# Accelerating ordered-subsets image reconstruction for X-ray CT using double surrogates

Jang Hwan Cho and Jeffrey A. Fessler

Department of Electrical Engineering and Computer Science
University of Michigan, 1301 Beal Ave., Ann Arbor, MI, USA 48109-2122

## ABSTRACT

Conventional ordered-subsets (*OS*) methods for regularized image reconstruction involve computing the gradient of the regularizer for every subset update. When dealing with large problems with many subsets, such as in 3D X-ray CT, computing the gradient for each subset update can be very computationally expensive. To mitigate this issue, some investigators use unregularized iterations followed by a denoising operation after updating all subsets.[1] Although such methods save computation, their convergence properties are uncertain, and since they may not be minimizing any particular cost function it becomes more difficult to design regularization parameters. Furthermore, it is known that inserting filtering steps into unregularized algorithms can lead to undesirable spatial resolution properties.[2] Our goal here is to reduce the computational cost without inducing such problems. We propose a new OS-type algorithm that is derived using optimization transfer principles. The proposed method allows the gradient of the regularizer to be updated less frequently, and thus reduces the computational expense when many subsets are used. Our derivation leads to a correction term that accounts for the fact that the regularizer gradient is updated less frequent than every sub-iteration. Simulations and a phantom experiment show that the proposed method reconstructed images with compatible image quality within reduced computation time.

**Keywords:** ordered-subsets, double surrogates, cone-beam tomography, iterative reconstruction, regularization

## 1. INTRODUCTION

Model-based image reconstruction (*MBIR*) methods, also known as statistical image reconstruction methods, can incorporate accurate system models and take the stochastic characteristics of noise into account, and thus have the potential to improve CT image quality and reduce patient dose compared to conventional methods such as filtered back-projection (*FBP*). However, iterative algorithms for penalized-likelihood (*PL*) image reconstruction require considerable computations, and the computational expense is one of the great challenges for practical utilization of MBIR methods.

Ordered-subsets (*OS*) algorithms, also known as block-iterative or incremental gradient methods, are popular in the field of statistical image reconstruction due to their significant acceleration in initial iterations. The basic idea of OS methods is to group the measurement data into an ordered sequence of subsets or blocks and utilize only one subset of the data for each update instead of using the entire measurements. Unregularized OS reconstruction methods are used routinely for PET and SPECT scans.[3] For regularized image reconstruction problems, conventional OS algorithms calculate the gradient of the regularizer for every subset update.[4] For large problems with large number of subsets, such as cone-beam or helical CT image reconstruction, calculating the gradient for each subset update can be very computationally expensive.

In this paper, we propose a new OS algorithm that is derived from optimization transfer principles and that allows us to compute the gradient of the regularizer less frequently. The proposed method reduces computational cost with little impact on the convergence rate leading to an overall acceleration. We apply the proposed algorithm to penalized weighted least squares (*PWLS*) image reconstruction for cone-beam X-ray computed tomography (*CT*). The method easily could be adapted for other statistical models.

## 2. METHODS

### 2.1 Ordered-Subsets (*OS*) algorithms

For most penalized-likelihood image reconstruction problems, the objective function and its gradient can be rewritten as the following partially separable forms:[5]

$$\Psi(\boldsymbol{x}) = \sum_{m=1}^{M} \Psi_m(\boldsymbol{x}), \quad \nabla_{\boldsymbol{x}} \Psi(\boldsymbol{x}) = \sum_{m=1}^{M} \nabla_{\boldsymbol{x}} \Psi_m(\boldsymbol{x}),$$

where typically $\Psi_m(\boldsymbol{x})$ corresponds to a subset of the projection data. Most iterative algorithms use the gradient of the objective function to obtain a minimizer, and many can be written as the form

$$\boldsymbol{x}^{(n+1)} = \boldsymbol{x}^{(n)} - \alpha_n \boldsymbol{D}_0(\boldsymbol{x}^{(n)}) \nabla_{\boldsymbol{x}} \Psi(\boldsymbol{x}^{(n)}), \tag{1}$$

where $\alpha_n > 0$ is a relaxation parameter and $\boldsymbol{D}_0$ is a scaling matrix, which is typically diagonal. For many algorithms, the empirical findings have suggested that in early iterations, one can replace the gradient of the entire objective function with that of only a part of the objective function and still have $\boldsymbol{x}^{(n+1)}$ be better than $\boldsymbol{x}^{(n)}$ but with less computational cost. Such methods are called incremental gradient methods in the optimization literature.[6,7] The term ordered-subsets has been used in tomography fields, because only a subset of projection views are used in the update steps and the ordering of the views is important.[3,4,8] For $\boldsymbol{x}^{(n)}$ far from the minimizer, $\hat{\boldsymbol{x}}$, if we select the subsets to be "balanced" in some appropriate sense then the following conditions hold:

$$\nabla_{\boldsymbol{x}} \Psi_1(\boldsymbol{x}) \cong \nabla_{\boldsymbol{x}} \Psi_2(\boldsymbol{x}) \cong \cdots \cong \nabla_{\boldsymbol{x}} \Psi_M(\boldsymbol{x}),$$

or equivalently,

$$\nabla_{\boldsymbol{x}} \Psi(\boldsymbol{x}) \approx M \nabla_{\boldsymbol{x}} \Psi_m(\boldsymbol{x}), \quad \forall m. \tag{2}$$

Thus, instead of (1), a typical OS algorithm has the form shown in Table 1.

