MR Image Reconstruction via Sparse Representation: Modeling and Algorithm

Xiaojing Ye[1], **Yunmei Chen**[1], **and Feng Huang**[2]
[1]Department of Mathematics, University of Florida, Gainesville, FL 32611, USA
[2]Advanced Concept Development, Invivo Corporation, Gainesville, FL 32608, USA

Abstract— *To reduce acquisition time in magnetic resonance (MR) imaging, compressive sensing and sparse representation techniques have been developed to reconstruct MR images with partially acquired data. Although this has been a hot research topic in the field, it has not been used clinically due to three inherent problems of its current framework: potential loss of fine structures, difficulty to predefine model parameters, and long reconstruction time. The aim of this work is to tackle these problems. We propose to minimize the total variation of the underlying image, together with the ℓ_1 norm of the coefficients in its representation using a trained dictionary, as well as a fidelity term. Using a trained dictionary can take the advantage of prior knowledge and hence improve accuracy in reconstruction. Our data fidelity constraint is derived from the likelihood estimator of the recovering error in partial k-space to improve the robustness of the model to parameter selection. Moreover, a simple and efficient numerical scheme is provided to solve this model faster. The consequent experiments on both synthetic and in vivo data indicate the improvement of the proposed model in preserving fine structure, reducing computational cost, and flexibility of parameter decision.*

Keywords: compressed sensing, dictionary, sparse representation, convex optimization, MRI reconstruction.

1. Introduction

Magnetic resonance imaging (MRI) is a technique that allows visualization of structures and functions of a body by non-invasive means. However, MRI takes much longer acquisition time than some other imaging modalities, which limits the application of MRI. Many advanced techniques have been developed to reduce acquisition time of MRI. One category of these techniques is reducing the amount of acquired data, and then, applying appropriate reconstruction algorithm to recover high quality images. The compressive sensing (CS) technique [1], [2], [3], [4], [5], [6], [7], [8], [9], [10] is one of the several effective methods to do this. The CS method requires the to-be-reconstructed image having a sparse representation in certain transform domain. The reconstruction schemes enforce both sparsity of the image representation and consistency with the acquired data. For complex medical images, the sparse representation is not necessarily in image domain, but could be in certain transform domains. This means that the image itself may not be sparse in terms of its pixel intensities, but it may have a sparse representation in terms of spatial finite differences, or its wavelet coefficients, or other transforms. In [11], a total variation (TV) based model was proposed for MR image reconstruction with partially acquired k-space data. This model indeed requires a sparsity in terms of spatial finite differences, which is good for piecewise homogeneous images [12]. For images with inhomogeneous intensity and fine structures, TV based model does not work well when the undersampling ratio is high. To overcome the shortcomings of only using TV regularization, Lustig et al. [2] proposed to minimize the TV norm of the underlying image together with the ℓ_1 norm of the wavelet transformed image, subject to data fidelity constraint:

$$\min_u \mu TV(u) + \|\Psi^\top u\|_1, \quad \text{s.t. } \|\mathcal{F}_p u - f_p\|_2 < \sigma, \quad (1)$$

where Ψ represents the wavelet transform, the superscript $^\top$ is the conjugate transpose of a matrix, $\mathcal{F}_p$ denotes the partial Fourier operator, f_p is the partially scanned k-space data and σ estimates the noise level during data sampling. Of note $TV(u)$ can be chosen as either $\|Du\|_1 = |D_x u| + |D_y u|$ or $\|Du\|_2 = \sqrt{|D_x u|^2 + |D_y u|^2}$. Besides the TV and wavelet transform, where the basis for image representation is defined globally on the entire image domain, there are many other transforms to sparsely represent various types of images [13]. So called dictionary is one of them, which has been shown effective in limited data reconstruction [10]. A redundant dictionary, that contains prototype signal-atoms, can represent a signal/image by sparse linear combinations of these atoms. The idea of using dictionary to sparsely represent an image is considering small overlapping image patches, and assuming each image patch has a sparse representation with respect to a basis (i.e. signal-atoms), that can be known or learned or trained using prior information. Since dictionary deals with small image patches, and can be learned or trained, it is more adequate for the recovery of fine structures.

Moreover, a number of numerical algorithms have been developed to solve CS models (e.g. see [3], [4], [5], [6], [7], [8], [9], [2], [14]). Especially, in [14], Yang et al. introduced a fast algorithm to solve the unconstrained version of model

(1)

$$\min_u \mu\|Du\|_2 + \|\Psi^\top u\|_1 + \frac{\lambda}{2}\|\mathcal{F}_p u - f_p\|_2^2. \qquad (2)$$

They provided an alternating minimization scheme, in which the clever use of the property of wavelet transform matrix, and diagonalizing the gradient operator using Fourier transform made the main computation of the algorithm involve only shrinkage operator and fast Fourier transform.

