

Advanced sparsity techniques in magnetic resonance imaging

7

J. Huang, Y. Li

University of Texas at Arlington, Arlington, TX, United States

CHAPTER OUTLINE

7.1 Introduction	183
7.2 Standard Sparsity in CS-MRI	186
7.2.1 Model and Algorithm	187
7.2.2 Evaluation	191
7.2.3 Summary	194
7.3 Group Sparsity in Multicontrast MRI	195
7.3.1 Model and Algorithm	196
7.3.2 Evaluation	202
7.3.3 Summary	211
7.4 Tree Sparsity in Accelerated MRI	211
7.4.1 Model and Algorithm	211
7.4.2 Evaluation	215
7.4.3 Summary	220
7.5 Forest Sparsity in Multichannel CS-MRI.....	220
7.5.1 Model and Algorithm	221
7.5.2 Evaluation	227
7.5.3 Summary	232
7.6 Conclusion	232
References.....	232

7.1 INTRODUCTION

Magnetic resonance imaging (MRI) has been widely used in medical diagnosis because of its noninvasive manner and excellent depiction of soft tissue changes. Recent developments in compressive sensing (CS) theory (Candès et al., 2006c; Donoho, 2006) show that it is possible to accurately reconstruct the magnetic

resonance (MR) images from highly undersampled K-space data and therefore significantly reduce the scanning duration.

Suppose x is an MR image and R is a partial Fourier transform, the sampling measurement b of x in K-space is defined as $b = Ax$. The compressed MR image reconstruction problem is to reconstruct x given the measurement b and the sampling matrix A . Sometimes the data is not sparse but compressible under some base Φ such as wavelet, and the corresponding problem is $A\Phi^{-1}\theta = b$, where θ denotes the set of wavelet coefficients. Although the problem is underdetermined, the data can be perfectly reconstructed if the sampling matrix satisfies the restricted isometry property (RIP) (Candès, 2006) and the number of measurements is larger than $\mathcal{O}(k+k \log(N/k))$ for k -sparse data. The term “ k -sparse data” means there are at most k nonzero components in the data. (Candès and Romberg, 2007; Candès et al., 2006b).

To solve the underdetermined problem, we may find the sparsest solution via ℓ_0 -norm regularization. However, because the problem is NP-hard (Natarajan, 1995) and impractical for most applications, ℓ_1 -norm regularization methods such as the lasso (Tibshirani, 1996) and basis pursuit (BP) (Chen et al., 1998) are first used to pursue the sparse solution. It has been proved that the ℓ_1 -norm regularization can exactly recover the sparse data for the CS inverse problem under mild conditions (Candès et al., 2006a; Donoho and Elad, 2002). Therefore a lot of efficient algorithms have been proposed for standard sparse recovery. Generally speaking, those algorithms can be classified into three groups: greedy algorithms (Needell and Tropp, 2009; Tropp, 2004), convex programming (Beck and Teboulle, 2009b; Figueiredo et al., 2007; Koh et al., 2007), and probability-based methods (Donoho et al., 2009; Ji et al., 2008).

Beyond standard sparsity, the nonzero components of x often tend to be in some structures. This leads to the concept of *structured sparsity* or model-based compressed sensing (Baraniuk et al., 2010; Huang et al., 2011c; Huang, 2011). In contrast to standard sparsity that only relies on the sparseness of the data, structured sparsity models exploit both the nonzero values and the corresponding locations. For example, in the multiple measurement vector (MMV) problem, the data consists of several vectors that share the same support. The set of indices corresponding to the nonzero entries is often called the support. This is called *joint sparsity* that is common in cognitive radio networks (Meng et al., 2011), direction-of-arrival estimation in radar (Krim and Viberg, 1996), multichannel compressed sensing (Baron et al., 2005; Majumdar and Ward, 2010), remote sensing (Chen et al., 2014b), and medical imaging (Bilgic et al., 2011; Huang et al., 2012). If the data $X \in \mathbb{R}^{TN \times 1}$ consists of T k -sparse vectors, the measurement bound could be substantially reduced to $\mathcal{O}(Tk + k \log(N/q))$ instead of $\mathcal{O}(Tk + Tk \log(N/q))$ for standard sparsity (Baraniuk et al., 2010; Huang et al., 2011c, 2009, 2010).

A common way to implement joint sparsity in convex programming is to replace the ℓ_1 -norm with $\ell_{2,1}$ -norm, which is the summation of ℓ_2 -norms of the correlated entries (Bach, 2008; Yuan and Lin, 2005). $\ell_{2,1}$ -norm for joint sparsity has been used in many convex solvers and algorithms (Cotter et al., 2005; Van Den Berg and Friedlander, 2008; Huang et al., 2012; Deng et al., 2011). In Bayesian sparse

learning or approximate message passing (Ji et al., 2009; Wipf and Rao, 2007; Ziniel and Schniter, 2011), data from all channels contribute to the estimation of parameters or hidden variables in the sparse prior model.

Another common structure would be the hierarchical tree structure, which has already been successfully utilized in image compression (Manduca, 1996), compressed imaging (Chen and Huang, 2012b; He and Carin, 2009; Som and Schniter, 2012; Rao et al., 2011), and machine learning (Kim and Xing, 2012). Most nature signals/images are approximately tree-sparse under the wavelet basis. A typical relationship with *tree sparsity* is that, if a node on the tree is nonzero, all of its ancestors leading to the root should be nonzero. For multichannel data $X = [x_1; x_2; \dots; x_T] \in \mathbb{R}^{NT \times 1}$, In this chapter, the notation [:] denotes concatenating the data vertically. $\mathcal{O}(Tk + T \log(N/k))$ measurements are required if each channel x_t is tree-sparse.

Due to the overlapping and intricate structure of tree sparsity, it is much harder to implement. For greedy algorithms, structured orthogonal matching pursuit (StructOMP) (Huang et al., 2011c) and tree-based orthogonal matching pursuit (TOMP) (La and Do, 2006) have been developed for exploiting tree structure where the coefficients are updated by only searching the subtree blocks instead of all subspace. In statistical models (He and Carin, 2009; Som and Schniter, 2012), hierarchical inference is used to model the tree structure, where the value of a node is not independent but relies on the distribution or state of its parent. In convex programming (Chen and Huang, 2014a; Rao et al., 2011), due to the trade-off between the recovery accuracy and computational complexity, this is often approximated as overlapping group sparsity (Jacob et al., 2009), where each node and its parent are assigned into one group.

Although both joint sparsity and tree sparsity have been widely studied, unfortunately there is no work that studies the benefit of their combinations so far. Actually, in many multichannel compressed sensing or MMV problems, the data has joint sparsity across different channels and each channel itself is tree-sparse. Note that this differs from C-HiLasso (Sprechmann et al., 2011), where sparsity is assumed inside the groups. No method has fully exploited both priors and no theory guarantees the performance. In practical applications, researchers and engineers have to choose either joint sparsity algorithms by giving up their intra tree-sparse prior, or tree sparsity algorithms by ignoring their intercorrelations.

A new sparsity model called *forest sparsity* is proposed to bridge this gap (Chen et al., 2014a). It is a natural extension of existing structured sparsity models by assuming that the data can be represented by a forest of mutually connected trees. Based on compressed sensing theory, it is proved that for a forest of T k -sparse trees, only $\mathcal{O}(Tk + \log(N/k))$ measurements are required for successful recovery with high probability (Chen et al., 2014a). That is much less than the bounds of joint sparsity $\mathcal{O}(Tk + k \log(N/k))$ and tree sparsity $\mathcal{O}(Tk + T \log(N/k))$ on the same data. The theory is further extended to the case of MMV problems, which is ignored in existing structured sparsity theories (Baraniuk et al., 2010; Huang et al., 2011c; Huang, 2011). In this chapter we will also show an efficient algorithm to optimize

the forest sparsity model. The algorithm is applied to medical imaging applications such as multicontrast MRI and parallel MRI (pMRI).

The rest of this chapter is organized as follows. In [Section 7.2](#) we discuss the application of the standard sparsity to MR image reconstruction. Then, [Section 7.3](#) introduces the benefits of the group sparsity in multicontrast MRI reconstruction, while [Section 7.4](#) discusses the benefit of tree sparsity in accelerated MRI. An extension of the tree sparsity named forest sparsity is discussed for multichannel CS-MRI in [Section 7.5](#). We conclude this chapter in [Section 7.6](#).

7.2 STANDARD SPARSITY IN CS-MRI

In this section we present an efficient algorithm for MR image reconstruction. The algorithm minimizes a linear combination of three terms corresponding to a least square data fitting, total variation (TV), and L_1 -norm regularization. This has been shown to be very powerful for the MR image reconstruction.

Suppose x is an MR image and R is a partial Fourier transform, the sampling measurement b of x in K-space is defined as $b = Rx$. The compressed MR image reconstruction problem is to reconstruct x given the measurement b and the sampling matrix R . Motivated by the compressive sensing theory, [Lustig et al. \(2007\)](#) proposed their pioneering work for MR image reconstruction. Their method can effectively reconstruct MR images with only 20% sampling. Improved results were obtained by having both a wavelet transform and a discrete gradient in the objective, which is formulated as follows:

$$\hat{x} = \arg \min_x \left\{ \frac{1}{2} \|Rx - b\|^2 + \alpha \|x\|_{TV} + \beta \|\Phi, x\|_1 \right\} \quad (7.1)$$

where α and β are two positive parameters, b is the undersampled measurements of K-space data, R is a partial Fourier transform, and Φ is a wavelet transform. It is based on the fact that the piecewise smooth MR images of organs can be sparsely represented by the wavelet basis and should have small total variations. The TV was defined discretely as $\|x\|_{TV} = \sum_i \sum_j \sqrt{(\nabla_1 x_{ij})^2 + (\nabla_2 x_{ij})^2}$, where ∇_1 and ∇_2 denote the forward finite difference operators on the first and second coordinates respectively. Since both L_1 - and TV-norm regularization terms are nonsmooth, this problem is very difficult to solve. The conjugate gradient (CG) ([Lustig et al., 2007](#)) and partial differential equation ([He et al., 2006](#)) methods have been used to do this. However, they are very slow and impractical for real MR images. Computation became the bottleneck that prevented this good model (7.1) from being used in practical MR image reconstruction. Therefore the key problem in compressed MR image reconstruction is to develop efficient algorithms to solve problem (7.1) with nearly optimal reconstruction accuracy.

7.2.1 MODEL AND ALGORITHM

7.2.1.1 Related acceleration algorithm

We first briefly review the fast iterative shrinkage-thresholding algorithm (FISTA) from (Beck and Teboulle, 2009b), since our methods are motivated by this. FISTA considers minimizing the following problem:

$$\min\{F(x) \equiv f(x) + g(x), x \in \mathbf{R}^p, \} \quad (7.2)$$

where f is a smooth convex function with Lipschitz constant L_f , and g is a convex function which may be nonsmooth.

ϵ -Optimal solution: Suppose x^* is an optimal solution to (7.2). $x \in \mathbf{R}^p$ is called an ϵ -optimal solution to (7.2) if $F(x) - F(x^*) \leq \epsilon$ holds.

Gradient: $\nabla f(x)$ denotes the gradient of the function f at the point x .

The proximal map: Given a continuous convex function $g(x)$ and any scalar $\rho > 0$, the proximal map associated with function g is defined as follows (Beck and Teboulle, 2009a,b):

$$\text{prox}_\rho(g)(x) := \arg \min_u \left\{ g(u) + \frac{1}{2\rho} \|u - x\|^2 \right\}, \quad (7.3)$$

Algorithm 7.1 outlines the FISTA. It can obtain an ϵ -optimal solution in $\mathcal{O}(1/\sqrt{\epsilon})$ iterations.

Theorem 7.1 (Theorem 4.1 in Beck and Teboulle (2009b)). *Suppose $\{x^k\}$ and $\{r^k\}$ are iteratively obtained by the FISTA, then we have*

$$F(x^k) - F(x^*) \leq \frac{2L_f \|x^0 - x^*\|^2}{(k+1)^2}, \forall x^* \in X_*$$

ALGORITHM 7.1 FISTA Beck and Teboulle (2009b)

```

series Input:  $\rho = 1/L_f$ ,  $r^1 = x^0$ ,  $t^1 = 1$ 
for  $k = 1$  series to  $K$  do
     $x_g^k = r^k - \rho \nabla f(r^k)$ 
     $x_{\rho}^k = \text{prox}_{\rho}(g)(x_g^k)$ 
     $t^{k+1} = \frac{1 + \sqrt{1 + 4(t^k)^2}}{2}$ 
     $r^{k+1} = x^k + \frac{t^k - 1}{t^{k+1}}(x^k - x^{k-1})$ 
end for

```

ALGORITHM 7.2 CSD

```

series Input:  $\rho = 1/L$ ,  $\alpha$ ,  $\beta$ ,  $z_1^0 = z_2^0 = x_g$ 
for  $j = 1$  series to  $J$  do
     $x_1 = \text{prox}_{\rho}(2\alpha \|x\|_{TV})(z_1^{j-1})$ 
     $x_2 = \text{prox}_{\rho}(2\beta \|\Phi x\|_1)(z_2^{j-1})$ 

```

```

 $x^j = (x_1 + x_2)/2$ 
 $z_1^j = z_1^{j-1} + x^j - x_1$ 
 $z_2^j = z_2^{j-1} + x^j - x_2$ 
end for

```

The efficiency of the FISTA depends on being able to quickly solve its second step $x^k = prox_\rho(g)(x_g)$. For simpler regularization problems, this is possible, that is, the FISTA can rapidly solve the l_1 regularization problem with cost $\mathcal{O}(p \log(p))$ (Beck and Teboulle, 2009b) (where p is the dimension of x), since the second step $x^k = prox_\rho(\beta\|\Phi x\|_1)(x_g)$ has a closed-form solution; it can also quickly solve the TV regularization problem, since the step $x^k = prox_\rho(\alpha\|x\|_{TV})(x_g)$ can be computed with cost $\mathcal{O}(p)$ (Beck and Teboulle, 2009a). However, the FISTA cannot efficiently solve the composite l_1 and TV regularization problem (7.1), since no efficient algorithm exists to solve the step

$$x^k = prox_\rho(\alpha\|x\|_{TV} + \beta\|\Phi x\|_1)(x_g). \quad (7.4)$$

To solve problem (7.1), the key problem is thus to develop an efficient algorithm to solve problem (7.4). In the following section, we will show that a scheme based on composite splitting techniques can be used to do this.

7.2.1.2 CSA and FCSA

From the above introduction, we know that, if we can develop a fast algorithm to solve problem (7.4), the MR image reconstruction problem can then be efficiently solved by the FISTA, which obtains an ϵ -optimal solution in $\mathcal{O}(1/\sqrt{\epsilon})$ iterations. Actually, problem (7.4) can be considered as a denoising problem:

$$x^k = \arg \min_x \left\{ \frac{1}{2} \|x - x_g\|^2 + \rho\alpha\|x\|_{TV} + \rho\beta\|\Phi x\|_1 \right\}. \quad (7.5)$$

We use composite splitting techniques to solve this problem: (1) splitting variable x into two variables $\{x_i\}_{i=1,2}$; (2) performing operator splitting over each of $\{x_i\}_{i=1,2}$ independently; and (3) obtaining the solution x by linear combination of $\{x_i\}_{i=1,2}$. We call this the composite splitting denoising (CSD) method, which is outlined in [Algorithm 7.2](#). Its validity is guaranteed by the following theorem:

Theorem 7.2. Suppose $\{x^j\}$ is the sequence generated by the CSD. Then, x^j will converge to $prox_\rho(\alpha\|x\|_{TV} + \beta\|\Phi x\|_1)(x_g)$, which means that we have $x^j \rightarrow prox_\rho(\alpha\|x\|_{TV} + \beta\|\Phi x\|_1)(x_g)$.

Sketch Proof of Theorem 7.2

Consider a more general formulation:

$$\min_{x \in \mathbf{R}^p} F(x) \equiv f(x) + \sum_{i=1}^m g_i(B_i x), \quad (7.6)$$

where f is the loss function and $\{g_i\}_{i=1,\dots,m}$ are the prior models, both of which are convex functions; $\{B_i\}_{i=1,\dots,m}$ are orthogonal matrices.

Proposition 7.1 (Theorem 3.4 in [Combettes and Pesquet \(2008\)](#)). *Let \mathcal{H} be a real Hilbert space, and let $g = \sum_{i=1}^m g_i$ in $\Gamma_0(\mathcal{H})$ such that $\text{dom}g_i \cap \text{dom}g_j \neq \emptyset$. Let $r \in \mathcal{H}$ and $\{x_j\}$ be generated by [Algorithm 7.3](#). Then, x_j will converge to $\text{prox}(g)(r)$.*

The detailed proof for this proposition can be found in [Combettes and Pesquet \(2008\)](#) and [Combettes \(2009\)](#).

ALGORITHM 7.3 ALGORITHM 3.1 IN [Combettes and Pesquet \(2008\)](#)

```

series Input:  $\rho, \{z_i\}_{i=1,\dots,m} = r, \{w_i\}_{i=1,\dots,m} =$ 
 $1/m,$ 
for  $j = 1$  series to  $J$  do
  for  $i = 1$  series to  $m$  do
     $p_{i,j} = \text{prox}_\rho(g_i/w_i)(z_j)$ 
  end for
   $p_j = \sum_{i=1}^m w_i p_{i,j}$ 
   $\lambda_j \in [0, 2]$ 
  for  $i = 1$  series to  $m$  do
     $z_{i,j+1} = z_{i,j+1} + \lambda_j(2p_j - x_j - p_{i,j})$ 
  end for
   $x_{j+1} = x_j + \lambda_j(p_j - x_j)$ 
end for

```

Suppose that $y_i = B_i x$, $s_i = B_i^T r$, and $h_i(y_i) = m\rho g_i(B_i x)$. Because the operators $\{B_i\}_{i=1,\dots,m}$ are orthogonal, we can easily obtain that $\frac{1}{2\rho} \|x - r\|^2 = \sum_{i=1}^m \frac{1}{2m\rho} \|y_i - s_i\|^2$. The above problem is transferred to:

$$\hat{y}_i = \arg \min_{y_i} \sum_{i=1}^m \left[\frac{1}{2} \|y_i - s_i\|^2 + h_i(y_i) \right], \quad x = B_i^T y_i, \quad i = 1, \dots, m. \quad (7.7)$$

Obviously, this problem can be solved by [Algorithm 7.3](#). According to [Proposition 7.1](#), we know that x will converge to $\text{prox}(g)(r)$. Assuming $g_1(x) = \alpha \|x\|_{\text{TV}}$, $g_2(x) = \beta \|x\|_1$, $m = 2$, $w_1 = w_2 = 1/2$ and $\lambda_j = 1$, we obtain the proposed CSD algorithm. x will converge to $\text{prox}(g)(r)$, where $g = g_1 + g_2 = \alpha \|x\|_{\text{TV}} + \beta \|\Phi x\|_1$.

End of Proof

Combining the CSD with FISTA, a new algorithm, the fast composite splitting algorithm (FCSA), is proposed for MR image reconstruction problem (7.1). In practice, we found that a small iteration number J in the CSD is enough for the FCSA to obtain good reconstruction results. In particular, it is set as 1 in our algorithm.

Numerous experimental results in the next section will show that it is good enough for real MR image reconstruction.

[Algorithm 7.5](#) outlines the proposed FCSA. In this algorithm, if we remove the acceleration step by setting $t^{k+1} \equiv 1$ in each iteration, we will obtain the composite splitting algorithm (CSA), which is outlined in [Algorithm 7.4](#). A key feature of the FCSA is its fast convergence performance borrowed from the FISTA. From [Theorem 7.1](#), we know that the FISTA can obtain an ϵ -optimal solution in $\mathcal{O}(1/\sqrt{\epsilon})$ iterations.

