CT Image Reconstruction by Spatial-Radon Domain Data-Driven Tight Frame Regularization

Ruohan Zhan* Bin Dong†

April 20, 2016

Abstract

This paper proposes a spatial-Radon domain CT image reconstruction model based on data-driven tight frames (SRD-DDTF). The proposed SRD-DDTF model combines the idea of joint image and Radon domain inpainting model of [1] and that of the data-driven tight frames for image denoising [2]. It is different from existing models in that both CT image and its corresponding high quality projection image are reconstructed simultaneously using sparsity priors by tight frames that are adaptively learned from the data to provide optimal sparse approximations. An alternative minimization algorithm is designed to solve the proposed model which is nonsmooth and nonconvex. Convergence analysis of the algorithm is provided. Numerical experiments showed that the SRD-DDTF model is superior to the model by [1] especially in recovering some subtle structures in the images.

Keywords. Computed tomography, data-driven tight frames, sparse approximation, spatial-Radon domain reconstruction.

1 Introduction

X-ray computed tomography (CT) has been widely used in clinic due to its great ability in visualizing interior structures. However, additional imaging dose to patients' healthy radiosensitive cells or organs has always been a serious clinical concern [3–5]. Low-dose CT is highly desirable if satisfactory image quality can be maintained for a specific clinical task. One commonly adopted strategy to achieve low-dose CT imaging, especially for cone beam CT (CBCT), is to reduce the total number of projections. However, this may also lead to degraded restored images if the reconstruction algorithm is not properly designed to incorporate missing information due to incomplete angular sampling and system noise. Therefore, many classical algorithms based on a complete angular sampling such as filtered back projection (FBP) [6] will generate undesirable artifacts due to lack of measurements. Other inversion techniques such as pseudo-inverse based methods [7, 8] also perform poorly at the presence of noise which is inevitable in the practice. As a consequence, a more effective and robust method is needed to achieve satisfactory reconstruction for clinical purposes. In this paper, we shall focus on the problem of low-dose planer fan beam CT reconstruction

[1]Yuanpei College, Peking University, Beijing, CHINA (zrhan@pku.edu.cn).

[2]Corresponding author. Beijing International Center for Mathematical Research, Peking University, Beijing, CHINA (dongbin@math.pku.edu.cn). Research supported in part by the Thousand Talents Plan of China.

of 2D images. However, the same modeling concept can be easily applied to 3D CBCT image reconstruction.

Assumed that the X-ray point source with a fixed milliampere-second (mAs) setting rotates along a circle centered at the object, and a linear detector array is used. CT image reconstruction can be casted as the following linear inverse problem in discrete setting

$$f = Pu + \epsilon, \tag{1.1}$$

where P is the projection matrix generated by the Sidden's Algorithm [9], f is the projection image whose rows indicate the data collected by each detector and columns indicate data collected from different projection angles, and ϵ is additive Gaussian white noise. To reduce radiation dose, one common way is to reduce the number of projection angles which leads to an under-determined (or rank deficient) linear system (or matrix P). This is the main challenge of reconstructing a desirable CT image u from its projections f via (1.1), and also the reason why traditional CT reconstruction algorithms such as FBP and pseudo-inverse based methods do not perform well.

1.1 Image Restoration Methods

In image restoration, many problems can be formulated as the same linear inverse problem (1.1) with P taking different forms for different image restoration problems. Image restoration has the same challenge as CT image reconstruction, which is the (nearly) rank deficiency of the matrix P. In fact, many CT image reconstruction methods were either originated from or motivated by methods developed for image restoration. We refer the interested readers to [10] for a review of classical and recent developments of image restoration methods.

Most of the existing models and algorithms for image restoration are transformation based. One of the most successful transformations in image restoration is the wavelet frame transform. It has been implemented with excellent results in both classical [11–18] and some more challenging image restoration problems [19–22]. Frames provide vast flexibility in designing adaptive and non-local filters with improved performance in applications [2, 23–25]. The application of wavelet frames has gone beyond image restoration. They have been successfully used in video processing [26], image segmentation [27, 28] and classifications [29, 30]. More recently, wavelet frames are constructed on non-flat domains such as surfaces [31, 32] and graphs [33–36] with applications to denoising [31, 32, 36] and graph clustering [36].

Another class of successful methods for image restoration that have been developed through a rather different path is the PDE based approach [37–39] which started with the refined total variation (TV) model [40] and anisotropic diffusion [41]. The PDE based approach includes variational and (nonlinear) PDE based methods. Both variational and PDE methods can be understood as transformation based methods as well, where the transformations are the differential operators involved in the models [42].

In recent work by [43–45], fundamental connections between wavelet frame based approach and variational methods were established. Furthermore, in [42], the authors established a generic connection between iterative wavelet frame shrinkage and general nonlinear evolution PDEs. The series of four papers [42–45] showed that wavelet frame transforms are discretization of differential operators in both variational and PDE frameworks, and such discretization is superior to some of the traditional finite difference schemes for image restoration. This new understanding essentially merged the two seemingly unrelated areas: wavelet frame base approach and PDE based approach. It also gave birth to many innovative and more effective image restoration models and algorithms.

1.2 CT Image Reconstruction Methods

The concept of sparse approximation via linear transformations originated from image restoration was also applied to CT image reconstruction due to the similarity of the problems in nature. For example, wavelet frame based methods are developed for standard CT image reconstruction [46], for 4D CT image reconstruction [47–49] and spectral CT reconstruction [50]. TV-based regularization model was also applied to CT image reconstruction in [51–56]. Many other regularization based methods for CT image reconstruction have also been introduced [57–62], as well as dictionary learning based methods [63–66].

However, all methods mentioned above attempted to recover a good CT image u with the given projection image f. Various sparsity based prior knowledge on the CT image u have been used, while the prior knowledge on f is yet to be fully exploited. The projection image f we collect using under-sampled angles will suffer from lack of angular resolution and measurement noise. Therefore, to reconstruct a high quality CT image u from (1.1), we need to restore a high quality (improved angular resolution and reduced noise) projection image f using properly chosen prior knowledge on f. Since f and u are linked by the linear inverse problem (1.1), it is more effective to restore both u and f simultaneously. Such modeling philosophy was first introduced in [1] with success, where the authors proposed the following optimization model based on sparse approximation of tight wavelet frames

$$\min_{f,u} \ \frac{1}{2}\|R_{\Lambda^c}(Pu-f)\|_2^2+\frac{1}{2}\|R_\Lambda(Pu)-f_0\|_2^2+\frac{\kappa}{2}\|R_\Lambda f-f_0\|_2^2+\lambda_1\|W_1 f\|_1+\lambda_2\|W_2 u\|_1. \quad (1.2)$$

Here, f_0 is the projection image we collect from the scanner defined on the grid Λ of size $N_D \times N_P$, where N_D is the total number of detectors and N_P is the number of angular projections. The projection image f that (1.2) tries to recover is defined on a grid $\Omega \supset \Lambda$ of size $N_D \times \widetilde{N}_P$. In this paper, we focus on the case $\widetilde{N}_P = 2N_P$, which means we want to recover a projection image f that has twice the angular resolution as that of f_0. The operator R_Λ is the restriction operator associated to the set Λ. The first three terms of (1.2) makes sure that f is consistent with f_0 on Λ and $Pu \approx f$. The last two terms are the sparsity priors assumed on u and f, where W_1 and W_2 are two (possibly different) tight wavelet frame transforms. We refer the interested readers to [1] for more details. Later in [67], simultaneous total variation regularization on the image space and sinogram was adopted for PET reconstruction.

1.3 Motivations and Organization of the Paper

Although positive results were reported in [1], the sparsity priors based on W_1 and W_2 can be further improved. It is known in the literature of image restoration that wavelet frames can sparsely approximate images or piecewise smooth functions in general. However, for a specifically given image, the sparse approximation by any pre-constructed wavelet frame system may not be ideal. This is the main reason why data-driven tight frames or bi-frames generally outperforms regular wavelet frames in image restoration [2, 24, 25]. In this paper, we propose a new model that uses data-driven tight frames as our sparsity priors for both u and f.

The rest of the paper is organized as follows. In Section 2 we review the basic knowledge of wavelet frames and data-driven tight frames. In Section 3, we introduce our spatial-Radon domain CT image reconstruction model based on data-driven tight frames, followed by an efficient algorithm and its convergence analysis. In Section 4, we present some numerical simulations, and the concluding remarks are given in Section 5 at the end.

2 Reviews and Preliminaries

2.1 Tight Wavelet Frames

In this section, we briefly introduce the concept of tight wavelet frames. The interested readers should consult [68–71] for theories of frames and wavelet frames, [10, 72] for a short survey on the theory and applications of frames, and [73] for a more detailed survey.

For a given set of functions $\Psi = \{\psi_1, \psi_2, \ldots, \psi_r\} \subset L_2(\mathbb{R})$, the quasi-affine wavelet system is defined as

$$X(\Psi) = \{\psi_{j,n,k} : \ 1 \le j \le r; n \in \mathbb{Z}, k \in \mathbb{Z}\},$$

where $\psi_{j,n,k}$ is defined by

$$\psi_{j,n,k} := \begin{cases} 2^{\frac{n}{2}} \psi_j(2^n \cdot -k), & n \ge 0; \\ 2^n \psi_j(2^n \cdot -2^n k), & n < 0. \end{cases} \tag{2.1}$$

The system $X(\Psi)$ is called a *tight wavelet frame* of $L_2(\mathbb{R})$ if

$$f = \sum_{g \in X(\Psi)} \langle f, g \rangle g$$

holds for all $f \in L_2(\mathbb{R})$, where $\langle \cdot, \cdot \rangle$ is the inner product in $L_2(\mathbb{R})$. When $X(\Psi)$ forms a tight frame of $L_2(\mathbb{R})$, each function ψ_j, $j = 1, \ldots, r$, is called a (tight) framelet and the whole system $X(\Psi)$ is called a tight wavelet frame.

