ADAPTIVE SPARSE MODELING AND SHIFTED-POISSON LIKELIHOOD BASED APPROACH FOR LOW-DOSE CT IMAGE RECONSTRUCTION

*Siqi Ye[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

Recent research in computed tomographic imaging has focused on developing techniques that enable reduction of the X-ray radiation dose without loss of quality of the reconstructed images or volumes. While penalized weighted-least squares (PWLS) approaches have been popular for CT image reconstruction, their performance degrades for very low dose levels due to the inaccuracy of the underlying WLS statistical model. We propose a new formulation for low-dose CT image reconstruction based on a shifted-Poisson model based likelihood function and a data-adaptive regularizer using the sparsifying transform model for images. The sparsifying transform is pre-learned from a dataset of patches extracted from CT images. The nonconvex cost function of the proposed penalized-likelihood reconstruction with sparsifying transforms regularizer (PL-ST) is optimized by alternating between a sparse coding step and an image update step. The image update step deploys a series of convex quadratic majorizers that are optimized using a relaxed linearized augmented Lagrangian method with ordered-subsets, reducing the number of (expensive) forward and backward projection operations. Numerical experiments show that for low dose levels, the proposed data-driven PL-ST approach outperforms prior methods employing a nonadaptive edge-preserving regularizer. PL-ST also outperforms prior PWLS-ST approach at very low X-ray doses.

Index Terms— Transform learning, low-dose CT, shifted-Poisson statistical model, sparse representations, machine learning

1. INTRODUCTION

X-ray Computed Tomography (CT) is a popular imaging technique in modern medical diagnosis and treatment. However, X-ray radiation can be harmful to human health, so there has been growing interest in techniques that enable reduced radiation dose without compromising image quality.

Two common strategies to reduce dose include decreasing the photon intensity at the X-ray source, and acquiring fewer projection views (angles) aka sparse-view CT. Both approaches yield degraded projection data, for which the conventional algebraic filtered back-projection (FBP) image reconstruction method yields poor image quality. In these settings, statistical image reconstruction (SIR) methods have been proposed and are popular.

Most SIR methods involve optimizing a cost composed of a data-fidelity term accounting for the statistical model of the measurement data and a regularization term modeling prior knowledge or expected characteristics of the target image. Often a Poisson + Gaussian model is used for the CT measurements, where the Poisson distribution models photon counting statistics and the electronic noise at the detector is assumed Gaussian. However, the associated likelihood function in this case lacks an elegant form. An approximate alternative but popular SIR method is penalized (or regularized) weighted least-squares (PWLS), where the logarithm of the measurements is modeled as a linear function of the image, up to additive Gaussian noise with projection view-dependent variance (hence WLS).

While PWLS works well at regular X-ray doses, it becomes inapplicable for very low-dose or noisy cases, when measurements need not be positive, and their logarithm would not exist. Techniques to resolve this problem include applying zero-weighting (in PWLS) on the non-positive measurements [1], replacing the non-positive components by some artificial positive values [2], or corrections using some recursive mean-preserving operations [3] and interpolation by sinogram smoothing or denoising [4, 5] to eliminate non-positive values. Nevertheless, these methods would introduce bias in the reconstructed image, and the spatial resolution of the reconstruction would be heavily degraded when the CT dose is ultra-low. Moreover, the correction of non-positive values and the non-linearity of the logarithm create challenges for estimation of the weighting parameters in PWLS. To address these challenges, a more robust statistical model for the CT (pre-log) measurements, the shifted-Poisson model that better approximates the Poisson + Gaussian model was proposed recently [6–8].

The quality of reconstructed images for SIR methods also depends highly on the efficacy of the image prior (regularizer) in modeling CT image properties. While total variation regularization has been successfully exploited for low-dose CT image reconstruction, there has been increasing interest in designing data-driven regularizers. The parameters of such image priors can be effectively learned from big data sets. Xu et al. [9] proposed a PWLS approach with a regularizer promoting sparsity of image patches in a redundant learned synthesis dictionary. An alternative approach uses a sparsifying transform model for image patches, wherein the patches are assumed approximately sparse in a learned transform domain. In contrast to the often non-convex and NP-hard synthesis dictionary learning problems [10, 11] and expensive learning algorithms, efficient methods have been proposed for learning sparsifying transforms (ST) with good convergence properties [12–14]. Moreover, sparse coding in the ST model involves simple thresholding, which is

an exact solution and is less expensive than synthesis sparse coding. Pfister et al. [15, 16] combined the PWLS method with transform learning and showed promising results. However, their approach learns transforms from the limited noisy measurements, which increases computation, and it does not exploit available big databases for learning the image prior.

Very recently, Zheng et al. [17] proposed a fast PWLS-ST approach for low-dose CT image reconstruction that uses a transform-based regularizer with the transform learned offline from a related dataset. Although the PWLS-ST approach benefits from the learned sparsifying transform, problems of bias in the reconstructed image remain due to the intrinsic downsides of the PWLS model. Here, we propose a new CT image reconstruction framework employing a shifted-Poisson model for measurements combined with an adaptive sparsifying transform-based prior for the image. The adaptive ST is learned from the patches of images from training CT datasets, which saves runtime during reconstruction. We refer to the proposed approach as Penalized-likelihood (PL) with ST regularizer, or PL-ST. Although the PL-ST objective is highly nonconvex, we present an efficient alternating optimization algorithm for the problem and demonstrate its advantages over prior approaches including PWLS methods in numerical experiments with low-dose CT data.

