Published in final edited form as:

IEEE Trans Med Imaging. 2016 April ; 35(4): 1090–1098. doi:10.1109/TMI.2015.2508780.

Relaxed Linearized Algorithms for Faster X-Ray CT Image Reconstruction

Hung Nien [Member IEEE] and Jeffrey A. Fessler [Fellow IEEE]

Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109, USA

Hung Nien: hungnien@umich.edu; Jeffrey A. Fessler: fessler@umich.edu

Abstract

Statistical image reconstruction (SIR) methods are studied extensively for X-ray computed tomography (CT) due to the potential of acquiring CT scans with reduced X-ray dose while maintaining image quality. However, the longer reconstruction time of SIR methods hinders their use in X-ray CT in practice. To accelerate statistical methods, many optimization techniques have been investigated. Over-relaxation is a common technique to speed up convergence of iterative algorithms. For instance, using a relaxation parameter that is close to two in alternating direction method of multipliers (ADMM) has been shown to speed up convergence significantly. This paper proposes a relaxed linearized augmented Lagrangian (AL) method that shows theoretical faster convergence rate with over-relaxation and applies the proposed relaxed linearized AL method to X-ray CT image reconstruction problems. Experimental results with both simulated and real CT scan data show that the proposed relaxed algorithm (with ordered-subsets [OS] acceleration) is about twice as fast as the existing unrelaxed fast algorithms, with negligible computation and memory overhead.

Index Terms

Statistical image reconstruction; computed tomography; ordered subsets; augmented Lagrangian; relaxation

I. Introduction

Statistical image reconstruction (SIR) methods [1, 2] have been studied extensively and used widely in medical imaging. In SIR methods, one models the physics of the imaging system, the statistics of noisy measurements, and the prior information of the object to be imaged, and then finds the best fitted estimate by minimizing a cost function using iterative algorithms. By considering noise statistics when reconstructing images, SIR methods have better bias-variance performance and noise robustness. However, the iterative nature of algorithms in SIR methods also increases the reconstruction time, hindering their ubiquitous use in X-ray CT in practice.

Penalized weighted least-squares (PWLS) cost functions with a statistically weighted quadratic data-fidelity term are commonly used in SIR methods for X-ray CT [3]. Conventional SIR methods include the preconditioned conjugate gradient (PCG) method [4] and the separable quadratic surrogate (SQS) method with ordered-subsets (OS) acceleration [5]. These first-order methods update the image based on the gradient of the cost function at the current estimate. Due to the time-consuming forward/back-projection operations in X-ray CT when computing gradients, conventional first-order methods are typically very slow. The efficiency of PCG relies on choosing an appropriate preconditioner of the highly shift-variant Hessian caused by the huge dynamic range of the statistical weighting. In 2-D CT, one can introduce an auxiliary variable that separates the shift-variant and approximately shift-invariant components of the weighted quadratic data-fidelity term using a variable splitting technique [6], leading to better conditioned inner least-squares problems. However, this method has not worked well in 3-D CT, probably due to the 3-D cone-beam geometry and helical trajectory.

OS-SQS accelerates convergence using more frequent image updates by incremental gradients, i.e., computing image gradients with only a subset of data. This method usually exhibits fast convergence behavior in early iterations and becomes faster by using more subsets. However, it is not convergent in general [7, 8]. When more subsets are used, larger limit cycles can be observed. Unlike methods that update all voxels simultaneously, the iterative coordinate descent (ICD) method [9] updates one voxel at a time. Experimental results show that ICD approximately minimizes the PWLS cost function in several passes of the image volume if initialized appropriately; however, the sequential nature of ICD makes it difficult to parallelize and restrains the use of modern parallel computing architectures like GPU for speed-up.

OS-mom [10] and OS-LALM [11] are two recently proposed iterative algorithms that demonstrate promising fast convergence speed when solving 3-D X-ray CT image reconstruction problems. In short, OS-mom combines Nesterov's momentum techniques [12, 13] with the conventional OS-SQS algorithm, greatly accelerating convergence in early iterations. OS-LALM, on the other hand, is a linearized augmented Lagrangian (AL) method [14] that does not require inverting an enormous Hessian matrix involving the forward projection matrix when updating images, unlike typical splitting-based algorithms [6], but still enjoys the empirical fast convergence speed and error tolerance of AL methods such as the alternating direction method of multipliers (ADMM) [15–17]. Further acceleration from an algorithmic perspective is possible but seems to be more challenging. Kim et al. [18, 19] proposed two optimal gradient methods (OGM's) that use a new momentum term and showed a $\sqrt{2}$ -times speed-up for minimizing smooth convex functions, comparing to existing fast gradient methods (FGM's) [12, 13, 20, 21].

Over-relaxation is a common technique to speed up convergence of iterative algorithms. For example, it is very effective for accelerating ADMM [16, 17]. The same relaxation technique was also applied to linearized ADMM very recently [22], but the speed-up was less significant than expected. Chambolle et al. proposed a relaxed primal-dual algorithm (whose unrelaxed variant happens to be a linearized ADMM [23, Section 4.3]) and showed the first theoretical justification for speeding up convergence with over-relaxation [24, Theorem 2].

However, their theorem also pointed out that when the smooth explicit term (majorization of the Lipschitz part in the cost function mentioned later) is not zero, one must use smaller primal step size to ensure convergence with over-relaxation, precluding the use of larger relaxation parameter (close to two) for more acceleration. This paper proposes a non-trivial relaxed variant of linearized AL methods that improves the convergence rate by using larger relaxation parameter values (close to two) but does not require the step-size adjustment in [24]. We apply the proposed relaxed linearized algorithm to X-ray CT image reconstruction problems, and experimental results show that our proposed relaxation works much better than the simple relaxation [22] and significantly accelerates X-ray CT image reconstruction, even with ordered-subsets (OS) acceleration.

This paper is organized as follows. Section II shows the convergence rate of a linearized AL method (LALM) with simple relaxation and proposes a novel relaxed LALM whose convergence rate scales better with the relaxation parameter. Section III applies the proposed relaxed LALM to X-ray CT image reconstruction and uses a second-order recursive system analysis to derive a continuation sequence that speeds up the proposed algorithm. Section IV reports the experimental results of X-ray CT image reconstruction using the proposed algorithm. Finally, we draw conclusions in Section V. Online supplementary material contains many additional results and derivation details.

II. Relaxed linearized AL methods

We begin by discussing a more general constrained minimization problem for which X-ray CT image reconstruction is a special case considered in Section III. Consider an equality-constrained minimization problem:

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

where $g_{\mathbf{y}}$ and h are closed and proper convex functions. In particular, $g_{\mathbf{y}}$ is a loss function that measures the discrepancy between the linear model $\mathbf{A}\mathbf{x}$ and noisy measurement $\mathbf{y}$, and h is a regularization term that introduces prior knowledge of $\mathbf{x}$ to the reconstruction. We assume that the regularizer $h \triangleq \varphi + \psi$ is the sum of two convex components φ and ψ , where φ has inexpensive proximal mapping (prox-operator) defined as

$$\text{prox}_{\varphi}(\mathbf{x}) \triangleq \arg \min_{\mathbf{z}} \left\{ \varphi(\mathbf{z}) + \frac{1}{2} \|\mathbf{z} - \mathbf{x}\|_2^2 \right\}, \quad (2)$$

e.g., soft-shrinkage for the ℓ_1 -norm and truncating zeros for non-negativity constraints, and where ψ is continuously differentiable with L_{ψ} -Lipschitz gradients [25, p. 48], i.e.,

$$\|\nabla \psi(\mathbf{x}_1) - \nabla \psi(\mathbf{x}_2)\|_2 \leq L_{\psi} \|\mathbf{x}_1 - \mathbf{x}_2\|_2 \quad (3)$$

for any $\mathbf{x}_1$ and $\mathbf{x}_2$ in the domain of ψ . The Lipschitz condition of $\nabla \psi$ implies the “(quadratic) majorization condition” of ψ :

$$\psi(\mathbf{x}_2) \leq \psi(\mathbf{x}_1) + \langle \nabla \psi(\mathbf{x}_1), \mathbf{x}_2 - \mathbf{x}_1 \rangle + \frac{L_\psi}{2} \|\mathbf{x}_2 - \mathbf{x}_1\|_2^2. \quad (4)$$

