Sparse-View X-Ray CT Reconstruction Using ℓ_1 Regularization with Learned Sparsifying Transform

Il Yong Chun, Xuehang Zheng, Yong Long*, and Jeffrey A. Fessler

***Abstract*—A major challenge in X-ray computed tomography (CT) is to reduce radiation dose while maintaining high quality of reconstructed images. To reduce the radiation dose, one can reduce the number of projection views (sparse-view CT); however, it becomes difficult to achieve high quality image reconstruction as the number of projection views decreases. Researchers have shed light on applying the concept of learning sparse representations from (high-quality) CT image dataset to the sparse-view CT reconstruction. We propose a new statistical CT reconstruction model that combines penalized weighted-least squares (PWLS) and ℓ_1 regularization with learned sparsifying transform (PWLS-ST-ℓ_1), and an algorithm for PWLS-ST-ℓ_1. Numerical experiments for sparse-view CT show that our model significantly improves the sharpness of edges of reconstructed images compared to the CT reconstruction methods using edge-preserving hyperbola regularizer and ℓ_2 regularization with learned ST.**

I. Introduction

Radiation dose reduction is a major challenge in X-ray computed tomography (CT). Sparse-view CT reduces dose by acquiring fewer projection views [1], [2]. However, as the number of projection views decreases, it becomes harder to achieve high quality (high resolution, contrast, and signal-to-noise ratio) image reconstruction. There have been extensive studies for sparse-view CT reconstruction with total variation [3], [4] or other sparsity promoting regularizers [1], [2]. This paper investigates learned sparsifying transforms for regularization.

Learning prior information from big datasets of CT images and exploiting it for CT reconstruction is a fascinating idea. In particular, patch-based sparse representation learning frameworks [5], [6] have been successfully applied to improve low-dose CT reconstruction [7], [8]. However, CT reconstruction with a ℓ_2 regularizer using a learned sparsifying transform (ST) had difficulty in reconstructing sharp edges [8].

This paper proposes *1)* a new (statistical) CT reconstruction model that combines penalized weighted-least squares (PWLS) and ℓ_1 regularization with learned ST (PWLS-ST-ℓ_1) and *2)* a corresponding algorithm based on Alternating Direction Method of Multipliers (ADMM) [9]. Numerical experiments with the XCAT phantom show that, for sparse-view CT, the proposed PWLS-ST-ℓ_1 model significantly improves the edge sharpness of reconstructed images compared to a PWLS reconstruction method with an edge-preserving (EP) hyperbola regularizer (PWLS-EP) and to ℓ_2 regularization with a learned ST (PWLS-ST-ℓ_2 [8]).

The first two authors contributed equally to this work. This work is supported in part by UM-SJTU Collaborative Research Program, NSFC (61501292), Shanghai Pujiang Talent Program (15PJ1403900), and NIH Grant U01 EB018753. *Asterisk indicates the corresponding author.*

Il Yong Chun and Jeffrey A. Fessler are with the Department of Electrical Engineering and Computer Science, The University of Michigan, Ann Arbor, MI 48019 USA (email: {iychun, fessler}@umich.edu). Xuehang Zheng and Yong Long are with the University of Michigan - Shanghai Jiao Tong University Joint Institute, Shanghai Jiao Tong University, Shanghai 200240, China (email: {zhxhang, yong.long}@sjtu.edu.cn).

II. Methods

A. *Offline Learning Sparsifying Transform*

We pre-learn a ST by solving the following problem [6]:

$$\min_{\substack{\Psi \in \mathbb{R}^{n\times n},\\ \{\mathbf{z}'_j \in \mathbb{R}^n\}}} \sum_{j=1}^{J'} \left\| \Psi \mathbf{x}'_j - \mathbf{z}'_j \right\|_2^2 + \gamma' \left\| \mathbf{z}'_j \right\|_0 + \tau \left(\xi \|\Psi\|_F^2 - \log |\det \Psi| \right) \tag{1}$$

where $\Psi \in \mathbb{R}^{n\times n}$ is a square ST, $\{\mathbf{x}'_j \in \mathbb{R}^n : j = 1,\ldots,J'\}$ is a set of patches extracted from training data, $\mathbf{z}'_j \in \mathbb{R}^n$ is the sparse code corresponding to the jth patch $\mathbf{x}'_j$, J' is the total number of the image patches, and $\gamma', \tau, \xi \in \mathbb{R}$ are regularization parameters. The ℓ_0 function $\|\cdot\|_0$ counts the nonzero elements in a vector.