Since 2005, using CS technique to reconstruct MR image with partially acquired k-space (frequency domain) data has been a hot research topic in MRI field and achieved great progress [2], [15], [16]. However, it has not been used clinically due to three inherent problems of its current framework: potential loss of fine structures, difficulty to predefine reconstruction parameters, and long reconstruction time. In this work, we present a novel variational framework to tackle these problems. Our effort focuses on the following three strategies. First, we propose to use TV norm together with an overcomplete dictionary trained from prior information of to-be-reconstructed image for sparse representation. In many medical applications, it is possible to exploit some specific prior information of the to-be-reconstructed images, such as previous scans and scans for other time frames, and to use them for training an overcomplete dictionary [17], [18]. In this work, we use K-SVD algorithm with a database consisting of patches of high quality brain MR images to train a dictionary for brain MR image reconstruction. K-SVD is an iterative method that alternates between sparse coding of the examples based on the current dictionary and a process of updating the dictionary atoms to better fit the given database. The output is a trained dictionary that can represent all signals in the database under strict sparsity constraints and error tolerance [19], [20]. The trained dictionary we obtained by K-SVD algorithm is shown in Fig 1. A comparison of the sparse representation of a brain image using wavelet transform and the trained dictionary are shown in Fig 2. In both cases the images are reconstructed by the linear combination of the basis, where only the largest 12.5% transform coefficients are used. The image reconstructed using dictionary have higher signal to noise ratio (SNR) than the one by wavelet, while have comparably well preserved edges and fine structures. The reason is that the trained dictionary absorbed prior knowledge by learning features of the same type of images. This also demonstrates the adaptivity of using trained dictionaries to sparsely represent image patches. Although dictionary has the advantages as mentioned above, it takes more computational time than wavelet transform. Therefore the development of efficient algorithms involving the use of a dictionary is very important. The fast algorithm developed in [14] cannot be directly applied to the models using dictionary, since, in general, a dictionary can not be formulated as an orthogonal matrix like the wavelet transform. In this study we found that if each pixel in the image is covered by the same number of patches,

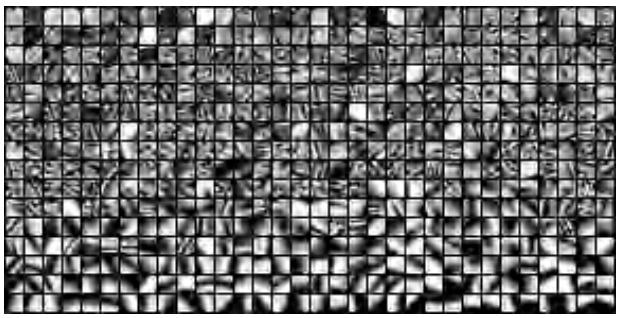

Fig. 1: Dictionary trained by K-SVD algorithm. The database used for training consists of patches of high quality MR brain images but excludes the to-be-construct image. Each block represents an atom of size 8×8. Atoms are sorted by ascending their standard deviations.

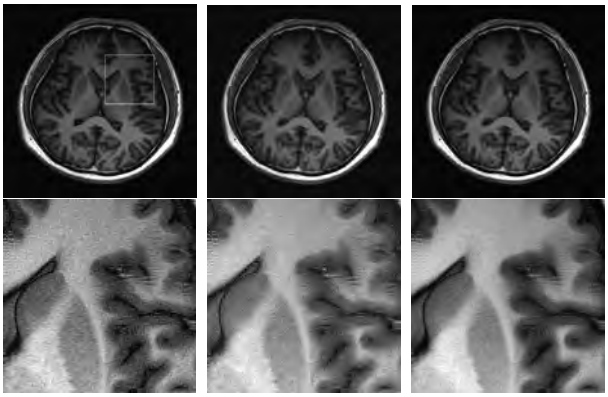

Fig. 2: Compare the accuracy of sparse representations by wavelet and trained dictionary. In both cases images are represented by picking up the largest 12.5% transform coefficients. Bottom images are corresponding zoomed-in square area shown on the top left image. Left column: original image. Middle column: representation by wavelet. Right column: representation by trained dictionary in Fig 1.

then we have $\sum_j R_j^\top R_j = mI$, where R_j is a binary matrix that extracts the j-th patch of the image u, all the pathes $\{R_j u\}_{j=1}^J$ together cover the entire image and might be overlapped, $m \in \mathbb{N}$ counts the number of the patches covering each pixel, and I denotes the identity matrix. Here we would like to point out that selecting image patches such that each pixel has the same number of covering is not difficult. For instance, the periodic/symmetric boundary condition of the image can make the patch construction straightforward (see section 3 below). Based on this observation and several techniques developed in [14] we provide a simple and fast numerical algorithm, which can be applied to CS models involving dictionary.