More generally, one can replace the Lipschitz constant L_ψ by a diagonal majorizing matrix $\mathbf{D}_\psi$ based on the maximum curvature [26] or Huber's optimal curvature [27, p. 184] of ψ while still guaranteeing the majorization condition:

$$\psi(\mathbf{x}_2) \leq \psi(\mathbf{x}_1) + \langle \nabla \psi(\mathbf{x}_1), \mathbf{x}_2 - \mathbf{x}_1 \rangle + \frac{1}{2} \|\mathbf{x}_2 - \mathbf{x}_1\|_{\mathbf{D}_\psi}^2. \quad (5)$$

We show later that decomposing h into the proximal part φ and the Lipschitz part ψ is useful when solving minimization problems with composite regularization. For example, Section III writes iterative X-ray CT image reconstruction as a special case of (1), where g_y is a weighted quadratic function, and h is an edge-preserving regularizer with a non-negativity constraint on the reconstructed image.

A. Preliminaries

Solving the equality-constrained minimization problem (1) is equivalent to finding a saddle-point of the Lagrangian:

$$\mathcal{L}(\mathbf{x}, \mathbf{u}, \boldsymbol{\mu}) \triangleq f(\mathbf{x}, \mathbf{u}) - \langle \boldsymbol{\mu}, \mathbf{A}\mathbf{x} - \mathbf{u} \rangle, \quad (6)$$

where $f(\mathbf{x}, \mathbf{u}) \triangleq g_y(\mathbf{u}) + h(\mathbf{x})$, and $\boldsymbol{\mu}$ is the Lagrange multiplier of the equality constraint [28, p. 237]. In other words, $(\hat{\mathbf{x}}, \hat{\mathbf{u}}, \hat{\boldsymbol{\mu}})$ solves the minimax problem:

$$(\hat{\mathbf{x}}, \hat{\mathbf{u}}, \hat{\boldsymbol{\mu}}) \in \arg \min_{\mathbf{x}, \mathbf{u}} \max_{\boldsymbol{\mu}} \mathcal{L}(\mathbf{x}, \mathbf{u}, \boldsymbol{\mu}). \quad (7)$$

Moreover, since $(\hat{\mathbf{x}}, \hat{\mathbf{u}}, \hat{\boldsymbol{\mu}})$ is a saddle-point of $\mathcal{L}$, the following inequalities hold for any $\mathbf{x}, \mathbf{u}$, and $\boldsymbol{\mu}$:

$$\mathcal{L}(\mathbf{x}, \mathbf{u}, \hat{\boldsymbol{\mu}}) \geq \mathcal{L}(\hat{\mathbf{x}}, \hat{\mathbf{u}}, \hat{\boldsymbol{\mu}}) \geq \mathcal{L}(\hat{\mathbf{x}}, \hat{\mathbf{u}}, \boldsymbol{\mu}). \quad (8)$$

The non-negative duality gap function:

$$\mathcal{G}(\mathbf{x}, \mathbf{u}, \boldsymbol{\mu}; \hat{\mathbf{x}}, \hat{\mathbf{u}}, \hat{\boldsymbol{\mu}}) \triangleq \mathcal{L}(\mathbf{x}, \mathbf{u}, \hat{\boldsymbol{\mu}}) - \mathcal{L}(\hat{\mathbf{x}}, \hat{\mathbf{u}}, \boldsymbol{\mu}) = [f(\mathbf{x}, \mathbf{u}) - f(\hat{\mathbf{x}}, \hat{\mathbf{u}})] - \langle \hat{\boldsymbol{\mu}}, \mathbf{A}\mathbf{x} - \mathbf{u} \rangle \quad (9)$$

characterizes the accuracy of an approximate solution $(\mathbf{x}, \mathbf{u}, \boldsymbol{\mu})$ to the saddle-point problem (7). Note that $\hat{\mathbf{u}} = \mathbf{A}\hat{\mathbf{x}}$ due to the equality constraint. Besides solving the classic Lagrangian minimax problem (7), $(\hat{\mathbf{x}}, \hat{\mathbf{u}}, \hat{\boldsymbol{\mu}})$ also solves a family of minimax problems:

$$(\hat{\mathbf{x}}, \hat{\mathbf{u}}, \hat{\boldsymbol{\mu}}) \in \arg \min_{\mathbf{x}, \mathbf{u}} \max_{\boldsymbol{\mu}} \mathcal{L}_{\text{AL}}(\mathbf{x}, \mathbf{u}, \boldsymbol{\mu}), \quad (10)$$

where the augmented Lagrangian (AL) [25, p. 297] is

$$\mathcal{L}_{\text{AL}}(\mathbf{x}, \mathbf{u}, \boldsymbol{\mu}) \triangleq \mathcal{L}(\mathbf{x}, \mathbf{u}, \boldsymbol{\mu}) + \frac{\rho}{2} \|\mathbf{Ax} - \mathbf{u}\|_2^2. \quad (11)$$

The augmented quadratic penalty term penalizes the feasibility violation of the equality constraint, and the AL penalty parameter $\rho > 0$ controls the curvature of $\mathcal{L}_{\text{AL}}$ but does not change the solution, sometimes leading to better conditioned minimax problems.

One popular iterative algorithm for solving equality-constrained minimization problems based on the AL theory is ADMM, which solves the AL minimax problem (10), and thus the equality-constrained minimization problem (1), in an alternating direction manner. More precisely, ADMM minimizes AL (11) with respect to $\mathbf{x}$ and $\mathbf{u}$ alternatingly, followed by a gradient ascent of $\boldsymbol{\mu}$ with step size ρ . One can also interpolate or extrapolate variables in subproblems, leading to a relaxed AL method [16, Theorem 8]:

$$\begin{cases} \mathbf{x}^{(k+1)} \in \arg \min_{\mathbf{x}} \left\{ h(\mathbf{x}) - \langle \boldsymbol{\mu}^{(k)}, \mathbf{Ax} \rangle + \frac{\rho}{2} \|\mathbf{Ax} - \mathbf{u}^{(k)}\|_2^2 \right\} \\ \mathbf{u}^{(k+1)} \in \arg \min_{\mathbf{u}} \left\{ g_y(\mathbf{u}) + \langle \boldsymbol{\mu}^{(k)}, \mathbf{u} \rangle + \frac{\rho}{2} \|\mathbf{r}_{\mathbf{u},\alpha}^{(k+1)} - \mathbf{u}\|_2^2 \right\} \\ \boldsymbol{\mu}^{(k+1)} = \boldsymbol{\mu}^{(k)} - \rho(\mathbf{r}_{\mathbf{u},\alpha}^{(k+1)} - \mathbf{u}^{(k+1)}), \end{cases} \quad (12)$$

where the relaxation variable of $\mathbf{u}$ is:

$$\mathbf{r}_{\mathbf{u},\alpha}^{(k+1)} \triangleq \alpha \mathbf{Ax}^{(k+1)} + (1-\alpha) \mathbf{u}^{(k)}, \quad (13)$$

and $0 < \alpha < 2$ is the relaxation parameter. It is called over-relaxation when $\alpha > 1$ and under-relaxation when $\alpha < 1$. When α is unity, (12) reverts to the standard (alternating direction) AL method [15]. Experimental results suggest that over-relaxation with $\alpha \in [1.5, 1.8]$ can accelerate convergence [17].

