Accelerating Ordered-Subsets X-Ray CT Image Reconstruction Using the Linearized Augmented Lagrangian Framework

Hung Nien and Jeffrey A. Fessler
Department of Electrical Engineering and Computer Science
University of Michigan, Ann Arbor

INTRODUCTION

The model-based image reconstruction (MBIR) technique has been explored extensively in CT community due to its potential for low-dose CT. However, the much longer computation time of MBIR still restrains its applicability in practice. To accelerate MBIR, lots of methods have been investigated, and the augmented Lagrangian (AL) method has drawn more attention recently. By introducing auxiliary variables, the AL method decomposes a convex optimization problem into a series of easier penalized least-squares problems that either have closed-form solutions or can be solved efficiently using proximal mappings. However, the AL method can be slow sometimes, e.g., in X-ray CT image reconstruction problems. In this poster, we present a new ordered-subsets (OS) algorithm based on a linearized AL framework that is free from the slow convergence problem in standard AL methods and greatly accelerates MBIR with excellent gradient error tolerance and negligible overhead.

BACKGROUND

Consider solving a regularized least-squares problem:

$$\hat{\mathbf{x}} \in \arg\min_{\mathbf{x}} \left\{ \Psi(\mathbf{x}) \triangleq \tfrac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + h(\mathbf{x}) \right\}, \tag{1}$$

where $\mathbf{A}$ is the system matrix, $\mathbf{y}$ is the noisy measurement, and h is a convex (and possibly non-smooth) regularization term. Instead of solving (1) directly, we consider solving an equivalent *constrained* minimization problem:

$$(\hat{\mathbf{x}}, \hat{\mathbf{u}}) \in \arg\min_{\mathbf{x},\mathbf{u}} \left\{ \tfrac{1}{2} \|\mathbf{u}\|_2^2 + h(\mathbf{x}) \right\} \text{ s.t. } \mathbf{u} = \mathbf{A}\mathbf{x} - \mathbf{y} \tag{2}$$

using the (alternating direction) AL method:

$$\begin{cases} \mathbf{x}^{(k+1)} \in \arg\min_{\mathbf{x}} \left\{ h(\mathbf{x}) + \frac{\rho}{2} \left\| \mathbf{A}\mathbf{x} - \mathbf{y} - \mathbf{u}^{(k)} - \mathbf{d}^{(k)} \right\|_2^2 \right\} \\ \mathbf{u}^{(k+1)} \in \arg\min_{\mathbf{u}} \left\{ \frac{1}{2} \|\mathbf{u}\|_2^2 + \frac{\rho}{2} \left\| \mathbf{A}\mathbf{x}^{(k+1)} - \mathbf{y} - \mathbf{u} - \mathbf{d}^{(k)} \right\|_2^2 \right\} \\ \mathbf{d}^{(k+1)} = \mathbf{d}^{(k)} - \left(\mathbf{A}\mathbf{x}^{(k+1)} - \mathbf{y} \right) + \mathbf{u}^{(k+1)}, \end{cases} \tag{3}$$

where $\mathbf{d}$ is the scaled Lagrange multiplier of $\mathbf{u}$ (a split variable of the residual), and $\rho > 0$ is the corresponding AL penalty parameter that affects only the convergence rate. As can be seen in (3), the $\mathbf{x}$-update is a penalized least-squares problem, and the Hessian of its quadratic term is $\mathbf{A}'\mathbf{A}$. This penalized least-squares problem can be solved efficiently using the preconditioned conjugate gradient (PCG) method if we can find an appropriate preconditioner for the Hessian. When the Hessian is highly shift-variant (just like in most CT reconstruction problems), it might take *many* forward/back-projection operations for a *single* image update, and this greatly reduces the number of image updates one can perform in a given reconstruction time, leading to a very slow convergence rate in practice [1]!

