

Research Article

SART-Type Image Reconstruction from a Limited Number of Projections with the Sparsity Constraint

Hengyong Yu and Ge Wang

SBES Division and ICTAS Center for Biomedical Imaging, VT-WFU School of Biomedical Engineering and Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA

Correspondence should be addressed to Hengyong Yu, hengyong-yu@ieee.org and Ge Wang, ge-wang@ieee.org

Received 31 December 2009; Accepted 10 February 2010

Academic Editor: Yibin Zheng

Copyright © 2010 H. Yu and G. Wang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Based on the recent mathematical findings on solving the linear inverse problems with sparsity constraints by Daubechiesx et al., here we adapt a simultaneous algebraic reconstruction technique (SART) for image reconstruction from a limited number of projections subject to a sparsity constraint in terms of an invertible compression transform. The algorithm is implemented with an exemplary Haar wavelet transform and tested with a modified Shepp-Logan phantom. Our preliminary results demonstrate that the sparsity constraint helps effectively improve the quality of reconstructed images and reduce the number of necessary projections.

1. Introduction

Worldwide there are growing concerns on radiation induced genetic, cancerous, and other diseases [1–3]. Computed tomography (CT) is considered as a radiation-intensive procedure, yet it becomes more and more common. In the mid-1990s, CT scans only accounted for 4% of the total X-ray procedures but they contributed 40% of the collective dose [4]. The introduction of helical, multislice, and cone-beam technologies has increased and continues the increasing usage of CT [5, 6]. In US, the number of CT examinations has been estimated as high as nearly 60 million in 2002, which account for 15% of imaging procedures and 75% of the radiation exposure [4]. In June 2007, the *New York Times* reported that “the per-capita dose of ionizing radiation from clinical imaging exams in the U.S. increased almost 600% from 1980 to 2006.” More recently, in a high-profile article on the rapid growth in CT use and its associated radiation risks [3], Brenner and Hall estimated that “on the basis of such risk estimates and data on CT use from 1991 through 1996, it was estimated that about 0.4% of all cancers in the United States may be attributable to the radiation from CT studies. By adjusting this estimate for current CT use, this estimate might now be in the range of 1.5 to 2.0%.” Facing

the increasing radiation risk, the well-known As Low As Reasonably Achievable (ALARA) principle is widely accepted in the medical community. One of the practical strategies is to reduce the number of necessary projection.

Very interestingly, an elegant theory of compressive sampling or compressive sensing (CS) has recently emerged which shows that high-quality signals and images can be reconstructed from far fewer measurements than what is usually considered necessary according to the Nyquist sampling theorem [7, 8]. The main idea of CS is that most signals are sparse in an appropriate orthonormal system; that is, a majority of their coefficients are close or equal to zero, when represented in the proper domain. Typically, CS starts with taking a limited amount of samples in a much less correlated basis, and then the signal is exactly recovered with an overwhelming probability from the limited amount of data via the ℓ_1 norm minimization. For example, the discrete gradient sparsifying transform has been widely utilized in CS-inspired CT reconstruction [9, 10], which was also referred to as the total variation minimization [11]. However, because the discrete gradient transform does not satisfy the restricted isometry property (RIP) required by the CS theory and is not invertible in general, such a reconstruction does not always convey the medically precise information.

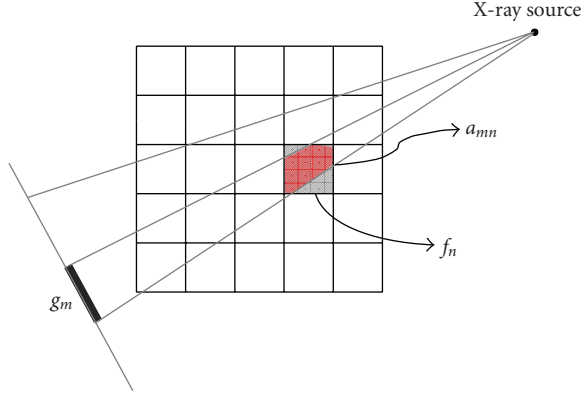


FIGURE 1: Projection model of a discrete image in fan-beam geometry.

In particular, when a small number of projections are collected by a CT scanner, data noise may hide tumor-like structures in the TV-minimization-based reconstruction [12].

The above problem can be overcome using an invertible sparsifying transform such as a wavelet transform for image compression. For an object of interest such as a medical image, we can find an orthonormal basis (in a more general setting, a frame) to make the object sparse in terms of significant transform coefficients. Then, we can perform image reconstruction from a limited number of projections by minimizing the corresponding ℓ_1 norm. Based on the recent mathematical findings made by Daubechies et al. [13, 14], here we will adapt a simultaneous algebraic reconstruction technique (SART) [15] for image reconstruction from a limited number of projections subject to a sparsity constraint in terms of an invertible sparsifying transform.

This paper is organized as follows. In the next section, the mathematical principles will be summarized. In the third section, a SART-type reconstruction algorithm will be developed with a sparsity constraint. In the fourth section, preliminary numerical simulation results will be presented. In the last section, the related issues will be discussed.

2. Mathematical Principles

Daubechies and her collaborators proposed a general iterative thresholding algorithm to solve linear inverse problems regularized by a sparsity constraint and proved its convergence [13, 14]. Their approach can be directly applied for the CT reconstruction from a limited number of projections. Their major results can be summarized as follows.