B. *CT Reconstruction Model Using ℓ_1-Regularization with Learned Sparsifying Transform: PWLS-ST-ℓ_1*

To reconstruct a linear attenuation coefficient image $\mathbf{x} \in \mathbb{R}^N$ from post-log measurement $\mathbf{y} \in \mathbb{R}^m$ [2], [10], we solve the following non-convex optimization problem using PWLS and the ST Ψ learned via (1):[a]

$$\min_{\mathbf{x},\mathbf{z} \in \mathbb{R}^{nJ}} \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_{\mathbf{W}}^2 + \lambda \left\| \widetilde{\Psi}\mathbf{x} - \mathbf{z} \right\|_1 + \gamma \|\mathbf{z}\|_0 , \tag{2}$$

where

$$\widetilde{\Psi} = \begin{bmatrix} \Psi \mathbf{P}_1 \\ \vdots \\ \Psi \mathbf{P}_J \end{bmatrix} \quad \text{and} \quad \mathbf{z} = \begin{bmatrix} \mathbf{z}_1 \\ \vdots \\ \mathbf{z}_J \end{bmatrix}.$$

Here, $\mathbf{A} \in \mathbb{R}^{m\times N}$ is a CT scan system matrix, $\mathbf{W} \in \mathbb{R}^{m\times m}$ is a diagonal weighting matrix with elements $\{W_{l,l} = \rho_l^2/(\rho_l + \sigma^2) : l = 1,\ldots,m\}$ based on a Poisson-Gaussian model for the pre-log measurements $\boldsymbol{\rho} \in \mathbb{R}^m$ with electronic readout noise variance σ^2 [2], [12], $\mathbf{P}_j \in \mathbb{R}^{n\times N}$ is a patch-extraction operator for the jth patch, $\mathbf{z}_j \in \mathbb{R}^n$ is unknown sparse code for the jth patch, J is the number of extracted patches, and $\lambda, \gamma \in \mathbb{R}$ are regularization parameters.

The term $\|\widetilde{\Psi}\mathbf{x} - \mathbf{z}\|_1$ denotes an ℓ_1-based sparsification error [13]. We expect ℓ_1 to be more robust to model mismatch than the ℓ_2-based sparsification error used in [8]. In particular,

[a]In [11], a similar approach is introduced with a "dictionary" (or synthesis) perspective and solved by a reweighted-ℓ_2 minimization. Here we directly attack ℓ_1 minimization—see Section II-C.

DOI: 10.12059/Fully3D.2017-11-3109002

the proposed ℓ_1-based sparsification error term, $\|\widetilde{\mathbf{\Psi}}\mathbf{x}-\mathbf{z}\|_1$, preserves edge sharpness better than $\|\widetilde{\mathbf{\Psi}}\mathbf{x}-\mathbf{z}\|_2^2$ in [8]; see Fig. 1 and Table I.

C. Proposed Algorithm for PWLS-ST-ℓ_1

To solve (2), our proposed algorithm alternates between updating the image $\mathbf{x}$ (*image update step*) and the sparse codes $\mathbf{z}$ (*sparse coding step*). For the image update, we apply ADMM [2], [4], [9]—simply put, it introduces an auxiliary variable to separate the effects of a certain variable or combinations of variables (called variable splitting in [4], [14]). For efficient sparse coding, we apply an analytical solution for $\mathbf{z}$. The next three subsections provide algorithmic details for solving (2), summarize them in Algorithm 1, and provide underlying intuitions.

1) Image Update - ADMM: Using the current sparse code estimates $\mathbf{z}$, we update the image $\mathbf{x}$ by augmenting (2) with auxiliary variables:

$$\min_{\substack{\mathbf{x},\mathbf{d}_a\in\mathbb{R}^m,\\ \mathbf{d}_\psi\in\mathbb{R}^{nC}}} \frac{1}{2}\|\mathbf{y}-\mathbf{d}_a\|_{\mathbf{W}}^2+\lambda\|\mathbf{d}_\psi\|_1$$

$$\text{subject to} \quad \begin{bmatrix}\mathbf{d}_a\\ \mathbf{d}_\psi\end{bmatrix}=\begin{bmatrix}\mathbf{A}\\ \widetilde{\mathbf{\Psi}}\end{bmatrix}\mathbf{x}-\begin{bmatrix}\mathbf{0}\\ \mathbf{z}\end{bmatrix}.$$

The corresponding augmented Lagrangian has the form

$$\frac{1}{2}\|\mathbf{y}-\mathbf{d}_a\|_{\mathbf{W}}^2+\lambda\|\mathbf{d}_\psi\|_1+\frac{\mu}{2}\|\mathbf{d}_a-\mathbf{A}\mathbf{x}-\mathbf{b}_a\|_2^2 + \frac{\mu\nu}{2}\left\|\mathbf{d}_\psi-\left(\widetilde{\mathbf{\Psi}}\mathbf{x}-\mathbf{z}\right)-\mathbf{b}_\psi\right\|^2.$$