The constructions of compactly supported and desirably (anti)symmetric framelets Ψ are usually based on the multiresolution analysis (MRA) generated by some refinable function ϕ with refinement mask a_0 satisfying

$$\phi = 2 \sum_{k \in \mathbb{Z}} a_0[k] \phi(2 \cdot -k).$$

The idea of an MRA-based construction of framelets $\Psi = \{\psi_1, \ldots, \psi_r\}$ is to find masks a_j, which are finite sequences (or filters), such that

$$\psi_j = 2 \sum_{k \in \mathbb{Z}} a_j[k] \phi(2 \cdot -k), \quad j = 1, 2, \ldots, r. \tag{2.2}$$

The sequences $a_1, \ldots, a_r$ are called wavelet frame masks, or the high pass filters associated to the tight wavelet frame system, and a_0 is also known as the low pass filter.

The unitary extension principle (UEP) [68] provides a rather general characterization of MRA-based tight wavelet frames. Roughly speaking, as long as $\{a_1, \ldots, a_r\}$ are finitely supported and their Fourier series satisfy

$$\sum_{j=0}^{r} |\widehat{a}_j(\xi)|^2 = 1 \quad \text{and} \quad \sum_{j=0}^{r} \widehat{a}_j(\xi) \overline{\widehat{a}_j(\xi + \pi)} = 0, \tag{2.3}$$

for all $\xi \in [-\pi, \pi]$, the quasi-affine system $X(\Psi)$ with $\Psi = \{\psi_1, \ldots, \psi_r\}$ defined by (2.2) forms a tight frame of $L_2(\mathbb{R})$. Note that, some filters used in image restoration, such as those constructed in [2, 24] and some filter banks in [42], only satisfy the first condition of (2.3). In this case, the wavelet systems associated to these filter banks are not tight frames of $L_2(\mathbb{R})$ in general. However,

these filter banks form tight frames for sequence space $\ell_2(\mathbb{Z})$ instead, which is sufficient for many image restoration problems.

In discrete setting, we denote W as the fast decomposition transform and its adjoint $W^\top$ as the fast reconstruction transform. Note that the wavelet frame transforms we use in this paper are known as the translation-invariant wavelet transform [74] whose corresponding wavelet system is the quasi-affine system (2.1) [75]. Both W and $W^\top$ can be formed by convolution operators with kernels $\{a_j\}_{j=0}^m$. Let a be a filter in $\ell_2(\mathbb{Z})$. The convolution operator $\mathcal{S}_a : \ell_2(\mathbb{Z}) \to \ell_2(\mathbb{Z})$ associated to kernel a is defined by

$$[\mathcal{S}_a u](n) := [a * u](n) = \sum_{k\in\mathbb{Z}} a(n-k)u(k).$$

Given a set of filters $\{a_j\}_{j=0}^m$, the associated analysis operator W and its adjoint $W^\top$ are given by

$$\begin{aligned} W &= [\mathcal{S}^\top_{a_0(-\cdot)}, \mathcal{S}^\top_{a_1(-\cdot)}, \ldots, \mathcal{S}^\top_{a_m(-\cdot)}]^\top, \\ W^\top &= [\mathcal{S}_{a_0}, \mathcal{S}_{a_1}, \ldots, \mathcal{S}_{a_m}]. \end{aligned} \tag{2.4}$$

It is not hard to verify that the filters $\{a_i\}_{j=0}^m$ satisfy the first condition of (2.3) if and only if

$$W^\top W = I. \tag{2.5}$$

2.2 Data-Driven Tight Frames

This subsection is to briefly review the data-driven tight frames. Interested readers should refer to [2, 24] for details.

To learn a good tight frame W, taking the form of (2.4), for a given image u, we solve the following optimization problem:

$$\min_{v,W} \quad \lambda^2\|v\|_0 + \|Wu - v\|_2^2, \quad W^\top W = I, \tag{2.6}$$

where $\|\cdot\|_0$ is the ℓ_0-"norm" that returns the number of non-zero entries of a vector.

To solve (2.6), let us start with reformulating (2.6). Reshape all $N \times N$ patches of u into vectors and put them together as column vectors of the matrix $G \in \mathbb{R}^{N^2\times p}$, where p is the total number of patches. We put the filters $\{a_j\}_{j=0}^m$ associated to W as column vectors of the matrix $D \in \mathbb{R}^{N^2\times m}$. For simplicity, we focus on the case $m = N^2$, i.e. $D \in \mathbb{R}^{N^2\times N^2}$. Denote $V \in \mathbb{R}^{N^2\times p}$ as the tight frame coefficients. So, we have

$$\begin{aligned} G &= (g_1, g_2, \ldots, g_p) \in \mathbb{R}^{N^2\times p}, \\ D &= (a_1, a_2, \ldots, a_{N^2}) \in \mathbb{R}^{N^2\times N^2}, \\ V &= (v_1, v_2, \ldots, v_p) \in \mathbb{R}^{N^2\times p}. \end{aligned} \tag{2.7}$$

Thus, the decomposition operation can be written as $V = D^T G$, and the reconstruction operation can be written as $\tilde{G} = DV$. The condition $W^\top W = I$ is satisfied whenever $DD^\top = I$. Now, we rewrite (2.6) as

$$\min_{V,D} \quad \lambda^2\|V\|_0 + \|D^T G - V\|_2^2, \quad DD^\top = I. \tag{2.8}$$

In [2], an alternative optimization algorithm is proposed to solve (2.8):

$$\begin{cases} D^{k+1} \in \underset{D,DD^T=I}{\operatorname{argmin}} \lambda^2\|V^k\|_0 + \|D^TG - V^k\|_2^2; \\ V^{k+1} \in \underset{V}{\operatorname{argmin}} \lambda^2\|V\|_0 + \|(D^{k+1})^TG - V\|_2^2. \end{cases} \tag{2.9}$$

What makes this algorithm efficient is that both the subproblems for solving D and V respectively have closed-form solutions that can be efficiently computed:

$$\begin{cases} D^{k+1} = XY^\top, \\ V^{k+1} = \mathcal{T}_\lambda((D^{k+1})^TG), \end{cases} \tag{2.10}$$

where X and Y are obtained by taking SVD of $G(V^k)^\top$, i.e. $G(V^k)^\top = X\Sigma Y^\top$, and $\mathcal{T}_\lambda$ is the hard-thresholding operator defined by

$$(\mathcal{T}_\lambda(V))[i,j] = \begin{cases} 0, & \text{if } |x| < \lambda, \\ \{0, V[i,j]\}, & \text{if } |x| = \lambda, \\ V[i,j], & \text{otherwise.} \end{cases} \tag{2.11}$$

The subsequence convergence property of the algorithm (2.10) was later provided in [24]. It has been proved in [24] that any convergent subsequence generated by algorithm (2.10) converges to a stationary point of (2.8). Furthermore, a modified algorithm that solves (2.8) was also proposed in [24] by introducing proximal terms:

$$\begin{cases} D^{k+1} \in \underset{D,DD^T=I}{\operatorname{argmin}} \lambda^2\|V^k\|_0 + \|D^TG - V^k\|_2^2 + \lambda^k\|D - D^k\|_2^2; \\ V^{k+1} \in \underset{V}{\operatorname{argmin}} \lambda^2\|V\|_0 + \|(D^{k+1})^TG - V\|_2^2 + \mu^k\|V - V^k\|_2^2, \end{cases} \tag{2.12}$$

where $\lambda^k, \mu^k \in (a,b)$ with $a, b > 0$. Both of the subproblems of (2.12) have closed-form solutions:

$$\begin{cases} D^{k+1} = XY^\top, \\ V^{k+1} = \mathcal{T}_{\lambda/\sqrt{1+\mu^k}}(\dfrac{(D^{k+1})^TG + \mu^kV^k}{1+\mu^k}), \end{cases} \tag{2.13}$$

where X and Y are obtained by taking SVD of $G(V^k)^\top + \lambda^kD^k$, i.e. $G(V^k)^\top + \lambda^kD^k = X\Sigma Y^\top$, and $\mathcal{T}_\lambda$ is the hard-thresholding operator (2.11).

Algorithm (2.12) was shown to achieve global convergence to a stationary point of (2.8) by [24] based on Kurdyka-Lojasiewicz (KL) property [76–79]. Algorithm (2.12) can be further extended to solve a wide range of nonsmooth and nonconvex optimization problems [77, 80] with guaranteed global convergence to a stationary point.

3 Models and Algorithms

3.1 CT Image Reconstruction Model

We first introduce some basic notation. Denote P_0 as the projection operator (computed using Sidden's algorithm [9]) with N_P projections and N_D detectors, and f_0 as the observed projection

image. Suppose f_0 is supported on the grid Λ of size $N_D \times N_P$, with each pixel value representing the data received from each detector at each projection angle. Given f_0, our objective is to reconstruct a projection image f with less noise and higher angular resolution than f_0, together with its corresponding high quality CT image u at the same time. Let f be supported on the grid $\Omega \supset \Lambda$ of size $N_D \times \widetilde{N}_P$ with $\widetilde{N}_p > N_p$. For simplicity, we focus on the case $\widetilde{N}_p = 2N_p$, i.e. we want to restore an f from f_0 with doubled angular resolution.