Table 1. Ordered-subsets Algorithm

| |
| --- |
| Initialize $\boldsymbol{x}^{(0)}$ <br> for $n = 0, 1, \cdots,$ <br><br> $$\boldsymbol{x}^{(n)} := \boldsymbol{x}^{(n,0)}$$ <br><br>    for $m = 1, \cdots, M$ <br><br> $$\boldsymbol{x}^{(n,m)} = \boldsymbol{x}^{(n,m-1)} - \alpha_n M \boldsymbol{D}(\boldsymbol{x}^{(n,m-1)}) \nabla_{\boldsymbol{x}} \Psi_m \Big( \boldsymbol{x}^{(n,m-1)} \Big),$$ <br><br>    end <br><br> $$\boldsymbol{x}^{(n+1)} := \boldsymbol{x}^{(n,M)}$$ <br><br> end |

We refer to each update in Table 1 as the $m$th subset update, or $m$th sub-iteration of the $n$th iteration. One complete iteration is performed by cycling through all the subsets indexed by $m$ so that all data is utilized. In tomography problems, the subsets are selected so that projections within each subset correspond to angularly downsampled projections. It was suggested that the ordering of the subsets that makes projections corresponding to one subset as "orthogonal" as possible to previously used projections is preferable.[9,10] Despite their success in speeding up the initial convergence, ordinary OS algorithms are not convergent in general but rather approach a suboptimal limit-cycle without relaxation, i.e., when $\alpha_n = \alpha$. To address this issue, several families of convergent OS type algorithms have been proposed,[11–14] although those modifications tend to slow down convergence. In this paper, we focus on the initial convergence characteristics of OS algorithms rather than their final convergence properties.

Consider a PWLS objective function of the form

$$\Psi(\boldsymbol{x}) = \mathsf{L}(\boldsymbol{x}) + \mathsf{R}(\boldsymbol{x}), \quad \mathsf{L}(\boldsymbol{x}) = \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_{\boldsymbol{W}}^2, \quad \mathsf{R}(\boldsymbol{x}) = \beta \sum_{k=1}^{K} \psi_k([\boldsymbol{C}\boldsymbol{x}]_k), \tag{3}$$

where $\boldsymbol{A}$ is the system matrix, $\boldsymbol{x}$ is the discretized version of the object being imaged, $\boldsymbol{W} = \mathrm{diag}\{w_i\}$ is a statistical weighting matrix, $\beta$ is the regularization coefficient that controls the resolution-noise tradeoff, $\psi_k$ is the potential function, and $\boldsymbol{C}$ is a matrix performs finite differences between neighboring voxels. The data-fit term can be rewritten as the following separable form:

$$\mathsf{L}(\boldsymbol{x}) = \sum_{m=1}^{M} \mathsf{L}_m(\boldsymbol{x}),$$

where $\mathsf{L}_m$ is the data-fit term associated with the $m$th subset. OS algorithm for PWLS objective function is shown in Table 2. For *separable quadratic surrogates* method, we can define the scaling matrix $\boldsymbol{D}$ as follows:

$$\boldsymbol{D}(\boldsymbol{x}) = [\boldsymbol{D}_L(\boldsymbol{x}) + \boldsymbol{D}_R(\boldsymbol{x})]^{-1}.$$

Typical choices[4] for $\boldsymbol{D}_L$ and $\boldsymbol{D}_R$ are $\boldsymbol{D}_L = \mathrm{diag}(\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}\boldsymbol{1})$ and $\boldsymbol{D}_R(\boldsymbol{x}) = \mathrm{diag}\left(|\boldsymbol{C}|'\mathrm{diag}(|\omega(\boldsymbol{C}\boldsymbol{x})|)\,|\boldsymbol{C}|\,\boldsymbol{1}\right)$ where $\omega(t) = \dot{\psi}(t)/t$ is Huber's curvature.[15] Notice that we need to calculate the regularization gradient, $\boldsymbol{\nabla}_{\boldsymbol{x}}\,\mathsf{R}\left(\boldsymbol{x}^{(n,m-1)}\right)$, for every sub-iteration, which may cause each update to be very computationally expensive when dealing with large problems with many subsets.

Table 2. Ordered-subsets Algorithm for PWLS problem

| |
|---|
| Initialize $\boldsymbol{x}^{(0)}$ <br> for $n = 0, 1, \cdots,$ <br><br> $\qquad\qquad\qquad\qquad \boldsymbol{x}^{(n)} := \boldsymbol{x}^{(n,0)}$ <br><br> $\quad$ for $m = 1, \cdots, M$ <br><br> $\qquad \boldsymbol{x}^{(n,m)} = \boldsymbol{x}^{(n,m-1)} - \alpha_n \boldsymbol{D}(\boldsymbol{x}^{(n,m-1)}) \left( M\boldsymbol{\nabla}_{\boldsymbol{x}}\,\mathsf{L}_m\left(\boldsymbol{x}^{(n,m-1)}\right) + \boldsymbol{\nabla}_{\boldsymbol{x}}\,\mathsf{R}\left(\boldsymbol{x}^{(n,m-1)}\right)\right),$ <br><br> $\quad$ end <br><br> $\qquad\qquad\qquad\qquad \boldsymbol{x}^{(n+1)} := \boldsymbol{x}^{(n,M)}$ <br><br> end |