We descend/ascend this augmented Lagrangian using the following iterative updates of the primal, auxiliary, dual variables—$\mathbf{x}$, $\{\mathbf{d}_a, \mathbf{d}_\psi\}$, and $\{\mathbf{b}_a, \mathbf{b}_\psi\}$, respectively:

$$\mathbf{x}^{(i+1)}=\left(\mathbf{A}^T\mathbf{A}+\nu\sum_{j=1}^J\mathbf{P}_j^T\mathbf{\Psi}^T\mathbf{\Psi}\mathbf{P}_j\right)^{-1}\cdot \left(\mathbf{A}^T\left(\mathbf{d}_a^{(i)}-\mathbf{b}_a^{(i)}\right)+\nu\widetilde{\mathbf{\Psi}}^T\left(\mathbf{d}_\psi^{(i)}-\mathbf{b}_\psi^{(i)}+\mathbf{z}\right)\right); \tag{3}$$

$$\mathbf{d}_a^{(i+1)}=(\mathbf{W}+\mu\mathbf{I}_m)^{-1}\left(\mathbf{W}\mathbf{y}+\mu\left(\mathbf{A}\mathbf{x}^{(i+1)}+\mathbf{b}_a^{(i)}\right)\right); \tag{4}$$

$$d_{\psi,u}^{(i+1)}=\mathrm{softshrink}\left(\left(\widetilde{\mathbf{\Psi}}\mathbf{x}^{(i+1)}-\mathbf{z}+\mathbf{b}_\psi^{(i)}\right)_u,\frac{\lambda}{\mu\nu}\right), \quad u=1,\ldots,nJ;$$

$$\mathbf{b}_a^{(i+1)}=\mathbf{b}_a^{(i)}-\left(\mathbf{d}_a^{(i+1)}-\mathbf{A}\mathbf{x}^{(i+1)}\right);$$

$$\mathbf{b}_\psi^{(i+1)}=\mathbf{b}_\psi^{(i)}-\left(\mathbf{d}_\psi^{(i+1)}-\left(\widetilde{\mathbf{\Psi}}\mathbf{x}^{(i+1)}-\mathbf{z}\right)\right),$$

where the soft-shrinkage operator is defined by $\mathrm{softshrink}(\alpha,\beta) := \mathrm{sign}(\alpha)\max(|\alpha|-\beta,0)$. To approximately solve (3), we use the preconditioned conjugate gradient (PCG) method with a circulant preconditioner $\mathbf{M}$ for $\mathbf{A}^T\mathbf{A}+\nu\sum_{j=1}^J\mathbf{P}_j^T\mathbf{\Psi}^T\mathbf{\Psi}\mathbf{P}_j$. For the two-dimensional (2D) CT problem, a circulant preconditioner is well suited because *1)* it is effective for the "nearly" shift-invariant matrix $\mathbf{A}^T\mathbf{A}$ [2], [4] and *2)* $\sum_{j=1}^J\mathbf{P}_j^T\mathbf{\Psi}^T\mathbf{\Psi}\mathbf{P}_j$ is a block circulant circulant block (BCCB) matrix when we use the overlapping "stride" 1 and the "wrap around" image patch assumption [15, Prop. 3.3].[b] PCG($\mathbf{M}$) denotes PCG method using a preconditioner $\mathbf{M}$; see Algorithm 1.

Algorithm 1 PWLS-ST-ℓ_1 CT Reconstruction

Require: $\mathbf{y}$, $\mathbf{x}^{(1)}$, $\mathbf{z}^{(1)}$, $\mathbf{\Psi}$ learned from (1), $\mathbf{M}$, $\mathbf{W}$, $\lambda,\gamma,\mu,\nu\geq 0$, $i=1$
 while a stopping criterion is not satisfied **do**
  for $i'=1,\ldots,\mathrm{Iter}_{\mathrm{ADMM}}$ **do**
   Obtain $\tilde{\mathbf{x}}^{(i'+1)}$ by solving (3) with PCG($\mathbf{M}$)
   $\mathbf{d}_a^{(i'+1)}=\left(\mathbf{W}+\mu\mathbf{I}_m\right)^{-1}\left(\mathbf{W}\mathbf{y}+\mu\left(\mathbf{A}\tilde{\mathbf{x}}^{(i'+1)}+\mathbf{b}_a^{(i')}\right)\right)$
   $d_{\psi,j}^{(i'+1)}=\mathrm{softshrink}\left(\left(\widetilde{\mathbf{\Psi}}\tilde{\mathbf{x}}^{(i'+1)}-\mathbf{z}^{(i)}+\mathbf{b}_\psi^{(i')}\right)_j,\frac{\lambda}{\mu\nu}\right)$, $j=1,\ldots,nJ$
   $\mathbf{b}_a^{(i'+1)}=\mathbf{b}_a^{(i')}-\left(\mathbf{d}_a^{(i'+1)}-\mathbf{A}\tilde{\mathbf{x}}^{(i'+1)}\right)$
   $\mathbf{b}_\psi^{(i'+1)}=\mathbf{b}_\psi^{(i')}-\left(\mathbf{d}_\psi^{(i'+1)}-\left(\widetilde{\mathbf{\Psi}}\tilde{\mathbf{x}}^{(i'+1)}-\mathbf{z}^{(i)}\right)\right)$,
  end for
  $\mathbf{x}^{(i+1)}=\tilde{\mathbf{x}}^{(\mathrm{Iter}_{\mathrm{ADMM}}+1)}$
  $z_j^{(i+1)}=\mathrm{hardshrink}\left(\left(\widetilde{\mathbf{\Psi}}\mathbf{x}^{(i+1)}\right)_j,\frac{\gamma}{\lambda}\right)$, $j=1,\ldots,nJ$
  $i=i+1$
 end while

