Sparse-View X-Ray CT Reconstruction Using \( \ell_1 \) 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 \( \ell_1 \) regularization with learned sparsifying transform (PWLS-ST-\( \ell_1 \)), and an algorithm for PWLS-ST-\( \ell_1 \). 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 \( \ell_2 \) 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 \( \ell_2 \) 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 \( \ell_1 \) regularization with learned ST (PWLS-ST-\( \ell_1 \)) 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-\( \ell_1 \) 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 \( \ell_2 \) regularization with a learned ST (PWLS-ST-\( \ell_2 \)) [8].

II. METHODS

A. Offline Learning Sparsifying Transform

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

\[
\min_{\Psi \in \mathbb{R}^{n \times n}, \{z_j' \in \mathbb{R}^n\}} \sum_{j=1}^{J'} \| \Psi x_j' - z_j' \|_2^2 + \gamma' \| z_j' \|_0 + \tau \left( \xi \| \Psi \|_F^2 - \log |\det \Psi| \right)
\]

where \( \Psi \in \mathbb{R}^{n \times n} \) is a square ST, \( \{x_j' \in \mathbb{R}^n : j = 1, \ldots, J'\} \) is a set of patches extracted from training data, \( z_j' \in \mathbb{R}^n \) is the sparse code corresponding to the \( j \)'th patch \( x_j' \), \( J' \) is the total number of the image patches, and \( \gamma', \tau, \xi \in \mathbb{R} \) are regularization parameters. The \( \ell_0 \) function \( \| \cdot \|_0 \) counts the nonzero elements in a vector.

B. CT Reconstruction Model Using \( \ell_1 \)-Regularization with Learned Sparsifying Transform: PWLS-ST-\( \ell_1 \)

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

\[
\min_{x \in \mathbb{R}^N, z \in \mathbb{R}^J} \frac{1}{2} \| y - \mathbf{Ax} \|_W^2 + \lambda \left( \| \Psi x - z \|_1 + \gamma \| z \|_0 \right),
\]

where

\[
\Psi = \begin{bmatrix} \Psi P_1 & \vdots & \vdots \\ \vdots & \ddots & \vdots \\ \vdots & \vdots & \Psi P_J \end{bmatrix}
\quad \text{and} \quad z = \begin{bmatrix} z_1 \\ \vdots \\ z_J \end{bmatrix}.
\]

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

The term \( \| \Psi x - z \|_1 \) denotes an \( \ell_1 \)-based sparsification error [13]. We expect \( \ell_1 \) to be more robust to model mismatch than the \( \ell_2 \)-based sparsification error used in [8]. In particular,
the proposed \( \ell_1 \)-based sparsification error term, \( \| \widetilde{\Psi} x - z \|_1 \), preserves edge sharpness better than \( \| \Psi x - z \|_2^2 \) in [8]; see Fig. 1 and Table I.

C. Proposed Algorithm for PWLS-\( \ell_1 \)

To solve (2), our proposed algorithm alternates between updating the image \( x \) (image update step) and the sparse codes \( z \) (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 \( z \). 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 \( z \), we update the image \( x \) by augmenting (2) with auxiliary variables:

\[
\min_{x,d_a,d_{\psi}} \frac{1}{2} \| y - d_a \|_W^2 + \lambda \| d_{\psi} \|_1
\]

subject to

\[
\begin{bmatrix}
  d_a \\
  d_{\psi}
\end{bmatrix} = \begin{bmatrix}
  A \\
  \widetilde{\Psi}
\end{bmatrix} x - \begin{bmatrix}
  0 \\
  z
\end{bmatrix}.
\]

The corresponding augmented Lagrangian has the form

\[
\frac{1}{2} \| y - d_a \|_W^2 + \lambda \| d_{\psi} \|_1 + \mu \frac{1}{2} \| d_a - Ax - b_a \|_2^2 + \frac{\mu \lambda}{2} \| d_{\psi} - (\widetilde{\Psi} x - z) - b_{\psi} \|_2^2.
\]

