converges much faster than the existing Jacobi PCG. These experiments demonstrate that the inner loop subproblem in our method is solved efficiently with the proposed preconditioner.

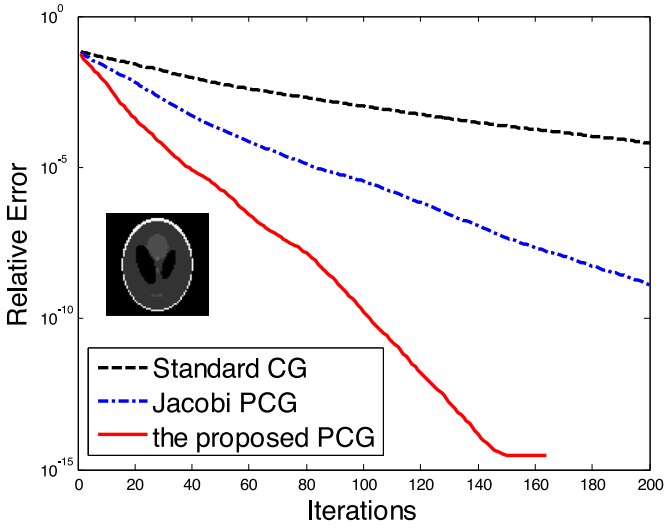


Fig. 2. Convergence rate comparison among standard CG, Jacobi PCG and the proposed PCG for TV minimization.

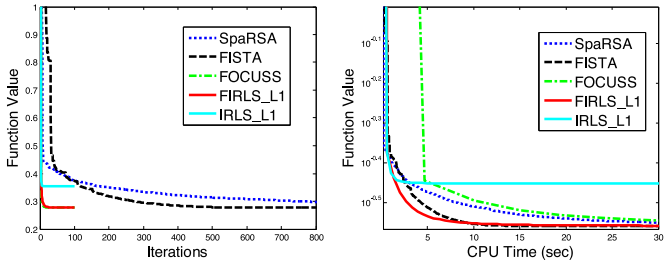


Fig. 3. Convergence rate comparison among FOCUSS, FISTA and SpaRSA for ℓ_1 norm minimization.

Table 1
Computational cost comparison between FOCUSS (Gorodnitsky and Rao, 1997) and the proposed method.

	FOCUSS			FIRLS_L1		
Time (seconds)	47.5	127.2	485.7	10.3	24.4	49.4
Function value	0.2810	0.2774	0.2767	0.2809	0.2774	0.2767

5.3. Convergence rate and computational complexity

One of the properties of the proposed FIRLS is its fast convergence rate, i.e., only a small number of iterations can achieve high reconstruction accuracy. In addition, each iteration has low 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 (Wright et al., 2009), FISTA (Beck and Teboulle, 2009b) and IRLS algorithm FOCUSS (Gorodnitsky and Rao, 1997), with $\mathcal{O}(1/k)$, $\mathcal{O}(1/k^2)$ and exponential convergence rates, respectively. In addition, the IRLS without our preconditioning is compared.

The test data is a random 1D signal of length 4000, with 10% elements being non-zeros. The number of measurements are 800. Fig. 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 100 iterations in terms of the object function value (1). 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 slightly different function values 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 not known how to solve the overlapping group sparsity problem with FOCUSS.

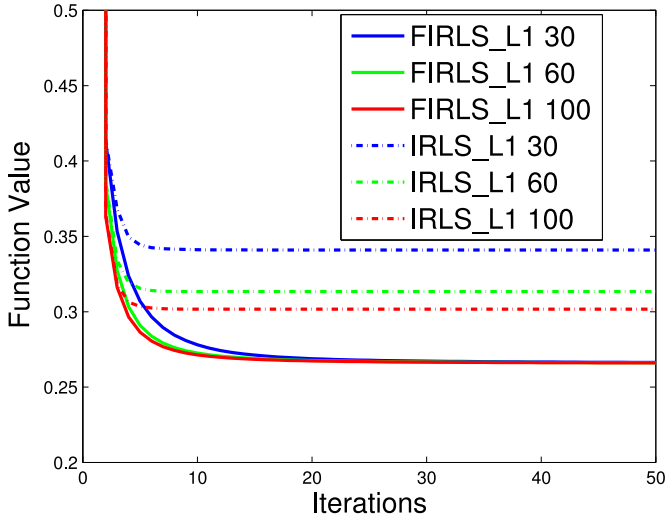


Fig. 4. Comparisons between FIRLS_L1 and IRLS_L1 with different inner PCG/CG iterations.

