COMPRESSED SENSING MRI WITH COMBINED SPARSIFYING TRANSFORMS AND SMOOTHED $\ell_0$ NORM MINIMIZATION

Xiaobo Qu, Xue Cao, Di Guo, Changwei Hu, Zhong Chen*

Departments of Communication Engineering, Software Engineering, and Physics, Fujian Key Laboratory of Plasma and Magnetic Resonance, Xiamen University, Xiamen, 361005, P.R. of China

Email: {quxiaobo,guodi,chenz}@xmu.edu.cn, snowcx@gmail.com, ianhoo65@gmail.com

ABSTRACT

Undersampling the k-space is an efficient way to speed up the magnetic resonance imaging (MRI). Recently emerged compressed sensing MRI shows promising results. However, most of them only enforce the sparsity of images in single transform, e.g. total variation, wavelet, etc. In this paper, based on the principle of basis pursuit, we propose a new framework to combine sparsifying transforms in compressed sensing MRI. Each transform can efficiently represent specific feature that the other can not. This framework is implemented via the state-of-art smoothed $\ell_0$ norm in overcomplete sparse decomposition. Simulation results demonstrate that the proposed method can improve image quality when comparing to single sparsifying transform.

Index Terms— Sparse decomposition, compressed sensing, MRI, medical imaging, multiscale transform

1. INTRODUCTION

Undersampling the k-space is a good way to speed up magnetic resonance imaging (MRI). However, this will violate the Nyquist sampling rule and results in artifacts. Recently emerged compressed sensing [1,2] provides a firm foundation to reconstruct signal from fewer measurements than the Nyquist sampling rule requests. For the signal $x \in \mathbb{R}^N$ that can be represented by $K$ nonzero terms, the signal can be reconstructed exactly with overwhelming probability when the number of acquired data $M$ satisfies $M \geq \text{Const} \cdot K \cdot \log N$. For a compressible signal $x \in \mathbb{R}^N$, the reconstruction error of compressed sensing is proportional to the error of approximating image with $K$ largest nonzero terms in specific sparsifying transform domain.

Lustig et al first proposed the basic mathematical model of compressed sensing MRI (CS-MRI) [3]. However, only one sparsifying transform is applied in his model. Diverse sparsifying transforms that can sparsely represent different types of features of MRI images are discussed in [3] and the recommended transform is 2D wavelet. But traditional 2D wavelet obtained by tensor products of 1D wavelets is good at isolating the point discontinuities, but fails in sparsely representing the curve-like image features[4,5]. To overcome this shortage, wavelet can be replaced by other geometric image transforms, curvelet[4] or contourlet[5,6], to sparsely represent curves. But they are not good at representing point-like image features.

Since each transform can only sparsely represent one type of features, a combination of them is a good choice. In this paper, we combine these sparsifying transforms to provide a overcomplete dictionaries. Our method is directly inspired by the principle of basis pursuit[7], which tries to find an optimal superposition of dictionary elements. For compressed sensing MRI, images are reconstructed from undersampled k-space data by searching the sparsest representation via $\ell_0$ quasi-norm (it is not exactly a metric) minimization.

Compared with previous work, the advantages of our method are: (i) quality of reconstructed image is improved by enforcing its sparsity in combined sparsifying transforms; (ii) our combined transforms can avoid the bias of selected single transforms which may greatly reduce the image quality. We only consider noiseless measurements right now.

2. SMOOTHED $\ell_0$ NORM IN CS-MRI

Compressed sensing is a new theory to reconstruct signal from undersampled measurements. Suppose that a noiseless signal $x \in \mathbb{C}^N$ is sampled by the sensing matrix $\Phi_{M \times N} (M<N)$, the measurements $y \in \mathbb{C}^M$ of $x$ are

$$y_{M \times 1} = \Phi_{M \times N} x_{N \times 1}$$

Under the assumption that $x$ can be sparsely represented in transform $\Psi$ domain and $\alpha$ is the coefficient with respect to $\Psi$, $x$ can be presented as

$$x = \Psi \alpha$$

CS tries to reconstruct the signal from undersampled measurements by minimizing $\ell_0$ norm optimization. Let $\hat{\alpha}$ denote the estimation of $\alpha$, the $\ell_0$ norm optimization is

$$\hat{\alpha} = \min_{\alpha} \| \alpha \|_0 \quad s.t. \quad y = \Phi x$$
Counts the non-zero elements of $\alpha$.