Another key feature of the FCSA is that the cost of each iteration is $\mathcal{O}(p \log(p))$, as confirmed by the following observations. Steps 4, 6, and 7 only involve adding vectors or scalars, thus cost only $\mathcal{O}(p)$ or $\mathcal{O}(1)$. In step 1, $\nabla f(r^k) = R^T(Rr^k - b)$ since $f(r^k) = \frac{1}{2}\|Rr^k - b\|^2$ in this case. Thus this step only costs $\mathcal{O}(p \log(p))$. As introduced above, the step $x^k = prox_\rho(2\alpha\|x\|_{TV})(x_g)$ can be computed quickly with cost $\mathcal{O}(p)$ ([Beck and Teboulle, 2009a](#)); the step $x^k = prox_\rho(2\beta\|\Phi x\|_1)(x_g)$ has a closed-form solution and can be computed with cost $\mathcal{O}(p \log(p))$. In the step $x^k = project(x^k, [l, u])$, the function $x = project(x, [l, u])$ is defined as: (1) $x = x$ if $l \leq x \leq u$; (2) $x = l$ if $x < u$; and (3) $x = u$ if $x > u$, where $[l, u]$ is the range of x . For example, in the case of MR image reconstruction, we can let $l = 0$ and $u = 255$ for 8-bit gray MR images. This step costs $\mathcal{O}(p)$. Thus the total cost of each iteration in the FCSA is $\mathcal{O}(p \log(p))$.

With these two key features, the FCSA efficiently solves the MR image reconstruction problem [\(7.1\)](#) and obtains better reconstruction results in terms of both the reconstruction accuracy and computation complexity. The experimental results in the next section demonstrate its superior performance compared with all previous methods for compressed MR image reconstruction.

ALGORITHM 7.4 CSA

```

series Input:  $\rho = 1/L$ ,  $\alpha$ ,  $\beta$ ,  $t^1 = 1$   $x^0 = r^1$ 
for  $k = 1$  series to  $K$  do
     $x_g = r^k - \rho \nabla f(r^k)$ 
     $x_1 = prox_\rho(2\alpha\|x\|_{TV})(x_g)$ 
     $x_2 = prox_\rho(2\beta\|\Phi x\|_1)(x_g)$ 
     $x^k = (x_1 + x_2)/2$ 
     $x^k = project(x^k, [l, u])$ 
     $r^{k+1} = x^k$ 
end for

```

ALGORITHM 7.5 FCSA

```

series Input:  $\rho = 1/L$ ,  $\alpha$ ,  $\beta$ ,  $t^1 = 1$   $x^0 = r^1$ 
for  $k = 1$  series to  $K$  do
     $x_g = r^k - \rho \nabla f(r^k)$ 
     $x_1 = prox_\rho(2\alpha \|x\|_{TV})(x_g)$ 
     $x_2 = prox_\rho(2\beta \|\Phi x\|_1)(x_g)$ 
     $x^k = (x_1 + x_2)/2$ ;  $x^k = \text{project}(x^k, [\mathbf{l}, \mathbf{u}])$ 
     $t^{k+1} = (1 + \sqrt{1 + 4(t^k)^2})/2$ 
     $r^{k+1} = x^k + ((t^k - 1)/t^{k+1})(x^k - x^{k-1})$ 
end for

```

7.2.2 EVALUATION

7.2.2.1 Experimental setup

Suppose an MR image x has n pixels, the partial Fourier transform R in problem (7.1) consists of m rows of an $n \times n$ matrix corresponding to the full 2D discrete Fourier transform. The m selected rows correspond to the acquired b . The sampling ratio is defined as m/n . The scanning duration is shorter if the sampling ratio is smaller. In MR imaging, we have certain freedom to select rows, which correspond to certain frequencies. In the following experiments, we select the corresponding frequencies in the following manner. In the K-space, we randomly obtain more samples at low frequencies and fewer samples at higher frequencies. This sampling scheme has been widely used for compressed MR image reconstruction (Lustig et al., 2007; Ma et al., 2008; Yang et al., 2010). Practically, the sampling scheme and speed in MR imaging also depend on the physical and physiological limitations (Lustig et al., 2007).

We implement our CSA and FCSA for problem (7.1) and apply them on 2D real MR images. The code that was used for the experiment is available for download http://ranger.uta.edu/~huang/R_FCSAMRI.htm. All experiments were conducted on a 2.4-GHz PC in a Matlab environment. We compare the CSA and FCSA with the classic MR image reconstruction method based on the CG (Lustig et al., 2007). We also compare them with two of the fastest MR image reconstruction methods, total variation-based compressed MRI (TVCMRI) <http://www1.se.cuhk.edu.hk/~sqma/TVCMRI.html>. (Ma et al., 2008) and reconstruction from partial Fourier data (RecPF) <http://www.caam.rice.edu/~optimization/L1/RecPF/>. (Yang et al., 2010). For fair comparisons, we download the codes from their websites and carefully follow their experiment setup. For example, the observation measurement b is synthesized as $b = Rx + \mathbf{n}$, where \mathbf{n} is the Gaussian white noise with standard deviation $\sigma = 0.01$. The regularization parameters α and β are set as 0.001 and 0.035. R and b are given as inputs, and x is the unknown target. For quantitative evaluation, the signal-to-noise ratio (SNR) is computed for each reconstruction result. Let x_0 be the original image and x a reconstructed image, the SNR is computed as: $\text{SNR} = 10 \log_{10}(V_s/V_n)$, where V_n is the mean square error between the original

image x_0 and the reconstructed image x ; $V_s = \text{var}(x_0)$ denotes the power level of the original image where $\text{var}(x_0)$ denotes the variance of the values in x_0 .

7.2.2.2 Visual comparisons

We apply all methods on four 2D MR images: cardiac, brain, chest, and artery. Fig. 7.1 shows these images. For convenience, they have the same size of 256×256 . The sample ratio is set to be approximately 20%. To perform fair comparisons, all methods run 50 iterations except that the CG runs only 8 iterations due to its higher computational complexity.

Figs. 7.2–7.5 show the visual comparisons of the reconstructed results by different methods. The FCSA always obtains the best visual effects on all MR images in less CPU time. The CSA is always inferior to the FCSA, which shows the effectiveness of acceleration steps in the FCSA for MR image reconstruction. The classical CG (Lustig et al., 2007) is far worse than the others because of its higher cost in each iteration, and the RecPF is slightly better than the TVCMRI, which is consistent with the observations in Ma et al. (2008) and Yang et al. (2010).

In our experiments, these methods have also been applied to the test images with the sample ratio set to 100%. We observe that all methods obtain almost the same reconstruction results, with SNR 64.8, after sufficient iterations. This was to be expected, since all methods are essentially solving the same formulation “Model (7.1).”

7.2.2.3 CPU time and SNRs

Fig. 7.6 gives the performance comparisons between different methods in terms of the CPU time over the SNR. Tables 7.1 and 7.2 tabulate the SNR and CPU time respectively by different methods, averaged over 100 runs for each experiment. The FCSA always obtains the best reconstruction results on all MR images by achieving the highest SNR in less CPU time. The CSA is always inferior to the FCSA, which shows the effectiveness of acceleration steps in the FCSA for MR image reconstruction. While the classical CG (Lustig et al., 2007) is far worse than the others because of its higher cost in each iteration, the RecPF is slightly better than

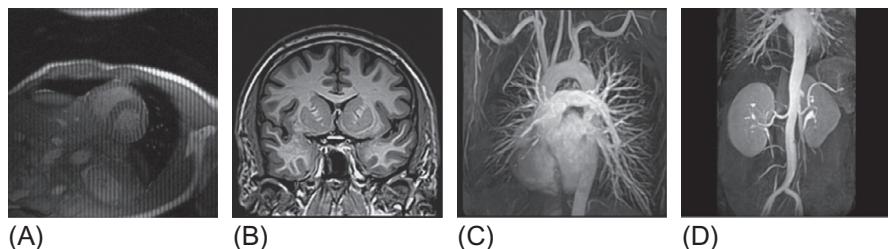
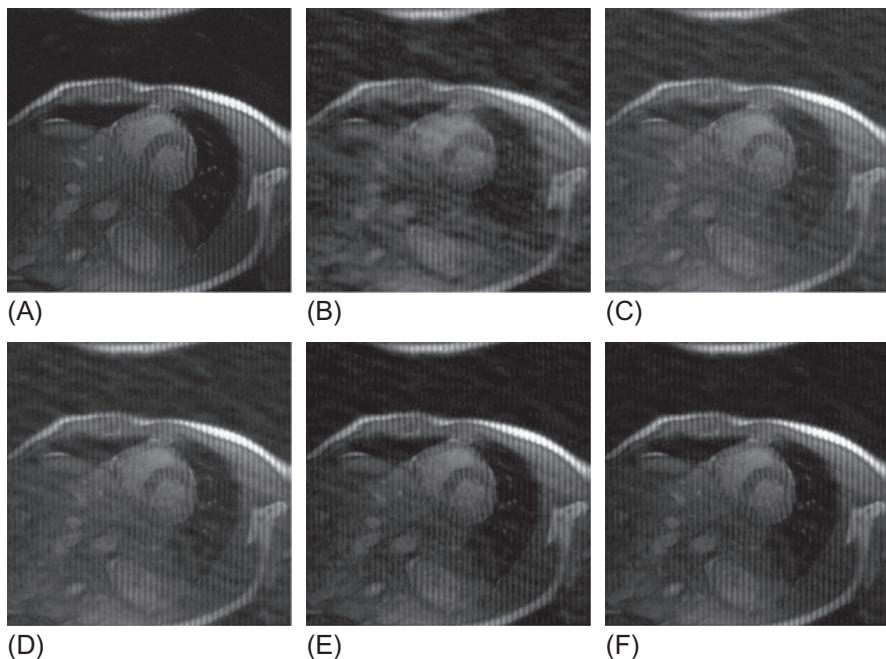


FIG. 7.1

MR images: (A) cardiac; (B) brain; (C) chest; (D) artery.

**FIG. 7.2**

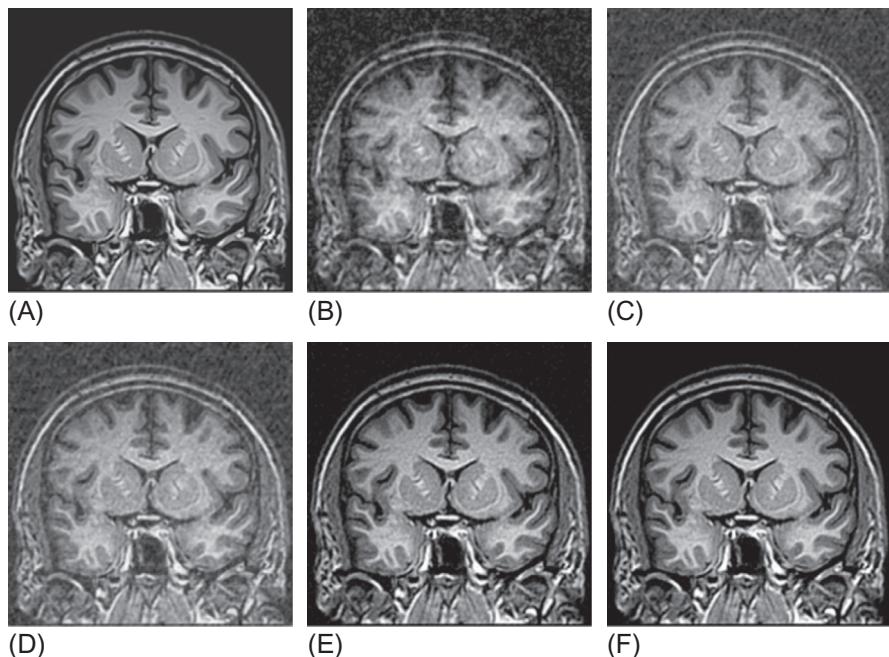
Cardiac MR image reconstruction from 20% sampling: (A) Original image; (B)–(F) are the reconstructed images by the CG ([Lustig et al., 2007](#)), TVCMRI ([Ma et al., 2008](#)), RecPF ([Yang et al., 2010](#)), CSA, and FCSA. Their SNRs are 9.86, 14.43, 15.20, 16.46, and 17.57 (dB). Their CPU times are 2.87, 3.14, 3.07, 2.22, and 2.29 (s).

the TVCMRI, which is consistent with the observations in [Ma et al. \(2008\)](#) and [Yang et al. \(2010\)](#).

7.2.2.4 Sample ratios

To test the efficiency of the discussed method, we further perform experiments on a full-body MR image with size of 924×208 . Each algorithm runs 50 iterations. Since we have shown that the CG method is far less efficient than the other methods, we will not include it in this experiment. The sample ratio is set to be approximately 25%. To reduce the randomness, we run each experiment 100 times for each parameter setting of each method. Examples of the original and recovered images by different algorithms are shown in [Fig. 7.7](#). From there, we can observe that the results obtained by the FCSA are not only visibly better, but also superior in terms of both the SNR and CPU time.

To evaluate the reconstruction performance with different sampling ratios, we use sampling ratios of 36%, 25%, and 20% to obtain the measurement b . Different methods are then used to perform reconstruction. To reduce the randomness, we run

**FIG. 7.3**

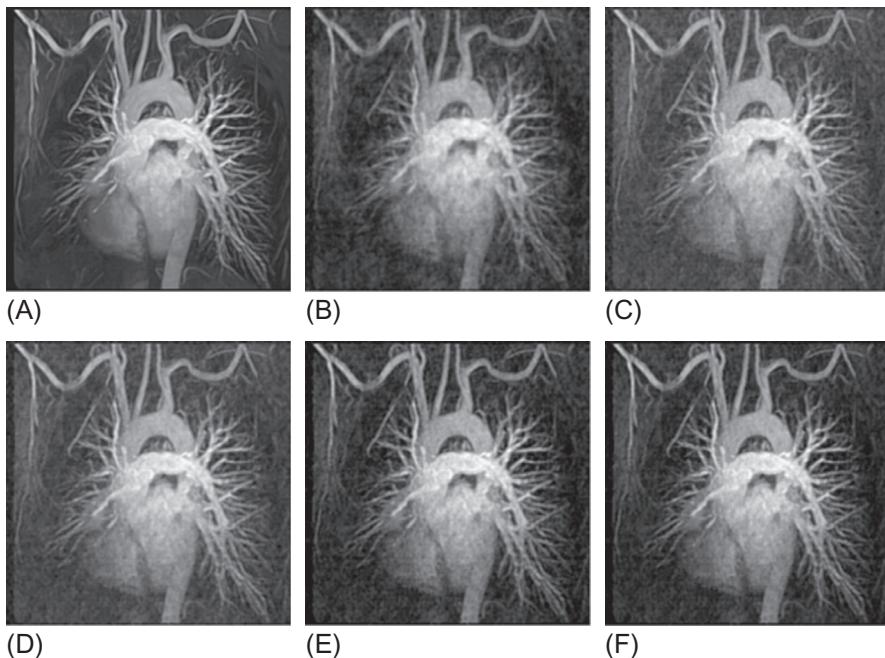
Brain MR image reconstruction from 20% sampling: (A) Original image; (B)–(F) are the reconstructed images by the CG ([Lustig et al., 2007](#)), TVCMRI ([Ma et al., 2008](#)), RecPF ([Yang et al., 2010](#)), CSA, and FCSA. Their SNRs are 8.71, 12.12, 12.40, 18.68, and 20.35 (dB). Their CPU times are 2.75, 3.03, 3.00, 2.22, and 2.20 (s).

each experiment 100 times for each parameter setting of each method. The SNR and CPU time are traced in each iteration for each method.

[Fig. 7.8](#) gives the performance comparisons between different methods in terms of the CPU time and SNR for sampling ratios of 36%, 25%, and 20%. The reconstruction results produced by the FCSA are far better than those produced by the CG, TVCMRI, and RecPF. The reconstruction performance of the FCSA is always the best in terms of both the reconstruction accuracy and the computational complexity, which further demonstrates the effectiveness and efficiency of the FCSA for compressed MR image construction.

7.2.3 SUMMARY

We have discussed an efficient algorithm for compressed MR image reconstruction. This work has the following benefits. First, the FCSA can efficiently solve a composite regularization problem including both TV term and l_1 -norm term, which can be easily extended to other medical image applications. Second, the

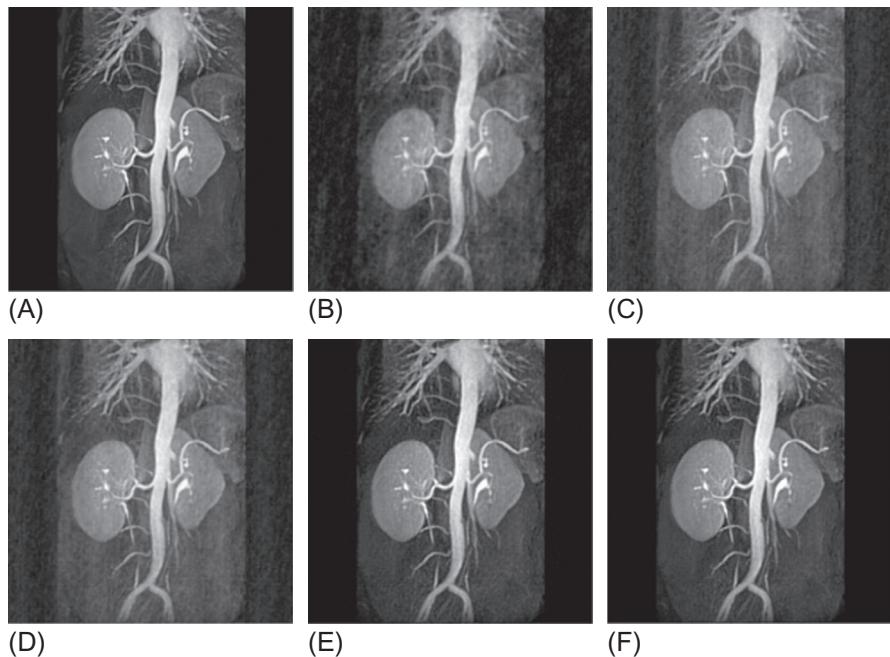
**FIG. 7.4**

Chest MR image reconstruction from 20% sampling: (A) Original image; (B)–(F) are the reconstructed images by the CG ([Lustig et al., 2007](#)), TVCMRI ([Ma et al., 2008](#)), RecPF ([Yang et al., 2010](#)), CSA, and FCSA. Their SNRs are 11.80, 15.06, 15.37, 16.53, and 16.07 (dB). Their CPU times are 2.95, 3.03, 3.00, 2.29, and 2.234 (s).

computational complexity of the FCSA is only $\mathcal{O}(p \log(p))$ in each iteration where p is the pixel number of the reconstructed image. It also has strong convergence properties. Experimental results further validate these benefits.

7.3 GROUP SPARSITY IN MULTICONTRAST MRI

This section investigates the benefits of the group sparsity in multicontrast MRI reconstruction. Multicontrast MRI is a useful technique to aid clinical diagnosis since it can achieve superior performance for clinical diagnosis over individual T1, T2, or proton-density weighted images. The superiority of multicontrast MRI lies in the fact that different contrasts emphasize different kinds of materials, which gives richer information for diagnosis. For group sparsity models, this section discusses an efficient algorithm to jointly reconstruct multiple T1/T2-weighted images of the same anatomical cross-section from partially sampled K-space data.