## 2.2 Ordered-subsets with Double Surrogates

Consider a general PL objective function of the form

$$\Psi(\boldsymbol{x}) = \mathsf{L}(\boldsymbol{x}) + \mathsf{R}(\boldsymbol{x}), \tag{4}$$

where $\mathsf{L}(\boldsymbol{x})$ is the data-fit term and $\mathsf{R}(\boldsymbol{x})$ is the regularizer. We assume that the data-fit term $\mathsf{L}(\boldsymbol{x})$ has a quadratic surrogate of the form:

$$\mathsf{L}(\boldsymbol{x}) \leq \phi_L(\boldsymbol{x}; \tilde{\boldsymbol{x}}) = \mathsf{L}(\tilde{\boldsymbol{x}}) + \nabla\,\mathsf{L}(\tilde{\boldsymbol{x}})(\boldsymbol{x} - \tilde{\boldsymbol{x}}) + \frac{1}{2}(\boldsymbol{x} - \tilde{\boldsymbol{x}})'\boldsymbol{D}_L(\tilde{\boldsymbol{x}})(\boldsymbol{x} - \tilde{\boldsymbol{x}}), \quad \forall \tilde{\boldsymbol{x}},$$

with an appropriate diagonal matrix $\boldsymbol{D}_L$. We also assume the regularizer $\boldsymbol{R}\,(\boldsymbol{x})$ has a quadratic surrogate of the form:

$$\mathsf{R}(\boldsymbol{x}) \leq \phi_R(\boldsymbol{x}; \bar{\boldsymbol{x}}) = \mathsf{R}(\bar{\boldsymbol{x}}) + \nabla\,\mathsf{R}(\bar{\boldsymbol{x}})(\boldsymbol{x} - \bar{\boldsymbol{x}}) + \frac{1}{2}(\boldsymbol{x} - \bar{\boldsymbol{x}})'\boldsymbol{D}_R(\bar{\boldsymbol{x}})(\boldsymbol{x} - \bar{\boldsymbol{x}}), \quad \forall \bar{\boldsymbol{x}},$$

with an appropriate diagonal matrix $\boldsymbol{D}_R$. Then we define the following double-surrogate function:

$$\phi(\boldsymbol{x}; \tilde{\boldsymbol{x}}, \bar{\boldsymbol{x}}) \triangleq \phi_L(\boldsymbol{x}; \tilde{\boldsymbol{x}}) + \phi_R(\boldsymbol{x}; \bar{\boldsymbol{x}}). \tag{5}$$

By construction, this quadratic surrogate has the following properties:

$$\begin{aligned} \Psi(\boldsymbol{x}) &= \phi(\boldsymbol{x}; \boldsymbol{x}, \boldsymbol{x}) \\ \Psi(\boldsymbol{x}) &\leq \phi(\boldsymbol{x}; \tilde{\boldsymbol{x}}, \bar{\boldsymbol{x}}), \quad \forall \tilde{\boldsymbol{x}}, \bar{\boldsymbol{x}}. \end{aligned}$$

These properties generalize those of usual optimization transfer methods.[16,17] For subsequent use, note that from (5):

$$\boldsymbol{\nabla}_{\boldsymbol{x}}\,\phi(\boldsymbol{x}; \tilde{\boldsymbol{x}}, \bar{\boldsymbol{x}}) = \nabla\,\mathsf{L}(\tilde{\boldsymbol{x}}) + \boldsymbol{D}_L(\tilde{\boldsymbol{x}})(\boldsymbol{x} - \tilde{\boldsymbol{x}}) + \nabla\,\mathsf{R}(\bar{\boldsymbol{x}}) + \boldsymbol{D}_R(\bar{\boldsymbol{x}})(\boldsymbol{x} - \bar{\boldsymbol{x}}),$$

so the minimizer of the double surrogate function is given as follows:

$$\arg\min_{\boldsymbol{x}} \phi(\boldsymbol{x}; \tilde{\boldsymbol{x}}, \bar{\boldsymbol{x}}) = [\boldsymbol{D}_L(\tilde{\boldsymbol{x}}) + \boldsymbol{D}_R(\bar{\boldsymbol{x}})]^{-1} \left( \boldsymbol{D}_L(\tilde{\boldsymbol{x}})\tilde{\boldsymbol{x}} + \boldsymbol{D}_R(\bar{\boldsymbol{x}})\bar{\boldsymbol{x}} - \nabla \mathsf{L}(\tilde{\boldsymbol{x}}) - \nabla \mathsf{R}(\bar{\boldsymbol{x}}) \right).$$

A standard optimization transfer algorithm works as follows:

$$\begin{aligned}
\boldsymbol{x}^{(n+1)} &= \arg\min_{\boldsymbol{x}} \phi(\boldsymbol{x}; \boldsymbol{x}^{(n)}, \boldsymbol{x}^{(n)}) \\
&= \boldsymbol{x}^{(n)} - [\boldsymbol{D}_L(\boldsymbol{x}^{(n)}) + \boldsymbol{D}_R(\boldsymbol{x}^{(n)})]^{-1} \left( \nabla \mathsf{L}(\boldsymbol{x}^{(n)}) + \nabla \mathsf{R}(\boldsymbol{x}^{(n)}) \right).
\end{aligned}$$

One can show this also decreases $\Psi$ monotonically, i.e., $\Psi(\boldsymbol{x}^{(n+1)}) \le \Psi(\boldsymbol{x}^{(n)})$. Furthermore, this algorithm converges under suitable conditions.[16] For the case of *separable quadratic surrogates* (*SQS*), usually this type of algorithm converges undesirably slowly.[4] The conventional *ordered-subsets* (*OS*) approach to accelerate convergence is to make the following approximation:

$$\nabla \mathsf{L}(\tilde{\boldsymbol{x}}) \approx \boldsymbol{D}_m \nabla \mathsf{L}_m(\tilde{\boldsymbol{x}}),$$

where $\mathsf{L}_m$ is the data-fit term corresponding to the $m$th subset of the projection views, and $\boldsymbol{D}_m$ is a suitable diagonal matrix, which often simply is $M\boldsymbol{I}$ for $M$ subsets as suggested in (2). Using this approximation, we define the following *approximate* surrogate function:

$$\phi_m(\boldsymbol{x}; \tilde{\boldsymbol{x}}, \bar{\boldsymbol{x}}) \triangleq \mathsf{L}(\tilde{\boldsymbol{x}}) + \boldsymbol{D}_m \nabla \mathsf{L}_m(\tilde{\boldsymbol{x}})(\boldsymbol{x} - \tilde{\boldsymbol{x}}) + \frac{1}{2}(\boldsymbol{x} - \tilde{\boldsymbol{x}})' \boldsymbol{D}_L(\tilde{\boldsymbol{x}})(\boldsymbol{x} - \tilde{\boldsymbol{x}}) + \phi_R(\boldsymbol{x}; \bar{\boldsymbol{x}}).$$

