LOW DOSE CT IMAGE RECONSTRUCTION WITH LEARNED SPARSIFYING TRANSFORM

Xuehang Zheng[1], Zening Lu[1], Saiprasad Ravishankar[2], Yong Long*[1], Jeffrey A. Fessler[2]

[1]University of Michigan - Shanghai Jiao Tong University Joint Institute,
Shanghai Jiao Tong University, Shanghai, China
[2]Department of Electrical Engineering and Computer Science, University of Michigan, MI, USA

ABSTRACT

A major challenge in computed tomography (CT) is to reduce X-ray dose to a low or even ultra-low level while maintaining the high quality of reconstructed images. We propose a new method for CT reconstruction that combines penalized weighted-least squares reconstruction (PWLS) with regularization based on a sparsifying transform (PWLS-ST) learned from a dataset of numerous CT images. We adopt an alternating algorithm to optimize the PWLS-ST cost function that alternates between a CT image update step and a sparse coding step. We adopt a relaxed linearized augmented Lagrangian method with ordered-subsets (relaxed OS-LALM) to accelerate the CT image update step by reducing the number of forward and backward projections. Numerical experiments on the XCAT phantom show that for low dose levels, the proposed PWLS-ST method dramatically improves the quality of reconstructed images compared to PWLS reconstruction with a nonadaptive edge-preserving regularizer (PWLS-EP).

Index Terms— Low dose CT, Sparsifying transform learning, Statistical image reconstruction, Sparse representation, Dictionary learning

1. INTRODUCTION

A major challenge in computed tomography (CT) is to reduce X-ray dose to a low or even ultra-low level while maintaining the high quality of reconstructed images. Low dose CT (LDCT) scans that still provide good image quality could significantly improve the benefits of CT scans and open up numerous entirely new clinical applications.

Currently, most commercial CT scanners use a technique called filtered back-projection (FBP) for image reconstruction. FBP requires undesirably high doses of radiation to produce high-quality diagnostic images. Model-based image reconstruction (MBIR) methods, also known as statistical image reconstruction methods, produce high-quality and accurate images, while reducing patient radiation exposure. Weighted-least squares (WLS) estimation is commonly used for CT [1]. WLS estimation with proper weighting that gives less weight to measurements that are noisier and more weight to the more reliable data reduces noise in the reconstructed image. Penalized weighted-least squares (PWLS) reconstruction with added regularization based on prior knowledge of the underlying unknown object improves image quality in LDCT reconstruction [2]. Thus, MBIR with better image priors is a promising way to develop improved reconstruction methods for achieving high quality LDCT imaging.

Prior information extracted from big datasets of CT images could potentially enable dramatic improvements in image reconstruction from LDCT measurements. It is well known that natural signals are sparse in certain transform domains, such as wavelets and discrete gradient domain. A sparsifying transform (ST) converts signals into these domains where they can be represented using a few non-zero coefficients. Ravishankar and Bresler [3, 4] proposed a generalized analysis dictionary learning method, called transform learning, to efficiently find sparse representations of data. The transform learning method avoids optimization of highly non-convex or NP-hard cost functions involved in both synthesis [5, 6] and previous analysis [7] dictionary learning methods, and shows promising performance and speed-ups over the popular synthesis K-SVD [6] algorithm in applications such as image denoising.

Xu et al. [8] first applied dictionary learning to CT image reconstruction by proposing a PWLS approach with regularization based on a redundant synthesis dictionary. Their method uses a global dictionary trained from image patches extracted from one normal-dose FBP image, or an adaptive dictionary jointly estimated with the low-dose image. Pfister and Bresler [2, 9] proposed a model-based iterative reconstruction method with adaptive sparsifying transforms to jointly estimate the ST and the image, showing the promise of PWLS reconstruction with ST regularization. Existing CT image reconstruction methods based on dictionary/transform learning have two common downsides. Firstly, the dictionary is often learned from a very small number of prior images or from the current measurements themselves, which does not

take advantage of the existing big databases of CT images acquired from thousands of patients. Secondly, the prior images are often from the same patient at regular dose. When such a prior image is not available, these methods could not be used.