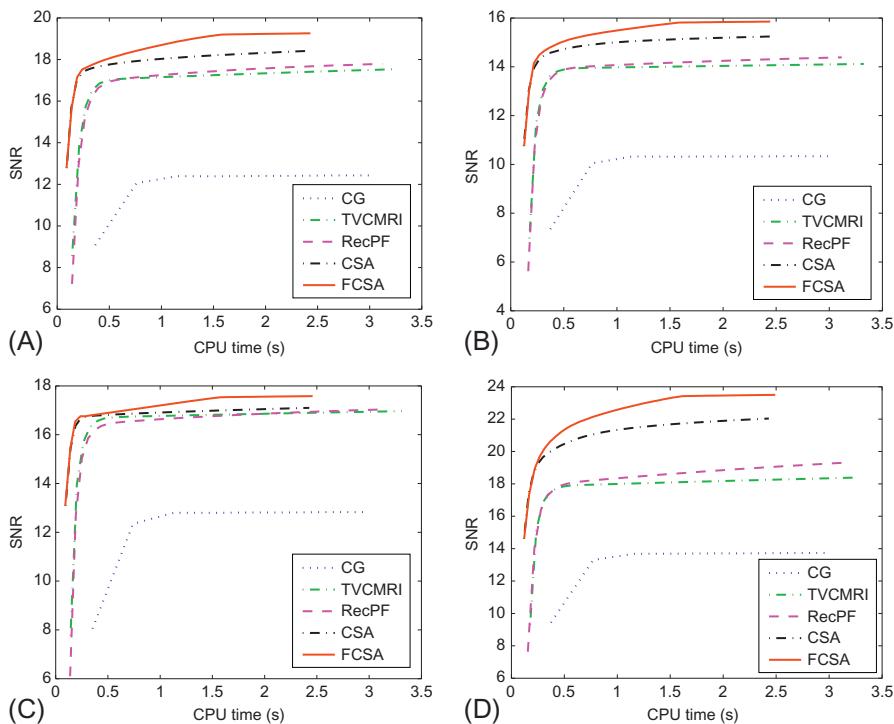
**FIG. 7.5**

Artery MR image reconstruction from 20% sampling: (A) Original image; (B)–(F) are the reconstructed images by the CG ([Lustig et al., 2007](#)), TVCMRI ([Ma et al., 2008](#)), RecPF ([Yang et al., 2010](#)), CSA, and FCSA. Their SNRs are 11.73, 15.49, 16.05, 22.27, and 23.70 (dB). Their CPU times are 2.78, 3.06, 3.20, 2.22, and 2.20 (s).

7.3.1 MODEL AND ALGORITHM

7.3.1.1 *Proposed fast multicontrast reconstruction*

In the multicontrast imaging setting, the different MR images denote MRI scans with different imaging weights. We make two observations about them: (1) the relative magnitudes of the gradients of images should be similar for the same spatial positions across multiple contrasts; (2) the wavelet coefficients of all MR images from the same spatial positions should have similar sparse modes. Intuitively, better performance can be achieved by fully exploiting group sparsity in both the wavelet and gradient domains. Motivated by these considerations, the joint reconstruction problem can be formulated as follows:

**FIG. 7.6**

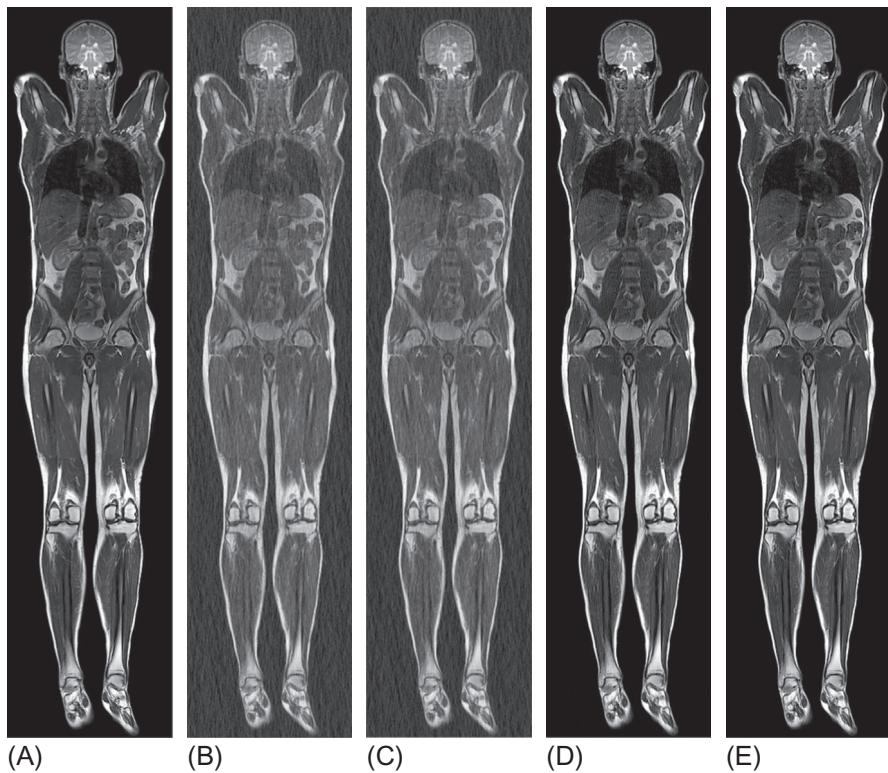
Performance comparisons (CPU time vs. SNR) on different MR images: (A) cardiac image; (B) brain image; (C) chest image; (D) artery image.

Table 7.1 Comparisons of the SNR (dB) Over 100 Runs

	CG	TVCMRI	RecPF	CSA	FCSA
Cardiac	12.43±1.53	17.54±0.94	17.79±2.33	18.41±0.73	19.26±0.78
Brain	10.33±1.63	14.11±0.34	14.39±2.17	15.25±0.23	15.86±0.22
Chest	12.83±2.05	16.97±0.32	17.03±2.36	17.10±0.31	17.58±0.32
Artery	13.74±2.28	18.39±0.47	19.30±2.55	22.03±0.18	23.50±0.20

Table 7.2 Comparisons of the CPU Time (s) Over 100 Runs

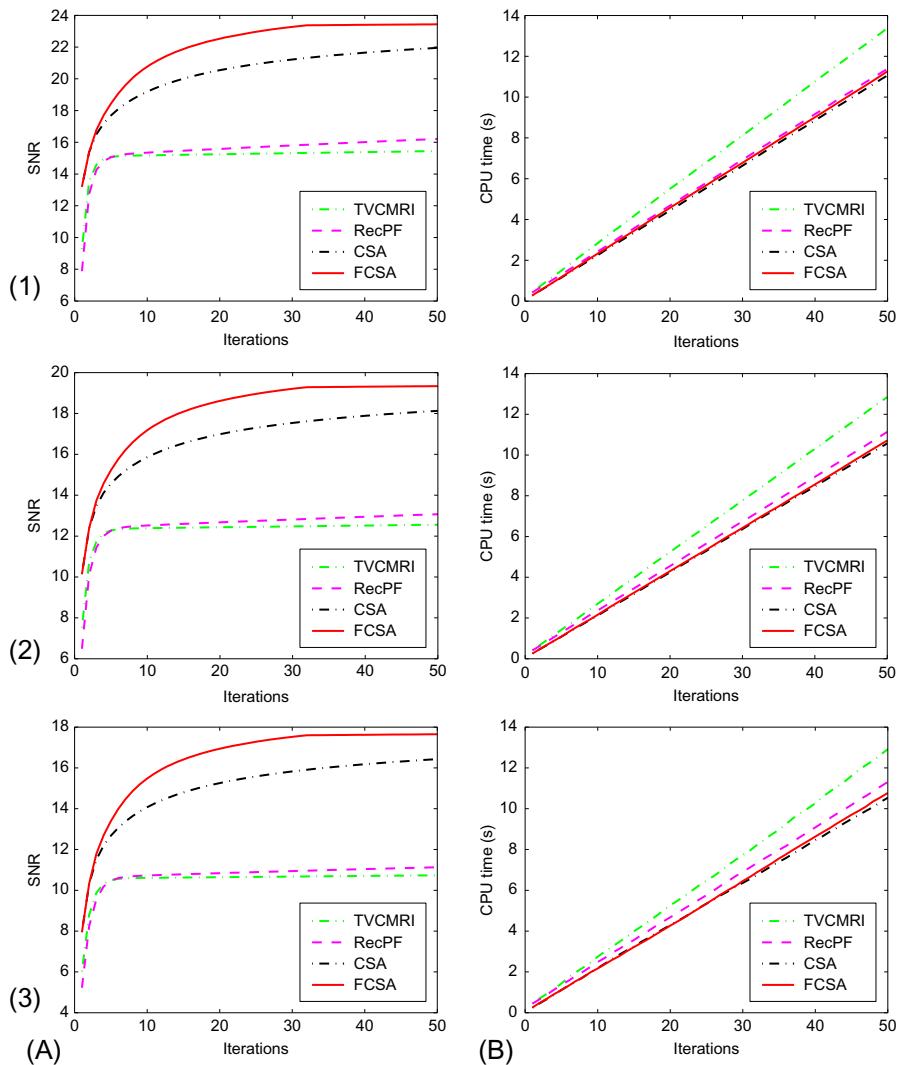
	CG	TVCMRI	RecPF	CSA	FCSA
Cardiac	2.82±0.16	3.16±0.10	2.97±0.12	2.27±0.08	2.30±0.08
Brain	2.81±0.15	3.12±0.15	2.95±0.10	2.27±0.12	2.31±0.13
Chest	2.79±0.16	3.00±0.11	2.89±0.07	2.21±0.06	2.26±0.07
Artery	2.81±0.17	3.04±0.13	2.94±0.09	2.22±0.07	2.27±0.13

**FIG. 7.7**

Full-Body MR image reconstruction from 25% sampling: (A) Original image; (B)–(E) are the reconstructed images by the TVCMRI ([Ma et al., 2008](#)), RecPF ([Yang et al., 2010](#)), CSA, and FCSA. Their SNRs are 12.56, 13.06, 18.21, and 19.45 (dB). Their CPU times are 12.57, 11.14, 10.20, and 10.64 (s).

$$\hat{X} = \arg \min_X \left\{ F(x) = \frac{1}{2} \sum_{s=1}^T \|F_s X(:, s) - y_s\|^2 + \alpha \|X\|_{JTV} + \beta \|\Phi X\|_{2,1} \right\}, \quad (7.8)$$

where $X = [x_1, x_2, x_T]$ is the set of all multicontrast images, and F_s and y_s are the sampling matrix and K-space measurements for the s th image respectively. The $\ell_{2,1}$ -norm is defined as $\|X\|_{2,1} = \sum_{i=1}^N \left(\sqrt{\sum_{s=1}^T (\Phi X_{is})^2} \right)$, which is the summation of the ℓ_2 -norm for each row. The JTV of X is defined as

**FIG. 7.8**

Performance comparisons on the full-body MR image with different sampling ratios. The sample ratios are: (1) 36%; (2) 25%, and (3) 20%. (A) Iterations vs. SNR (dB). (B) Iterations vs. CPU time (s).

$$\|X\|_{JTV} = \sum_{i=1}^N \sqrt{\sum_{s=1}^T ((\nabla_1 X_{is})^2 + (\nabla_2 X_{is})^2)}.$$

The JTV is also known as color total variation for color image denoising (Blomgren and Chan, 1998; Saito et al., 2011). To solve problem (7.8), we follow the FCSA (Huang et al., 2011b) scheme to decompose it into two subproblems. Let $f(X) = \frac{1}{2} \sum_{s=1}^T \|F_s X_s - y_s\|_2^2$, $g_1(X) = \alpha \|X\|_{JTV}$, and $g_2(X) = \beta \|\Phi X\|_{2,1}$. **Algorithm 7.6** outlines the whole algorithm for this problem. Here, $\nabla f(X)$ denotes the gradient of $f(X)$ and L_f denotes its Lipschitz constant.

ALGORITHM 7.6 PROPOSED FCSA-MT

```

series Input:  $\rho = \frac{1}{L_f}$ ,  $\alpha$ ,  $\beta$ ,  $t^1 = 1$   $Z = X^0$ 
for  $k = 1$  series to  $K$  do
     $Y = Z - \rho \nabla f(Z)$ ,  $s = 1, \dots, T$ 
     $X_1 = \arg \min_X \left\{ \frac{1}{4\rho} \|X - Y\|^2 + \alpha \|X\|_{JTV} \right\}$ 
     $X_2 = \arg \min_X \left\{ \frac{1}{4\rho} \|X - Y\|^2 + \beta \|\Phi X\|_{2,1} \right\}$ 
     $X^k = \frac{X_1 + X_2}{2}$ ;  $t^{k+1} = \frac{1 + \sqrt{1 + 4(t^k)^2}}{2}$ 
     $Z = X^k + \frac{t^k - 1}{t^{k+1}} [X^k - X^{k-1}]$ 
end for

```

For the subproblem of X_2 :

$$X_2 = \arg \min_X \left\{ \frac{1}{4\rho} \|X - Y\|^2 + \beta \|\Phi X\|_{2,1} \right\}, \quad (7.9)$$

there is a closed-form solution, by soft thresholding:

$$(X_2)_i = \Phi^T \left(\max \left(1 - \frac{2\rho\beta}{\|(\Phi X)_i\|_2}, 0 \right) (\Phi X)_i \right), \quad (7.10)$$

where $(\cdot)_i$ denotes the i th row of the matrix. The efficiency of the whole algorithm is highly dependent on how quickly we can solve the X_1 subproblem in each iteration.

For the subproblem of X_1 :

$$X_1 = \arg \min_X \left\{ \frac{1}{4\rho} \|X - Y\|^2 + \alpha \|X\|_{JTV} \right\}. \quad (7.11)$$

There is no closed-form solution. The fast gradient projection (FGP) algorithm for TV (Beck and Teboulle, 2009a) previously proposed cannot directly solve it, due to the different formulation. Fortunately, we have developed a new method, called the fast joint-gradient projection (FJGP) algorithm, for this JTV problem. Following the FGP (Beck and Teboulle, 2009a), we consider a dual method for problem (7.11). Supposing the size of each image is m by n with $m \times n = N$, we reshape the image

matrices X, Y to $m \times n \times T$ for convenience. Let P be an $(m - 1) \times n \times T$ matrix, Q be an $m \times (n - 1) \times T$ matrix, and that they satisfy:

$$\begin{cases} \sum_{s=1}^T (P_{i,j,s}^2 + Q_{i,j,s}^2) \leq 1 & i = 1, 2, \dots, m - 1; j = 1, 2, \dots, n - 1, \\ |P_{i,n,s}| \leq 1 & i = 1, 2, \dots, m - 1; s = 1, 2, \dots, T, \\ |Q_{m,j,s}| \leq 1 & j = 1, 2, \dots, n - 1; s = 1, 2, \dots, T. \end{cases} \quad (7.12)$$

A linear operator is defined as $\mathcal{L}(P, Q)_{i,j,s} = P_{i,j,s} - P_{i-1,j,s} + Q_{i,j,s} - Q_{i,j-1,s}$, where $i = 1, \dots, n_1, j = 1, \dots, n_2$, and $s = 1, \dots, T$. The \mathcal{L}^T is defined as $\mathcal{L}^T(X) = (P, Q)$, where $P \in \mathbf{R}^{(n_1-1) \times n_2 \times T}$ and $Q \in \mathbf{R}^{n_1 \times (n_2-1) \times T}$ are defined as

$$\begin{aligned} P_{i,j,s} &= x_{i,j,s} - x_{i+1,j,s}, & i = 1, \dots, n_1 - 1, j = 1, \dots, n_2, s = 1, \dots, T, \\ Q_{i,j,s} &= x_{i,j,s} - x_{i,j+1,s}, & i = 1, \dots, n_1, j = 1, \dots, n_2 - 1, s = 1, \dots, T. \end{aligned}$$

Therefore the optimal solution for problem (7.11) is $X^* = Y - 2\alpha\rho\mathcal{L}(P^*, Q^*)$, where (P^*, Q^*) is the optimal solution for

$$\min_{P,Q} \left\{ h(P, Q) = Y - 2\alpha\rho\mathcal{L}(P, Q)_F^2 \right\}, \quad (7.13)$$

where $\|\cdot\|_F$ denotes the Frobenius norm. Note that problem (7.13) could be accelerated by FISTA (Beck and Teboulle, 2009a). The whole algorithm for problem (7.11) is summarized in Algorithm 7.7. Our recent study shows that the JTV can also improve parallel MRI reconstruction (Chen et al., 2013).

ALGORITHM 7.7 PROPOSED FJGP FOR JOINT TOTAL VARIATION

```

series Input:  $\rho, \alpha, Y, P^0, Q^0, U, V$ 
for  $k = 1$  series to  $K$  do
     $t^{k+1} = \frac{1 + \sqrt{1 + 4(t^k)^2}}{2}$ 
    for  $s = 1$  series to  $T$  do
         $(P^k, Q^k) = \text{Proj}[(U, V) + \frac{1}{16\rho\alpha}\mathcal{L}^T[Y - 2\rho\alpha\mathcal{L}(U, V)]]$ 
         $(U, V) = (P^k, Q^k) + \frac{t^{k+1}}{t^k}((P^k - P^{k-1}, Q^k - Q^{k-1}))$ 
    end for
end for
 $X = Y - 2\rho\alpha\mathcal{L}(P^K, Q^K)$ 

```

The projection operator $\text{Proj}(P, Q) = (U, V)$ is used to force (P, Q) to satisfy the conditions (7.12):

$$U_{i,j,s} = \begin{cases} \frac{P_{i,j,s}}{\max(1, \sqrt{\sum_{s=1}^T P_{i,j,s}^2 + Q_{i,j,s}^2})} & i = 1, 2, \dots, m - 1; j = 1, 2, \dots, n - 1, \\ \frac{P_{i,n,s}}{\max(1, \sqrt{\sum_{s=1}^T P_{i,n,s}^2})} & i = 1, 2, \dots, m - 1 \end{cases} \quad (7.14)$$

and

$$V_{i,j,s} = \begin{cases} \frac{Q_{i,j,s}}{\max(1, \sqrt{\sum_{s=1}^T P_{i,j,s}^2 + Q_{i,j,s}^2})} & i = 1, 2, \dots, m-1; j = 1, 2, \dots, n-1, \\ \frac{Q_{m,j,s}}{\max(1, \sqrt{\sum_{s=1}^T Q_{m,j,s}^2})} & j = 1, 2, \dots, n-1. \end{cases} \quad (7.15)$$

It can be observed that all operations in Algorithm 7.2 are linear. Therefore the total computational complexity is $\mathcal{O}(TN)$. It can be easily proved that Algorithm 7.2 achieves the optimal convergence rate $F(X^k) - F(X^*) \leq \mathcal{O}(1/k^2)$ (Beck and Teboulle, 2009a,b). Due to the trade-off between efficiency and effectiveness, the FJGP algorithm only runs for one iteration in our implementation. For the entire Algorithm 7.1, step 1 takes $\mathcal{O}(TN \log N)$ if the fast Fourier transform (FFT) is applied. Steps 2 and 3 for the two subproblems take $\mathcal{O}(TN)$. Therefore the computational complexity for the whole algorithm is $\mathcal{O}(TN \log N)$. In addition, it is accelerated by FISTA, which has very fast convergence speed.

7.3.2 EVALUATION

The proposed method was evaluated on three datasets: the SRI24 brain atlas (Rohlfing et al., 2009), the complex-valued Shepp-Logan phantom data, and in vivo brain data. All reconstruction methods were implemented in Matlab (MathWorks, Natick, MA) on a desktop computer with an Intel i7-3770 central processing unit and 12-GB random-access memory. We compared our algorithm with the conventional CS-MRI methods SparseMRI (Lustig et al., 2007), TVCMRI (Ma et al., 2008), RecPF (Yang et al., 2010), FCSA (Huang et al., 2011b), and multicontrast MRI methods Bayesian CS (BCS) (Bilgic et al., 2011) and GroupSparseMRI (Majumdar and Ward, 2011) (GSMRI for short). For fair comparison, all codes were downloaded from the authors' websites and we carefully followed their experimental setup. All the convex relaxation methods were run for 100 iterations and the parameters were set the same for them. Due to the slow convergence speed of BCS (eg, 26.4 h for the SRI24 brain atlas (Rohlfing et al., 2009)), we only ran it for 6000 iterations as we are interested in fast reconstruction. The BCS method has been accelerated in a recent work (Cauley et al., 2013), but it still requires 9 min to reconstruct the images. The accelerated version is much less accurate than the original version and is therefore not compared. We added Gaussian white noise with 0.01 standard deviation for the simulated data. The images with Nyquist rate sampling were used as reference images. SNR and relative error (RE) were used for result evaluation. $SNR = 10 \log 10(V_s/V_n)$, where V_n is the mean square error between the reference image x_0 and the reconstructed image x ; $V_s = var(x_0)$ denotes the power level of the original image, where $var(x_0)$ denotes the variance of the values in x_0 . $RE = 100\% \times \|x - x_0\|_2 / \|x_0\|_2$.

