Accelerating X-ray CT ordered subsets image reconstruction with Nesterov's first-order methods

Donghwan Kim, Sathish Ramani, and Jeffrey A. Fessler

***Abstract*—Low-dose X-ray CT can reduce the risk of cancer to patients. However, it requires computationally expensive statistical image reconstruction methods for improved image quality. Iterative algorithms require long compute times, so we focus on algorithms that "converge" in few iterations. This paper proposes to apply ordered subsets (OS) methods to Nesterov's fast first-order methods for 3D X-ray CT problems. Nesterov's algorithms use previous iterates to provide momentum towards the optimum and thus achieve a fast convergence rate of $O(1/n^2)$, where n counts the number of iterations. We also propose to use separable quadratic surrogates (SQS) (with a non-uniform (NU) approach) in Nesterov's algorithms. We use a real patient helical CT scan to show that the proposed algorithms converge rapidly, and we investigate the behavior of OS methods in Nesterov's algorithms.**

I. Introduction

Based on the statistics of X-ray CT, we reconstruct a (non-negative) image $x \in \mathbb{R}_+^{N_p}$ from noisy measurements $y \in \mathbb{R}^{N_d}$ by minimizing a convex and continuously differentiable objective function $\Psi(x)$. This paper focuses on a penalized weighted least squares (PWLS) problem [1]:

$$\hat{x} = \arg\min_{x \succeq 0} \left\{ \Psi(x) \triangleq \frac{1}{2}||y - Ax||_W^2 + \beta R(x) \right\}, \quad (1)$$

where A is a projection operator, and the diagonal matrix W provides statistical weighting. $R(x)$ is a (edge-preserving) regularization function and β balances the data-fit term and $R(x)$. Due to the large scale of the problem (in 3D CT), iterative algorithms for minimizing $\Psi(x)$ require considerable compute time. Thus, the goal of this paper is to develop iterative algorithms that "converge" in fewer iterations.

This paper focuses on Nesterov's fast first-order algorithms [2], [3] that use previous iterates as momentum for additional acceleration towards the optimum. The former [2] uses two previous iterates as momentum, while the latter [3] uses all accumulated previous iterates. Both provide a fast convergence rate of $O(1/n^2)$ where n counts the number of iterations, whereas usual gradient-based methods have $O(1/n)$ convergence rate [4].

In our recent work [5], we combined ordered subsets (OS) methods [6], [7] with Nesterov's early work [2] that has been used to develop a fast iterative shrinkage-thresholding algorithm (FISTA) [4]. We also used a separable quadratic surrogates (SQS) method [7] (and a non-uniform approach [8]) in Nesterov's algorithm (in [5]). These combinations provided very promising results as they converged very rapidly even with relatively small number of subsets. (Using fewer subsets is preferable, as it decreases inexactness in OS methods and also reduces the overhead of computing the regularizer.) In addition, the overhead needed for proposed algorithms with OS-SQS is minimal as Nesterov's algorithms are simple. In this paper, we apply OS and (NU-)SQS methods to the more recent Nesterov's algorithm (2005) [3] and observe that this combination achieves as fast a convergence as the method in [5] but with improved stability.

We propose to use OS methods here as they can initially accelerate any gradient-based algorithms dramatically by approximating $\nabla\Psi(x)$ using only a subset of measurements. But, OS methods usually approach a limit-cycle looping around the optimum [6], [7]. (The more the subsets, the more the initial acceleration but with increased inexactness in the iterates.) However, the stability of OS methods in the proposed Nesterov's algorithms is unknown. Therefore, we experimentally investigated the behavior of OS in Nesterov's algorithms with respect to the number of subsets. We found that our newly proposed Nesterov's algorithm based on [3] with OS-SQS is more stable than the previous combination in [5].

In this paper, we propose to combine OS-SQS methods with Nesterov's fast first-order algorithms for X-ray CT image reconstruction. We first explain two of Nesterov's algorithms and illustrate their application to the X-ray CT problem in (1) with OS-SQS algorithms. Then we show the results for accelerated convergence of the two proposed algorithms using a real patient CT scan. We also discuss the stability of OS in Nesterov's algorithms.

II. Nesterov's algorithms

Nesterov published a fast first-order method using two previous iterates as a momentum for smooth functions[1] in [2], and it was extended later for non-smooth functions by Beck *et al.* [4], which is one of the state-of-the-art methods in image restoration. In [3], Nesterov also proposed new formulation of a fast first-order method using all previous iterates.

Both algorithms [2], [3] have been used widely for various optimization problems. They have also been used for X-ray CT reconstruction showing a noticeable acceleration [9], [10]. However, Nesterov's algorithms by themselves are not very

D. Kim, S. Ramani, and J. A. Fessler are with the Dept. of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109 USA (e-mail:kimdongh@umich.edu, sramani@umich.edu, fessler@umich.edu).