To ensure a high quality reconstruction of both u and f, we shall enforce sparsity based regularization on both of the variables. In [1], sparsity regularization based on tight wavelet frames was used and their numerical experiments showed the advantage of recovering both u and f simultaneously over the classical approach where f is fixed, i.e. setting $f = f_0$. In this paper, instead of using a pre-constructed system as sparse approximation to u and f, we adopt the idea of data-driven tight frames of [2] to actively learn the optimal sparse representation for u and f for the given data f_0. Our spatial-Radon domain CT image reconstruction model based on data-driven tight frames (SRD-DDTF) reads as follows:

$$\begin{aligned} \min_{f,u,v_1,W_1,v_2,W_2} \ & \frac{1}{2}\|R_{\Lambda^C}(Pu-f)\|_2^2 + \frac{1}{2}\|R_\Lambda Pu - f_0\|_2^2 + \frac{\kappa}{2}\|R_\Lambda f - f_0\|_2^2 \\ & + \lambda_1\|v_1\|_0 + \frac{\mu_1}{2}\|W_1 f - v_1\|_2^2 + \lambda_2\|v_2\|_0 + \frac{\mu_2}{2}\|W_2 u - v_2\|_2^2, \\ \text{s.t.} \quad & W_i^\top W_i = I, \ i = 1, 2. \end{aligned} \tag{3.1}$$

where R_{Λ^C} denotes the restriction on $\Omega \setminus \Lambda$, and R_Λ denotes the restriction on Λ.

The first two terms $\frac{1}{2}\|R_{\Lambda^C}(Pu-f)\|_2^2 + \frac{1}{2}\|R_\Lambda Pu - f_0\|_2^2$ is to ensure that $Pu \approx f$ on Λ^C and $Pu \approx f_0$ on Λ, while the third term $\frac{\kappa}{2}\|R_\Lambda f - f_0\|_2^2$ is to ensure with restriction on Λ, $f \approx f_0$. The reason that we are not using the simpler fidelity term $\frac{1}{2}\|Pu - f\|_2^2$ to enforce $Pu \approx f$ is because f is the estimated projection data which may not be as reliable as f_0 on Λ. Therefore, in the domain Λ where the actual projection image f_0 is available, we should make sure that $R_\Lambda Pu \approx f_0$.

The transforms W_1 and W_2 are tight frames (due to the constraints $W_i^\top W_i = I, i = 1, 2$), with frame coefficients v_1 and v_2, that are learned from u and f respectively. The use of the ℓ_0-"norm" is to enforce sparsity of v_1 and v_2 which in turn grants sparse approximation to u and f by the transforms W_1 and W_2. The special structure of W_i given by (2.4) and the constraints $W_i^\top W_i = I$ make the dictionary learning component of (3.1) different from the popular K-SVD method [81], where neither of the aforementioned properties is guaranteed to be satisfied. These properties we assume make the learning of W_i much faster than the K-SVD method, because the size of the problem is much smaller, while the performance is still comparable to the K-SVD method. Another drawback of the K-SVD method is that the learned dictionary is not guaranteed to be complete in the underlying Euclidean space, i.e. $W_i^\top W_i \neq I$. We refer the interested readers to [2] for more details on the comparison between data-driven tight frames and the K-SVD method.

3.2 Alternative Optimization Algorithms

Given a projected data f_0, we first solve the following analysis based model [13, 82, 83]

$$\min_u \frac{1}{2}\|P_0 u - f_0\|_2^2 + \lambda\|Wu\|_1 \tag{3.2}$$

to obtain an initial reconstruction u^0. Then, we let $f^0 = Pu^0$ to be the initial estimation of the higher quality projection image. The initial estimations on the variables v_1, W_1, v_2, W_2 are obtained

by solving the following problems

$$\min_{v_1,W_1} \quad \lambda^2\|v_1\|_0 + \|W_1u^0 - v_1\|_2^2, \quad W_1^\top W_1 = I \tag{3.3}$$

and

$$\min_{v_2,W_2} \quad \lambda^2\|v_2\|_0 + \|W_2f^0 - v_2\|_2^2, \quad W_2^\top W_2 = I \tag{3.4}$$

using algorithm (2.10). After the initializations, we optimize the variables f,u,$\{W_1, W_2\}$,$\{v_1, v_2\}$ in the SRD-DDTF model (3.1) alternatively and iterate until convergence. Full details of the proposed algorithm is given in Algorithm 1. Convergence analysis of the algorithm is given in the next subsection.

Algorithm 1: Adaptive Frames Based CT Image Reconstruction

Step 1. Initialization: Compute u^0 from (3.2) and set $f^0 = Pu^0$. Compute $v_1^0, W_1^0, v_2^0, W_2^0$ from (3.3) and (3.4).

Step 2. Main Loop:

while *stopping criteria are not met* **do**

(1) optimize f

$$f^{k+1} \leftarrow \text{argmin}_f \ \frac{\kappa}{2}\|R_\Lambda f - f_0\|_2^2 + \frac{1}{2}\|R_{\Lambda^C}(Pu^k - f)\|_2^2 + \frac{\mu_1}{2}\|W_1^k f - v_1^k\|_2^2 + \frac{a}{2}\|f - f^k\|_2^2 \tag{3.5}$$

(2) optimize u

$$u^{k+1} \leftarrow \text{argmin}_u \ \frac{1}{2}\|R_{\Lambda^C}(Pu - f^{k+1})\|_2^2 + \frac{1}{2}\|R_\Lambda Pu - f_0\|_2^2 + \frac{\mu_2}{2}\|W_2^k u - v_2^k\|_2^2 + \frac{b}{2}\|u - u^k\|_2^2 \tag{3.6}$$

(3) optimize W_1, W_2

$$\begin{aligned} W_1^{k+1} &\leftarrow \text{argmin}_{W_1^\top W_1 = I} \ \frac{\mu_1}{2}\|W_1 f^{k+1} - v_1^k\|_2^2 + \frac{c_1}{2}\|W_1 - W_1^k\|_2^2, \\ W_2^{k+1} &\leftarrow \text{argmin}_{W_2^\top W_2 = I} \ \frac{\mu_2}{2}\|W_2 u^{k+1} - v_2^k\|_2^2 + \frac{c_2}{2}\|W_2 - W_2^k\|_2^2 \end{aligned} \tag{3.7}$$

(4) optimize v_1, v_2

$$\begin{aligned} v_1^{k+1} &\leftarrow \text{argmin}_{v_1} \ \lambda_1\|v_1\|_0 + \frac{\mu_1}{2}\|W_1^{k+1} f^{k+1} - v_1\|_2^2 + \frac{d_1}{2}\|v_1 - v_1^k\|_2^2, \\ v_2^{k+1} &\leftarrow \text{argmin}_{v_2} \ \lambda_2\|v_2\|_0 + \frac{\mu_2}{2}\|W_2^{k+1} u^{k+1} - v_2\|_2^2 + \frac{d_2}{2}\|v_2 - v_2^k\|_2^2 \end{aligned} \tag{3.8}$$

Note that in step 2 of Algorithm 1 where variables are updated, we added additional ℓ_2 terms, $\frac{a}{2}\|f - f^k\|_2^2$, $\frac{b}{2}\|u - u^k\|_2^2$,$\frac{c_1}{2}\|W_1 - W_1^k\|_2^2$, $\frac{c_2}{2}\|W_2 - W_2^k\|_2^2$, $\frac{d_1}{2}\|v_1 - v_1^k\|_2^2$,$\frac{d_2}{2}\|v_2 - v_2^k\|_2^2$, so that we can theoretically justify the convergence of the algorithm. Numerically, however, Algorithm 1 still converges with $a = b = c_1 = c_2 = d_1 = d_2 = 0$.

Problem (3.5) in Algorithm 1 has the following closed-form solution:

$$f^{k+1} = (R_{\Lambda^c} + \kappa R_\Lambda + (\mu_1 + a)I)^{-1}(R_{\Lambda^c}Pu^k + \kappa R_\Lambda f_0 + \mu_1 {W_1^k}^T v_1^k + af^k),$$

where $R_{\Lambda^c} + \mu_1 R_\Lambda + (\mu_3 + a)I$ is simply a diagonal matrix and hence no matrix inversion is needed. Problem (3.6) also has a closed-form solution:

$$u^{k+1} = (P^TP + (\mu_2 + b)I)^{-1}(P^T R_{\Lambda^c} f^{k+1} + P^T R_\Lambda f_0 + \mu_2 {W_2^k}^\top {v_2}^k + bu^k),$$

which can be efficiently solved by the conjugate gradient method.

The updates on the variables v_1, W_1, v_2, W_2 can be implemented by reformulating the problem in the form (2.7) and solving them by a variant algorithm of (2.10) [24]. To be more specific, we first make the following reformulations:

$$\begin{aligned} \{f, v_1, W_1\} \quad &\Leftrightarrow \quad \{F, V_1, D_1\}, \\ \{u, v_2, W_2\} \quad &\Leftrightarrow \quad \{U, V_2, D_2\}. \end{aligned} \tag{3.9}$$

Thus, to solve problem (3.7), we can simply compute

$$\begin{cases} D_1^{k+1} = X_1 Y_1^\top, \quad \text{where } X_1 \Sigma_1 Y_1^\top = F^{k+1}(V_1^k)^\top + \dfrac{c_1}{\mu_1} D_1^k; \\ D_2^{k+1} = X_2 Y_2^\top, \quad \text{where } X_2 \Sigma_2 Y_2^\top = U^{k+1}(V_2^k)^\top + \dfrac{c_2}{\mu_2} D_2^k. \end{cases} \tag{3.10}$$

To solve problem (3.8), we can simply compute

$$\begin{aligned} V_1^{k+1} &= \mathcal{T}_{\sqrt{2\lambda_1/(\mu_1+d_1)}}((\mu_1 (D_1^{k+1})^T F^{k+1}) + d_1 V_1^k)/(\mu_1 + d_1)), \\ V_2^{k+1} &= \mathcal{T}_{\sqrt{2\lambda_2/(\mu_2+d_2)}}((\mu_2 (D_2^{k+1})^T U^{k+1}) + d_2 V_2^k)/(\mu_2 + d_2)), \end{aligned} \tag{3.11}$$

where $\mathcal{T}_a(\cdot)$ is the hard-thresholding operator defined by (2.11).