7.3.2.1 SRI24 multichannel brain atlas data

This experiment was conducted on an MR image extracted from the SRI24 atlas (Rohlfing et al., 2009). Structural scans of the atlas features were obtained on a 3.0T GE scanner with an eight-channel head coil with three different contrast settings:

- i. For T1-weighted structural images: 3D axial IR-prep SPoiled Gradient Recalled (SPGR), TR = 6.5 ms, TE = 1.54 ms, number of slices = 124, slice thickness = 1.25 mm.
- ii. For proton density-weighted (early-echo) and T2-weighted (late-echo) images: 2D axial dual-echo fast spin echo (FSE), TR = 10,000 ms, TE = 14/98 ms, number of slices = 62, slice thickness = 2.5 mm.

The field-of-view covers a region of $240\text{ mm} \times 240\text{ mm}$ with resolution of 256×256 pixels. The sampling mask is Gaussian random with variable density and reduction factor $R = 4$. More samples were acquired at low frequencies and fewer samples were acquired at higher frequencies (Huang et al., 2011b; Lustig et al., 2007). Reconstructions were performed using conventional CS-MRI methods (Huang et al., 2011b; Lustig et al., 2007; Ma et al., 2008; Yang et al., 2010), multicontrast MRI methods (Bilgic et al., 2011; Majumdar and Ward, 2011), and the proposed method. All the convex relaxation methods ran for 100 iterations and BCS ran for 6000 iterations, due to its higher computational complexity.

7.3.2.2 Complex-valued Shepp-Logan phantoms data

To validate the proposed method on complex-valued data, we conducted experiments on two complex-valued numerical phantoms (Bilgic et al., 2011). Each of these has a resolution of 128×128 pixels. The real parts and imaginary parts tend to be similar but not exactly the same, which was used to validate the stability of the proposed method. The phantom images are piecewise smooth, where the TV or JTV would significantly increase the reconstruction accuracy. The sampling mask is single-slice (Bilgic et al., 2011; Lustig et al., 2007) with variable density and reduction factor $R = 4$. Reconstructions were performed with BCS (Bilgic et al., 2011), GSMRI (Majumdar and Ward, 2011), and the proposed method.

7.3.2.3 Complex-valued turbo spin echo slices with early and late TEs data

In vivo images were used to further validate the performance of the proposed method. These images were obtained with two different TE settings, using a TSE sequence ($1\text{ mm} \times 1\text{ mm}$ in-plane spatial resolution with 3-mm-thick contiguous slices, TR = 6000 ms, TE1 = 27 ms, TE2 = 94 ms) (Bilgic et al., 2011). This dataset contains 38 slices with 256×256 resolution. They are much less compressible than the previous two datasets. The sampling mask consisted of multiple radial lines (Block et al., 2007; Chang et al., 2006; Yang et al., 2010; Ye et al., 2007) with reduction factor $R = 4$. Reconstructions were performed using BCS (Bilgic et al., 2011), GSMRI (Majumdar and Ward, 2011), and the proposed method.

7.3.2.4 The benefit of group sparsity on both wavelet and gradient domains

Fig. 7.9 demonstrates the benefit of the proposed method when utilizing JTV and group wavelet sparsity on the SRI24 dataset. Comparing the proposed FJGP and conventional FGP (Beck and Teboulle, 2009a), the JTV denoising is much more powerful on multicontrast MR data. With more iterations, our algorithm achieves higher accuracy and less computational time. The results produced by the combination of group wavelet sparsity and JTV are always better than those produced by the group wavelet sparsity or JTV only. This confirms the benefit of the combination of group wavelet sparsity and JTV for multicontrast CS-MRI.

7.3.2.5 Results on SRI24 multichannel brain atlas data

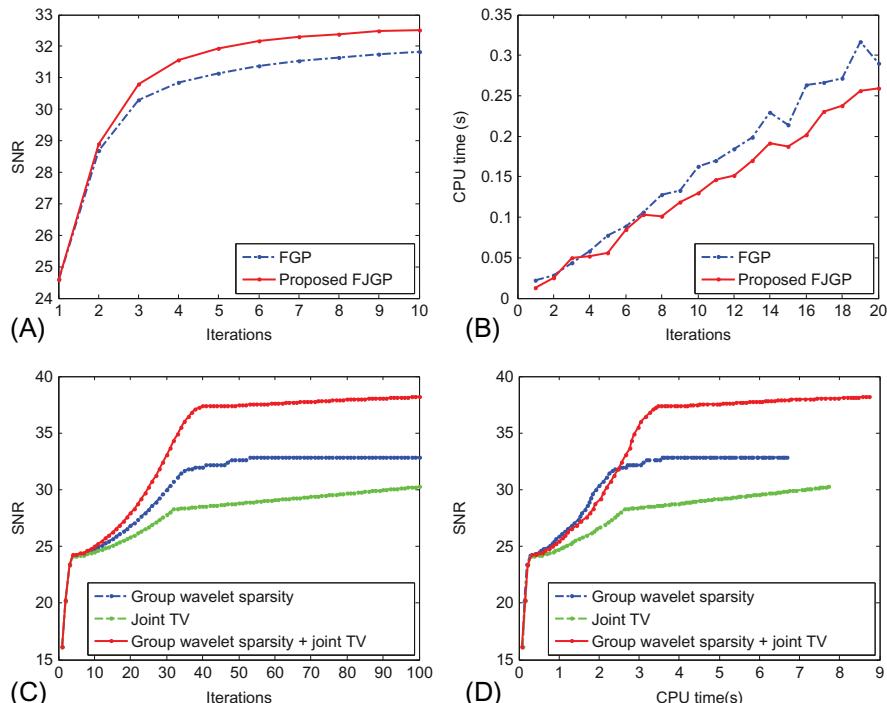


FIG. 7.9

Demonstration of the benefit of the proposed method for SRI24 dataset. (A) The SNRs of the denoising result with FGP (18) and the proposed FJGP. (B) Comparison of the computational time between FGP (18) and the proposed FJGP. (C) Comparison of the SNR in terms of iterations among only group wavelet sparsity, only joint TV, and their combination. (D) Comparison of the SNR in terms of CPU time among only group wavelet sparsity, only joint TV, and their combination.

Fig. 7.10 depicts the SRI24 reconstruction results obtained by the conventional CS-MRI methods (Huang et al., 2011b; Lustig et al., 2007; Ma et al., 2008; Yang et al., 2010), multicontrast MRI methods (Bilgic et al., 2011; Majumdar and Ward, 2011), and the proposed method. These images are very compressible. It is difficult to observe visible artifacts for all reconstruction results. For better visualization, the reconstruction errors are also shown on the same scale. We found the reconstruction with our method had the smallest error. **Table 7.3** shows all the SNRs, REs as well as the reconstruction time for each algorithm. Our algorithm had the highest accuracy with comparable computational cost to the fastest CS-MRI algorithms. Due to the inherent shortcoming of the Bayesian CS framework (Ji et al., 2009), it has a huge computational cost that may make it practically impossible to use in most applications. GSMRI applies SPGL1 (Van Den Berg and Friedlander, 2008) to solve the group wavelet sparsity problem. It had similar accuracy with BCS on this data, while the computational cost was much less.

Fig. 7.11 presents the convergence speed of each algorithm with convex relaxation in terms of iterations and CPU time. Inherent from the fast convergence rate of FISTA, FCSA, and the proposed algorithm outperformed all other algorithms on this dataset. However, FCSA cannot reconstruct the multicontrast MR images jointly, but only individually. That is why it is always inferior to the proposed method. In general, the conventional CS-MRI methods are not good as joint reconstruction methods, which has also been validated in previous works (Bilgic et al., 2011; Majumdar and Ward, 2011). In later experiments, we only compared the proposed method with multicontrast reconstruction methods (Bilgic et al., 2011; Majumdar and Ward, 2011).

7.3.2.6 Results on Complex-valued Shepp-Logan phantoms data

Absolute values of the reconstruction results after undersampling with $R = 4$ are presented in **Fig. 7.12**. The proposed method was better than BCS in terms of SNR and RE on this dataset, and far better than GSMRI. At first glance, it seems that BCS achieved similar reconstruction results from visual observation. This is because such images are ideal examples for BCS, where the images are extremely piecewise smooth. However, for in vivo data, as we will show later, our method will be much more stable and robust. In addition, BCS converges the fastest on these ideal images among all those used in the experiments, but it is still substantially slower than our method. Visible artifacts can be found in the results reconstructed by GSMRI. This is because these images are piecewise smooth, which is perfect for using gradient-based methods. However, it is still unknown how to combine TV or JTV in SPGL1 for GSMRI. Across different contrasts, the gradients and wavelet coefficients of these phantom images are not exactly the same, which validates the stability of the proposed method. When the sparsity patterns of different images tend to be the same, our algorithm can obtain very good results.

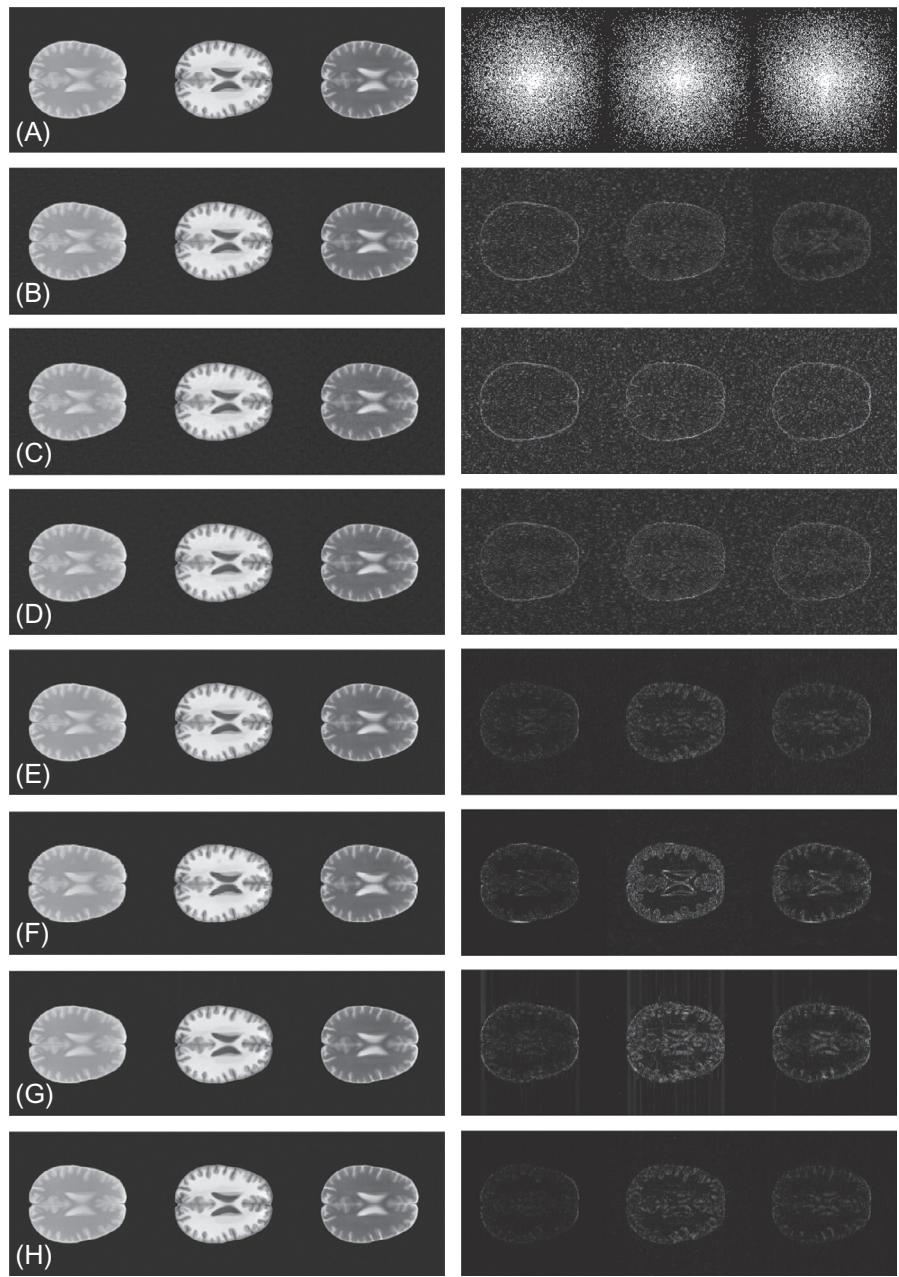


FIG. 7.10

Representative reconstruction results for SRI24 atlas images after undersampling with $R = 4$. (A) Atlas images at Nyquist rate sampling (left) and the sampling masks (right). (B)–(H) The reconstruction results (left) and the absolute errors (right) with SparseMRI (Lustig et al., 2007), TVCMRI (Ma et al., 2008), RecPF (Yang et al., 2010), FCSA (Huang et al., 2011b), BCS (Bilgic et al., 2011), GSMRI (Majumdar and Ward, 2011), and the proposed method respectively.

Table 7.3 The SNRs, Computational Times, and REs for the Reconstructions in Fig. 7.10

	SparseMRI	TVCMRI	RecPF	FCSA	BCS	GSMRI	Proposed
SNR (dB)	25.13	24.03	25.80	35.64	31.49	30.40	37.97
RE (%)	4.3	4.9	4.0	1.3	2.3	2.1	0.9
Time (s)	44.96	7.60	8.53	8.36	7134.1	11.81	8.78

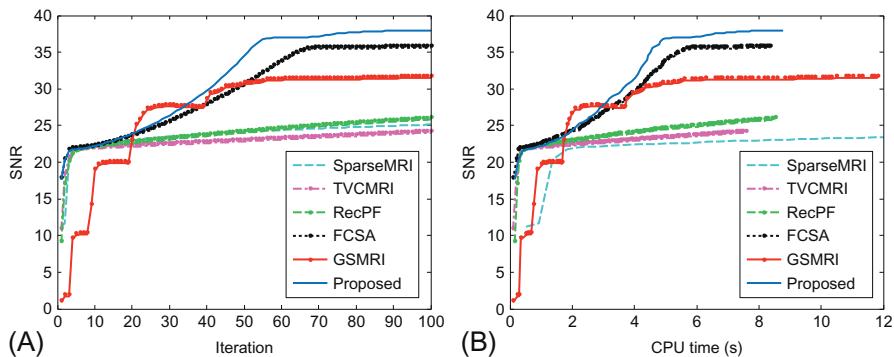


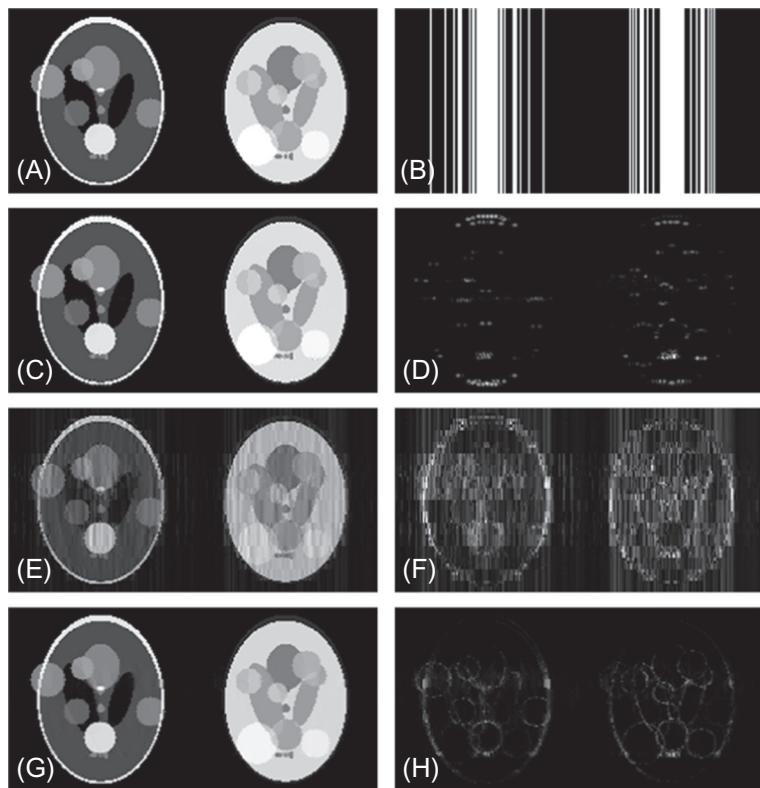
FIG. 7.11

Performance comparisons of reconstruction results for SRI24 atlas among algorithms with convex relaxation. (A) SNR vs. iteration. (B) SNR vs. CPU time (s). Due to the high computational cost of SparseMRI, we only show its SNRs for the first 12 s.

7.3.2.7 Results on Complex-valued turbo spin echo slices with early and late TEs

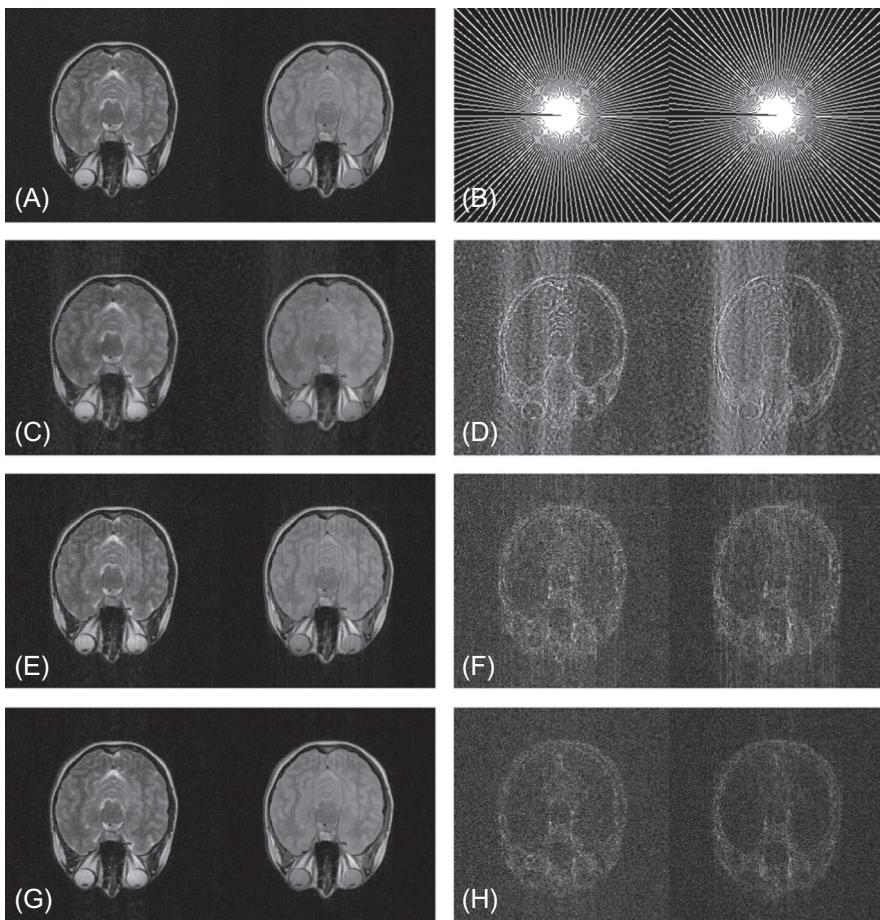
The reconstruction results on the in vivo human brain MRI data are shown in Fig. 7.13. These images are less compressible than the previous two datasets. With the same reduction factor, all algorithms performed worse on this dataset than they performed on the previous datasets. The result of BCS is significantly blurred at the edges, while the result of GSMRI contains obvious artifacts. By contrast, the result of our method is closest to the original one. It takes more than 10,000 s for BCS for this reconstruction (it was still not converged), while our algorithm obtained an acceptable result within only around 12 s.