The third effort is proposing a fidelity constraint derived from maximum likelihood estimator (MLE) for the error between the partially scanned k-space data and partial Fourier-

transformed-reconstruction. In this way the ratio between the consistency with acquired data and sparse representation of the reconstruction is not a prefixed parameter λ, but λ divided by the sample variance of the reconstruction error in partial k-space, which varies during iterations. Therefore, the choice for λ is more flexible. Moreover, when the accuracy of the reconstruction gets better, the weight on the fidelity term increases, and hence, the accuracy is more enhanced.

The rest of this paper is organized as follows. A description of the proposed model is given in the next section. In section 3 a fast algorithm to solve the proposed model is introduced. Experimental results are presented in section 4. Then we conclude this paper in the last section.

2. Proposed Model

Let $u \in \mathbb{R}^N$ be the objective image consisting of N pixels, and $\mathcal{F}$ be the Fourier transform, which is in fact an $N \times N$ unitary matrix. Let $f_p \in \mathbb{C}^M$ with $M < N$ be the partially scanned data, and $P \in \mathbb{R}^{M\times N}$ be the binary matrix representing the sampling pattern. Thus $\mathcal{F}_p = P\mathcal{F}$ is the partial Fourier transform operator. For short notation, we use $\|\cdot\|$ to represent ℓ_2 norm $\|\cdot\|_2$, and $(\cdot;\cdots;\cdot)$ to stack components in column.

One difficulty of solving unconstrained energy minimization problem as (2), instead of the constrained one (1), is in determining the parameter that balances the consistency of reconstructed image with the acquired data and sparsity of the image representation. To tackle this problem our consistent measure (i.e. fidelity term) is derived from MLE approach. Let $\zeta = (\zeta_1, \cdots, \zeta_M)^\top \in \mathbb{C}^M$ be the reconstruction error, i.e. the difference between Fourier transform of the reconstruction u and partially scanned data f_p. Thus,

$$f_p = P\mathcal{F}u + \zeta.$$

Consider ζ_i $(h = 1, \ldots, M)$ as independent random variables indexed by h. Since we wish each ζ_h to be small, our fidelity constraint is designed to force all the ζ_h's obeying a normal distribution of mean zero and variance σ^2 to be determined. Namely, the probability density function (p.d.f.) of each ζ_h is $p(z|\sigma) = (2\pi\sigma^2)^{-1/2}e^{-z^2/2\sigma^2}$. Then, the joint density, or alternatively, the likelihood function is given as

$$\begin{aligned}\mathscr{L}(\sigma|\zeta) &= p(\zeta|\sigma) = \prod_{h=1}^{M}\left(\frac{1}{\sqrt{2\pi}\sigma}e^{-\zeta_h^2/2\sigma^2}\right)\\ &= \left(2\pi\sigma^2\right)^{-M/2}e^{-\|\zeta\|^2/2\sigma^2}.\end{aligned}$$

Thus, negative log-likelihood is

$$-\log\mathscr{L}(\sigma|\zeta) = \|\zeta\|^2/2\sigma^2 + M\log\sqrt{2\pi}\sigma. \qquad (3)$$

Substituting ζ by $\mathcal{F}_p u - f_p$, and omitting constant $M\log\sqrt{2\pi}$, we obtain a MLE based consistency estimation with the partially acquired data:

$$F(u,\sigma,f_p) = \|\mathcal{F}_p u - f_p\|^2/2\sigma^2 + M\log\sigma. \qquad (4)$$

This is a generalization of the least square estimation, which is just the case when $\sigma = 1$. We will use (4) as fidelity term in our energy functional. One advantage of using this form is that in the construction of u the weight on the ℓ^2 norm of ζ is not a prefixed parameter, but that divided by σ^2, which can be automatically updated during iterations. In the Euler-Lagrange (EL) equations associated with the proposed energy function below, one can see that σ is the standard deviation of ζ. Hence, when the construction error ζ decreases, the weight on minimizing ℓ_2 norm of ζ increases. This self-adjusting weight process makes faster convergence faster and better accuracy in reconstruction.