3.3 Convergence Analysis

In this subsection, we prove that under the boundedness assumption of $\{f^k, u^k\}$, the whole sequence $\{f^k, u^k, W_1^k, W_2^k, v_1^k, v_2^k\}$ generated by Algorithm 1 converges globally and the limit is a stationary point of the proposed model (3.1). Algorithm 1 is essentially a special case of the more general multi-block proximal alternating algorithm proposed by [80] solving a general nonsmooth and nonconvex optimization problem. Therefore, the global convergence of our algorithm is based on one of the results of [80]. However, our analysis starts from a weaker assumption of boundedness of $\{f^k, u^k\}$, instead of the boundedness of all variables $\{f^k, u^k, W_1^k, W_2^k, v_1^k, v_2^k\}$ which is required by the analysis of [80]. Also, we show that the sequence generated by our proposed algorithm converges to a stationary point whose definition is stronger than that of [80]. Therefore, we will present necessary convergence proofs for our algorithm and omit the details on the parts that can be directly implied by the analysis of [24, 80].

Our convergence analysis is based on the following assumption:

Assumption 3.1. *The sequence $\{u^k, f^k\}$ generated by Algorithm 1 is bounded.*

Under this assumption, we will show that:

1. Global convergence of $\{f^k, u^k, W_1^k, W_2^k, v_1^k, v_2^k\}$;
2. The limit is a stationary point of the SRD-DDTF model (3.1).

We start with some basic notation and definitions.

Definition 3.1. (Critical Point) *Let $f : \mathbb{R}^n \mapsto \mathbb{R} \cup \{\pm\infty\}$ be a proper and lower semi-continuous function.*

1. The Fréchet subdifferential of f at x is defined by

$$\partial_F f(x) := \{w \in \mathbb{R}^n : \liminf_{y \to x} \frac{f(y) - f(x) - \langle w, y - x\rangle}{\|y - x\|} \geqslant 0\}$$

for any x with $|f(x)| < \infty$ and $\partial_F f(x) = \emptyset$ if $|f(x)| = \infty$. Denote $\partial_F f = \{x : \partial_F f(x) \neq \emptyset\}$.

*2. For each $x \in domf$, x is called the **stationary point** of f if it satisfies $0 \in \partial_F f(x)$.*

Remark 3.1. *Our definition of stationary point is the same with the one used in [24], which is stronger than the definition used by [76, 77, 79, 80].*

Suppose the patches of f have size $n \times n$, and the patches of u have size $m \times m$. For simplicity, define $\mathcal{D}_1 = \{W \in R^{n^2 \times n^2} : W^\top W = I_{n^2}\}$ and $\mathcal{D}_2 = \{W \in R^{m^2 \times m^2} : W^\top W = I_{m^2}\}$. Define

$$\begin{aligned} Q(f, u, W_1, W_2, v_1, v_2) =& \frac{1}{2}\|R_{\Lambda^C}(Pu - f)\|_2^2 + \frac{1}{2}\|R_\Lambda Pu - f_0\|_2^2 + \frac{\kappa}{2}\|R_\Lambda f - f_0\|_2^2 \\ &+ \frac{\mu_1}{2}\|W_1 f - v_1\|_2^2 + \frac{\mu_2}{2}\|W_2 u - v_2\|_2^2, \end{aligned}$$

and

$$f_1(v_1) = \lambda_1\|v_1\|_0, \quad f_2(v_2) = \lambda_2\|v_2\|_0, \quad g_1(W_1) = I_{\mathcal{D}_1}(W_1), \quad g_2(W_2) = I_{\mathcal{D}_2}(W_2),$$

where $I_{\mathcal{D}}(W) = 0$, if $W \in \mathcal{D}$ and $+\infty$ otherwise.

Then, problem (3.1) can be reformulated as

$$\min_{f,u,W_1,W_2,v_1,v_2} F(f, u, W_1, W_2, v_1, v_2) := f_1(v_1) + f_2(v_2) + g_1(W_1) + g_2(W_2) + Q(f, u, W_1, W_2, v_1, v_2). \tag{3.12}$$

For convenience, let $X^k = (f^k, u^k, W_1^k, W_2^k, v_1^k, v_2^k), Q^k = Q(X^k), F^k = F(X^k)$.

Proposition 3.1. *Based on Assumption 3.1, the sequence $\{F^k\}$ and the sequence*

$$X^k = (f^k, u^k, W_1^k, W_2^k, v_1^k, v_2^k)$$

generated by Algorithm 1 are both bounded, and $\{F^k\}$ monotonically decreases to a limit point.

Proof. According to Algorithm 1, we have

$$\begin{aligned} & F(f^{k+1}, u^{k+1}, W_1^{k+1}, W_2^{k+1}, v_1^{k+1}, v_2^{k+1}) \\ \leq\ & F(f^{k+1}, u^{k+1}, W_1^{k+1}, W_2^{k+1}, v_1^{k+1}, v_2^{k+1}) + \frac{a}{2}\|f^{k+1} - f^k\|_2^2 + \frac{b}{2}\|u^{k+1} - u^k\|_2^2 \\ & + \frac{c_1}{2}\|W_1^{k+1} - W_1^k\|_2^2 + \frac{c_2}{2}\|W_2^{k+1} - W_2^k\|_2^2 + \frac{d_1}{2}\|v_1^{k+1} - v_1^k\|_2^2 + \frac{d_2}{2}\|v_2^{k+1} - v_2^k\|_2^2 \\ \leq\ & F(f^k, u^k, W_1^k, W_2^k, v_1^k, v_2^k) \\ \leq\ & \cdots \leq F(f^0, u^0, W_1^0, W_2^0, v_1^0, v_2^0). \end{aligned} \tag{3.13}$$

Therefore, the sequence $\{F^k \geq 0\}$ is bounded and monotonically decreasing, thus convergent.

By Assumption 3.1, the sequence $\{u^k, f^k\}$ is bounded. Also, $W_1^k \in \mathcal{D}_1, W_2^k \in \mathcal{D}_2$ are bounded too. Then, we have

$$\|v_1^k\|_2 - \|W_1^k f^k\|_2 \leq \|W_1^k f - v_1^k\|_2 \leq \sqrt{\frac{2F^k}{\mu_1}} \leq \sqrt{\frac{2F^0}{\mu_1}}, \quad \forall k,$$

hence v_1^k is also bounded. Similarly, v_2^k is bounded. Thus, X^k is bounded. □

It's easy to verify that our proposed algorithm is a special case of the multi-block hybrid proximal alternating method in [80]. Therefore, the global convergence of our algorithm is guaranteed by the convergence analysis of [80].

Theorem 3.2. *(Global Convergence) The sequence* $\{X^k = (f^k, u^k, W_1^k, W_2^k, v_1^k, v_2^k)\}$ *generated by Algorithm 1 is globally convergent.*

Proof. By Proposition 3.1, the sequence X^k is bounded. It is easy to see that all assumptions of [80, Theorem 3.7] are satisfied. Therefore, we have global convergence of X^k (see [80] for details). □

Lemma 3.3. *Under the Assumption 3.1, denote the limit point of* $\{X^k\}$ *generated by Algorithm 1 as* $X^* = (f^*, u^*, W_1^*, W_2^*, v_1^*, v_2^*)$*, we have*

$$\lim_{k\to\infty} f_1(v_1^k) + f_2(v_2^k) = f_1(v_1^*) + f_2(v_2^*), \text{ and } \lim_{k\to\infty} F(X^k) = F(X^*).$$

Proof. From Theorem 3.2, we have $\lim_{k\to\infty} \|X^{k-1} - X^k\|_2 = 0$ and $\lim_{k\to\infty} X^k = X^*$.

From (3.8) in Algorithm 1, we have

$$\begin{aligned}
&Q(u^k, f^k, v_1^k, W_1^k, v_2^k, W_2^k) + f_1(v_1^k) + f_2(v_2^k) + \frac{d_1}{2}\|v_1^k - v_1^{k-1}\|_2^2 + \frac{d_2}{2}\|v_2^k - v_2^{k-1}\|_2^2 \\
\leq &Q(u^k, f^k, v_1, W_1^k, v_2, W_2^k) + f_1(v_1) + f_2(v_2) + \frac{d_1}{2}\|v_1 - v_1^{k-1}\|_2^2 + \frac{d_2}{2}\|v_2 - v_2^{k-1}\|_2^2, \quad \forall v_1, v_2.
\end{aligned}$$

Replacing v_1, v_2 with v_1^*, v_2^*, and taking k to infinity, we have

$$\liminf_{k'\to\infty} f_1(v_1^k) + f_2(v_2^k) \leq f_1(v_1^*) + f_2(v_2^*). \tag{3.14}$$

Note that f_1 and f_2 are lower semi-continuous. Together with (3.14), we have

$$\liminf_{k\to+\infty} f_1(v_1^k) + f_2(v_2^k) = f_1(v_1^*) + f_2(v_2^*). \tag{3.15}$$

On the other hand, we have $W_1^k \in \mathcal{D}_1, W_2^k \in \mathcal{D}_2$. Since $\mathcal{D}_1, \mathcal{D}_2$ are compact, we have $g_1(W_1^*) = g_1(W_1^k) = g_2(W_2^*) = g_2(W_2^k) = 0$ for all k. Since Q is continuous, we have

$$\lim_{k\to+\infty} Q(u^k, f^k, v_1^k, W_1^k, v_2^k, W_2^k) = Q(u^*, f^*, v_1^*, W_1^*, v_2^*, W_2^*).$$

Note that $\{F^k\}$ is monotonically decreasing and bounded. Thus, F^k is convergent which means that the liminf in (3.15) is in fact a regular limit. Consequently, we have $\lim_{k\to\infty} F(X^k) = F(X^*)$. □

In the following, we will prove that the limit of $\{X^k\}$ is a stationary point of our SRD-DDTF model.

Theorem 3.4. (Stationary Point) *Under Assumption 3.1, the sequence* $\{X^k = (f^k, u^k, W_1^k, W_2^k, v_1^k, v_2^k)\}$ *globally converges to a stationary point of the SRD-DDTF model* (3.1).