We propose a new method for low dose CT reconstructions that combines conventional PWLS reconstruction with regularization based on a sparsifying transform (PWLS-ST) learned from a dataset of numerous CT images. Numerical experiments on the XCAT phantom show that for low dose levels, our method dramatically improves the quality of reconstructed images compared to PWLS reconstruction with an edge-preserving hyperbola regularizer (PWLS-EP). We adopt a relaxed linearized augmented Lagrangian method with ordered-subsets (relaxed OS-LALM) [10] to accelerate the image reconstruction process.

2. PROBLEM FORMULATION

We solve the following optimization problem to reconstruct an image $\mathbf{x} \in \mathbb{R}^{N_p}$ from noisy sinogram data $\mathbf{y} \in \mathbb{R}^{N_d}$ using a pre-learned [4] ST matrix $\mathbf{\Omega} \in \mathbb{R}^{l \times l}$:

$$\min_{\mathbf{x} \in \mathcal{C}} \frac{1}{2} \|\mathbf{y} - \mathbf{Ax}\|_{\mathbf{W}}^2 + \mathsf{R}(\mathbf{x}) \qquad \text{(P0)}$$

where the regularizer R based on the sparsifying transform $\mathbf{\Omega}$ is defined as:

$$\mathsf{R}(\mathbf{x}) \triangleq \min_{\{\mathbf{z}_j\}} \beta \sum_{j=1}^{N} \left\{ \|\mathbf{\Omega} \mathbf{P}_j \mathbf{x} - \mathbf{z}_j\|_2^2 + \gamma^2 \|\mathbf{z}_j\|_0 \right\} \qquad (1)$$

$\mathbf{W} = \mathsf{diag}\{w_i\}$ is the diagonal weighting matrix with elements being the reciprocal of the variance, i.e., $w_i \approx 1/\sigma^2(y_i)$, $\mathbf{A} \in \mathbb{R}^{N_d \times N_p}$ is the system matrix of a CT scan, and $\mathcal{C} \triangleq \{\mathbf{x} | x_j \geq 0, 1 \leq j \leq N_p\}$. N is the number of image patches, the operator $\mathbf{P}_j \in \mathbb{R}^{l \times N_p}$ extracts the jth patch of l voxels of $\mathbf{x}$ as a vector $\mathbf{P}_j \mathbf{x}$, $\mathbf{z}_j \in \mathbb{R}^l$ denotes the sparse representation of $\mathbf{P}_j \mathbf{x}$, β is a positive parameter to control the noise and resolution trade-off, γ is a weight to control sparsity in the model, and $\| \cdot \|_0$ is the ℓ_0 "norm" that counts the number of nonzero elements in a vector.

In (P0), the patches of the underlying image are assumed to be approximately sparse in the learned transform $\mathbf{\Omega}$ domain. We estimate both the image $\mathbf{x}$ and the sparse coefficients $\{\mathbf{z}_j\}$ from LDCT data.

3. ALGORITHM

3.1. Sparsifying Transform (ST) Learning

We learn a sparsifying transform (ST) matrix $\mathbf{\Omega}$ from the patches extracted from a dataset of regular dose CT images,

i.e., we solve the following problem:

$$\min_{\mathbf{\Omega}, \mathbf{Z}} \|\mathbf{\Omega Y} - \mathbf{Z}\|_F^2 + \lambda \left(\|\mathbf{\Omega}\|_F^2 - \log |\det \mathbf{\Omega}| \right) + \sum_{i=1}^{N'} \gamma^2 \|\mathbf{Z}_i\|_0$$
$$\text{(P1)}$$

where N' is the number of training patches, $\mathbf{Z} \in \mathbb{R}^{l \times N'}$ is a matrix whose columns $\{\mathbf{Z}_i\}$ are the sparse codes of the corresponding training signals (vectorized patches) in $\mathbf{Y} \in \mathbb{R}^{l \times N'}$, λ and γ are positive scalar parameters, and regularizer $\|\mathbf{\Omega}\|_F^2 - \log |\det \mathbf{\Omega}|$ prevents trivial solutions and enables control over the condition number of $\mathbf{\Omega}$.

3.2. Optimization algorithm

We propose an alternating algorithm to solve (P0) that alternates between updating $\mathbf{x}$ (*image update step*) and $\{\mathbf{z}_j\}$ (*sparse coding step*) with other variables kept fixed.

