Accelerating Iterative Image Reconstruction via Adaptive Surrogate Functions

Ayan Mitra[1], David G. Politte[2], and Joseph A. O'Sullivan[1]
[1]Deparment of Electrical and Systems Engineering, Washington University
One Brookings Drive, Saint Louis, MO, USA, 63130
[2]Mallinckrodt Institute of Radiology, Washington University School of Medicine
510 South Kingshighway Blvd, Saint Louis, MO, USA, 63110

Abstract

Three-dimensional statistical iterative reconstruction (SIR) algorithms have the potential to significantly reduce image artifacts by minimizing a cost function that models the physics and statistics of the data acquisition process in x-ray CT. SIR algorithms are important for a wide range of applications including nonstandard geometries arising from irregular sampling, limited angular range, missing data, and low-dose CT. For iterative image reconstruction algorithms to be deployed in clinical settings, the images must be quantitatively accurate and computed in clinically useful times. We describe an acceleration method that is based on adaptively varying an update factor of the additive step of the alternating minimization (AM) algorithm. Our implementation combines this method with other acceleration techniques like ordered subsets (OS) which was originally proposed for transmission tomography by Ahn, Fessler et. al [1]. Results on both an NCAT phantom and real clinical data from a Siemens Sensation 16 scanner demonstrate an improved convergence rate compared to the straightforward implementations of the alternating minimization (AM) algorithm of O'Sullivan and Benac [2] with a Huber-type edge-preserving penalty, originally proposed by Lange [3]. Our proposed acceleration method on average yields 2X acceleration of the convergence rate for both baseline and ordered subset implementations of the AM algorithm.

Introduction

Every image reconstruction method for X-ray CT can be broadly classified into two major families: linear methods such as filtered backprojection (FBP) and statistical methods. Statistical methods are mostly iterative in nature where the next image estimate is computed based on the measure of error between measured data and predicted data from the current image estimate along with a regularization function. SIR algorithms may produce quantitatively better images with lower x-ray dose which is important in many clinical applications [4, 5]. However, in most clinical applications, linear reconstruction algorithms like FBP or Feldkamp-Davis-Kress (FDK) are used due to their simplicity and low computational burden [6, 7].

Several statistical image reconstruction algorithms for x-ray CT have been proposed throughout the years (Sukovic and Clinthorne [8], De Man, et al. [9], Williamson, et al. [10], Fessler, et al. [11, 12]). The main motivations for shifting towards these iterative algorithms are that they can realistically model the nonlinear detector response in the attenuation line integral caused by beam hardening, scatter, and stochastic properties of the measured data. Other advantages include the adaptability of the method for specific detector-response models and the possibility of incorporating additional constraints.

The main hurdles for the adoption of SIR algorithms in practice are the iterative nature of these algorithms and the larger computation time. The actual computation time demand varies depending on the field of application which in turn dependent on the volume of the measured data and the amount of desired accuracy of the reconstructed images. In security applications, the reconstruction time of three-dimensional image volumes must satisfy the rate at which bags travel through the scanner. For many medical applications, the time depends on the availability of radiologists, which can vary widely. There are various pathways to decrease the time required for iterative image reconstruction. One of the effective pathways is to use multiple graphical processing units (GPUs) to parallelize the computationally intensive parts of the algorithm [13-17]. A second pathway is through advanced algorithms from convex optimization theory [18]. Another way is by accelerating the convergence rate of existing algorithms at the expense of not having guaranteed convergence properties [19-21]. A new method named *adaptive auxiliary variable* is investigated in this article for accelerating the convergence rate of the AM algorithm and is evaluated using a simulated phantom and real clinical data obtained from a Siemens Sensation 16 scanner.

In our current work, we first assume a Poisson distribution model for the measured transmission data. Then we calculate a maximum likelihood estimate of the measured data and data model by reformulating the estimation problem as a double minimization of an I-divergence problem. A Huber-type penalty is then added to the divergence term. Finally, we formulate an objective function with the I-divergence and regularization terms. As the optimization space is quite large, we have reformulated the objective function as N one-dimensional convex optimization problems, where N is the number of voxels in the image being reconstructed. After that, we have provided pseudocodes for the general AM algorithm and its accelerated version with the ordered-subset technique. Next, we derive our proposed auxiliary variable based acceleration method and present a pseudocode for its efficient parallel implementation. Finally, we have validated our proposed acceleration technique with NCAT phantoms and Siemens Sensation 16 helical scan data by comparing the convergence rates of a straightforward implementation of the AM algorithm with its accelerated version.

Convex Optimization Technique for X-ray CT

Mathematical Model

In this article, we consider a mono-energetic scatter-free statistical model to account for the x-ray photon randomness as was done previously [2, 22]. At the basis of our statistical model, we assume the photons arrive at the detectors in accordance with a photon counting process. Let the 3D image volume of linear attenuation coefficients (in mm^{-1}) be represented by the vector μ. Let i denote a

ray path between the x-ray source and a pixel in the multi-row detector array and j denote a voxel in the image volume. The measured transmission data, d_i, is modeled as originating from independent Poisson counting processes. In discretized form, the mean value of d_i is

$$g_i(\mu) = I_{0,i}\exp\left(-\sum_j h_{ij}\mu_j\right) + \beta_i, \quad (1)$$

where $I_{0,i}$ is the mean number of counts in the absence of an attenuating medium, β_i is the mean number of background events, assumed to be nonnegative and known, and μ_j is the linear attenuation coefficient in voxel j. The system matrix elements h_{ij} comprise the appropriately discretized point-spread function relating the projection space to the image space. If projection i does not pass through voxel j, then h_{ij} is zero. The Poisson log-likelihood function is

$$l(d;\mu) = \sum_i\left[d_i\log\left(g_i(\mu)\right) - g_i(\mu)\right]. \quad (2)$$