Although (12) is used widely in applications, two concerns about the relaxed AL method (12) arise in practice. First, the cost function of the $\mathbf{x}$ -subproblem in (12) contains the augmented quadratic penalty of AL that involves $\mathbf{A}$, deeply coupling elements of $\mathbf{x}$ and often leading to an expensive iterative $\mathbf{x}$ -update, especially when $\mathbf{A}$ is large and unstructured, e.g., in X-ray CT. This motivates alternative methods like LALM [11, 14]. Second, even though LALM removes the $\mathbf{x}$ -coupling due to the augmented quadratic penalty, the regularization term h might not have inexpensive proximal mapping and still require an iterative $\mathbf{x}$ -update (albeit without using $\mathbf{A}$). This consideration inspires the decomposition $h \triangleq \varphi + \psi$ used in the algorithms discussed next.

B. Linearized AL methods with simple relaxation

In LALM¹, one adds an iteration-dependent proximity term:

$$\frac{1}{2} \|\mathbf{x} - \mathbf{x}^{(k)}\|_{\mathbf{P}}^2 \quad (14)$$

¹Because (15) is quadratic, not linear, a more apt term would be “majorized” rather than “linearized.” We stick with the term linearized for consistency with the literature on LALM.

to the $\mathbf{x}$ -update in (12) with $\alpha = 1$, where $\mathbf{P}$ is a positive semi-definite matrix. Choosing $\mathbf{P} = \rho \mathbf{G}$, where $\mathbf{G} \triangleq L_{\mathbf{A}} \mathbf{I} - \mathbf{A}' \mathbf{A}$, and $L_{\mathbf{A}}$ denotes the maximum eigenvalue of $\mathbf{A}' \mathbf{A}$, the non-separable Hessian of the augmented quadratic penalty of AL is cancelled, and the Hessian of

$$\frac{\rho}{2} \|\mathbf{A}\mathbf{x} - \mathbf{u}^{(k)}\|_2^2 + \frac{\rho}{2} \|\mathbf{x} - \mathbf{x}^{(k)}\|_{\mathbf{G}}^2 \quad (15)$$

becomes a diagonal matrix $\rho L_{\mathbf{A}} \mathbf{I}$, decoupling $\mathbf{x}$ in the $\mathbf{x}$ -update except for the effect of h . This technique is known as linearization (more precisely, majorization) because it majorizes a non-separable quadratic term by its linear component plus some separable quadratic proximity term. In general, one can also use

$$\mathbf{G} \triangleq \mathbf{D}_{\mathbf{A}} - \mathbf{A}' \mathbf{A}, \quad (16)$$

where $\mathbf{D}_{\mathbf{A}} \succeq \mathbf{A}' \mathbf{A}$ is a diagonal majorizing matrix of $\mathbf{A}' \mathbf{A}$, e.g., $\mathbf{D}_{\mathbf{A}} = \text{diag}\{|\mathbf{A}|' |\mathbf{A}| \mathbf{1}\} \succeq \mathbf{A}' \mathbf{A}$ [5], and still guarantee the positive semi-definiteness of $\mathbf{P}$. This trick can be applied to (12) when $\alpha = 1$, too.

To remove the possible coupling due to the regularization term h , we replace the Lipschitz part of $h \triangleq \varphi + \psi$ in the $\mathbf{x}$ -update of (12) with its separable quadratic surrogate (SQS):

$$Q_{\psi}(\mathbf{x}; \mathbf{x}^{(k)}) \triangleq \psi(\mathbf{x}^{(k)}) + \langle \nabla \psi(\mathbf{x}^{(k)}), \mathbf{x} - \mathbf{x}^{(k)} \rangle + \frac{1}{2} \|\mathbf{x} - \mathbf{x}^{(k)}\|_{\mathbf{D}_{\psi}}^2 \quad (17)$$

shown in (4) and (5). Note that (4) is just a special case of (5) when $\mathbf{D}_{\psi} = L_{\psi} \mathbf{I}$. Incorporating all techniques mentioned above, the $\mathbf{x}$ -update becomes simply a proximal mapping of φ , which by assumption is inexpensive. The resulting ‘‘LALM with simple relaxation’’ algorithm is:

$$\left\{ \begin{array}{l} \mathbf{x}^{(k+1)} \in \arg \min_{\mathbf{x}} \left\{ \phi(\mathbf{x}) + Q_{\psi}(\mathbf{x}; \mathbf{x}^{(k)}) - \langle \boldsymbol{\mu}^{(k)}, \mathbf{A}\mathbf{x} \rangle + \frac{\rho}{2} \|\mathbf{A}\mathbf{x} - \mathbf{u}^{(k)}\|_2^2 + \frac{\rho}{2} \|\mathbf{x} - \mathbf{x}^{(k)}\|_{\mathbf{G}}^2 \right\} \\ \mathbf{u}^{(k+1)} \in \arg \min_{\mathbf{u}} \left\{ g_{\mathbf{y}}(\mathbf{u}) + \langle \boldsymbol{\mu}^{(k)}, \mathbf{u} \rangle + \frac{\rho}{2} \|\mathbf{r}_{\mathbf{u}, \alpha}^{(k+1)} - \mathbf{u}\|_2^2 \right\} \\ \boldsymbol{\mu}^{(k+1)} = \boldsymbol{\mu}^{(k)} - \rho(\mathbf{r}_{\mathbf{u}, \alpha}^{(k+1)} - \mathbf{u}^{(k+1)}). \end{array} \right. \quad (18)$$

When $\psi = 0$, (18) reverts to the L-GADMM algorithm proposed in [22]. In [22], the authors analyzed the convergence rate of L-GADMM (for solving an equivalent variational inequality problem; however, there is no analysis on how relaxation parameter α affects the convergence rate) and investigated solving problems in statistical learning using L-GADMM. The speed-up resulting from over-relaxation was less significant than expected (e.g., when solving an X-ray CT image reconstruction problem discussed later). To explain the small speed-up, the following theorem shows that the duality gap (9) of the time-averaged approximate solution $\mathbf{w}_K = (\mathbf{x}_K, \mathbf{u}_K, \boldsymbol{\mu}_K)$ generated by (18) vanishes at rate $\mathcal{O}(1/K)$, where K is the number of iterations, and

$$\mathbf{c}_K \triangleq \frac{1}{K} \sum_{k=1}^K \mathbf{c}^{(k)} \quad (19)$$

denotes the time-average of some iterate $\mathbf{c}^{(k)}$ for $k = 1$ to K .

Theorem 1—Let $\mathbf{w}_K = (\mathbf{x}_K, \mathbf{u}_K, \boldsymbol{\mu}_K)$ be the time-averages of the iterates of LALM with simple relaxation in (18), where $\rho > 0$ and $0 < \alpha < 2$. We have

$$\mathcal{G}(\mathbf{w}_K; \hat{\mathbf{w}}) \leq \frac{1}{K} (A_{\mathbf{D}_\psi} + B_{\rho, \mathbf{D}_\mathbf{A}} + C_{\alpha, \rho}), \quad (20)$$

where the first two constants

$$A_{\mathbf{D}_\psi} \triangleq \frac{1}{2} \|\mathbf{x}^{(0)} - \hat{\mathbf{x}}\|_{\mathbf{D}_\psi}^2 \quad (21)$$

$$B_{\rho, \mathbf{D}_\mathbf{A}} \triangleq \frac{\rho}{2} \|\mathbf{x}^{(0)} - \hat{\mathbf{x}}\|_{\mathbf{D}_\mathbf{A} - \mathbf{A}'\mathbf{A}}^2 \quad (22)$$

depend on how far the initial guess is from a minimizer, and the last constant depends on the relaxation parameter

$$C_{\alpha, \rho} \triangleq \frac{1}{2\alpha} \left[\sqrt{\rho} \|\mathbf{u}^{(0)} - \hat{\mathbf{u}}\|_2 + \frac{1}{\sqrt{\rho}} \|\boldsymbol{\mu}^{(0)} - \hat{\boldsymbol{\mu}}\|_2 \right]^2. \quad (23)$$

Proof: The proof is in the supplementary material.