Let $\mathbf{f} = [f_1, f_2, \dots, f_N]^T \in \mathbb{R}^N$ be an object function and $\mathbf{g} = [g_1, g_2, \dots, g_M]^T \in \mathbb{R}^M$ a set of measurements. Usually, they are linked by

$$\mathbf{g} = \mathbf{A}\mathbf{f} + \mathbf{e}, \quad (1)$$

where $\mathbf{A} = (a_{mn}) \in \mathbb{R}^M \times \mathbb{R}^N$ is the linear measurement matrix, and $\mathbf{e} \in \mathbb{R}^M$ the measurement noise. Let us define

the ℓ_p norm of the vector \mathbf{g} as

$$\|\mathbf{g}\|_p = \left(\sum_{m=1}^M g_m^p \right)^{1/p}. \quad (2)$$

In practical applications, we usually omit the subscript p when $p = 2$. To estimate \mathbf{f} from \mathbf{g} , one can minimize the discrepancy $\Delta(\mathbf{f})$

$$\Delta(\mathbf{f}) = \|\mathbf{g} - \mathbf{A}\mathbf{f}\|^2. \quad (3)$$

When system (1) is ill posed, the solution to (3) is not satisfactory, and additional constraints are required to regularize the solution. Particularly, given a complete basis $(\boldsymbol{\varphi}_\gamma)_{\gamma \in \Gamma}$ of the space \mathbb{R}^N satisfying $\mathbf{f} = \sum_{\gamma \in \Gamma} \langle \mathbf{f}, \boldsymbol{\varphi}_\gamma \rangle \boldsymbol{\varphi}_\gamma$, and a sequence of strictly positive weights $\mathbf{w} = (w_\gamma)_{\gamma \in \Gamma}$, we define the functional $\Phi_{\mathbf{w},p}(\mathbf{f})$ by

$$\Phi_{\mathbf{w},p}(\mathbf{f}) = \Delta(\mathbf{f}) + \sum_{\gamma \in \Gamma} 2w_\gamma \left| \langle \mathbf{f}, \boldsymbol{\varphi}_\gamma \rangle \right|^p, \quad (4)$$

where $\langle \cdot, \cdot \rangle$ represents the inner product and $1 \leq p \leq 2$. Using the ℓ_p norm definition (2), let us define the ℓ_p norm of a matrix operator \mathbf{A} as

$$\|\mathbf{A}\|_p = \max_{\mathbf{f} \neq 0} \left(\frac{\|\mathbf{A}\mathbf{f}\|_p}{\|\mathbf{f}\|_p} \right). \quad (5)$$

Let \mathbf{A}^T be the transpose matrix of \mathbf{A} , the operator \mathbf{A} in (1) is bounded, and $\|\mathbf{A}^T\mathbf{A}\| < C$. In the following, we will assume $C = 1$ because \mathbf{A} can always be renormalized. To find an estimate of \mathbf{f} from \mathbf{g} under the ℓ_p norm regularization term $\sum_{\gamma \in \Gamma} 2w_\gamma |\langle \mathbf{f}, \boldsymbol{\varphi}_\gamma \rangle|^p$, we can minimize $\Phi_{\mathbf{w},p}(\mathbf{f})$ defined in (4). The minimizer of $\Phi_{\mathbf{w},p}(\mathbf{f})$ can be recursively determined by the soft-thresholding algorithm:

$$\mathbf{f}^k = \mathcal{S}_{\mathbf{w},p} \left(\mathbf{f}^{k-1} + \mathbf{A}^T (\mathbf{g} - \mathbf{A}\mathbf{f}^{k-1}) \right), \quad (6)$$

where $k = 1, 2, \dots$ is the iteration number, \mathbf{f}^0 the initial value in \mathbb{R}^N , and

$$\mathcal{S}_{\mathbf{w},p}(\mathbf{f}) = \sum_{\gamma \in \Gamma} S_{w_\gamma,p} \left(\langle \mathbf{f}, \boldsymbol{\varphi}_\gamma \rangle \right) \boldsymbol{\varphi}_\gamma \quad (7)$$

with $S_{w,p} = (F_{w,p})^{-1}$ being a one-to-one map from \mathbb{R} to its self for $p > 1$ with

$$F_{w,p}(x) = x + wp \operatorname{sgn}(x) |x|^{p-1}. \quad (8)$$

Particularly,

$$S_{w,3/2}(x) = \begin{cases} x - \frac{3w(\sqrt{9w^2 + 16|x|} - 3w)}{8} & \text{if } x \geq 0, \\ x + \frac{3w(\sqrt{9w^2 + 16|x|} - 3w)}{8} & \text{if } x < 0. \end{cases} \quad (9)$$