2) Sparse Coding: Given the current estimates of the image $\mathbf{x}$, we update the sparse codes $\mathbf{z}$ by solving the following optimization problem:

$$\min_{\mathbf{z}\in\mathbb{R}^{nJ}} \lambda\left\|\widetilde{\mathbf{\Psi}}\mathbf{x}-\mathbf{z}\right\|_1+\gamma\|\mathbf{z}\|_0. \tag{5}$$

We efficiently solve (5) by an element-wise operator:

$$z_j^\star=\mathrm{hardshrink}\left(\left(\widetilde{\mathbf{\Psi}}\mathbf{x}\right)_j,\frac{\gamma}{\lambda}\right),\quad j=1,\ldots,nJ, \tag{6}$$

where the hard-shrinkage operator is given as follows: $\mathrm{hardshrink}(\alpha,\beta)$ is equal to α if $|\alpha|\geq\beta$, and is 0 otherwise. Note that γ should be properly determined based on the (estimated) intensity of $\widetilde{\mathbf{\Psi}}\mathbf{x}$. If γ is too small compared to the intensity, the operator in (6) may remove the sparse code coefficients corresponding to some edges in low-contrast regions (e.g., soft tissues); if γ is relatively too large, it does not properly remove noise (or unwanted artifacts).

3) Parameter selection based on condition numbers: In practice, the ADMM methods can require difficult parameter tuning processes for fast and stable convergence. We moderate this problem by selecting ADMM parameters (e.g., ν,μ) based on condition numbers [4]. Observe that, for two square Hermitian matrices $\mathbf{A}$ and $\mathbf{B}$,

$$\kappa(\mathbf{A}+\mathbf{B}):=\frac{\sigma_{\max}(\mathbf{A}+\mathbf{B})}{\sigma_{\min}(\mathbf{A}+\mathbf{B})}\leq\frac{\sigma_{\max}(\mathbf{A})+\sigma_{\max}(\mathbf{B})}{\sigma_{\min}(\mathbf{A})+\sigma_{\min}(\mathbf{B})}, \tag{7}$$

by Weyl's inequality, where the notations $\kappa(\cdot)$, $\sigma_{\max}(\cdot)$, and $\sigma_{\min}(\cdot)$ denote the condition number, the largest eigenvalue,

[b]For an orthogonal transform $\mathbf{\Psi}$, $\sum_{j=1}^J\mathbf{P}_j^T\mathbf{\Psi}^T\mathbf{\Psi}\mathbf{P}_j$ can be approximated by $(n/\iota)\mathbf{I}_N$ (ι denotes the stride parameter), i.e., a circulant preconditioner is still a reasonable choice for the 2D projection matrix $\mathbf{A}$.

and the smallest eigenvalue of a matrix, respectively. Applying the bound (7) to (3), we select ν by

$$\nu = \frac{\sigma_{\max}(\mathbf{\Lambda_A}) - \kappa_{\mathrm{des},\nu} \cdot \sigma_{\min}(\mathbf{\Lambda_A})}{\kappa_{\mathrm{des},\nu} \cdot \sigma_{\min}(\mathbf{\Lambda}_{\widetilde{\mathbf{\Psi}}}) - \sigma_{\max}(\mathbf{\Lambda}_{\widetilde{\mathbf{\Psi}}})} \tag{8}$$

where $\kappa_{\mathrm{des},\nu}$ denotes the desired "upper bounded" condition number of $\mathbf{A}^T\mathbf{A} + \nu \sum_{j=1}^J \mathbf{P}_j^T \mathbf{\Psi}^T \mathbf{\Psi} \mathbf{P}_j$, and $\mathbf{\Lambda_A}$ and $\mathbf{\Lambda}_{\widetilde{\mathbf{\Psi}}}$ are approximated diagonal eigenvalue matrices of $\mathbf{A}^T\mathbf{A}$ and $\widetilde{\mathbf{\Psi}}^T\widetilde{\mathbf{\Psi}}$ by using their preconditioners in Section II-C1, respectively. Note that equality holds in (7) when either $\mathbf{A}$ or $\mathbf{B}$ is a scaled identity matrix. In other words, $\kappa_{\mathrm{des},\nu}$ becomes close to the condition number of $\mathbf{A}^T\mathbf{A} + \nu \sum_{j=1}^J \mathbf{P}_j^T \mathbf{\Psi}^T \mathbf{\Psi} \mathbf{P}_j$, when the learned ST $\mathbf{\Psi}$ is close to orthogonal. We select μ for (4) by

$$\mu = \frac{\sigma_{\max}(\mathbf{W}) - \kappa_{\mathrm{des},\mu} \cdot \sigma_{\min}(\mathbf{W})}{\kappa_{\mathrm{des},\mu} - 1}, \tag{9}$$

where $\kappa_{\mathrm{des},\mu}$ denotes the desired condition number of $\mathbf{W} + \mu \mathbf{I}_m$ in (4). We empirically found that $\kappa_{\mathrm{des},\nu}, \kappa_{\mathrm{des},\mu} \in [10, 40]$ are reasonable values for fast and stable convergence.

