Relaxed Linearized Algorithms for Faster X-Ray CT Image Reconstruction

Hung Nien and Jeffrey A. Fessler

Abstract—**Statistical image reconstruction (SIR) methods enable acquiring computed tomography (CT) scans with lower X-ray dose while maintaining image quality. However, the iterative nature of algorithms in SIR methods increases the reconstruction time, hindering their use in practice. Relaxation (especially over-relaxation) is a common technique to accelerate convergence of iterative algorithms. For instance, using a relaxation parameter that is close to two in the alternating direction method of multipliers (ADMM) has been shown to speed up convergence significantly. When we consider a linearized ADMM (or augmented Lagrangian [AL] method), applying relaxation is not trivial, and the simple relaxation approach that extends the existing relaxed ADMM to its linearized variant directly does not work well. This paper proposes one way to properly use relaxation in linearized AL methods and applies the proposed relaxed linearized AL method to X-ray CT image reconstruction problems. Experimental results show that the proposed relaxed algorithm (with moderate ordered-subsets [OS] acceleration) is about twice as fast as the unrelaxed counterpart in a 3D XCAT phantom simulation.**

I. INTRODUCTION

Relaxation (usually over-relaxation) is a common technique to speed up convergence of iterative algorithms. For example, over-relaxation is very effective for accelerating the alternating direction method of multipliers (ADMM) [1, 2], provided that all inner minimization problems of ADMM can be solved efficiently. However, in problems like X-ray computed tomography (CT) image reconstruction, exact image updates in typical ADMM methods would require inverting an enormous non-circulant Hessian matrix involving the system matrix $\mathbf{A}$, precluding the use of exact updates in practice.

In such cases, a linearized ADMM or augmented Lagrangian (AL) method [3, 4] becomes more attractive because these methods replace the image update by a proxlinear step that one can solve exactly or iteratively without using $\mathbf{A}$ and $\mathbf{A}'$. We can interpret the linearized AL method (LALM) as a proximal-point variant of AL methods with an additional iteration-dependent proximity term adding to the image subproblem. However, adding the extra proximity term changes the original alternating direction AL framework, so adapting the original relaxation approach [1, Theorem 8] to the linearized AL methods is not trivial.

This paper proposes a non-trivial relaxed variant of linearized AL methods and applies the proposed relaxed LALM

to X-ray CT image reconstruction problems. Experimental results show that our proposed relaxation works much better than the simple relaxation [5] and significantly accelerates X-ray CT image reconstruction, even with ordered-subsets (OS) acceleration.

II. METHOD

Consider an equality-constrained minimization problem:

$$(\hat{\mathbf{x}}, \hat{\mathbf{u}}) \in \arg\min_{\mathbf{x},\mathbf{u}} \left\{ g(\mathbf{u} - \mathbf{y}) + h(\mathbf{x}) \right\} \text{ s.t. } \mathbf{u} = \mathbf{A}\mathbf{x}, \quad (1)$$

where g and h are closed and proper convex functions. In particular, g is called the loss function that measures the fitness between the linear model $\mathbf{A}\mathbf{x}$ and noisy measurement $\mathbf{y}$, and h is a regularization term introducing the prior knowledge of $\mathbf{x}$ to the reconstruction. For example, one can write statistical X-ray CT image reconstruction as a special case of (1), where g is a weighted quadratic function, and h is an edge-preserving regularizer (often with a non-negativity constraint on the image). The relaxed AL method [1, Theorem 8] solves this problem in an alternating direction manner:

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

where

$$\mathbf{r}_\mathbf{u}^{(k+1)} \triangleq \alpha \mathbf{A}\mathbf{x}^{(k+1)} + (1 - \alpha)\,\mathbf{u}^{(k)} \quad (3)$$

is the relaxation variable of $\mathbf{u}$, $\rho > 0$ is the AL penalty parameter, and $0 < \alpha < 2$ is the relaxation parameter. When $\alpha > 1$, it is called over-relaxation, and it is called under-relaxation when $\alpha < 1$. When $\alpha = 1$, (2) reverts to the standard (alternating direction) AL method. Experimental results suggest that over-relaxation with $\alpha \in [1.5, 1.8]$ can improve convergence [2]. However, the $\mathbf{x}$-update in (2) requires solving a penalized least-squares problem involving the system matrix $\mathbf{A}$ and thus is expensive in X-ray CT, motivating alternative methods like LALM.