2. PROBLEM FORMULATIONS

2.1. Low-Dose CT Image Reconstruction Formulation

Our goal is to find the linear attenuation coefficients $\mathbf{x} \in \mathbb{R}^{N_p}$ or (vectorized) image that best fits the measured CT data $\mathbf{y} \in \mathbb{R}^{N_d}$. We propose to solve the following penalized-likelihood objective function $\Phi(\mathbf{x})$:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x} \geq 0} \Phi(\mathbf{x}), \ \ \Phi(\mathbf{x}) = -\mathsf{L}(\mathbf{x}) + \mathsf{R}(\mathbf{x}), \quad \text{(P0)}$$

where $-\mathsf{L}(\mathbf{x})$ is a negative log-likelihood function or data-fidelity term of the measurements that are modeled as shifted-Poisson random variables, i.e., $Y_i \sim \text{Poisson}\{I_0 e^{-[\mathbf{Ax}]_i} + \sigma^2\}$. Here, I_0 represents the number of incident photons per ray from the source, $\mathbf{A}$ is the $N_d \times N_p$ system matrix for CT, and σ^2 is the variance of the electronic noise. By denoting $l_i \triangleq [\mathbf{Ax}]_i$ and introducing functions $h_i(l_i)$ defined as:

$$h_i(l_i) \triangleq (I_0 e^{-l_i} + \sigma^2) - Y_i \log(I_0 e^{-l_i} + \sigma^2), \quad \text{(1)}$$

the negative log-likelihood is written as

$$-\mathsf{L}(\mathbf{x}) = \sum_{i=1}^{N_d} h_i(l_i). \quad \text{(2)}$$

The penalty or regularizer $\mathsf{R}(\mathbf{x})$ in (P0) is based on a pre-learned sparsifying transform $\mathbf{\Omega} \in \mathbb{R}^{v \times v}$ [13] for image patches as follows:

$$\mathsf{R}(\mathbf{x}) \triangleq \min_{\{\mathbf{z}_j\}} \beta \sum_{j=1}^{N} \left\{ \|\mathbf{\Omega P}_j \mathbf{x} - \mathbf{z}_j\|_2^2 + \gamma_c^2 \|\mathbf{z}_j\|_0 \right\}. \quad \text{(3)}$$

We assume N vectorized patches, each with v voxels that are extracted from $\mathbf{x}$, with $\mathbf{P}_j \in \mathbb{R}^{v \times N_p}$ being the patch extraction operator. The subscript j indexes the jth patch. Vector $\mathbf{z}_j \in \mathbb{R}^v$ denotes the sparse approximation in the transform domain of $\mathbf{P}_j \mathbf{x}$ and the weight γ_c^2 (with $\gamma_c > 0$) controls the sparsity level. The ℓ_0 "norm" counts the number of nonzero elements in $\mathbf{z}_j$. The positive parameter β in (3) helps balance the data-fidelity term and the regularizer, providing a trade-off between noise and image resolution.

2.2. Sparsifying Transform Learning

We learn the sparsifying transform $\mathbf{\Omega}$ from a dataset of patches extracted from regular dose CT images. This involves solving the following transform learning problem [12, 13]:

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

Here, N' is the number of patches used for training, $\mathbf{X} \in \mathbb{R}^{v \times N'}$ is a matrix whose columns are the vectorized training patches, and $\mathbf{Z} \in \mathbb{R}^{v \times N'}$ is a matrix of sparse coefficients, whose columns $\{\mathbf{Z}_i\}$ denote the sparse approximations in the transform domain of the training patches. The penalty $\|\mathbf{\Omega}\|_F^2 - \log|\det \mathbf{\Omega}|$ prevents trivial solutions in (P1) and helps control the condition number of the learned transform [12]. Parameters λ and γ are positive scalars.

3. ALGORITHM

In the following, first we briefly describe the transform learning algorithm for (P1) and then present the CT image reconstruction method for optimizing Problem (P0).

3.1. Training Sparsifying Transforms

The transform learning problem (P1) involves optimization with respect to the variables $\mathbf{\Omega}$ and $\mathbf{Z}$. We adopt the recent efficient alternating minimization approach for (P1) [13] that is proven to converge to (at least) the local minimizers in the nonconvex problem. The procedure involves a *transform update step* (updating $\mathbf{\Omega}$) and a *sparse coding step* (updating $\mathbf{Z}$), which are discussed next.

3.1.1. Transform update step:

With $\mathbf{Z}$ fixed, the optimization with respect to $\mathbf{\Omega}$ results in the following sub-problem:

$$\min_{\mathbf{\Omega}} \|\mathbf{\Omega X} - \mathbf{Z}\|_F^2 + \lambda\big(\|\mathbf{\Omega}\|_F^2 - \log|\det \mathbf{\Omega}|\big). \quad \text{(4)}$$

The optimal solution is computed via a singular value decomposition (SVD) [13]. Let $\mathbf{L}$ be the square root matrix in $\mathbf{XX}^T + \lambda \mathbf{I} = \mathbf{LL}^T$, where $\mathbf{I}$ denotes the identity matrix. Then, denoting the full SVD of $\mathbf{L}^{-1}\mathbf{XZ}^T$ as $\mathbf{Q\Sigma R}^T$, a global minimizer of (4) is as follows: $\hat{\mathbf{\Omega}} = 0.5\mathbf{R}(\mathbf{\Sigma} + (\mathbf{\Sigma}^2 + 2\lambda \mathbf{I})^{\frac{1}{2}})\mathbf{Q}^T \mathbf{L}^{-1}$.