Theorem 1 shows that (18) converges at rate $\mathcal{O}(1/K)$, and the constant multiplying $1/K$ consists of three terms: $A_{\mathbf{D}_\psi}$, $B_{\rho, \mathbf{D}_\mathbf{A}}$, and $C_{\alpha, \rho}$. The first term $A_{\mathbf{D}_\psi}$ comes from the majorization of ψ , and it is large when ψ has large curvature. The second term $B_{\rho, \mathbf{D}_\mathbf{A}}$ comes from the linearization trick in (15). One can always decrease its value by decreasing ρ . The third term $C_{\alpha, \rho}$ is the only α -dependent component. The trend of $C_{\alpha, \rho}$ when varying ρ depends on the norms of $\mathbf{u}^{(0)} - \hat{\mathbf{u}}$ and $\boldsymbol{\mu}^{(0)} - \hat{\boldsymbol{\mu}}$, i.e., how one initializes the algorithm. Finally, the convergence rate of (18) scales well with α iff $C_{\alpha, \rho} \gg A_{\mathbf{D}_\psi}$ and $C_{\alpha, \rho} \gg B_{\rho, \mathbf{D}_\mathbf{A}}$. When ψ has large curvature or $\mathbf{D}_\mathbf{A}$ is a loose majorizing matrix of $\mathbf{A}'\mathbf{A}$ (like in X-ray CT), the above inequalities do not hold, leading to poor scalability of convergence rate with the relaxation parameter α .

C. Linearized AL methods with proposed relaxation

To better scale the convergence rate of relaxed LALM with α , we want to design an algorithm that replaces the α -independent components by α -dependent ones in the constant multiplying $1/K$ in (20). This can be (partially) done by linearizing (more precisely, majorizing) the non-separable AL penalty term in (12) implicitly. Instead of explicitly adding a $\mathbf{G}$ -weighted proximity term, where $\mathbf{G}$ is defined in (16), to the $\mathbf{x}$ -update like (18),

we consider solving an equality-constrained minimization problem equivalent to (1) with an additional redundant equality constraint $\mathbf{v} = \mathbf{G}^{1/2}\mathbf{x}$, i.e.,

$$(\hat{\mathbf{x}}, \hat{\mathbf{u}}, \hat{\mathbf{v}}) \in \arg \min_{\mathbf{x}, \mathbf{u}, \mathbf{v}} \{g_{\mathbf{y}}(\mathbf{u}) + h(\mathbf{x})\} \text{ s. t. } \mathbf{u} = \mathbf{A}\mathbf{x} \text{ and } \mathbf{v} = \mathbf{G}^{1/2}\mathbf{x}, \quad (24)$$

using the relaxed AL method (12) as follows:

$$\left\{ \begin{array}{l} \mathbf{x}^{(k+1)} \in \arg \min_{\mathbf{x}} \left\{ \phi(\mathbf{x}) + Q_{\psi}(\mathbf{x}; \mathbf{x}^{(k)}) - \langle \boldsymbol{\mu}^{(k)}, \mathbf{A}\mathbf{x} \rangle - \langle \boldsymbol{\nu}^{(k)}, \mathbf{G}^{1/2}\mathbf{x} \rangle + \frac{\rho}{2} \|\mathbf{A}\mathbf{x} - \mathbf{u}^{(k)}\|_2^2 + \frac{\rho}{2} \|\mathbf{G}^{1/2}\mathbf{x} - \mathbf{v}^{(k)}\|_2^2 \right\} \\ \mathbf{u}^{(k+1)} \in \arg \min_{\mathbf{u}} \left\{ g_{\mathbf{y}}(\mathbf{u}) + \langle \boldsymbol{\mu}^{(k)}, \mathbf{u} \rangle + \frac{\rho}{2} \|\mathbf{r}_{\mathbf{u}, \alpha}^{(k+1)} - \mathbf{u}\|_2^2 \right\} \\ \boldsymbol{\mu}^{(k+1)} = \boldsymbol{\mu}^{(k)} - \rho(\mathbf{r}_{\mathbf{u}, \alpha}^{(k+1)} - \mathbf{u}^{(k+1)}) \\ \mathbf{v}^{(k+1)} = \mathbf{r}_{\mathbf{v}, \alpha}^{(k+1)} - \rho^{-1}\boldsymbol{\nu}^{(k)} \\ \boldsymbol{\nu}^{(k+1)} = \boldsymbol{\nu}^{(k)} - \rho(\mathbf{r}_{\mathbf{v}, \alpha}^{(k+1)} - \mathbf{v}^{(k+1)}), \end{array} \right. \quad (25)$$

where the relaxation variable of $\mathbf{v}$ is:

$$\mathbf{r}_{\mathbf{v}, \alpha}^{(k+1)} \triangleq \alpha \mathbf{G}^{1/2}\mathbf{x}^{(k+1)} + (1-\alpha)\mathbf{v}^{(k)}, \quad (26)$$

and $\boldsymbol{\nu}$ is the Lagrange multiplier of the redundant equality constraint. One can easily verify that $\boldsymbol{\nu}^{(k)} = \mathbf{0}$ for $k = 0, 1, \dots$ if we initialize $\boldsymbol{\nu}$ as $\boldsymbol{\nu}^{(0)} = \mathbf{0}$.

The additional equality constraint introduces an additional inner-product term and a quadratic penalty term to the $\mathbf{x}$ -update. The latter can be used to cancel the non-separable Hessian of the AL penalty term as in explicit linearization. By choosing the same AL penalty parameter $\rho > 0$ for the additional constraint, the Hessian matrix of the quadratic penalty term in the $\mathbf{x}$ -update of (25) is $\rho\mathbf{A}'\mathbf{A} + \rho\mathbf{G} = \rho\mathbf{D}_{\mathbf{A}}$. In other words, by choosing $\mathbf{G}$ in (16), the quadratic penalty term in the $\mathbf{x}$ -update of (25) becomes separable, and the $\mathbf{x}$ -update becomes an efficient proximal mapping of φ , as seen in (30) below.

Next we analyze the convergence rate of the proposed relaxed LALM method (25). With the additional redundant equality constraint, the Lagrangian becomes

$$\mathcal{L}'(\mathbf{x}, \mathbf{u}, \boldsymbol{\mu}, \mathbf{v}, \boldsymbol{\nu}) \triangleq \mathcal{L}(\mathbf{x}, \mathbf{u}, \boldsymbol{\mu}) - \langle \boldsymbol{\nu}, \mathbf{G}^{1/2}\mathbf{x} - \mathbf{v} \rangle. \quad (27)$$

Setting gradients of $\mathcal{L}'$ with respect to $\mathbf{x}$, $\mathbf{u}$, $\boldsymbol{\mu}$, $\mathbf{v}$, and $\boldsymbol{\nu}$ to be zero yields a necessary condition for a saddle-point $\hat{\boldsymbol{\omega}} = (\hat{\mathbf{x}}, \hat{\mathbf{u}}, \hat{\boldsymbol{\mu}}, \hat{\mathbf{v}}, \hat{\boldsymbol{\nu}})$ of $\mathcal{L}'$. It follows that $\hat{\boldsymbol{\nu}} = \nabla_{\mathbf{v}}\mathcal{L}'(\hat{\boldsymbol{\omega}}) = \mathbf{0}$. Therefore, setting $\boldsymbol{\nu}^{(0)} = \mathbf{0}$ is indeed a natural choice for initializing $\boldsymbol{\nu}$. Moreover, since $\hat{\boldsymbol{\nu}} = \mathbf{0}$, the gap function $\mathcal{G}'$ of the new problem (24) coincides with (9), and we can compare the convergence rate of the simple and proposed relaxed algorithms directly.

