ORDERED SUBSETS WITH MOMENTUM FOR ACCELERATED X-RAY CT IMAGE RECONSTRUCTION

Donghwan Kim, Sathish Ramani, and Jeffrey A. Fessler

The University of Michigan
Department of Electrical Engineering and Computer Science
1301 Beal Avenue, Ann Arbor, MI 48109-2122

ABSTRACT

Statistical image reconstruction methods provide improved image quality in low-dose X-ray CT. However, the long computation time of iterative algorithms limits their clinical use. Ordered subsets algorithms based on separable quadratic surrogates (OS-SQS) are attractive as they are simple and amenable for massive parallelization in modern computing architecture, but require many iterations to converge. Here, we further accelerate OS algorithms by using momentum techniques. We use real patient CT scan to illustrate that the proposed algorithms converge rapidly compared to previous OS algorithms.

1. INTRODUCTION

In X-ray CT reconstruction, we reconstruct an image $\hat{x} \in \mathbb{R}^{N_p}$ from noisy measured data $y \in \mathbb{R}^{N_d}$ by minimizing a convex (and differentiable) cost function $\Psi(x)$, possibly with a non-negativity constraint:

$$\hat{x} = \arg\min_{x \succeq 0} \Psi(x). \tag{1}$$

One standard choice of $\Psi(x)$ for X-ray CT reconstruction is a penalized weighted least squares (PWLS) cost function [1]:

$$\Psi(x) = \frac{1}{2}||y - Ax||_W^2 + \beta R(x), \tag{2}$$

where A is a projection operator, a diagonal matrix W provides statistical weighting, $R(x)$ is a (often non-quadratic and usually differentiable) regularization function, and β is a regularization parameter that balances the data-fitting term and the regularizer $R(x)$. Minimizing $\Psi(x)$ requires iterative algorithms that need substantial computation time (especially in 3D CT), and the goal of this work is to describe faster iterative algorithms.

There are many general iterative algorithms, but only some of them are well suited to X-ray CT reconstruction, because of the large-scale of the problem. Particularly, the projection operators A and A' in X-ray CT require significant computation time and thus computing the gradient $\nabla\Psi(x) = A'W(Ax - y) + \beta\nabla R(x)$ of the cost function repeatedly in an iterative algorithm becomes expensive. (Computation of $\nabla R(x)$ is less intense than computing products of the form Ax in CT.) Ordered subsets (OS) algorithms are widely used in X-ray CT research (and are already used in clinical PET and SPECT), because they approximate $\nabla\Psi(x)$ using only a subset of measured data [2, 3]. Even though OS methods involve approximations and only nearly approach $\hat{x}$, they can provide dramatic initial acceleration and are thus useful for large-scale problems such as 3D CT.

OS algorithms are naturally suited to optimization transfer methods. One well-known variation is an OS algorithm based on separable quadratic surrogates (OS-SQS) that is simple and amenable for massive parallelization in modern parallel computing architectures [3]. However, the algorithm requires many iterations to converge. Recently, we developed a non-uniform (NU) optimization transfer that reduced the number of iterations to converge [4]. Here, we further accelerate OS methods by using momentum techniques that provide additional acceleration toward the optimum. For non-OS methods, the fast iterative shrinkage-thresholding algorithm (FISTA) [5] is a representative method for using momentum; FISTA provides $O(1/n^2)$ convergence in $\Psi(x)$ as opposed to $O(1/n)$ rate for conventional optimization transfer methods, where n counts the number of iterations.

In this work, we propose to combine momentum technique from [5] with OS methods for image reconstruction, particularly for optimization transfer methods such as SQS and NU-SQS. We apply the algorithms to X-ray CT reconstruction. We first explain the optimization transfer and OS methods in general and introduce the momentum technique to optimization transfer and to OS methods. Then we investigate the acceleration of OS techniques when combined with momentum on a real patient CT scan. The results show that the proposed algorithms highly accelerate the iterative reconstruction for 3D X-ray CT.

Supported in part by NIH grant R01-HL-098686 and equipment donations from Intel. Thank GE for a real patient data.