3.1.2. Sparse coding step:

With fixed $\mathbf{\Omega}$, we solve (P1) for $\mathbf{Z}$ as follows:

$$\min_{\mathbf{Z}} \|\mathbf{\Omega X} - \mathbf{Z}\|_F^2 + \sum_{i=1}^{N'} \gamma^2 \|\mathbf{Z}_i\|_0. \quad \text{(5)}$$

The optimal $\mathbf{Z}$ is computed via thresholding as $\hat{\mathbf{Z}}_i = H_\gamma(\mathbf{\Omega X}_i) \ \forall i$, where $H_\gamma(\cdot)$ is the hard-thresholding operator that sets vector entries with magnitude below γ to zero and leaves other entries unaffected.

3.2. Optimization Algorithm for (P0)

We optimize the reconstruction problem (P0) with regularizer (3) by alternating minimization that alternates between updating $\mathbf{x}$ (*image update step*) and $\{\mathbf{z}_j\}$ (*sparse coding step*). These steps are discussed next.

3.2.1. Image Update Step

In this step, we fix the sparse codes $\{\mathbf{z}_j\}$ and find a non-negative image $\hat{\mathbf{x}}$ by optimizing the following cost consisting of the negative log-likelihood and a sparsification error (of patches):

$$\Phi_1(\mathbf{x}) = -\mathsf{L}(\mathbf{x}) + \beta \sum_{j=1}^{N} \|\mathbf{\Omega P}_j \mathbf{x} - \mathbf{z}_j\|_2^2. \tag{6}$$

To optimize $\Phi_1(\mathbf{x})$, we design a series of quadratic surrogate functions (majorizers) for the shifted Poisson data-fit term $-\mathsf{L}(\mathbf{x})$ (which is not convex when σ^2 is nonzero) and minimize the (convex) surrogate functions together with the sparsification error using the relaxed linearized augmented Lagrangian method with ordered-subsets (relaxed OS-LALM method) [18].

The image $\mathbf{x}$ is updated iteratively, and the majorizer for the negative log-likelihood in the nth such inner iteration is:

$$\phi(\mathbf{x}; \mathbf{x}^n) = -\mathsf{L}(\mathbf{x}^n) + \boldsymbol{d}_h(\mathbf{l}^n)\mathbf{A}(\mathbf{x} - \mathbf{x}^n) \\ + \frac{1}{2}(\mathbf{x} - \mathbf{x}^n)^T \mathbf{A}^T \mathbf{D}(c_i(l_i^n))\mathbf{A}(\mathbf{x} - \mathbf{x}^n), \tag{7}$$

where $(\cdot)^n$ represents (estimated) quantities for the nth (inner) iteration, and $\boldsymbol{d}_h(\mathbf{l}^n)$ is a length N_d row vector capturing gradient information, whose entries are $[\dot{h}_i(l_i^n)]_{i=1}^{N_d}$. Matrix $\mathbf{D}(c_i(l_i^n))$ is a N_d-dimensional diagonal matrix whose pivots (diagonal entries) are curvatures $c_i(l_i^n)$ of the parabola. We use the optimum curvatures defined as follows [19], where $\ddot{h}$ denotes the second (scalar-valued) derivative:

$$c_i(l_i^n) = \begin{cases} \left[2\frac{h_i(0) - h_i(l_i^n) + \dot{h}_i(l_i^n)(l_i^n)}{(l_i^n)^2}\right]_+, & l_i^n > 0 \\ \left[\ddot{h}_i(0)\right]_+, & l_i^n = 0. \end{cases} \tag{8}$$

Using the above majorizer, we replace minimizing (6) with iteratively minimizing the following problem for each n:

$$\min_{\mathbf{x} \geq 0} \left\{ \phi(\mathbf{x}; \mathbf{x}^n) + \beta \sum_{j=1}^{N} \|\mathbf{\Omega P}_j \mathbf{x} - \mathbf{z}_j\|_2^2 \right\}. \tag{9}$$

For simplicity, we replace the notation $\mathbf{D}(c_i(l_i^n))$ with $\mathbf{D}^n$ hereafter. To solve (9), we rewrite (7) as follows:

$$\phi(\mathbf{x}; \mathbf{x}^n) = \frac{1}{2} \Big\{ (\mathbf{x} - \mathbf{x}^n)^T \mathbf{A}^T \mathbf{D}^n \mathbf{A}(\mathbf{x} - \mathbf{x}^n) \\ + 2\boldsymbol{d}_h(\mathbf{l}^n)(\mathbf{D}^n)^{-\frac{1}{2}}(\mathbf{D}^n)^{\frac{1}{2}} \mathbf{A}(\mathbf{x} - \mathbf{x}^n) \\ + \boldsymbol{d}_h(\mathbf{l}^n)(\mathbf{D}^n)^{-\frac{1}{2}}[\boldsymbol{d}_h(\mathbf{l}^n)(\mathbf{D}^n)^{-\frac{1}{2}}]^T \Big\} \\ - \mathsf{L}(\mathbf{x}^n) - \frac{1}{2}\boldsymbol{d}_h(\mathbf{l}^n)(\mathbf{D}^n)^{-\frac{1}{2}}[\boldsymbol{d}_h(\mathbf{l}^n)(\mathbf{D}^n)^{-\frac{1}{2}}]^T. \tag{10}$$