METHOD

To solve the problem of the challenging inner minimization problem, we proposed to *linearize* the quadratic AL penalty term of the $\mathbf{x}$-update in (3) by its separable quadratic surrogate (SQS) function around $\mathbf{x} = \mathbf{x}^{(k)}$. This majorization removes the entanglement of $\mathbf{x}$ introduced by the system matrix $\mathbf{A}$, leading to a simple $\mathbf{x}$-update. Furthermore, since we focus on solving the regularized least-squares problem (1) with variable splitting scheme (2), in this case, we can rewrite our linearized AL method so that all the updates depend only on the gradients of quadratic data-fitting term. That is, letting $\ell(\mathbf{x})$ denote the quadratic data-fitting term in (1), we have the *gradient-based* linearized AL method (a detailed derivation is in the paper):

$$\begin{cases} \mathbf{s}^{(k+1)} = \rho \nabla \ell\left(\mathbf{x}^{(k)}\right) + (1-\rho)\mathbf{g}^{(k)} \\ \mathbf{x}^{(k+1)} \in \operatorname{prox}_{(\rho^{-1}t)h}\left(\mathbf{x}^{(k)} - (\rho^{-1}t)\,\mathbf{s}^{(k+1)}\right) \\ \mathbf{g}^{(k+1)} = \frac{\rho}{\rho+1} \nabla \ell\left(\mathbf{x}^{(k+1)}\right) + \frac{1}{\rho+1}\mathbf{g}^{(k)}. \end{cases} \tag{4}$$

The $\mathbf{x}$-update in (4) can be interpreted as a generalized proximal gradient descent of the cost function Ψ with step size $\rho^{-1}t$ and search direction $\mathbf{s}$, where t is the reciprocal of the maximum eigenvalue of $\mathbf{A}'\mathbf{A}$. A smaller ρ leads to a larger step size.

METHOD (CONT'D)

Thanks to the ρ-dependent correction of $\mathbf{s}$ and $\mathbf{g}$, we can arbitrarily increase the step size of the proximal gradient method by decreasing ρ. Remember that the updates in (4) depend only on the gradients of ℓ. Suppose ℓ is suitable for OS acceleration; that is, ℓ can be decomposed into M smaller quadratic functions $\ell_1, \ldots, \ell_M$ that satisfy the "subset balance condition":

$$\nabla \ell(\mathbf{x}) \approx M \nabla \ell_1(\mathbf{x}) \approx \cdots \approx M \nabla \ell_M(\mathbf{x}). \tag{5}$$

We can further accelerate the proposed algorithm using OS. To enable OS acceleration, we simply replace the gradients in (4) with the approximations in (5) and incrementally perform (4) for M times as one outer iteration, leading to our proposed *OS-accelerable* linearized AL method (OS-LALM):

$$\begin{cases} \mathbf{s}^{(k,m+1)} = \rho M \nabla \ell_m\left(\mathbf{x}^{(k,m)}\right) + (1-\rho)\mathbf{g}^{(k,m)} \\ \mathbf{x}^{(k,m+1)} \in \operatorname{prox}_{(\rho^{-1}t)h}\left(\mathbf{x}^{(k,m)} - (\rho^{-1}t)\,\mathbf{s}^{(k,m+1)}\right) \\ \mathbf{g}^{(k,m+1)} = \frac{\rho}{\rho+1} M \nabla \ell_{m+1}\left(\mathbf{x}^{(k,m+1)}\right) + \frac{1}{\rho+1}\mathbf{g}^{(k,m)}, \end{cases} \tag{6}$$

where $\mathbf{c}^{(k,M+1)} = \mathbf{c}^{(k+1)} = \mathbf{c}^{(k+1,1)}$ for $\mathbf{c} = \{\mathbf{s}, \mathbf{x}, \mathbf{g}\}$ and $\ell_{M+1} = \ell_1$.

One drawback of the AL method with a fixed AL penalty parameter ρ is the difficulty of finding the value that provides the fastest convergence. We proposed a *deterministic downward continuation* approach that decreases ρ in a certain rate so that the updates generated by (6) and (7) never over-shoot or oscillate! Here, we consider a decreasing sequence:

$$\rho_l = \begin{cases} 1 & , \text{ if } l = 0 \\ \frac{\pi}{l+1}\sqrt{1 - \left(\frac{\pi}{2l+2}\right)^2} & , \text{ otherwise}, \end{cases} \tag{7}$$

where l is a counter that starts from zero and increases by one.

RESULTS

We evaluated our proposed algorithm using the statistically weighted X-ray CT image reconstruction problem:

$$\hat{\mathbf{x}} \in \arg\min_{\mathbf{x}\in\Omega} \left\{ \tfrac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_{\mathbf{W}}^2 + \mathsf{R}(\mathbf{x}) \right\}, \tag{8}$$

where $\mathbf{A}$ is the system matrix of a CT scan, $\mathbf{y}$ is the noisy sinogram, $\mathbf{W}$ is the statistical weighting matrix, R denotes an edge-preserving regularizer, and Ω denotes the convex set for a box constraint (usually the non-negativity constraint) on $\mathbf{x}$. To solve (8) using the proposed algorithm, we simply replace $\mathbf{A}$ and $\mathbf{y}$ in (1) with the *weighted forward projection matrix* ($\mathbf{W}^{1/2}\mathbf{A}$) and the *weighted noisy sinogram* ($\mathbf{W}^{1/2}\mathbf{y}$), respectively, and combine the edge-preserving regularizer R and the box constraint Ω as a general convex regularization term h in (1).