With the same datasets, additional reconstructions were performed to quantify the performance of the three multicontrast MRI methods. The efficiency and effectiveness of the proposed method were validated under different sampling schemes with various reduction factors. The comparisons with GSMRI (Majumdar and Ward, 2011) are tabulated in Table 7.4, and the comparisons with BCS (Bilgic et al., 2011) are shown in Table 7.5. For speed, we used 128×128 low-resolution images when comparing the results with BCS. For the same data, the results on the non-Cartesian

**FIG. 7.12**

Reconstruction results for the complex-valued Shepp-Logan phantoms after undersampling with reduction factor $R = 4$. (A) Magnitudes of phantoms at Nyquist rate sampling. (B) The sampling masks. (C) The reconstruction result with Bayesian CS. Its SNR is 26.43. (D) Absolute error of Bayesian CS reconstruction. $RE = 4.3\%$. (E) The reconstruction result with GSMRI. Its SNR is 14.32. (F) Absolute error of GSMRI reconstruction. $RE = 17.7\%$. (G) The reconstruction result with the proposed method. Its SNR is 27.15. (H) Absolute error of the proposed reconstruction. $RE = 4.2\%$. The reconstruction times were 271.2, 10.37, and 8.64 s, respectively.

masks were much better than those on the Cartesian mask. This is because the Cartesian sampling matrix is less incoherent. Also, with the same sampling scheme, the accuracy of each algorithm is higher with more measurements. When compared with different algorithms in the same condition (same sampling scheme and same reduction factor), our algorithm always outperformed the other two in terms of both reconstruction accuracy and computational complexity.

**FIG. 7.13**

Reconstruction results for the complex-valued in vivo TSE images after undersampling with reduction factor $R = 4$. (A) Magnitude images at Nyquist rate sampling. (B) The sampling masks. (C) The reconstruction result with Bayesian CS. Its SNR is 12.34. (D) Absolute error of Bayesian CS reconstruction. $RE = 27.4\%$. (E) The reconstruction result with GSMRI. Its SNR is 15.07. (F) Absolute error of GSMRI reconstruction. $RE = 18.9\%$. (G) The reconstruction result with the proposed method. Its SNR is 17.07. (H) Absolute error of the proposed reconstruction. $RE = 14.7\%$. The reconstruction times were: 10,074.3, 18.21, and 12.04 s, respectively.

7.3.2.8 Discussion

One may note that there are two parameters that need to be tuned for most convex relaxation methods including the proposed method, while no parameters are required for BCS (Bilgic et al., 2011) and GSMRI (Majumdar and Ward, 2011). Fortunately,

Table 7.4 Additional Reconstruction Results on the SRI24 and TSE Datasets, With Comparisons Between GSMRI ([Majumdar and Ward, 2011](#)) and the Proposed Method

Dataset	Schemes	R	GSMRI			Proposed		
			SNR (dB)	RE (%)	Time (s)	SNR (dB)	RE (%)	Time (s)
SRI24	Random (Fig. 7.10)	4	30.40	2.1	11.81	37.97	0.9	8.78
	Random	5	29.19	2.7	11.40	34.33	1.5	8.31
	Slice (Fig. 7.12)	3	22.16	6.1	11.65	26.54	3.7	8.68
	Slice	4	20.08	7.8	11.54	24.77	4.5	9.39
	Radial (Fig. 7.13)	4	28.98	2.8	11.58	33.63	1.6	9.06
	Radial	5	27.09	3.4	11.20	32.33	1.9	9.25
TSE	Random	3	17.86	14.0	18.88	19.08	12.0	12.85
	Random	4	16.26	16.4	17.86	17.08	13.7	11.47
	Slice	2	16.34	16.2	20.19	17.28	14.2	12.39
	Slice	3	13.33	22.3	18.30	14.38	19.4	11.51
	Radial	3	17.18	15.1	20.19	18.88	12.3	13.17
	Radial	4	15.07	18.9	18.21	17.07	14.7	12.04

Table 7.5 Additional Reconstruction Results on the SRI24 and TSE Datasets, With Comparisons Between BCS ([Bilgic et al., 2011](#)) and the Proposed Method (for Speed, We Used 128×128 Low-Resolution Images for These Comparisons)

Dataset	Schemes	R	BCS			Proposed		
			SNR (dB)	RE (%)	Time (s)	SNR (dB)	RE (%)	Time (s)
SRI24	Random	3.5	32.29	2.0	2706.8	36.21	0.4	2.48
	Random	4	30.13	1.1	2129.9	33.53	0.6	2.46
	Slice	3	23.04	5.8	1919.8	24.68	4.8	2.02
	Slice	4	19.78	8.5	1838.4	20.70	7.6	2.09
	Radial	4	30.65	2.4	2384.4	32.72	1.9	2.15
	Radial	5	27.73	3.3	2335.0	29.82	2.6	2.09
TSE	Random	3	15.63	17.2	9262.7	23.87	6.7	3.25
	Random	4	12.95	23.1	5306.9	21.39	8.5	3.24
	Slice	3	15.44	16.0	177.6	15.72	15.1	2.65
	Slice	3.5	13.86	18.7	378.3	14.02	18.2	2.97
	Radial	3	15.91	16.5	3384.1	22.73	7.6	3.25
	Radial	4	13.52	21.3	4759.3	19.67	10.5	3.18

the parameters are easy to tune when the sampling scheme is fixed. The parameters determine the relative weights between the least squares fitting and the sparsity terms. The user could manually set how important the sparseness of the data is relative to the least squares fitting. However, BCS and GSMRI always seek the sparsest solution under a fixed tolerance of the least squares fitting, which may result in oversmoothing, especially for those less compressible images. The parameters to obtain the experimental results are set as $\alpha = 0.01$ or 0.1 , $\beta = 0.035$. Although they are not necessarily the optimal values, they were good enough to demonstrate the superiority of the proposed method. Interested readers may refer to existing methods ([Liang et al., 2009](#); [Weller et al., 2010](#)) for more information on how to combine parallel imaging techniques.

7.3.3 SUMMARY

We have discussed an efficient algorithm for multicontrast CS-MRI. In the discussion, we demonstrated the model that applied group sparsity in both wavelet and gradient domains and the algorithm to efficiently solve the optimization problem. The presented FCSA algorithm has fast convergence speed borrowed from FISTA, and each iteration only costs $\mathcal{O}(TN \log N)$.

7.4 TREE SPARSITY IN ACCELERATED MRI

This section investigates the benefits of tree sparsity in accelerated MRI. In contrast to conventional CS-MRI that only relies on the sparsity of MR images in the wavelet or gradient domain, the tree sparsity exploits the wavelet tree structure to improve CS-MRI, which can further improve the reconstruction quality.

7.4.1 MODEL AND ALGORITHM

To validate the benefit of tree sparsity in accelerated MRI, we demonstrate two algorithms that efficiently solve the constrained and unconstrained tree-based CSMRI problems. The tree structure in MR images is approximated as overlapping groups ([Jacob et al., 2009](#); [Rao et al., 2011](#)). The unconstrained problem is solved in a FISTA ([Beck and Teboulle, 2009a](#)) framework and the constrained problem is solved in a NEsterov Shrinkage Thresholding Algorithm ([Becker et al., 2011](#)) framework. Both FISTA and NESTA have the optimal convergence rate for first-order methods, that is, $\mathcal{O}(1/k^2)$ in function value where k is the iteration number ([Nesterov, 1983](#)).

7.4.1.1 Unconstrained tree-based MRI

Following overlapping group sparsity algorithms ([Jacob et al., 2009](#); [Rao et al., 2011](#)), the unconstrained MRI problem with tree sparsity can be formulated as

$$\hat{x} = \min_x \left\{ \frac{1}{2} \|Rx - b\|_2^2 + \beta \sum_{g \in \mathcal{G}} \|\Phi x\|_2 \right\}, \quad (7.16)$$

where x is the MR image to be reconstructed, R is the partial Fourier transform, b is the measurement vector, Φ denotes the wavelet transform, and β is a positive parameter that needs to be tuned. Here, g denotes one of the groups that encourages tree sparsity (eg, one node and its parent) and \mathcal{G} denotes the set of all such groups. Due to the nonsmoothness and nonseparability of the overlapping group penalty, it is not easy to solve the problem directly. Instead, we introduce a variable z to constrain the problem:

$$\hat{x} = \arg \min_{x,z} \left\{ \frac{1}{2} \|Rx - b\|_2^2 + \beta \sum_{g \in \mathcal{G}} \|z_g\|_2 + \frac{\lambda}{2} \|z - G\Phi x\|_2^2 \right\}, \quad (7.17)$$

where λ is another positive parameter and G is a binary matrix to duplicate the overlapped entries. z is the extended vector of wavelet coefficients x without overlapping.

All terms in our model are convex. For the z subproblem:

$$z_g = \arg \min_{z_g} \left\{ \beta \|z_g\|_2 + \frac{\lambda}{2} \|z_g - (G\Phi x)_g\|_2^2 \right\}, \quad g \in \mathcal{G}. \quad (7.18)$$

It has closed-form solution by soft thresholding:

$$z_g = \max \left(\|r\|_2 - \frac{\beta}{\lambda}, 0 \right) \frac{r}{\|r\|_2}, \quad g \in \mathcal{G}, \quad (7.19)$$

where $r = (G\Phi x)_g$. We denote this step by $z = \text{shrinkgroup}(G\Phi x, \frac{\beta}{\lambda})$ for convenience. For the x subproblem:

$$x = \arg \min_x \left\{ \frac{1}{2} \|Rx - b\|_2^2 + \frac{\lambda}{2} \|z - G\Phi x\|_2^2 \right\}. \quad (7.20)$$

This is a combination of two quadratic terms and has closed-form solution: $x = (R^T R + \lambda \Phi^T G^T G \Phi)^{-1} (R^T b + \Phi^T G^T z)$. However, the inverse of $R^T R + \lambda \Phi^T G^T G \Phi$ is not easily obtained. In order to validate the benefit tree structure, we apply FISTA to solve the x subproblem, which can match the convergence rate of FCSA. Let $f(x) = \frac{1}{2} \|Rx - b\|_2^2 + \frac{\lambda}{2} \|z - G\Phi x\|_2^2$, which is a convex and smooth function with Lipschitz constant L_f , and $g(x) = 0$. Then our algorithm can be summarized in [Algorithm 7.8](#), which is called FISTA_Tree. Here $\nabla f(r^k) = R^T(Rr^k - b) + \lambda \Phi^T G^T(G\Phi r^k - z)$. R^T and Φ^T denote the inverse partial Fourier transform and the inverse wavelet transform.

ALGORITHM 7.8 FISTA_TREE

```

series Input:  $\rho = 1/L_f$ ,  $r^1 = x^0$ ,  $t^1 = 1, \beta, \lambda, N$ 
for  $k = 1$  series to  $N$  do
     $z = \text{shrinkgroup}(G\Phi x^{k-1}, \beta/\lambda)$ 
     $x^k = r^k - \rho \nabla f(r^k)$ 
     $r^{k+1} = [1 + \sqrt{1 + 4(r^k)^2}] / 2$ 
     $r^{k+1} = x^k + \frac{r^{k-1}}{r^{k+1}}(x^k - x^{k-1})$ 
end for

```

Computational complexity. Note that $G \in \mathbb{R}^{N' \times N}$ is a sparse matrix with each row containing only one nonzero element 1. Therefore the multiplication by G only cost $\mathcal{O}(N') = \mathcal{O}(N)$ with our group configuration. Suppose x is an image with N pixels. The *shrinkgroup* step can be implemented in only $\mathcal{O}(N \log N)$ time and the gradient step also takes $\mathcal{O}(N \log N)$. We can find that the total time complexity in each iteration is still $\mathcal{O}(N \log N)$, the same as that of TVCMRI, RecPF, and FCSA. This good feature guarantees that the proposed algorithm is comparable with the fastest MRI algorithms in terms of execution speed.

7.4.1.2 Constrained tree-based MRI

NESTA (Becker et al., 2011) solves the constrained problem of standard sparsity:

$$\min_{\theta} \|\theta\|_1, \quad \text{s.t.} \quad \|b - A\theta\|_2 \leq \epsilon, \quad (7.21)$$

where θ denotes the set of wavelet coefficients with $\theta = \Phi x$, $A = R\Phi^T$, Φ^T denotes the inverse wavelet transform, and ϵ is a small constant. It reaches the optimal convergence rate for first-order methods. Similar to the previous section, we extend it to solve the tree-based MRI problem:

$$\min_{\theta} \|G\theta\|_{2,1}, \quad \text{s.t.} \quad \|b - A\theta\|_2 \leq \epsilon, \quad (7.22)$$

where $\|G\theta\|_{2,1} = \sum_{g \in \mathcal{G}} \|(G\theta)_g\|_2$, and g, \mathcal{G} are the same as those in [Algorithm 7.8](#). Recall that the $\ell_{2,1}$ -norm also has the form:

$$\|G\theta\|_{2,1} = \max_{u \in \mathcal{Q}} \langle u, Gx \rangle, \quad (7.23)$$

where the dual feasible set is

$$\mathcal{Q} = \{u : \|u\|_{2,\infty} \leq 1\} = \left\{u : \max_{g \in \mathcal{G}} \|u_g\|_2 \leq 1\right\}. \quad (7.24)$$

We relax the nonsmooth $\ell_{2,1}$ -norm to a smooth function with:

$$f_\mu(\theta) = \max_{u \in \mathcal{Q}} \left(\langle u, G\theta \rangle - \frac{\mu}{2} \|u\|_2^2 \right), \quad (7.25)$$

where μ is a small fixed number.

Note that $(G\theta)_g = G_g\theta$, where G_g is the row of G , to group g . The first-order gradient of $f_\mu(\theta)$ with Lipschitz constant L_μ is given by

$$\nabla f_\mu(\theta)_g = \begin{cases} \mu^{-1} G_g^T G_g \theta, & \|G_g \theta\|_2 < \mu \\ G_g^T G_g \theta / \|G_g \theta\|_2, & \text{otherwise,} \end{cases} \quad (7.26)$$

NESTA assumes the rows of the sampling matrix A are orthogonal, that is, $AA^T = I$, where I denotes the identical matrix. Fortunately, the partial Fourier transform in compressed sensing MRI satisfies this assumption: $AA^T = R\Phi^T\Phi R^T = RR^T = I$, where R^T denotes the inverse operator of R . The whole algorithm based on the NESTA (Becker et al., 2011) framework is given in [Algorithm 7.9](#).

ALGORITHM 7.9 NESTA_TREE

```

series Input:  $\theta_0, \epsilon, k = 1, L_\mu, \mu$ 
while not meeting the stopping criterion do
    1. Compute  $\nabla f_\mu(\theta)$ 
    2. Compute  $y^k$ 
         $q = \theta^k - L_\mu^{-1} \nabla f_\mu(\theta)$ 
         $\lambda_\epsilon = \max(0, \epsilon^{-1} \|b - Aq\|_2 - L_\mu)$ 
         $y^k = (I - \frac{\lambda_\epsilon}{\lambda_\epsilon + L_\mu} A^T A)(\frac{\lambda_\epsilon}{L_\mu} A^T b + q)$ 
    3. Compute  $z^k$ 
         $\alpha^k = 1/2(k + 1)$ 
         $q = x_0 - L_\mu^{-1} \sum_{i \leq k} \nabla \alpha_i f_\mu(\theta)$ 
         $\lambda_\epsilon = \max(0, \epsilon^{-1} \|b - Aq\|_2 - L_\mu)$ 
         $z^k = (I - \frac{\lambda_\epsilon}{\lambda_\epsilon + L_\mu} A^T A)(\frac{\lambda_\epsilon}{L_\mu} A^T b + q)$ 
    4. Update  $\theta^k$ 
         $\tau^k = 2(k + 3)$ 
         $\theta^k = \tau^k z^k + (1 - \tau^k) y^k$ 
    5.  $k = k + 1$ 
end while

```

Computational complexity. As shown in [Algorithm 7.9](#), the complexity of the proposed algorithm is the same as the original NESTA algorithm (Becker et al., 2011). It is $6\mathcal{C} + \mathcal{O}(N)$, where \mathcal{C} denotes the complexity of applying A or A^T . In CSMRI, $\mathcal{C} = \mathcal{O}(N \log N)$ if FFT is applied. Therefore the total computational complexity is $\mathcal{O}(N \log N)$ for each iteration, the same as that of [Algorithm 7.8](#).

If we compare the two types of algorithms, the parameters can be manually set in the unconstrained algorithm to determine how sparse the data is. Or the weights between sparseness and the least squares fitting can be controlled. However, the constrained algorithm always seeks the sparsest solution that satisfies the constraint. In the application of MRI, we find that if good parameters can be tuned, the unconstrained algorithm ([Algorithm 7.8](#)) performs better, or vice versa. In contrast,

the constrained algorithm (Algorithm 7.9) has the convenience of not requiring tuning of the parameter.

7.4.2 EVALUATION

7.4.2.1 Experimental setup

We compare the unconstrained algorithm FISTA_Tree with CG (Lustig et al., 2007), TVCMRI (Ma et al., 2008), RecPF (Yang et al., 2010), and FCSA (Huang et al., 2011b), and compare the constrained algorithm NESTA_Tree with the original NESTA (Becker et al., 2011) algorithm for CSMRI. For fair comparisons, all code were downloaded from the authors' websites and we carefully followed their experimental setup. We applied all these methods on four real-valued MR images: cardiac, brain, chest, and shoulder (shown in Fig. 7.14). In addition, a complex-valued MR brain image <http://www.eecs.berkeley.edu/~mlustig/CS.html>. is added to validate the benefit of tree sparsity on complex-valued data. Suppose R is a partial Fourier transform with M rows and N columns. The sampling ratio is defined as M/N . For simulations with real-valued images, we follow the sampling strategy of previous works (Ma et al., 2008; Huang et al., 2011b), which randomly choose more Fourier coefficients at low frequency and less at high frequency. For complex-valued

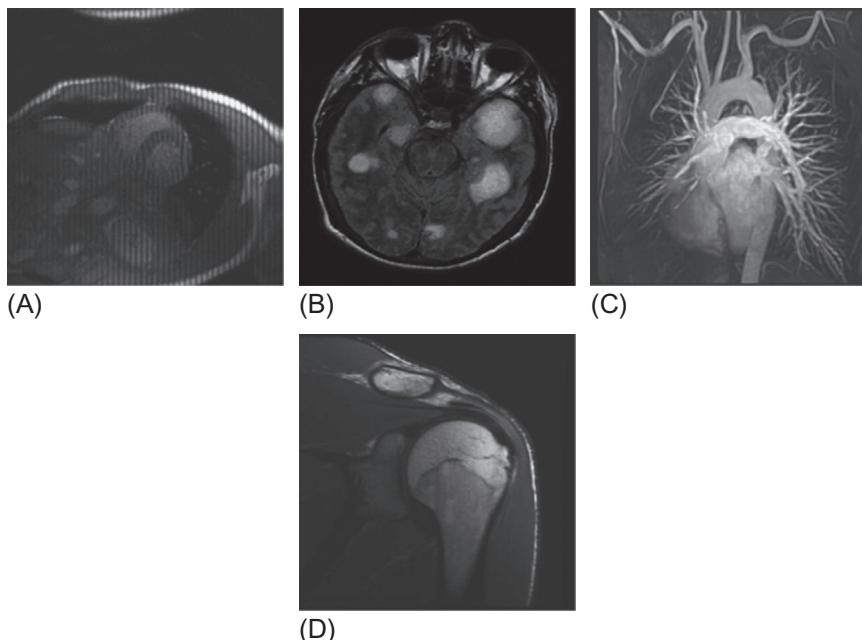


FIG. 7.14

MR images: (A) cardiac; (B) brain; (C) chest; (D) shoulder.

data, the radial sampling mask is used (Yang et al., 2010), which is more feasible in practice.

In order to study the benefit of the tree structure in CSMRI, we remove the TV term in all algorithms. The parameters for real-valued images and complex-valued images are tuned separately. There is no continuation step (Becker et al., 2011) in the NESTA_Tree algorithm. All experiments are on a desktop with 3.4-GHz Intel core i7 3770 CPU. Matlab version is 7.8 (2009a). Measurements are added by Gaussian white noise with 0.01 standard deviation. The SNR is used for result evaluation:

$$\text{SNR} = 10 \log_{10}(V_s/V_n), \quad (7.27)$$

where V_n is the mean square error between the original image x_0 and the solution x ; V_s denotes the variance of the values in x_0 .