2. OPTIMIZATION TRANSFER METHOD

When an objective function $\Psi(x)$ is difficult to minimize, we replace it by a surrogate $\phi^{(n)}(x)$ at nth iteration that is easier to minimize [3]. The basic iteration of optimization transfer is

$$x^{(n+1)} = \arg\min_{x \succeq 0} \phi^{(n)}(x). \tag{3}$$

To monotonically decrease $\Psi(x)$, we design surrogates $\phi^{(n)}(x)$ that satisfy the following conditions:

$$\begin{aligned} \Psi(x^{(n)}) &= \phi^{(n)}(x^{(n)}), \\ \Psi(x) &\le \phi^{(n)}(x), \quad \forall x \in \mathbb{R}^{N_p}, x \succeq 0. \end{aligned} \tag{4}$$

We usually majorize the cost function $\Psi(x)$ by the following quadratic surrogate function $\phi^{(n)}(x)$:

$$\begin{aligned} \Psi(x) \le \phi^{(n)}(x) \triangleq \Psi(x^{(n)}) + \nabla\Psi(x^{(n)})'(x - x^{(n)}) \\ + \frac{1}{2}(x - x^{(n)})' D (x - x^{(n)}), \end{aligned} \tag{5}$$

which satisfies the condition (4). The majorizing matrix D can simply be an identity matrix scaled by the Lipschitz constant of $\Psi(x)$ [5] or a diagonal matrix derived by (NU-)SQS algorithm [3, 4], or can be generated by other methods that yield a matrix that is easier to invert than the Hessian of $\Psi(x)$. The smallest possible Lipschitz constant of the cost function $\Psi(x)$ in (2) is not easily computable, and thus it is usually preferable to use (NU-)SQS algorithms in X-ray CT.

Then the iteration (3) turns out to be

$$x^{(n+1)} = \left[x^{(n)} - D^{-1} \nabla\Psi(x^{(n)}) \right]_+, \tag{6}$$

where a clipping $[\cdot]_+$ enforces the non-negativity constraint. The optimization transfer method with a diagonal majorizer D has convergence rate $O(1/n)$:

Theorem 1 *For a diagonal majorizer D, the sequence $\left\{x^{(n)}\right\}$ generated by* (6) *satisfies*

$$\Psi(x^{(n)}) - \Psi(\hat{x}) \le \frac{||x^{(0)} - \hat{x}||_D^2}{2n}. \tag{7}$$

This result is a simple generalization of Theorem 3.1 in [5].

Our NU approach [4] accelerated the optimization transfer method by reducing the numerator $||x^{(0)} - \hat{x}||_D^2$ with respect to D in Theorem 1, subject to the condition (5). This helped the algorithm to converge faster but the convergence rate remained $O(1/n)$. Applying momentum techniques to optimization transfer method can provide acceleration and achieve a faster convergence rate $O(1/n^2)$, as detailed in the next section.

3. OPTIMIZATION TRANSFER WITH MOMENTUM

One can accelerate optimization transfer algorithms by introducing a momentum term that uses an image estimate from the previous iteration [5]. Fig. 1 shows optimization transfer method with momentum. In Fig. 1, $x^{(n)}$ is a normal optimization transfer update and $v^{(n)}$ is the momentum update term that depends on previous iteration, which helps the algorithm converge faster. The momentum update $v^{(n)}$ is computationally negligible compared with the $x^{(n)}$ update, so introducing momentum to optimization transfer is very practical. The algorithm in Fig. 1 reduces to the standard optimization transfer method in (6) when $t_n = 1$ for all $n \ge 0$.