The objective of our iterative reconstruction algorithm is to maximize the log-likelihood function in (2) subject to μ_j being nonnegative, due to the nature of linear attenuation coefficients. It turns out that maximizing $l(d;\mu)$ is equivalent to minimizing the I-divergence between d and $g(\mu)$. In other words

$$\mu_{ML}^* = \underset{\mu \geq 0}{\mathrm{argmax}}\, l(d;\mu) = \underset{\mu \geq 0}{\mathrm{argmin}}\, \mathrm{I}(d||g(\mu)), \quad (3)$$

where the I-divergence $I[d||g(\mu)]$ is defined as

$$I[d||g(\mu)] \triangleq \sum_i\left[d_i\ln\left(\frac{d_i}{g_i(\mu)}\right) + g_i(\mu) - d_i\right]. \quad (4)$$

The objective function presented in (3) cannot be optimized directly over μ since the optimization space is large. One of the best approaches is to develop surrogate functions that approximate the original function at every iteration and are easy to minimize. This approach leads to iterative algorithms where different surrogate functions are formed and minimized at each iteration and yet the original function decreases monotonically.

Because this is an ill-posed inverse problem, we add a penalty term, $R(\mu)$, to the objective function used in the AM reconstruction, and weight it by a regularization parameter λ, where λ is a scalar that reflects the amount of smoothing desired. A larger value will give emphasis to the penalty term (i.e., the prior expectation that the image will be smooth), whereas a smaller value will give more emphasis to the I-divergence term (i.e., the discrepancy between the measured data and the data estimated by the model). The added penalty term is defined as

$$R(\mu_j) \triangleq \sum_{j' \in N_j} \omega_{j,j'}\psi(\mu_j - \mu_{j'}). \quad (5)$$

For 3D regularization, we use the 26-voxel neighborhood N_j surrounding voxel j. The weights $\omega_{j,j'}$ control the relative contribution of each neighbour and are chosen as the inverse distances between voxels j and j'. The potential function $\psi(t)$ is a symmetric convex function that penalizes the difference between the values of neighboring voxels. For computational simplicity, we use a modified potential function used by Lange [3],

$$\psi(t) \triangleq \delta^2\left[\left|\frac{t}{\delta}\right| - \ln\left(1 + \left|\frac{t}{\delta}\right|\right)\right], \quad (6)$$

where δ is a parameter that controls the transition between a quadratic region (for smaller $\left|\frac{t}{\delta}\right|$) and a linear region (for larger $\left|\frac{t}{\delta}\right|$). For our specific reconstruction, we exclude a few image slices from the beginning and end in the penalty calculation because those slices will have severe artifacts due to cone beam truncation. Calculating the penalty for those slices could negatively impact reconstruction of the inner slices since the artifacts do not have any type of structure that can meaningfully be penalized by $R(\mu)$. The overall problem is then to find the penalized likelihood estimate

$$\mu_{PML}^* = \underset{\mu \geq 0}{\mathrm{argmin}}\, I[d||g(\mu)] + \lambda R(\mu). \quad (7)$$

This approach is also called penalized maximum likelihood estimation. It is worth noting that (3) is a special case of (7) when $\lambda = 0$.

Derivation of the Surrogate Function for the Regularized AM Algorithm

In this section, we present the derivation of the surrogate function for an AM algorithm with a Huber-type penalty term. First, we start with a nonnegative initial image, μ_j^0, where the superscript represents the iteration index, and create surrogate functions for both the I-divergence and penalty term at each iteration and update the image by minimizing the surrogate functions. Special properties of the surrogate function guarantee a monotonic decrease of the original function, which will be explained later in this section.

We define our data term as the I-divergence between the data d_i and the estimated mean $g_i(\mu)$. The I-divergence has some terms that depend on the data only, which do not affect the minimization problem. We represent those data-dependent terms by $constant(i)$.

Assuming $\beta_i = 0$, we have

$$I[d||g;\mu] \triangleq \sum_i d_i \sum_j h_{ij}\mu_j + \sum_i I_{0,i}\exp\left(-\sum_j h_{ij}\mu_j\right) + constant(i). \quad (8)$$

Focusing on the terms that include μ, we construct the surrogate function as

$$I[d||g;\mu] \triangleq \sum_i d_i \sum_j h_{ij}\mu_j + \sum_i I_{0,i}\exp\left(-\sum_j h_{ij}\mu_j + \sum_j h_{ij}\hat{\mu}_j - \sum_j h_{ij}\hat{\mu}_j\right), \quad (9)$$

$$= \sum_j \mu_j \sum_i h_{ij}d_i + \sum_i I_{0,i}\exp\left(-\sum_j h_{ij}\hat{\mu}_j\right)\exp\left(-\sum_j h_{ij}\left(\mu_j - \hat{\mu}_j\right)\right). \quad (10)$$

We define the exponentiated forward projection of image estimate as

$$\hat{q}_i = I_{0,i}\exp\left(-\sum_j h_{ij}\hat{\mu}_j\right). \quad (11)$$

The backprojection of $\hat{q}_i$ is defined as

$$\hat{b}_j = \sum_i \hat{q}_i h_{ij}, \quad (12)$$

and the backprojection of the data is defined as

$$\tilde{b}_j = \sum_i d_i h_{ij} \, . \tag{13}$$

Therefore, the I-divergence term equals

$$I[d||g;\mu] = \sum_j \mu_j \, \tilde{b}_j \; + \; \sum_i \hat{q}_i \exp\left(-\sum_j h_{ij}\left(\mu_j - \hat{\mu}_j\right)\right). \tag{14}$$