In Fig. 3, one may notice that IRLS_L1 also converges very fast, but the converged function value is a higher than the others. To investigate this observation, we compare IRLS_L1 and FIRLS_L1 with different inner PCG/CG iterations. The results on the same data are shown in Fig. 4. When the subproblem is not precisely solved, IRLS_L1 may not converge to the same result of FIRLS_L1. With more number of CG iterations, the IRLS_L1 can converge to better solutions. In contrast, FIRLS_L1 can always converge to similar solutions, which correspond to much lower function values than those by IRLS_L1. For 30 inner iterations, the computational costs of IRLS_L1 and FIRLS_L1 are 28.1 s and 28.5 s. With a similar computational cost, our method can produce more accurate results.

5.4. Approximation of the diagonal dominant structure

We do not have strong assumption on the type of the $\mathbf{A}$ matrix, but only require $\mathbf{A}^T\mathbf{A}$ has a diagonally dominant structure. If $\mathbf{A}^T\mathbf{A}$ is less diagonally dominant, we will get more approximation errors using the proposed preconditioning. An experiment is conducted to evaluate the behaviors of our preconditioning on different $\mathbf{A}$ matrices.

We create three types of random projection $\mathbf{A}$ matrices. The number of rows are N, $0.5N$ and $0.1N$, and $\mathbf{A}$ is less and less diagonally dominant. The test data is created using the same method as in Section 5.3. We compare the convergence speeds of our method for solving the subproblem (14). The PCG algorithm is terminated when a tolerance 10^{-6} of the residual is reached. The average numbers of iterations to reach the tolerance are 87, 152 and 226, respectively. This experiment demonstrates that our method converges much faster when $\mathbf{A}^T\mathbf{A}$ is more diagonally dominant.

6. Application: compressive sensing MRI

Compressive sensing MRI (CS-MRI) (Lustig et al., 2007) is one of the most successful applications of compressive sensing and sparsity regularization. There are various sparsity patterns on MR images. Therefore, we validate the performance of different versions

of our method on CS-MRI as a concrete reconstruction instance. 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 can significantly accelerate MRI acquisition. In CS-MRI, $\mathbf{A} = \mathbf{RF}$ is an undersampled Fourier operator, where $\mathbf{F}$ is the Fourier transform and $R \in \mathbb{R}^{M\times N}$ is a selection matrix containing M rows of the identity matrix. Therefore, $\mathbf{A}^T\mathbf{A} = \mathbf{F}^T\mathbf{R}^T\mathbf{RF}$ is diagonally dominant as $\mathbf{R}^T\mathbf{R}$ is diagonal. Based on (16), a is identical to the sampling ratio (a fixed scalar). The examples in this article are single coil CS-MRI. For parallel MRI reconstruction, one can apply our method in the CS-SENSE framework (Liang et al., 2009; Otazo et al., 2010; Ramani and Fessler, 2011).

Following previous works, Signal-to-Noise Ratio (SNR) is used as metric for result evaluation:

$$SNR = 10\log_{10}(V_s/V_n), \tag{44}$$

where V_n is the Mean Square Error between the original image $\mathbf{x}_0$ and the reconstructed $\mathbf{x}$; $V_s = var(\mathbf{x}_0)$ denotes the variance of the values in $\mathbf{x}_0$.

6.1. CS-MRI

6.1.1. CS-MRI with wavelet tree sparsity

The MR images are often piecewise smooth, which are widely assumed to be sparse under the wavelet basis or in the gradient domain (Lustig et al., 2007; Ma et al., 2008; Yang et al., 2010; Huang et al., 2011a). 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, approximated by the overlapping group sparsity (Chen and Huang, 2012; 2013). Tree-structured CS-MRI method (Chen and Huang, 2012; 2013) has been shown to be superior to standard CS-MRI methods (Lustig et al., 2007; Ma et al., 2008; Huang et al., 2011a). Therefore, we compare our algorithm with two latest and fastest tree-based algorithms, turbo AMP (Som and Schniter, 2012) and WaTMRI (Chen and Huang, 2012). In addition, overlapping group sparsity solvers SLEP (Liu et al., 2009; Yuan et al., 2013) and YALL1 (Deng et al., 2011) are also compared. The total number of iterations is 100 except that turbo AMP only runs 10 iterations due to its higher time complexity. Followed by the previous works (Ma et al., 2008; Huang et al., 2011a; Chen and Huang, 2012), four MR images with the same size 256×256 are used for testing, which are shown in Fig. 5. Using a similar sampling strategy, we 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.

The convergence speeds of different algorithms on the Brain image are illustrated in Fig. 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 FIRLS. 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 (Chen and Huang, 2012). The IRLS without preconditioning has the similar performance as SLEP. 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 with about 43 s. 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. Our results are consistently more accurate.