4) Intuitions behind Algorithm 1: The underlying idea of the image reconstruction model (2) is that the signal is very sparse in the learned transform ($\mathbf{\Psi}$)-domain, i.e., $\widetilde{\mathbf{\Psi}}\mathbf{x}$ has a few large coefficients, usually corresponding to local high-frequency features (e.g., edges). Thresholding in the sparse coding step, i.e., (6), removes the noise in the other components while preserving the large signal coefficients. Substituting the denoised sparse codes $\mathbf{z}$ to the image updating optimization, we estimate an image $\mathbf{x}$ close to the denoised sparse codes in $\widetilde{\mathbf{\Psi}}$-domain, while being robust to the model mismatch in $\widetilde{\mathbf{\Psi}}\mathbf{x}$ and $\mathbf{z}$. Repeating these processes, we expect to obtain reconstructed images with higher accuracy.

III. Experimental Results and Discussion

A. Experiment Setup

We pre-learned square STs from 8×8 image patches extracted from five different slices of an XCAT phantom (with 1×1 overlapping stride) [16]. We chose a large enough τ, e.g., $\tau = 5.85 \times 10^{15}$, to learn well-conditioned transforms. We chose $\gamma' = 110$ and $\xi = 1$. We ran 1000 iterations of the alternating minimization algorithm proposed in [6] to ensure learned transforms are completely converged.

Our experiments used a simulated (2D) fan-beam CT scan of a 1024×1024 slice of the XCAT phantom, which is different from the learning slices, and $\Delta_x = \Delta_y = 0.4883$ mm. We simulated sinograms of size 888 (detectors or rays) $\times$ 246, 123 (regularly spaced projection views or angles; 984 is the number of full views) with GE LightSpeed fan-beam geometry corresponding to a monoenergetic source with $\rho_0 = 10^5$ incident photons per ray and no background events, the conversion gain $\vartheta = 1000$ (electrons per incident X-ray photon) [17], and electronic noise variance $\sigma^2 = 330^2$. We reconstructed a 512×512 image with a coarser grid, where $\Delta_x = \Delta_y = 0.9766$ mm.

For PWLS-EP with hyperbola regularizer $\varphi(t) := \delta^2(\sqrt{1 + |t/\delta|^2} - 1)$ ($\delta = 10$ in Hounsfield units, HU), we used relaxed linearized augmented Lagrangian method with

TABLE I

RMSE (HU) of different X-ray CT reconstructions with different number of projection views ($\rho_0 = 10^5$)

Views	Measure	FBP	PWLS-EP	PWLS-ST-ℓ_2	PWLS-ST-ℓ_1
246	RMSE	60.3	32.6	30.3	**25.2**
123	RMSE	82.0	37.0	33.2	**28.7**

ordered-subsets proposed in [18] to accelerate the reconstruction. Initialized with filtered back projection (FBP) reconstructions (Hanning window), we chose the regularization parameter as 2^{21} and $2^{20.5}$ for 246 and 123 views, respectively. We ran the algorithm with 100 iterations and 10 subsets.

We evaluated PWLS-ST-ℓ_2 with the algorithm in [8] but without the non-nonnegativity constraint for image update. For both PWLS-ST-ℓ_1 and PWLS-ST-ℓ_2, we used converged PWLS-EP reconstructions for initialization and set a stopping criterion by meeting the maximum number of iterations, e.g., $\mathrm{Iter} = 200$. For the image update, we set $\mathrm{Iter}_{\mathrm{ADMM}}$ as 2 (2 PCG iterations) for PWLS-ST-ℓ_1; and set 12 rLALM iterations (rLALM stands for relaxed linearized augmented Lagrangian method [8]) without ordered subsets for PWLS-ST-ℓ_2.

We finely tuned the parameters λ, γ to achieve the lowest root mean squared error (RMSE) in image reconstruction. We tuned ν, μ through the condition number based selection schemes, i.e., $\kappa_{\mathrm{des},\nu}$ in (8) and $\kappa_{\mathrm{des},\mu}$ in (9). For PWLS-ST-ℓ_1, we chose $\{\lambda, \gamma/\lambda, \kappa_{\mathrm{des},\nu}, \kappa_{\mathrm{des},\mu}\}$ as follows: for 246 views, $\{1.1 \times 10^9, 80, 30, 30\}$; for 123 views, $\{9 \times 10^8, 80, 30, 30\}$. For PWLS-ST-$\ell_2$ [8], we chose $\{\lambda, \gamma\}$ as follows: for 246 views, $\{1.5 \times 10^{13}, 20\}$; for 123 views, $\{8 \times 10^{12}, 20\}$. Note that λ and γ are in HU.

We evaluated the reconstruction quality by the RMSE (in HU) in a region of interest (ROI).[c] The RMSE is defined by $\mathrm{RMSE} := (\sum_{j=1}^{N_{\mathrm{ROI}}} (\hat{x}_j - x_j^*)^2 / N_{\mathrm{ROI}})^{1/2}$, where $\hat{\mathbf{x}}$ is the reconstructed image (after thresholding non-negative values), $\mathbf{x}^*$ is the ground truth image, and N_{ROI} is the number of pixels in a ROI.