Proof. Theorem 3.2 tells us that X^k is globally convergent. Denote its limit point by $X^* = (f^*, u^*, W_1^*, W_2^*, v_1^*, v_2^*)$ and $F(X^*) = F^*$. From Algorithm 1, we have the following four inequalities:

$$
\begin{aligned}
& Q(f^{k+1}, u^k, W_1^k, W_2^k, v_1^k, v_2^k) + \frac{a}{2}\|f^{k+1} - f^k\|_2^2 \\
\leq\ & Q(f, u^k, W_1^k, W_2^k, v_1^k, v_2^k) + \frac{a}{2}\|f - f^k\|_2^2, \quad \forall f; \\
& Q(f^{k+1}, u^{k+1}, W_1^k, W_2^k, v_1^k, v_2^k) + \frac{b}{2}\|u^{k+1} - u^k\|_2^2 \\
\leq\ & Q(f^{k+1}, u, W_1^k, W_2^k, v_1^k, v_2^k) + \frac{b}{2}\|u - u^k\|_2^2, \quad \forall u; \\
& Q(f^{k+1}, u^{k+1}, W_1^{k+1}, W_2^{k+1}, v_1^k, v_2^k) + g_1(W_1^{k+1}) + g_2(W_2^{k+1}) + \frac{c_1}{2}\|W_1^{k+1} - W_1^k\|_2^2 + \frac{c_2}{2}\|W_2^{k+1} - W_2^k\|_2^2 \\
\leq\ & Q(f^{k+1}, u^{k+1}, W_1, W_2, v_1^k, v_2^k) + g_1(W_1) + g_2(W_2) + \frac{c_1}{2}\|W_1 - W_1^k\|_2^2 + \frac{c_2}{2}\|W_2 - W_2^k\|_2^2, \quad \forall W_1, W_2; \\
& Q(f^{k+1}, u^{k+1}, W_1^{k+1}, W_2^{k+1}, v_1^{k+1}, v_2^{k+1}) + f_1(v_1^{k+1}) + f_2(v_2^{k+1}) + \frac{d_1}{2}\|v_1^{k+1} - v_1^k\|_2^2 + \frac{d_2}{2}\|v_2^{k+1} - v_2^k\|_2^2 \\
\leq\ & Q(f^{k+1}, u^{k+1}, W_1^{k+1}, W_2^{k+1}, v_1, v_2) + f_2(v_2) + f_1(v_1) + \frac{d_1}{2}\|v_1 - v_1^k\|_2^2 + \frac{d_2}{2}\|v_2 - v_2^k\|_2^2, \quad \forall v_1, v_2.
\end{aligned}
$$

Taking $k \to \infty$, we have

$$
\begin{cases}
F^* \leq F(f^* + \delta f, u^*, W_1^*, W_2^*, v_1^*, v_2^*) + \dfrac{a}{2}\|\delta f\|_2^2, \quad \forall \delta f; \\
F^* \leq F(f^*, u^* + \delta u, W_1^*, W_2^*, v_1^*, v_2^*) + \dfrac{b}{2}\|\delta u\|_2^2, \quad \forall \delta u; \\
F^* \leq F(f^*, u^*, W_1^* + \delta W_1, W_2^* + \delta W_2, v_1^*, v_2^*) + \dfrac{c_1}{2}\|\delta W_1\|_2^2 + \dfrac{c_2}{2}\|\delta W_2\|_2^2, \quad \forall \delta W_1, \delta W_2; \\
F^* \leq F(f^*, u^*, W_1^*, W_2^*, v_1^* + \delta v_1, v_2^* + \delta v_2) + \dfrac{d_1}{2}\|\delta v_1\|_2^2 + \dfrac{d_2}{2}\|\delta v_2\|_2^2, \quad \forall \delta v_1, \delta v_2.
\end{cases}
\tag{3.16}
$$

Therefore, for any $\delta X = (\delta f, \delta u, \delta W_1, \delta W_2, \delta v_1, \delta v_2)$, we have

$$
\begin{aligned}
&\liminf_{\|\delta X\|\to 0} \frac{F(X^*+\delta X)-F(X^*)}{\|\delta X\|} \\
&= \liminf_{\|\delta X\|\to 0} \frac{Q(X^*+\delta X)-Q(X^*)+(f_1(v_1^*+\delta v_1)+f_2(v_2^*+\delta v_2))-(f_1(v_1^*)+f_2(v_2^*))}{\|\delta X\|} \\
&+ \frac{(g_1(W_1^*+\delta W_1)+g_2(W_2^*+\delta W_2))-(g_1(W_1^*)+g_2(W_2^*))}{\|\delta X\|} \\
&= \liminf_{\|\delta X\|\to 0} \frac{\langle \nabla Q(X^*), \delta X\rangle + (f_1(v_1^*+\delta v_1)+f_2(v_2^*+\delta v_2))-(f_1(v_1^*)+f_2(v_2^*))}{\|\delta X\|} \\
&+ \frac{(g_1(W_1^*+\delta W_1)+g_2(W_2^*+\delta W_2))-(g_1(W_1^*)+g_2(W_2^*))+o(\|\delta X\|)}{\|\delta X\|} \\
&= \liminf_{\|\delta X\|\to 0} \Bigg(\frac{Q(f^*,u^*,W_1^*,W_2^*,v_1^*+\delta v_1,v_2^*+\delta v_2)-Q(X^*)}{\|\delta X\|} \\
&\quad + \frac{f_1(v_1^*+\delta v_1)+f_2(v_2^*+\delta v_2)-f_1(v_1^*)-f_2(v_2^*)+o(\|\delta v_1\|+\|\delta v_2\|)+o(\|\delta X\|)}{\|\delta X\|} \\
&+ \frac{Q(f^*,u^*,W_1^*+\delta W_1,W_2^*+\delta W_2,v_1^*,v_2^*)-Q(X^*)}{\|\delta X\|} \\
&\quad + \frac{g_1(W_1^*+\delta W_1)+g_2(W_2^*+\delta W_2)-g_1(W_1^*)-g_2(W_2^*)+o(\|\delta W_1\|+\|\delta W_2\|)}{\|\delta X\|} \\
&+ \frac{Q(f^*,u^*+\delta u,W_1^*,W_2^*,v_1^*,v_2^*)-Q(X^*)+o(\|\delta u\|)}{\|\delta X\|} \\
&+ \frac{Q(f^*+\delta f,u^*,W_1^*,W_2^*,v_1^*,v_2^*)-Q(X^*)+o(\|\delta f\|)}{\|\delta X\|} \Bigg) \\
&\geqslant \liminf_{\|\delta X\|\to 0} \frac{-\frac{a}{2}\|\delta f\|^2-\frac{b}{2}\|\delta u\|^2-\frac{c_1}{2}\|\delta W_1\|^2-\frac{c_2}{2}\|\delta W_2\|^2-\frac{d_1}{2}\|\delta v_1\|^2-\frac{d_2}{2}\|\delta v_2\|^2}{\|\delta X\|} \\
&+ \frac{o(\|\delta v_1\|+\|\delta v_2\|)+o(\|\delta u\|+\|\delta f\|+\|\delta W_1\|+\|\delta W_2\|)+o(\|\delta X\|)}{\|\delta X\|} \\
&= 0.
\end{aligned}
\tag{3.17}
$$

where the second and third identity follow from the first order Taylor expansion of Q and the inequality follows from the inequalities (3.16). By definition, the limit point X^* is a stationary point of our model (3.1). □

4 Numerical Experiments

It has been shown in [1] that wavelets based inpainting model (1.2) has better image restoration performance than TV-based model and wavelet analysis model. Therefore, in this section, we will focus on comparing our proposed SRD-DDTF model (3.1) with wavelet frame based model (1.2) using the same initial value given by analysis model (3.2). We show that the SRD-DDTF model can achieve noticeably better image reconstruction results.

Throughout our experiments, all data is synthesized by $f = Pu + \epsilon$, where ϵ is some Gaussian white noise. The standard deviation of noise is chosen to be $\max(|f|)/300$. In our model (3.1), we always set $\kappa = 1$, which is the same as the parameter κ in (1.2). Empirically, we observe that $\mu_1 \approx 5200\lambda_1$, $\mu_2 \approx 8400\lambda_2$ is a good choice. The patch size for u is 8×8, and the patch size for f is 8×2 to properly adapt to the shape of matrix f which has much more rows (number of detectors) than columns (number of angular projections). Parameters λ_1 and λ_2, which are the same as the parameter λ_1 and λ_2 in model (1.2), vary case by case and are chosen manually for optimal image reconstruction results. We use the analysis model (3.2) to obtain initial estimates for both our algorithm and model (1.2). We find in our experiments that the value of λ in the analysis model has small effects on the performance of both algorithms.

The experiments are conducted, with different configurations, on a real patient's image data (provided by Dr. Xun Jia from Department of Radiation Oncology, University of Texas, Southwestern Medical Center) and the popular test data set NURBS-based cardiac-torso (NCAT) phantom [84]. We shall refer to the former simply as "head" and the latter as "NCAT". In addition to visual observation given by Figure 1 and Figure 3, we use relative error, correlation and computation time to quantify the quality of the model (1.2) and our SRD-DDTF model (3.1) (see Table 1). The relative error and correlation for reconstructed u based on ground truth image u_t are defined as follows:

$$err(u_t, u) = \frac{\|u - u_t\|_2}{\|u_t\|_2}, \quad corr(u_t, u) = \frac{(u - \bar{u})(u_t - \bar{u}_t)}{\|u - \bar{u}\|_2 \|u_t - \bar{u}_t\|_2} \tag{4.1}$$

where $\bar{u}$, $\bar{u}_t$ denote the mean value of u and u_t. As we can see from both Figure 1, Figure 3 and Table 1 that our SRD-DDTF model (3.1) managed to achieve better image reconstruction results than the model (1.2) of [1] for all configurations.