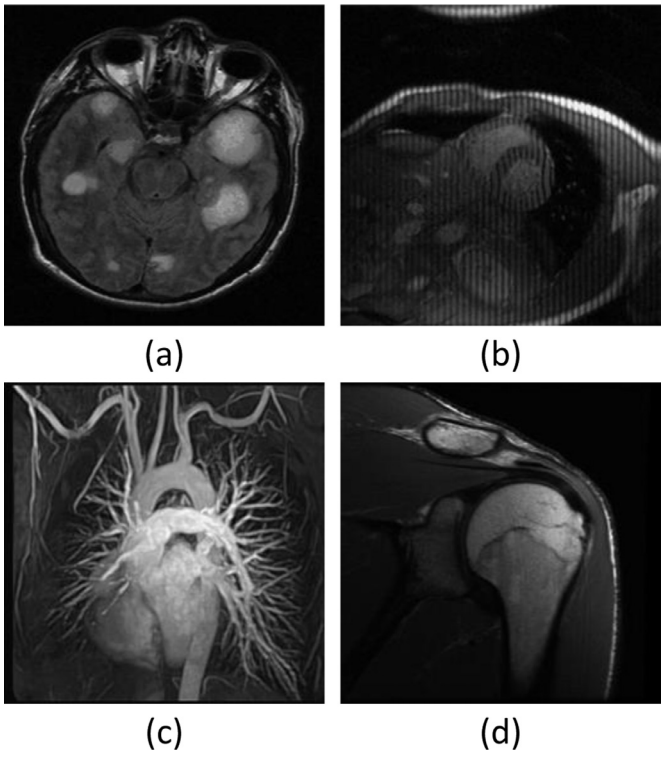


Fig. 5. The original images: (a) Brain; (b) Cardiac; (c) Chest; (d) Shoulder.

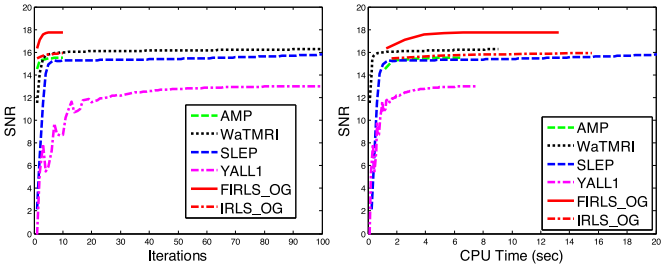


Fig. 6. Convergence speed comparison on the Brain image with 25% sampling. Left: SNR vs outer loop iterations. Right: SNR vs CPU time.

Table 2
Average SNR (dB) comparisons on the four MR images with wavelet tree sparsity.

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

A visual comparison on the Brain image is shown in Fig. 7, with a 25% sampling ratio. Visible artifacts can be found on the results by YALL1 (Deng et al., 2011). The image reconstructed by AMP (Som and Schniter, 2012) tends to be blurry when compared with the original. The image recovered by SLEP (Liu et al., 2009) is noisy. Our method and WaTMRI (Chen and Huang, 2012) produce the most accurate results in terms of SNR. Note that WaTMRI has more parameters required to be tuned due to its variable splitting strategy. Besides SNR, we also compare the mean structural similarity (Wang et al., 2004) (MSSIM) of different images, which mimics the human visual system. The MSSIM for the images recovered by AMP (Som and Schniter, 2012), WaTMRI (Chen and Huang, 2012), SLEP (Liu et al., 2009), YALL1 (Deng et al., 2011) and the proposed