3.2.1. Image Update Step

With $\{\mathbf{z}_j\}$ fixed, (P0) reduces to the following weighted least squares problem:

$$\min_{\mathbf{x} \in \mathcal{C}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{Ax}\|_{\mathbf{W}}^2 + \beta \sum_{j=1}^{N} \|\mathbf{\Omega} \mathbf{P}_j \mathbf{x} - \mathbf{z}_j\|_2^2 \right\} \qquad (2)$$

We solve this problem using the relaxed OS-LALM [10] by iterating over the following update steps:

$$\begin{cases} \mathbf{s}^{(k+1)} = \rho(\mathbf{D_A} \mathbf{x}^{(k)} - \mathbf{h}^{(k)}) + (1-\rho)\mathbf{g}^{(k)} \\ \mathbf{x}^{(k+1)} = [\mathbf{x}^{(k)} - (\rho \mathbf{D_A} + \mathbf{D_R})^{-1}(\mathbf{s}^{(k+1)} + \nabla\mathsf{R}(\mathbf{x}^{(k)}))]_{\mathcal{C}} \\ \boldsymbol{\zeta}^{(k+1)} \triangleq M \mathbf{A}_m' \mathbf{W}(\mathbf{A}_m \mathbf{x}^{(k+1)} - \mathbf{y}) \\ \mathbf{g}^{(k+1)} = \frac{\rho}{\rho+1}(\alpha \boldsymbol{\zeta}^{(k+1)} + (1-\alpha)\mathbf{g}^{(k)}) + \frac{1}{\rho+1}\mathbf{g}^{(k)} \\ \mathbf{h}^{(k+1)} = \alpha(\mathbf{D_A}\mathbf{x}^{(k+1)} - \boldsymbol{\zeta}^{(k+1)}) + (1-\alpha)\mathbf{h}^{(k)} \end{cases}$$
$$(3)$$

where $\mathbf{D_A} \succeq \mathbf{A}'\mathbf{WA}$ is a diagonal majorizing matrix of $\mathbf{A}'\mathbf{WA}$, e.g., $\mathbf{D_A} \triangleq \mathsf{diag}\{|\mathbf{A}|'|\mathbf{W}||\mathbf{A}|\mathbf{1}\} \succeq \mathbf{A}'\mathbf{WA}$ [11], $[\cdot]_{\mathcal{C}}$ is an operator that projects the input vector onto the convex set $\mathcal{C}$, M is the number of ordered subsets, $\mathbf{A}_m$ is the subset forward projection matrix, $1 \leq \alpha < 2$ is the (over-)relaxation parameter, and $\rho > 0$ is the AL penalty parameter decreasing gradually as iterations progress [10], i.e.,

$$\rho_n(\alpha) = \begin{cases} 1 & , n = 0 \\ \frac{\pi}{\alpha(n+1)} \sqrt{1 - \left(\frac{\pi}{2\alpha(n+1)}\right)^2} & , \text{otherwise.} \end{cases} \qquad (4)$$

The learned sparsifying transform $\mathbf{\Omega}$ is a well-conditioned square matrix [4]. $\mathbf{D_R} \succeq \nabla^2\mathsf{R}(\mathbf{x}) = 2\beta \sum_{j=1}^{N} \mathbf{P}_j'\mathbf{\Omega}'\mathbf{\Omega}\mathbf{P}_j$ is a diagonal majorizing matrix of the Hessian of the regularizer $\mathsf{R}(\mathbf{x})$, e.g.,

$$\mathbf{D_R} \triangleq 2\beta \sum_{j=1}^{N} \mathbf{P}_j'\mathbf{P}_j \lambda_{\max}(\mathbf{\Omega}'\mathbf{\Omega}) \succeq \nabla^2\mathsf{R}(\mathbf{x}) \qquad (5)$$

The term $\sum_{j=1}^{N} \mathbf{P}_j' \mathbf{P}_j \in \mathbb{C}^{N_p \times N_p}$ is a diagonal matrix with the diagonal entries corresponding to image pixel locations and their values being the number of patches overlapping each pixel [12]. If we assume periodically positioned overlapping image patches that wrap around at image boundaries, then the diagonal entries are equal, i.e., $\sum_{j=1}^{N} \mathbf{P}_j' \mathbf{P}_j = \kappa \mathbf{I}$ ($\mathbf{I} \in \mathbb{C}^{N_p \times N_p}$), where κ is a scalar. In particular, when the overlap stride is 1, κ is equal to the image patch size l. Therefore, $\mathbf{D_R}$ simplifies to:

$$\mathbf{D_R} = 2\beta l \lambda_{\max}(\mathbf{\Omega'\Omega})\mathbf{I} \tag{6}$$

Since $\mathbf{D_R}$ is independent of $\mathbf{x}$ and $\mathbf{z}_j$, we precompute it prior to iterating.