Using the convex decomposition lemma, we can derive the inequality

$$I[d||g;\mu] \leq \sum_j \mu_j \, \tilde{b}_j \; + \; \sum_i \hat{q}_i \sum_j r_{ij} \exp\left(-\frac{h_{ij}}{r_{ij}}\left(\mu_j - \hat{\mu}_j\right)\right), \tag{15}$$

where

$$r_{ij} \geq 0, \forall \, i, j \tag{16}$$

$$\sum_j r_{ij} \leq 1 \; \forall \, i. \tag{17}$$

If we choose

$$r_{ij} = \frac{h_{ij}}{Z}, \forall \, i, j \tag{18}$$

where Z is defined as

$$Z = \max_i \sum_j h_{ij} \, , \tag{19}$$

then we can satisfy the conditions denoted by equations (16) and (17). Finally, we define the surrogate function of the data fit term $\hat{I}[d||g;\mu,\hat{\mu}]$ using equations (12) (15) and (18) as

$$\hat{I}[d||g;\mu,\hat{\mu}] = \sum_j \mu_j \, \tilde{b}_j \; + \; \sum_i \hat{q}_i \sum_j \frac{h_{ij}}{Z} \exp\left(-Z\left(\mu_j - \hat{\mu}_j\right)\right) \tag{20}$$

$$= \sum_j \mu_j \, \tilde{b}_j \; + \; \frac{1}{Z}\sum_j \left(\sum_i \hat{q}_i \, h_{ij}\right) \exp\left(-Z\left(\mu_j - \hat{\mu}_j\right)\right) \tag{21}$$

$$= \sum_j \mu_j \, \tilde{b}_j \; + \; \frac{1}{Z}\sum_j \hat{b}_j \exp\left(-Z\left(\mu_j - \hat{\mu}_j\right)\right). \tag{22}$$

This surrogate function has the following two majorization properties

$$I[d||g;\mu] \; = \; \hat{I}[d||g;\mu,\mu] \; \forall \, \mu, \tag{23}$$

and

$$I[d||g;\mu] \; \leq \; \hat{I}[d||g;\mu,\hat{\mu}] \; \forall \, \mu,\hat{\mu} \, . \tag{24}$$

Using the properties from equations (23) and (24), we have

$$I[d||g;\hat{\mu}] - I[d||g;\mu] \; \geq \; \hat{I}[d||g;\hat{\mu},\hat{\mu}] \; - \; \hat{I}[d||g;\mu,\hat{\mu}]. \tag{25}$$

In other words, if one can find some μ that makes the right-hand side of (25) positive (some μ that decreases the surrogate function value), then the original objective function also decreases. This is the key idea for forming iterative algorithms using any kind of surrogate function, including the Jensen type for our case. With a proper choice of r_{ij}, the surrogate can be "decoupled"; in other words, minimizing $\hat{I}[d||g;\mu,\hat{\mu}]$ can become N one-dimensional

independent convex minimization problems (one for each μ_j), which are easy to parallelize.

Using the convex decomposition lemma again we can write the potential function as

$$\psi(\mu_j - \mu_{j'}) = \psi\left\{\alpha\left[\frac{1}{\alpha}\left(\mu_j - \hat{\mu}_j\right) + \left(\hat{\mu}_j - \hat{\mu}_{j'}\right)\right] + \right.$$
$$\left. (1-\alpha)\left[\frac{-1}{(1-\alpha)}\left(\mu_{j'} - \hat{\mu}_{j'}\right) + \left(\hat{\mu}_j - \hat{\mu}_{j'}\right)\right]\right\} \; \forall \, j \tag{26}$$

$$\leq \alpha\psi\left[\frac{1}{\alpha}\left(\mu_j - \hat{\mu}_j\right) + \left(\hat{\mu}_j - \hat{\mu}_{j'}\right)\right] +$$
$$(1-\alpha)\psi\left[\frac{-1}{(1-\alpha)}\left(\mu_{j'} - \hat{\mu}_{j'}\right) + \left(\hat{\mu}_j - \hat{\mu}_{j'}\right)\right] \; \forall \, j. \tag{27}$$

To simplify equation (27), we choose $\alpha \triangleq 1/2$ to obtain

$$\psi(\mu_j - \mu_{j'}) \leq \frac{1}{2}\psi\left[2\left(\mu_j - \hat{\mu}_j\right) + \left(\hat{\mu}_j - \hat{\mu}_{j'}\right)\right] +$$
$$\frac{1}{2}\psi\left[-2\left(\mu_{j'} - \hat{\mu}_{j'}\right) + \left(\hat{\mu}_j - \hat{\mu}_{j'}\right)\right] \tag{28}$$

$$= \frac{1}{2}\psi\left[2\mu_j - \hat{\mu}_j - \hat{\mu}_{j'}\right] \; + \frac{1}{2}\psi\left[2\mu_{j'} - \hat{\mu}_j - \hat{\mu}_{j'}\right]. \tag{29}$$

We substitute the potential function in (6) into the surrogate function in (29) and ignore the part of (29) that is independent of μ_j to obtain the modified penalty function, $\hat{R}(\mu)$, given by

$$\hat{R}(\mu,\hat{\mu}) = \sum_j \sum_{j' \in N_j} \frac{\omega_{jj'}}{2} \delta^2 \left(\left|\frac{2\mu_j - \hat{\mu}_j - \hat{\mu}_{j'}}{\delta}\right| - \log\left(1 + \right.\right.$$
$$\left.\left. \left|\frac{2\mu_j - \hat{\mu}_j - \hat{\mu}_{j'}}{\delta}\right|\right)\right). \tag{30}$$

A necessary condition for the solution of the penalized likelihood function is given by

$$\frac{\partial \hat{I}[d||g;\mu,\hat{\mu}]}{\partial \mu_j} \; + \lambda \frac{\partial \hat{R}(\mu)}{\partial \mu_j} = 0 \; \forall \, j \, . \tag{31}$$

From equation (22) we can derive the derivative of the I-divergence terms as

$$\frac{\partial \hat{I}[d||g;\mu,\hat{\mu}]}{\partial \mu_j} = \tilde{b}_j \; - \; \hat{b}_j \exp\left(-Z\left(\mu_j - \hat{\mu}_j\right)\right) \; \forall \, j \, . \tag{32}$$

If we choose $\lambda = 0$, we can derive a closed form solution of the unpenalized likelihood estimate as

$$\mu_j = \hat{\mu}_j \; + \; \frac{1}{Z}\log\frac{\hat{b}_j}{\tilde{b}_j} \; \forall \, j \, . \tag{33}$$

Implementation of the Regularized AM Algorithm
The decoupling steps provide an iterative algorithm that is guaranteed not to increase. Also, it creates many one-parameter convex functions (one for each voxel) that can be minimized in parallel using GPU threads. The pseudocode for the regularized AM Algorithm is shown in Figure 1.