Ignoring constant terms, we get the following equivalent form:

$$\phi(\mathbf{x}; \mathbf{x}^n) \equiv \frac{1}{2}\|\mathbf{y}_e^n - \mathbf{A}\mathbf{x}\|_\mathbf{W}^2, \tag{11}$$

where "$\equiv$" means equal to within irrelevant constants independent of $\mathbf{x}$, $\mathbf{y}_e^n = \mathbf{A}\mathbf{x}^n - (\mathbf{D}^n)^{-1}[\boldsymbol{d}_h(\mathbf{l}^n)]^T$ and $\mathbf{W} \triangleq \mathbf{D}^n = \mathrm{diag}\{c_i(l_i^n)\}$ (we drop the superscript n in $\mathbf{W}$ for simplicity). Thus, (9) becomes the following quadratic problem:

$$\min_{\mathbf{x} \geq 0} \left\{ \frac{1}{2}\|\mathbf{y}_e^n - \mathbf{A}\mathbf{x}\|_\mathbf{W}^2 + \beta \sum_{j=1}^{N} \|\mathbf{\Omega P}_j \mathbf{x} - \mathbf{z}_j\|_2^2 \right\}. \tag{12}$$

With $R_2(\mathbf{x})$ denoting the second term in (12), we optimize (12) using the relaxed OS-LALM method [18] by iterating over the following updates, where for each k, we further iterate over $1 \leq m \leq M$,

the M ordered subsets:

$$\begin{cases} \mathbf{s}^{(k+1)} = \rho(\mathbf{D_A}\tilde{\mathbf{x}}^{(k)} - \boldsymbol{\eta}^{(k)}) + (1-\rho)\mathbf{g}^{(k)}, \\ \tilde{\mathbf{x}}^{(k+1)} = [\tilde{\mathbf{x}}^{(k)} - (\rho\mathbf{D_A} + \mathbf{D_R})^{-1}(\mathbf{s}^{(k+1)} + \nabla R_2(\tilde{\mathbf{x}}^{(k)}))]_\mathcal{C}, \\ \boldsymbol{\zeta}^{(k+1)} \triangleq M\mathbf{A}_m^T \mathbf{W}_m(\mathbf{A}_m\tilde{\mathbf{x}}^{(k+1)} - (\mathbf{y}_e^n)_m), \\ \mathbf{g}^{(k+1)} = \frac{\rho}{\rho+1}(\alpha\boldsymbol{\zeta}^{(k+1)} + (1-\alpha)\mathbf{g}^{(k)}) + \frac{1}{\rho+1}\mathbf{g}^{(k)}, \\ \boldsymbol{\eta}^{(k+1)} = \alpha(\mathbf{D_A}\tilde{\mathbf{x}}^{(k+1)} - \boldsymbol{\zeta}^{(k+1)}) + (1-\alpha)\boldsymbol{\eta}^{(k)}. \end{cases} \tag{13}$$

The image $\mathbf{x}^{n+1}$ is then updated as the output ($\tilde{\mathbf{x}}$ part) of relaxed OS-LALM, whose iterates are initialized based on the (previous inner iterate) image $\mathbf{x}^n$. In (13), $\mathbf{D_A}$ is a diagonal majorizing matrix of $\mathbf{A}^T\mathbf{W}\mathbf{A}$, e.g., $\mathbf{D_A} \triangleq \mathrm{diag}\{|\mathbf{A}|^T|\mathbf{W}||\mathbf{A}|\mathbf{1}\}$ [17], and $[\cdot]_\mathcal{C}$ projects the input vector onto the constraint $\mathbf{x} \geq 0$. Matrix $\mathbf{A}_m$ is the subset forward projection matrix (mth subset of rows of $\mathbf{A}$) and $\mathbf{W}_m$ is the corresponding sub-matrix of $\mathbf{W}$. The (over-)relaxation parameter $\alpha \in [1,2)$ and $\rho > 0$ is the AL penalty parameter decreasing with the subiterations t (function of k and m) in relaxed OS-LALM as

$$\rho_t(\alpha) = \begin{cases} 1 & , t = 0 \\ \frac{\pi}{\alpha(t+1)}\sqrt{1 - \left(\frac{\pi}{2\alpha(t+1)}\right)^2} & , \text{otherwise.} \end{cases} \tag{14}$$