We descend/ascend this augmented Lagrangian using the following iterative updates of the primal, auxiliary, dual variables—\( x \), \( \{d_a, d_{\psi}\} \), and \( \{b_a, b_{\psi}\} \), respectively:

\[
x^{(i+1)} = \left( A^T A + \nu \sum_{j=1}^J P_j \Psi^T \Psi P_j \right)^{-1} \left( A^T (d_a^{(i)} - b_{\psi}^{(i)}) + \nu \widetilde{\Psi}^T (d_{\psi}^{(i)} - b_{\psi}^{(i)} + z) \right);
\]

\[
d_a^{(i+1)} = (W + \mu I_m)^{-1} \left( W y + \mu \left( A x^{(i+1)} + b_a^{(i)} \right) \right);
\]

\[
d_{\psi}^{(i+1)} = \text{softshrink} \left( \left( \Psi x^{(i+1)} - z + b_{\psi}^{(i)} \right) u, \frac{\lambda}{\mu \nu} \right),
\]

\[
u = 1, \ldots, nJ;
\]

\[
b_a^{(i+1)} = b_a^{(i)} - \left( d_a^{(i+1)} - Ax^{(i+1)} \right);
\]

\[
b_{\psi}^{(i+1)} = b_{\psi}^{(i)} - \left( d_{\psi}^{(i+1)} - \left( \Psi x^{(i+1)} - z \right) \right),
\]

where the soft-shrinkage operator is defined by

\[
\text{softshrink}(\alpha, \beta) := \text{sign}(\alpha) \max(\alpha - \beta, 0).
\]

To approximately solve (3), we use the preconditioned conjugate gradient (PCG) method with a circulant preconditioner \( M \) for \( A^T A + \nu \sum_{j=1}^J P_j \Psi^T \Psi P_j \). For the two-dimensional (2D) CT problem, a circulant preconditioner is well suited because it is effective for the “nearly” shift-invariant matrix \( A^T A \) [2], [4] and 2) \( \sum_{j=1}^J P_j \Psi^T \Psi P_j \) is a block

circulant block (BCCB) matrix when we use the overlapping “stride” 1 and the “wrap around” image patch assumption [15, Prop. 3.3].

Algorithm 1 PWLS-\( \ell_1 \) CT Reconstruction

Require: \( y \), \( x^{(1)} \), \( z^{(1)} \), \( \Psi \) learned from (1), \( M \), \( W \), \( \lambda, \gamma, \mu, \nu \geq 0 \), \( i = 1 \)

while a stopping criterion is not satisfied do

for \( i = 1, \ldots, \text{iter}_{\text{ADMM}} \) do

\[\begin{align*}
\text{Obtain } \tilde{x}^{(i+1)} & \text{ by solving (3) with PCG(M)} \\
\tilde{d}_a^{(i+1)} & = (W + \mu I_m)^{-1} \left( W y + \mu \left( A \tilde{x}^{(i+1)} + b_a^{(i)} \right) \right) \\
\tilde{d}_{\psi}^{(i+1)} & = \text{softshrink} \left( \left( \Psi \tilde{x}^{(i+1)} - z^{(i)} + b_{\psi}^{(i)} \right), \frac{\lambda}{\mu \nu} \right),
\end{align*}\]

end for

\[\begin{align*}
x^{(i+1)} & = \tilde{x}^{(\text{iter}_{\text{ADMM}} + 1)} \\
z^{(i+1)} & = \text{hardshrink} \left( \left( \Psi x^{(i+1)} \right), \frac{\gamma}{\lambda} \right),
\end{align*}\]

end while

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

\[
z_j^{(i)} = \text{hardshrink} \left( \left( \Psi x_j^{(i+1)} \right), \frac{\gamma}{\lambda} \right),
\]

where the hard-shrinkage operator is given as follows: hardshrink(\( \alpha, \beta \)) is equal to \( \alpha \) if \( |\alpha| \geq \beta \), and is 0 otherwise. Note that \( \gamma \) should be properly determined based on the (estimated) intensity of \( \Psi x \). If \( \gamma \) 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 \( \gamma \) is relatively too large, it does not properly remove noise (or unwanted artifacts).