A. Simple relaxation

In LALM, one uses $\alpha = 1$ in (2) and just replaces the quadratic AL penalty term in the $\mathbf{x}$-update of (2):

$$\theta_k(\mathbf{x}) \triangleq \frac{\rho}{2} \left\| \mathbf{A}\mathbf{x} - \mathbf{u}^{(k)} - \mathbf{d}^{(k)} \right\|_2^2 \quad (4)$$

by its separable quadratic surrogate (SQS) function [4]:

$$\breve{\theta}_k\big(\mathbf{x}; \mathbf{x}^{(k)}\big) \propto \big\langle \nabla\theta_k\big(\mathbf{x}^{(k)}\big), \mathbf{x} - \mathbf{x}^{(k)} \big\rangle + \frac{\rho}{2} \big\| \mathbf{x} - \mathbf{x}^{(k)} \big\|_{\mathbf{D}_\mathbf{L}}^2, \quad (5)$$
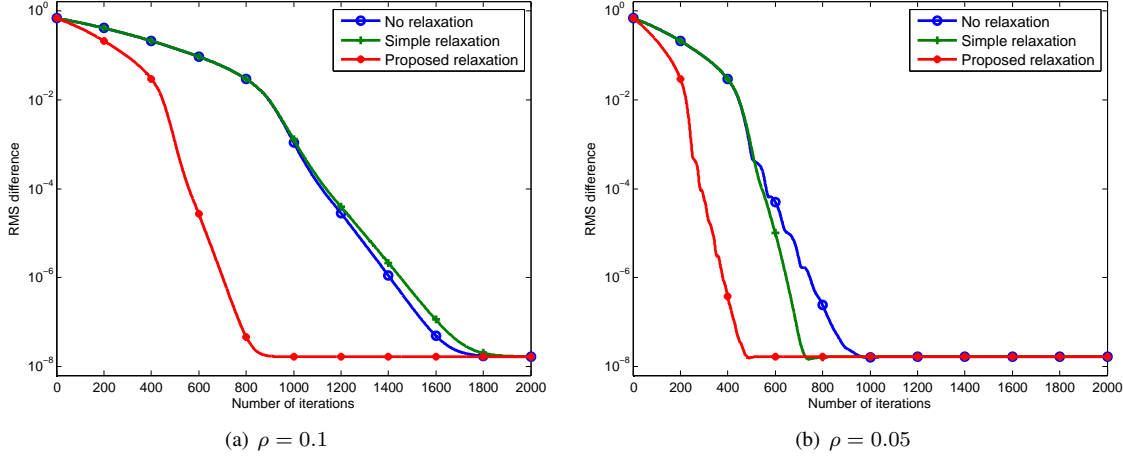
(a) $\rho = 0.1$
(b) $\rho = 0.05$

Fig. 1: LASSO regression: RMS differences between the iterate $\mathbf{x}^{(k)}$ and the solution $\hat{\mathbf{x}}$ when solving a LASSO regression problem (12) using proposed relaxed LALM ($\alpha = 2$) with AL penalty parameters (a) $\rho = 0.1$ and (b) $\rho = 0.05$.

where $\mathbf{D_L}$ is a diagonal majorizing matrix of $\mathbf{A}'\mathbf{A}$. This replacement makes the $\mathbf{x}$-update far simpler and more efficient. Unlike standard AL methods that have column-rank constraint on $\mathbf{A}$, LALM converges unconditionally for any fixed AL penalty parameter ρ and any $\mathbf{A}$ [4]. The simple (and straightforward) relaxed variant of LALM just substitutes the SQS AL penalty term (5) for the quadratic one (4) in (2) with $\alpha \neq 1$ [5]. Note that LALM can also be thought of as an inexact AL method (with inexact image updates). The updated image $\mathbf{x}$ with simple relaxation might be still far away from the solution to the $\mathbf{x}$-subproblem in (2). Hence, the extrapolated relaxation variable $\mathbf{r_u}^{(k+1)}$ could be less useful for speed-up.