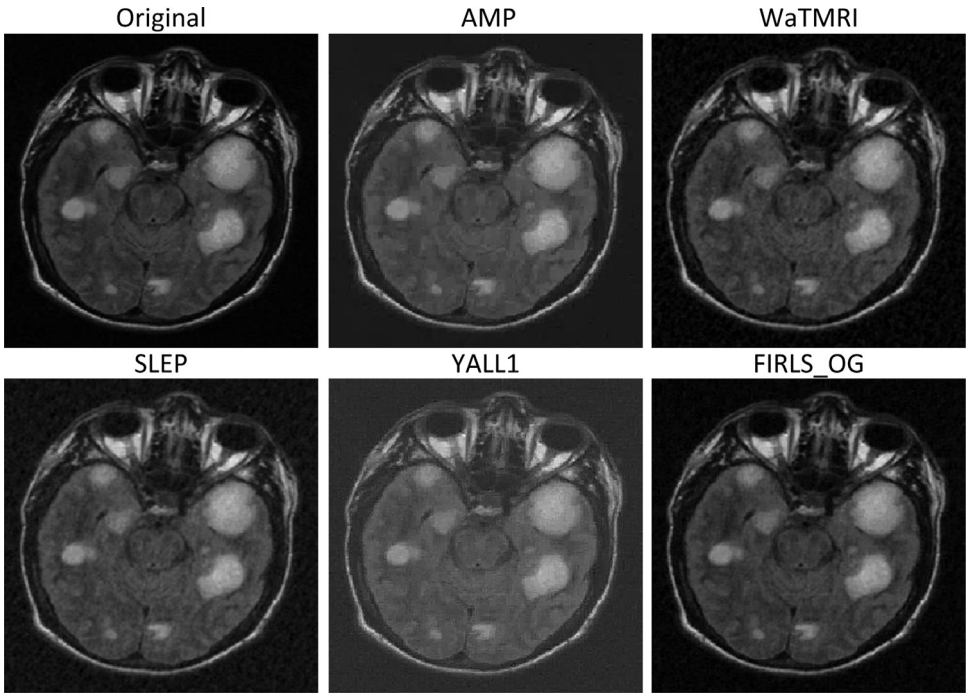


Fig. 7. Visual comparison on the Brain image with 25% sampling. The SNRs of AMP (Som and Schniter, 2012), WaTMRI (Chen and Huang, 2012), SLEP (Liu et al., 2009), YALL1 (Deng et al., 2011) and the proposed method are 15.91, 16.72, 16.49, 12.86 and 18.39, respectively.

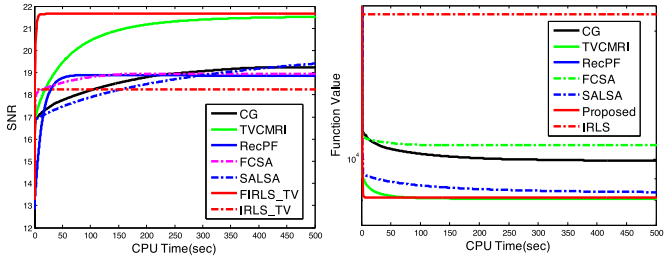


Fig. 8. Convergence rate comparison for TV minimization on the Chest image with 25% sampling. Left: SNR vs CPU time. Right: Function Value vs CPU time. The converged function value of RecPF is approximately 2.6×10^5, which is out of range of this display.

Table 3
Quantitative comparison of convergence speed on the Chest image by TV regularization with 25% sampling.

	Iterations	CPU time (sec)	SNR (dB)
CG	3181	397.8	19.23
TVCMRI	21392	495.1	21.54
RecPF	7974	163.4	18.86
FCSA	1971	39.6	18.96
SALSA	9646	882.4	20.13
FIRLS_TV	**29**	**6.9**	**21.65**

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 is consistent with the observation in terms of SNR.

6.1.2. CS-MRI by TV reconstruction

TV is another popular regularizer for MRI reconstruction and the images recovered by TV tend to be less noisy (Lustig et al., 2007). For TV based reconstruction, we compare our method with classical method CG (Lustig et al., 2007), TVCMRI (Ma et al., 2008), RecPF (Yang et al., 2010), FCSA (Huang et al., 2011a) and SALSA (Afonso et al., 2010).

The convergence speeds of different algorithms on the Chest image are presented in Fig. 8. It is worthwhile to note that no closed form solutions exist for the subproblems of these algorithms. The subproblems in these algorithms are often solved in an approximate way. From the figure, the final results of our method and TVCMRI are almost the same while the others converge to different results. The right pane demonstrates that our method and TVCMRI can minimize the object function (29) more successfully. IRLS_TV meets the same problem as in Fig. 3, i.e., converge to a poor solution with a low SNR with a high function value. We further found that only TVCMRI has analyzed their global convergence (in Section 2.3 of Ma et al., 2008), while the accuracy of all the other methods (Lustig et al., 2007; Yang et al., 2010; Huang et al., 2011a; Afonso et al., 2010) has not been discussed in details. For the four MR images, the average SNRs of CG (Lustig et al., 2007), TVCMRI (Ma et al., 2008), RecPF (Yang et al., 2010), FCSA (Huang et al., 2011a), SALSA (Afonso et al., 2010) and the proposed algorithm are 19.45, 21.78, 21.70, 21.53 21.95 and 23.07, respectively.

We terminate each algorithm after a fixed toleration is reached, e.g., 10^{-3} of the relative solution change. The final SNR and convergence speeds of different methods are listed in Table 3. To produce a similar result of TVCMRI, our method only requires about its 1/70 computational time. These convergence performances are not surprising.FIRLS converges with a similar rate as the conventional IRLS algorithm FOCUSS (as shown in Fig. 3) and require the fewest iterations. FCSA is a FISTA based algorithm, which has $\mathcal{O}(1/k^2)$ convergence rate. It converges with the second fewest iterations. For the rest algorithms, there is no known convergence rate better than $\mathcal{O}(1/k)$.