Matrix $\mathbf{D_R} \succeq \nabla^2 R_2(\mathbf{x}) = 2\beta \sum_{j=1}^{N} \mathbf{P}_j^T \mathbf{\Omega}^T \mathbf{\Omega} \mathbf{P}_j$, e.g., $\mathbf{D_R} \triangleq 2\beta \sum_{j=1}^{N} \mathbf{P}_j^T \mathbf{P}_j \lambda_{\max}(\mathbf{\Omega}^T\mathbf{\Omega})$, where $\lambda_{\max}(\mathbf{\Omega}^T\mathbf{\Omega})$ denotes the maximal eigenvalue of $\mathbf{\Omega}^T\mathbf{\Omega}$ and $\sum_{j=1}^{N} \mathbf{P}_j^T \mathbf{P}_j$ is a diagonal matrix with the diagonal entries corresponding to pixel locations and with values equal to the number of patches overlapping each pixel. If periodically positioned overlapping patches with patch stride 1 (pixel) are used in (P0), and patches overlapping image boundaries are assumed to wrap around the opposite side of the image, then $\sum_{j=1}^{N} \mathbf{P}_j^T \mathbf{P}_j = v\mathbf{I}$ with v the number of pixels in a patch. In this case, $\mathbf{D_R} = 2\beta v\lambda_{\max}(\mathbf{\Omega}^T\mathbf{\Omega})\mathbf{I}$.

3.2.2. Sparse Coding Step

Here, we update $\{\mathbf{z}_j\}$ by solving the following problem with $\mathbf{x}$ fixed to its most recent estimate:

$$\min_{\{\mathbf{z}_j\}} \sum_{j=1}^{N} \left\{ \|\mathbf{\Omega P}_j \mathbf{x} - \mathbf{z}_j\|_2^2 + \gamma_c^2 \|\mathbf{z}_j\|_0 \right\}. \tag{15}$$

The optimal sparse codes $\hat{\mathbf{z}}_j$ are computed in closed-form via hard-thresholding in a similar manner as in Section 3.1.2, i.e., $\hat{\mathbf{z}}_j = H_{\gamma_c}(\mathbf{\Omega P}_j \mathbf{x}) \, \forall \, j$.

4. NUMERICAL EXPERIMENTS

4.1. Framework

We simulated a 2D fan-beam CT scan using a 1024×1024 XCAT phantom slice [20] with $\Delta_x = \Delta_y = 0.4883$ mm. An 888×984 sinogram was numerically generated with GE LightSpeed fan-beam geometry and with $\sigma = 5$. We reconstructed an image of size 512×512 with $\Delta_x = \Delta_y = 0.9766$ mm.

For training the sparsifying transform, we extracted 8×8 overlapping patches with a stride of 1×1 from five 512×512 XCAT phantom slices (different from the testing data). The algorithm in Section 3.1 for optimizing (P1) was initialized with the 2D DCT (for $\mathbf{\Omega}$) and was run for 1000 iterations to ensure convergence. Several choices for γ (in (P1)) were tested, and $\gamma = 110$ was observed to provide the best reconstruction results. λ was set to be 5.85×10^{14}.
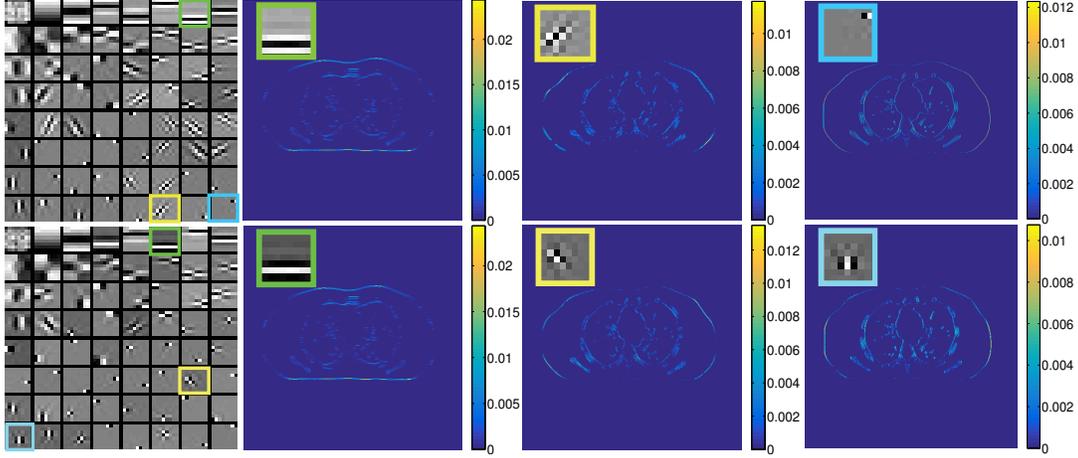
Fig. 1: Learned sparsifying transforms (in first column) at $\gamma = 125$ (top) and $\gamma = 110$ (bottom), with the transform rows shown as 8×8 patches. The sparse codes learned in (P0) at $I_0 = 10^4$ are visualized as images (magnitudes shown) for selected rows (or filters) of the transforms. Each pixel in the images corresponds to a specific component (corresponding to the specific transform row) of the sparse code of an image patch.
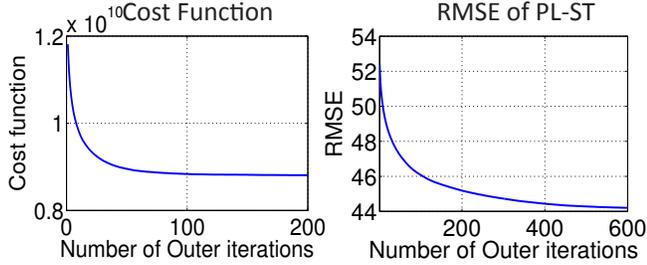
Fig. 2: Evolution of the objective function in (P0) and the reconstruction RMSE for $I_0 = 2 \times 10^3$.