B. Proposed relaxation

To properly apply relaxation to LALM, instead of treating LALM as a proximal-point variant of AL methods, we write it as the AL method with an extra redundant equality constraint $\mathbf{v} = \mathbf{G}^{1/2}\mathbf{x}$, where $\mathbf{G} \triangleq \mathbf{D_L} - \mathbf{A}'\mathbf{A}$ is positive-definite, using the same AL penalty parameter ρ [4]. This expression allows us to use the same relaxation approach of standard AL methods [1, Theorem 8] for accelerating LALM and still guarantees convergence (and speed-up) as follows:

$$
\begin{cases}
\mathbf{x}^{(k+1)} \in \arg\min_{\mathbf{x}} \left\{ h(\mathbf{x}) + \frac{\rho}{2}\left\| \mathbf{A}\mathbf{x} - \mathbf{u}^{(k)} - \mathbf{d}^{(k)} \right\|_2^2 \\ \qquad\qquad + \frac{\rho}{2}\left\| \mathbf{G}^{1/2}\mathbf{x} - \mathbf{v}^{(k)} - \mathbf{e}^{(k)} \right\|_2^2 \right\} \\
\mathbf{u}^{(k+1)} \in \arg\min_{\mathbf{u}} \left\{ g(\mathbf{u} - \mathbf{y}) + \frac{\rho}{2}\left\| \mathbf{r_u}^{(k+1)} - \mathbf{u} - \mathbf{d}^{(k)} \right\|_2^2 \right\} \\
\mathbf{d}^{(k+1)} = \mathbf{d}^{(k)} - \mathbf{r_u}^{(k+1)} + \mathbf{u}^{(k+1)} \\
\mathbf{v}^{(k+1)} = \mathbf{r_v}^{(k+1)} - \mathbf{e}^{(k)} \\
\mathbf{e}^{(k+1)} = \mathbf{e}^{(k)} - \mathbf{r_v}^{(k+1)} + \mathbf{v}^{(k+1)},
\end{cases}
$$
(6)

where

$$
\mathbf{r_v}^{(k+1)} \triangleq \alpha \mathbf{G}^{1/2}\mathbf{x}^{(k+1)} + (1 - \alpha)\mathbf{v}^{(k)} \tag{7}
$$

is the relaxation variable of $\mathbf{v}$. As seen in (6), the Hessian $\rho \mathbf{A}'\mathbf{A}$ of (4) is cancelled by the additional AL penalty term.

One can verify that $\mathbf{e}^{(k+1)} = \mathbf{0}$ if we initialize $\mathbf{e}$ as $\mathbf{e}^{(0)} = \mathbf{0}$. Let $\mathbf{h}^{(k)} \triangleq \mathbf{G}^{1/2}\mathbf{v}^{(k)} + \mathbf{A}'\mathbf{y}$. We rewrite (6) so that no explicit multiplication by $\mathbf{G}^{1/2}$ has to be computed, leading to the following proposed relaxed LALM:

$$
\begin{cases}
\mathbf{x}^{(k+1)} \in \arg\min_{\mathbf{x}} \left\{ h(\mathbf{x}) + \frac{1}{2}\left\| \mathbf{x} - (\rho \mathbf{D_L})^{-1}\boldsymbol{\gamma}^{(k+1)} \right\|_{\rho \mathbf{D_L}}^2 \right\} \\
\mathbf{u}^{(k+1)} \in \arg\min_{\mathbf{u}} \left\{ g(\mathbf{u} - \mathbf{y}) + \frac{\rho}{2}\left\| \mathbf{r_u}^{(k+1)} - \mathbf{u} - \mathbf{d}^{(k)} \right\|_2^2 \right\} \\
\mathbf{d}^{(k+1)} = \mathbf{d}^{(k)} - \mathbf{r_u}^{(k+1)} + \mathbf{u}^{(k+1)} \\
\mathbf{h}^{(k+1)} = \alpha\boldsymbol{\eta}^{(k+1)} + (1 - \alpha)\mathbf{h}^{(k)},
\end{cases}
$$
(8)