In the field of MRI, the sampling matrix $\Phi$ is replaced by under-sampling Fourier operator $F_x$, which means partial Fourier coefficients are sensed. CS-MRI can be expressed as

$$\hat{y} = \min_{\alpha} \| \alpha \|_0 \quad s.t. \quad y = F_x \alpha$$  \hspace{1cm} (1)$$

where $y$ is the acquired k-space data and $\alpha$ be the image and $\Psi$ be the dictionary.

Because $\ell_0$ quasi-norm minimization is an NP hard problem, it is replaced by its closest linear counterpart, $\ell_1$ norm minimization, and the results hold. However, the simpler solution is at the cost of increasing the number of required measurements for exact reconstruction [1, 2, 8].

A recent work proposed a relaxation that uses continuous function to approximate $\ell_0$ quasi-norm [9] as

$$\| \alpha \|_0 \approx H_\alpha(\alpha) = M - \sum_{i=1}^{M} f_{\alpha_i}(\alpha_i) = M - \sum_{i=1}^{M} \exp\left(-\frac{\alpha_i^2}{2\sigma^2}\right)$$

where $\alpha_i$ is the $i^{th}$ element of vector $\alpha$ with length $M$.

When smoothed $\ell_0$ is applied in CS-MRI, we can reconstruct the MR images from undersampled k-space by solving the following problem

$$\hat{y} = \arg \min_{\alpha} H_\alpha(\alpha) \quad s.t. \quad y = F_x \alpha$$

and the reconstructed image is $\hat{x} = \Psi \alpha$.

Smoothed $\ell_0$ minimization can reduce the required sampling rate to gain expected reconstruction quality with a given dictionary. However, it is limited to the sparsity of the image. This motivates us to construct a bigger and more expressive dictionary to enhance the sparsity of an image in it. In this paper, we extend the smoothed $\ell_0$ approach to 2D compress sensing MRI and propose a method using combined sparsifying transforms.

### 3. COMBINED SPARSIFYING TRANSFORMS

As a major approach to solve CS, basis pursuit [7] suggests improving the sparsity of signal $x$ with length $N$ in overcomplete waveform dictionaries $\Psi=[\Psi_f, \Psi_\alpha, \cdots \Psi_{wT}, \cdots \Psi_{wM}]$ $\ (M>N)$ as concatenation of the subsets $\{\psi_i, i=1,2,\cdots , I\}$, where $\Lambda_i$ is the indices of waveforms in the $i^{th}$ subset.

$$\|z\|_0 = \sum_{j=1}^{M} \|\psi_j, x\|_0 = \sum_{j=1}^{M} \|\psi_j, x\|_0$$

This indicates that $\ell_0$ norm minimization in global overcomplete dictionary is equivalent to minimize the sum of $\ell_1$ norm of the dictionary’s subsets. So, (2) can be written as

$$\arg \min_{\Psi} \sum_{j=1}^{M} \|\psi_j, x\|_0 \quad s.t. \quad y = F_x \alpha$$

This provides us with an opportunity to recover the MR image from union of subsets of waveforms dictionary.

The model can be extended to $\ell_0$ norm seamlessly. $\ell_0$ norm minimization in global overcomplete dictionary is equivalent to minimize the sum of $\ell_0$ norm of the dictionary’s subsets. So, the combined sparsifying transforms for $\ell_0$ norm can be expressed as

$$\arg \min_{\Psi} \sum_{j=1}^{M} \|\psi_j, x\|_0 \quad s.t. \quad y = F_x \alpha$$

In this paper, we consider the condition each sub-dictionary $\Psi_j$ comes from commonly used transforms which sparsify different types of image features. Unfortunately, the matrix of dictionary $\Psi$ and its adjoint $\Psi^T$ are rarely explicitly constructed in memory, e.g. curvelet and contourlet. Instead, they are implemented as fast implicit analysis and synthesis operators. Let $T$ and $T^*$ denote the operator pair, $\Psi^Tx = Tx$ and $T^*x = Ix$. So that even if $\Psi$ is a tight frame and it may not be orthogonal, i.e. $\Psi^T \Psi = cI, c > 0$, we don’t have to normalize $T$ and $T^*[10]$.