For a conventional regularized ordered-subsets method, the minimization step for each subset is given as follows:

$$\begin{aligned}
\boldsymbol{x}^{(n,m)} &= \arg\min_{\boldsymbol{x}} \phi_m\left(\boldsymbol{x}; \boldsymbol{x}^{(n,m-1)}, \boldsymbol{x}^{(n,m-1)}\right) \\
&= \boldsymbol{x}^{(n,m-1)} - \left[\boldsymbol{D}_L(\boldsymbol{x}^{(n,m-1)}) + \boldsymbol{D}_R(\boldsymbol{x}^{(n,m-1)})\right]^{-1} \left(\boldsymbol{D}_m \nabla \mathsf{L}\left(\boldsymbol{x}^{(n,m-1)}\right) + \nabla \mathsf{R}\left(\boldsymbol{x}^{(n,m-1)}\right)\right), \\
&\qquad \boldsymbol{x}^{(n,0)} := \boldsymbol{x}^{(n)}, \quad \boldsymbol{x}^{(n+1)} := \boldsymbol{x}^{(n,M)},
\end{aligned} \tag{6}$$

for $m = 1, \ldots, M$.

This iteration is undesirably slow because it computes the regularization gradient $\nabla \boldsymbol{R}$ for every subset. To reduce this expense, we propose to exploit the generality of the double surrogate (5) by using the following novel update:

$$\begin{aligned}
\boldsymbol{x}^{(n,m)} &= \arg\min_{\boldsymbol{x}} \phi_m\left(\boldsymbol{x}; \boldsymbol{x}^{(n,m-1)}, \boldsymbol{x}^{(n)}\right) \\
&= \boldsymbol{D}_\phi(\boldsymbol{x}^{(n,m-1)}, \boldsymbol{x}^{(n)}) \left(\boldsymbol{D}_L(\boldsymbol{x}^{(n,m-1)})\boldsymbol{x}^{(n,m-1)} + \boldsymbol{D}_R(\boldsymbol{x}^{(n)})\boldsymbol{x}^{(n)} - \boldsymbol{D}_m \nabla \mathsf{L}\left(\boldsymbol{x}^{(n,m-1)}\right) - \nabla \mathsf{R}(\boldsymbol{x}^{(n)})\right) \\
&= \boldsymbol{x}^{(n,m-1)} - \boldsymbol{D}_\phi(\boldsymbol{x}^{(n,m-1)}, \boldsymbol{x}^{(n)}) \left(\boldsymbol{D}_m \nabla \mathsf{L}\left(\boldsymbol{x}^{(n,m-1)}\right) + \nabla \mathsf{R}(\boldsymbol{x}^{(n)}) + \underbrace{\boldsymbol{D}_R(\boldsymbol{x}^{(n)})(\boldsymbol{x}^{(n,m-1)} - \boldsymbol{x}^{(n)})}_{\text{new term}}\right),
\end{aligned} \tag{7}$$

where $\boldsymbol{D}_\phi(\boldsymbol{x}^{(n,m-1)}, \boldsymbol{x}^{(n)}) = \left[\boldsymbol{D}_L(\boldsymbol{x}^{(n,m-1)}) + \boldsymbol{D}_R(\boldsymbol{x}^{(n)})\right]^{-1}$. This new iteration (7) utilizes the *same* regularizer gradient for all subsets. Compared to (6), the updates in (7) are quite similar except for an extra term that compensates for not updating the regularizer gradient. Table 3 summarizes the proposed algorithm. It requires storing the previous image $\boldsymbol{x}^{\text{last}}$.

In above description, we updated the regularizer gradient only after *all* subsets were updated. Obviously, the regularizer gradient can be updated as frequently as needed and we denote the update frequency as $U_f$ (see algorithm in Table 3). Updating the regularizer gradient less often will reduce the computational cost at the expense of the convergence rate in early iterations.

The proposed algorithm was evaluated on a PWLS image reconstruction problem for cone-beam X-ray CT. We have investigated the trade-off between the update frequency and the computational expense per iteration.