where

$$
\boldsymbol{\gamma}^{(k+1)} \triangleq \rho \mathbf{A}'\big(\mathbf{u}^{(k)} - \mathbf{y} + \mathbf{d}^{(k)}\big) + \rho \mathbf{h}^{(k)}, \tag{9}
$$

and

$$
\boldsymbol{\eta}^{(k+1)} \triangleq \mathbf{D_L}\mathbf{x}^{(k+1)} - \mathbf{A}'\big(\mathbf{A}\mathbf{x}^{(k+1)} - \mathbf{y}\big). \tag{10}
$$

When g is a quadratic loss, i.e., $g(\mathbf{z}) \triangleq (1/2)\|\mathbf{z}\|_2^2$, the relaxed LALM can be further simplified by manipulations similar to those in [4] (that are omitted here due to space constraints) as:

$$
\begin{cases}
\boldsymbol{\gamma}^{(k+1)} = (\rho - 1)\mathbf{g}^{(k)} + \rho \mathbf{h}^{(k)} \\
\mathbf{x}^{(k+1)} \in \arg\min_{\mathbf{x}} \left\{ h(\mathbf{x}) + \frac{1}{2}\left\| \mathbf{x} - (\rho \mathbf{D_L})^{-1}\boldsymbol{\gamma}^{(k+1)} \right\|_{\rho \mathbf{D_L}}^2 \right\} \\
\boldsymbol{\zeta}^{(k+1)} = \nabla \mathsf{L}\big(\mathbf{x}^{(k+1)}\big) \triangleq \mathbf{A}'\big(\mathbf{A}\mathbf{x}^{(k+1)} - \mathbf{y}\big) \\
\mathbf{g}^{(k+1)} = \frac{\rho}{\rho+1}\big(\alpha\boldsymbol{\zeta}^{(k+1)} + (1 - \alpha)\mathbf{g}^{(k)}\big) + \frac{1}{\rho+1}\mathbf{g}^{(k)} \\
\mathbf{h}^{(k+1)} = \alpha\big(\mathbf{D_L}\mathbf{x}^{(k+1)} - \boldsymbol{\zeta}^{(k+1)}\big) + (1 - \alpha)\mathbf{h}^{(k)},
\end{cases}
$$
(11)

where $\mathsf{L}(\mathbf{x}) \triangleq g(\mathbf{A}\mathbf{x} - \mathbf{y})$ is the quadratic data-fidelity term. Throughout the algorithm, one only has to compute multiplications by $\mathbf{A}$ and $\mathbf{A}'$ once per iteration and does not have to invert $\mathbf{A}'\mathbf{A}$ as in standard relaxed AL methods like (2). This property is especially useful in cases where $\mathbf{A}'\mathbf{A}$ is large and non-structured like in CT reconstruction.

To illustrate the speed-up of the proposed relaxation (11), Figure 1 shows the convergence rate curves (RMS differences
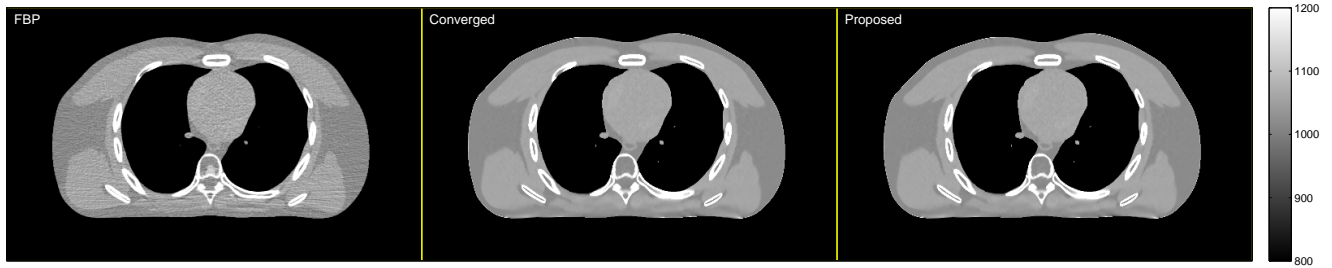
Fig. 2: XCAT phantom: cropped images (displayed from 800 to 1200 HU) from the central transaxial plane of the initial FBP image $\mathbf{x}^{(0)}$ (left), the reference reconstruction $\mathbf{x}^\star$ (center), and the reconstructed image $\mathbf{x}^{(10)}$ using the proposed algorithm (relaxed OS-LALM with 24 subsets) after 10 iterations (right).