Now, the inner minimization problem in (4) and (6) becomes a *constrained denoising problem*. In our implementation, we simply solve the inner constrained denoising problem using n iterations of the fast iterative shrinkage/thresholding algorithm (FISTA) starting from the previous update as a warm-start. Throughout the experiment, we use $\mathbf{G} = \text{diag}\{\mathbf{A}'\mathbf{W}\mathbf{A}\mathbf{1}\}$ to majorize the quadratic AL penalty term. When $n = 1$, the computational complexity of the proposed algorithm (with OS acceleration) is one forward/back-projection pair and M gradient evaluation of the regularization term, same as existing OS-based algorithms. Therefore, comparing the convergence rate as a function of iteration with other OS-based algorithms is fair.

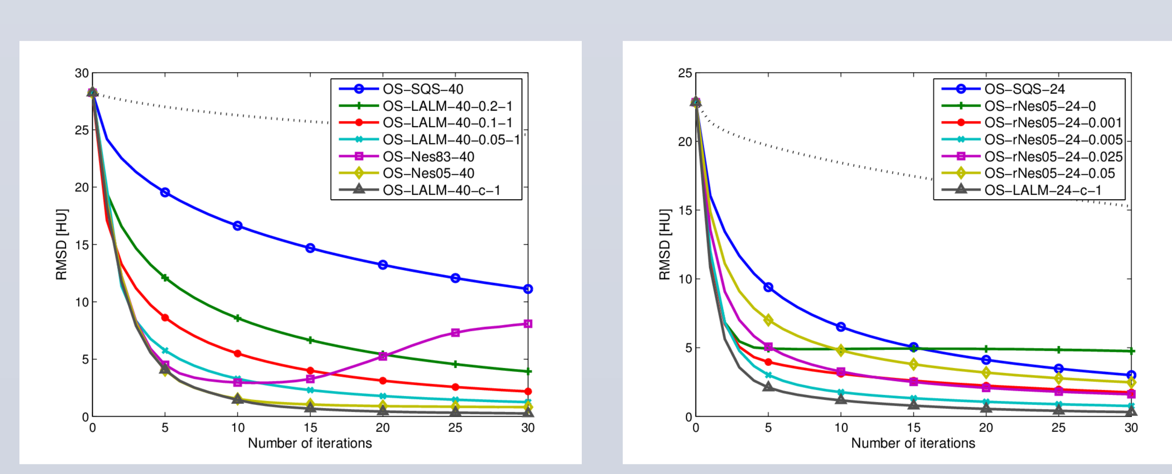

Figure 1. RMS differences between the reconstructed image $\mathbf{x}^{(k)}$ and the reference reconstruction $\mathbf{x}^\star$ as a function of iteration using OS-based algorithms for (a) the shoulder helical scan and (b) the GEPP axial scan, respectively. The dotted lines show the RMS difference using the one-subset OS algorithm [2] as the baseline convergence rate.

RESULTS (CONT'D)