B. Results and Discussion

Table I shows that the proposed PWLS-ST-ℓ_1 model outperforms PWLS-EP and PWLS-ST-ℓ_2 in terms of RMSE. In particular, PWLS-ST-ℓ_1 resolves the blurry edge problem in PWLS-ST-ℓ_2; see Fig. 1. The edge-preserving benefit of PWLS-ST-ℓ_1 over PWLS-ST-ℓ_2 can be explained when there exist some outliers for some $\mathbf{z}^{(i+1)}$: $\|\widetilde{\mathbf{\Psi}}\mathbf{x} - \mathbf{z}^{(i+1)}\|_1$ in (2) gives equal emphasis to all sparse code coefficients—e.g., the components corresponding to edges from low-contrast (e.g., soft tissue) to high-contrast (e.g., bone) regions—in estimating $\mathbf{x}$; however, PWLS-ST-ℓ_2 adjusts $\mathbf{x}$ to mainly minimize the outliers, i.e., it may not pay enough attention to reconstruct edges on soft tissues. The proposed PWLS-ST-ℓ_1 model can accomplish the both benefits of edge-preserving (achieved by PWLS-EP) and image denoising (achieved by PWLS-ST-ℓ_2).

[c]The ROI in our experiment was a circular (around center) region containing all the phantom tissues.

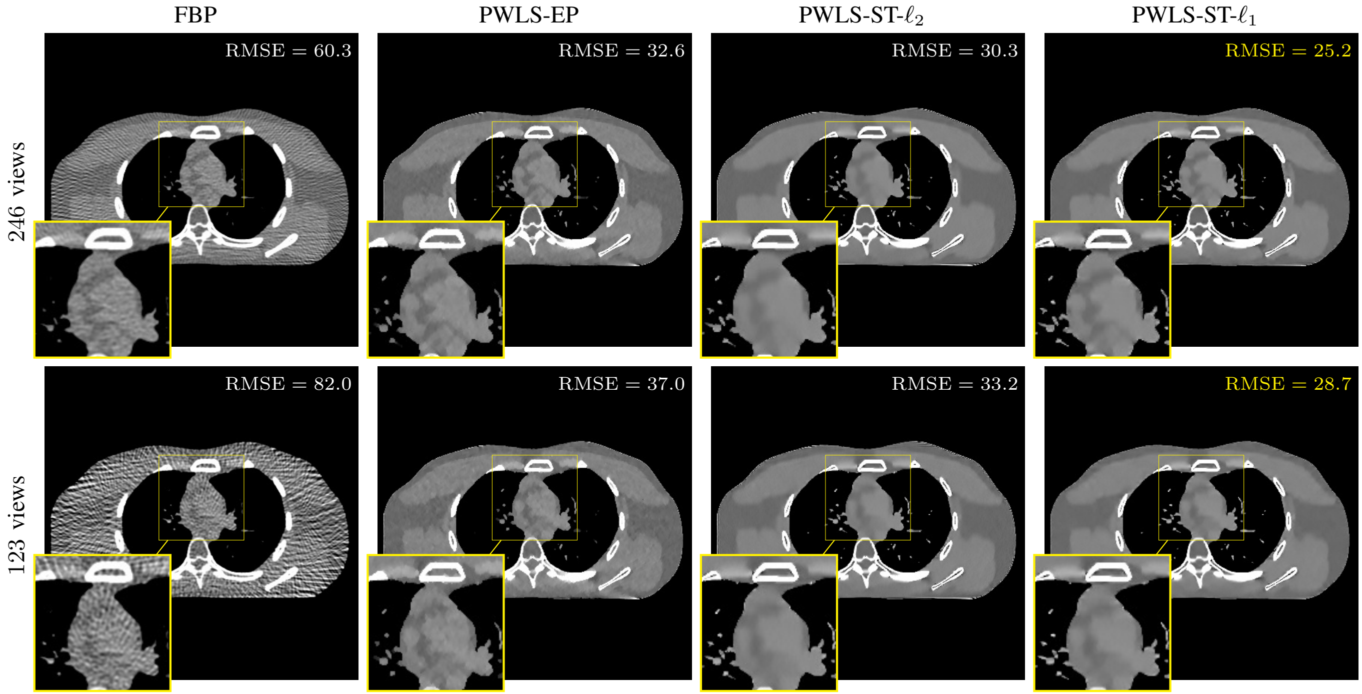

Fig. 1. Comparison of reconstructed images from different X-ray CT reconstruction models with different number of views ($\rho_0 = 10^5$ and display window is within $[800, 1200]$ HU). The proposed PWLS-ST-ℓ_1 consistently improves the sharpness of the reconstructed images over PWLS-ST-ℓ_2.