between the iterate $\mathbf{x}^{(k)}$ and the solution $\hat{\mathbf{x}}$) of the simple and proposed relaxed LALM's when solving a LASSO regression problem (in an equality-constrained form):

$$(\hat{\mathbf{x}}, \hat{\mathbf{u}}) \in \arg\min_{\mathbf{x},\mathbf{u}} \left\{ \tfrac{1}{2} \|\mathbf{u} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{x}\|_1 \right\} \text{ s.t. } \mathbf{u} = \mathbf{A}\mathbf{x}, \tag{12}$$

where $\mathbf{A} \in \mathbb{R}^{250 \times 1000}$ is an i.i.d. Gaussian random matrix with variance one, and $\mathbf{y}$ is a noisy projection (with noise standard deviation 0.1) from a 50-sparse vector $\mathbf{x}_{\text{true}}$. The regularization parameter λ is tuned ($\lambda = 1$) for the best reconstruction. We set the relaxation parameter α to be two, which is the theoretical upper limit of α. Note that when $\mathbf{A}$ is large (for example, in the distributed LASSO regression problem), inverting $\mathbf{A}'\mathbf{A}$ requires a time-consuming matrix factorization at the beginning, whereas computing the maximum eigenvalue of $\mathbf{A}'\mathbf{A}$ is cheaper [6]. Hence, it is reasonable to solve LASSO regression using linearized algorithms.

As can be seen in Figure 1, the simple relaxation does not provide much acceleration. In contrast, the proposed relaxation accelerates convergence about twice (α-times). This behavior is consistent with the convergence rate analysis of a closely related relaxed primal-dual algorithm shown by Chambolle and Pock [9, Theorem 2].

C. Relaxed OS-LALM for faster CT reconstruction

Consider the X-ray CT image reconstruction problem:

$$\hat{\mathbf{x}} \in \arg\min_{\mathbf{x} \in \Omega} \left\{ \tfrac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_{\mathbf{W}}^2 + \mathsf{R}(\mathbf{x}) \right\}, \tag{13}$$

where $\mathbf{A}$ is the forward projection 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 a box-constraint on the image $\mathbf{x}$. To solve it using the proposed relaxed LALM (11), we apply the following substitution:

$$\begin{cases} \mathbf{A} \leftarrow \mathbf{W}^{1/2}\mathbf{A} \\ \mathbf{y} \leftarrow \mathbf{W}^{1/2}\mathbf{y} \\ h \leftarrow \mathsf{R} + \iota_\Omega, \end{cases} \tag{14}$$

where $\iota_\mathcal{C}$ denotes the characteristic function of a convex set $\mathcal{C}$. The image ($\mathbf{x}$-)update now is a constrained diagonally weighted denoising problem. We solve it by a single projected gradient descent from $\mathbf{x}^{(k)}$ (equivalently, further majorizing

the smooth regularizer R) and use Huber's curvature of R for the fastest convergence. We leave the convergence rate analysis with further majorization as future work [10]. Since the updates in (11) depend only on the gradients of L, we can further accelerate it using OS. When $\alpha = 1$, the proposed algorithm is the same as the OS-LALM algorithm [4], and we expect to see twice acceleration if we set α to be (close to) two ($\alpha = 1.999$ in the experiment).