Next we show how to represent an image using a overcomplete dictionary. Let $R_j \in \mathbb{R}^{n\times N}$ be the binary matrix that extracts the j-th patch of image u, $j = 1, \cdots, J$. The union of these pathes $\{R_j u\}_{j=1}^J$, that might be overlapped, covers the entire image u. Let $\mathcal{T}$ denote the trained dictionary mentioned in the previous section with K atoms. The dictionary $\mathcal{T}$ can be explicitly written as a matrix of size $n \times K$, with each column representing one atom. Note that the sizes of atom and image patch are compatible, which ensures that any patch can be represented as a linear combination of some atoms.

As shown earlier, provided this trained dictionary $\mathcal{T}$, we are able to sparsely represent all image patches $R_j u$. In other words, for any j, there exists a *sparse* representation of $R_j u$:

$$\|\alpha_j\|_0 << n < K \quad \text{s.t. } \mathcal{T}\alpha_j = R_j u,$$

where the ℓ_0 norm $\|\alpha_j\|_0$ counts the number of nonzero elements in α_j, and is significantly less than n. As n linearly independent atoms can *exactly* represent any patches but we here want a *sparse* representation using few atoms from an overcomplete dictionary, whose atoms must be linearly dependent since $n < K$. However, minimizing the nonconvex ℓ_0 is generally a NP-hard problem and hence is not tractable in practice. A common substitute is to use ℓ_1 norm. And it has been proved that minimizing ℓ_1 leads to the same solution of ℓ_0 under certain conditions [1], [15].

Now we present our model. We propose to sparsely represent image in terms of a trained dictionary and its spatial finite differences. Moreover we use (4) as the fidelity constraint. Then, our model is formulated as an energy minimization problem:

$$\begin{aligned}\min_{u,\alpha,\sigma}\ \mu\|Du\| + \sum_{j=1}^{J}\left(\|\alpha_j\|_1 + \frac{\nu}{2}\|\mathcal{T}\alpha_j - R_j u\|^2\right) \qquad (5)\\ +\lambda F(u,\sigma,f_p),\end{aligned}$$

and $\alpha = (\alpha_1;\alpha_2;\cdots;\alpha_J) \in \mathbb{R}^{KJ}$. The first term in (5) is the TV norm of u in the form $\|Du\| = \sqrt{|D_x u|^2 + |D_y u|^2}$. The second term is for a sparse representation of u using trained dictionary $\mathcal{T}$. For each patch $R_j u$, it is expected to use the linear combination of a few atoms in $\mathcal{T}$ to match the patch. The last term is from (4).

3. Algorithm

There have been a number of efficient algorithms developed for dealing with the difficulties related to the TV and ℓ_1 terms in (5) [11], [21], [22], [23]. The algorithm provided in this section is inspired by the work in [14], [24], where a classical quadratic penalty method [25] is used to make the algorithm fast and efficient.

3.1 Derivation of Fast Algorithm for Proposed Model

We first introduce two auxiliary variables $w = (w_1^\top; w_2^\top; \cdots; w_N^\top) \in \mathbb{R}^{N\times 2}$ and $\beta = (\beta_1; \beta_2; \cdots; \beta_J) \in \mathbb{R}^{KJ}$ where $w_i \in \mathbb{R}^2$ and $\beta_j \in \mathbb{R}^K$ for all $i = 1, \cdots, N$ and $j = 1, \cdots, J$. The minimization problem (5) is equivalent to

$$\begin{aligned}\min_{u,w,\alpha,\beta,\sigma} \mu\sum_{i=1}^{N}\|w_i\| + \sum_{j=1}^{J}\left(\|\beta_j\|_1 + \frac{\nu}{2}\|\mathcal{T}\alpha_j - R_j u\|^2\right) \quad (6)\\ +\lambda F(u,\sigma,f_p)\\ \text{s.t. } w_i = D_i u,\ \beta_j = \alpha_j,\ \forall\, i = 1,\cdots,N, j = 1,\cdots,J.\end{aligned}$$

where $D_i u \in \mathbb{R}^2$ represents the gradient of u at the i-th pixel. Converting the equality constraints to quadratic penalties, we obtain an approximation of (6):

$$\begin{aligned}\min_{u,w,\alpha,\beta,\sigma} \mu\sum_{i=1}^{N}\phi(w_i, D_i u) + \psi(\beta,\alpha) \quad (7)\\ +\sum_{j=1}^{J}\frac{\nu}{2}\|\mathcal{T}\alpha_j - R_j u\|^2 + \lambda F(u,\sigma,f_p)\end{aligned}$$

where functions ϕ and ψ are defined as

$$\phi(s,t) = \|s\| + \frac{\eta}{2}\|s-t\|^2, \quad s,t \in \mathbb{R}^2$$

and

$$\psi(s,t) = \|s\|_1 + \frac{\theta}{2}\|s-t\|^2, \quad s,t \in \mathbb{R}^{KJ}$$