3.2.2. Sparse Coding Step

With $\mathbf{x}$ fixed, we update $\{\mathbf{z}_j\}$ by solving

$$\min_{\{\mathbf{z}_j\}} \sum_{j=1}^{N} \left\{ \|\mathbf{\Omega P}_j\mathbf{x} - \mathbf{z}_j\|_2^2 + \gamma^2 \|\mathbf{z}_j\|_0 \right\} \tag{7}$$

The optimal sparse codes are given in closed form as $\hat{\mathbf{z}}_j = H_\gamma(\mathbf{\Omega P}_j\mathbf{x}) \, \forall j$, i.e., setting the entries with magnitude less than γ to zero. The *hard-thresholding* operator $H_\gamma(\cdot)$ is applied to each entry b in a vector as

$$H_\gamma(b) \triangleq \begin{cases} 0, & |b| < \gamma \\ b, & |b| \geq \gamma. \end{cases} \tag{8}$$

Algorithm 1 describes the proposed optimization algorithm for solving the PWLS-ST reconstruction problem (P0).

4. EXPERIMENTAL RESULTS

We evaluate the proposed PWLS-ST method and compare its image reconstruction quality with those of conventional FBP with a Hanning window, PWLS reconstruction with regularization based on DCT in (1) (PWLS-DCT), and PWLS reconstruction with edge-preserving hyperbola regularization (PWLS-EP). The PWLS-EP reconstruction is optimized using relaxed OS-LALM algorithm [10].

(a) DCT   (b) $\gamma = 125$
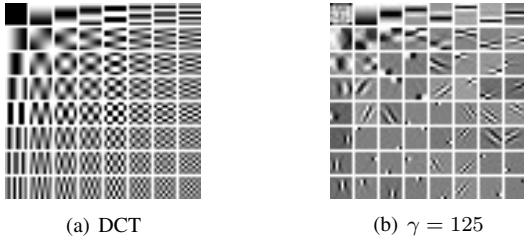
Fig. 1: Rows of the sparsifying transform shown as 8×8 patches for (a) 2D DCT, and (b) learned $\mathbf{\Omega}$ with $\gamma = 125$.

We pre-learned a ST matrix from 5 different slices of an XCAT phantom [13] using (P1). We extracted 8×8 image

---

Algorithm 1 PWLS-ST Algorithm

Input: initial image $\tilde{\mathbf{x}}^{(0)}$, $\alpha = 1.999$, $\rho = 1$,
$\mathbf{D_A} = \text{diag}\{|\mathbf{A}|'|\mathbf{W}||\mathbf{A}|\mathbf{1}\}$, $\mathbf{D_R} = 2\beta l \lambda_{\max}(\mathbf{\Omega'\Omega})\mathbf{I}$.
Output: $\tilde{\mathbf{x}}^{(I)}$ - reconstructed image.
for $i = 0, 1, 2, \cdots, I - 1$ **do**
    (1) Image Reconstruction: with $\{\tilde{\mathbf{z}}_j^{(i)}\}$ fixed,
      Initialization: $\mathbf{g}^{(0)} = M\mathbf{A}_m'\mathbf{W}(\mathbf{A}_m\tilde{\mathbf{x}}^{(i)} - \mathbf{y})$,
$\boldsymbol{\zeta}^{(0)} = \mathbf{g}^{(0)}$, $\mathbf{h}^{(0)} = \mathbf{D_A}\tilde{\mathbf{x}}^{(i)} - \boldsymbol{\zeta}^{(0)}$, $\mathbf{x}^{(0)} = \tilde{\mathbf{x}}^{(i)}$,
$\nabla\mathsf{R}(\mathbf{x}^{(0)}) = 2\beta \sum_{j=1}^{N} \{\mathbf{P}_j'\mathbf{\Omega'\Omega P}_j\tilde{\mathbf{x}}^{(i)} - \mathbf{P}_j'\mathbf{\Omega}'\tilde{\mathbf{z}}_j^{(i)}\}$.
    for $k = 0, 1, 2, \cdots, K - 1$ **do**
      for $m = 0, 1, 2, \cdots, M - 1$ **do**