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

\[
\min_{z \in \mathbb{R}^{nJ}} \lambda \| \Psi x - z \|_1 + \gamma \| z \|_0.
\]

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

\[
z_j^{(i)} = \text{hardshrink} \left( \left( \Psi x_j^{(i+1)} \right), \frac{\gamma}{\lambda} \right),
\]

where the hard-shrinkage operator is given as follows: hardshrink(\( \alpha, \beta \)) is equal to \( \alpha \) if \( |\alpha| \geq \beta \), and is 0 otherwise. Note that \( \gamma \) should be properly determined based on the (estimated) intensity of \( \Psi x \). If \( \gamma \) 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 \( \gamma \) 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., \( \nu, \mu \)) based on condition numbers [4]. Observe that, for two square Hermitian matrices \( A \) and \( B \),

\[
\kappa(A + B) := \frac{\sigma_{\text{max}}(A + B)}{\sigma_{\text{min}}(A + B)} \leq \frac{\sigma_{\text{max}}(A) + \sigma_{\text{max}}(B)}{\sigma_{\text{min}}(A) + \sigma_{\text{min}}(B)},
\]

by Weyl’s inequality, where the notations \( \kappa(\cdot) \), \( \sigma_{\text{max}}(\cdot) \), and \( \sigma_{\text{min}}(\cdot) \) denote the condition number, the largest eigenvalue,
and the smallest eigenvalue of a matrix, respectively. Applying the bound (7) to (3), we select \( \nu \) by

\[
\nu = \frac{\sigma_{\max}(\mathbf{A}) - \kappa_{\text{des,}\nu} \cdot \sigma_{\min}(\mathbf{A})}{\kappa_{\text{des,}\nu} \cdot \sigma_{\min}(\mathbf{\Phi}) - \sigma_{\max}(\mathbf{\Phi})},
\]

(8)

where \( \kappa_{\text{des,}\nu} \) denotes the desired “upper bounded” condition number of \( \mathbf{A}^T \mathbf{A} + \nu \sum_{j=1}^{J} \mathbf{P}_j^T \mathbf{\Psi}_j \mathbf{P}_j \), and \( \mathbf{A} \) and \( \mathbf{\Phi} \) are approximated diagonal eigenvalue matrices of \( \mathbf{A}^T \mathbf{A} \) and \( \mathbf{\Psi}^T \mathbf{\Psi} \) 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, \( \kappa_{\text{des,}\nu} \) becomes close to the condition number of \( \mathbf{A}^T \mathbf{A} + \nu \sum_{j=1}^{J} \mathbf{P}_j^T \mathbf{\Psi}_j \mathbf{P}_j \), when the learned ST \( \mathbf{\Psi} \) is close to orthogonal. We select \( \mu \) for (4) by

\[
\mu = \frac{\sigma_{\max}(\mathbf{W}) - \kappa_{\text{des,}\nu} \cdot \sigma_{\min}(\mathbf{W})}{\kappa_{\text{des,}\mu} - 1},
\]

(9)

where \( \kappa_{\text{des,}\mu} \) denotes the desired condition number of \( \mathbf{W} + \mu \mathbf{I}_m \) in (4). We empirically found that \( \kappa_{\text{des,}\nu}, \kappa_{\text{des,}\mu} \in [10, 40] \) 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 \( \langle \mathbf{\Psi} \rangle \)-domain, i.e., \( \mathbf{\Psi} \mathbf{x} \) has a few large coefficients, usually corresponding to low-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 \( \langle \mathbf{\Psi} \rangle \)-domain, while being robust to the model mismatch in \( \mathbf{\Psi} \mathbf{x} \) 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 \( \tau \), 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) × 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 \( \varphi(t) := \delta^2 (\sqrt{t + \delta^2} - 1) \) (\( \delta = 10 \) in Hounsfield units, HU), we used relaxed linearized augmented Lagrangian method with 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-\( \ell_2 \) with the algorithm in [8] but without the non-nonnegativity constraint for image update. For both PWLS-ST-\( \ell_1 \) and PWLS-ST-\( \ell_2 \), we used converged PWLS-EP reconstructions for initialization and set a stopping criterion by meeting the maximum number of iterations, e.g., \( \text{Iter} = 200 \). For the image update, we set \( \text{Iter}_{\text{ADMM}} = 2 \) (2 PCG iterations) for PWLS-ST-\( \ell_1 \); and set 12 rLALM iterations (rLALM stands for relaxed linearized augmented Lagrangian method [8]) without ordered subsets for PWLS-ST-\( \ell_2 \).