IV. Conclusion

The proposed PWLS-ST-ℓ_1 model achieves more accurate sparse-view (2D) CT reconstruction compared to PWLS-EP and PWLS-ST-ℓ_2; in particular, it leads to sharper edge reconstruction compared to PWLS-ST-ℓ_2. Future work will explore PWLS-ST-ℓ_1 with the technique controlling local spatial resolution or noise in the reconstructed images [19], [20] in 3D CT to reduce blur, particularly around the center of reconstructed image. The Appendix introduces the PWLS-ST-ℓ_1 model encouraging uniform spatial resolution or noise; see our preliminary results showing its effectiveness for 3D CT in Fig. 2. On the algorithmic side, we plan to apply *block proximal gradient method using majorizer* [21] to solve nonconvex problem (2) faster.

Appendix: PWLS-ST-ℓ_1 Encouraging Uniform Spatial Resolution or Noise

We first obtain parameter $\boldsymbol{\omega} \in \mathbb{R}^N_{>0}$ that controls local spatial resolution or noise in reconstructed image [19], [20]:

$$\omega_u = \sqrt{\frac{\sum_{l=1}^m A_{l,u} W_{l,l}}{\sum_{l=1}^m A_{l,u}}}, \quad u = 1, \ldots, N.$$

Using $\boldsymbol{\omega}$, we compute the weighting parameters $\boldsymbol{\lambda}' \in \mathbb{R}^J_{>0}$ by $\{\lambda'_j = \|\mathbf{P}_j \boldsymbol{\omega}\|_1 / n : j = 1, \ldots, J\}$. Similar to [20, (5)], we apply $\{\lambda'_j : j = 1, \ldots, J\}$ to a regularizer in image update step (problem related to $\mathbf{x}$) of (2). We propose PWLS-ST-ℓ_1 model promoting uniform spatial resolution or noise as follows:

$$\min_{\mathbf{x}, \mathbf{z} \in \mathbb{R}^{nJ}} \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_{\mathbf{W}}^2 + \sum_{j=1}^{J} \lambda'_0 \lambda'_j \|\boldsymbol{\Psi} \mathbf{P}_j \mathbf{x} - \mathbf{z}_j\|_1 + \gamma \|\mathbf{z}_j\|_0 . \tag{10}$$

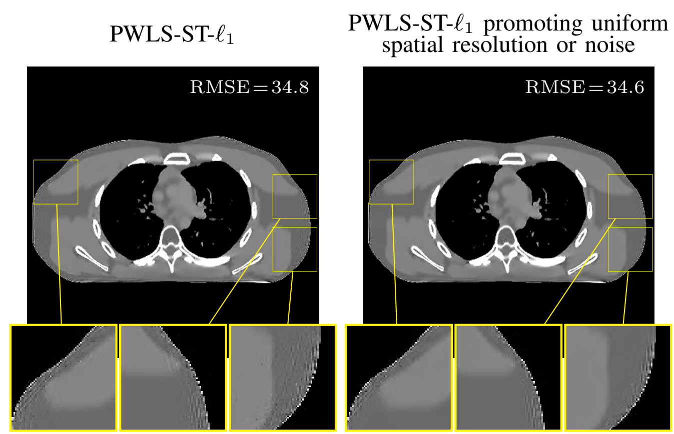

Fig. 2. Comparison of 3D reconstructed images from PWLS-ST-ℓ_1 and PWLS-ST-ℓ_1 integrating uniform spatial resolution approach [19], [20] (420×420-sized images shown for central axial plane with $[800, 1200]$ HU display window; 3D cone-beam CT with $\rho_0 = 10^4$ and $888 \times 64 \times 123$ sinograms; PWLS-ST-ℓ_1 using $8 \times 8 \times 8$ overlapping image patches with $3 \times 3 \times 3$ patch stride, PCG with diagonal majorizer $\mathbf{M}$ and 2 iterations for solving (3), $\text{Iter}_{\text{ADMM}} = 2$, and $\text{Iter} = 200$; the error measurements are calculated from $420 \times 420 \times 64$ reconstructed images). For 3D sparse-view CT, integrating the approach controlling local spatial resolution or noise [19], [20] into PWLS-ST-ℓ_1 further improves the quality of reconstructed images.

Fig. 2 shows that, for 3D CT reconstruction, PWLS-ST-ℓ_1 encouraging uniform spatial resolution or noise (10) improves the accuracy of reconstructed images compared to the PWLS-ST-ℓ_1 (2). Alternatively, one can promote uniform spatial resolution or noise as follows:

$$\min_{\mathbf{x}, \mathbf{z} \in \mathbb{R}^{nJ}} \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_{\mathbf{W}}^2 + \sum_{j=1}^{J} \lambda'_0 \|\boldsymbol{\Psi} \mathbf{P}_j \mathbf{x} - \mathbf{z}_j\|_1 + \lambda'_j \gamma \|\mathbf{z}_j\|_0 .$$

In other words, the sparse codes at the jth patch are thresholded less when the corresponding λ_j has low values. This is expected to be useful to preserve edges around the center (which has low λ_j values).