$$\begin{cases} \mathbf{s}^{(k+1)} = \rho(\mathbf{D_A}\mathbf{x}^{(k)} - \mathbf{h}^{(k)}) + (1-\rho)\mathbf{g}^{(k)} \\ \mathbf{x}^{(k+1)} = [\mathbf{x}^{(k)} - (\rho\mathbf{D_A} + \mathbf{D_R})^{-1}(\mathbf{s}^{(k+1)} + \nabla\mathsf{R}(\mathbf{x}^{(k)}))]_{\mathcal{C}} \\ \boldsymbol{\zeta}^{(k+1)} \triangleq M\mathbf{A}_m'\mathbf{W}(\mathbf{A}_m\mathbf{x}^{(k+1)} - \mathbf{y}) \\ \mathbf{g}^{(k+1)} = \frac{\rho}{\rho+1}(\alpha\boldsymbol{\zeta}^{(k+1)} + (1-\alpha)\mathbf{g}^{(k)}) + \frac{1}{\rho+1}\mathbf{g}^{(k)} \\ \mathbf{h}^{(k+1)} = \alpha(\mathbf{D_A}\mathbf{x}^{(k+1)} - \boldsymbol{\zeta}^{(k+1)}) + (1-\alpha)\mathbf{h}^{(k)} \end{cases}$$

    decrease ρ using (4) with $n = kM + m$.
      end for
    end for
    $\tilde{\mathbf{x}}^{(i+1)} = \mathbf{x}^{(K)}$
    (2) Sparse Coding: with $\tilde{\mathbf{x}}^{(i+1)}$ fixed, the optimal sparse codes are $\tilde{\mathbf{z}}_j^{(i+1)} = H_\gamma(\mathbf{\Omega P}_j\tilde{\mathbf{x}}^{(i+1)}) \, \forall j$.
**end for**

---

patches with overlapping stride of 1 from the five 512×512 XCAT slices. We ran 2000 iterations of the alternating minimization algorithm proposed in [4] to make sure the learned ST is completely converged with $\lambda = 5.85 \times 10^{15}$, and $\gamma = 125$. Figure 1 shows 2D DCT and the well-conditioned pre-learned ST. Each row of these transforms is displayed as an 8×8 patch.

We simulated a 2D fan-beam CT scan using a 1024×1024 XCAT phantom slice, which is different from the learning slices, and $\Delta_x = \Delta_y = 0.4883$ mm. Noisy (Poisson noise) sinograms of size 888×984 were numerically generated with GE LightSpeed fan-beam geometry corresponding to a monoenergetic source with 10^5, 10^4 and 10^3 incident photons per ray and no scatter, respectively. We reconstructed a 512×512 image with a coarser grid, where $\Delta_x = \Delta_y = 0.9766$ mm. The statistical weighting matrix $\mathbf{W}$ is defined as a diagonal matrix with diagonal entries $w_i \triangleq \exp(-y_i)$.

We computed the root mean square error (RMSE) in Hounsfield units (HU), which is defined for a reconstructed CT image $\hat{\mathbf{x}} \in \mathbb{R}^{N_p}$ as RMSE $= \sqrt{\sum_{i=1}^{N_p}(\hat{x}_i - x_i^*)^2/N_p}$, where $\mathbf{x}^*$ is the ground truth image.