Furthermore, storing and computation of $\alpha$ is expensive because dimension of $\alpha$ is higher than dimension of signal $x$ for overcomplete dictionary. So we apply fast forward transform $T_f$ on the image and get another version of (3)

$$\arg \min_{\Psi} \sum_{j=1}^{M} \|\psi_j, x\|_0 \quad s.t. \quad y = F_x \alpha$$

This model assumes each transform is regularized, i.e. that the $\ell_1$ norm of the dictionary’s each column equals to 1. In practice, the condition is more complicated. An effective solution to solve this problem is to enforce the sparsity in each transform domain individually.

For the transforms under consideration, an ideal case is to construct an infinite dictionary that contains any possible waveform. In this case, if a waveform exactly the same with target image is included, such that the image we are to reconstruct can be represented with only one coefficient. However, this case is obviously unrealistic. Instead, we try to find several complement transforms, in other words, the

---

coherence between them is expected to be as low as possible, namely the uncertainty principle. The Morphological Component Analysis (MCA) [10] framework assumes an image \( x \) can be composed into several components, each component \( x_j \) can be sparsely represented in an associated basis \( \Psi_j \). For each \( j \), the representation of \( x_j \) in \( \Psi_j \) is sparsest, and in any other \( \Psi_i (i \neq j) \), it is not or at least not as sparse as another sparsifying transform.

Considering talents in representing different image features and the computing complexity [5,6], we adopted wavelet and an improved contourlet [6] to represent point-like and curve-like image features respectively.

### 4. IMPLEMENTATION OF THE COMBINED APPROACH

In order to solve equation (4), we extend sigma annealing method proposed in [9] to compressed sensing MRI with combined sparsifying transforms.

Without loss of generality, suppose we have \( K \) different sparsifying transforms combined. When minimizing the coefficients in the transform domain of \( T_k \), which consist in the \( k \)-th subset of the global overcomplete dictionary, we assume the coefficients corresponding to other transform domains are fixed. Theoretically, the order of transforms is arbitrary as long as we traverse all the selected transforms in \( \{ T_k | k = 1, 2, \ldots, K \} \). During the calculation process, we keep projecting the coefficients back onto the feasible set so that the data consistency is ensured.

Given an annealing sequence of \( [\sigma_1, \sigma_2, \ldots, \sigma_J] \), we minimize smoothed \( \ell_0 \) norm iteratively for each \( \sigma \) with general gradient descent method. In implementation of smoothed \( \ell_0 \) norm, value of \( \sigma \) at \( j + 1 \) iteration is \( \sigma_{j+1} = d \star \sigma_j \) ( \( 0 < d < 1 \) ). A decreasing sequence of sigma is carefully selected and the minimization of smoothed \( \ell_0 \) norm is roughly calculated for each sigma in the sequence. Simulations in [9] demonstrate the smoothed \( \ell_0 \) norm algorithm converges fast. In [8] [4], the authors proposed to solve the Lagrange form of (1). But as mentioned before, the equivalent dictionary of some geometric transform’s fast algorithm is not regularized, i.e. the \( \ell_2 \) norm of each column is not equal to 1. So we had better force the sparsity of image in each transform domain individually and keep projecting the results back to feasible set after each optimization step.