Fan-beam CT measurements were generated at different (low) dose levels (number of incident photons per ray) $I_0 = 10^4, 5 \times 10^3, 3 \times 10^3$, and 2×10^3. The attenuation was measured in units of mm^{-1}. For each dose level, we investigated the performance of several reconstruction methods: the FBP method, the PWLS statistical model with a non-adaptive edge-preserving regularizer (PWLS-EP), the PL model with EP regularizer (PL-EP), the PWLS-ST approach, and the proposed PL-ST method. The reconstructed images were evaluated both visually and using the root mean square error (RMSE) and structural similarity index (SSIM) metrics [9, 21]. For displaying the results, the unit mm^{-1} was converted to Hounsfield unit (HU). All simulations were run on a workstation with two 2.7 GHz 12-core Intel Xeon E5-2697 processors. The regularizer $R(\mathbf{x}) = \sum_{j=1}^{N_p} \sum_{k \in N_j} \kappa_j \kappa_k \varphi(x_j - x_k)$ is used for EP methods with $\varphi(t) \triangleq \delta^2(\sqrt{1 + |t/\delta|^2} - 1)$ ($\delta = 2 \times 10^{-4}$mm), N_j is the neighborhood, and κ_j denotes parameter encouraging noise uniformity [22]. The PWLS-EP and PL-EP methods were initialized with the FBP reconstruction and the regularization parameter β was empirically set in each experiment to achieve good reconstruction quality.

The PWLS-ST and PL-ST methods (that exploit the learned transform $\mathbf{\Omega}$) used 8×8 patches with a stride of 1 pixel, and were initialized with PWLS-EP reconstructions. The parameters β and γ_c for the ST methods were set as 8×10^4 and 2×10^{-4}, respectively, which worked well in our experiments. Similar to the EP methods, we used patch-based weights $\{\gamma_c^2 \tau_j\}$ for the ℓ_0 terms in (3), where $\tau_j \triangleq \|\mathbf{P}_j \mathbf{\kappa}\|_1 / v$ with $\mathbf{\kappa}$ (same size as $\mathbf{x}$) whose elements κ_j are

defined similarly as for the EP methods [22]. Adding patch-based weights in (3) only modifies the thresholds in the sparse coding step of the ST-based reconstruction algorithms.

4.2. Behavior of the Proposed Method

This section examines the behavior of the learned sparsifying transform in the image reconstruction process. The empirical convergence behavior of the algorithm for (P0) is also illustrated.

4.2.1. How do the Learned Models work?

Fig. 1 shows the transforms learned for $\gamma = 125$ and $\gamma = 110$ in (P1). The rows of the learned $\mathbf{\Omega}$ are shown as 8×8 patches called transform atoms. The transforms show various directional or gradient like features adapted based on training data.

To study the effect of the learned models in (P0), we visualize entries of the learned sparse codes $\{\mathbf{z}_j\}$ (in (P0)) corresponding to specific rows of $\mathbf{\Omega}$ in Fig. 1. We work with $I_0 = 10^4$ and show sparse codes obtained using two different learned transforms (with $\gamma = 125$ and $\gamma = 110$). Each pixel in the sparse code image for a specific transform row corresponds to the sparse coefficient (entry in $\mathbf{z}_j$) for an image patch (the coefficient is placed in the top left corner of the patch). Because each row of the learned transforms is applied to all the overlapping image patches and the result is thresholded to generate the patch coefficients, it is clear that the rows of $\mathbf{\Omega}$ act like sparsifying filters. The sparse code images capture different kinds of edges depending on the characteristics of the transform row. Some filters capture horizontal edges, while others emphasize directed edges in the reconstruction. Importantly, the learned transforms with a variety of filters/features ensure very sparse representations.

The transform $\mathbf{\Omega}$ learned with $\gamma = 110$ led to better reconstruction quality for the testing data. Hence, we used it in the remainder of this section.