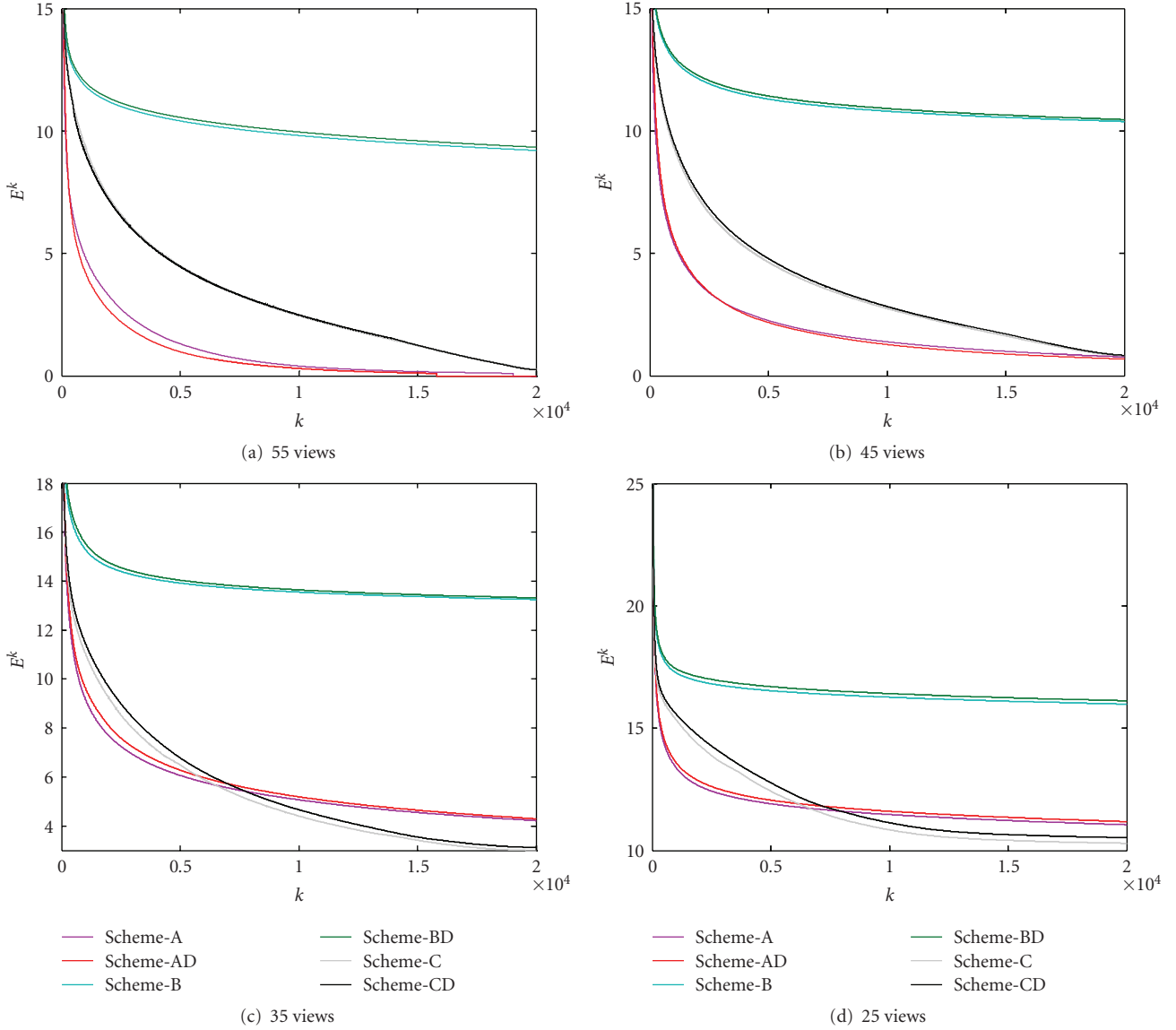


FIGURE 2: Relative error curves for various reconstructions from noise-free projections. The numbers of projections were set to (a) 55, (b) 45, (c) 35, and (d) 25, respectively, where the horizontal and vertical axes represent the iteration index k and the corresponding log error E^k , respectively.

When $p = 1$, we can set [13]

$$S_{w,1}(x) = \begin{cases} x - w & \text{if } x \geq w, \\ 0 & \text{if } |x| < w, \\ x + w & \text{if } x \leq -w. \end{cases} \quad (10)$$

The main result of Daubechies et al. in [13] is that the solution of (6) is convergent.

Unfortunately, the convergence speed of (6) is very slow. To facilitate practical applications, an accelerated projected gradient method was then developed [14]. When $w_\gamma = \tau$ for all $\gamma \in \Gamma$, $\Phi_{w,p}(\mathbf{f})$ can be rewritten as

$$\Phi_{w,p}(\mathbf{f}) = \Phi_{\tau,p}(\mathbf{f}) = \Delta(\mathbf{f}) + \sum_{\gamma \in \Gamma} 2\tau \left| \langle \mathbf{f}, \boldsymbol{\varphi}_\gamma \rangle \right|^p. \quad (11)$$

Denote the minimizer of (11) as \mathbf{f}^* and define

$$R(\mathbf{f}^*, p) = \left(\sum_{\gamma \in \Gamma} \left| \langle \mathbf{f}^*, \boldsymbol{\varphi}_\gamma \rangle \right|^p \right)^{1/p}, \quad (12)$$

which is the ℓ_p norm radius of \mathbf{f}^* in the sparse space, we have the accelerated projected gradient algorithm

$$\mathbf{f}^k = \mathbb{P}_{R(\mathbf{f}^*, p)}(\mathbf{f}^{k-1} + \beta^k \mathbf{r}^k), \quad (13)$$

where

$$\mathbf{r}^k = \mathbf{A}^T(\mathbf{g} - \mathbf{A}\mathbf{f}^{k-1}), \quad \beta^k = \frac{\|\mathbf{r}^k\|^2}{\|\mathbf{A}\mathbf{r}^k\|^2}, \quad (14)$$

$$\mathbb{P}_{R(\mathbf{f}^*, p)}(\mathbf{f}) = S_{\mu,p}(\mathbf{f}) = \sum_{\gamma \in \Gamma} S_{\mu,p}(\langle \mathbf{f}, \boldsymbol{\varphi}_\gamma \rangle) \boldsymbol{\varphi}_\gamma,$$