for given $\eta, \theta > 0$. As η and θ gradually increase, the solution of (7) approximates to that of(5). Meanwhile, the minimization of (7) with respect to each variable is much easier: First, for fixed u and α, the minimization with respect to w and β can be carried out in parallel:

$$w_i = \mathcal{S}_2(D_i u),\ \forall\, i \quad (8)$$

where $\mathcal{S}_2(t)$ minimizes $\phi(s,t)$ for fixed t by two-dimensional shrinkage

$$\mathcal{S}_2(t) = \max\{\|t\| - 1/\eta, 0\}\cdot(t/\|t\|), \quad t \in \mathbb{R}^2$$

and

$$\beta = \mathcal{S}_c(\alpha) \quad (9)$$

where $\mathcal{S}_c(t)$ minimizes $\psi(s,t)$ for fixed t by componentwise shrinkage

$$\mathcal{S}_c(t) = \{s : s_i = \max\{|t_i| - 1/\theta, 0\}\cdot\mathrm{sign}(t_i)\}.$$

with assumption $0\cdot(0/0) = 0$. Both computational costs are linear in N with given patch size.

Secondly, for fixed u and β, we can minimize (7) with respect to $\alpha = (\alpha_1; \cdots; \alpha_J)$ by

$$\min_{\alpha}\sum_{j=1}^{J}\left(\theta\|\alpha_j - \beta_j\|^2 + \nu\|\mathcal{T}\alpha_j - R_j u\|^2\right). \quad (10)$$

The solution can be obtained by setting α_j as

$$\alpha_j = V(\theta I + \nu\Lambda)^{-1}V^\top\left(\theta\beta_j + \nu\mathcal{T}^\top R_j u\right) \quad (11)$$

where the spectral decomposition $\mathcal{T}^\top\mathcal{T} = V\Lambda V^\top$ is computationally inexpensive since the dictionary $\mathcal{T}$ is prepared before any experiments and its largest dimension K is usually much less than N. Again, all α_j's can be computed in parallel.

Thirdly, for fixed w, α and σ, the minimization of u is

$$\begin{aligned}\min_u \|w_x - D_x u\|^2 + \|w_y - D_y u\|^2 \quad (12)\\ +\sum_{j=1}^{J}\gamma\|\mathcal{T}\alpha_j - R_j u\|^2 + \xi\|\mathcal{F}_p u - f_p\|^2,\end{aligned}$$

where w_x and w_y are the first and second column of w, respectively, and $\gamma = \nu/\mu\eta$, $\xi = \xi(\sigma) = \lambda/\mu\eta\sigma^2$. Thus the normal equation of (12) becomes

$$Lu = r, \quad (13)$$

where

$$L = D_x^\top D_x + D_y^\top D_y + \sum_{j=1}^{J}\gamma R_j^\top R_j + \xi\mathcal{F}_p^\top\mathcal{F}_p$$

and

$$r = D_x^\top w_x + D_y^\top w_y + \sum_{j=1}^{J}\gamma R_j^\top\mathcal{T}\alpha_j + \xi\mathcal{F}_p^\top f_p.$$

Under the periodic boundary condition for u, the finite difference operators D_x and D_y are block circulant matrices with circulant blocks and hence can be diagonalized by Fourier transform $\mathcal{F}$. Thus, $\hat{D}_x = \mathcal{F}D_x\mathcal{F}^\top$ and $\hat{D}_y = \mathcal{F}D_y\mathcal{F}^\top$ are diagonal. Also, periodic boundary condition enables us to extract patches that cover each pixel m times, where $m = n/d^2$ and d is the sliding distance between all concatenated patches. Therefore $\sum_j R_j^\top R_j$, which is a diagonal matrix with i-th diagonal entry counting the number of times the i-th pixel covered by patches, is just mI. So multiplying $\mathcal{F}$ on both sides of (13) gives

$$\hat{L}\mathcal{F}(u) = \hat{r}, \quad (14)$$

where

$$\hat{L} = \hat{D}_x^\top\hat{D}_x + \hat{D}_y^\top\hat{D}_y + m\gamma I + \xi P^\top P$$

is a diagonal matrix since $P^\top P$ is diagonal, and

$$\hat{r} = \hat{D}_x^\top\mathcal{F}(w_x) + \hat{D}_y^\top\mathcal{F}(w_y) + \gamma\mathcal{F}(u_\alpha) + \xi P^\top f_p$$

where $u_\alpha = \sum_j R_j^\top \mathcal{T}\alpha_j$ is an "image" assembled using patches that are represented by dictionary $\mathcal{T}$ and α.