Supported in part by NIH grant R01-HL-098686 and equipment donations from Intel. Helical CT data provided by GE Healthcare.

[1]A smooth function $f(x)$ is continuously differentiable with Lipschitz continuous gradient L satisfying $||\nabla f(x) - \nabla f(z)|| \leq L||x - z||$ for all $x, z \in \mathbb{R}^{N_p}$.

attractive in CT, as the cost function $\Psi(x)$ in (1) has a large Lipschitz constant that slows down the convergence [11]. Here, we suggest new combinations of Nesterov's algorithms and OS-SQS methods that show very promising results.

We first review Nesterov's algorithms briefly. Both [2] and [3] begin by using an optimization transfer technique [12]. Nesterov uses a convex cost function $\Psi(x)$ that is continuously differentiable with Lipschitz constant L, which can be majorized at the nth iteration as:

$$\Psi(x) \leq \phi_L^{(n)}(x) \tag{2}$$
$$\triangleq \Psi(x^{(n)}) + \nabla\Psi(x^{(n)})'(x - x^{(n)}) + \frac{L}{2}||x - x^{(n)}||^2.$$

The optimization transfer step minimizes the surrogate $\phi_L^{(n)}(x)$ at nth iteration:

$$x^{(n+1)} = \arg\min_{x \succeq 0} \phi_L^{(n)}(x) = \left[x^{(n)} - \frac{1}{L}\nabla\Psi(x^{(n)})\right]_+, \tag{3}$$

where $[\cdot]_+$ enforces a non-negativity constraint. Then the algorithm (3) is accelerated using previous iterates as shown in Figs. 1 and 2 [2], [3]. We use the choice of parameters suggested in [13] for the algorithm in Fig. 2, which provides faster convergence than the choice in [3].

Initialize $x^{(0)} = v^{(0)},\ t_0 = 1$
for $n = 0, 1, 2, \cdots$
$$t_{n+1} = \left(1 + \sqrt{1 + 4t_n^2}\right)/2$$
$$x^{(n+1)} = \left[z^{(n)} - \frac{1}{L}\nabla\Psi(z^{(n)})\right]_+$$
$$z^{(n+1)} = x^{(n+1)} + \frac{t_n - 1}{t_{n+1}}(x^{(n+1)} - x^{(n)})$$

Fig. 1. Nesterov's algorithm (1983) [2].

Initialize $x^{(0)} = v^{(0)} = z^{(0)},\ t_0 = 1$
for $n = 0, 1, 2, \cdots$
$$t_{n+1} = \left(1 + \sqrt{1 + 4t_n^2}\right)/2$$
$$x^{(n+1)} = \left[z^{(n)} - \frac{1}{L}\nabla\Psi(z^{(n)})\right]_+$$
$$v^{(n+1)} = \left[z^{(0)} - \frac{1}{L}\sum_{k=0}^{n} t_k \nabla\Psi(z^{(k)})\right]_+$$
$$z^{(n+1)} = \left(1 - \frac{1}{t_{n+1}}\right)x^{(n+1)} + \frac{1}{t_{n+1}}v^{(n+1)}$$

Fig. 2. Nesterov's algorithm (2005) [3].

The sequences $\left\{x^{(n)}\right\}$ generated by both algorithms are proven to have the following convergence rate [2], [3]:

$$\Psi(x^{(n)}) - \Psi(\hat{x}) \leq O\left(\frac{L}{n^2}\right). \tag{4}$$

This is promising since ordinary optimization transfer in (3) provides only $O(1/n)$ rate [4]. However, the large Lipschitz constant L in CT problem causes slow convergence even with the $O(1/n^2)$ rate.

III. Proposed Nesterov's algorithms with ordered subsets

We suggest combining ordered subsets with Nesterov's fast first-order algorithms. Ordered subsets algorithms group projection views into M subsets evenly, and assume

$$\nabla\Psi(x) \approx M\nabla\Psi_0(x) \approx \cdots \approx M\nabla\Psi_{M-1}(x), \tag{5}$$

where we define the subset gradient:

$$\nabla\Psi_m(x) \triangleq A_m' W_m(A_m x - y_m) + \frac{\beta}{M}\nabla R(x) \tag{6}$$

for $m = 0, \cdots, M-1$. The matrices A_m, y_m and W_m are sub-matrices of A, y, and W corresponding to mth subset. We accelerate Nesterov's algorithms by replacing $\nabla\Psi(\cdot)$ in Figs. 1 and 2 with $M\nabla\Psi_m(\cdot)$. We count each mth sub-iteration as $1/M$ iteration, since $M\nabla\Psi_m(\cdot)$ requires roughly $1/M$ amount of computation of $\nabla\Psi(\cdot)$. Then we expect to have the following convergence rate in early iterations:

$$\Psi(x^{(n+\frac{m}{M})}) - \Psi(\hat{x}) \lessapprox O\left(\frac{L}{(nM+m)^2}\right). \tag{7}$$

This rate will not hold as the sequence $\left\{x^{(n+\frac{m}{M})}\right\}$ nears the optimum where the condition (5) fails.