7.4.2.2 Group configuration for tree sparsity

In all previous works, the tree structures are approximated as overlapping groups (Chen and Huang, 2012a; Rao et al., 2011). In addition, for each of them the wavelet coefficient and its parent are assigned into one group. However, the relationship between a coefficient and its grandparent is not exploited. We first conduct an experiment to validate the influence of the group size on the reconstruction result. Four group sizes are compared: (a) each group contains one coefficient, which is the same case as standard sparsity; (b) each group contains a coefficient and its parent, which is the same as previous works (Chen and Huang, 2012a; Rao et al., 2011); (c) each group contains a coefficient, its parent, and its grandparent; (d) each group contains four coefficients, where the grandparent's parent is also assigned in the same group.

With these group configurations, we test their performance on the FISTA _Tree algorithm, except the standard sparsity case is performed on FISTA. The parameter β determines how strong the tree sparsity assumption is. Tables 7.6 and 7.7 show the average SNRs and CPU time respectively for the four MR images with various parameter settings. With smaller parameters, the third group configuration performs the best, while the second group configuration is the best with larger parameters. The computational time increases monotonously as the size of the group becomes

Table 7.6 Comparisons of SNR (dB) for Different Group Sizes With Tree Sparsity

$\beta \setminus \text{Group Size}$	1	2	3	4
5×10^{-2}	17.30	17.94	16.45	15.33
10^{-2}	16.49	16.99	16.95	16.53
5×10^{-3}	16.36	16.62	16.66	16.48
10^{-3}	16.21	16.27	16.29	16.27

Table 7.7 Comparisons of Computational Cost (s) for Different Group Sizes With Tree Sparsity

$\beta \setminus \text{Group Size}$	1	2	3	4
5×10^{-2}	0.69	0.99	1.11	1.17
10^{-2}	0.70	0.95	1.09	1.15
5×10^{-3}	0.72	0.97	1.07	1.11
10^{-3}	0.70	0.97	1.07	1.11

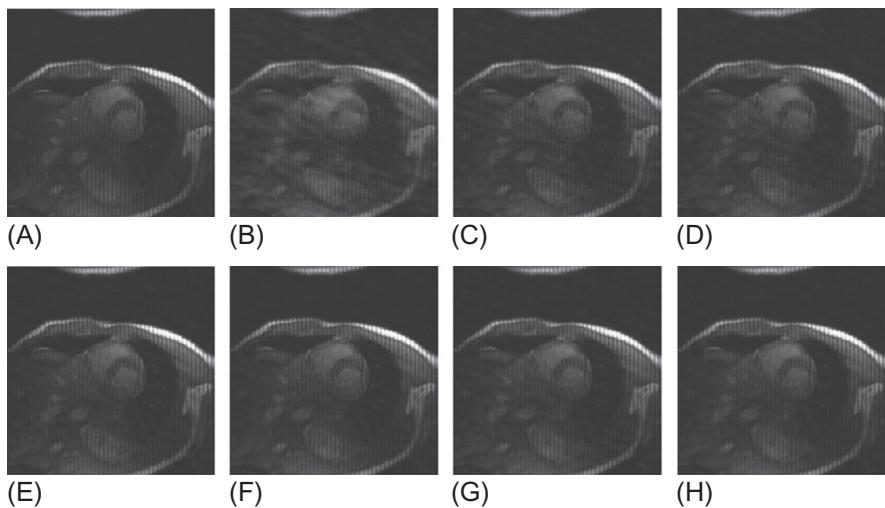
larger. Due to the above two reasons, we encourage the use of the second group configuration on CS-MRI.

7.4.2.3 Visual comparisons

We compare the proposed tree-based algorithms with the fastest MRI algorithms to validate how much the tree structure can improve existing results. To perform fair comparisons, all methods run 50 iterations except that the CG runs only eight iterations due to its higher computational complexity. Total variation terms are removed in all algorithms, as we only want to validate how much benefit the wavelet tree sparsity can bring compared to standard wavelet sparsity. In this case, FCSA (Huang et al., 2011b) is similar to FISTA (Beck and Teboulle, 2009a). Figs. 7.15–7.18 show the visual results for the four MR images with 20% sampling. It can be seen that the proposed unconstrained algorithm FISTA_Tree is always better than CG (Lustig et al., 2007), TVCMRI (Ma et al., 2008), RecPF (Yang et al., 2010), and FCSA (Huang et al., 2011b). These results are consistent with previous observations (Huang et al., 2011b). Compared to the proposed NESTA_Tree with NESTA, our method is still much better. These results are reasonable because no structured prior information has been exploited in previous algorithms other than sparsity, while the tree structure in our algorithms is utilized. Any coefficient that disobeys the tree structure will be penalized in our algorithms, which makes the results closer to the original ones.

7.4.2.4 SNRs and CPU time

Fig. 7.19 gives the performance comparisons between different methods in terms of SNR with 50 iterations. Due to the faster convergence rate of FISTA and NESTA, they always outperform CG, TVCMRI, and RecPF. Moreover, the tree-based algorithms approximated by overlapping group sparsity are always better than those with standard sparsity. Table 7.8 shows all computational costs for the different algorithms. CG has the highest computational complexity. TVCMRI and RecPF are much faster than CG and slower than FCSA. It is to be expected that the tree-based algorithms FISTA_Tree and NESTA_Tree are slower than FISTA and NESTA respectively, since the overlapping structure requires more time for computing than the nonoverlapping structure. However, applying the wavelet transform and the

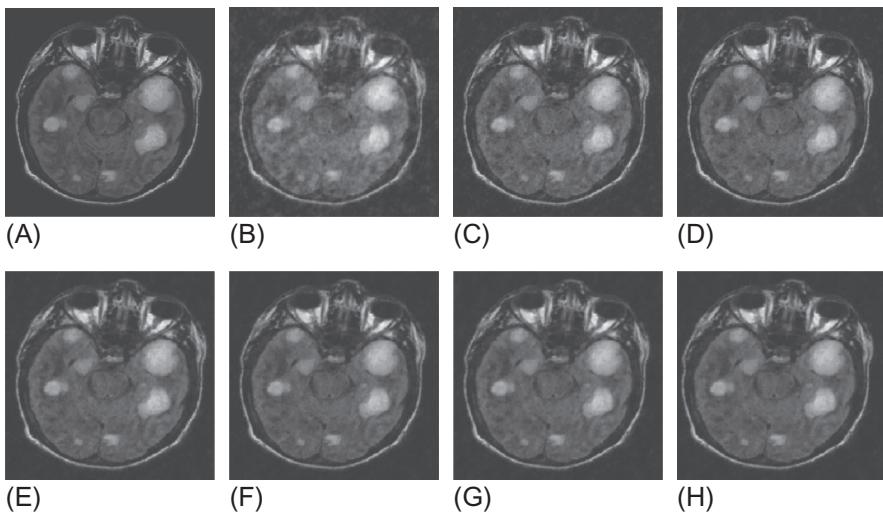
**FIG. 7.15**

Cardiac MR image reconstruction from 20% sampling. (A) The original image. Also shown are images recovered by: (B) CG ([Lustig et al., 2007](#)); (C) TVCMRI ([Ma et al., 2008](#)); (D) RecPF ([Yang et al., 2010](#)); (E) FCSA ([Huang et al., 2011b](#)); (F) FISTA_Tree; (G) NESTA ([Becker et al., 2011](#)); (H) NESTA_Tree. All algorithms are without total variation regularization. Their SNRs are 9.86, 14.70, 15.14, 17.31, 17.93, 16.31, and 16.96 respectively. Their computational time costs are 1.34, 1.12, 1.25, 0.67, 0.85, 0.88, and 1.05 s respectively.

Fourier transform is still the dominant cost, which is the same for all algorithms. As a result, FISTA_Tree and NESTA_Tree are comparable to the corresponding standard sparsity algorithms in terms of reconstruction speed, and provide greater improvement in terms of accuracy.

7.4.2.5 Sampling ratios

All algorithms are compared under different sampling ratios for the four MR images. Since we have shown that the CG method is far less efficient than the other methods, we do not include it in this experiment. To reduce the randomness, we run each experiments 100 times to obtain the average results for each method. The sampling ratio ranges from 17% to 25%. [Fig. 7.20](#) shows these results for the four images. We can observe that TVCMRI and RecPF are not comparable to recent algorithms with fast convergence rates. Under the same framework and with similar convergence rate, the tree-based algorithms (ie, FISTA_Tree and NESTA_Tree) are always better than the corresponding standard sparsity algorithms (ie, FCSA and NESTA). These results further demonstrate the benefit of tree sparsity in accelerated MRI.

**FIG. 7.16**

Brain MR image reconstruction from 20% sampling. (A) The original image. Also shown are images recovered by: (B) CG ([Lustig et al., 2007](#)); (C) TVCMRI ([Ma et al., 2008](#)); (D) RecPF ([Yang et al., 2010](#)); (E) FCSA ([Huang et al., 2011b](#)); (F) FISTA_Tree; (G) NESTA ([Becker et al., 2011](#)); (H) NESTA_Tree. None of the algorithms have total variation regularization. Their SNRs are 10.25, 13.81, 14.22, 15.65, 16.13, 15.05, and 15.52 respectively. Their computational time costs are 1.36, 1.11, 1.17, 0.71, 1.02, 0.91, and 1.03 s respectively.

7.4.2.6 Complex-valued image with radial sampling mask

We have observed superior performance for tree-based MRI algorithms from numerical simulations. In this section, we validate their performance on a complex-valued MR image with 512×512 pixels. The sampling mask is a radial mask, which is more feasible than the random sampling mask in practice. Here, we only compare the classical method CG ([Lustig et al., 2007](#)) and the fastest algorithm FCSA ([Huang et al., 2011b](#)) with the proposed tree-based algorithms.

[Fig. 7.21](#) presents the visual results reconstructed by difference methods. The image with full sampling is used as referent image. We can observe that the tree-based algorithms FISTA_Tree and NESTA_Tree achieve higher SNRs than the standard sparsity algorithms CG and FCSA. Due to the relatively slow convergence rate of CG, it has still not converged after 50 iterations. That is why it has inferior performance to FCSA. This data is scanned with noise. Therefore we also compare image quality besides SNR. From the zoomed-in areas, image details are lost in the image reconstructed by CG and blurred in that reconstructed by FCSA. However, both tree-based algorithms can preserve significant features on the MR image even with a low sampling ratio.

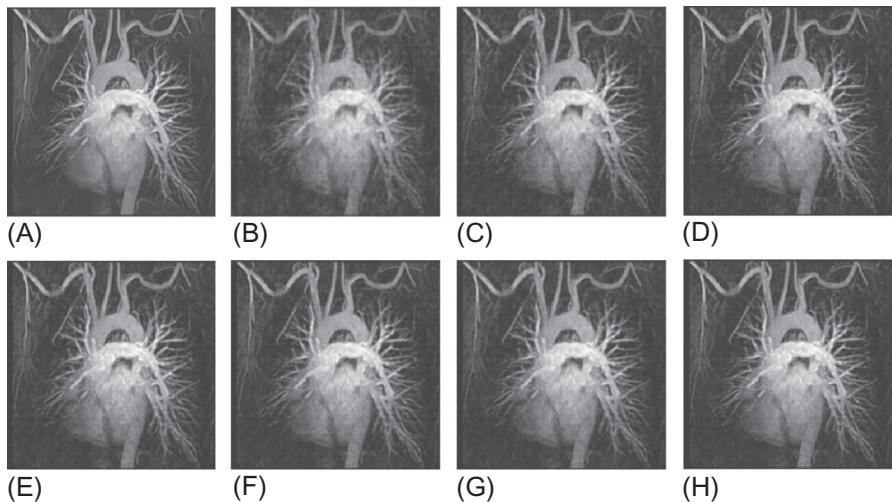


FIG. 7.17

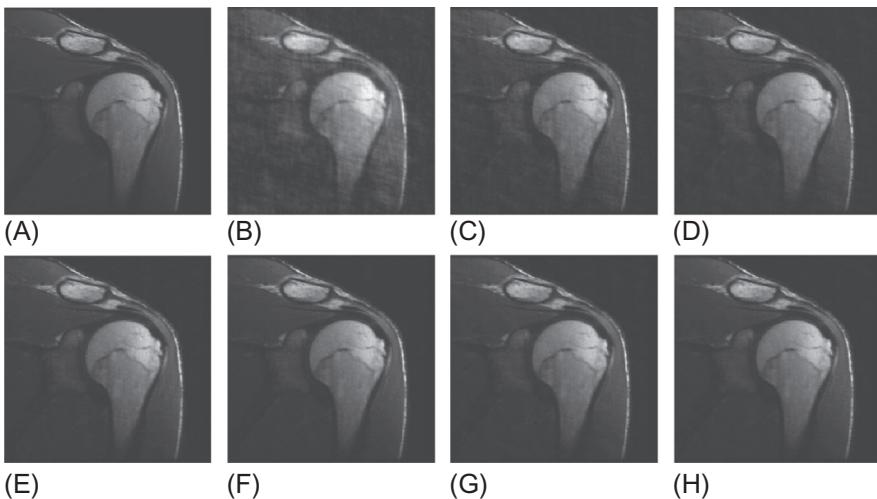
Chest MR image reconstruction from 20% sampling. (A) The original image. Also shown are images recovered by: (B) CG ([Lustig et al., 2007](#)); (C) TVCMRI ([Ma et al., 2008](#)); (D) RecPF ([Yang et al., 2010](#)); (E) FCSA ([Huang et al., 2011b](#)); (F) FISTA_Tree; (G) NESTA ([Becker et al., 2011](#)); (H) NESTA_Tree. None of the algorithms have total variation regularization. Their SNRs are 11.82, 15.09, 15.36, 15.98, 16.35, 15.91, and 16.30 respectively. Their computational time costs are 1.28, 1.12, 1.23, 0.67, 0.96, 0.84, and 1.07 s respectively.

7.4.3 SUMMARY

We discussed two tree-sparsity-based algorithms for CS-MRI and compared them with the state-of-the-art algorithms based on standard sparsity. In order to observe the benefit of tree sparsity more clearly, total variation terms were removed in all algorithms. Evaluation results demonstrated the practical improvement of the tree-sparsity-based algorithm on MR images. The results show that the benefit of the presented algorithm is greater than predicted by structured sparsity theory. That is because the tree structure is not adhered to as strictly as the structured sparsity theories assumed for practical data.

7.5 FOREST SPARSITY IN MULTICHANNEL CS-MRI

This section discusses the application of forest sparsity ([Chen et al., 2014a](#)) in CS-MRI. In practical applications, it is usually observed that multichannel images, such as color images, multispectral images and MR images, have joint sparsity and tree sparsity simultaneously. Therefore the support of such data consists of several connected trees and is like a forest. [Fig. 7.22](#) shows the forest structure in

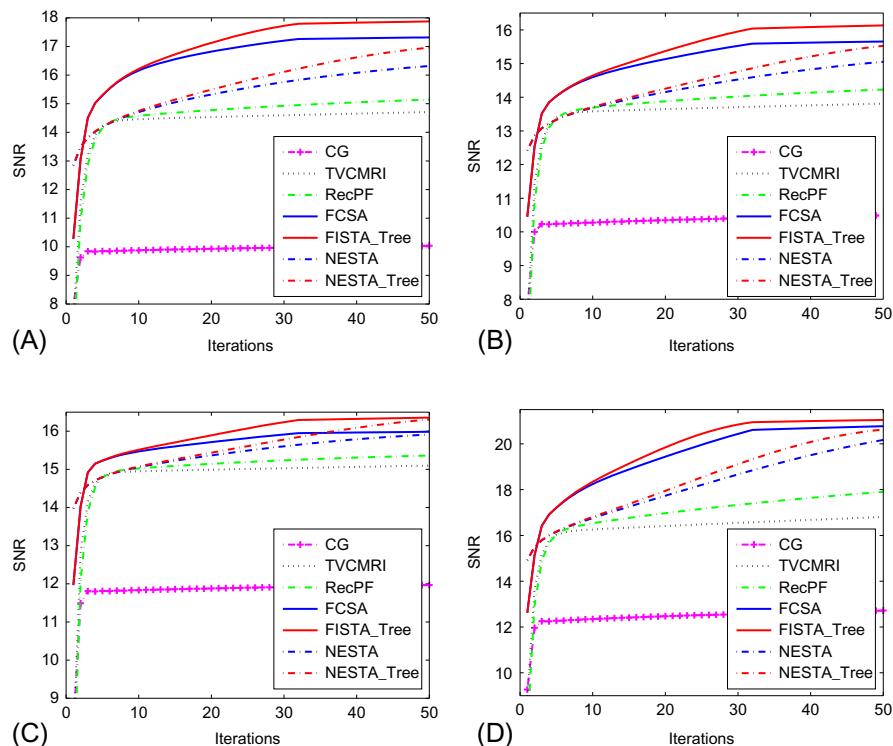
**FIG. 7.18**

Shoulder MR image reconstruction from 20% sampling. (A) The original image. Also shown are images recovered by: (B) CG ([Lustig et al., 2007](#)); (C) TVCMRI ([Ma et al., 2008](#)); (D) RecPF ([Yang et al., 2010](#)); (E) FCSA ([Huang et al., 2011b](#)); (F) FISTA_Tree; (G) NESTA ([Becker et al., 2011](#)); (H) NESTA_Tree. None of the algorithms have total variation regularization. Their SNR are 12.31, 16.80, 17.90, 20.77, 21.04, 20.17, and 20.62 respectively. Their computational time costs are 1.36, 1.07, 1.25, 0.67, 0.95, 0.82, and 1.07 s respectively.

multicontrast MR images. We can see that the nonzero coefficients are not randomly distributed but form a connected forest. Therefore forest sparsity is a favorable model for these kinds of problems.

7.5.1 MODEL AND ALGORITHM

In this section, the forest structure is approximated as overlapping group sparsity ([Jacob et al., 2009](#)) with mixed $\ell_{2,1}$ -norm. Although it may not be the best approximation, it is enough to demonstrate the benefit of forest sparsity. To evaluate the forest sparsity model, we need to compare different models via a similar framework. From the definition of forest-sparse data, we can find that if a coefficient is large/small, its parent and “neighbors” Parent denotes the parent node on the same channel while neighbors mean coefficients at the same position on other channels. also tend to be large/small. All parent-child pairs in the same position across different channels are assigned into one group, and the problem becomes overlapping group sparsity regularization. A similar scheme has been used in approximating tree sparsity ([Chen and Huang, 2012b](#); [Rao et al., 2011](#)), where each node and its parent are assigned

**FIG. 7.19**

Performance comparisons (SNRs) for different MR images: (A) cardiac image; (B) brain image; (C) chest image; (D) shoulder image.