4.2.2. Convergence of Algorithm for (P0)

Here, we investigate the behavior of the proposed PL-ST scheme further. Fig. 2 shows the evolution of the objective function (in (P0))
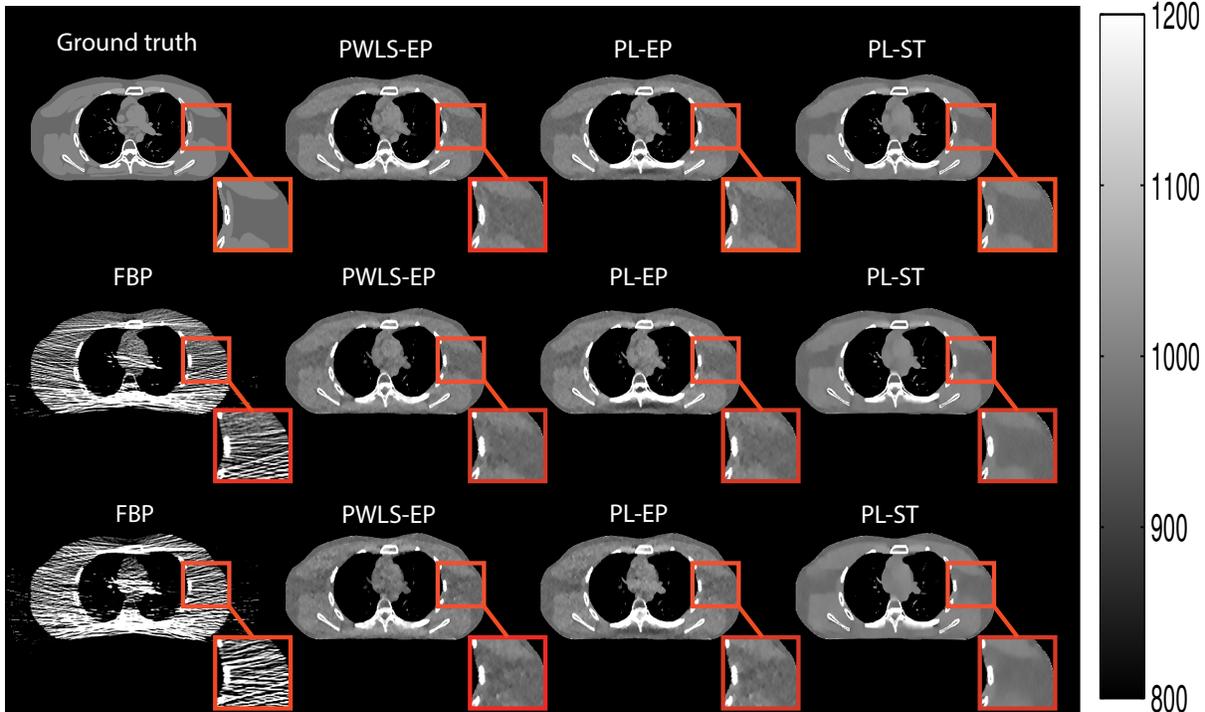
Fig. 3: Reconstructed images for different methods for $I_0 = 10^4, 5 \times 10^3$, and 3×10^3 (top to bottom), respectively. The ground-truth image is shown as the leftmost image in the first row.

and reconstruction RMSE over the outer iterations (alternations between image update and sparse coding) of the PL-ST algorithm for dose level $I_0 = 2 \times 10^3$. In each outer iteration, we ran 5 inner (n) iterations of image update (surrogate update), where the relaxed OS-LALM method was run for 3 iterations with 3 ordered subsets. The cost function for PL-ST converged quickly. The RMSE metric improved (or decreased) significantly during the initial $(100 - 200)$ iterations, with more iterations providing only a marginal additional improvement.

4.3. Results and Comparisons

Table 1 shows the RMSE and SSIM values of reconstructions obtained with various methods for multiple dose levels. It is clear that the proposed PL-ST method provides the best RMSE and SSIM values at all doses, with larger improvements achieved over competing methods at smaller dose levels. Both PWLS-ST and PL-ST outperform the conventional FBP and the non-adaptive PWLS-EP and PL-EP methods. Importantly, the proposed PL-ST outperforms PWLS-ST, with nearly 4 HU better RMSE at $I_0 = 2 \times 10^3$. These results indicate the superiority of the shifted-Poisson statistical model and the data-driven transform-based regularization in the proposed framework.

Fig. 3 shows the ground truth test image and the reconstructed images for different methods at various dose levels. Clearly, when the photon dose decreases, the FBP reconstruction shows heavy streak artifacts and noise. Statistical approaches using the nonadaptive edge-preserving regularizer, i.e., PLWS-EP and PL-EP, significantly reduce the artifacts in the FBP result. But the EP results, especially at low doses, still suffer from high noise and loss of details such as tissue boundaries. On the other hand, reconstructions with the proposed PL-ST display reduced noise or artifacts. For example, in the first row of Fig. 3, the noise is mostly eliminated in

the PL-ST result compared with PWLS-EP or PL-EP, and the tissue boundaries are clearer. Smaller structures such as blood vessels (within the central air regions) are also reconstructed better by the proposed PL-ST at various doses.

Fig. 4 compares the reconstructions for PWLS-ST and PL-ST at the low dose $I_0 = 2 \times 10^3$. Zoom-ins of the reconstructions show that PL-ST reconstructs tissue edges better than the PWLS-ST approach. The reconstruction errors (magnitudes) are also shown in Fig. 4. Clearly, PL-ST provides much smaller reconstruction errors for various soft tissues and edges compared to PWLS-ST.

5. CONCLUSION

We presented a novel PL-ST approach for low-dose CT image reconstruction based on a shifted-Poisson likelihood function and a learned transform-based regularizer. The proposed alternating algorithm for image reconstruction involves a simple and efficient sparse coding step and an iterative image update step that optimizes quadratic majorizing functions (of a nonconvex cost). Numerical experiments with low-dose fan beam CT scans of the XCAT phantom demonstrate that PL-ST outperforms prior nonadaptive image reconstruction techniques based on the edge-preserving regularizer, and moreover outperforms PWLS-ST especially at very low doses. For future work, we plan to apply the proposed approach to low-dose 3D CT data and evaluate its performance. Currently PL-ST takes longer time per outer iteration than PWLS-ST due to the surrogate function updates. We will investigate accelerating the PL-ST method in future work.