Due to the relatively slower convergence speed, we note that previous methods (Lustig et al., 2007; Ma et al., 2008; Yang et al., 2010; Huang et al., 2011a) often terminate after a fixed number of iterations (e.g., 200) in practice. This is because the exactly convergence is time consuming that may not be feasible for clinic applications. Following by this scheme, we run TVCMRI 200 iter-

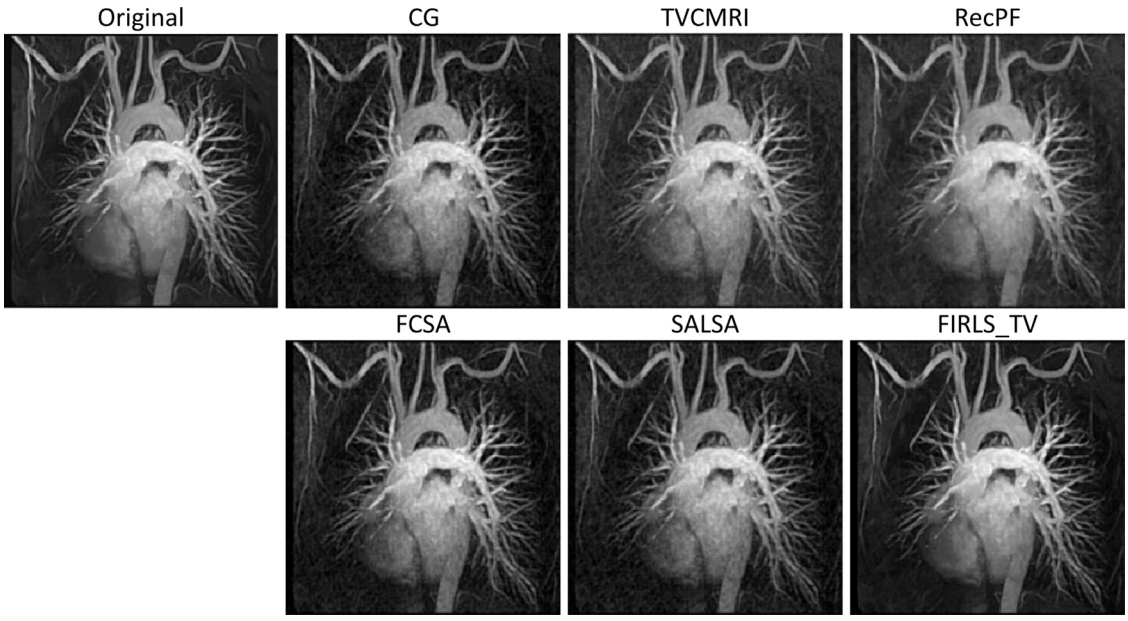


Fig. 9. Chest MR image reconstruction from 25% sampling. All methods terminate after 4 s. The SNRs for CG, TVCMRI, RecPF, FCSA, SALSA and the proposed are 17.13, 17.32, 16.18, 18.28, 16.96 and 21.63, respectively.

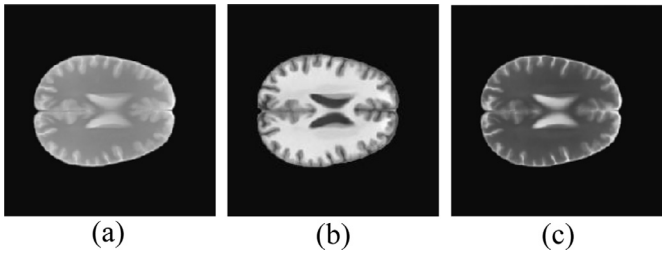


Fig. 10. The original images for multi-contrast MRI.

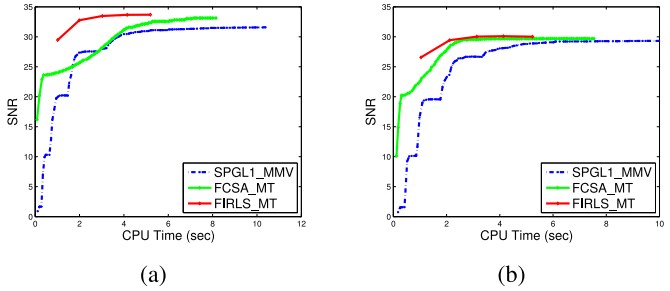


Fig. 11. (a) Performance comparison for multi-contrast MRI with 25% sampling. The average time costs of SPGL1_MMV, FCSA_MT, and the proposed method are 10.38 s, 8.15 s, 5.19 s. Their average SNRs are 31.58, 33.12 and 33.69. (b) Performance comparison for multi-contrast MRI with 20% sampling. Their average time costs are 9.98 s, 7.54 s, 5.23 s. Their average SNRs are 29.31, 29.69 and 30.01.

ations. All the other algorithms terminate after the same running time of TVCMRI (i.e., around 4 s). The reconstruction results on the Chest MR image are shown in Fig. 9. A close look shows that our method preserves highest organ-to-background contrast. Such results are expected if we take a review on Fig. 8. Similar results can be obtained on the Brain, Cardiac and Artery images.