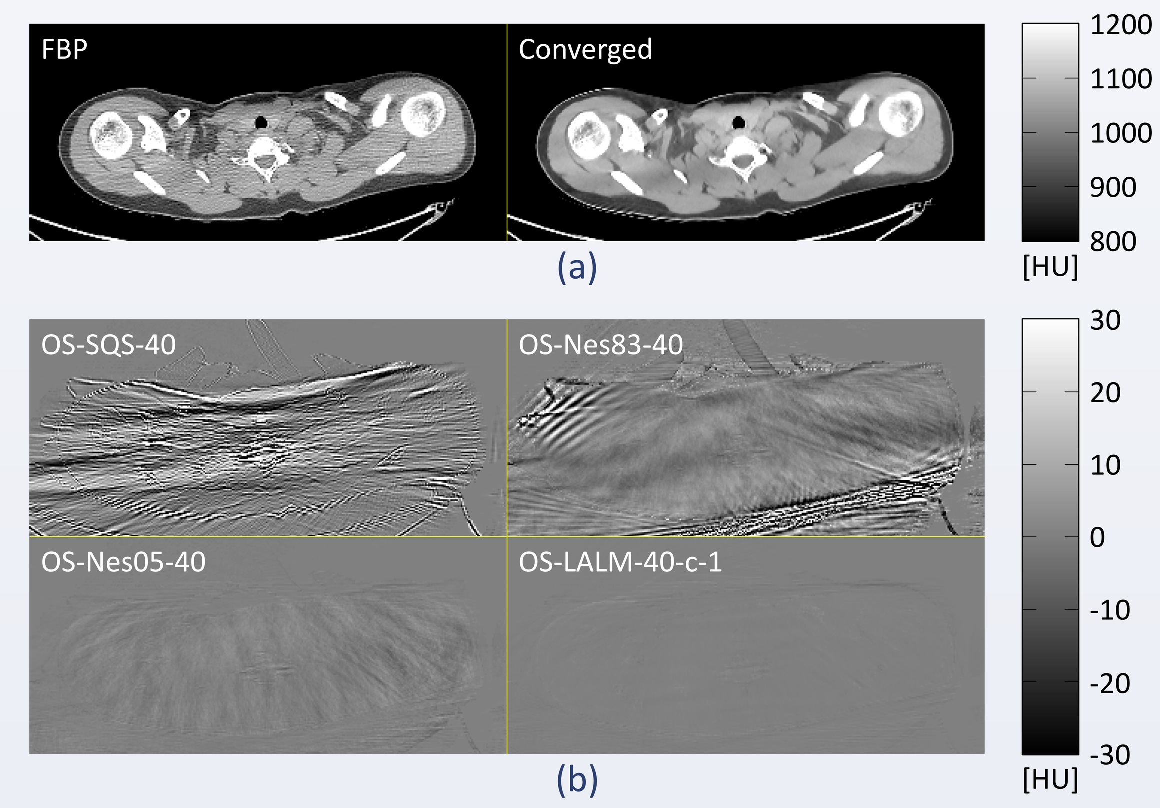

Figure 2. Shoulder: cropped images from the central transaxial plane of (a) the initial FBP image $\mathbf{x}^{(0)}$ and the reference reconstruction $\mathbf{x}^\star$ [displayed from 800 to 1200 HU], and (b) the difference images at the 30th iteration $\mathbf{x}^{(30)} - \mathbf{x}^\star$ using OS-based algorithms [displayed from -30 to 30 HU].