We also use a continuation technique to speed up convergence; that is, we decrease ρ gradually as iteration progresses [4]. The difference between the proposed algorithm and the previous OS-LALM in [4] is that we decrease ρ twice as fast as the decreasing sequence ρ_k proposed in [4], i.e. ,

$$\bar{\rho}_k = \begin{cases} 1, & \text{if } k = 0 \\ \frac{\pi}{2(k+1)} \sqrt{1 - \left(\frac{\pi}{4(k+1)}\right)^2}, & \text{otherwise}. \end{cases} \tag{15}$$

The faster decreasing sequence $\bar{\rho}_k$ comes from the fact that LALM is accelerated by two-times with $\alpha \approx 2$ for any fixed ρ empirically. When ρ changes every iteration (i.e., continuation), the AL penalty parameter $\bar{\rho}_k$ of the relaxed (OS-)LALM at the kth iteration should be the same as the AL penalty parameter ρ_{2k} of the original (OS-)LALM at the $(2k)$th iteration, leading to the same instantaneous ρ at the corresponding iterations.

III. RESULTS: 3D X-RAY CT IMAGE RECONSTRUCTION

This section reports numerical results for 3D X-ray CT image reconstruction. We simulated an axial CT scan using a $1024 \times 1024 \times 154$ XCAT phantom [11] for 500 mm transaxial field-of-view (FOV), where $\Delta_x = \Delta_y = 0.4883$ mm and $\Delta_z = 0.6250$ mm. An $888 \times 64 \times 984$ noisy (with Poisson noise) sinogram is numerically generated with GE LightSpeed fan-beam geometry corresponding to a monoenergetic source at 70 keV with 10^5 incident photons per ray and no scatter. We reconstructed a $512 \times 512 \times 90$ image volume with a coarser grid, where $\Delta_x = \Delta_y = 0.9776$ mm and $\Delta_z = 0.6250$ mm. We defined the statistical weighting matrix $\mathbf{W}$ as a diagonal matrix with diagonal entries $w_j \triangleq \exp(-y_j)$. An edge-preserving regularizer (based on a scaled Fair potential function) is used with parameters set to achieve the best noise-
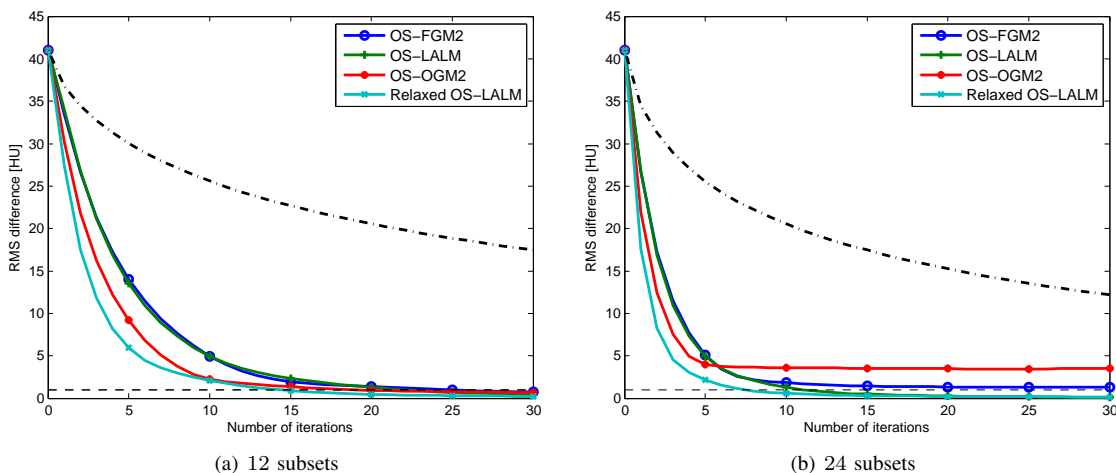
(a) 12 subsets

(b) 24 subsets

Fig. 3: XCAT phantom: RMS differences between the reference reconstruction $\mathbf{x}^\star$ and the reconstructed image $\mathbf{x}^{(k)}$ using different algorithms as a function of iteration with (a) 12 subsets and (b) 24 subsets, where OS-FGM2 and OS-OGM2 are the OS variants of the state-of-the-art fast gradient methods proposed in [7]. The dashed lines show the 1 HU RMS difference as the stopping criteria, and the dash-dot lines show the convergence rate curves of OS-SQS [8].

resolution trade-off.