Initialize $x^{(0)} = v^{(0)}$, $t_0 = 1$

for $n = 0, 1, 2, \cdots$

$$t_{n+1} = \left(1 + \sqrt{1 + 4t_n^2}\right)/2$$

$$x^{(n+1)} = \left[v^{(n)} - D^{-1} \nabla\Psi(v^{(n)}) \right]_+$$

$$v^{(n+1)} = x^{(n+1)} + \frac{t_n - 1}{t_{n+1}} (x^{(n+1)} - x^{(n)})$$

Fig. 1. Optimization transfer method with momentum

This algorithm has a convergence proof by Theorem 2 for the sequence $\left\{x^{(n)}\right\}$ with convergence rate $O(1/n^2)$:

Theorem 2 *For a diagonal majorizer D, the sequence $\left\{x^{(n)}\right\}$ generated by Fig. 1 satisfies*

$$\Psi(x^{(n)}) - \Psi(\hat{x}) \le \frac{2||x^{(0)} - \hat{x}||_D^2}{(n+1)^2}. \tag{8}$$

Theorem 2 is a simple generation of Theorem 4.4 in [5], which was shown for a surrogate with a scaled identity Hessian (using Lipschitz constant).

We further accelerate the algorithm in Fig. 1 by applying ordered subsets (OS) methods in the next section.

4. PROPOSED ALGORITHM COMBINING ORDERED SUBSETS AND MOMENTUM

4.1. Ordered subsets (OS) algorithm

We first review the standard ordered subsets (OS) approach for accelerating iterative image reconstruction. The cost function $\Psi(x) = \sum_{m=0}^{M-1} \Psi_m(x)$ is a sum of

$$\Psi_m(x) = \frac{1}{2}||y_m - A_m x||_{W_m}^2 + \frac{\beta}{M} R(x), \tag{9}$$

which is a function of mth subset of measurement data, where M is a number of subsets. A_m, y_m and W_m are sub-matrices of A, y and W that correspond to the mth subset of measured data. In an ordered subsets method, we use the approximation

$$\nabla\Psi(x) \approx M\nabla\Psi_0(x) \approx M\nabla\Psi_1(x) \approx \cdots \approx M\nabla\Psi_{M-1}(x), \tag{10}$$

when each subset consists of measurement data (*e.g.*, projection views) approximately uniformly down-sampled by M.

With the assumption (10), we replace $\nabla\Psi(x)$ by $M\nabla\Psi_m(x)$ (or with similar approximations [6]) in optimization transfer method in (6). So, with only $1/M$ of the amount of computation by using a subset of measured data, we can approximate the original gradient of the cost function and have about M times acceleration in early iterations. However, the approximation (10) becomes very inaccurate as the iterates approach the minimizer, and the OS algorithm loses the convergence property.

Optimization transfer with ordered subsets generates a sequence $\left\{x^{(n+\frac{m}{M})}\right\}$:

$$x^{(n+\frac{m+1}{M})} = \left[x^{(n+\frac{m}{M})} - D^{-1}M\nabla\Psi_m(x^{(n+\frac{m}{M})})\right]_+, \quad (11)$$

where each mth sub-iteration is counted as $1/M$ iteration of (6), as (11) uses only $1/M$ amount of computation for the forward and back projection required in (6). But, OS methods with large M can be slow in run time due to the increased computation for $\nabla R(x)$.

4.2. Proposed OS algorithms with momentum

To provide more acceleration, we propose to combine OS algorithms with the momentum technique described in Fig. 1. The outline of proposed OS algorithms with momentum is presented in Fig. 2. The sequence $\left\{x^{(n+\frac{m}{M})}\right\}$ generated by Fig. 2 is expected to have a M^2-times accelerated convergence rate $O(1/(nM+m)^2)$ with the assumption (10).

Like standard OS methods, OS algorithms with momentum are not guaranteed to converge to the optimum $\hat{x}$. The approximation (10) will fail as the iterates reach near the optimum, but the OS aspects of the proposed algorithm will still speed up the convergence initially.