Table 3. General PL Ordered Subsets Method with Double Surrogate

Initialize $\boldsymbol{x}^{(0)}$
for $n = 0, 1, \cdots,$

$$\boldsymbol{x}^{(n)} := \boldsymbol{x}^{(n,0)}, \quad \boldsymbol{x}^{\text{last}} := \boldsymbol{x}^{(n)}$$

   for $m = 1, \cdots, M$
     if $\mathrm{mod}(m - 1, U_f) = 0$

$$\boldsymbol{x}^{\text{last}} = \boldsymbol{x}^{(n,m-1))}$$

   end

$$\boldsymbol{x}^{(n,m)} = \arg\min_{\boldsymbol{x}} \phi_m\big(\boldsymbol{x}; \boldsymbol{x}^{(n,m-1)}, \boldsymbol{x}^{\text{last}}\big)$$

   end

$$\boldsymbol{x}^{(n+1)} := \boldsymbol{x}^{(n,M)}$$

end



      (a)               (b)               (c)
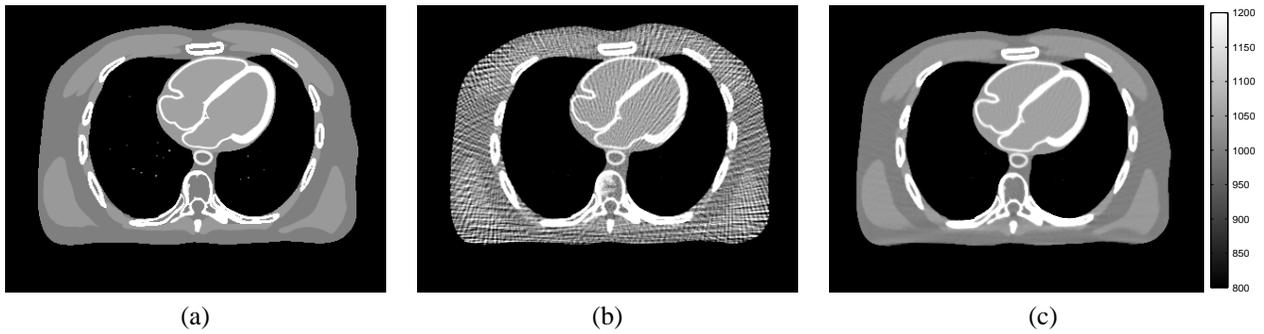
Figure 1. Images of (a) XCAT phantom and (b) FBP reconstruction (c) Converged Image ($\boldsymbol{x}^\infty$) from cone-beam CT data with 164 projection views.

## 3. RESULTS

The proposed algorithm was investigated on a 3D cone-beam CT image reconstruction problem with limited view angles. The XCAT phantom[18] was used as a static object, and the image was reconstructed to a $512 \times 512 \times 50$ grid with pixel size $\Delta_x = \Delta_y = 0.9766$ mm and $\Delta_z = 0.625$ mm.

We simulated a 3rd-generation axial cone-beam CT system using the separable footprint projector.[19] The simulated system has $N_s = 888$ channels and $N_t = 32$ detector rows spaced by $\Delta_s = 1.0239$ mm and $\Delta_t = 1.09878$ mm per view, and 164 evenly spaced view angles over $360°$, which corresponds to an undersampling factor of 6. The source to detector distance was 949 mm, and the source to rotation center distance was 541 mm. We also included a quarter detector offset in the channel direction to reduce aliasing.

For the edge-preserving regularizer, we used a certainty-based penalty[20] to obtain more uniform resolution and a q-Generalized Gaussian MRF (q-GGMRF)[21] as the penalty function to provide edge-preserving regularizer. The regularization parameter $\beta$ was selected such that the target PSF has a full-width at half-maximum (FWHM) of approximately 1.4 mm.[20] We generated the noisy sinogram with Poisson noise, and used weighting $w_i = \exp(-[\boldsymbol{A}\boldsymbol{x}]_i)$.

To assess the convergence speed of the proposed method, we computed the root mean squared difference between the image estimate at the $n$th iteration, $\boldsymbol{x}^{(n)}$, with both the fully converged solution, $\boldsymbol{x}^\infty$, and true image, $\boldsymbol{x}^{\text{true}}$. We calculated $\boldsymbol{x}^\infty$ using consecutive steps of ordered-subsets with decreasing number of subsets. We used 41, 10, and 1 subsets with 100, 100, and 1000 iterations respectively.

This type of "relaxed" OS is guaranteed to converge because the final stage uses just 1 subset for which (6) is convergent.[4] Fig. 1 shows the images of true phantom, Filtered Back Projection (FBP) reconstruction,[22] and fully converged image ($\boldsymbol{x}^\infty$). For the FBP reconstruction, the ramp filter was associated with a Hanning apodization window of 2048-point in length to attenuate the high frequency noise. Due to limited view angles, conventional FBP reconstruction, which was used as the initial condition for our OS reconstruction, shows severe streaking artifacts compared to the true phantom. The fully converged image has much less artifacts, thus illustrating the benefits of statistical image reconstruction in limited view
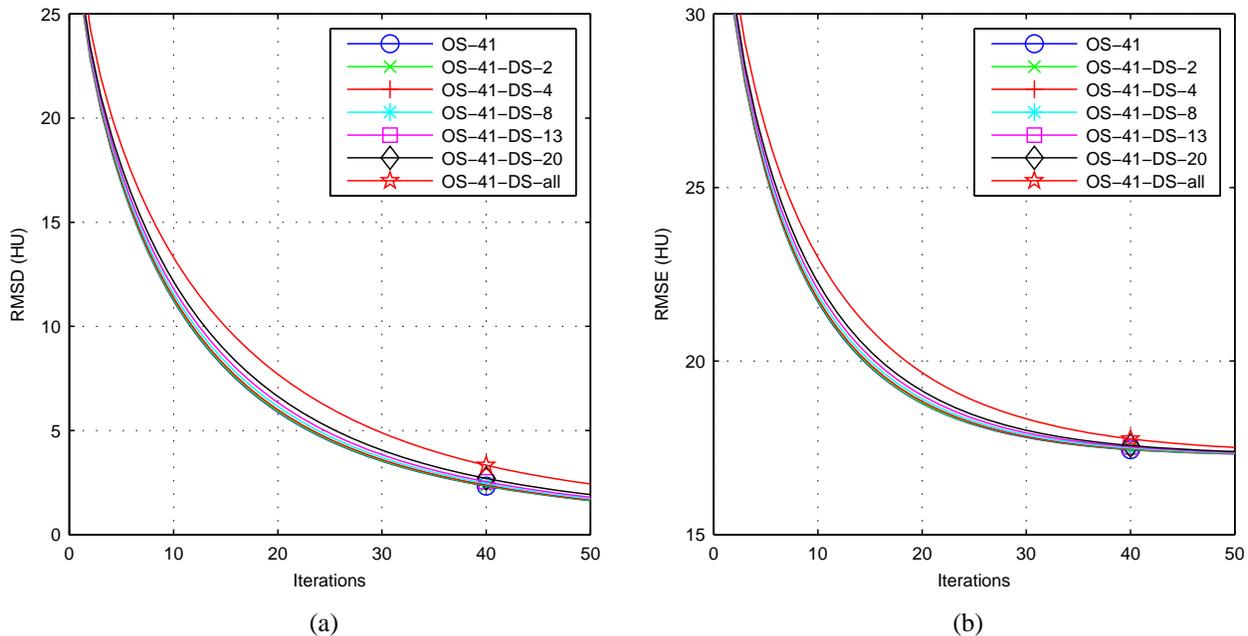
Figure 2. Convergence rates at each iteration for different regularizer update frequency. OS-41-DS-n indicates OS with 41 subsets and $U_f$ = n, and "n = all" means only updating once after all subset updates are done. (a) $\boldsymbol{x}^{(n)}$ with respect to $\boldsymbol{x}^\infty$ and (b) $\boldsymbol{x}^{(n)}$ with respect to $\boldsymbol{x}^{\text{true}}$