Figure 2 shows the cropped image from the central transaxial plane of the initial FBP image $\mathbf{x}^{(0)}$, the reference reconstruction $\mathbf{x}^\star$, and the reconstructed image $\mathbf{x}^{(10)}$ using the proposed algorithm (relaxed OS-LALM with 24 subsets) after 10 iterations. There is no visible difference between the reference reconstruction and our reconstruction. To analyze the proposed algorithm quantitatively, Figure 3 shows the RMS differences between the reference reconstruction $\mathbf{x}^\star$ and the reconstructed image $\mathbf{x}^{(k)}$ using different algorithms as a function of iteration with 12 and 24 subsets, where OS-FGM2 and OS-OGM2 are the OS variants of the state-of-the-art fast gradient methods proposed in [7]. As seen in Figure 3, the proposed algorithm (cyan curves) is approximately twice as fast as the unrelaxed OS-LALM (green curves) at least in early iterations. Furthermore, comparing with OS-FGM2 and OS-OGM2, the proposed algorithm converges faster and is more stable when using more subsets for acceleration.

Trade-off between speed and stability always exists. Using over-relaxation lets OS-LALM converge faster but might introduce instability. For instance, the RMS difference of the proposed algorithm with 24 subsets decreases slower than the unrelaxed OS-LALM after 20 iterations. In practice, we do not observe stability issues when using moderate numbers of subsets for acceleration. Lastly, to ensure convergence, one might sacrifice some speed by using smaller step size. As an example, Chambolle and Pock showed that one has to shrink the primal-dual step sizes according to the relaxation parameter to ensure convergence if further majorizing the smooth term in the x-updates [9, Remark 6]. We also majorize the smooth regularizer empirically when solving image updates but did not have any convergence problem. This inspires our convergence rate analysis as future work.

IV. CONCLUSION

In this paper, we proposed a non-trivial relaxed variant of (OS-)LALM. Experimental results showed that our proposed algorithm converges twice as fast as its unrelaxed counterpart. Empirically, the method is reasonably stable when we use moderate numbers of subsets. For future work, we will work on the convergence rate analysis of the proposed algorithm and evaluate the proposed algorithm using real CT scans.

REFERENCES

[1] J. Eckstein and D. P. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Mathematical Programming*, vol. 55, pp. 293–318, Apr. 1992.

[2] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. & Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.

[3] X. Zhang, M. Burger, and S. Osher, "A unified primal-dual algorithm framework based on Bregman iteration," *Journal of Scientific Computing*, vol. 46, no. 1, pp. 20–46, 2011.

[4] H. Nien and J. A. Fessler, "Fast X-ray CT image reconstruction using a linearized augmented Lagrangian method with ordered subsets," *IEEE Trans. Med. Imag.*, vol. 34, pp. 388–99, Feb. 2015.

[5] E. X. Fang, B. He, H. Liu, and X. Yuan, "Generalized alternating direction method of multipliers: New theoretical insight and application," *Math. Prog. Comp.*, 2015. To appear.

[6] Z. Peng, M. Yan, and W. Yin, "Parallel and distributed sparse optimization," in *Proc. Asilomar Conf. Sig. Sys. Comp.*, 2013.

[7] D. Kim and J. A. Fessler, "Optimized first-order methods for smooth convex minimization," 2014. arXiv 1406.5468.

[8] H. Erdoğan and J. A. Fessler, "Ordered subsets algorithms for transmission tomography," *Phys. Med. Biol.*, vol. 44, pp. 2835–51, Nov. 1999.

[9] A. Chambolle and T. Pock, "On the ergodic convergence rates of a first-order primal-dual algorithm," 2014. Optimization Online.

[10] H. Nien and J. A. Fessler, "Relaxed linearized algorithms for faster X-ray CT image reconstruction," 2015. In preparation.

[11] 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, pp. 3800–8, Aug. 2008.