Figure 1. Regularized AM Algorithm

Acceleration Methods

Ordered Subsets

A method called ordered subsets is widely used to increase the convergence speed by using a subset of data at each sub-iteration. The subsets are constructed to be balanced, disjoint, and exhaustive. If the data are partitioned into L subsets, at sub-iteration l a surrogate function for the data-fitting term with only data indices in the corresponding subset is created and minimized with a proportional regularization term. Since the original data-fitting term for which we create surrogate functions changes at each iteration, convergence is not guaranteed. Denoting all source-detector pairs as $\mathbb{Y}$ and source-detector pairs in subset l as $\mathbb{Y}_l$ for $l = 0,1, …, (L-1)$, the regularized ordered subsets algorithm (OS-AM) is presented in Figure 2.

Adaptive Auxiliary Variable

The AM algorithm in closed form solution yields additive updates for the linear attenuation coefficient values with step sizes that are chosen to guarantee convergence. This guaranteed convergence criterion results in step sizes that are unnecessarily conservative. Therefore, to accelerate the convergence of our algorithm, we will try to choose bigger step sizes using adaptive auxiliary variables Z_j such that $r_{ij} = \frac{h_{ij}}{Z_j}$.

For the derivation of these so-called adaptive auxiliary variables, we start with data fit term surrogate function and ignore the regularization term for simplicity, which yields

$$\hat{I}[d||g;\mu,\hat{\mu}] = \sum_j \mu_j\,\tilde{b}_j + \sum_i I_{0,i}\exp\left(-\sum_j h_{ij}\hat{\mu}_j\right)\sum_j \frac{h_{ij}}{Z_j}\exp\left(-Z_j(\mu_j - \hat{\mu}_j)\right). \quad (34)$$

The derivative of this function with respect to μ_j is

$$\frac{\partial\hat{I}[d||g;\mu,\hat{\mu}]}{\partial\mu_j} = \tilde{b}_j - \sum_i h_{ij}I_{0,i}\exp\left(-\sum_j h_{ij}\hat{\mu}_j\right)\exp\left(-Z_j(\mu_j - \hat{\mu}_j)\right). \quad (36)$$

Figure 2. Regularized AM Algorithm with ordered subsets

If our current estimate of $\hat{\mu}_j$ at the k-th iteration is $\hat{\mu}_j^k$ and we allow Z_j to vary with iteration, denoted by Z_j^k, then setting

$$\left.\frac{\partial\hat{I}[d||g;\mu,\hat{\mu}]}{\partial\mu_j}\right|_{\mu_j = \hat{\mu}_j^k} = 0 \quad (36)$$

gives

$$Z_j^k = \frac{\log\left(\frac{\sum_i h_{ij}I_{0,i}\exp\left(-\sum_j h_{ij}\hat{\mu}_j\right)}{\tilde{b}_j}\right)}{\hat{\mu}_j^k - \hat{\mu}_j} \quad \forall j. \quad (37)$$

Since we are minimizing the surrogate function around $\hat{\mu}_j$, any nonnegative value for this variable can be used. The inverse of the auxiliary variable acts as the weight in the closed form update. So, reducing the value of Z_j^k accelerates the convergence of the algorithm. One such choice would be to make $\hat{\mu}_j = 0 \,\forall j$, which results in

$$Z_j^k = \frac{\log\left(\frac{\sum_i h_{ij}I_{0,i}}{\tilde{b}_j}\right)}{\hat{\mu}_j^k} \quad \forall j. \quad (38)$$

The backprojection of the incident photon intensity is

$$\tilde{b}_{0,j} = \sum_i h_{ij}I_{0,i}. \quad (39)$$

Then the adaptive auxiliary variable is

$$Z_j^k = \frac{\log\left(\frac{\tilde{b}_{0,j}}{\tilde{b}_j}\right)}{\hat{\mu}_j^k} \ \forall j \ . \tag{40}$$

In words, we can express our auxiliary variable as

Current estimate of auxiliary variable =
$$\frac{\log\left(\frac{Backprojection\ of\ airscan\ sinogram}{Backprojection\ of\ measured\ data\ sinogram}\right)}{Current\ estimate\ of\ image\ (in\ mm^{-1})} \ . \tag{41}$$

According to our previous choice of Z from equation (19), we can use the reconstruction diameter as a threshold for our proposed adaptive auxiliary variable Z_j^k, according to

$$Z_j^k =
\begin{cases}
\frac{\log\left(\frac{\tilde{b}_{0,j}}{\tilde{b}_j}\right)}{\hat{\mu}_j^k} & \text{if } \frac{\log\left(\frac{\tilde{b}_{0,j}}{\tilde{b}_j}\right)}{\hat{\mu}_j^k} < 2R_{recon}, \frac{\tilde{b}_{0,j}}{\tilde{b}_j} > 1, \hat{\mu}_j^k > 0 \text{ and } k \geq 1 \\
2 \cdot R_{recon}, & \text{else.}
\end{cases} \tag{42}$$

Since we start our iterative algorithm with an initial image estimate derived from a linear reconstruction algorithm like filtered backprojection (FBP) [23] or Feldkamp-Davis-Kress (FDK) [24], we can use this initial image estimate to precompute the initial values of the auxiliary variable. The value of Z_j^{FDK} is shown in Figures 4c, 4d, and 4f for reconstructed data from a Siemens Sensation 16 scanner and a simulated NCAT phantom respectively, where $\hat{\mu}_j^0 = \hat{\mu}_j^{FDK}$. The region of the image with larger attenuation coefficients show a smaller value of the auxiliary variable, which in turn results in larger update steps and vice-versa. However, the requirement (17) in the derivation of the surrogate function of the AM algorithm which guarantees convergence is not satisfied as is evident in Figure 5. The histogram distribution of $\sum_j r_{ij} = \sum_j \frac{h_{ij}}{Z_j}$, which is the forward projection of the inverse auxiliary variable $1/Z_j$, shows values larger than 1 for our adaptive variable scheme. So, the loss of guaranteed convergence is one of the compromises of our proposed acceleration method.