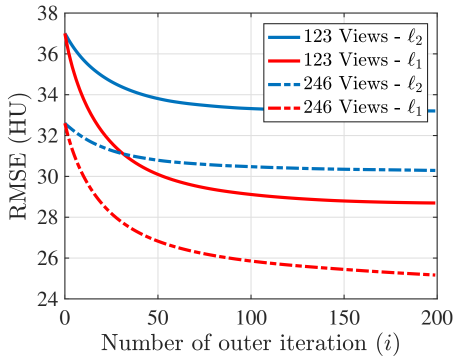

Fig. 3. RMSE convergence behavior for PWLS-ST-ℓ_1 and PWLS-ST-ℓ_2 (123 and 246 projection views, and $\rho_0 = 10^5$).

Acknowledgment

We thank Dr. Jang Hwan Cho and Dr. Saiprasad Ravishankar for a constructive discussion.

References

[1] G. H. Chen *et al.*, "Prior image constrained compressed sensing (PICCS): a method to accurately reconstruct dynamic ct images from highly undersampled projection data sets," *Med. Phys.*, vol. 35, no. 2, pp. 660–663, Feb. 2008.

[2] I. Y. Chun and T. Talavage, "Efficient compressed sensing statistical X-ray/CT reconstruction from fewer measurements," in *Proc.* 12^{th} *Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, Lake Tahoe, CA, Jun. 2013, pp. 30–33.

[3] E. Y. Sidky *et al.*, "Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT," *J. X-ray Sci. Technol.*, vol. 14, no. 2, pp. 119–139, 2006.

[4] S. Ramani and J. A. Fessler, "A splitting-based iterative algorithm for accelerated statistical X-ray CT reconstruction," *IEEE Trans. Med. Imag.*, vol. 31, no. 3, pp. 677–688, Mar. 2012.

[5] M. Aharon *et al.*, "*K*-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[6] S. Ravishankar and Y. Bresler, "ℓ_0 sparsifying transform learning with efficient optimal updates and convergence guarantees," *IEEE Trans. Signal Process.*, vol. 63, no. 9, pp. 2389–2404, May 2015.

[7] Q. Xu *et al.*, "Low-dose X-ray CT reconstruction via dictionary learning," *IEEE Trans. Med. Imag.*, vol. 31, no. 9, pp. 1682–1697, Sep. 2012.

[8] X. Zheng *et al.*, "Low dose CT image reconstruction with learned sparsifying transform," in *Proc.* 12^{th} *IEE IVMSP Workshop*, Bordeaux, France, Jul. 2016, pp. 1–5.

[9] S. Boyd *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. & Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

[10] Q. Ding *et al.*, "Modeling mixed Poisson-Gaussian noise in statistical image reconstruction for X-ray CT," in *Proc.* 4^{th} *Intl. Mtg. on Image Formation in X-ray CT*, Bamberg, Germany, pp. 399–402.

[11] C. Zhang *et al.*, "Low-dose ct reconstruction via L1 dictionary learning regularization using iteratively reweighted least-squares," *Biomed. Eng. OnLine*, vol. 15, no. 1, p. 66, Jun. 2016.

[12] J. B. Thibault *et al.*, "A recursive filter for noise reduction in statistical iterative tomographic imaging," in *Proc. SPIE 6065, Computational Imaging IV*, vol. 6065, Feb. 2006, p. 60650X.

[13] I. Y. Chun and B. Adcock, "Compressed sensing and parallel acquisition," to appear in *IEEE Trans. Inf. Theory*, Mar. 2017. [Online]. Available: http://arxiv.org/abs/1601.06214

[14] I. Y. Chun *et al.*, "Efficient compressed sensing SENSE pMRI reconstruction with joint sparsity promotion," *IEEE Trans. Med. Imag.*, vol. 35, no. 1, pp. 354–368, Jan. 2016.

[15] S. Ravishankar and Y. Bresler, "Efficient blind compressed sensing using sparsifying transforms with convergence guarantees and application to magnetic resonance imaging," *SIAM J. Imaging Sci.*, vol. 8, no. 4, pp. 2519–2557, Nov. 2015.

[16] W. P. Segars *et al.*, "Realistic CT simulation using the 4D XCAT phantom," *Med. Phys.*, vol. 35, no. 8, pp. 3800–3808, Jul. 2008.

[17] C. C. Shaw, *Cone beam computed tomography*. Taylor & Francis, 2014.

[18] H. Nien and J. A. Fessler, "Relaxed linearized algorithms for faster X-ray CT image reconstruction," pp. 1090–1098, Apr. 2016.

[19] J. A. Fessler and W. L. Rogers, "Spatial resolution properties of penalized-likelihood image reconstruction methods: Space-invariant tomographs," *IEEE Trans. Image Process.*, vol. 5, no. 9, pp. 1346–58, Sep. 1996.

[20] J. H. Cho and J. A. Fessler, "Regularization designs for uniform spatial resolution and noise properties in statistical image reconstruction for 3-D X-ray CT," *IEEE Trans. Med. Imag.*, vol. 34, no. 2, pp. 678–689, Feb. 2015.

[21] I. Y. Chun and J. A. Fessler, "Convolutional dictionary learning: Acceleration and convergence," under review for *IEEE Trans. Image Process.*, Apr. 2017.