Initialized with FBP reconstructed images, the PWLS-EP method converges quickly using relaxed OS-LALM with 20 subsets. For PWLS-DCT and the proposed PWLS-ST meth-
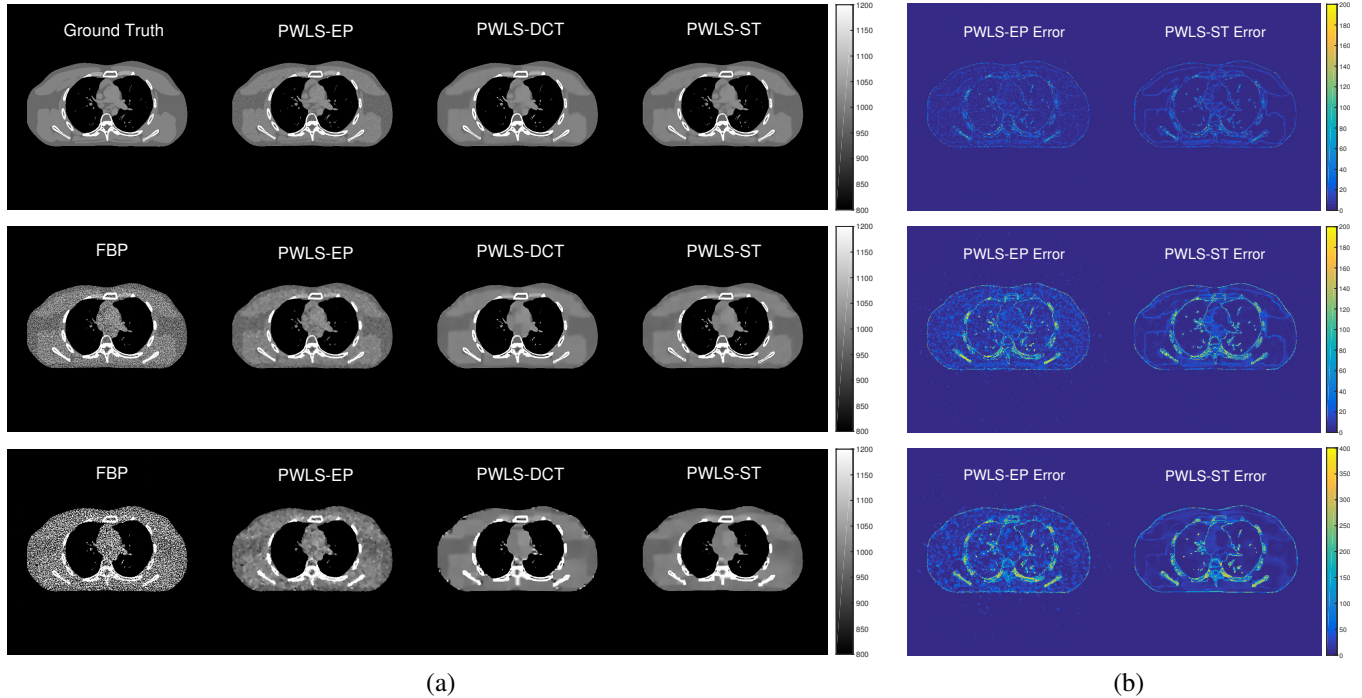
Fig. 2: From top to bottom the incident photon intensities are 10^5, 10^4, and 10^3. (a) Reconstructions. (b) Difference images (magnitudes) between PWLS-EP and PWLS-ST reconstructions and the ground truth image. All units in HU.

ods, we used the image obtained after a few iterations of the PWLS-EP method as initialization to realize faster convergence. We set γ as 25 to achieve a good trade-off between reconstructed image quality and convergence speed. The parameter β was determined empirically by sweeping over a large range of values and choosing the parameter that corresponded to the lowest RMSE, i.e., β was set as 3.75×10^5, 1.10×10^5, and 5.00×10^4 for incident photon intensities of 10^5, 10^4 and 10^3 respectively. In each iteration, we ran 3 inner iterations of the image update step with 4 subsets, i.e., $K = 3$, $M = 4$ in Algorithm 1. For PWLS-EP, β was set as $2^{5.7}$, $2^{7.3}$, and $2^{9.3}$ respectively for the three incident photon intensities, and the edge-preserving regularizer is $\varphi(t) \triangleq \delta^2 \left(\sqrt{1 + |t/\delta|^2} - 1 \right)$ ($\delta = 10$ HU).

Table 1: RMSE (HU) of reconstructions with FBP, PWLS-EP, PWLS-DCT and PWLS-ST for three levels of incident photon intensities.

Intensity	FBP	PWLS-EP	PWLS-DCT	PWLS-ST
10^5	34.7	**10.5**	19.5	13.3
10^4	69.5	24.6	29.3	**23.0**
10^3	202.3	49.1	53.1	**46.7**

Figure 2(a) shows the reconstructed images by FBP, PWLS-EP, PWLS-DCT and PWLS-ST. When the incident photon intensity is 10^5, although the PWLS-EP image has lower RMSE than the PWLS-ST image, the latter has no vis-

ible noise. When the incident photon intensities are 10^4 and 10^3, compared with FBP and PWLS-EP, PWLS-ST greatly improves image quality in terms of decreasing noise and retaining small structures. Figure 2(b) shows the difference images (magnitudes) between the PWLS-EP and PWLS-ST reconstructions and the ground truth image, for the three incident photon intensities.