Finally, the computation of first variation of $F(u, \sigma, f_p)$ gives an update of σ in each iteration:

$$\sigma = \sqrt{\|\mathcal{F}_p u - f_p\|^2/M}. \tag{15}$$

3.2 Algorithm and Convergence Analysis

For σ in a large range $(0, 3(\|\mathcal{F}_p u_0 - f_p\|^2/M)^{1/2}]$ where u_0 is the original image, energy functional (7) is convex and coercive, and hence has a unique minimizer. Numerical proof of convergence is similar to [24] with little modifications and thus is omitted here. For stopping criterion, we let "res" be the maximum absolute/norm value of increments of w, α, β, u, and terminate each inner loop once $\text{res} < \epsilon$ for a predefined error tolerance ϵ, then update u and start a new loop with doubled η and θ. The upper bound 2^{12} for η, θ is chosen empirically so it is sufficiently large that solutions to (7) is a close approximation of (5). Based on derivations above, we summarize the algorithm for our model as Alg. 1 (**recMRI**).

Algorithm 1 MR Image Reconstruction via Sparse Representation (**recMRI**)

Input P, f_p, and $\mu, \nu, \lambda, \epsilon > 0$. Initialize $u = \mathcal{F}_p^\top f_p$, $\eta = \theta = 2^6$ and $\alpha = 0$.
while $\eta, \theta < 2^{12}$ **do**
  repeat
    Given u and α, compute w and β using (8) and (9).
    for $j = 1$ to J **do**
      Given u and β, compute α_j using (11).
    end for
    Given w and α, compute u by solving (14) and update σ by (15).
  until $\text{res} < \epsilon$
  return $u^{\eta,\theta}$
  $u \leftarrow u^{\eta,\theta}$, $(\eta, \theta) \leftarrow (2\eta, 2\theta)$
end while

4. Experimental Results

In this section, we present the experimental results of proposed model and comparisons with using wavelet transform on both phantom and in vivo MRI data. All implementations involved in the experiments were coded in Matlab v7.3 (R2006b), except the component-wise shrinkage and wavelet transform operators, which were coded in C++ and downloaded from Rice Wavelet Toolbox (http://www.dsp.rice.edu/software/rwt.shtml) with Haar wavelet transform and other settings as default. Computations were performed on a Linux (version 2.6.16) workstation with Intel Core 2 CPU at 1.86GHz and 2GB memory.

First, a Shepp-Logan phantom with size 256×256, as shown on the very left of Fig. 3, was used to simulate the synthetic MRI data. The full k-space data were produced by a 2-dimensional fast Fourier transform (**fft2** in Matlab) on this phantom image. Then we applied a radial sampling pattern with 22 radial lines, as shown next to the phantom in Fig. 3, and add a random Gaussian noise with mean zero and standard deviation 3.2 to simulate the partially acquired data for reconstruction. This is approximately 20% of standard deviation of high frequency signals. The third image in Fig. 3 is the reconstruction obtained by zero filling, which results in a serious artifacts aliasing. The last image in Fig. 3 is the reconstruction obtained by the proposed model (5), in which the dictionary was set to be an overcomplete Discrete Cosine Transform (DCT) consisting 256 atoms of size 8×8 [19], with parameters $(\mu, \nu, \lambda, \epsilon) = (1, 1, 10^3, 10^{-3})$ and no patch overlapping. The relative root of mean squared error (RMSE) of the reconstruction to the original image is 2.18%.

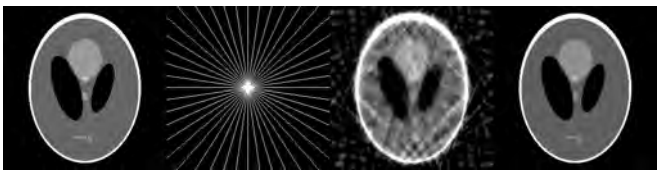

Fig. 3: Reconstruction of Shepp-Logan phantom using model (5). From left to right: Original image, sampling pattern with 22 radial lines (sample ratio 8.4%), reconstruction by zero-filling unscanned data, reconstruction by model (5) (RMSE=2.18%).

Table 1: Comparison of results of phantom reconstructions using nonlinear conjugate gradient (CG) with wavelet sparsity, recMRI with wavelet, and recMRI with dictionary. *Wavelet transformations are generated using optimized DWT packages for Matlab.