We finely tuned the parameters \( \lambda, \gamma \) to achieve the lowest root mean squared error (RMSE) in image reconstruction. We tuned \( \nu, \mu \) through the condition number based selection schemes, i.e., \( \kappa_{\text{des,}\nu} \) in (8) and \( \kappa_{\text{des,}\mu} \) in (9). For PWLS-ST-\( \ell_1 \), we chose \( \{\lambda, \gamma/\lambda, \kappa_{\text{des,}\nu}, \kappa_{\text{des,}\mu}\} \) as follows: for 246 views, \( \{1.1 \times 10^3, 30, 30, 30\} \); for 123 views, \( \{9 \times 10^2, 80, 30, 30\} \). 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 \( \lambda \) and \( \gamma \) are in HU.

We evaluated the reconstruction quality by the RMSE (in HU) in a region of interest (ROI).\cretion{The ROI in our experiment was a circular (around center) region containing all the phantom tissues.}

Table I shows that the proposed PWLS-ST-\( \ell_1 \) model outperforms PWLS-EP and PWLS-ST-\( \ell_2 \) in terms of RMSE. In particular, PWLS-ST-\( \ell_1 \) resolves the blurry edge problem in PWLS-ST-\( \ell_2 \); see Fig. 1. The edge-preserving benefit of PWLS-ST-\( \ell_1 \) over PWLS-ST-\( \ell_2 \) can be explained when there exist some outliers for some \( \mathbf{z}^{(\ell+1)} : ||\mathbf{\Psi} \mathbf{x} - \mathbf{z}^{(\ell+1)}||_1 \) 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-\( \ell_2 \) 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-\( \ell_1 \) model can accomplish the both benefits of edge-preserving (achieved by PWLS-EP) and image denoising (achieved by PWLS-ST-\( \ell_2 \)).

<table>
<thead>
<tr>
<th>TABLE I</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE (HU) of different X-ray CT reconstructions with different number of projection views (( \rho_0 = 10^5 ))</td>
</tr>
<tr>
<td>Views</td>
</tr>
<tr>
<td>246</td>
</tr>
<tr>
<td>123</td>
</tr>
</tbody>
</table>
The 14th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine

IV. Conclusion

The proposed PWLS-ST-ℓ₁ model achieves more accurate sparse-view (2D) CT reconstruction compared to PWLS-EP and PWLS-ST-ℓ₂; in particular, it leads to sharper edge reconstruction compared to PWLS-ST-ℓ₂. Future work will explore PWLS-ST-ℓ₁ 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-ℓ₁ 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-ℓ₁ ENCOURAGING UNIFORM SPATIAL RESOLUTION OR NOISE

We first obtain parameter \( \omega \in \mathbb{R}^N \) that controls local spatial resolution or noise in reconstructed image [19], [20]:

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

Using \( \omega \), we compute the weighting parameters \( \lambda' \in \mathbb{R}_+^J \) by \( \{ \lambda'_j = \| P_j \omega \|_1 : j = 1, \ldots, J \} \). 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-ℓ₁ model promoting uniform spatial resolution or noise as follows:

\[
\min_{\mathbf{x}, \mathbf{z} \in \mathbb{R}^{N \times J}} \frac{1}{2} \| \mathbf{y} - \mathbf{A} \mathbf{x} \|_W^2 + \sum_{j=1}^{J} \lambda'_j \| \mathbf{P}_j \mathbf{x} - \mathbf{z}_j \|_1 + \gamma \| \mathbf{z}_j \|_0.
\]

(10)

Fig. 2. Comparison of 3D reconstructed images from PWLS-ST-ℓ₁ and PWLS-ST-ℓ₁ promoting uniform spatial resolution or noise (10) improves the accuracy of reconstructed images compared to the PWLS-ST-ℓ₁ (2). Alternatively, one can promote uniform spatial resolution or noise as follows:

\[
\min_{\mathbf{x}, \mathbf{z} \in \mathbb{R}^{N \times J}} \frac{1}{2} \| \mathbf{y} - \mathbf{A} \mathbf{x} \|_W^2 + \sum_{j=1}^{J} \lambda'_j \| \mathbf{P}_j \mathbf{x} - \mathbf{z}_j \|_1 + \lambda'_j \gamma \| \mathbf{z}_j \|_0.
\]

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

![Fig. 3. RMSE convergence behavior for PWLS-ST-$\ell_1$ and PWLS-ST-$\ell_2$ (123 and 246 projection views, and $\rho_0 = 10^3$).](image)

**ACKNOWLEDGMENT**

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

**REFERENCES**