The reconstruction algorithm with smoothed \( \ell_0 \) norm and combined sparsifying transforms (SL0-CST) is as follows

```plaintext
Initialization:
1) Let \( x_0 \) be the minimum \( \ell_2 \) norm solution of \( y = A x \) obtained by pseudo-inverse of \( A \).
2) Choose a suitable annealing strategy for \( \sigma \) and get its sequence \( [\sigma_1, \sigma_2, \ldots, \sigma_j, \ldots, \sigma_J] \).

Main Loop:
for \( j = 1, 2, \ldots, J \)
   Set \( \sigma = \sigma_j, x = \hat{x}_j \)
   for \( k = 1, 2, \ldots, K \), where \( K \) is the number of transforms employed.
      1) \( \alpha_k = \psi_k x \)
      2) \( g_k = \alpha_k \exp \left( -\frac{\alpha_k}{2\sigma^2} \right) \)
      3) \( x = x - t \cdot \psi_k g_k \) with step length \( t \)
      4) \( x = x - \lambda F_k (F_k x - y) \) where \( \lambda \) controls the data consistency and \( \lambda = 1 \) by default.
   end for
Set \( \hat{x}_j = x \)
end for
Final answer \( \hat{x} = \hat{x}_j \)
```

### 5. SIMULATION RESULTS

To validate the performance, the proposed method is compared with the nonlinear conjugate (CG) \( \ell_1 \) norm when single and combined sparsifying transforms are employed [11]. Parameters of nonlinear conjugate \( \ell_1 \) norm are the same as [11]. Parameters of smoothed \( \ell_0 \) norm are set as decay factor \( d \) of \( \sigma \) is 0.5 and the stop criteria, smallest \( \sigma \), is \( 10^{-4} \). The limitations of outer and inner iterations are 50 and 5. Besides the visual appearance, peak signal-to-noise ratio (PSNR) is served as objective criteria and defined as

\[
PSNR = 20 \log_{10} \left( \frac{255}{\sqrt{MSE}} \right)
\]

where \( MSE = \frac{1}{P \times Q} \sum_{p=1}^{P} \sum_{q=1}^{Q} (\hat{x}(p,q) - x(p,q))^2 \) and \( (p,q) \) is pixel location of a \( P \times Q \) image. \( \hat{x} \) is the reconstructed image and \( x \) is reconstructed image from fully sampled k-space data. PSNR evaluates the difference in gray values between \( \hat{x} \) and \( x \).

We adopt the variable density sampling pattern, shown in Fig.1 (b), to acquire 15% of k-space data. The reconstructed curves via contourlet are much clearer than those of wavelet. No matter in the \( \ell_1 \) norm or \( \ell_0 \) norm, combined sparsifying transforms can improve the image quality than single transform. Due to the approximation of \( \ell_0 \) norm, smoothed...

\[ \ell_0 \] norm with combined sparsifying transforms obtains the best results. This conclusion is in accordance with the PSNR performance in Table 1.

![Figure 1. Comparisons on reconstructed images. CG denotes nonlinear conjugate $\ell_1$ norm and SL0 denotes smoothed $\ell_0$ norm, CT means contourlet and WT means wavelet.](image)

<table>
<thead>
<tr>
<th>Types of norm</th>
<th>Sparsifying transforms</th>
<th>PSNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonlinear conjugate $\ell_1$</td>
<td>Wavelet</td>
<td>29.02</td>
</tr>
<tr>
<td></td>
<td>Contourlet</td>
<td>31.81</td>
</tr>
<tr>
<td></td>
<td>Wavelet+Contourlet</td>
<td>32.56</td>
</tr>
<tr>
<td>Smoothed $\ell_0$</td>
<td>Wavelet</td>
<td>28.16</td>
</tr>
<tr>
<td></td>
<td>Contourlet</td>
<td>32.36</td>
</tr>
<tr>
<td></td>
<td>Wavelet+Contourlet</td>
<td>34.23</td>
</tr>
</tbody>
</table>

6. CONCLUSION

The combined sparsifying transforms for smooth $\ell_0$ norm minimization is proposed to reconstruct the magnetic resonance images from noiseless undersampled k-space data. Theoretical analysis and simulation results demonstrate that the proposed method can improve image quality than single sparsifying transform. However, the robust to noisy k-space data is the future work.

7. ACKNOWLEDGEMENT

Prof. Zhong Chen is the author to whom the correspondence should be addressed, his e-mail address is chenz@xmu.edu.cn. This work was partially supported by NNSF of China under Grants (10774125,10605019 and 10974164). Thank Mohimani H., Lustig M., Lu Y. and Trzasko J. for sharing their code.

8. REFERENCES


==========================================================================