Theorem 2—Let $\mathbf{w}_K = (\mathbf{x}_K, \mathbf{u}_K, \mathbf{v}_K, \boldsymbol{\mu}_K, \boldsymbol{\nu}_K)$ be the time-averages of the iterates of LALM with proposed relaxation in (25), where $\rho > 0$ and $0 < \alpha < 2$. When initializing $\mathbf{v}$ and $\boldsymbol{\nu}$ as $\mathbf{v}^{(0)} = \mathbf{G}^{1/2}\mathbf{x}^{(0)}$ and $\boldsymbol{\nu}^{(0)} = \mathbf{0}$, respectively, we have

$$\mathcal{G}'(\mathbf{w}_K; \hat{\mathbf{w}}) \leq \frac{1}{K}(A_{\mathbf{D}_\psi} + \bar{B}_{\alpha, \rho, \mathbf{D}_\mathbf{A}} + C_{\alpha, \rho}), \quad (28)$$

where $A_{\mathbf{D}_\psi}$ and $C_{\alpha, \rho}$ were defined in (21) and (23), and

$$\bar{B}_{\alpha, \rho, \mathbf{D}_\mathbf{A}} \triangleq \frac{\rho}{2\alpha} \|\mathbf{v}^{(0)} - \hat{\mathbf{v}}\|_2^2 = \frac{\rho}{2\alpha} \|\mathbf{x}^{(0)} - \hat{\mathbf{x}}\|_{\mathbf{D}_\mathbf{A} - \mathbf{A}'\mathbf{A}}^2. \quad (29)$$

Proof: The proof is in the supplementary material.

Theorem 2 shows the $\mathcal{O}(1/K)$ convergence rate of (25). Due to the different variable splitting scheme, the term introduced by the implicit linearization trick in (25) (i.e., $B_{\alpha, \rho, \mathbf{D}_\mathbf{A}}$) also depends on the relaxation parameter α , improving convergence rate scalability with α in (25) over (18). This theorem provides a theoretical explanation why (25) converges faster than (18) in the experiments shown later².

For practical implementation, the remaining concern is multiplications by $\mathbf{G}^{1/2}$ in (25). There is no efficient way to compute the square root of $\mathbf{G}$ for any $\mathbf{A}$ in general, especially when $\mathbf{A}$ is large and unstructured like in X-ray CT. To solve this problem, let $\mathbf{h} \triangleq \mathbf{G}^{1/2}\mathbf{v} + \mathbf{A}'\mathbf{y}$. We rewrite (25) so that no explicit multiplication by $\mathbf{G}^{1/2}$ is needed (the derivation is in the supplementary material), leading to the following ‘‘LALM with proposed relaxation’’ algorithm:

$$\left\{ \begin{array}{l} \mathbf{x}^{(k+1)} \in \arg \min_{\mathbf{x}} \left\{ \phi(\mathbf{x}) + Q_\psi(\mathbf{x}; \mathbf{x}^{(k)}) + \frac{1}{2} \|\mathbf{x} - (\rho \mathbf{D}_\mathbf{A})^{-1} \boldsymbol{\gamma}^{(k+1)}\|_{\rho \mathbf{D}_\mathbf{A}}^2 \right\} \\ \mathbf{u}^{(k+1)} \in \arg \min_{\mathbf{u}} \left\{ g_{\mathbf{y}}(\mathbf{u}) + \langle \boldsymbol{\mu}^{(k)}, \mathbf{u} \rangle + \frac{\rho}{2} \|\mathbf{r}_{\mathbf{u}, \alpha}^{(k+1)} - \mathbf{u}\|_2^2 \right\} \\ \boldsymbol{\mu}^{(k+1)} = \boldsymbol{\mu}^{(k)} - \rho(\mathbf{r}_{\mathbf{u}, \alpha}^{(k+1)} - \mathbf{u}^{(k+1)}) \\ \mathbf{h}^{(k+1)} = \alpha \boldsymbol{\eta}^{(k+1)} + (1 - \alpha) \mathbf{h}^{(k)}, \end{array} \right. \quad (30)$$

where

$$\boldsymbol{\gamma}^{(k+1)} \triangleq \rho \mathbf{A}'(\mathbf{u}^{(k)} - \mathbf{y} + \rho^{-1} \boldsymbol{\mu}^{(k)}) + \rho \mathbf{h}^{(k)}, \quad (31)$$

and

$$\boldsymbol{\eta}^{(k+1)} \triangleq \mathbf{D}_\mathbf{A} \mathbf{x}^{(k+1)} - \mathbf{A}'(\mathbf{A} \mathbf{x}^{(k+1)} - \mathbf{y}). \quad (32)$$

When $g_{\mathbf{y}}$ is a quadratic loss, i.e., $g_{\mathbf{y}}(\mathbf{z}) = (1/2) \|\mathbf{z} - \mathbf{y}\|_2^2$, we further simplify the proposed relaxed LALM by manipulations like those in [11] (omitted here for brevity) as:

²When ψ has large curvature (thus, α -dependent terms do not dominate the constant multiplying $1/K$), we can use techniques as in [29, 30] to reduce the ψ -dependent constant. In X-ray CT, the data-fidelity term often dominates the cost function, so $A_{\mathbf{D}_\psi} \ll B_{\alpha, \rho, \mathbf{D}_\mathbf{A}}$.

$$\left\{ \begin{array}{l} \gamma^{(k+1)} = (\rho-1)\mathbf{g}^{(k)} + \rho\mathbf{h}^{(k)} \\ \mathbf{x}^{(k+1)} \in \arg \min_{\mathbf{x}} \left\{ \phi(\mathbf{x}) + Q_{\psi}(\mathbf{x}; \mathbf{x}^{(k)}) + \frac{1}{2} \|\mathbf{x} - (\rho\mathbf{D}_{\mathbf{A}})^{-1} \gamma^{(k+1)}\|_{\rho\mathbf{D}_{\mathbf{A}}}^2 \right\} \\ \zeta^{(k+1)} \triangleq \nabla L(\mathbf{x}^{(k+1)}) = \mathbf{A}' (\mathbf{A}\mathbf{x}^{(k+1)} - \mathbf{y}) \\ \mathbf{g}^{(k+1)} = \frac{\rho}{\rho+1} (\alpha \zeta^{(k+1)} + (1-\alpha) \mathbf{g}^{(k)}) + \frac{1}{\rho+1} \mathbf{g}^{(k)} \\ \mathbf{h}^{(k+1)} = \alpha (\mathbf{D}_{\mathbf{A}} \mathbf{x}^{(k+1)} - \zeta^{(k+1)}) + (1-\alpha) \mathbf{h}^{(k)}, \end{array} \right. \quad (33)$$

where $L(\mathbf{x}) \triangleq g_{\mathbf{y}}(\mathbf{A}\mathbf{x})$ is the quadratic data-fidelity term, and $\mathbf{g} \triangleq \mathbf{A}'(\mathbf{u} - \mathbf{y})$ [11]. For initialization, we suggest using $\mathbf{g}^{(0)} = \zeta^{(0)}$ and $\mathbf{h}^{(0)} = \mathbf{D}_{\mathbf{A}}\mathbf{x}^{(0)} - \zeta^{(0)}$ (Theorem 2). The algorithm (33) computes multiplications by $\mathbf{A}$ and $\mathbf{A}'$ only once per iteration and does not have to invert $\mathbf{A}'\mathbf{A}$, unlike standard relaxed AL methods (12). This property is especially useful when $\mathbf{A}'\mathbf{A}$ is large and unstructured. When $\alpha = 1$, (33) reverts to the unrelaxed LALM in [11].

Lastly, we contrast our proposed relaxed LALM (30) with Chambolle's relaxed primal-dual algorithm [24, Algorithm 2]. Both algorithms exhibit $\mathcal{O}(1/K)$ ergodic (i.e., with respect to the time-averaged iterates) convergence rate and α -times speed-up when $\psi = 0$. Using (30) would require one more multiplication by $\mathbf{A}'$ per iteration than in Chambolle's relaxed algorithm; however, the additional $\mathbf{A}'$ is not required with quadratic loss in (33). When $\psi = 0$, unlike Chambolle's relaxed algorithm in which one has to adjust the primal step size according to the value of α (effectively, one scales $\mathbf{D}_{\psi}$ by $1/(2 - \alpha)$) [24, Remark 6], the proposed relaxed LALM (30) does not require such step-size adjustment, which is especially useful when using α that is close to two.