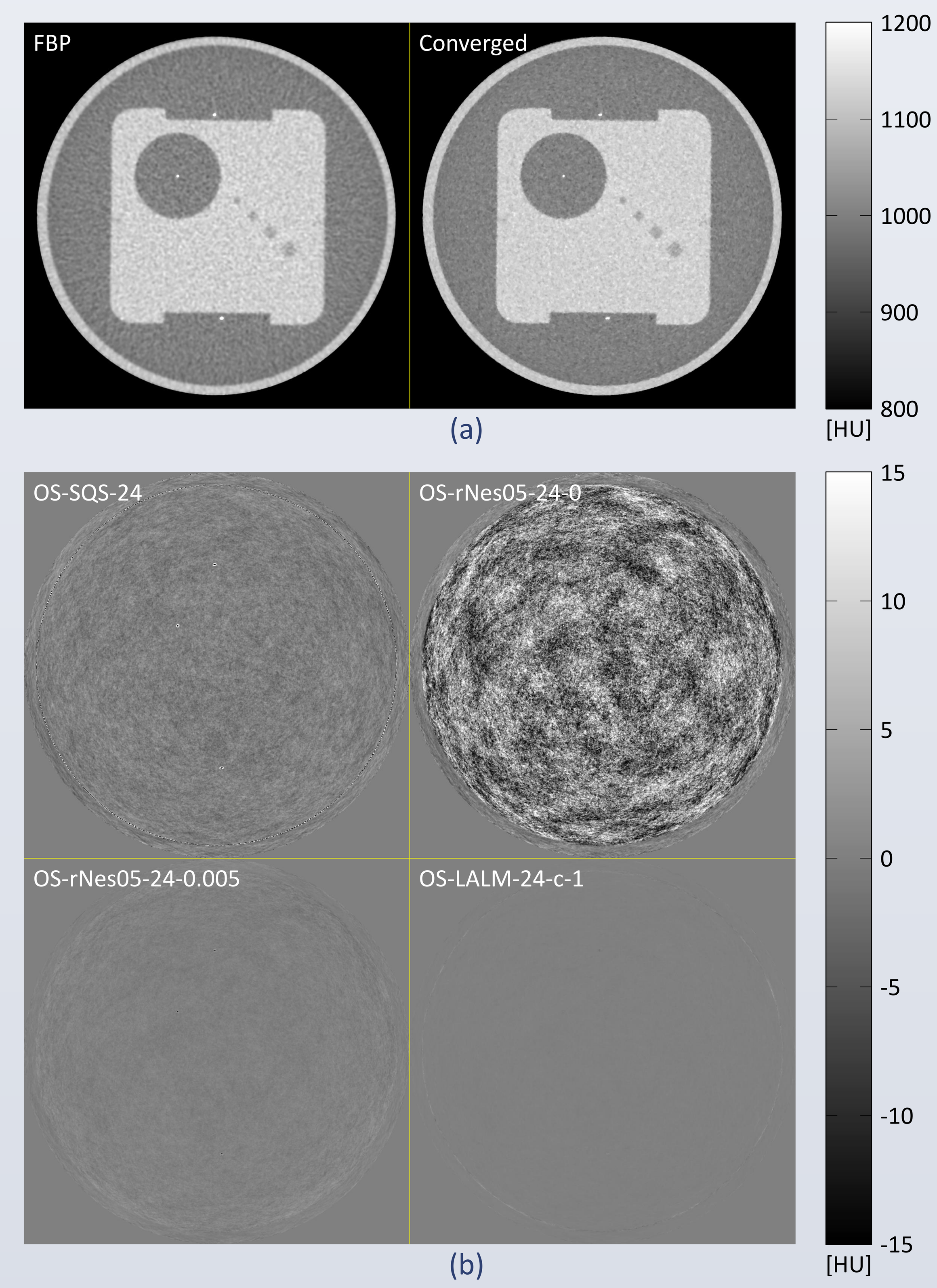

Figure 3. GEPP: cropped images from the central transaxial plane of (a) the initial FBP image $\mathbf{x}^{(0)}$ and the reference reconstruction $\mathbf{x}^\star$ [displayed from 800 to 1200 HU], and (b) the difference images at the 30th iteration $\mathbf{x}^{(30)} - \mathbf{x}^\star$ using OS-based algorithms [displayed from -15 to 15 HU].

OS-SQS-*M*: the OS algorithm proposed by Erdoğan *et al.* [2]
OS-Nes83-*M*: the first OS+momentum algorithm proposed by Kim *et al.* [3]
OS-Nes05-*M*: the second OS+momentum algorithm proposed by Kim *et al.* [4]
OS-rNes05-*M*-*γ*: the relaxed OS+momentum algorithm proposed by Kim *et al.* [5]
OS-LALM-*M*-*ρ*-*n*: the proposed algorithm with a fixed AL penalty parameter
OS-LALM-*M*-c-*n*: the proposed algorithm with downward continuation (7)

CONCLUSION

We proposed an OS-accelerable splitting-based algorithm, OS-LALM, for solving penalized weighted least-squares X-ray CT image reconstruction problems using a linearized AL framework. To further accelerate the proposed algorithm, we also proposed a deterministic downward continuation approach that avoids tedious parameter tuning for fast convergence. Experimental results showed that our proposed algorithm significantly accelerates the convergence of X-ray CT image reconstruction with negligible overhead and greatly reduces the OS artifacts in the reconstructed image when using many subsets for OS acceleration.

References: [1] S. Ramani and J. A. Fessler, "A splitting-based iterative algorithm for accelerated statistical X-ray CT reconstruction," *IEEE Trans. Med. Imag.*, vol. 31, pp. 677–88, Mar. 2012. [2] H. Erdоğan and J. A. Fessler, "Ordered subsets algorithms for transmission tomography," *Phys. Med. Biol.*, vol. 44, pp. 2835–51, Nov. 1999. [3] D. Kim, S. Ramani, and J. A. Fessler, "Ordered subsets with momentum for accelerated X-ray CT image reconstruction," in *Proc. IEEE Conf. Acoust. Speech Sig. Proc.*, pp. 920–3, 2013. [4] D. Kim, S. Ramani, and J. A. Fessler, "Accelerating X-ray CT ordered subsets image reconstruction with Nesterov's first-order methods," in *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med.*, pp. 22–5, 2013. [5] D. Kim and J. A. Fessler, "Ordered subsets acceleration using relaxed momentum for X-ray CT image reconstruction," in *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, 2013. To appear.