Method	**CG**(Wavelet*)		**recMRI**(Wavelet*)		**recMRI**(Dictionary)	
λ	RMSE	CPU	RMSE	CPU	RMSE	CPU
1e+2	5.93%	86.3	7.93%	28.2	5.21%	211
1e+3	2.47%	71.6	2.52%	27.7	2.18%	199
1e+4	5.05%	71.4	4.98%	26.9	3.47%	198
1e+5	25.9%	87.1	5.93%	27.0	3.67%	201
1e+6	37.0%	81.2	6.16%	28.7	5.52%	212

Table 1 shows the comparison of three methods on robustness, speed and accuracy. These three methods are: model (2) by conjugate gradient based numerical implementations, model (5) by the proposed numerical implementation with wavelet and DCT sparse transformation respectively. From the change of RMSE with λ, it can be clearly seen that the proposed model (5) is much less sensitive to the parameter than the conventional model (2). This demonstrates the robustness of the proposed method. From the consumed CPU time, it can be seen that the proposed numerical method was over 2.6 times faster than conjugate gradient based method. The DCT based sparse representation consistently produced images with lower RMSE than wavelet based method at the cost of longer reconstruction time.

A set of brain image acquired on a 3T GE system (GE healthcare, WI, USA) was used in the second experiment. The reference image reconstructed with full k-space is shown on the upper left of Fig. 4. A Cartesian random sampling pattern, shown under the brain image in Fig. 4, was used for artificial downsampling. The zoomed-in square of the brain image is shown on the upper middle image. With only 34.0% data for reconstruction, strong aliasing artifacts can be observed in the image reconstructed by zero-filling, whose zoomed-in is shown on the lower middle of Fig. 4. To produce the trained dictionary for the proposed model, K-SVD algorithm was applied on a database consisting high quality brain MR images (excluding the objective brain image), as shown in Fig. 1. In both models λ, μ and ϵ are set to be 2e+3, 1 and 5e-4, respectively. ν in model (5) is set to be 10^6. The zoomed-in area of reconstructed images by model (2) and proposed model (5) are shown on the right column of Fig. 4. It can be seen the image reconstructed by model (2) has oil-painting effect (due to the intrinsic smooth property of both wavelet and TV). On the contrary, the image reconstructed by the proposed model looks more natural and preserve fine structures much better. The RMSEs of these two reconstructions were 8.52% for model (2) and 7.74% for proposed model (5), respectively. This further confirms the accuracy of the proposed method.

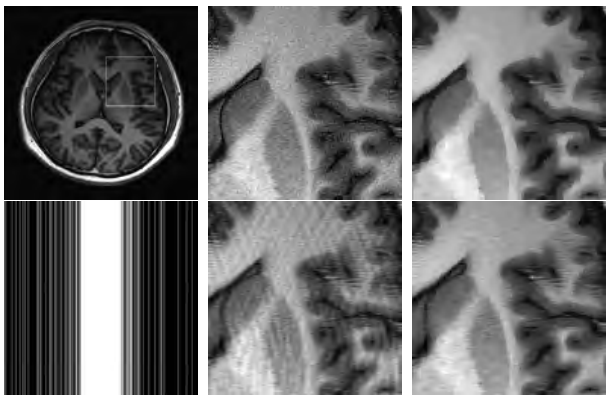

Fig. 4: Reconstruction of brain image in Fig. 2 using model (2) and (5). Left column: original image (upper) and sampling pattern in white (lower). Middle column: zoomed-in area of original image and reconstruction by zero-filling (lower). Right column: zoomed-in of reconstructions by model (2) (upper) and proposed model (5) (lower).

5. Conclusion

A novel framework of CS is introduced in this paper to tackle three shortcomings of the conventional framework. Trained dictionary is used to better preserve edges and fine structures by taking advantage of prior information; negative log-likelihood estimation of the recovering error is used to self-adjust the weight on the mean squared error of the reconstruction to improve the robustness of the model to the choice of the parameter; a quadratic penalty approach is adopted to speed up the reconstruction. The qualitative and quantitative comparisons on both phantom and in vivo data demonstrate the proposed method significantly improves the conventional framework in accuracy, robustness and stableness.

References

[1] E. J. Candes, J. K. Romberg, and T. Tao, "Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[2] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The Application of Compressed Sensing for Rapid MR Imaging," *Magn. Reson. Med.*, vol. 58, no. 6, pp. 1182–1195, 2007.

[3] T. Chang, L. He, and T. Fang, "MR Image Reconstruction from Sparse Radial Samples Using Bregman Iteration," *Proceedingss of Anual Meeting of ISMRM*, p. 696, 2006.

[4] J. Ye, S. Tak, Y. Han, and H. Park, "Projection Reconstruction MR Imaging Using FOCUSS," *Magn. Reson. Med.*, vol. 57, pp. 764–775, 2007.