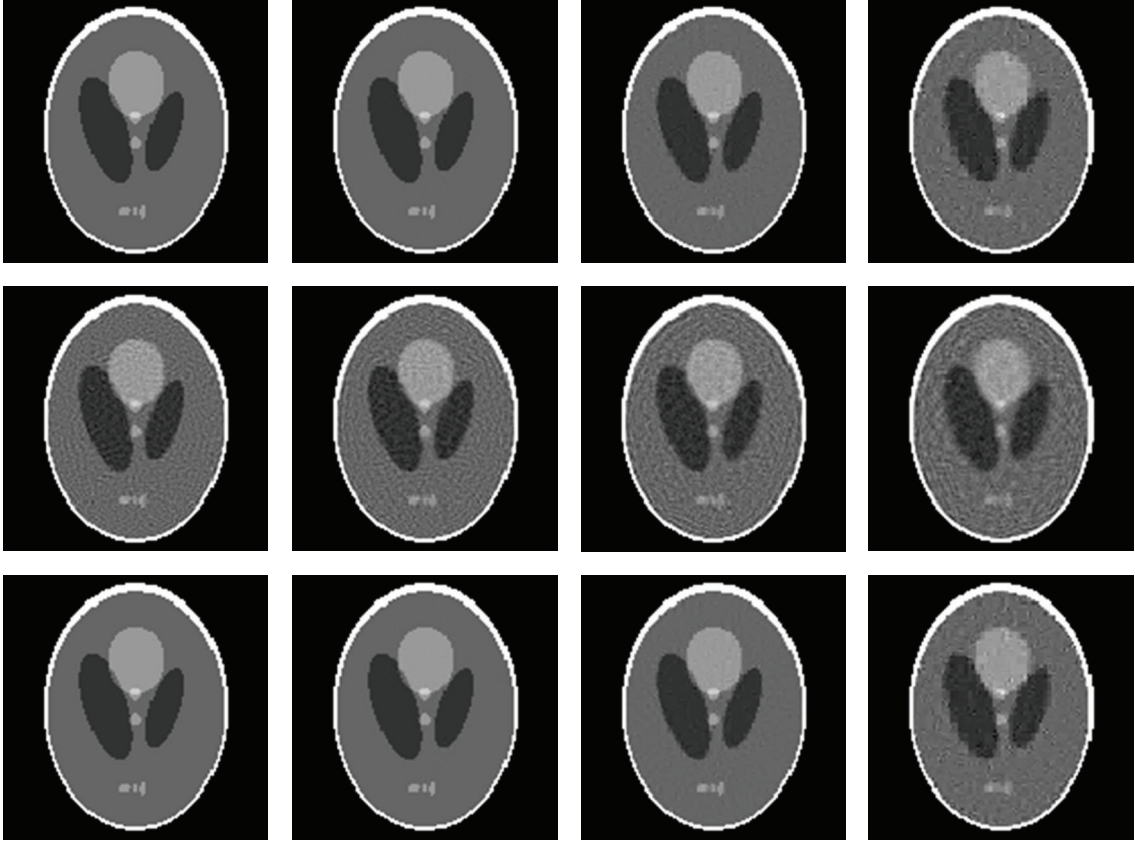


FIGURE 3: Reconstructed 128×128 images from noise-free projection datasets. The 1st, 2nd, and 3rd rows are reconstructed using Scheme-A, Scheme-B, and Scheme-C, respectively, and the 1st, 2nd, 3rd, and 4th columns are from 55, 45, 35, and 25 views respectively (the display window: $[0 \ 0.5]$).

with an adapted soft-threshold $\mu = \mu(R(\mathbf{f}^*, p), \mathbf{f})$ depending on $R(\mathbf{f}^*, p)$ and \mathbf{f} . When $R(\mathbf{f}, p) \leq R(\mathbf{f}^*, p)$, $\mu(R(\mathbf{f}^*, p), \mathbf{f}) = 0$ and $\mathbb{P}_{R(\mathbf{f}^*, p)}(\mathbf{f}) = \mathbf{f}$. When $R(\mathbf{f}, p) > R(\mathbf{f}^*, p)$, the adapted threshold μ should be chosen to satisfy

$$R(\mathbb{P}_{R(\mathbf{f}^*, p)}(\mathbf{f}), p) = R(\mathbb{S}_{\mu, p}(\mathbf{f}), p) = R(\mathbf{f}^*, p). \quad (15)$$

Regarding algorithm (13), we have several comments in order. First, although Daubechies et al. only proved the convergence for the case $p = 1$ [14], we believe that it should stand for $1 \leq p \leq 2$. Second, while we have previously assumed that $\|\mathbf{A}^T \mathbf{A}\| < C$ and $C = 1$, it can be proved that the algorithm (13) holds for any positive C . Third, it is generally impossible to know the exact value of $R(\mathbf{f}^*, p)$ but we can have an approximate estimate.

3. Algorithm Development

In the context of image reconstruction, each component of the function \mathbf{f} in (1) is interpreted as an image pixel with N being the total pixel number. Each component of the function \mathbf{g} is a measured datum with M being the product of the number of projections and the number of detector elements. In fan-beam geometry with a discrete image grid, the n th pixel can be viewed as a rectangular

region with a constant value f_n , the m th measured datum g_m as an integral of areas of pixels partially covered by a narrow beam from an X-ray source to a detector element and respectively weighted by the corresponding X-ray linear attenuation coefficients. Thus, the component a_{mn} in (1) can be understood as the interaction area between the n th pixel and the m th fan-beam path (Figure 1). While the whole matrix \mathbf{A} represents the forward projection, \mathbf{A}^T implements the back projection. The SART-type solution to (1) can be written as [15]

$$f_n^k = f_n^{k-1} + \lambda^k \frac{1}{a_{+n}} \sum_{m=1}^M \frac{a_{mn}}{a_{m+}} (g_m - \mathbf{A}_m \mathbf{f}^{k-1}), \quad (16)$$

where $a_{+n} = \sum_{m=1}^M a_{mn} > 0$, $a_{m+} = \sum_{n=1}^N a_{mn} > 0$, \mathbf{A}_m is the m th row of \mathbf{A} , k the iteration index, and $0 < \lambda^k < 2$ a free relaxation parameter. Let $\Lambda^{+N} \in \mathbb{R}^N \times \mathbb{R}^N$ be a diagonal matrix with $\Lambda_{nn}^{+N} = 1/a_{+n}$ and let $\Lambda^{M+} \in \mathbb{R}^M \times \mathbb{R}^M$ be a diagonal matrix with $\Lambda_{mm}^{M+} = 1/a_{m+}$, then (16) can be rewritten as

$$\mathbf{f}^k = \mathbf{f}^{k-1} + \lambda^k \tilde{\mathbf{r}}^k, \quad (17)$$

with

$$\tilde{\mathbf{r}}^k = \Lambda^{+N} \mathbf{A}^T \Lambda^{M+} (\mathbf{g} - \mathbf{A} \mathbf{f}^{k-1}). \quad (18)$$