The stopping criterion we used takes the form $err(u^k, u^{k-1}) \leq \epsilon$ together with a maximum allowable iteration 1000, where $\epsilon = 0.001$ for the image "head" and $\varepsilon = 0.005$ for "NCAT". We consider the configurations with $N_P = 15, 30, 45, 60$ for "head" and $N_P = 60, 75, 90$ for "NCAT". Table 1 shows that the SRD-DDTF model achieves noticeably better reconstruction with less relative errors and higher correlations based on ground truth images. It is remarkable that for "NCAT", the results of the SRD-DDTF model can even achieve better image restoration than that of model (1.2) using the next larger projection number. The reconstructed image for each configuration is shown in Figure 1 and Figure 3, and it is worth noticing that the SRD-DDTF model is managed to recover some key structures that are lost by model (1.2). In particular, we list some zoom-in views of the results in figure 2 for $N_p = 15$ to show that our model is capable of restoring subtle features.

5 Conclusion

In this paper, we proposed a new spatial-Radon domain CT image reconstruction model based on data-driven tight frames (SRD-DDTF), together with an efficient alternative minimization algorithm. Our convergence analysis on the proposed algorithm indicated that, under suitable assumptions, the sequence generated by the algorithm converges to a stationary point of the proposed model. Our numerical experiments showed that our model (3.1) can obtain noticeably better reconstruction results than those from the model (1.2), which showed that using data-driven tight frames as sparsity priors for both CT image and the projection image performs better than using pre-determined systems, such as tight wavelet frames, that may not be ideal for a specifically given image data.

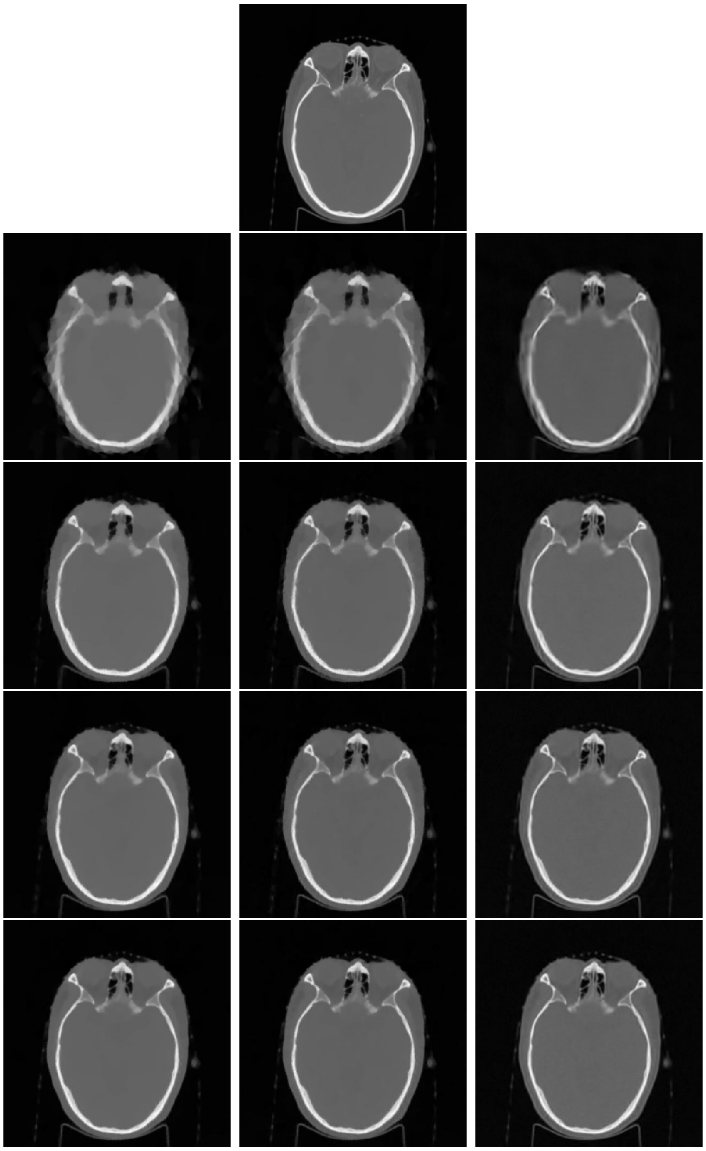

Figure 1: The tomographic results for the image "head". The image on the top is the true data. The following rows represent the results using 15,30,45,60 projections, respectively. Images from left to right in each row are results from initial value, wavelets based inpainting model and our SRD-DDTF model.

"Head"								
NP	initial value		Model (1.2) of [1]			SRD-DDTF model (3.1)		
	err	corr	err	corr	time	err	corr	time
15	14.09	98.29	12.70	98.61	288.91	10.50	99.05	629.92
30	6.79	99.61	6.25	99.67	1186.54	5.39	99.75	930.65
45	5.20	99.77	4.70	99.81	1550.63	4.24	99.85	1389.01
60	4.16	99.85	3.89	99.87	319.29	3.58	99.89	1785.50
"NCAT"								
NP	initial value		Model (1.2) of [1]			SRD-DDTF model (3.1)		
	err	corr	err	corr	time	err	corr	time
60	9.55	99.35	5.00	99.82	239.47	4.23	99.87	955.67
75	9.02	99.42	4.61	99.85	296.71	3.65	99.90	1681.94
90	8.81	99.45	4.21	99.87	303.98	3.41	99.91	2339.88

Table 1: Comparison of relative errors (in percentage), correlations (in percentage) and running time (in seconds).

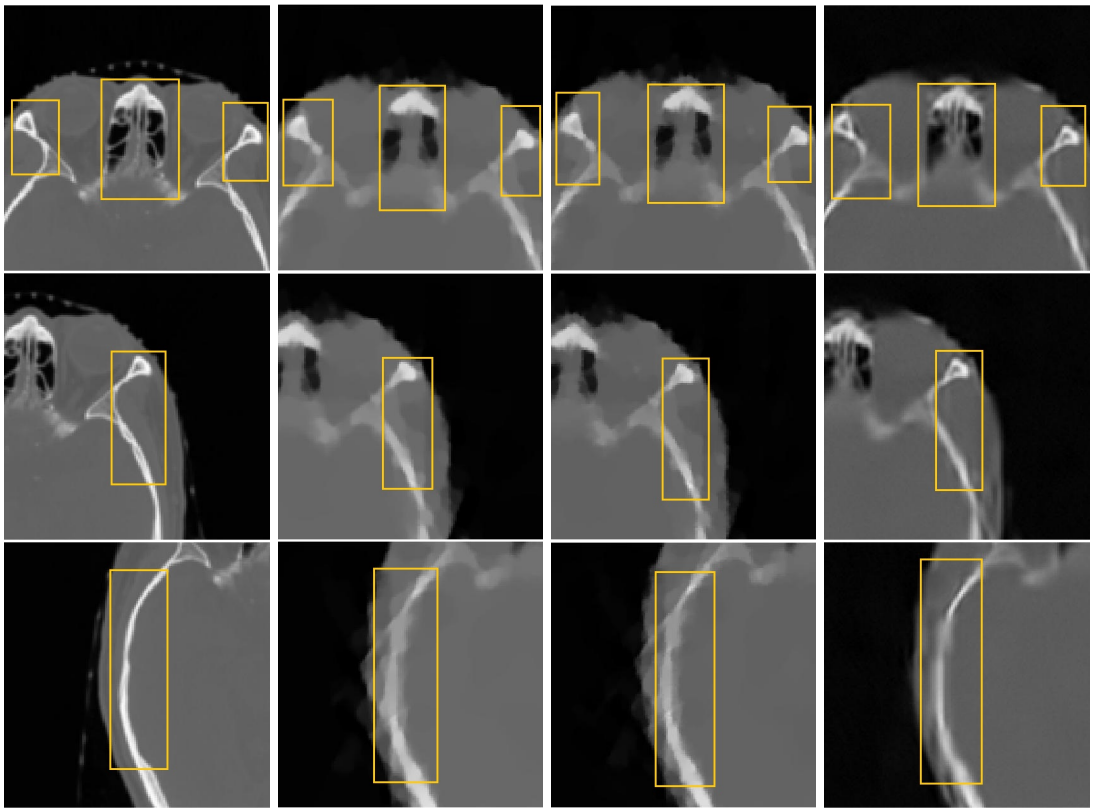

Figure 2: Local comparison of restored images for the image "head" with $N_p = 15$. Images from left to right in each row are zoom-in patterns from the ground truth, the initial value, wavelets based inpainting model restored image, and our SRD-DDTF model.