Also, it is evident from equation (41) that both backprojection arrays $\tilde{b}_{0,j}$ and $\tilde{b}_j$ can be precomputed. So, the adaptive nature of the auxiliary variable comes from the fact that after each iteration the value of the variable is updated with the current estimate of the reconstructed image. For parallel processing units like GPUs, this step doesn't add any significant burden to the overall computation time since the computation of each element of the auxiliary variable is independent of each other and GPU threads can compute all the elements efficiently. The regularized AM algorithm with ordered subset is presented in Figure 3 with the initial image estimate derived from the FDK algorithm.

Data Description
To validate the acceleration of the convergence rate of our algorithm, we used an NCAT phantom and raw sinogram data from a Siemens Sensation 16 scanner. To generate synthetic sinogram from the NCAT phantom image volume, we include Poisson noise in the forward projection data of the phantom image using equation (1). The parameter β_i denotes the mean of the Poisson distributed

number of background events added to the data at source-detector pair i, and is equal to 1% of the measured photon intensity of the i-th source-detector pair. The parameters of the measured data and reconstructed images are shown in Table 1:

No. of views	13920
No. of detector channels	672
No. of detector rows	16
No. of image slices	164
No. of pixels/slice	512x512

Table 1. Parameters of measured data and reconstructed image

Input: $\hat{\mu}_j^{(0,0)} = \hat{\mu}^{FDK} \in \mathbb{R}_+^N, \quad d_i, \quad I_{0,i} \in \mathbb{R}_+^M, \quad \lambda \geq 0, \delta > 0, \mathbb{Y}_l \text{ for } l = 0,1,2,\dots(L-1).$

Precompute $\tilde{b}_j^l = \sum_{i \in \mathbb{Y}_l} d_i h_{ij}, \forall l \text{ and } j$

Precompute $\tilde{b}_j = \sum_i d_i h_{ij}, \forall j$

Precompute $\tilde{b}_{0,j} = \sum_i d_i I_{0,i}, \forall j$

Precompute

$Z_j^0 = \begin{cases} \frac{\log\left(\frac{\tilde{b}_{0,j}}{\tilde{b}_j}\right)}{\hat{\mu}_j^{FDK}}, & \text{if } \frac{\log\left(\frac{\tilde{b}_{0,j}}{\tilde{b}_j}\right)}{\hat{\mu}_j^{FDK}} < 2*R_{recon}, \frac{\tilde{b}_{0,j}}{\tilde{b}_j} > 1, \hat{\mu}_j^{FDK} > 0 \\ 2*R_{recon}, & \text{else} \end{cases}$
$\forall j$

for k=0,1,2, …. **do**

    for $l = 0,1,2,\dots(L-1)$ **do**

    $\hat{q}_i^{(k,l)} = I_{0,i} \exp\left(-\sum_j h_{ij}\hat{\mu}_j^{(k,l)}\right), \forall i \in \mathbb{Y}_l$

    $\hat{b}_j^{(k,l)} = \sum_i h_{ij} \hat{q}_i^{(k,l)} \ \forall j$

    $\hat{\mu}_j^{(k,l+1)} = \underset{\mu_j \geq 0}{\operatorname{argmin}} \ \tilde{b}_j^l \left(\mu_j - \hat{\mu}_j^{(k,l)}\right) + \frac{\hat{b}_j^{(k,l)}}{Z_j^k} \exp\left(-Z_j^k\left(\mu_j - \hat{\mu}_j^{(k,l)}\right)\right) +$

    $\frac{\lambda}{L} \sum_{j' \in N_j} \frac{\omega_{jj'}}{2} \delta^2 \left(\left|\frac{2\mu_j - \hat{\mu}_j^{(k,l)} - \hat{\mu}_{j'}^{(k,l)}}{\delta}\right| - \log\left(1 + \left|\frac{2\mu_j - \hat{\mu}_j^{(k,l)} - \hat{\mu}_{j'}^{(k,l)}}{\delta}\right|\right)\right), \forall j$

    end for

    $\hat{\mu}_j^{(k+1,0)} = \hat{\mu}_j^{(k,L)} \ \forall j$

    $Z_j^{k+1} = \begin{cases} \frac{\log\left(\frac{\tilde{b}_{0,j}}{\tilde{b}_j}\right)}{\hat{\mu}_j^{(k+1,0)}}, & \text{if } \frac{\log\left(\frac{\tilde{b}_{0,j}}{\tilde{b}_j}\right)}{\hat{\mu}_j^{(k+1,0)}} < 2*R_{recon}, \frac{\tilde{b}_{0,j}}{\tilde{b}_j} > 1, \hat{\mu}_j^{(k+1,0)} > 0 \\ 2*R_{recon}, & \text{else} \end{cases}$
$\forall j$

end for

Figure 3. Regularized OS-AM Algorithm with adaptive auxiliary variable

Results
To quantify the accuracy of the reconstructed images, the quantity discussed below was measured on the reconstructed images. In the following definition, N denotes the total number of voxels in our region of interest (ROI), $\hat{\mu}_j^k$ is the reconstructed image after k iterations, and $\hat{\mu}_j^{true}$ is the phantom image from which the synthetic

projection data were generated. The percent absolute error (PAE) is defined as

$$\text{PAE} = 100 \times \frac{1}{N} \sum_{j=1}^{N} \left| \frac{\hat{\mu}_j^k}{\hat{\mu}_j^{true}} - 1 \right|. \tag{43}$$

However, for real data there is no true image that can be used to calculate PAE. Instead, we use the total value of the objective function from equation (7) as our performance measure.