angle problems.

We divided the projections into 41 subsets, which corresponds to 4 views per subset. This is a rather aggressive selection compared to conventional choices to try to accelerate convergence significantly.

The regularizer gradient was updated at different frequencies to see its effect to convergence and computation time. Fig. 2 illustrates that our proposed method gives similar root-mean-square (RMS) differences, $\sqrt{\frac{1}{N}\sum_{j=1}^{N}(x_j^{(n)} - x_j^\infty)}$, as the conventional OS even when we update the regularizer infrequently per iteration.

On the other hand, Fig. 3 illustrates that the computational expense required to obtain the same level of RMS differences was reduced by the proposed method. With $U_f = 13$, which gives the best result in this case, the proposed method converges about 3 times faster than the conventional OS. By observing the reconstructed image at the same time point, Fig. 4 clearly shows that the proposed method is converging faster. There exists tradeoff between convergence rates and computational expense, and for our case calculating the regularizer gradient for every 13 subset updates gave the most efficient results. For different problems, the optimal update frequency will differ. However, it is noticeable that regardless of the update frequency, the proposed method is converging faster than the conventional OS. As the problem gets larger and the number of subsets increase, the computational expenses required to calculate the gradient of the regularizer becomes much more dominant. Therefore, we can expect substantial benefits from our proposed method for such problems.

## 4. CONCLUSIONS

We presented a simple modification of the conventional OS method that allows the regularizer gradient to be less frequently updated. The method provides very significant acceleration when applied to large problems with many subsets, such as cone-beam or helical CT image reconstruction, while still providing good reconstructed images.

Simulations demonstrated that a good reconstruction with compatible quality was achieved within much less computation time. However, the reduction in the computational expense may depend on the size of the problem and the complexity of updating the regularizer compared to the complexity of forward and back-projection. When implemented on graphics processing unit (GPU), the trade-offs may differ.

We only focussed on the initial convergence characteristics of the ordered-subsets method. Since the conventional OS algorithm is inherently not globally convergent, neither is the proposed method. We can apply the double surrogate idea to
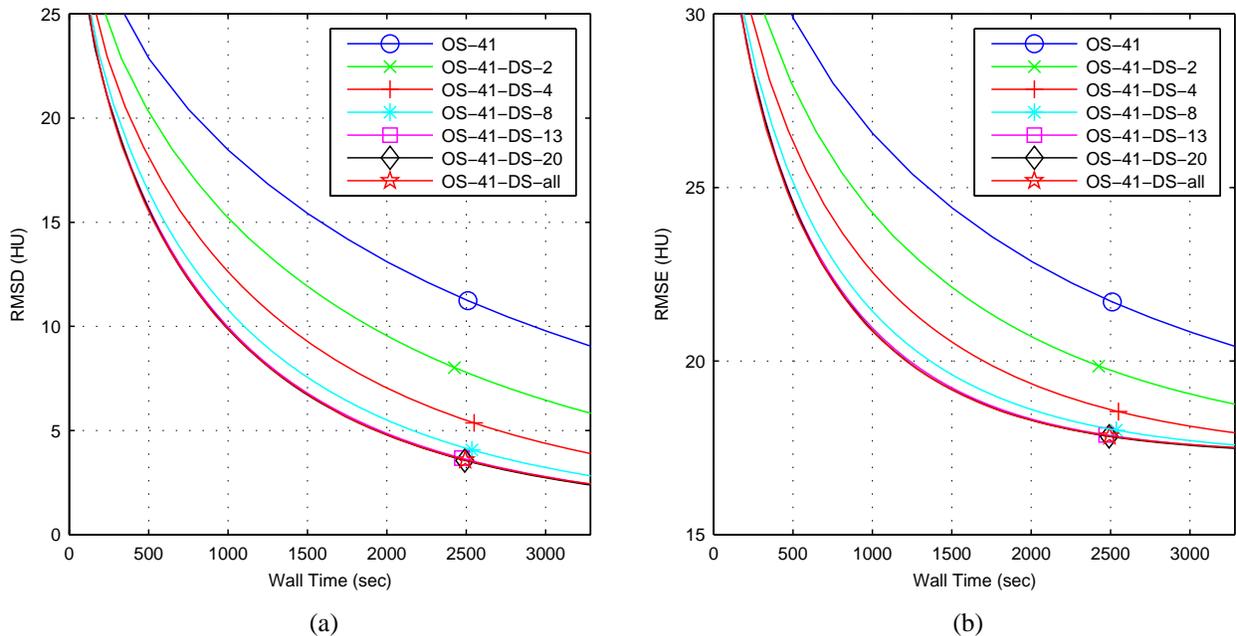
Figure 3. Convergence rates versus time for different regularizer update frequency. (a) $x^{(n)}$ with respect to $x^{\infty}$ and (b) $x^{(n)}$ with respect to $x^{\text{true}}$