6.2. Multi-contrast MRI

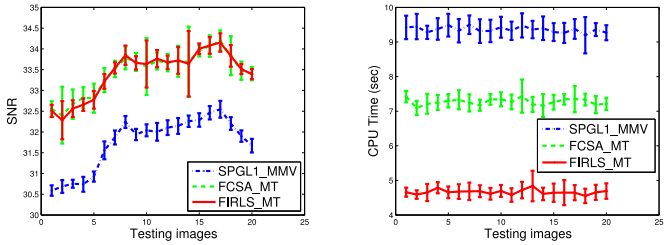
6.2.1. Multi-contrast MRI with wavelet joint sparsity

To assist clinic diagnose, multiple MR images with different contrasts 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 of the wavelet coefficients and JTV across different contrasts have been used in recent multi-contrast reconstruction methods (Majumdar and Ward, 2011; Huang et al., 2012).

Here, the multi-contrast MR images are extracted from the SRI24 Multi-Channel Brain Atlas Data (Rohlfing et al., 2010). An example of the test images is shown in Fig. 10. We compare our method with the fastest multi-contrast MRI methods (Majumdar and Ward, 2011; Huang et al., 2012), which use the algorithms SPGL1_MMV (Berg and Friedlander, 2008) and FCSA to solve the corresponding sub-problems, respectively. FCSA_MT and FIRLS_MT denote the algorithm in Huang et al. (2012) and the proposed method in this setting.

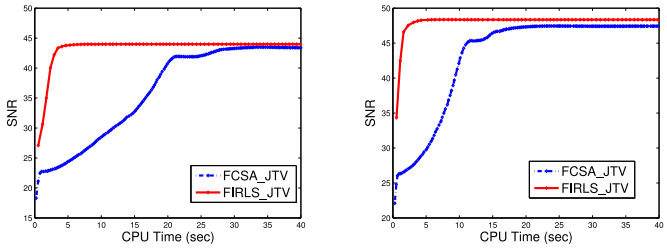
Fig. 11 shows the performance comparisons among SPGL1_MMV (Berg and Friedlander, 2008), FCSA_MT (Huang et al., 2012) and FIRLS_MT on the example images shown in Fig. 10. Each algorithm runs 100 iterations in total. After convergence, three algorithms achieve similar accuracy for 20% sampling and SPGL1 is only slightly worse than others for 25% sampling. From the curves, our method always ourperforms SPGL1_MMV and FCSA_MT, i.e., higher accuracy for the same reconstruction time.

To quantitatively compare the convergence speed of these three methods, we conduct an experiment on 20 set images (i.e., total 60 images) that are from SRI24. Different from the tree-based CS-MRI, each algorithm for non-overlapping group sparsity converges much faster. To reduce randomness, all algorithms run 100 times and the reconstruction results are shown in Fig. 12. With 25% sampling, the accuracy of our method is almost the same as FCSA_MT, and always better than SPGL1. In the process to achieve the convergence, our method is consistently faster than the other two algorithms. These results demonstrate the efficiency of proposed method.



(a) (b)

Fig. 12. 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 s, 7.2 s, 4.6 s, respectively.



(a) (b)

Fig. 13. Multi-contrast MRI with JTV reconstruction. (a) The performance comparison with 25% sampling. (b) The performance comparison with 30% sampling.

6.2.2. *Multi-contrast MRI with JTV*

Finally, we reconstruct multi-contrast MR images with JTV. Here, we compare our method FIRLS_JTV with FCSA_JTV (Huang et al., 2012).

Fig. 13 shows the performance comparison on the example images (in Fig. 10) from 25% and 30% sampling, without setting stoping criteria. After convergence, the accuracy of our results are slightly higher than those of FCSA_JTV. Also, it is clearly that FIRLS_JTV requires much less time to converge in both cases. We then let each algorithm terminate with the 10^{-3} tolerance. FCSA_JTV cost 35.6 s and 19.7 s to converge for the two sampling cases, while the proposed FIRLS_JTV only requires 6.7 s and 4.6 s for the two cases, respectively.

6.3. *Discussion*

The first and second experiments validate the fast convergence speed of our method due to the proposed preconditioner. The superior performance of the proposed preconditioner attributes to the utilizing of the special structure of the matrix $\mathbf{A}^T\mathbf{A}$, which is often diagonally dominant in reconstruction problems (e.g., $\mathbf{A}$ is random projection or partial Fourier transform). We do not have strong assumption about the properties matrix $\mathbf{A}^T\mathbf{A}$ except its diagonally dominant structure. As shown in Section 5.4, our method converges faster if the $\mathbf{A}^T\mathbf{A}$ matrix is more diagonally dominant.