Initialize $x^{(0)} = v^{(0)}$, $t_0 = 1$
for $n = 0, 1, 2, \cdots$
for $m = 0, 1, \cdots, M-1$

$$t_{nM+m+1} = \left(1 + \sqrt{1 + 4t_{nM+m}^2}\right)/2$$

$$x^{(n+\frac{m+1}{M})} = \left[v^{(n+\frac{m}{M})} - D^{-1}M\nabla\Psi_m(v^{(n+\frac{m}{M})})\right]_+$$

$$v^{(n+\frac{m+1}{M})} = x^{(n+\frac{m+1}{M})} + \frac{t_{nM+m}-1}{t_{nM+m+1}}(x^{(n+\frac{m+1}{M})} - x^{(n+\frac{m}{M})})$$

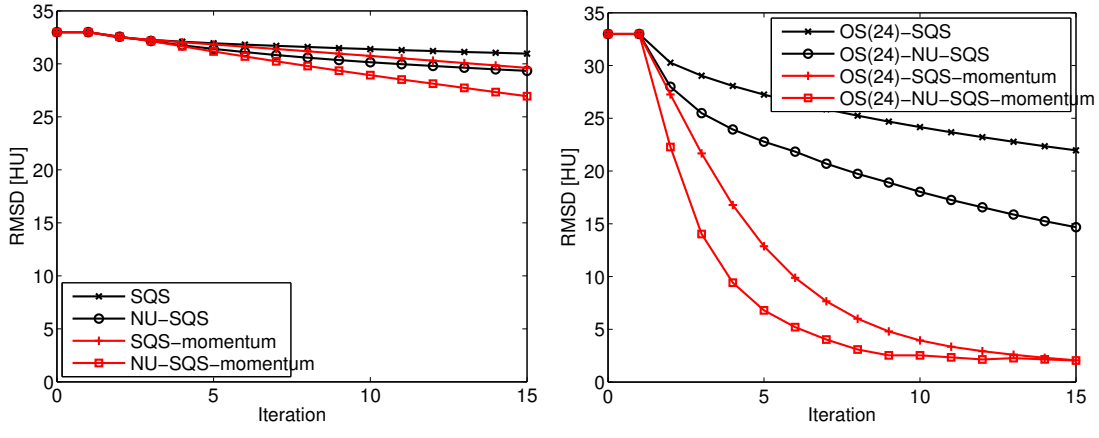
Fig. 2. Optimization transfer method with ordered subsets and momentum

Furthermore, owing to the acceleration provided by the momentum term, it is possible to use fewer subsets (smaller M) than with standard OS methods, so the approximation in (10) becomes more accurate and the limiting behavior of the proposed (OS-momentum) algorithm can be better stabilized compared to conventional OS methods. Fewer subsets also reduces overhead in computing the regularizer. Further theoretical investigation of the behavior of OS with momentum is future work.

5. RESULTS

We examined the convergence rate of the proposed (OS-momentum) algorithms compared with the ordinary OS-SQS algorithm on a 3D helical X-ray CT data set of a human shoulder. Fig. 3 shows the RMSD[1] (root mean square difference) between the current and converged image in Hounsfield Units

[1] RMSD $= ||x^{(n)} - \hat{x}||_2/\sqrt{N_p}$ [HU].



(a) SQS (b) OS-SQS (24 subsets)

Fig. 3. Plots of RMSD versus iterations for the standard (NU-)OS algorithms and the proposed (NU-)OS algorithms with momentum: (a) 1 subset and (b) 24 subsets version. (There are no changes in RMSD during first iteration, since we use one forward and back projections for precomputation of the denominator D before iterating.)