Table 1 summarizes the RMSE of reconstructions with FBP, PWLS-EP, PWLS-DCT and PWLS-ST for three different incident photon intensities. For low dose cases, PWLS-ST further decreases the lowest RMSE achieved by PWLS-EP.

5. CONCLUSION

We present a PWLS-ST method that combines conventional PWLS reconstruction with regularization based on a sparsifying transform that is pre-learned from a dataset of numerous CT images, to improve the quality of reconstructed images in low dose CT imaging. Numerical experiments show the proposed PWLS-ST method may help reduce X-ray dose to a low level while still providing high quality image reconstructions. For future work, we will investigate PWLS with a union of sparsifying transforms [14] that could be pre-learned in an online manner [15] from large datasets. We also plan to compare our methods to the recent transform blind reconstruction framework [2]. We will apply the proposed PWLS-ST method to clinical CT data.

6. REFERENCES

[1] J-B. Thibault, K. Sauer, C. Bouman, and J. Hsieh, "A three-dimensional statistical approach to improved image quality for multi-slice helical CT," *Med. Phys.*, vol. 34, no. 11, pp. 4526–44, Nov. 2007.

[2] L. Pfister and Y. Bresler, "Adaptive sparsifying transforms for iterative tomographic reconstruction," in *Proc. 3rd Intl. Mtg. on image formation in X-ray CT*, 2014, pp. 107–10.

[3] S. Ravishankar and Y. Bresler, "Learning sparsifying transforms," *IEEE Trans. Sig. Proc.*, vol. 61, no. 5, pp. 1072–86, Mar. 2013.

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

[5] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Im. Proc.*, vol. 15, no. 12, pp. 3736–45, Dec. 2006.

[6] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Sig. Proc.*, vol. 54, no. 11, pp. 4311–22, Nov. 2006.

[7] R. Rubinstein, T. Peleg, and M. Elad, "Analysis K-SVD: A dictionary-learning algorithm for the analysis sparse model," *IEEE Trans. Sig. Proc.*, vol. 61, no. 3, pp. 661–677, Feb. 2013.

[8] Q. Xu, H. Yu, X. Mou, L. Zhang, J. Hsieh, and G. Wang, "Low-dose X-ray CT reconstruction via dictionary learning," *IEEE Trans. Med. Imag.*, vol. 31, no. 9, pp. 1682–97, Sept. 2012.

[9] L. Pfister and Y. Bresler, "Model-based iterative tomographic reconstruction with adaptive sparsifying transforms," in *Proc. SPIE 9020 Computational Imaging XII*, 2014, pp. 90200H–1–90200H–11.

[10] H. Nien and J. A. Fessler, "Relaxed linearized algorithms for faster X-ray CT image reconstruction," *IEEE Trans. Med. Imag.*, vol. 35, no. 4, pp. 1090–1098, Apr. 2016.

[11] H. Erdoğan and J. A. Fessler, "Ordered subsets algorithms for transmission tomography," *Phys. Med. Biol.*, vol. 44, no. 11, pp. 2835–51, Nov. 1999.

[12] S. Ravishankar and Y. Bresler, "MR image reconstruction from highly undersampled k-space data by dictionary learning," *IEEE Trans. Med. Imag.*, vol. 30, no. 5, pp. 1028–41, May 2011.

[13] W. P. Segars, M. Mahesh, T. J. Beck, E. C. Frey, and B. M. W. Tsui, "Realistic CT simulation using the 4D XCAT phantom," *Med. Phys.*, vol. 35, no. 8, pp. 3800–8, Aug. 2008.

[14] B. Wen, S. Ravishankar, and Y. Bresler, "Structured overcomplete sparsifying transform learning with convergence guarantees and applications," *Intl. J. Comp. Vision*, vol. 114, no. 2-3, pp. 137–167, 2015.

[15] S. Ravishankar, B. Wen, and Y. Bresler, "Online sparsifying transform learning – Part I: Algorithms," *IEEE J. Sel. Top. Sig. Proc.*, vol. 9, no. 4, pp. 625–636.