Table 7.8 Comparison of Average Computational Costs (s) for Different MR Images With 20% Sampling

	Cardiac	Brain	Chest	Shoulder
CG	9.14	8.87	9.21	9.17
TVCMRI	1.12	1.11	1.12	1.07
RecPF	1.25	1.17	1.23	1.25
FCSA	0.67	0.71	0.67	0.67
FISTA_Tree	0.85	1.02	0.96	0.95
NESTA	0.88	0.91	0.84	0.82
NESTA_Tree	1.05	1.03	1.07	1.07

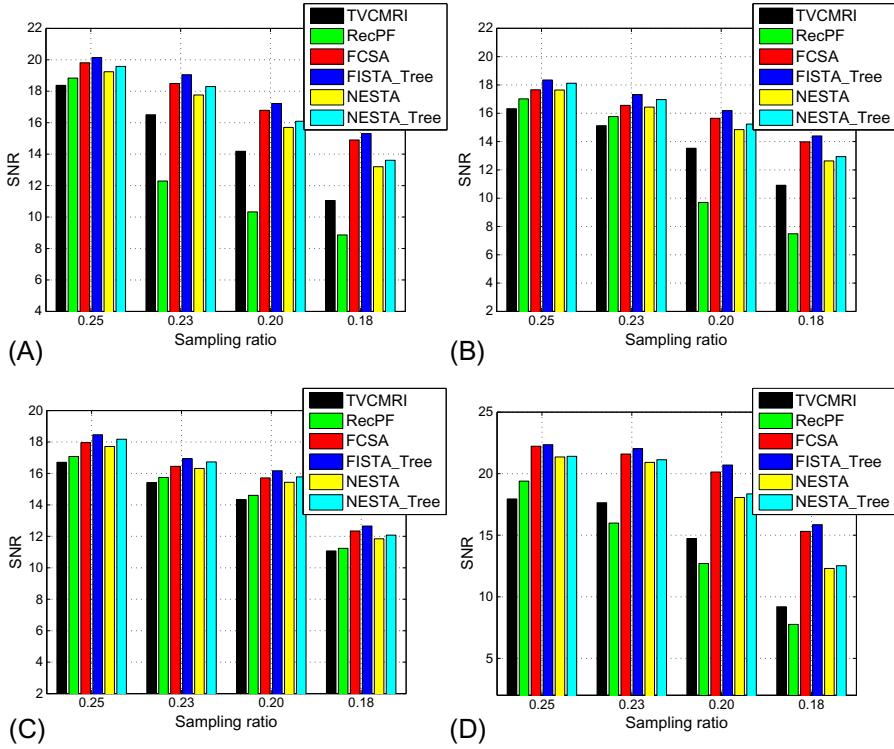


FIG. 7.20

Performance comparisons for the four MR images with different sampling ratios:
 (A) cardiac image; (B) brain image; (C) chest image; (D) shoulder image.

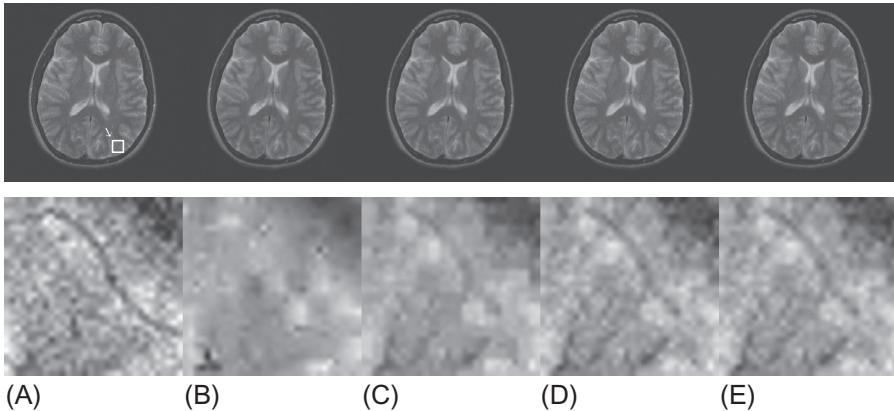
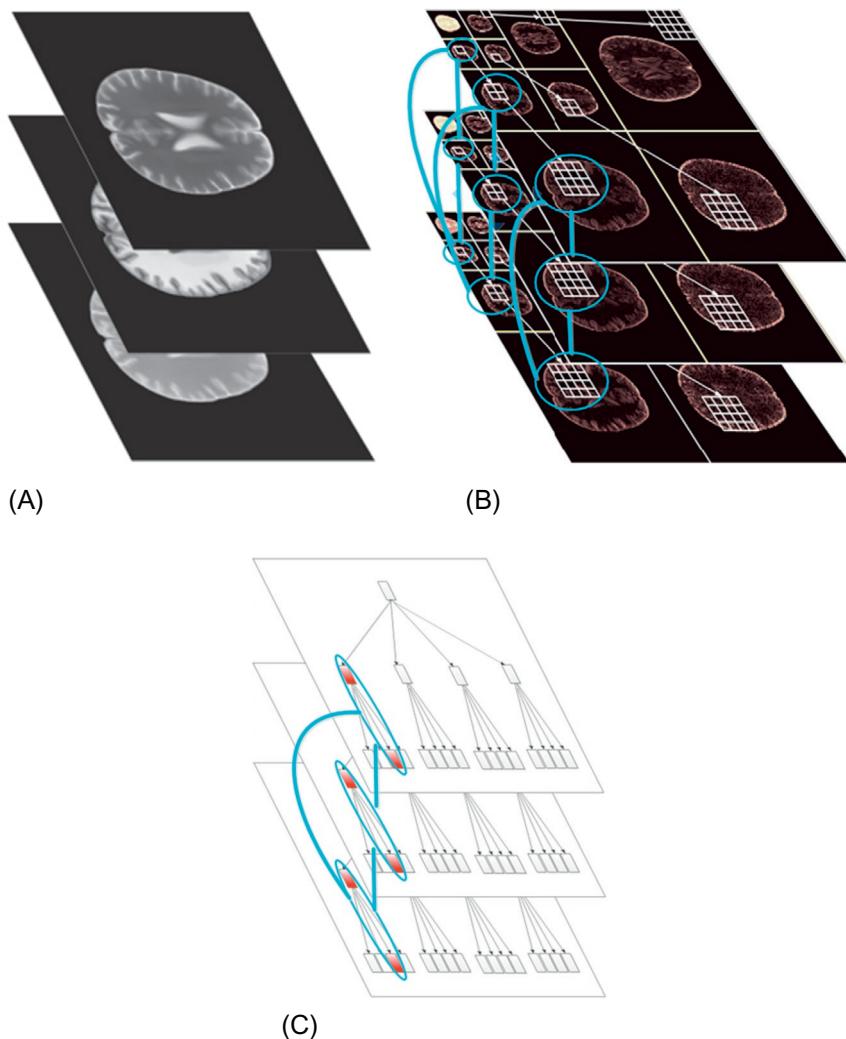


FIG. 7.21

Reconstruction on complex-valued MR image with 20% sampling. The first shows the visual results of by different algorithm. The second row shows the zoomed-in areas indicated by the white box. (A) Inverse FFT with full sampling. (B) CG. (C) FCSA. (D) FISTA_Tree. (E) NESTA_Tree. Their SNRs are 15.31, 16.32, 16.65, and 16.68, respectively.

**FIG. 7.22**

Forest structure on multicontrast MR images. (A) Three multicontrast MR images. (B) The wavelet coefficients of the images. Each coefficient tends to be consistent with its parent and children, and the coefficients across different trees at the same position. (C) One joint parent-child group across different trees that was used in our algorithm.

into one group. We write the approximated problem as

$$\min_x \frac{1}{2} \|Ax - b\|_2^2 + \lambda \sum_{g \in \mathcal{G}} \|(\Phi x)_g\|_2, \quad (7.28)$$

where g denotes one of the coefficient groups discussed above (an example is demonstrated in Fig. 7.22C), $(\cdot)_g$ denotes the coefficients in group g , and \mathcal{G} is the set of all groups.

The mixed $\ell_{2,1}$ -norm encourages all the components in the same group g to be zeros or nonzeros simultaneously. With our group configuration, this encourages forest sparsity. We present an efficient implementation based on the fast iterative shrinkage-thresholding algorithm (FISTA) (Beck and Teboulle, 2009b) framework for this problem. This is because FISTA can be easily applied for standard sparsity and joint sparsity, which makes the validation of the benefit of the proposed model more convenient. In addition, the formulation (7.28) can be easily extended to the combination of total variation (TV) via the fast composite splitting algorithm (FCSA) scheme (Huang et al., 2011b). Note that other algorithms may be used to solve the forest sparsity problems (eg, (Deng et al., 2011; Jacob et al., 2009; Kowalski et al., 2013)), but determining the optimal algorithm for forest sparsity is beyond the scope of this chapter.

FISTA (Beck and Teboulle, 2009b) is an accelerated version of the proximal method which minimizes the object function with the following form:

$$\min\{F(x) = f(x) + g(x)\}, \quad (7.29)$$

where $f(x)$ is a convex smooth function with Lipschitz constant L_f and $g(x)$ is a convex but usually nonsmooth function. It reverts to the original FISTA when $f(x) = \frac{1}{2}\|Ax - b\|_2^2$ and $g(x) = \lambda\|\Phi x\|_1$, which is summarized in Algorithm 7.10, where A^T denotes the transpose of A .

ALGORITHM 7.10 FISTA Beck and Teboulle (2009b)

```

series Input:  $\rho = 1/L_f$ ,  $\lambda$ ,  $n = 1$ ,  $t^1 = 1$ ,  $r^1 = x^0$ 
while not meeting the stopping criterion do
     $y = r^n - \rho A^T(Ar^n - b)$ 
     $x = \arg \min_x \left\{ \frac{1}{2\rho} \|x - y\|^2 + \lambda \|\Phi x\|_1 \right\}$ 
     $t^{n+1} = 1 + \sqrt{1 + 4(t^n)^2}/2$ 
     $r^{n+1} = x^n + \frac{t^n - 1}{t^{n+1}}(x^n - x^{n-1})$ 
     $n = n + 1$ 
end while

```

For the second step, there is a closed-form solution by soft-thresholding. For a joint sparsity problem where $g(x) = \lambda\|\Phi x\|_{2,1}$, the second step also has a closed-form solution. We call this version FISTA_Joint for joint sparsity. However, for the problem (7.28) with overlapping groups, we cannot directly apply FISTA to solve it.

In order to transfer the problem (7.28) to the nonoverlapping version, we introduce a binary matrix $G \in \mathbb{R}^{D \times TN}$ ($D > TN$) to duplicate the overlapping coefficients. Each row of G only contains one 1 and all others are 0s. The 1 appears in the i th column corresponds to the i th coefficient of Φx . Intuitively, if the coefficient is included in j groups, G will contain j such rows. An auxiliary variable z is used to constrain $G\Phi x$. This scheme is widely utilized in the alternating direction method (ADM) (Deng et al., 2011). The alternating formulation becomes:

$$\min_{x,z} \left\{ \frac{1}{2} \|Ax - b\|_2^2 + \lambda \sum_{g \in \mathcal{G}} \|z_g\|_2 + \frac{\gamma}{2} \|z - G\Phi x\|_2^2 \right\}, \quad (7.30)$$

where γ is another positive parameter. We iteratively solve this alternative formulation by minimizing x and z subproblems respectively. For the z subproblem:

$$\hat{z}_g = \arg \min_{z_g} \left\{ \lambda \|z_g\|_2 + \frac{\gamma}{2} \|z_g - (G\Phi x)_g\|_2^2 \right\}, \quad g \in \mathcal{G} \quad (7.31)$$

which has the closed-form solution:

$$\hat{z}_g = \max\left(\|(G\Phi x)_g\|_2 - \frac{\lambda}{\gamma}, 0\right) \frac{(G\Phi x)_g}{\|(G\Phi x)_g\|_2}, \quad g \in \mathcal{G}. \quad (7.32)$$

We denote it as a shrinkgroup operation. For the x -subproblem:

$$\hat{x} = \arg \min_x \left\{ \frac{1}{2} \|Ax - b\|_2^2 + \frac{\gamma}{2} \|z - G\Phi x\|_2^2 \right\} \quad (7.33)$$

The optimal solution is $x = (A^T A + \lambda \Phi^T G^T G \Phi)^{-1} (A^T b + \lambda \Phi^T G^T z)$, which contains a large-scale inverse problem. Actually, this problem can be efficiently solved by various methods. In order to compare with FISTA and FISTA_Joint, we apply FISTA to solve (7.33). This will demonstrate the benefit of forest sparsity more clearly. Let $f(x) = \frac{1}{2} \|Ax - b\|_2^2 + \frac{\lambda}{2} \|z - G\Phi x\|_2^2$ and $g(x) = 0$. Supposing its Lipschitz constant is L_f , the whole algorithm is summarized in Algorithm 7.11.

ALGORITHM 7.11 FISTA_FOREST

```

series Input:  $\rho = 1/L_f$ ,  $r^1 = x^0$ ,  $t^1 = 1$ ,  $\lambda, \gamma, n = 1$ 
while not meeting the stopping criterion do
     $z = \text{shrinkgroup}(G\Phi x^{n-1}, \lambda/\gamma)$ 
     $x^n = r^n - \rho[A^T(Ar^n - b) + \gamma \Phi^T G^T(G\Phi r^n - z)]$ 
     $t^{n+1} = [1 + \sqrt{1 + 4(t^n)^2}]/2$ 
     $r^{n+1} = x^n + \frac{t^n - 1}{t^{n+1}}(x^n - x^{n-1})$ 
     $n = n + 1$ 
end while

```

For the first step, we solve (7.31) while $\frac{1}{2} \|Ax - b\|_2^2$ stays the same. The object function value in (7.30) decreases. For the second step, (7.33) is solved by FISTA

iteratively while $\lambda \sum_{g \in \mathcal{G}} \|z_g\|_2$ stays the same. Therefore the object function value in (7.30) decreases in each iteration and the algorithm is convergent. [Algorithm 7.11](#) is also used to implement tree sparsity by recovering the data channel by channel separately. We call it FISTA_Tree.

In some practical applications, the data tends to be forest-sparse but not strictly so. We can soften and complement the forest assumption with other penalties, such as joint $\ell_{2,1}$ -norm or TV. For example, after combining TV, problem (7.30) becomes:

$$\min_{x,z} \left\{ \frac{1}{2} \|Ax - b\|_2^2 + \lambda \sum_{g \in \mathcal{G}} \|z_g\|_2 + \frac{\gamma}{2} \|z - G\Phi x\|_2^2 + \mu \|x\|_{TV} \right\}, \quad (7.34)$$

where $\|x\|_{TV} = \sum_{i=1}^{TN} \sqrt{(\nabla_1 x_i)^2 + (\nabla_2 x_i)^2}$; ∇_1 and ∇_2 denote the forward finite difference operators on the first and second coordinates respectively; μ is a positive parameter. Compared with [Algorithm 7.11](#), we only need to set $g(x) = \mu \|x\|_{TV}$ and the corresponding subproblem has already been solved ([Beck and Teboulle, 2009b](#); [Huang et al., 2011a,b](#)). This TV combined algorithm is called FCSA_Forest, which will be used in the experiments. To avoid repetition, it is not listed.

7.5.2 EVALUATION

We conduct experiments on RGB color images, multicontrast MR images, and MR images of multichannel coils to validate the benefit of forest sparsity. All experiments are conducted on a desktop with 3.4-GHz Intel core i7 3770 CPU. Matlab version is 7.8 (2009a). If the sampling matrix A is M by N , the sampling ratio is defined as M/N . All measurements are mixed with Gaussian white noise of 0.01 standard deviation. The SNR is used as the metric for evaluations:

$$SNR = 10 \log_{10}(V_s/V_n) \quad (7.35)$$

where V_n is the Mean Square Error between the original data x_0 and the reconstructed x ; $V_s = var(x_0)$ denotes the power level of the original data where $var(x_0)$ denotes the variance of the values in x_0 .

7.5.2.1 Multicontrast MRI

Multicontrast MRI is a popular technique to aid clinical diagnosis. For example T1 weighted MR images can distinguish fat from water, with water appearing darker and fat brighter. In T2 weighted images fat is darker and water is lighter, which is better suited to imaging edema. Although having different intensities, T1/T2 or proton-density weighted MR images are scanned at the same anatomical position. Therefore they are not independent but highly correlated. Multicontrast MR images are typically forest-sparse under the wavelet basis. Suppose $\{x_t\}_{t=1}^T \in \mathbb{R}^N$

are the multicontrast images for the same anatomical cross-section and $\{b_t\}_{t=1}^T$ are the corresponding undersampled data in the Fourier domain, the forest-sparse reconstruction can be formulated as

$$\hat{x} = \arg \min_x \|\Phi x\|_{\mathcal{F}, T} + \lambda \sum_{s=1}^T \|R_t x_t - b_t\|^2, \quad (7.36)$$

where x is the vertorized data of $[x_1, \dots, x_T]$ and R_t is the measurement matrix for the image x_t . This is an extension of conventional CS-MRI (Lustig et al., 2007). Fig. 7.22 shows an example of the forest structure in multicontrast MR images.

The data is extracted from the SRI24 Multichannel Brain Atlas Dataset (Rohlfing et al., 2009). In the Fourier domain, we randomly obtain more samples at low frequencies and fewer samples at higher frequencies. This sampling scheme has been widely used for CS-MRI (Huang et al., 2011b; Lustig et al., 2007; Ma et al., 2008). Fig. 7.23 shows the original multicontrast MR images and the sampling mask.

We compare four algorithms on this dataset: FISTA, FISTA_Joint, FISTA_Tree, and FISTA_Forest. The parameter λ is set to 0.035 and γ is set to 0.5λ . We run each algorithm for 400 iterations. Fig. 7.24A demonstrates the performance comparisons among different algorithms. From the figure, we can observe that modeling with forest sparsity achieves the highest SNR after convergence. Although the algorithm for forest sparsity takes more time due to the overlapping structure, it always outperforms all other methods in terms of accuracy.

In addition, as total variation is very popular in CS-MRI (Huang et al., 2012, 2011b; Lustig et al., 2007), we compare our FCSA_Forest algorithm with FCSA (Huang et al., 2011b) (TV is combined in FISTA), FCSA_Joint (Huang et al., 2012) (TV is combined in FISTA_Joint), and FCSA_Tree. The parameter μ for TV is set to 0.001, the same as in previous works (Huang et al., 2011b; Ma et al., 2008). Fig. 7.24B demonstrates the performance comparison including TV regularization. Compared with Fig. 7.24A, all algorithms improve at different degrees. However, the ranking does not change, which validates the superiority of forest sparsity. As FCSA has been proved to be better than other algorithms for general compressed

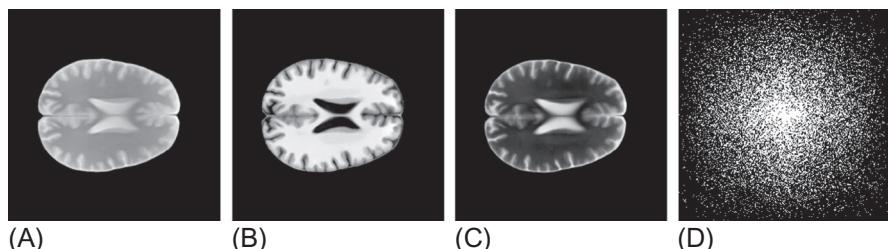
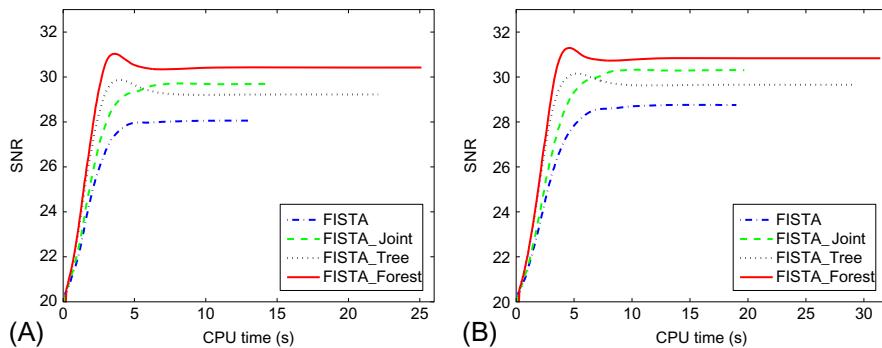


FIG. 7.23

(A)–(C) The original multicontrast images. (D) The sampling mask.