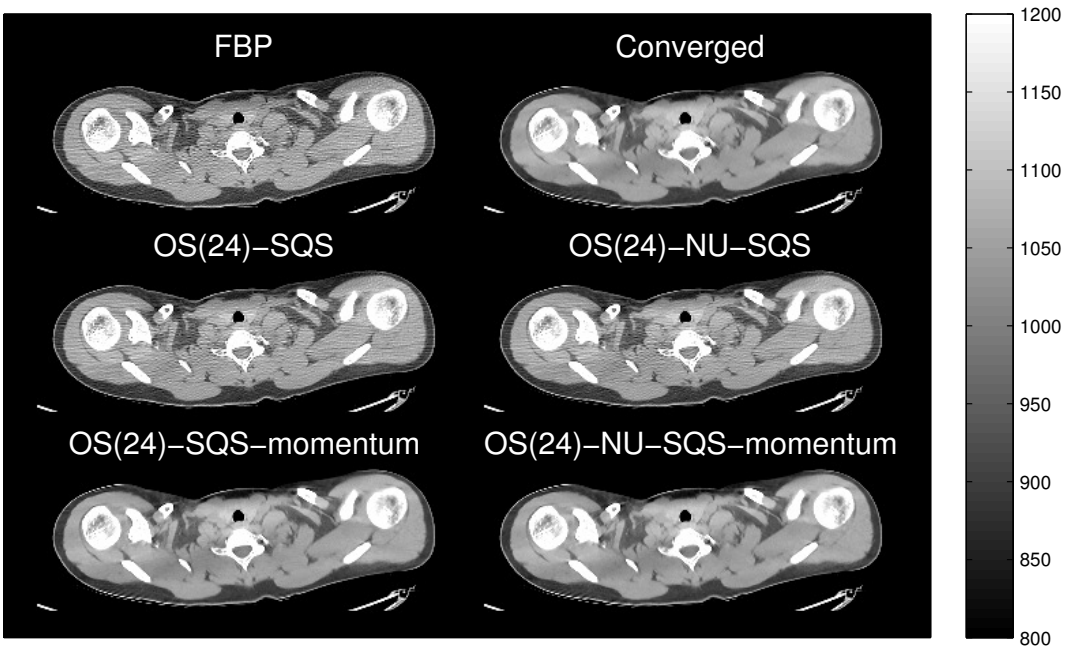


Fig. 4. Center slice of FBP image $x^{(0)}$, converged image $\hat{x}$, and reconstructed images of OS(24 subsets)-(NU-)SQS algorithms (with momentum) at 10th iteration.

(HU) for CT, versus iteration. The results show that inclusion of momentum highly accelerated the standard OS-SQS algorithm, and its combination with NU approach provided the fastest convergence among all considered algorithms. Fig. 3(b) reaffirms that using OS highly accelerates the standard optimization transfer methods that have slow convergence as shown in Fig. 3(a).

OS methods with 24 subsets required about 20% increased computation time per iteration compared to the standard SQS, and the NU approach needed additional 13% run time in our implementation. (This will vary depending on the implementation and computing environments.) In contrast, incorporating the momentum term requires negligible additional computation time, providing significant acceleration in terms of both number of iterations and computation time.

Fig. 4 presents the filtered back projection (FBP) that was used as initialization for all iterative algorithms in this work and converged image $\hat{x}$ as a reference. The reconstructed images of the standard (NU-)OS algorithms and the proposed (NU-)OS algorithms with momentum are from the 10th iteration. The results confirm that the momentum term greatly improves the speed of OS algorithms, and the combination of OS, NU and momentum gives the best image among others, very close to the converged image.

6. CONCLUSION

We have proposed to combine OS algorithms and momentum technique on top of optimization transfer methods. The proposed algorithms converge much faster than conventional OS algorithms, and are of practical interest. The next step is to examine the performance of proposed algorithms on additional real patient CT scans and to study theoretical convergence properties.

7. REFERENCES

[1] J-B. Thibault, K. Sauer, C. Bouman, and J. Hsieh, "A three-dimensional statistical approach to improved image quality for multi-slice helical CT," *Med. Phys.*, vol. 34, no. 11, pp. 4526–44, Nov. 2007.

[2] H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Trans. Med. Imag.*, vol. 13, no. 4, pp. 601–9, Dec. 1994.

[3] H. Erdoğan and J. A. Fessler, "Ordered subsets algorithms for transmission tomography," *Phys. Med. Biol.*, vol. 44, no. 11, pp. 2835–51, Nov. 1999.

[4] D. Kim and J. A. Fessler, "Parallelizable algorithms for X-ray CT image reconstruction with spatially non-uniform updates," in *Proc. 2nd Intl. Mtg. on image formation in X-ray CT*, 2012, pp. 33–7.

[5] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[6] D. Kim, D. Pal, J-B. Thibault, and J. A. Fessler, "Improved ordered subsets algorithm for 3D X-ray CT image reconstruction," in *Proc. 2nd Intl. Mtg. on image formation in X-ray CT*, 2012, pp. 378–81.