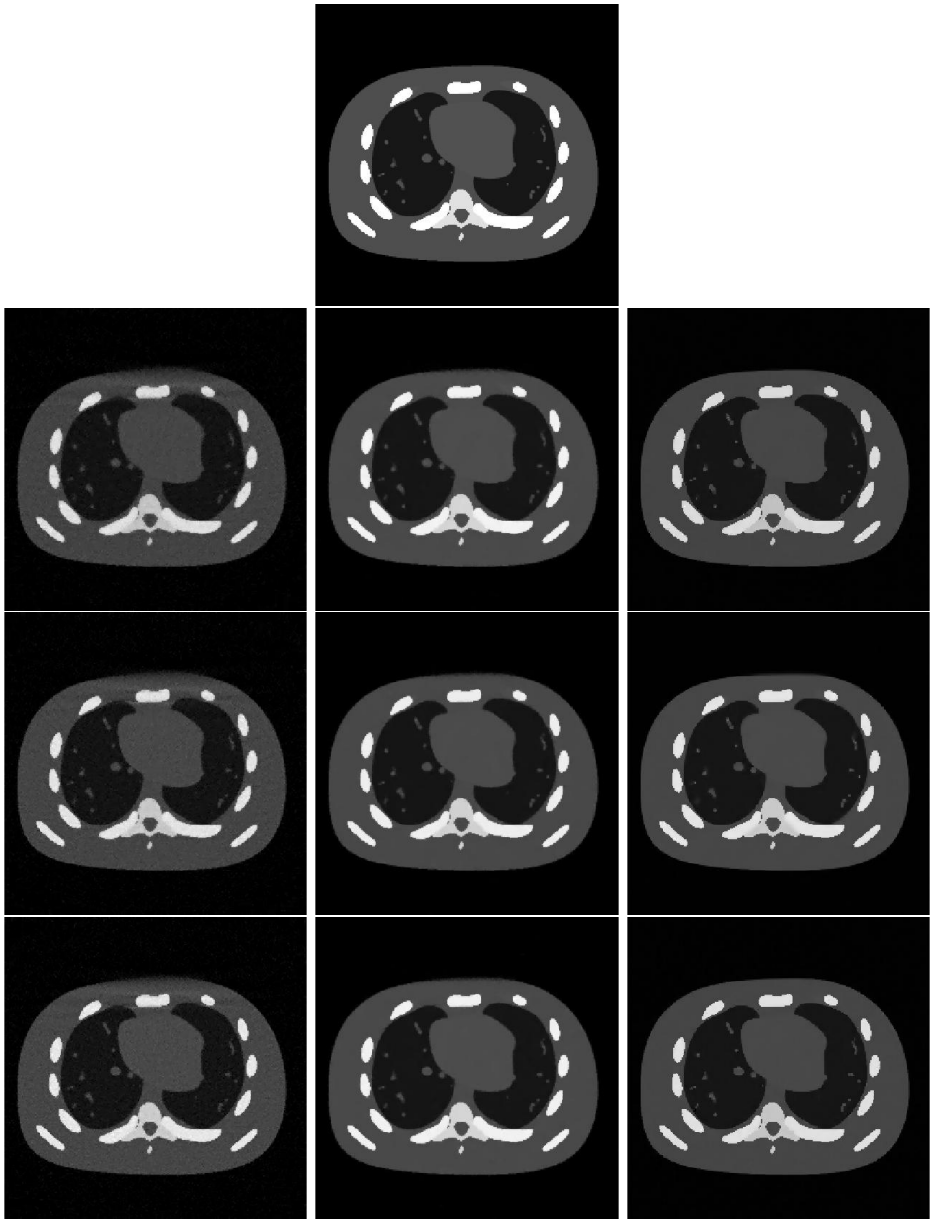

Figure 3: The tomographic results for "NCAT". The image on the top is the true data. The following rows represent the results using 60,75,90 projections, respectively. Images from left to right in each row are results from initial value, wavelets based inpainting model and our SRD-DDTF model.

References

[1] B. Dong, J. Li, and Z. Shen, "X-ray ct image reconstruction via wavelet frame based regularization and radon domain inpainting," *Journal of Scientific Computing*, vol. 54, no. 2-3, pp. 333–349, 2013.

[2] J. F. Cai, H. Ji, Z. Shen, and G. B. Ye, "Data-driven tight frame construction and image denoising," *Applied & Computational Harmonic Analysis*, vol. 37, no. 1, p. 89C105, 2014.

[3] M. K. Islam, B. D. Purdie TGNorrlinger, H. Alasti, D. J. Moseley, M. B. Sharpe, J. H. Siewerdsen, and D. A. Jaffray, "Patient dose from kilovoltage cone beam computed tomography imaging in radiation therapy," *Medical Physics*, vol. 33, no. 6, pp. 1573–1582, 2006.

[4] D. J. Brenner and E. J. Hall, "Computed tomographyan increasing source of radiation exposure," *New England Journal of Medicine*, vol. 357, no. 22, pp. 2277–2284, 2007.

[5] G. X. Ding and C. W. Coffey, "Radiation dose from kilovoltage cone beam computed tomography in an image-guided radiotherapy procedure," *International Journal of Radiation Oncology Biology Physics*, vol. 73, no. 2, p. 610C617, 2009.

[6] A. Katsevich, "Theoretically exact filtered backprojection-type inversion algorithm for spiral ct," *Siam Journal on Applied Mathematics*, vol. 62, no. 6, pp. pgs. 2012–2026, 2002.

[7] Z. G. P. J. Antsaklis, "Stability of the pseudo-inverse method for reconfigurable control systems," *International Journal of Control*, vol. 53, no. 3, pp. 717–729, 1991.

[8] A. N. Tikhonov and V. Y. Arsenin, "Solutions of ill-posed problems," *Mathematics of Computation*, vol. 32, no. 5, pp. 491–491, 1978.

[9] R. Siddon, "Fast calculation of the exact radiological path for a three-dimensional CT array," *Medical Physics*, vol. 12, p. 252, 1985.

[10] B. Dong and Z. Shen, "Image restoration: a data-driven perspective.," *Proceedings of the International Congress of Industrial and Applied Mathematics (ICIAM)*, pp. 65–108, 2015.

[11] J. Cai, R. Chan, and Z. Shen, "A framelet-based image inpainting algorithm," *Applied and Computational Harmonic Analysis*, vol. 24, no. 2, pp. 131–149, 2008.

[12] R. Chan, T. Chan, L. Shen, and Z. Shen, "Wavelet algorithms for high-resolution image reconstruction," *SIAM Journal on Scientific Computing*, vol. 24, no. 4, pp. 1408–1432, 2003.

[13] J. Cai, S. Osher, and Z. Shen, "Split Bregman methods and frame based image restoration," *Multiscale Modeling and Simulation: A SIAM Interdisciplinary Journal*, vol. 8, no. 2, pp. 337–369, 2009.

[14] J. Cai, S. Osher, and Z. Shen, "Linearized Bregman iterations for frame-based image deblurring," *SIAM J. Imaging Sci*, vol. 2, no. 1, pp. 226–252, 2009.

[15] Y. Zhang, B. Dong, and Z. Lu, "ℓ_0 minimization of wavelet frame based image restoration," *Mathematics of Computation*, vol. 82, pp. 995–1015, 2013.

[16] B. Dong and Y. Zhang, "An efficient algorithm for ℓ_0 minimization in wavelet frame based image restoration," *Journal of Scientific Computing*, vol. 54 (2-3), pp. 350–368, 2013.

[17] J. Liang, J. Li, Z. Shen, and X. Zhang, "Wavelet frame based color image demosaicing," *Inverse Problems and Imaging*, vol. 7, no. 3, pp. 777–794, 2013.

[18] L. Hou, H. Ji, and Z. Shen, "Recovering over-/underexposed regions in photographs.," *SIAM J. Imaging Sciences*, vol. 6, no. 4, pp. 2213–2235, 2013.

[19] J. Cai, H. Ji, C. Liu, and Z. Shen, "Blind motion deblurring using multiple images," *Journal of Computational Physics*, vol. 228, no. 14, pp. 5057–5071, 2009.

[20] J. Cai, H. Ji, C. Liu, and Z. Shen, "Blind motion deblurring from a single image using sparse approximation," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pp. 104–111, IEEE, 2009.

[21] B. Dong, H. Ji, J. Li, Z. Shen, and Y. Xu, "Wavelet frame based blind image inpainting," *Applied and Computational Harmonic Analysis*, vol. 32, no. 2, pp. 268–279, 2012.

[22] Z. Gong, Z. Shen, and K.-C. Toh, "Image restoration with mixed or unknown noises," *Multiscale Modeling & Simulation*, vol. 12, no. 2, pp. 458–487, 2014.

[23] Y. Quan, H. Ji, and Z. Shen, "Data-driven multi-scale non-local wavelet frame construction and image recovery," *Journal of Scientific Computing*, pp. 1–23, 2014.

[24] C. Bao, H. Ji, and Z. Shen, "Convergence analysis for iterative data-driven tight frame construction scheme," *Applied and Computational Harmonic Analysis*, 2014.

[25] C. Tai and W. E, "Multiscale adaptive representation of signals: I. the basic framework," *Preprint*, 2015.

[26] H. Ji, S. Huang, Z. Shen, and Y. Xu, "Robust video restoration by joint sparse and low rank matrix approximation," *SIAM Journal on Imaging Sciences*, vol. 4, no. 4, pp. 1122–1142, 2011.

[27] B. Dong, A. Chien, and Z. Shen, "Frame based segmentation for medical images," *Communications in Mathematical Sciences*, vol. 9(2), pp. 551–559, 2010.

[28] C. Tai, X. Zhang, and Z. Shen, "Wavelet frame based multiphase image segmentation," *SIAM Journal on Imaging Sciences*, vol. 6, no. 4, pp. 2521–2546, 2013.

[29] H. Wendt, P. Abry, S. Jaffard, H. Ji, and Z. Shen, "Wavelet leader multifractal analysis for texture classification," in *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pp. 3829–3832, IEEE, 2009.

[30] C. Bao, H. Ji, Y. Quan, and Z. Shen, "ℓ_0 norm based dictionary learning by proximal methods with global convergence," in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pp. 3858–3865, IEEE, 2014.

[31] Q. Jiang and D. K. Pounds, "Highly symmetric bi-frames for triangle surface multiresolution processing," *Applied and Computational Harmonic Analysis*, vol. 31, no. 3, pp. 370–391, 2011.

[32] B. Dong, Q. Jiang, C. Liu, and Z. Shen, "Multiscale representation of surfaces by tight wavelet frames with applications to denoising," *Applied and Computational Harmonic Analysis, doi:10.1016/j.acha.2015.03.005*, 2015.

[33] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 129–150, 2011.

[34] M. Gavish, B. Nadler, and R. R. Coifman, "Multiscale wavelets on trees, graphs and high dimensional data: Theory and applications to semi supervised learning," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pp. 367–374, 2010.

[35] N. Leonardi and D. Van De Ville, "Tight wavelet frames on multislice graphs," *Signal Processing, IEEE Transactions on*, vol. 61, no. 13, pp. 3357–3367, 2013.

[36] B. Dong, "Sparse representation on graphs by tight wavelet frames and applications," *Applied and Computational Harmonic Analysis, doi:10.1016/j.acha.2015.09.005*, 2015.

[37] G. Sapiro, *Geometric partial differential equations and image analysis*. Cambridge University Press, 2001.