III. X-ray CT image reconstruction

Consider the X-ray CT image reconstruction problem [3]:

$$\hat{\mathbf{x}} \in \arg \min_{\mathbf{x} \in \Omega} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_{\mathbf{W}}^2 + R(\mathbf{x}) \right\}, \quad (34)$$

where $\mathbf{A}$ is the forward projection matrix of a CT scan [31], $\mathbf{y}$ is the noisy sinogram, $\mathbf{W}$ is the statistical diagonal weighting matrix, R denotes an edge-preserving regularizer, and Ω denotes a box-constraint on the image $\mathbf{x}$. We focus on the edge-preserving regularizer R defined as:

$$R(\mathbf{x}) \triangleq \sum_i \beta_i \sum_n \kappa_n \kappa_{n+s_i} \varphi_i([\mathbf{C}_i \mathbf{x}]_n), \quad (35)$$

where β_i , s_i , φ_i , and $\mathbf{C}_i$ denote the regularization parameter, spatial offset, potential function, and finite difference matrix in the i th direction, respectively, and κ_n is a voxel-dependent weight for improving resolution uniformity [32, 33]. In our experiments, we used 13 directions to include all 26 neighbors in 3-D CT.

A. Relaxed OS-LALM for faster CT reconstruction

To solve X-ray CT image reconstruction (34) using the proposed relaxed LALM (33), we apply the following substitution:

$$\begin{cases} \mathbf{A} \leftarrow \mathbf{W}^{1/2} \mathbf{A} \\ \mathbf{y} \leftarrow \mathbf{W}^{1/2} \mathbf{y}, \end{cases} \quad (36)$$

and we set $\varphi = \iota_\Omega$ and $\psi = \mathbf{R}$, where $\iota_\Omega(\mathbf{x}) = 0$ if $\mathbf{x} \in \Omega$, and $\iota_\Omega(\mathbf{x}) = +\infty$ otherwise. The proximal mapping of ι_Ω simply projects the input vector to the convex set Ω , e.g., clipping negative values of $\mathbf{x}$ to zero for a non-negativity constraint. Theorems developed in Section II considered the ergodic convergence rate of the non-negative duality gap, which is not a common convergence metric for X-ray CT image reconstruction. However, the ergodic convergence rate analysis suggests how factors like α , ρ , $\mathbf{D}_\mathbf{A}$, and $\mathbf{D}_\psi$ affect convergence speed (a LASSO regression example can be found in the supplementary material) and motivates our “more practical” (over-)relaxed OS-LALM summarized below.

Algorithm 1

Proposed (over-)relaxed OS-LALM for (34).

Input: $M \geq 1$, $1 \leq \alpha < 2$, and an initial (FBP) image $\mathbf{x}$.

set $\rho = 1$, $\boldsymbol{\zeta} = \mathbf{g} = M \nabla \mathbf{L}_M(\mathbf{x})$, $\mathbf{h} = \mathbf{D}_\mathbf{L} \mathbf{x} - \boldsymbol{\zeta}$

for $k = 1, 2, \dots$ **do**

for $m = 1, 2, \dots, M$ **do**

$\mathbf{s} = \rho (\mathbf{D}_\mathbf{L} \mathbf{x} - \mathbf{h}) + (1 - \rho) \mathbf{g}$

$\mathbf{x}^+ = [\mathbf{x} - (\rho \mathbf{D}_\mathbf{L} + \mathbf{D}_\mathbf{R})^{-1} (\mathbf{s} + \nabla \mathbf{R}(\mathbf{x}))]_\Omega$

$\boldsymbol{\zeta} = M \nabla \mathbf{L}_m(\mathbf{x}^+)$

$\mathbf{g}^+ = \frac{\rho}{\rho+1} (\alpha \boldsymbol{\zeta} + (1 - \alpha) \mathbf{g}) + \frac{1}{\rho+1} \mathbf{g}$

$\mathbf{h}^+ = \alpha (\mathbf{D}_\mathbf{L} \mathbf{x}^+ - \boldsymbol{\zeta}) + (1 - \alpha) \mathbf{h}$

decrease ρ using (37)

end

end

Algorithm 1 describes the proposed relaxed algorithm for solving the X-ray CT image reconstruction problem (34), where $\mathbf{L}_m$ denotes the data-fidelity term of the m th subset, and $[\cdot]_\Omega$ is an operator that projects the input vector onto the convex set Ω , e.g., truncating zeros for $\Omega \triangleq \{\mathbf{x} / x_i \geq 0 \text{ for all } i\}$. All variables are updated in-place, and we use the superscript $(\cdot)^+$ to denote the new values that replace the old values. We also use the substitution $\mathbf{s} \triangleq \rho \mathbf{D}_\mathbf{L} \mathbf{x} - \boldsymbol{\gamma}^\dagger$ in the proposed method, so Algorithm 1 has comparable form with the unrelaxed OS-LALM [11]; however, such substitution is not necessary.

As seen in Algorithm 1, the proposed relaxed OS-LALM has the form of (33) but uses some modifications that violate assumptions in our theorems but speed up “convergence” in practice. First, although Theorem 2 assumes a constant majorizing matrix $\mathbf{D}_\mathbf{R}$ for the Lipschitz term $\mathbf{R}$ (e.g., the maximum curvature of $\mathbf{R}$), we use the iteration-dependent Huber’s curvature of $\mathbf{R}$ [26] for faster convergence (the same in other algorithms for comparison). Second, since the updates in (33) depend only on the gradients of $\mathbf{L}$, we can further accelerate the gradient computation by using partial projection data, i.e., ordered

subsets. Lastly, we incorporate continuation technique (i.e., decreasing the AL penalty parameter ρ every iteration) in the proposed algorithm as described in the next subsection.

To select the number of subsets, we used the rule suggested in [11, Eqn. 55 and 57]. However, since over-relaxation provides two-times acceleration, we used 50% of the suggested number of subsets (for the unrelaxed OS-LALM) yet achieved similar convergence speed (faster in runtime since fewer regularizer gradient evaluations are performed) and more stable reconstruction. For the implicit linearization, we use the diagonal majorizing matrix $\text{diag}\{\mathbf{A}'\mathbf{W}\mathbf{A}\mathbf{1}\}$ for $\mathbf{A}'\mathbf{W}\mathbf{A}$ [5], the same diagonal majorizing matrix $\mathbf{D}_L$ for the quadratic loss function used in OS algorithms.

Furthermore, Algorithm 2 depicts the OS version of the simple relaxed algorithm (18) for solving (34) (derivation is omitted here). The main difference between Algorithm 1 and Algorithm 2 is the extra recursion of variable $\mathbf{h}$. When $\alpha = 1$, both algorithms revert to the unrelaxed OS-LALM [11].

Algorithm 2

Simple (over-)relaxed OS-LALM for (34).