6. REFERENCES

[1] I. A. Elbakri and J. A. Fessler, "Statistical image reconstruction for polyenergetic X-ray computed tomography," *IEEE Trans.*

Table 1: RMSE in HU and SSIM (at convergence) of reconstructions at various dose levels (I_0) using the FBP, PWLS-EP, PL-EP, PWLS-ST, and PL-ST methods. The best results are in bold.

I_0	FBP		PWLS-EP		PL-EP		PWLS-ST		PL-ST	
	RMSE	SSIM	RMSE	SSIM	RMSE	SSIM	RMSE	SSIM	RMSE	SSIM
1×10^4	141.9	0.414	32.5	0.957	32.8	0.955	29.3	0.968	**29.1**	**0.968**
5×10^3	199.8	0.384	38.7	0.941	40.1	0.940	34.6	0.964	**33.9**	**0.965**
3×10^3	257.6	0.378	44.4	0.923	43.4	0.923	40.3	0.956	**38.6**	**0.959**
2×10^3	327.7	0.341	52.4	0.906	50.3	0.908	48.0	0.948	**44.1**	**0.954**
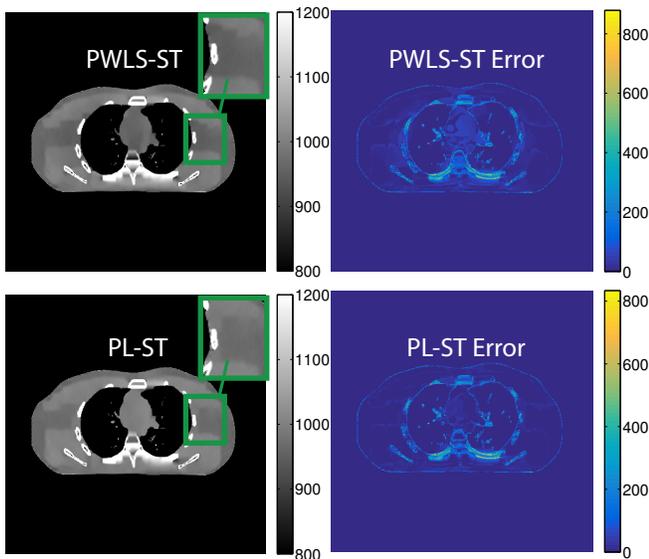
Fig. 4: Reconstructed images and the corresponding error images with PWLS-ST and PL-ST for $I_0 = 2 \times 10^3$.

Med. Imag., vol. 21, no. 2, pp. 89–99, 2002.

[2] J. Wang, T. Li, H. B. Lu, and Z. R. Liang, "Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose x-ray computed tomography," *IEEE Trans. Med. Imag.*, vol. 25, no. 10, pp. 1272–1283, 2006.

[3] J.-B. Thibault, C. A. Bouman, K. D. Sauer, and J. Hsieh, "A recursive filter for noise reduction in statistical iterative tomographic imaging," in *Proc. SPIE 6065, Computational Imaging IV*, 2006, vol. 6065, pp. 60650X–60650X–10.

[4] J. Hsieh, "Adaptive streak artifact reduction in computed tomography resulting from excessive xray photon noise," *Med. Phys.*, vol. 25, no. 11, pp. 2139–2147, 1998.

[5] O. Demirkaya, "Reduction of noise and image artifacts in computed tomography by nonlinear filtration of projection images," in *Proc. SPIE 4322, Medical Imaging*, 2001, vol. 4322, pp. 917–923.

[6] P. J. La Riviere, "Penalized-likelihood sinogram smoothing for low-dose CT," *Med. Phys.*, vol. 32, no. 6, pp. 1676–1683, 2005.

[7] P. J. La Riviere, J. G. Bian, and P. A. Vargas, "Penalized-likelihood sinogram restoration for computed tomography," *IEEE Trans. Med. Imag.*, vol. 25, no. 8, pp. 1022–1036, 2006.

[8] J. Y. Xu and B. M. W. Tsui, "Electronic noise modeling in statistical iterative reconstruction," *IEEE Trans. Image Process.*, vol. 18, no. 6, pp. 1228–1238, 2009.

[9] 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.

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

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

[12] S. Ravishankar and Y. Bresler, "Learning sparsifying transforms," *IEEE Trans. Signal Process.*, vol. 61, no. 5, pp. 1072–1086, 2013.

[13] 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, 2015.

[14] 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, may 2015.

[15] 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.

[16] 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.

[17] X. Zheng, Z. Lu, S. Ravishankar, Y. Long, and J. A. Fessler, "Low dose CT image reconstruction with learned sparsifying transform," in *IEEE 12th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*, 2016, pp. 1–5.

[18] 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.

[19] H. Erdogan and J. A. Fessler, "Monotonic algorithms for transmission tomography," *IEEE Trans. Med. Imag.*, vol. 18, no. 9, pp. 801–814, 1999.

[20] 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.

[21] Y. Zhang, X. Mou, G. Wang, and H. Yu, "Tensor-based dictionary learning for spectral CT reconstruction," *IEEE Trans. Med. Imag.*, vol. 36, no. 1, pp. 142–154, 2017.

[22] 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, 2015.