The advantages of our method over the state-of-the-arts are validated on practical application CS-MRI with four sparsity patterns: overlapping groups with tree sparsity, non-overlapping groups with joint sparsity, TV and JTV. Although results on these 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, TV and JTV reconstruction. In this work, we focus on fast minimization of the given functions, which is our major contribution.

7. Conclusion

We have proposed a novel method for analysis-based sparsity reconstruction, which includes structured sparsity with an orthogonal basis and TV. It is of the IRLS type and preserves the *fast convergence rate*. The subproblem in our scheme is accelerated by the PCG method with a new incomplete Cholesky 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

Afonso, M.V., Bioucas-Dias, J.M., Figueiredo, M.A., 2010. Fast image recovery using variable splitting and constrained optimization. IEEE Trans. Image Process. 19 (9), 2345–2356.
Bach, F., Jenatton, R., Mairal, J., Obozinski, G., 2011. Optimization with sparsity-inducing penalties. Found. Trends Mach. Learn. 4 (1), 1–106.
Baraniuk, R.G., Cevher, V., Duarte, M.F., Hegde, C., 2010. Model-based compressive sensing. IEEE Trans. Inf. Theory 56, 1982–2001.
Beck, A., Teboulle, M., 2009. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. IEEE Trans. Image Process. 18 (11), 2419–2434.
Beck, A., Teboulle, M., 2009. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci. 2 (1), 183–202.
Berg, E., Friedlander, M., 2008. Probing the Pareto frontier for basis pursuit solutions. SIAM J. Scient. Comput. 31, 890–912.
Bresson, X., Chan, T.F., 2008. Fast dual minimization of the vectorial total variation norm and applications to color image processing. Inverse Probl. Imaging 2 (4), 455–484.
Candes, E., Romberg, J., Tao, T., 2006. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans. Inf. Theory 52 (2), 489–509.
Chan, S.H., Khoshabeh, R., Gibson, K.B., Gill, P.E., Nguyen, T.Q., 2011. An augmented Lagrangian method for total variation video restoration. IEEE Trans. Image Process. 20 (11), 3097–3111.
Chartrand, R., Yin, W., 2008. Iteratively reweighted algorithms for compressive sensing. In: Proc. Int. Conf. Accoust., Speech, Signal Process.
Chen, C., Huang, J., 2012. Compressive sensing MRI with wavelet tree sparsity. In: Proc. Advances in Neural Information Processing Systems.
Chen, C., Huang, J., 2013. The benefit of tree sparsity in accelerated MRI. Med. Image Anal.
Chen, C., Huang, J., He, L., Li, H., 2014. Preconditioning for accelerated iteratively reweighted least squares in structured sparsity reconstruction. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 2713–2720.
Chen, P.-Y., Selesnick, I.W., 2014. Group-sparse signal denoising: non-convex regularization, convex optimization. IEEE Trans. Signal Process. 62 (13), 3464–3478.
Chen, P.-Y., Selesnick, I.W., 2014. Translation-invariant shrinkage/thresholding of group sparse signals. Signal Process. 94, 476–489.
Cleju, N., Jafari, M.G., Plumbley, M.D., 2012. Choosing analysis or synthesis recovery for sparse reconstruction. In: Proceedings of the 20th European Signal Processing Conference. IEEE, pp. 869–873.
Daubechies, I., DeVore, R., Fornasier, M., Gunturk, S., 2010. Iteratively reweighted least squares minimization for sparse recovery. Commun. Pure Appl. Math. 63 (1), 1–38.
Deng, W., Wotao, Y., Zhang, Y., 2011. Group Sparse Optimization by Alternating Direction Method. Technical Report. Rice University.
Donoho, D., 2006. Compressed sensing. IEEE Trans. on Inf. Theory 52 (4), 1289–1306.
Goldstein, T., Osher, S., 2009. The split Bregman method for l1-regularized problems. SIAM J. Imaging Sci. 2 (2), 323–343.
Gorodnitsky, I.F., Rao, B.D., 1997. Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm. IEEE Trans. Signal Process. 45 (3), 600–616.
Huang, J., Chen, C., Axel, L., 2012. Fast multi-contrast MRI reconstruction. In: Proc. Int. Conf. Med. Image Computing Computer-Assisted Interventions, pp. 281–288.
Huang, J., Zhang, S., Metaxas, D., 2011. Efficient MR image reconstruction for compressed MR imaging. Med. Image Anal. 15 (5), 670–679.
Huang, J., Zhang, T., Metaxas, D., 2011. Learning with structured sparsity. J. Mach. Learn. Res. 12, 3371–3412.
Hunter, D., Lange, K., 2004. A tutorial on MM algorithms. Am. Stat. 58, 30–37.
Jacob, L., Obozinski, G., Vert, J.-P., 2009. Group lasso with overlap and graph lasso. In: Proc. Intïl Conf. Machine Learning, pp. 433–440.
Jenatton, R., Mairal, J., Obozinski, G., Bach, F., 2011. Proximal methods for hierarchical sparse coding. J. Mach. Learn. Res. 12, 2297–2334.
Kaasschieter, E.F., 1988. Preconditioned conjugate gradients for solving singular systems. J. Comput. Appl. Math. 24 (1–2), 265–275.