Input: $M \geq 1$, $1 \leq \alpha < 2$, and an initial (FBP) image $\mathbf{x}$.
 set $\rho = 1$, $\zeta = \mathbf{g} = M\nabla L_M(\mathbf{x})$
for $k = 1, 2, \dots$ **do**
 for $m = 1, 2, \dots, M$ **do**
 $\mathbf{s} = \rho \zeta + (1 - \rho) \mathbf{g}$
 $\mathbf{x}^+ = [\mathbf{x} - (\rho \mathbf{D}_L + \mathbf{D}_R)^{-1} (\mathbf{s} + \nabla R(\mathbf{x}))]_{\Omega}$
 $\zeta^+ = M\nabla L_m(\mathbf{x}^+)$
 $\mathbf{g}^+ = \frac{\rho}{\rho+1} (\alpha \zeta^+ + (1 - \alpha) \mathbf{g}) + \frac{1}{\rho+1} \mathbf{g}$
 decrease ρ using (37)
 end
end

B. Further speed-up with continuation

We also use a continuation technique [11] to speed up convergence; that is, we decrease ρ gradually with iteration. Note that $\rho \mathbf{D}_L + \mathbf{D}_R$ is the inverse of the voxel-dependent step size of image updates; decreasing ρ increases step sizes gradually as iteration progress. Due to the extra relaxation parameter α , the good decreasing continuation sequence differs from that in [11]. We use the following α -dependent continuation sequence for the proposed relaxed LALM ($1 \leq \alpha < 2$):

$$\rho_k(\alpha) = \begin{cases} 1, & \text{if } k=0 \\ \frac{\pi}{\alpha(k+1)} \sqrt{1 - \left(\frac{\pi}{2\alpha(k+1)}\right)^2}, & \text{otherwise.} \end{cases} \quad (37)$$

The supplementary material describes the rationale for this continuation sequence. When using OS, ρ decreases every subiteration, and the counter k in (37) denotes the number of subiterations, instead of the number of iterations.

IV. Experimental results

This section reports numerical results for 3-D X-ray CT image reconstruction using one conventional algorithm (OS-SQS [5]) and four contemporary algorithms:

- **OS-FGM2**: the OS variant of the standard fast gradient method proposed in [10, 19],
- **OS-LALM**: the OS variant of the unrelaxed linearized AL method proposed in [11],
- **OS-OGM2**: the OS variant of the optimal fast gradient method proposed in [19], and
- **Relaxed OS-LALM**: the OS variants of the proposed relaxed linearized AL methods given in Algorithm 1 (proposed) and Algorithm 2 (simple) above ($\alpha = 1.999$ unless otherwise specified).

A. XCAT phantom

We simulated an axial CT scan using a $1024 \times 1024 \times 154$ XCAT phantom [34] for 500 mm transaxial field-of-view (FOV), where $x = y = 0.4883$ mm and $z = 0.625$ mm. An $888 \times 64 \times 984$ ([detector columns] $\times$ [detector rows] $\times$ [projection views]) 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 $x = y = 0.9776$ mm and $z = 0.625$ mm. The statistical weighting matrix $\mathbf{W}$ is defined as a diagonal matrix with diagonal entries $w_j \triangleq \exp(-y_j)$, and an edge-preserving regularizer is used with $\phi(\delta) \triangleq \delta^2 (1/\delta - \log(1 + 1/\delta))$ ($\delta = 10$ HU) and parameters β_i set to achieve a reasonable noise-resolution trade-off. We used 12 subsets for the relaxed OS-LALM, while [11, Eqn. 55] suggests using about 24 subsets for the unrelaxed OS-LALM.

Figure 1 shows the cropped images (displayed from 800 to 1200 HU [modified so that air is 0]) from the central transaxial plane of the initial FBP image $\mathbf{x}^{(0)}$, the reference reconstruction $\mathbf{x}^*$ (generated by running thousands of iterations of the convergent FGM with adaptive restart [35]), and the reconstructed image $\mathbf{x}^{(20)}$ using the proposed algorithm (relaxed OS-LALM with 12 subsets) after 20 iterations. There is no visible difference between the reference reconstruction and our reconstruction. To analyze the proposed algorithm quantitatively, Figure 2 shows the RMS differences between the reference reconstruction $\mathbf{x}^*$ and the reconstructed image $\mathbf{x}^{(k)}$ using different algorithms as a function of iteration³ with 12 and 24 subsets. As seen in Figure 2, the proposed algorithm (cyan curves) is approximately twice as fast as the unrelaxed OS-LALM (green curves) at least in early

³All algorithms listed above require one forward/back-projection pair and M (the number of subsets) regularizer gradient evaluations (plus some negligible overhead) per iteration, so comparing the convergence rate as a function of iteration is fair.

iterations. Furthermore, comparing with OS-FGM2 and OS-OGM2, the proposed algorithm converges faster and is more stable when using more subsets for acceleration. Difference images using different algorithms and additional experimental results are shown in the supplementary material.

To illustrate the improved speed-up of the proposed relaxation (Algorithm 1) over the simple one (Algorithm 2), Figure 3 shows convergence rate curves of different relaxed algorithms (12 subsets and $\alpha = 1.999$) with (a) a fixed AL penalty parameter $\rho = 0.05$ and (b) the decreasing sequence ρ_k in (37). As seen in Figure 3(a), the simple relaxation does not provide much acceleration, especially after 10 iterations. In contrast, the proposed relaxation accelerates convergence about twice (i.e., α -times), as predicted by Theorem 2. When the decreasing sequence of ρ_k is used, as seen in Figure 3(b), the simple relaxation seems to provide somewhat more acceleration than before; however, the proposed relaxation still outperforms the simple one, illustrating approximately twofold speed-up over the unrelaxed counterpart.

B. Chest scan

We reconstructed a $600 \times 600 \times 222$ image volume, where $x = y = 1.1667$ mm and $z = 0.625$ mm, from a chest region helical CT scan. The size of sinogram is $888 \times 64 \times 3611$ and pitch 1.0 (about 3.7 rotations with rotation time 0.4 seconds). The tube current and tube voltage of the X-ray source are 750 mA and 120 kVp, respectively. We started from a smoothed FBP image $\mathbf{x}^{(0)}$ and tuned the statistical weights [36] and the q -generalized Gaussian MRF regularization parameters [33] to emulate the MBIR method [3, 37]. We used 10 subsets for the relaxed OS-LALM, while [11, Eqn. 57] suggests using about 20 subsets for the unrelaxed OS-LALM. Figure 4 shows the cropped images from the central transaxial plane of the initial FBP image $\mathbf{x}^{(0)}$, the reference reconstruction $\mathbf{x}^*$, and the reconstructed image $\mathbf{x}^{(20)}$ using the proposed algorithm (relaxed OS-LALM with 10 subsets) after 20 iterations. Figure 5 shows the RMS differences between the reference reconstruction $\mathbf{x}^*$ and the reconstructed image $\mathbf{x}^{(k)}$ using different algorithms as a function of iteration with 10 and 20 subsets. The proposed relaxed OS-LALM shows about two-times faster convergence rate, comparing to its unrelaxed counterpart, with moderate number of subsets. The speed-up diminishes as the iterate approaches the solution. Furthermore, the faster relaxed OS-LALM seems likely to be more sensitive to gradient approximation errors and exhibits ripples in convergence rate curves when using too many subsets for acceleration. In contrast, the slower unrelaxed OS-LALM is less sensitive to gradient error when using more subsets and does not exhibit such ripples in convergence rate curves. Compared with OS-FGM2 and OS-OGM2, the proposed relaxed OS-LALM has smaller limit cycles and might be more stable for practical use.

V. Discussion and conclusions

In this paper, we proposed a non-trivial relaxed variant of LALM and applied it to X-ray CT image reconstruction. Experimental results with simulated and real CT scan data showed that our proposed relaxed algorithm “converges” about twice as fast as its unrelaxed counterpart, outperforming state-of-the-art fast iterative algorithms using momentum [10, 19]. This speed-up means that one needs fewer subsets to reach an RMS difference criteria

like 1 HU in a given number of iterations. For instance, we used 50% of the number of subsets suggested by [11] (for the unrelaxed OS-LALM) in our experiment but found similar convergence speed with over-relaxation. Moreover, using fewer subsets can be beneficial for distributed computing [38], reducing communication overhead required after every update.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work is supported in part by National Institutes of Health (NIH) grant U01-EB-018753 and by equipment donations from Intel Corporation.

The authors thank GE Healthcare for providing sinogram data in our experiments. The authors would also like to thank the anonymous reviewers for their comments and suggestions.