**FIG. 7.24**

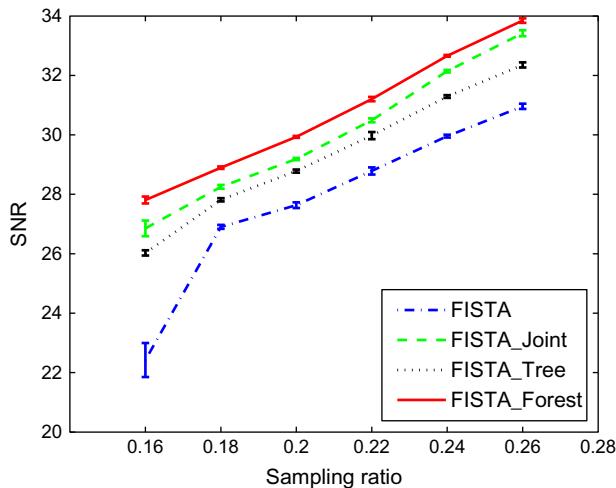
Performance comparisons among different algorithms. (A) Multic平st MRI reconstruction with 20% sampling. Their final SNRs are 28.05, 29.69, 29.22, and 30.42 respectively. The time costs are 13.11, 14.43, 22.08, and 25.11 s respectively. (B) Multic平st MRI images reconstruction with 20% sampling by both wavelet sparsity and TV regularization. Their final SNRs are 28.75, 30.30, 29.65, and 30.83 respectively. The time costs are 19.00, 19.68, 29.11, and 31.41 s, respectively.

sensing MRI (CS-MRI) (Lustig et al., 2007; Ma et al., 2008; Yang et al., 2010) and FCSA_Joint (Huang et al., 2012) is better (Bilgic et al., 2011; Majumdar and Ward, 2011) in multic平st MRI, the proposed method further improves CS-MRI and make it more feasible than before.

In order to validate the benefit of forest sparsity in terms of measurement number, we conducted an experiment to reconstruct multic平st MR images from different sampling ratios. Fig. 7.25 demonstrates the final results of four algorithms with sampling ratio from 16% to 26%. With more sampling, all algorithms show better performance. However, the forest sparsity algorithm always achieves the best reconstruction. For the same reconstruction accuracy, the FISTA_Forest algorithm only requires about 16% measurements to achieve SNR 28, which is approximately 2%, 3%, and 5% less than those of FISTA_Joint, FISTA_Tree, and FISTA respectively. More results of forest sparsity on multic平st MRI can be found in Chen and Huang (2014b).

7.5.2.2 Parallel MRI

To improve the scanning speed of MRI, an efficient and feasible way is to acquire the data in parallel with multichannel coils. The scanning time depends on the number of measurements in the Fourier domain, and it will be significantly reduced when each coil only acquires a small fraction of the whole measurements. The bottleneck is how to reconstruct the original MR image efficiently and precisely. This is called parallel MRI (pMRI) in the literature. Sparsity techniques have been used to improve the classical method SENSE (Pruessmann et al., 1999). However, when the coil

**FIG. 7.25**

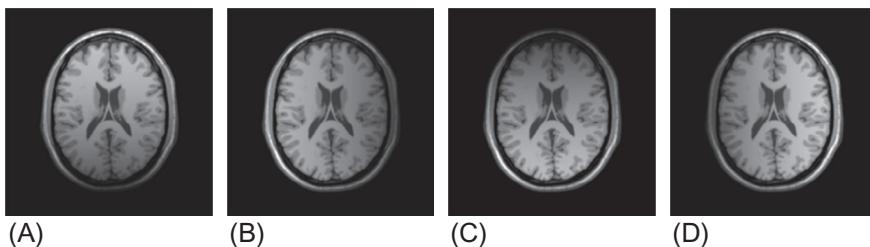
Reconstruction performance with different sampling ratios.

sensitivity cannot be estimated precisely, the final image will contain visual artifacts. Unlike the previous CS-SENSE (Liang et al., 2009), which reconstructs the images of multicoils individually, calibrationless parallel MRI (Chen et al., 2013; Majumdar and Ward, 2012) recovers the aliased images of all coils jointly by assuming the data is jointly sparse.

Let T equal the number of coils and b_t be the measurement vector from coil t . It is therefore the same CS problem as (7.36). The final result of CaLM-MRI is obtained by a sum-of-square (SoS) approach without coil sensitivity and SENSE encoding. It shows comparable results with those methods which need precise coil configuration. As shown in Fig. 7.26, the appearances of different images obtained from multicoils are very similar. This method can be improved with forest sparsity, since the images follow the forest sparsity assumption.

There are two steps for compressed sensing pMRI reconstruction in CaLM-MRI (Majumdar and Ward, 2012): (1) the aliased images are recovered from the undersampled Fourier signals of different coil channels by CS methods; (2) the final image for clinical diagnosis is synthesized by the recovered aliased images using the SoS approach. As discussed above, these aliased images should be forest-sparse under the wavelet basis. We compare our algorithm with FISTA_Joint and SPGL1 (Van Den Berg and Friedlander, 2008), which solves the joint $\ell_{2,1}$ -norm problem in CaLM-MRI. For the second step, all methods use the SoS approach from the aliased images that they recovered. All algorithms run for sufficient time until it has converged.

Tables 7.9 and 7.10 show all the comprehensive comparisons among these algorithms. For the same algorithm, more measurements or a greater number of

**FIG. 7.26**

The aliased MR images of multicoils. Due to the different locations of the coils, they have different sensitivities to the same image.

coils tend to increase the SNRs of aliased images, although this does not result in linear improvement for the final image reconstruction. Another observation is that FISTA_Joint and SPGL1 show similar performance in terms of SNR on this data. This is because both solve the same joint sparsity problem, even with different schemes. Upgrading the model to forest sparsity, significant improvement can be gained. Finally, it is unknown how to combine TV in SPGL1. However, both

Table 7.9 Comparison of SNRs (dB) at Different Sampling Ratios With Four Coils

	Sampling Ratios	25%	20%	17%	15%
SNR of aliased images	SPGL1	26.72	24.59	23.08	22.31
	FISTA_Joint	26.95	24.73	23.06	22.21
	FISTA_Forest	27.47	25.22	23.37	22.59
SNR of final image	SPGL1	20.64	20.35	19.12	18.64
	FISTA_Joint	20.79	20.41	19.75	18.49
	FISTA_Forest	22.62	22.29	21.03	20.47

Table 7.10 Comparison of SNRs (dB) for Different Number of Coils With 20% Sampling Ratio

	Number of Coils	2	4	6	8
SNR of aliased images	SPGL1	23.33	24.61	24.74	25.16
	FISTA_Joint	23.41	24.71	24.89	25.23
	FISTA_Forest	24.25	25.12	25.29	25.52
SNR of final image	SPGL1	21.76	18.95	21.05	21.32
	FISTA_Joint	21.90	18.94	21.15	21.87
	FISTA_Forest	22.44	22.22	22.52	22.52

FISTA_Joint and FISTA_Forest can easily combine TV, which can further enhance the results (Huang et al., 2012).

7.5.3 SUMMARY

In this section, we have discussed the *forest sparsity* model for sparse learning and compressed sensing. This model enriches the family of structured sparsity and can be widely applied to numerous fields of sparse regularization problems. The benefit of the forest sparsity model has been theoretically proved and empirically validated in practical applications. Under compressed sensing assumptions, significant reduction of measurements is achieved with forest sparsity compared with standard sparsity, joint sparsity or independent tree sparsity. We also discuss a fast algorithm for efficiently solving the forest sparsity problem.

7.6 CONCLUSION

In this chapter, we have presented several sparsity models for MRI reconstruction. A set of efficient algorithms is discussed to solve sparsity regularization MRI reconstruction problems. First, we introduce standard sparsity for MR image reconstruction. Second, we discussed an efficient algorithm for fast multicontrast MRI reconstruction based on group sparsity constraints. Third, we introduce wavelet tree sparsity in accelerated MRI. Finally, we discuss forest sparsity on multichannel data, which is a natural extension of tree sparsity.

REFERENCES

- Bach, F., 2008. Consistency of the group lasso and multiple kernel learning. *J. Mach. Learn. Res.* 9, 1179–1225.
- Baraniuk, R., Cevher, V., Duarte, M., Hegde, C., 2010. Model-based compressive sensing. *IEEE Trans. Inform. Theory* 56 (4), 1982–2001.
- Baron, D., Wakin, M., Duarte, M., Sarvotham, S., Baraniuk, R., 2005. Distributed compressed sensing. *arXiv* 0901.3403
- Beck, A., Teboulle, M., 2009a. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Trans. Image Process.* 18 (113), 2419–2434.
- Beck, A., Teboulle, M., 2009b. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* 2 (1), 183–202.
- Becker, S., Bobin, J., Candès, E.J., 2011. Nesta: a fast and accurate first-order method for sparse recovery. *SIAM J. Imaging Sci.* 4 (1), 1–39.
- Bilgic, B., Goyal, V., Adalsteinsson, E., 2011. Multi-contrast reconstruction with Bayesian compressed sensing. *Magn. Reson. Med.* 66 (6), 1601–1615.
- Block, K.T., Uecker, M., Frahm, J., 2007. Undersampled radial MRI with multiple coils. iterative image reconstruction using a total variation constraint. *Magn. Reson. Med.* 57 (6), 1086–1098.

- Blomgren, P., Chan, T.F., 1998. Color TV: total variation methods for restoration of vector-valued images. *IEEE Trans. Image Process.* 7 (3), 304–309.
- Candès, E., 2006. Compressive sampling. In: *Proceedings of the International Congress of Mathematicians*, pp. 1433–1452.
- Candès, E., Romberg, J., 2007. Sparsity and incoherence in compressive sampling. *Inverse Problems* 23 (3), 969.
- Candès, E., Romberg, J., Tao, T., 2006a. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory* 52 (2), 489–509.
- Candès, E., Romberg, J., Tao, T., 2006b. Stable signal recovery from incomplete and inaccurate measurements. *Commun. Pure Appl. Math.* 59 (8), 1207–1223.
- Candès, E.J., Romberg, J., Tao, T., 2006c. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory* 52, 489–509.
- Cauley, S.F., Xi, Y., Bilgic, B., Setsompop, K., Xia, J., Adalsteinsson, E., Balakrishnan, V.R., Wald, L.L., 2013. Scalable and accurate variance estimation (save) for joint Bayesian compressed sensing. In: *21st Annual Meeting of ISMRM*, p. 2603.
- Chang, T.-C., He, L., Fang, T., 2006. MR image reconstruction from sparse radial samples using Bregman iteration. In: *Proceedings of the 13th Annual Meeting of ISMRM*, Seattle, p. 696.
- Chen, C., Huang, J., 2012a. The benefit of tree sparsity in accelerated MRI. In: *MICCAI Workshop on Sparsity Techniques in Medical Imaging*. IEEE, Piscataway, NJ.
- Chen, C., Huang, J., 2012b. Compressive sensing MRI with wavelet tree sparsity. In: *Proceedings of the Annual Conference on Advances in Neural Information Processing Systems (NIPS)*, pp. 1124–1132.
- Chen, C., Huang, J., 2014a. The benefit of tree sparsity in accelerated MRI. *Med. Image Anal.* 18 (6), 834–842.
- Chen, C., Huang, J., 2014b. Exploiting both intra-quadtree and inter-spatial structures for multi-contrast MRI. In: *Proceedings of the International Symposium on Biomedical Imaging (ISBI)*.
- Chen, S., Donoho, D., Saunders, M., 1998. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.* 20 (1), 33–61.
- Chen, C., Li, Y., Huang, J., 2013. Calibrationless parallel MRI with joint total variation regularization. In: *Proceedings of the Annual International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 106–114.
- Chen, C., Li, Y., Huang, J., 2014a. Forest sparsity for multi-channel compressive sensing. *IEEE Trans. Signal Process.* 62 (11), 2803–2813.
- Chen, C., Li, Y., Liu, W., Huang, J., 2014b. Image fusion with local spectral consistency and dynamic gradient sparsity. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2760–2765.
- Combettes, P.L., 2009. Iterative construction of the resolvent of a sum of maximal monotone operators. *J. Convex Anal.* 16, 727–748.
- Combettes, P.L., Pesquet, J.C., 2008. A proximal decomposition method for solving convex variational inverse problems. *Inverse Problems* 24, 1–27.
- Cotter, S., Rao, B., Engan, K., Kreutz-Delgado, K., 2005. Sparse solutions to linear inverse problems with multiple measurement vectors. *IEEE Trans. Signal Process.* 53 (7), 2477–2488.

- Deng, W., Yin, W., Zhang, Y., 2011. Group sparse optimization by alternating direction method, TR11-06. Department of Computational and Applied Mathematics, Rice University.
- Donoho, D., 2006. Compressed sensing. *IEEE Trans. Inform. Theory* 52 (4), 1289–1306.
- Donoho, D., Elad, M., 2002. Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization. *Optimally* 100 (5), 2197–2202.
- Donoho, D., Maleki, A., Montanari, A., 2009. Message-passing algorithms for compressed sensing, *Proc. Natl. Acad. Sci.* 106, (45), 18914–18919.
- Figueiredo, M., Nowak, R., Wright, S., 2007. Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems. *IEEE J. Sel. Top. Signal Process.* 1 (4), 586–597.
- He, L., Carin, L., 2009. Exploiting structure in wavelet-based Bayesian compressive sensing. *IEEE Trans. Signal Process.* 57 (9), 3488–3497.
- He, L., Chang, T.C., Osher, S., Fang, T., Speier, P., 2006. MR image reconstruction by using the iterative refinement method and nonlinear inverse scale space methods. Technical Report UCLA CAM 06-35. <ftp://ftp.math.ucla.edu/pub/camreport/cam06-35.pdf>.
- Huang, J., 2011. Structured sparsity: theorems, algorithms and applications. Ph.D. thesis, Rutgers University.
- Huang, J., Huang, X., Metaxas, D., 2009. Learning with dynamic group sparsity. In: Proceedings of IEEE International Conference on Computer Vision (ICCV), pp. 64–71.
- Huang, J., Zhang, T., et al., 2010. The benefit of group sparsity. *Ann. Stat.* 38 (4), 1978–2004.
- Huang, J., Zhang, S., Li, H., Metaxas, D., 2011a. Composite splitting algorithms for convex optimization. *Comput. Vis. Image Understand.* 115 (12), 1610–1622.
- Huang, J., Zhang, S., Metaxas, D., 2011b. Efficient MR image reconstruction for compressed MR imaging. *Med. Image Anal.* 15 (5), 670–679.
- Huang, J., Zhang, T., Metaxas, D., 2011c. Learning with structured sparsity. *J. Mach. Learn. Res.* 12, 3371–3412.
- Huang, J., Chen, C., Axel, L., 2012. Fast Multi-contrast MRI Reconstruction. In: Proceedings of Medical Image Computing and Computer-Assisted Intervention, pp. 281–288.
- Jacob, L., Obozinski, G., Vert, J., 2009. Group lasso with overlap and graph lasso. In: Proceedings of the International Conference on Machine Learning (ICML), pp. 433–440.
- Ji, S., Xue, Y., Carin, L., 2008. Bayesian compressive sensing. *IEEE Trans. Signal Process.* 56 (6), 2346–2356.
- Ji, S., Dunson, D., Carin, L., 2009. Multitask compressive sensing. *IEEE Trans. Signal Process.* 57 (1), 92–106.
- Kim, S., Xing, E., 2012. Tree-guided group lasso for multi-response regression with structured sparsity, with an application to eqtl mapping. *Ann. Appl. Stat.* 6 (3), 1095–1117.
- Koh, K., Kim, S., Boyd, S., 2007. An interior-point method for large-scale l_1 -regularized logistic regression. *J. Mach. Learn. Res.* 8 (8), 1519–1555.
- Kowalski, M., Siedenburg, K., Dörfler, M., 2013. Social sparsity! Neighborhood systems enrich structured shrinkage operators. *IEEE Trans. Signal Process.* 61 (10), 2498–2511.
- Krim, H., Viberg, M., 1996. Two decades of array signal processing research: the parametric approach. *IEEE Signal Process. Mag.* 13 (4), 67–94.
- La, C., Do, M., 2006. Tree-based orthogonal matching pursuit algorithm for signal reconstruction. In: Proceedings of the IEEE International Conference on Image Processing (ICIP), pp. 1277–1280.
- Liang, D., Liu, B., Wang, J., Ying, L., 2009. Accelerating sense using compressed sensing. *Magn. Reson. Med.* 62 (6), 1574–1584.

- Lustig, M., Donoho, D., Pauly, J., 2007. Sparse MRI: the application of compressed sensing for rapid MR imaging. *Magn. Reson. Med.* 58, 1182–1195.
- Ma, S., Yin, W., Zhang, Y., Chakraborty, A., 2008. An efficient algorithm for compressed MR imaging using total variation and wavelets. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–8.
- Majumdar, A., Ward, R., 2010. Compressive color imaging with group-sparsity on analysis prior. In: Proceedings of the 17th IEEE International Conference on Image Processing (ICIP), pp. 1337–1340.
- Majumdar, A., Ward, R., 2011. Joint reconstruction of multiecho MR images using correlated sparsity. *Magn. Reson. Imaging* 29 (7), 899–906.
- Majumdar, A., Ward, R., 2012. Calibration-less multi-coil MR image reconstruction. *Magn. Reson. Imaging*.
- Manduca, A., 1996. Wavelet compression of medical images with set partitioning in hierarchical trees. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, vol. 3, pp. 1224–1225.
- Meng, J., Yin, W., Li, H., Hossain, E., Han, Z., 2011. Collaborative spectrum sensing from sparse observations in cognitive radio networks. *IEEE J. Sel. Areas Commun.* 29 (2), 327–337.
- Natarajan, B., 1995. Sparse approximate solutions to linear systems. *SIAM J. Imaging Sci.* 24 (2), 227–234.
- Needell, D., Tropp, J., 2009. CoSaMP: iterative signal recovery from incomplete and inaccurate samples. *Appl. Comput. Harmon. Anal.* 26 (3), 301–321.
- Nesterov, Y., 1983. A method for unconstrained convex minimization problem with the rate of convergence $\mathcal{O}(1/k^2)$. *Sov. Math. Dokl.* 269, 543–547.
- Pruessmann, K., Weiger, M., Scheidegger, M., Boesiger, P., et al., 1999. SENSE: sensitivity encoding for fast MRI. *Magn. Reson. Med.* 42 (5), 952–962.
- Rao, N., Nowak, R., Wright, S., Kingsbury, N., 2011. Convex approaches to model wavelet sparsity patterns. In: Proceedings of the IEEE International Conference on Image Processing (ICIP), pp. 1917–1920.
- Rohlfing, T., Zahr, N., Sullivan, E., Pfefferbaum, A., 2009. The SRI24 multichannel atlas of normal adult human brain structure. *Hum. Brain Map.* 31 (5), 798–819.
- Saito, T., Takagaki, Y., Komatsu, T., 2011. Three kinds of color total-variation semi-norms and its application to color-image denoising. In: Proceedings of the IEEE International Conference on Image Processing (ICIP), pp. 1457–1460.
- Som, S., Schniter, P., 2012. Compressive imaging using approximate message passing and a Markov-tree prior. *IEEE Trans. Signal Process.* 60 (7), 3439–3448.
- Sprechmann, P., Ramirez, I., Sapiro, G., Eldar, Y.C., 2011. C-HiLasso: a collaborative hierarchical sparse modeling framework. *IEEE Trans. Signal Process.* 59 (9), 4183–4198.
- Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. B* 267–288.
- Tropp, J., 2004. Greed is good: algorithmic results for sparse approximation. *IEEE Trans. Inform. Theory* 50 (10), 2231–2242.
- Van Den Berg, E., Friedlander, M., 2008. Probing the Pareto frontier for basis pursuit solutions. *SIAM J. Imaging Sci.* 31 (2), 890–912.
- Weller, D., Polimeni, J., Grady, L., Wald, L., Adalsteinsson, E., Goyal, V., 2010. Combining nonconvex compressed sensing and grappa using the nullspace method. In: 18th Annual Meeting of ISMRM, p. 4880.

- Wipf, D., Rao, B., 2007. An empirical Bayesian strategy for solving the simultaneous sparse approximation problem. *IEEE Trans. Signal Process.* 55 (7), 3704–3716.
- 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.
- Ye, J., Tak, S., Han, Y., Park, H., 2007. Projection reconstruction MR imaging using FOCUSS. *Magn. Reson. Med.* 57 (4), 764–775.
- Yuan, M., Lin, Y., 2005. Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. B* 68 (1), 49–67.
- Ziniel, J., Schniter, P., 2011. Efficient high-dimensional inference in the multiple measurement vector problem. *arXiv* 1111.5272.