[38] S. Osher and R. Fedkiw, *Level set methods and dynamic implicit surfaces*. Springer, 2003.

[39] T. Chan and J. Shen, *Image processing and analysis: variational, PDE, wavelet, and stochastic methods*. Society for Industrial Mathematics, 2005.

[40] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*, vol. 60, pp. 259–268, 1992.

[41] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, 1990.

[42] B. Dong, Q. Jiang, and Z. Shen, "Image Restoration: Wavelet Frame Shrinkage, Nonlinear Evolution PDEs, and Beyond," *UCLA CAM Report*, vol. 13-78, 2013.

[43] J. Cai, B. Dong, S. Osher, and Z. Shen, "Image restorations: total variation, wavelet frames and beyond," *Journal of American Mathematical Society*, vol. 25(4), pp. 1033–1089, 2012.

[44] J. Cai, B. Dong, and Z. Shen, "Image restorations: a wavelet frame based model for piecewise smooth functions and beyond," *Applied and Computational Harmonic Analysis*, 2015. http://dx.doi.org/10.1016/j.acha.2015.06.009.

[45] B. Dong, Z. Shen, and P. Xie, "Image restoration: a general wavelet frame based model and its asymptotic analysis," *preprint*, 2016.

[46] X. Jia, B. Dong, Y. Lou, and S. Jiang, "GPU-based iterative cone-beam CT reconstruction using tight frame regularization," *Physics in Medicine and Biology*, vol. 56, pp. 3787–3807, 2011.

[47] H. Gao, J.-F. Cai, Z. Shen, and H. Zhao, "Robust principal component analysis-based four-dimensional computed tomography," *Physics in medicine and biology*, vol. 56, no. 11, p. 3181, 2011.

[48] H. Gao, R. Li, Y. Lin, and L. Xing, "4d cone beam ct via spatiotemporal tensor framelet," *Medical physics*, vol. 39, no. 11, pp. 6943–6946, 2012.

[49] J. Cai, X. Jia, H. Gao, S. Jiang, Z. Shen, and H. Zhao, "Cine cone beam ct reconstruction using low-rank matrix factorization: Algorithm and a proof-of-principle study.," *IEEE transactions on medical imaging*, vol. 33, no. 8, pp. 1581–1591, 2014.

[50] B. Zhao, H. Gao, H. Ding, and S. Molloi, "Tight-frame based iterative image reconstruction for spectral breast ct," *Medical physics*, vol. 40, no. 3, p. 031905, 2013.

[51] E. Sidky, C. Kao, and X. Pan, "Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT," *Journal of X-Ray Science and Technology*, vol. 14, no. 2, pp. 119–139, 2006.

[52] E. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Physics in medicine and biology*, vol. 53, p. 4777, 2008.

[53] X. Y. Jia, R. Li, W. Y. Song, and S. B. Jiang, "Gpu-based fast cone beam ct reconstruction from undersampled and noisy projection data via total variation.," *Medical Physics*, vol. 37, no. 4, pp. 1757–1760, 2010.

[54] X. Zhang, M. Burger, X. Bresson, and S. Osher, "Bregmanized nonlocal regularization for deconvolution and sparse reconstruction," *SIAM Journal on Imaging Sciences*, vol. 3, pp. 253–276, 2010.

[55] Z. Chen, X. Jin, L. Li, and G. Wang, "A limited-angle ct reconstruction method based on anisotropic tv minimization," *Physics in medicine and biology*, vol. 58, no. 7, p. 2119, 2013.

[56] Z. Han-Ming, W. Lin-Yuan, Y. Bin, L. Lei, X. Xiao-Qi, and L. Li-Zhong, "Image reconstruction based on total-variation minimization and alternating direction method in linear scan computed tomography," *Chinese Physics B*, vol. 22, no. 7, p. 078701, 2013.

[57] J.-B. Thibault, K. D. Sauer, C. A. Bouman, and J. Hsieh, "A three-dimensional statistical approach to improved image quality for multislice helical ct," *Medical physics*, vol. 34, no. 11, pp. 4526–4544, 2007.

[58] J. Wang, T. Li, and L. Xing, "Iterative image reconstruction for cbct using edge-preserving prior," *Medical physics*, vol. 36, no. 1, pp. 252–260, 2009.

[59] J. Tang, B. E. Nett, and G.-H. Chen, "Performance comparison between total variation (tv)-based compressed sensing and statistical iterative reconstruction algorithms," *Physics in medicine and biology*, vol. 54, no. 19, p. 5781, 2009.

[60] L. Ouyang, T. Solberg, and J. Wang, "Effects of the penalty on the penalized weighted least-squares image reconstruction for low-dose cbct," *Physics in medicine and biology*, vol. 56, no. 17, p. 5535, 2011.

[61] H. Lee, L. Xing, R. Davidi, R. Li, J. Qian, and R. Lee, "Improved compressed sensing-based cone-beam ct reconstruction using adaptive prior image constraints," *Physics in medicine and biology*, vol. 57, no. 8, p. 2287, 2012.

[62] P. T. Lauzier and G.-H. Chen, "Characterization of statistical prior image constrained compressed sensing (piccs): Ii. application to dose reduction," *Medical physics*, vol. 40, no. 2, p. 021902, 2013.

[63] Q. Xu, H. Yu, X. Mou, L. Zhang, J. Hsieh, and G. Wang, "Low-dose x-ray ct reconstruction via dictionary learning," *Medical Imaging, IEEE Transactions on*, vol. 31, no. 9, pp. 1682–1697, 2012.

[64] Y. Chen, X. Yin, L. Shi, H. Shu, L. Luo, J.-L. Coatrieux, and C. Toumoulin, "Improving abdomen tumor low-dose ct images using a fast dictionary learning based processing," *Physics in medicine and biology*, vol. 58, no. 16, p. 5803, 2013.

[65] W. Zhou, J.-F. Cai, and H. Gao, "Adaptive tight frame based medical image reconstruction: a proof-of-concept study for computed tomography," *Inverse problems*, vol. 29, no. 12, p. 125006, 2013.

[66] Y. Chen, L. Shi, Q. Feng, J. Yang, H. Shu, L. Luo, J.-L. Coatrieux, and W. Chen, "Artifact suppressed dictionary learning for low-dose ct image processing," *Medical Imaging, IEEE Transactions on*, vol. 33, no. 12, pp. 2271–2292, 2014.

[67] M. Burger, J. Müller, E. Papoutsellis, and C.-B. Schönlieb, "Total variation regularization in measurement and image space for pet reconstruction," *Inverse Problems*, vol. 30, no. 10, p. 105003, 2014.

[68] A. Ron and Z. Shen, "Affine systems in $L_2(\mathbb{R}^d)$: The analysis of the analysis operator," *Journal of Functional Analysis*, vol. 148, no. 2, pp. 408–447, 1997.

[69] A. Ron and Z. Shen, "Affine systems in $L_2(\mathbb{R}^d)$ ii: dual systems," *Journal of Fourier Analysis and Applications*, vol. 3, no. 5, pp. 617–638, 1997.

[70] I. Daubechies, *Ten lectures on wavelets*, vol. CBMS-NSF Lecture Notes, SIAM, nr. 61. Society for Industrial and Applied Mathematics, 1992.

[71] I. Daubechies, B. Han, A. Ron, and Z. Shen, "Framelets: MRA-based constructions of wavelet frames," *Applied and Computational Harmonic Analysis*, vol. 14, pp. 1–46, Jan 2003.

[72] Z. Shen, "Wavelet frames and image restorations," in *Proceedings of the International Congress of Mathematicians*, vol. 4, pp. 2834–2863, 2010.

[73] B. Dong and Z. Shen, "MRA-Based Wavelet Frames and Applications," *IAS Lecture Notes Series, Summer Program on "The Mathematics of Image Processing", Park City Mathematics Institute*, 2010.

[74] R. Coifman and D. Donoho, "Translation-invariant de-noising," *Wavelets and statistics*, vol. 103, p. 125, 1995.

[75] A. Chai and Z. Shen, "Deconvolution: A wavelet frame approach," *Numerische Mathematik*, vol. 106, no. 4, pp. 529–587, 2007.

[76] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, "Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the kurdyka-?ojasiewicz inequality," *Mathematics of Operations Research*, vol. 35, no. 2, pp. 438–457, 2010.

[77] J. Bolte, S. Sabach, and M. Teboulle, "Proximal alternating linearized minimization for nonconvex and nonsmooth problems," *Mathematical Programming*, vol. 146, no. 1-2, pp. 459–494, 2014.

[78] Y. Xu and W. Yin, "A globally convergent algorithm for nonconvex optimization based on block coordinate update," *arXiv preprint arXiv:1410.1386*, 2014.

[79] H. Attouch, J. Bolte, and B. F. Svaiter, "Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forwardbackward splitting, and regularized gaussseidel methods," *Mathematical Programming*, vol. 137, no. 1-2, pp. 91–129, 2011.

[80] C. Bao, H. Ji, Y. Quan, and Z. Shen, "Dictionary learning for sparse coding: Algorithms and analysis," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, pp. 1–1, 2015.

[81] M. Aharon, M. Elad, and A. Bruckstein, "K-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE TRANSACTIONS ON SIGNAL PROCESSING*, vol. 54, no. 11, p. 4311, 2006.

[82] M. Elad, J. Starck, P. Querre, and D. Donoho, "Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)," *Applied and Computational Harmonic Analysis*, vol. 19, no. 3, pp. 340–358, 2005.

[83] J. Starck, M. Elad, and D. Donoho, "Image decomposition via the combination of sparse representations and a variational approach," *IEEE transactions on image processing*, vol. 14, no. 10, pp. 1570–1582, 2005.

[84] W. P. Segars, D. S. Lalush, and B. M. W. Tsui, "Development of an interactive software application to model patient populations in the 4d nurbs-based cardiac torso phantom," in *Nuclear Science Symposium Conference Record, 2000 IEEE*, pp. 20/51–20/55 vol.3, 2000.