References

1. Fessler JA. Penalized weighted least-squares image reconstruction for positron emission tomography. *IEEE Trans Med Imag.* Jun.1994 13:290–300.
2. Nuyts J, De Man B, Fessler JA, Zbijewski W, Beekman FJ. Modelling the physics in iterative reconstruction for transmission computed tomography. *Phys Med Biol.* Jun.2013 58:R63–96. [PubMed: 23739261]
3. Thibault JB, Sauer K, Bouman C, Hsieh J. A three-dimensional statistical approach to improved image quality for multi-slice helical CT. *Med Phys.* Nov.2007 34:4526–44. [PubMed: 18072519]
4. Fessler JA, Booth SD. Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction. *IEEE Trans Im Proc.* May.1999 8:688–99.
5. Erdo an H, Fessler JA. Ordered subsets algorithms for transmission tomography. *Phys Med Biol.* Nov.1999 44:2835–51. [PubMed: 10588288]
6. Ramani S, Fessler JA. A splitting-based iterative algorithm for accelerated statistical X-ray CT reconstruction. *IEEE Trans Med Imag.* Mar.2012 31:677–88.
7. Ahn S, Fessler JA. Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms. *IEEE Trans Med Imag.* May.2003 22:613–26.
8. Ahn S, Fessler JA, Blatt D, Hero AO. Convergent incremental optimization transfer algorithms: Application to tomography. *IEEE Trans Med Imag.* Mar.2006 25:283–96.
9. Yu Z, Thibault JB, Bouman CA, Sauer KD, Hsieh J. Fast model-based X-ray CT reconstruction using spatially non-homogeneous ICD optimization. *IEEE Trans Im Proc.* Jan.2011 20:161–75.
10. Kim D, Ramani S, Fessler JA. Combining ordered subsets and momentum for accelerated X-ray CT image reconstruction. *IEEE Trans Med Imag.* Jan.2015 34:167–78.
11. Nien H, Fessler JA. Fast X-ray CT image reconstruction using a linearized augmented Lagrangian method with ordered subsets. *IEEE Trans Med Imag.* Feb.2015 34:388–99.
12. Nesterov Y. A method for unconstrained convex minimization problem with the rate of convergence $O(1/k^2)$. *Dokl Akad Nauk USSR.* 1983; 269(3):543–7.
13. Nesterov Y. Smooth minimization of non-smooth functions. *Mathematical Programming.* May. 2005 103:127–52.
14. Zhang X, Burger M, Osher S. A unified primal-dual algorithm framework based on Bregman iteration. *Journal of Scientific Computing.* 2011; 46(1):20–46.
15. Gabay D, Mercier B. A dual algorithm for the solution of nonlinear variational problems via finite-element approximations. *Comput Math Appl.* 1976; 2(1):17–40.
16. Eckstein J, Bertsekas DP. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming.* Apr.1992 55:293–318.

17. Boyd S, Parikh N, Chu E, Peleato B, Eckstein J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found & Trends in Machine Learning*. 2010; 3(1):1–122.
18. Kim D, Fessler JA. Optimized momentum steps for accelerating X-ray CT ordered subsets image reconstruction. *Proc 3rd Intl Mtg on image formation in X-ray CT*. 2014:103–6.
19. Kim D, Fessler JA. Optimized first-order methods for smooth convex minimization. *Mathematical Programming*. 2016
20. Nesterov Y. On an approach to the construction of optimal methods of minimization of smooth convex functions. *Ekonomika i Mateaticheskie Metody*. 1988; 24:509–17. In Russian.
21. Beck A, Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J Imaging Sci*. 2009; 2(1):183–202.
22. Fang EX, He B, Liu H, Yuan X. Generalized alternating direction method of multipliers: New theoretical insight and application. *Math Prog Comp*. Jun.2015 7:149–87.
23. Chambolle A, Pock T. A first-order primal-dual algorithm for convex problems with applications to imaging. *J Math Im Vision*. 2011; 40(1):120–45.
24. Chambolle A, Pock T. On the ergodic convergence rates of a first-order primal-dual algorithm. *Optimization Online*. 2014 Preprint ID 2014-09-4532.
25. Bertsekas, DP. *Nonlinear programming*. 2. Belmont: Athena Scientific; 1999.
26. Erdo an H, Fessler JA. Monotonic algorithms for transmission tomography. *IEEE Trans Med Imag*. Sep.1999 18:801–14.
27. Huber, PJ. *Robust statistics*. New York: Wiley; 1981.
28. Boyd, S.; Vandenberghe, L. *Convex optimization*. UK: Cambridge; 2004.
29. Azadi S, Sra S. Towards an optimal stochastic alternating direction method of multipliers. *Proc Intl Conf on Mach Learning*. 2014:620–8.
30. Ouyang Y, Chen Y, Lan G, Pasiliao E Jr. An accelerated linearized alternating direction method of multipliers. *SIAM J Imaging Sci*. 2015; 8(1):644–81.
31. Long Y, Fessler JA, Balter JM. 3D forward and back-projection for X-ray CT using separable footprints. *IEEE Trans Med Imag*. Nov.2010 29:1839–50.
32. Fessler JA, Rogers WL. Spatial resolution properties of penalized-likelihood image reconstruction methods: Space-invariant tomographs. *IEEE Trans Im Proc*. Sep.1996 5:1346–58.
33. Cho JH, Fessler JA. Regularization designs for uniform spatial resolution and noise properties in statistical image reconstruction for 3D X-ray CT. *IEEE Trans Med Imag*. Feb.2015 34:678–89.
34. Segars WP, Mahesh M, Beck TJ, Frey EC, Tsui BMW. Realistic CT simulation using the 4D XCAT phantom. *Med Phys*. Aug.2008 35:3800–8. [PubMed: 18777939]
35. O'Donoghue B, Candès E. Adaptive restart for accelerated gradient schemes. *Found Comp Math*. Jun.2015 15:715–32.
36. Chang Z, Zhang R, Thibault J-B, Sauer K, Bouman C. Statistical x-ray computed tomography from photon-starved measurements. *Proc SPIE 9020 Computational Imaging XII*. 2014:90200G.
37. Shuman WP, Green DE, Busey JM, Kolokythas O, Mitsumori LM, Koprowicz KM, Thibault JB, Hsieh J, Alessio AM, Choi E, Kinahan PE. Model-based iterative reconstruction versus adaptive statistical iterative reconstruction and filtered back projection in 64-MDCT: Focal lesion detection, lesion conspicuity, and image noise. *Am J Roentgenol*. May.2013 200:1071–6. [PubMed: 23617492]
38. Rosen JM, Wu J, Wenisch TF, Fessler JA. Iterative helical CT reconstruction in the cloud for ten dollars in five minutes. *Proc Intl Mtg on Fully 3D Image Recon in Rad and Nuc Med*. 2013:241–4.

Fig. 1.

XCAT: 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 $\mathbf{x}^*$ (center), and the reconstructed image $\mathbf{x}^{(20)}$ using the proposed algorithm (relaxed OS-LALM with 12 subsets) after 20 iterations (right).

Fig. 2.

XCAT: Convergence rate curves of different OS algorithms with (a) 12 subsets and (b) 24 subsets. The proposed relaxed OS-LALM with 12 subsets exhibits similar convergence rate as the unrelaxed OS-LALM with 24 subsets.

Fig. 3. XCAT: Convergence rate curves of different relaxed algorithms (12 subsets and $\alpha = 1.999$) with (a) a fixed AL penalty parameter $\rho = 0.05$ and (b) the decreasing sequence ρ_k in (37).

Fig. 4.

Chest: 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 $\mathbf{x}^*$ (center), and the reconstructed image $\mathbf{x}^{(20)}$ using the proposed algorithm (relaxed OS-LALM with 10 subsets) after 20 iterations (right).

Fig. 5.

Chest: Convergence rate curves of different OS algorithms with (a) 10 subsets and (b) 20 subsets. The proposed relaxed OS-LALM with 10 subsets exhibits similar convergence rate as the unrelaxed OS-LALM with 20 subsets.