For both PAE and I-divergence values, the rate of decrease doubles with the addition of an adaptive auxiliary variable for both baseline and ordered subset AM implementation as evident in Figure 6. For larger numbers of ordered subsets, the rate of acceleration is also larger. However, after a certain number of initial iterations, the straightforward implementation catches up with the adaptive auxiliary variable implementation and eventually surpasses it, possibly due to the loss of the convergence guarantee because of our aggressive update steps. After extensive experimentation with different datasets, scanner geometries and ordered subset configurations, we have repeatedly observed an average of two times acceleration in convergence rate for initial iterations but the rate of acceleration diminishes as we approach convergence.

Conclusions
In this paper, we have proposed a novel approach to adaptively compute the additive step in the AM algorithm. We have observed that our approach of using an adaptive auxiliary variable combined with ordered subsets creates no extra computation cost compared to the straightforward implementation of the OS-AM algorithm. From the image quality assessment parameters, we can conclude that our proposed adaptive auxiliary variable technique shows an average of 2X increase in convergence rate for every OS configuration.

Acknowledgments

References
[1] H. Erdogan and J. A. Fessler, "Ordered subsets algorithms for transmission tomography," *Physics in Medicine and Biology,* vol. 44, no. 11, p. 2835, 1999.

[2] J. A. O'Sullivan and J. Benac, "Alternating minimization algorithms for transmission tomography," *IEEE Transactions on Medical Imaging,* vol. 26, no. 3, pp. 283-297, 2007.

[3] K. Lange, "Convergence of EM image reconstruction algorithms with Gibbs smoothing," *IEEE Transactions on Medical Imaging,* vol. 9, no. 4, pp. 439-446, 1990.

[4] A. K. Hara, R. G. Paden, A. C. Silva, J. L. Kujak, H. J. Lawder, and W. Pavlicek, "Iterative reconstruction technique for reducing body radiation dose at CT: feasibility study," *American Journal of Roentgenology,* vol. 193, no. 3, pp. 764-771, 2009.

[5] P. Prakash *et al.*, "Reducing abdominal CT radiation dose with adaptive statistical iterative reconstruction technique," *Investigative Radiology,* vol. 45, no. 4, pp. 202-210, 2010.

[6] D. Jaffray and J. Siewerdsen, "Cone-beam computed tomography with a flat-panel imager: Initial performance characterization," *Medical Physics,* vol. 27, no. 6, pp. 1311-1323, 2000.

[7] L. Feldkamp, L. Davis, and J. Kress, "Practical cone-beam algorithm," *JOSA A,* vol. 1, no. 6, pp. 612-619, 1984.

[8] P. Sukovic and N. H. Clinthorne, "Penalized weighted least-squares image reconstruction for dual energy X-ray transmission tomography," *IEEE Transactions on Medical Imaging,* vol. 19, no. 11, pp. 1075-1081, 2000.

[9] B. De Man, J. Nuyts, P. Dupont, G. Marchal, and P. Suetens, "An iterative maximum-likelihood polychromatic algorithm for CT," *IEEE Transactions on Medical Imaging,* vol. 20, no. 10, pp. 999-1008, 2001.

[10] J. F. Williamson *et al.*, "Prospects for quantitative computed tomography imaging in the presence of foreign metal bodies using statistical image reconstruction," *Medical Physics,* vol. 29, no. 10, pp. 2404-2418, 2002.

[11] I. A. Elbakri and J. A. Fessler, "Segmentation-free statistical image reconstruction for polyenergetic X-ray computed tomography with experimental validation," *Physics in Medicine and Biology,* vol. 48, no. 15, p. 2453, 2003.

[12] J. A. Fessler, "Statistical image reconstruction methods for transmission tomography," *Handbook of Medical Imaging,* vol. 2, pp. 1-70, 2000.

[13] A. Mitra, D. G. Politte, B. R. Whiting, J. F. Williamson, and J. A. O'Sullivan, "Multi-GPU Acceleration of Branchless Distance Driven Projection and Backprojection for Clinical Helical CT," *Journal of Imaging Science and Technology,* vol. 61, no. 1, pp. 10405-1-10405-13, 2017.

[14] A. Mitra, S. Degirmenci, D. G. Politte, and J. A. O'Sullivan, "Fast Parallel GPU Implementation for Clinical Helical CT using Branchless DD," in *GPU Technology Conference*, 2016, p. 6234.

[15] M. Wu and J. A. Fessler, "GPU acceleration of 3D forward and backward projection using separable footprints for X-ray CT image reconstruction," in *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, 2011, vol. 6, p. 021911.

[16] X. Jia *et al.*, "GPU-based fast low-dose cone beam CT reconstruction via total variation," *Journal of X-ray Science and Technology,* vol. 19, no. 2, pp. 139-154, 2011.

[17] D. Schlifske and H. Medeiros, "A fast GPU-based approach to branchless distance-driven projection and back-projection in cone beam CT," in *Proc. SPIE*, 2016, vol. 9783, p. 97832W.

[18] Q. Xu, D. Yang, J. Tan, A. Sawatzky, and M. A. Anastasio, "Accelerated fast iterative shrinkage thresholding algorithms for sparsity-regularized cone-beam CT image reconstruction," *Medical Physics,* vol. 43, no. 4, pp. 1849-1872, 2016.

[19] S. Degirmenci, D. G. Politte, C. Bosch, N. Tricha, and J. A. O'Sullivan, "Acceleration of iterative image reconstruction for x-ray imaging for security applications," in *Computational Imaging*, 2015, p. 94010C.

[20] S. Degirmenci, *A General Framework of Large-Scale Convex Optimization Using Jensen Surrogates and Acceleration Techniques*. Washington University in St. Louis, 2016.

[21] C. Bosch, S. Degirmenci, J. Barlow, A. Mesika, D. G. Politte, and J. A. O'Sullivan, "Optimizing convergence rates of alternating minimization reconstruction algorithms for real-time explosive detection applications," in *SPIE Defense+ Security*, 2016, pp. 98470P-98470P-17: International Society for Optics and Photonics.

[22] M. G. McGaffin and J. A. Fessler, "Alternating dual updates algorithm for X-ray CT reconstruction on the GPU," *IEEE Transactions on Computational Imaging,* vol. 1, no. 3, pp. 186-199, 2015.