Kershaw, D.S., 1978. The incomplete Cholesky-conjugate gradient method for the iterative solution of systems of linear equations. J. Comput. Phys. 26 (1), 43–65.
Kim, S., Xing, E.P., 2010. Tree-guided group lasso for multi-task regression with structured sparsity. In: Proc. Intïl Conf. Machine Learning, pp. 543–550.
Lefkimmiatis, S., Bourquard, A., Unser, M., 2012. Hessian-based norm regularization for image restoration with biomedical applications. IEEE Trans. Image Process. 21 (3), 983–995.
Liang, D., Liu, B., Wang, J., Ying, L., 2009. Accelerating SENSE using compressed sensing. Magn. Reson. Med. 62 (6), 1574–1584.
Liu, J., Ji, S., Ye, J., 2009. SLEP: Sparse Learning with Efficient Projections. Arizona State University.
Liu, J., Ye, J., 2010. Moreau-Yosida regularization for grouped tree structure learning. In: Proc. Advances in Neural Information Processing Systems, 23, pp. 1459–1467.
Lustig, M., Donoho, D., Pauly, J., 2007. Sparse MRI: the application of compressed sensing for rapid MR imaging. Magn. Reson. Med. 58 (6), 1182–1195.
Ma, S., Yin, W., Zhang, Y., Chakraborty, A., 2008. An efficient algorithm for compressed MR imaging using total variation and wavelets. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 1–8.
Majumdar, A., Ward, R., 2011. Joint reconstruction of multiecho MR images using correlated sparsity. Magn. Reson. Imaging 29, 899–906.
Mosci, S., Villa, S., Verri, A., Rosasco, L., 2010. A primal-dual algorithm for group sparse regularization with overlapping groups. In: Proc. Advances in Neural Information Processing Systems.
Otazo, R., Kim, D., Axel, L., Sodickson, D.K., 2010. Combination of compressed sensing and parallel imaging for highly accelerated first-pass cardiac perfusion MRI. Magn. Reson. Med. 64 (3), 767–776.
Papandreou, G., Yuille, A., 2011. Efficient variational inference in large-scale Bayesian compressed sensing. In: IEEE Int',l Conf. Computer Vision Workshop.
Ramani, S., Fessler, J.A., 2011. Parallel mr image reconstruction using augmented lagrangian methods. IEEE Trans. Med. Imaging 30 (3), 694–706.
Rodríguez, P., Wohlberg, B., 2009. Efficient minimization method for a generalized total variation functional. IEEE Trans. Image Process. 18 (2), 322–332.
Rohlfing, T., NM, N.Z., Sullivan, E., Pfefferbaum, A., 2010. The SRI24 multichannel atlas of normal adult human brain structure. Hum. Brain Mapp. 31, 798–819.
Saad, Y., 2003. Iterative Methods for Sparse Linear Systems. SIAM.
Som, S., Schniter, P., 2012. Compressive imaging using approximate message passing and a Markov-tree prior. IEEE Trans. Signal Process. 60, 3439–3448.
Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004. Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. 13 (4), 600–612.
Wright, S., Nowak, R., Figueiredo, M., 2009. Sparse reconstruction by separable approximation. IEEE Trans. Signal Process. 57, 2479–2493.
Xiao, Y., Yang, J., 2010. A fast algorithm for total variation image reconstruction from random projections. arXiv preprint arXiv:1001.1774.
Yang, J., Yin, W., Zhang, Y., Wang, Y., 2009. A fast algorithm for edge-preserving variational multichannel image restoration. SIAM J. Imaging Sci. 2 (2), 569–592.
Yang, J., Zhang, Y., Yin, W., 2010. A fast alternating direction method for TVL1-L2 signal reconstruction from partial Fourier data. IEEE J. Sel. Top. Signal Process. 4 (2), 288–297.
Yuan, L., Liu, J., Ye, J., 2013. Efficient methods for overlapping group lasso. IEEE Trans. Pattern Anal. Mach.Intell. 35 (9), 2104–2116.
Yuan, M., Lin, Y., 2005. Model selection and estimation in regression with grouped variables. J. R. Stat. Soc. Ser. B Stat. Methodol. 68 (1), 49–67.