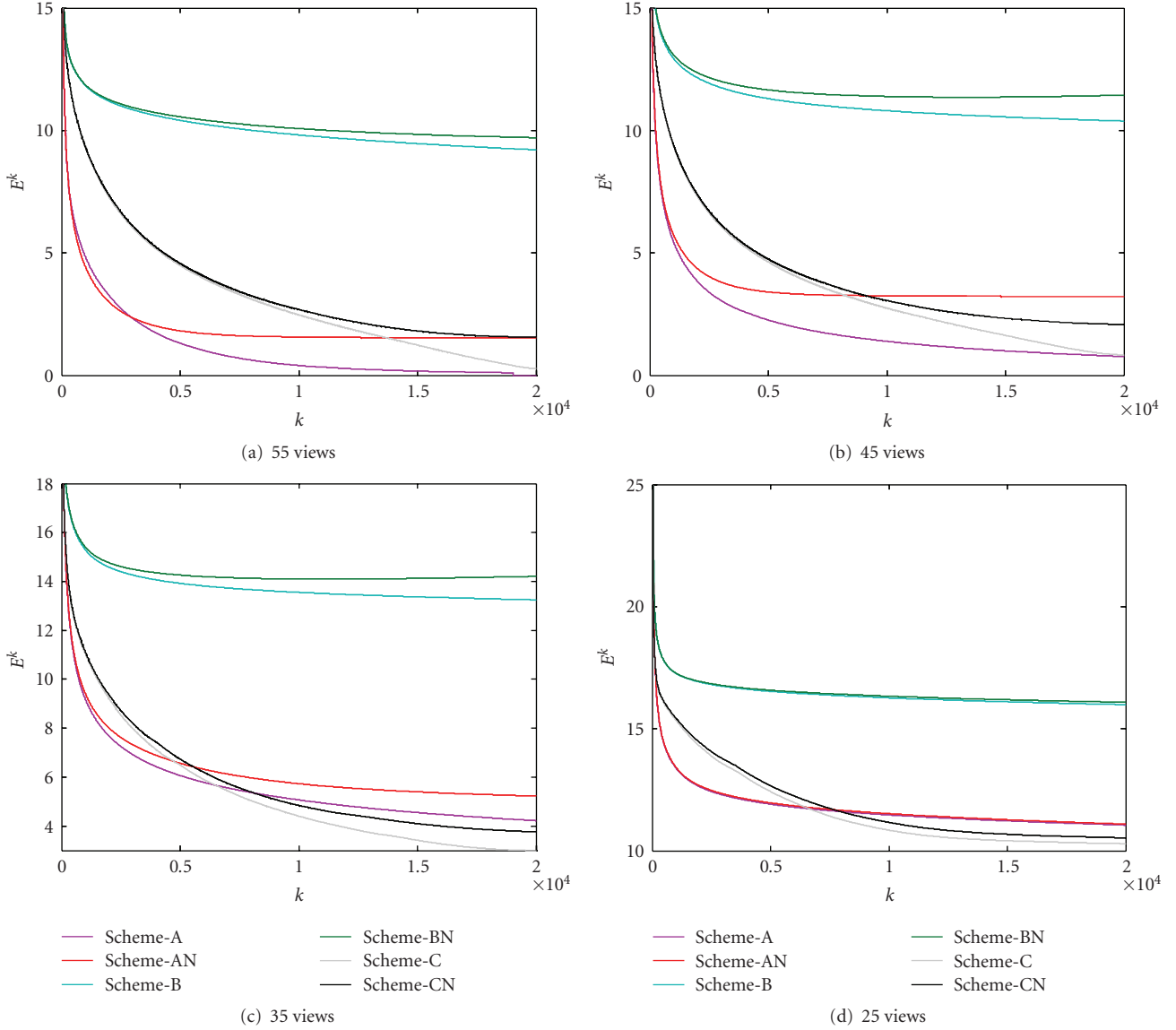


FIGURE 4: Relative error curves for various reconstructions from noisy projection. The numbers of projections were set to (a) 55, (b) 45, (c) 35, and (d) 25, respectively, where the horizontal and vertical axes represent the iteration index k and the corresponding log error E^k , respectively.

Due to the introduction of Λ^{+N} and Λ^{M+} , (18) cannot be directly applied to solve (13). However, we can modify (18) to obtain a new \mathbf{r}^k defined as

$$\mathbf{r}^k = \frac{\|\mathbf{A}^T\|}{\|\Lambda^{+N}\mathbf{A}^T\Lambda^{M+}\|} \tilde{\mathbf{r}}^k = \alpha \tilde{\mathbf{r}}^k. \quad (19)$$

Substituting (19) into (13), we have a SART-type algorithm

$$\mathbf{f}^k = \mathbb{P}_{R(\mathbf{f}^*, p)}(\mathbf{f}^{k-1} + \alpha \beta^k \tilde{\mathbf{r}}^k), \quad (20)$$

with $\beta^k = \|\tilde{\mathbf{r}}^k\|^2 / \|\mathbf{A}^T \tilde{\mathbf{r}}^k\|^2$. The heuristic rationale for the above modification is to incorporate the SART-type weighting

scheme for a more uniform convergence behavior. Now, our task is to estimate α . Since $\|\mathbf{A}^T\| = \|\mathbf{A}\|$, we have

$$\alpha^2 = \frac{\|\mathbf{A}^T\|^2}{\|\Lambda^{+N}\mathbf{A}^T\Lambda^{M+}\|^2} = \frac{\|\mathbf{A}^T\| \cdot \|\mathbf{A}\|}{\|\Lambda^{+N}\mathbf{A}^T\Lambda^{M+}\| \cdot \|\Lambda^{M+}\mathbf{A}\Lambda^{+N}\|}. \quad (21)$$

Let $\mathbf{I} \in \mathbb{R}^N$ be the vector with all whose components being “1”. We have $\|\mathbf{A}^T\mathbf{A}\|_1 = \|\mathbf{A}^T\mathbf{A}\|_\infty = \max_{1 \leq n \leq N} (\mathbf{A}^T\mathbf{A}\mathbf{I})$ because $\mathbf{A}^T\mathbf{A}$ is a symmetric matrix. Hence, we have

$$\|\mathbf{A}^T\mathbf{A}\| \leq \sqrt{\|\mathbf{A}^T\mathbf{A}\|_1 \|\mathbf{A}^T\mathbf{A}\|_\infty} = \max_{1 \leq n \leq N} (\mathbf{A}^T\mathbf{A}\mathbf{I}). \quad (22)$$