[23] L.-T. Chang, "A method for attenuation correction in radionuclide computed tomography," *IEEE Transactions on Nuclear Science,* vol. 25, no. 1, pp. 638-643, 1978.

[24] K. Stierstorfer, A. Rauscher, J. Boese, H. Bruder, S. Schaller, and T. Flohr, "Weighted FBP—a simple approximate 3D FBP algorithm for multislice spiral CT with good dose usage for arbitrary pitch," *Physics in Medicine and Biology,* vol. 49, no. 11, p. 2209, 2004.

Author Biography

Ayan Mitra received his B.E. in Electrical Engineering from the Jadavpur University in 2013. He is currently a PhD candidate in Electrical Engineering at Washington University in St. Louis. His research advisor is Prof. Joseph A. O'Sullivan. His research interests include GPU acceleration, image reconstruction, and deep learning.
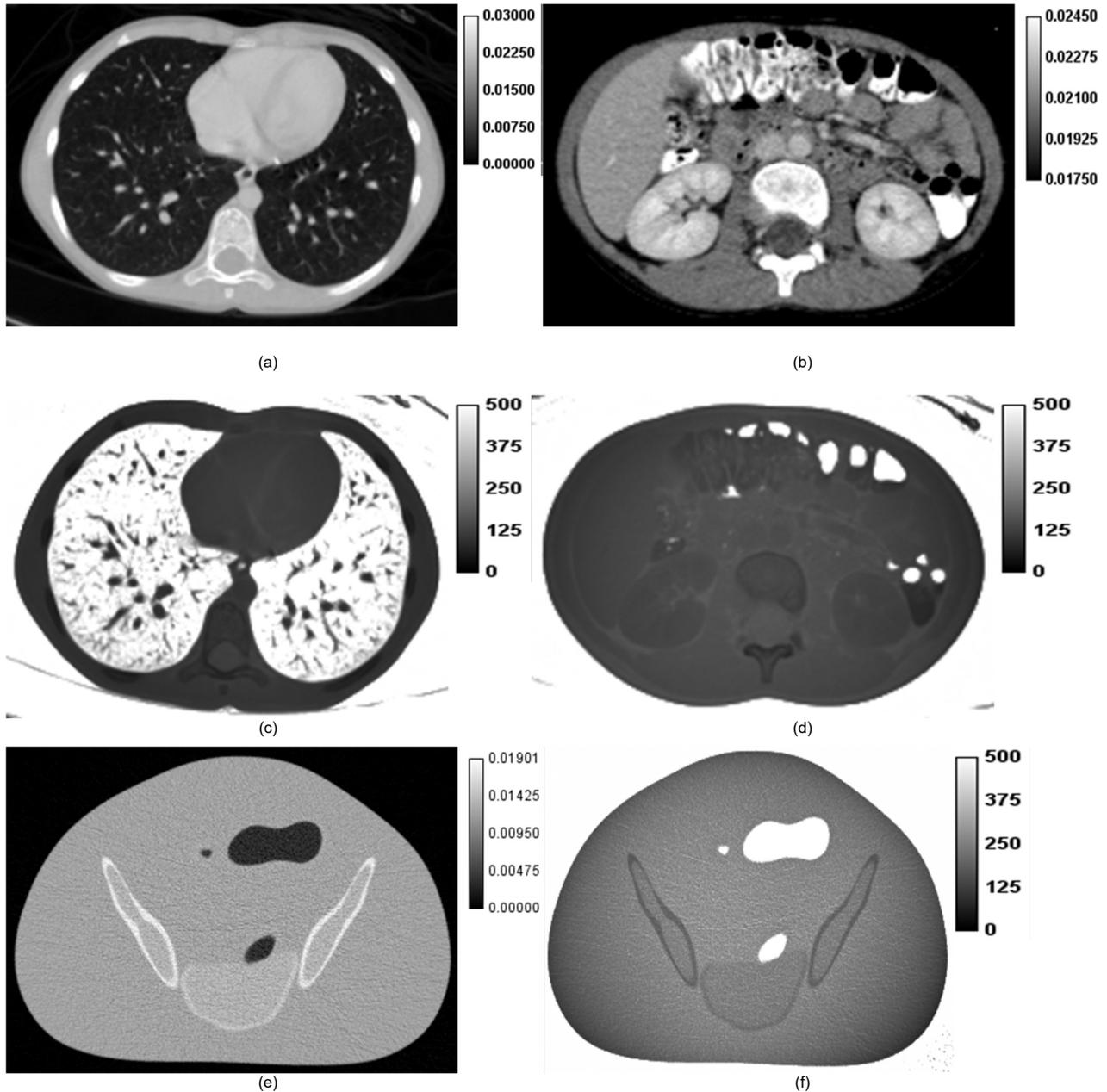
Figure 4 (a) and (b) Linear attenuation coefficient map reconstructed with FDK algorithm for real data obtained from Siemens Sensation 16 scanner, units of mm^{-1}. (c) and (d) Values of the auxiliary variable for the corresponding image slices. (e) Linear