other block iterative algorithms like incremental optimization transfer (IOT), which is globally convergent.[23]
Further research will address the systematic way to determine the optimal update frequency and extension of the idea to other algorithms.

## REFERENCES

[1] Xu, W. and Mueller, K., "Evaluating popular non-linear image processing filters for their use in regularized iterative CT," in [*Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*], 2864–5 (2010).

[2] Mustafovic, S. and Thielemans, K., "Object dependency of resolution in reconstruction algorithms with inter-iteration filtering applied to PET data," *IEEE Trans. Med. Imag.* **23**, 433–46 (Apr. 2004).

[3] Hudson, H. M. and Larkin, R. S., "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Trans. Med. Imag.* **13**, 601–9 (Dec. 1994).

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

[5] Nocedal, J., "Large scale unconstrained optimization," in [*The State of the Art in Numerical Analysis*], Watson, A. and Duff, I., eds., Oxford University Press, Oxford (1996).

[6] Kibardin, V. M., "Decomposition into functions in the minimization problem," *Avtomatika i Telemekhanika* **9**, 66–79 (Sept. 1979). Translation: p. 1311-23 in Plenum Publishing Co. "Adaptive Systems".

[7] Bertsekas, D. P., "A new class of incremental gradient methods for least squares problems," *SIAM J. Optim.* **7**, 913–26 (Nov. 1997).

[8] Kamphuis, C. and Beekman, F. J., "Accelerated iterative transmission CT reconstruction using an ordered subsets convex algorithm," *IEEE Trans. Med. Imag.* **17**, 1001–5 (Dec. 1998).

[9] Herman, G. T. and Meyer, L. B., "Algebraic reconstruction techniques can be made computationally efficient," *IEEE Trans. Med. Imag.* **12**, 600–9 (Sept. 1993).

[10] Hamaker, C. and Solmon, D. C., "The angles between the null spaces of X rays," *J. Math. Anal. Applic.* **62**, 1–22 (Jan. 1978).

[11] De Pierro, A. R. and Yamagishi, M. E. B., "Fast EM-like methods for maximum 'a posteriori' estimates in emission tomography," *IEEE Trans. Med. Imag.* **20**, 280–8 (Apr. 2001).

[12] Kudo, H., Nakazawa, H., and Saito, T., "Convergent block-iterative method for general convex cost functions," in [*Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*], 247–250 (1999).

[13] Neal, R. and Hinton, G. E., "A view of the EM algorithm that justifies incremental, sparse and other variants," in [*Learning in Graphical Models*], Jordan, M. I., ed., 255–68, Kluwer, Dordrencht (1998).

[14] Blatt, D., Hero, A. O., and Gauchman, H., "A convergent incremental gradient method with a constant step size," *SIAM J. Optim.* **18**(1), 29–51 (2007).

[15] Huber, P. J., [*Robust statistics*], Wiley, New York (1981).

[16] Jacobson, M. W. and Fessler, J. A., "An expanded theoretical treatment of iteration-dependent majorize-minimize algorithms," *IEEE Trans. Im. Proc.* **16**, 2411–22 (Oct. 2007).

[17] Hunter, D. R. and Lange, K., "A tutorial on MM algorithms," *American Statistician* **58**, 30–7 (Feb. 2004).

[18] Segars, W. P., Mahesh, M., Beck, T. J., Frey, E. C., and Tsui, B. M. W., "Realistic CT simulation using the 4D XCAT phantom," *Med. Phys.* **35**, 3800–8 (Aug. 2008).

[19] Long, Y., Fessler, J. A., and Balter, J. M., "3D forward and back-projection for X-ray CT using separable footprints," *IEEE Trans. Med. Imag.* **29**, 1839–50 (Nov. 2010).

[20] Fessler, J. A. and Rogers, W. L., "Spatial resolution properties of penalized-likelihood image reconstruction methods: Space-invariant tomographs," *IEEE Trans. Im. Proc.* **5**, 1346–58 (Sept. 1996).

[21] Thibault, J.-B., Sauer, K., Bouman, C., and Hsieh, J., "A three-dimensional statistical approach to improved image quality for multi-slice helical CT," *Med. Phys.* **34**, 4526–44 (Nov. 2007).

[22] Feldkamp, L. A., Davis, L. C., and Kress, J. W., "Practical cone beam algorithm," *J. Opt. Soc. Am. A* **1**, 612–9 (June 1984).