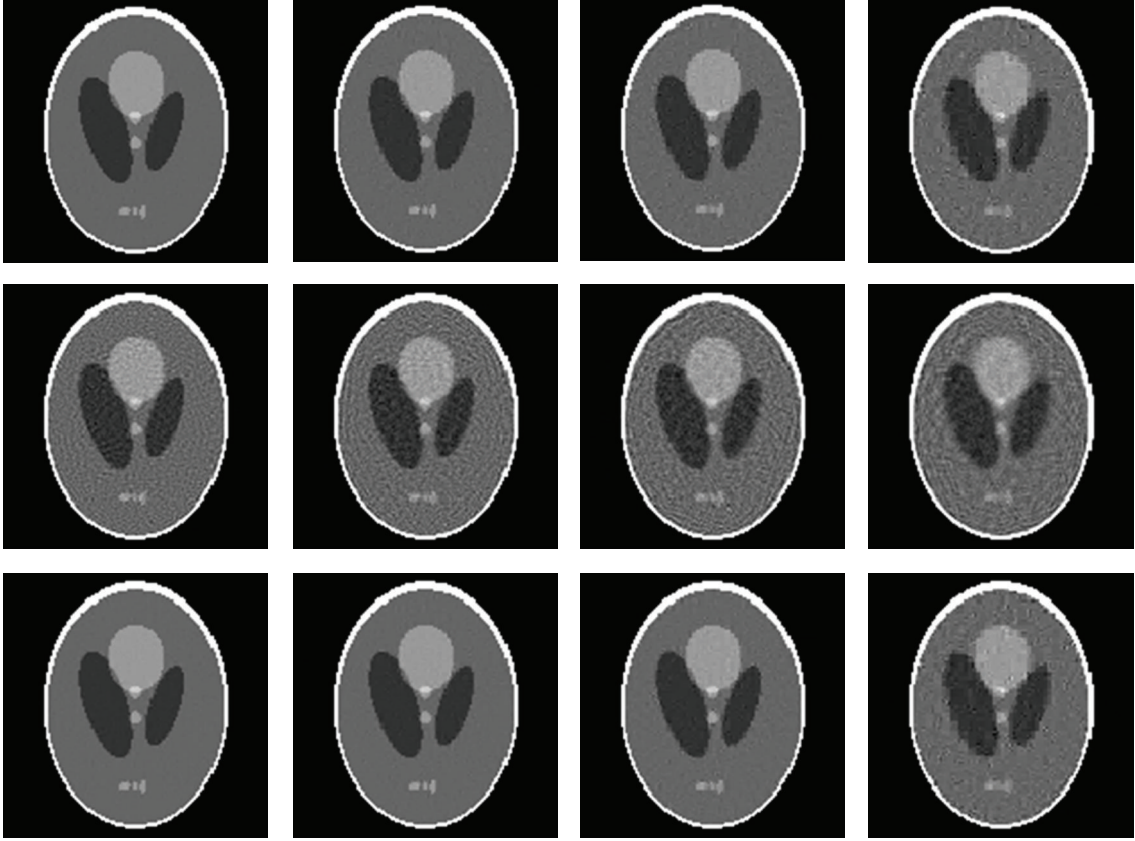


FIGURE 5: Counterpart of Figure 3 in the case of noisy projection datasets.

Similarly, we have

$$\begin{aligned}
 & \left\| \Lambda^{+N} \mathbf{A}^T \Lambda^{M+} \Lambda^{M+} \mathbf{A} \Lambda^{+N} \right\| \\
 & \leq \sqrt{\left\| \Lambda^{+N} \mathbf{A}^T \Lambda^{M+} \Lambda^{M+} \mathbf{A} \Lambda^{+N} \right\|_1 \left\| \Lambda^{+N} \mathbf{A}^T \Lambda^{M+} \Lambda^{M+} \mathbf{A} \Lambda^{+N} \right\|_\infty} \\
 & = \max_{1 \leq n \leq N} \left(\Lambda^{+N} \mathbf{A}^T \Lambda^{M+} \Lambda^{M+} \mathbf{A} \Lambda^{+N} \mathbf{I} \right). \tag{23}
 \end{aligned}$$

That is,

$$\alpha^2 = \frac{\max_{1 \leq n \leq N} (\mathbf{A}^T \mathbf{A} \mathbf{I})}{\max \left(\left\| \Lambda^{+N} \mathbf{A}^T \Lambda^{M+} \Lambda^{M+} \mathbf{A} \Lambda^{+N} \mathbf{I} \right\| \right)} \alpha_0^2, \tag{24}$$

with

$$\alpha_0^2 = \frac{\max_{1 \leq n \leq N} \left(\Lambda^{+N} \mathbf{A}^T \Lambda^{M+} \Lambda^{M+} \mathbf{A} \Lambda^{+N} \mathbf{I} \right)}{\left\| \Lambda^{+N} \mathbf{A}^T \Lambda^{M+} \right\| \cdot \left\| \Lambda^{M+} \mathbf{A} \Lambda^{+N} \right\|} \frac{\left\| \mathbf{A}^T \right\| \cdot \left\| \mathbf{A} \right\|}{\max_{1 \leq n \leq N} (\mathbf{A}^T \mathbf{A} \mathbf{I})}. \tag{25}$$

In practical applications, we can set α_0 to a reasonably large constant such as 2.0 in our simulation in the next section. If the algorithm does not converge, we can reduce α_0 until the algorithm converges.

For a basis $(\boldsymbol{\varphi}_\gamma)_{\gamma \in \Gamma}$ of the space \mathbb{R}^N , in which \mathbf{f} has a sparse representation. Our SART-type CT algorithm regularized by sparsity can be summarized in the following pseudocode:

- S1 Initialize $\alpha_0, \mathbf{f}^{(0)}$ and k ;
- S2 Estimate $R(\mathbf{f}^*, p)$;
- S3 Precompute α, a_{+n} and a_{m+} ;
- S4 Update the current estimation iteratively:
 - S4.1 $k := k + 1$;
 - S4.2 $\tilde{\mathbf{r}}^k := \Lambda^{+N} \mathbf{A}^T \Lambda^{M+} (\mathbf{g} - \mathbf{A} \mathbf{f}^{k-1})$;
 - S4.3 $\beta^k := \|\tilde{\mathbf{r}}^k\|^2 / \|\mathbf{A} \tilde{\mathbf{r}}^k\|^2$;
 - S4.4 $\tilde{\mathbf{f}}^k := \mathbf{f}^{k-1} + \alpha \beta^k \tilde{\mathbf{r}}^k$;
 - S4.5 Compute the sparse transform $\phi_\gamma := \langle \tilde{\mathbf{f}}^k, \boldsymbol{\varphi}_\gamma \rangle$ for $\gamma \in \Gamma$;
 - S4.6 Estimate the adapted threshold μ ;
 - S4.7 Perform the soft-thresholding $\tilde{\phi}_\gamma := S_{\mu, p}(\phi_\gamma)$;
 - S4.8 Perform the inverse sparse transform $\mathbf{f}^k := \sum_{\gamma \in \Gamma} \tilde{\phi}_\gamma \boldsymbol{\varphi}_\gamma$;
- S5 Go to S.4 until certain convergence criteria are satisfied.