[5] K. Block, M. Uecker, and J. Frahm, "Undersampled Radial MRI with Multiple Coils. Iterative Image Reconstruction Using a Total Variation Consraint," *Magn. Reson. Med.*, vol. 57, pp. 1086–1098, 2007.

[6] E. J. Candes and J. K. Romberg, "Signal Recovery from Random Projections," *Proceedings of SPIE Computational Imaging III*, vol. 5674, pp. 76–86, 2005.

[7] S. Chen, D. Donoho, and M. Saunders, "Atomic Decomposition by Basis Pursuit," *SIAM J. Sci. Comput.*, vol. 20, pp. 33–61, 1999.

[8] S. Kim, K. Koh, M. Lustig, and S. Boyd, "An Efficient Method for Compressed Sensing," *Proceedings of IEEE Intl. Conf. Image Process.(ICIP)*, vol. 3, pp. 117–120, 2007.

[9] D. L. Donoho, M. Elad, and V. Temlyakov, "Stable Recovery of Sparse Overcomplete Represenation in the Presence of Noise," *IEEE Trans. Inf. Theory*, vol. 52, pp. 6–18, 2006.

[10] H. Liao and G. Sapiro, "Sparse Representation for Limited Data Tomography," *Proceedings of IEEE Intl. Symp. Biomed. Imag. (ISBI)*, pp. 1375–1378, 2008.

[11] T. Goldstein and S. Osher, "The Split Bregman Method for L1 Regularized Problems, Tech. Rep. CAM Report 08-29, UCLA.

[12] L. Rudin, S. Osher, and E. Fatemi, "Non-linear Total Variation Noise Removal Algorithm," *Physics D.*, vol. 60, pp. 259–268, 1992.

[13] J. Starck, M. Elad, and D. Donoho, "Image Decomposition via the Combination of Sparse Representations and a Variational Approach," *IEEE Trans. Image Process.*, vol. 14, pp. 1570–1582, 2005.

[14] J. Yang, Y. Zhang, and W. Yin, "A Fast TVL1-L2 Minimization Algorithm for Signal Reconstruction from Partial Fourier Data," CAAM, Rice Univ., Tech. Rep. TR08-29, 2008.

[15] D. L. Donoho, "For Most Large Underdetermined Systems of Linear Equations, the Minimal l1-norm Solution is Also the Sparsest Solution," *Commun. Pure Appl. Math.*, vol. 59, no. 7, pp. 907–934, 2006.

[16] L. He, T.-C. Chang, S. Osher, T. Fang, and P. Speier, "MR Image Reconstruction by Using the Iterative Refinement Method and Nonlinear Inverse Scale Space Methods," *CAM Report 06-35, UCLA*, 2006.

[17] K. Kreutz-Delgado, J. F. Murray, B. D. Rao, K. Engan, T. Lee, and T. J. Sejnowski, "Dictionary Learning Algorithms for Sparse Representation," *Neural Comp.*, vol. 15, pp. 349–396, 2003.

[18] J. F. Murray and K. Kreutz-Delgado, "An Improved Focuss-based Learning Algorithm for Solving Sparse Linear Inverse Problems," *IEEE Int. Conf. Signals, Syst. Comput.*, pp. 347–351, 2001.

[19] M. Aharon, M. Elad, and A. Bruckstein, "The K-SVD: An algorithm for Designing of Overcomplete Dictionaries for Sparse Representation," *IEEE Trans. Signal Process*, vol. 54, no. 11, pp. 4311–4322, 2006.

[20] M. Elad and M. Aharon, "Image Denoising Via Sparse and Redundant Representations Over Learned Dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, 2006.

[21] W. Yin, S. Osher, D. Goldfarb, , and J. Darbon, "Bregman Iterative Algorithms for ℓ_1-Minimization with Applications to Compressed Sensing," *SIAM J. Imaging Science*, vol. 1, no. 1, pp. 143–168, 2008.
[22] S. Osher, Y. Mao, B. Dong, and W. Yin, "Fast Linearized Bregman Iteration for Compressive Sensing and Sparse Denoising, Tech. Rep. CAM Report 08-37, UCLA.
[23] J.-F. Cai, S. Osher, and Z. Shen, "Linearized Bregman Iterations for Compressed Sensing," *CAM Report 08-06, UCLA*, 2008.
[24] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A New Alternating Minimization Algorithm for Total Variation Image Reconstruction," *SIAM J. Imaging Science*, vol. 1, no. 3, pp. 248–272, 2008.
[25] R. Courant, "Variational Methods for the Solution of Problems with Equilibrium and Vibration," *Bull. Am. Math. Soc.*, vol. 49, pp. 1–23, 1943.