[23] Ahn, S., Fessler, J. A., Blatt, D., and Hero, A. O., "Convergent incremental optimization transfer algorithms: Application to tomography," *IEEE Trans. Med. Imag.* **25**, 283–96 (Mar. 2006).
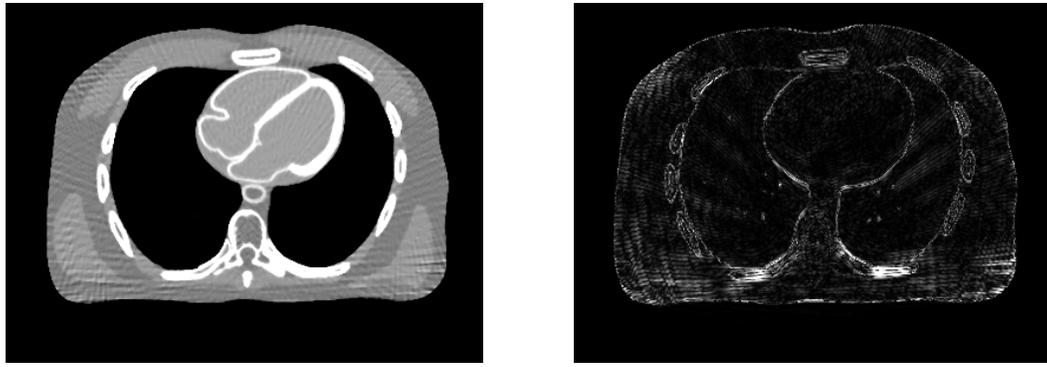
(a) OS-41 (conventional OS)
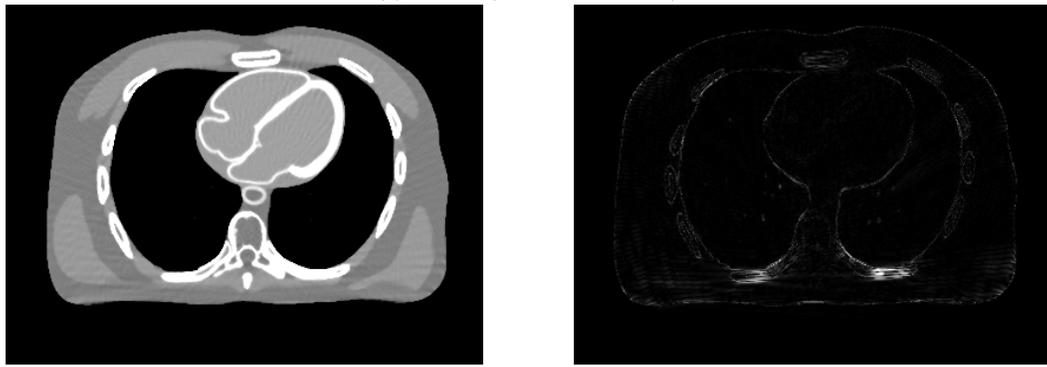


(b) OS-41-DS-4
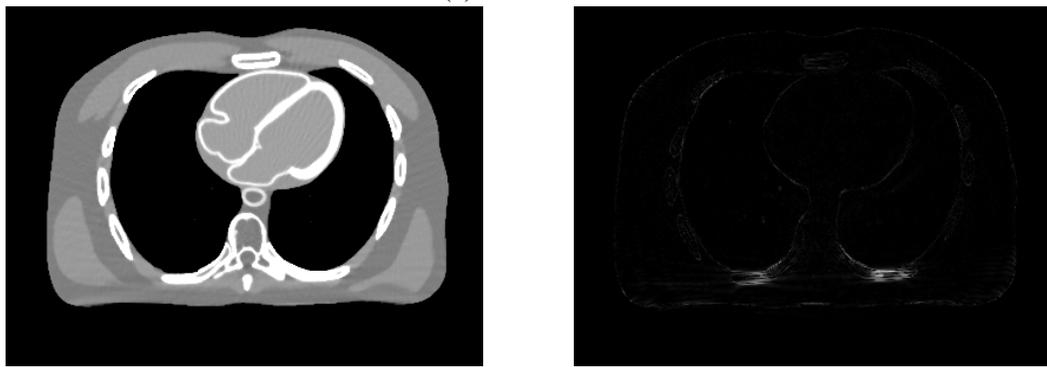


(c) OS-41-DS-13
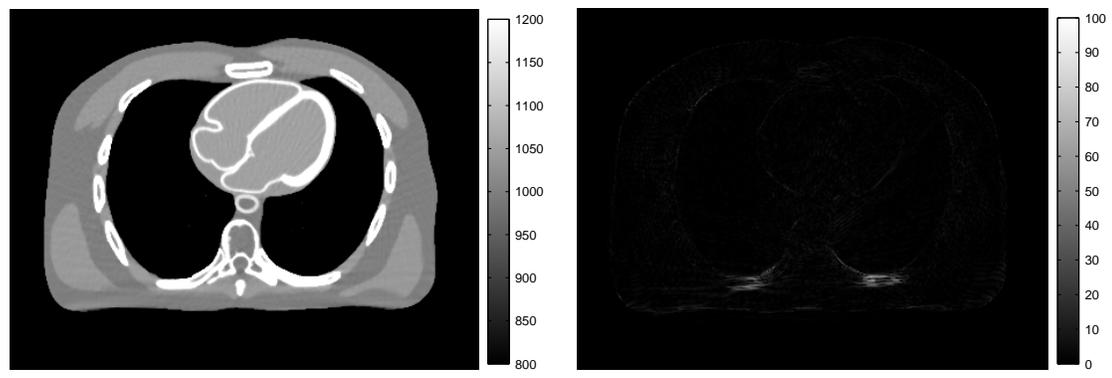


(d) OS-41-DS-all

Figure 4. Comparing the convergence speed of OSDS with different update frequencies. Left column: Images at the same time point (4000 sec after initialization). Right column: Absolute difference images with respect to $x^{(\infty)}$.