attenuation coefficient map reconstructed with FDK algorithm for NCAT data in units of mm^{-1}. (f) Values of the auxiliary variable for the corresponding image slice.
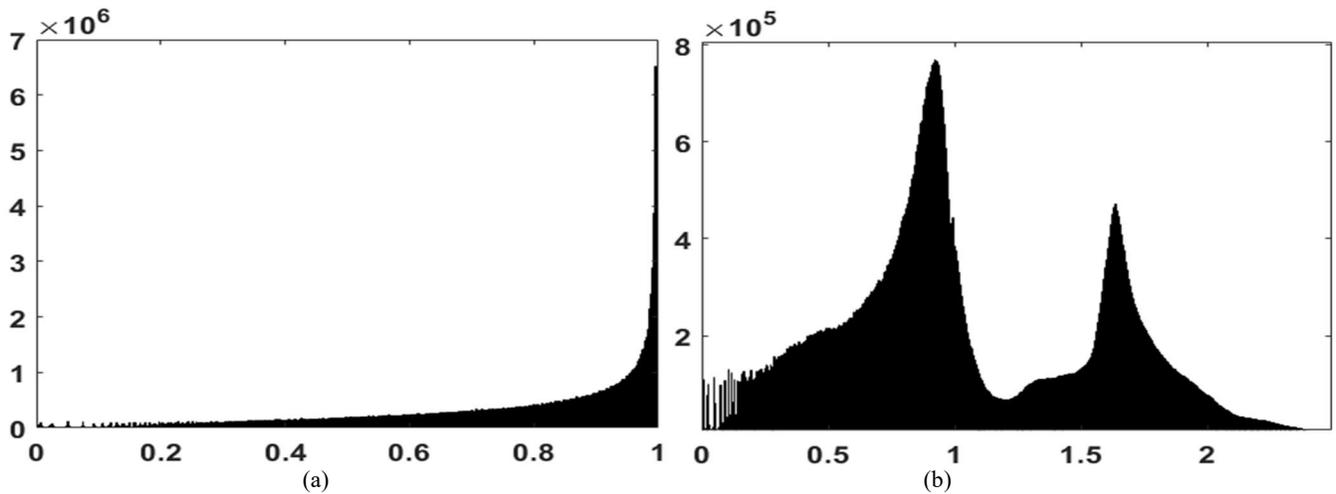
Figure 5. Histogram of the forward projection of the inverse auxiliary variable $\frac{1}{Z_j}$ where (a) Z_j is constant ($Z = \max_i \sum_j h_{ij} = 2 * R_{recon}$) and (b) Z_j is computed by adaptive auxiliary variable.
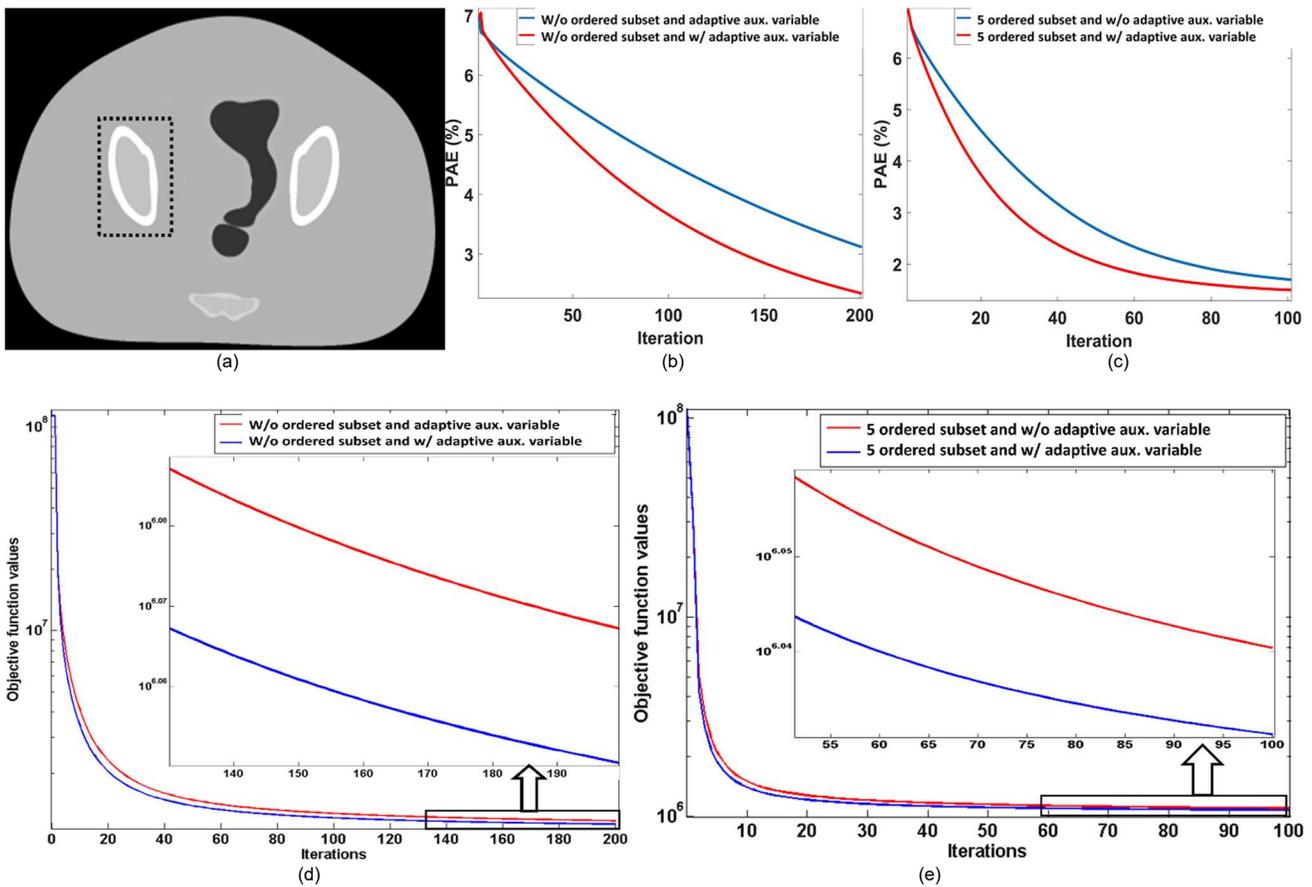
Figure 6. (a) ROI of the NCAT phantom image. PAE vs iteration number for the NCAT phantom in the given ROI (b) without ordered subsets and with (red) and without (blue) the adaptive auxiliary variable (c) with 5 ordered subsets and with (red) and without (blue) the adaptive auxiliary variable. Objective function values vs iteration number for Siemens Sensation 16 scanner reconstructed images (d) without ordered subset implementation and with (blue) and without (red) adaptive auxiliary variable, and (e) with 5 ordered subset implementation and with (blue) and without (red) adaptive auxiliary variable.