In the above pseudocode, S4.5 represents a sparse transform in a basis $(\boldsymbol{\varphi}_\gamma)_{\gamma \in \Gamma}$. It can be either orthonormal (e.g., Fourier transform) or nonorthonormal, and ϕ_γ is the corresponding coefficient in the sparse space. In S4.6, the adapted threshold μ can be estimated by a dichotomy

searching method. S.4.7 performs the inverse sparse transform. Finally, the stopping criteria for S.5 can be either the maximum iteration number is reached or the relative reconstruction error (RRE) comes under a predefined threshold [14]:

$$E^k = \frac{\|\mathbf{f}^k - \mathbf{f}^*\|}{\|\mathbf{f}^*\|} \times 100. \quad (26)$$

4. Numerical Simulation

The above-proposed algorithm was implemented in MatLab. To demonstrate its validity, we performed several numerical tests assuming a circular scanning locus of radius 57.0 cm in fan-beam geometry. The object was a 128×128 modified Shepp-Logan phantom in a compact support with a radius of 10.0 cm. We used an equispacial virtual detector array of length 20.0 cm. The detector was centered at the system origin and made perpendicular to the direction from the system origin to the X-ray source. The detector array consisted of 128 elements. The well-known ‘‘Haar’’ wavelet transform was selected to derive a sparse representation. While the pixel number of the original phantom image was 16384, there were only 1708 nonzero wavelet coefficients. In our preliminary study, the ℓ_1 -norm was focused as suggested by the CS theory. For each of our selected numbers of projections over a full-scan range, we equiangularly acquired the corresponding projection dataset based on the discrete projection model as shown in Figure 1. The stopping criterion in S.5 was defined as either the maximum iteration number 20,000 was reached or the RRE became less than 0.1%.

From each acquired dataset, we first reconstructed the image using the algorithm developed in Section 3, which is called ‘‘Scheme-A.’’ For comparison, we also run our codes without S.4.6–S.4.8. This strategy is an adapted SART-type reconstruction without regularization with the sparsity constraint, which is called ‘‘Scheme-B.’’ In reality, the real solution \mathbf{f}^* is usually unknown, Daubechies et al. suggested an interior algorithm that could slowly increase the radius of the solution in each iteration step [14]. Thus, we also modified our algorithm described in Section 3 by replacing $R(\mathbf{f}^*, p)$ as $R^k = (0.4 + 0.6(k/20000)^{0.05})R(\mathbf{f}^*, p)$ in each iteration step k to implement the corresponding version of interior algorithm, which is called ‘‘Scheme-C.’’

In the numerical simulation, after the stopping criteria were met, the iteration numbers and relative errors were listed in Table 1 with respect to different numbers of projections. The corresponding relative error convergence curves were plotted in Figure 2. The reconstructed images were shown in Figure 3. From the results in Table 1, Figures 2 and 3, we have several observations. When the number of projections was 55, Scheme-A reached a 0.1% RRE after 19040 iterations. Because 0.1% was really small, the corresponding reconstructed image would serve as a gold standard for all other reconstructed images. First of all, in any tested cases either Scheme-A or Scheme-C performed far better than Scheme-B, which confirmed that the sparse regularization did help improve the reconstructed image quality. Initially, the convergence speed of Scheme-A was

TABLE 1: Maximum iteration numbers and relative errors associated with each of three representative reconstruction schemes for different numbers of noise-free projections.

| View number | Scheme-A | Scheme-B | Scheme-C |
|-------------|-----------------|-----------------|-----------------|
| 55 | 19040 (0.1000) | 20000 (9.2080) | 20000 (0.2734) |
| 45 | 20000 (0.7689) | 20000 (10.3855) | 20000 (0.8261) |
| 35 | 20000 (4.2200) | 20000 (13.2479) | 20000 (2.9895) |
| 25 | 20000 (11.0556) | 20000 (15.9855) | 20000 (10.2940) |

TABLE 2: Same as Table 1 but for the direct implementation counterparts of the proposed algorithms.

| View number | Scheme-AD | Scheme-BD | Scheme-CD |
|-------------|-----------------|-----------------|-----------------|
| 55 | 15805 (0.1000) | 20000 (9.3391) | 20000 (0.2477) |
| 45 | 20000 (0.6837) | 20000 (10.4658) | 20000 (0.8357) |
| 35 | 20000 (4.2946) | 20000 (13.3163) | 20000 (3.1190) |
| 25 | 20000 (11.1846) | 20000 (16.1218) | 20000 (10.5271) |

faster than Scheme-C. However, after a number of iterations, the convergence speed of Scheme-A became slower than Scheme-C. If the ill posedness of the problem was not too bad, such as the cases of 55 and 45 projections, both Scheme-A and Scheme-C could perform well. When the problem was rather ill posed, such as the cases of 35 and 25 projections, Scheme-C would perform better than Scheme-A.

Compared to the original algorithm proposed by Daubechies et al., one unique feature of the proposed SART-type algorithm is the weighting functions Λ^{M+} , Λ^{+N} and the associated constant α for a more uniform converging behavior. To demonstrate this advantage, we modified our codes into a direct implementation of the algorithm described in [14] by forcing $\alpha = 1.0$ and setting Λ^{M+} and Λ^{+N} to unit diagonal matrices. The aforementioned reconstruction strategies were named as ‘‘Scheme-AD,’’ ‘‘Scheme-BD,’’ and ‘‘Scheme-CD’’ and tested, respectively. The corresponding stopping conditions were listed in Table 2 with respect to the number of projections. The relative error curves were plotted in Figure 2. It can be observed in Figure 2 that when the problem was not too under-determined, such as in the cases of 55 and 45 projections, the proposed methods did not perform significantly better, and might do even worse (e.g., Scheme-AD was actually better than Scheme-A in the case of 55 projections). When the problem was seriously under-determined, such as in the cases of 35 and 25 projections, the proposed algorithms performed better than their direct implementation counterparts.

In practical applications, measurement noise is unavoidable. It is always important to use a stable algorithm for noisy data. To test the noise characteristic and stability of the proposed algorithms, we repeated the aforementioned reconstruction tests using ‘‘Scheme-A,’’ ‘‘Scheme-B,’’ and ‘‘Scheme-C’’ with projections bearing 0.1% Gaussian noise, which are denoted as ‘‘Scheme-AN,’’ ‘‘Scheme-BN,’’ and ‘‘Scheme-CN,’’ respectively. The corresponding stopping

TABLE 3: Same as Table 1 but for projection data bearing 0.1% Gaussian noise.

| View number | Scheme-AD | Scheme-BD | Scheme-CD |
|-------------|-----------------|-----------------|-----------------|
| 55 | 20000 (1.5386) | 20000 (9.7042) | 20000 (1.5496) |
| 45 | 20000 (3.2240) | 20000 (11.4394) | 20000 (2.0746) |
| 35 | 20000 (5.2298) | 20000 (14.2091) | 20000 (3.7667) |
| 25 | 20000 (11.0959) | 20000 (16.0837) | 20000 (10.5335) |

conditions were listed in Table 3 with respect to the number of projections. The converging curves were plotted in Figure 4. The reconstructed images were in Figure 5. It can be seen from the above results that the proposed algorithms produced similar relative error curves for noisy datasets compared to the noise-free counterparts. Due to noise in the projections, all the reconstructed images from noisy datasets generally had larger RREs than those from noise-free datasets given an iteration number. Note that Scheme-AN had a better performance than “Scheme-A” in the initial iterations in the case of 55 views. Our interpretation to this phenomenon is that in the initial iterations the discrepancy $\Delta(\mathbf{f})$ may be relatively large, which implies that $\Delta(\mathbf{f})$ may dominate the total cost functional $\Phi_{w,p}(\mathbf{f})$, and the regularization effect of the ℓ_1 -norm would be small.

5. Discussions and Conclusion

Although the above CS-based reconstruction algorithms have been accelerated relative to the previous benchmark [14], the convergence speed is still slow for large-scale images and/or very ill-posed conditions. In the future, we could use the state-of-the-art computing techniques to speedup the convergence, such as ordered-subset [15] and multiscale computing [16] techniques. At the same time, we should optimize the reconstruction parameters and imaging protocols as well.

For the modified Shepp-Logan phantom, the ℓ_1 -norm seems giving the best performance than any other ℓ_p norm with $1 < p \leq 2$ in a Haar space. However, it does not imply that the ℓ_1 -norm is the best option for any application. In fact, our algorithm was implemented for any $1 \leq p \leq 2$. For a specific application, the optimal p may be studied.

Furthermore, this orthonormality of the wavelet transform used in this study is not necessary. If an image can be sparsely expanded in a certain basis or frame, the ℓ_p -norm minimization can be in principle performed to regularize the reconstruction process. Since there exist many compression methods for medical images, we should evaluate representative bases and frames for sparse representations and CS-based reconstruction methods. The heuristic rule is to achieve a minimal compression ratio.

In conclusion, we have developed a SART-type reconstruction algorithm based on the recent mathematical findings by Daubechies et al. Our preliminary simulation results have confirmed its merits and suggested research directions. Because the approach accommodates any $1 \leq p \leq 2$ and any

sparse expansion, there should be a large room for further improvements of the algorithm performance.

Acknowledgment

This work is partially supported by NIH/NIBIB Grants (EB002667, EB004287, and EB007288).

References

- [1] D. J. Brenner, C. D. Elliston, E. J. Hall, and W. E. Berdon, “Estimated risks of radiation-induced fatal cancer from pediatric CT,” *American Journal of Roentgenology*, vol. 176, no. 2, pp. 289–296, 2001.
- [2] A. B. de Gonzalez and S. Darby, “Risk of cancer from diagnostic X-rays: estimates for the UK and 14 other countries,” *The Lancet*, vol. 363, no. 9406, pp. 345–351, 2004.
- [3] D. J. Brenner and E. J. Hall, “Current concepts—computed tomography—an increasing source of radiation exposure,” *The New England Journal of Medicine*, vol. 357, no. 22, pp. 2277–2284, 2007.
- [4] O. W. Linton and F. A. Mettler Jr., “National conference on dose reduction in CT, with an emphasis on pediatric patients,” *American Journal of Roentgenology*, vol. 181, no. 2, pp. 321–329, 2003.
- [5] G. Wang, Y. Ye, and H. Yu, “Approximate and exact cone-beam reconstruction with standard and non-standard spiral scanning,” *Physics in Medicine and Biology*, vol. 52, no. 6, pp. R1–R13, 2007.
- [6] G. Wang, H. Yu, and B. De Man, “An outlook on x-ray CT research and development,” *Medical Physics*, vol. 35, no. 3, pp. 1051–1064, 2008.
- [7] D. L. Donoho, “Compressed sensing,” *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [8] E. J. Candes, J. Romberg, and T. Tao, “Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information,” *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [9] G.-H. Chen, J. Tang, and S. Leng, “Prior image constrained compressed sensing (PICCS): a method to accurately reconstruct dynamic CT images from highly undersampled projection data sets,” *Medical Physics*, vol. 35, no. 2, pp. 660–663, 2008.
- [10] H. Yu and G. Wang, “Compressed sensing based interior tomography,” *Physics in Medicine and Biology*, vol. 54, no. 9, pp. 2791–2805, 2009.
- [11] L. I. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” *Physica D*, vol. 60, no. 1–4, pp. 259–268, 1992.
- [12] G. T. Herman and R. Davidi, “Image reconstruction from a small number of projections,” *Inverse Problems*, vol. 24, no. 4, Article ID 045011, 17 pages, 2008.
- [13] I. Daubechies, M. Defrise, and C. De Mol, “An iterative thresholding algorithm for linear inverse problems with a sparsity constraint,” *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, pp. 1413–1457, 2004.
- [14] I. Daubechies, M. Fornasier, and I. Loris, “Accelerated projected gradient method for linear inverse problems with sparsity constraints,” *Journal of Fourier Analysis and Applications*, vol. 14, no. 5–6, pp. 764–792, 2008.

- [15] G. Wang and M. Jiang, "Ordered-subset simultaneous algebraic reconstruction techniques (OS-SART)," *Journal of X-Ray Science and Technology*, vol. 12, no. 3, pp. 169–177, 2004.
- [16] S. Oh, A. B. Milstein, C. A. Bouman, and K. J. Webb, "A general framework for nonlinear multigrid inversion," *IEEE Transactions on Image Processing*, vol. 14, no. 1, pp. 125–140, 2005.