Owing to the acceleration in proposed algorithms based on the M^2 effect of OS in (7), it is possible to use fewer subsets for better accuracy in OS. However, it is unknown how the inexactness in OS methods affect the behavior of the Nesterov's algorithms. (Ordinary OS methods are known to reach a limit-cycle looping around the optimum.) Therefore, we investigated OS algorithms with Nesterov's algorithms in Section IV, where we found that Nesterov's algorithm (2005) with OS methods is better stabilized than the earlier one.

For CT, it is computationally expensive to find the smallest possible Lipschitz constant L, and the backtracking line search scheme in [4] would be undesirably slow. Instead, we use a separable quadratic surrogate (SQS) method [7] for the optimization transfer step in (3), replacing $\phi_L^{(n)}$ in (2) by

$$\phi_{SQS}^{(n)}(x) \triangleq \Psi(x^{(n)}) + \nabla\Psi(x^{(n)})'(x - x^{(n)}) + \frac{1}{2}||x - x^{(n)}||_D^2, \tag{8}$$

where D is a diagonal matrix. The advantage of using SQS is that we can compute an exact surrogate $\phi_{SQS}^{(n)}(x)$ with modest computation. We can further accelerate the SQS-type algorithms by our recently proposed non-uniform approach [8].

We summarize the proposed algorithms, namely OS-SQS-Nes83 and OS-SQS-Nes05, in Figs. 3 and 4 that respectively combine OS-SQS with the two methods of Nesterov in Figs. 1 and 2.

IV. Results

We used a 3D helical X-ray CT data set of a human shoulder to show the acceleration of proposed algorithms. We computed the root mean square difference (RMSD) between the current and converged[2] image within the region-of-interest (ROI) in

[2]We generated an (almost) converged image by running 100 iterations of (convergent) NH-ABCD-SQS [8] followed by 2000 iterations of (convergent) SQS.

$$
\begin{aligned}
&\text{Initialize } x^{(0)} = v^{(0)}, t_0 = 1 \\
&\text{for } n = 0, 1, 2, \cdots \\
&\text{for } m = 0, 1, \cdots, M-1 \\
&\quad t_{nM+m+1} = \left(1 + \sqrt{1 + 4t_{nM+m}^2}\right)/2 \\
&\quad x^{(n+\frac{m+1}{M})} = \left[z^{(n+\frac{m}{M})} - D^{-1} M \nabla \Psi_m (z^{(n+\frac{m}{M})}) \right]_+ \\
&\quad z^{(n+\frac{m+1}{M})} = x^{(n+\frac{m+1}{M})} + \frac{t_{nM+m} - 1}{t_{nM+m+1}} (x^{(n+\frac{m+1}{M})} - x^{(n+\frac{m}{M})})
\end{aligned}
$$

Fig. 3. Proposed Nesterov's algorithm (1983) with ordered subsets (OS-Nes83).

$$
\begin{aligned}
&\text{Initialize } x^{(0)} = v^{(0)} = z^{(0)}, t_0 = 1 \\
&\text{for } n = 0, 1, 2, \cdots \\
&\text{for } m = 0, 1, \cdots, M-1 \\
&\quad t_{nM+m+1} = \left(1 + \sqrt{1 + 4t_{nM+m}^2}\right)/2 \\
&\quad x^{(n+\frac{m+1}{M})} = \left[z^{(n+\frac{m}{M})} - D^{-1} M \nabla \Psi_m (z^{(n+\frac{m}{M})}) \right]_+ \\
&\quad v^{(n+\frac{m+1}{M})} = \left[z^{(0)} - D^{-1} \sum_{k=0}^{nM+m} t_k M \nabla \Psi_{(k)_M} (z^{(\frac{k}{M})}) \right]_+ \\
&\quad z^{(n+\frac{m+1}{M})} = \left(1 - \frac{1}{t_{nM+m+1}}\right) x^{(n+\frac{m+1}{M})} + \frac{1}{t_{nM+m+1}} v^{(n+\frac{m+1}{M})}
\end{aligned}
$$

Fig. 4. Proposed Nesterov's algorithm (2005) with ordered subsets (OS-Nes05). The notation $(k)_M$ stands for $k \bmod M$.

Hounsfield Units (HU):

$$
\text{RMSD} = \frac{||x_{\text{ROI}}^{(n)} - \hat{x}_{\text{ROI}}||_2}{\sqrt{N_{\text{p,ROI}}}} [\text{HU}] \tag{9}
$$

versus iteration, to evaluate the convergence rate. In Fig. 5(a), we used different number of subsets such as 1, 24, and 48 subsets and observed that the ordered subsets highly accelerated both Nesterov's algorithms.

However, OS-Nes83 algorithms diverged when we used more than 40 subsets (as seen in the case of 48 subsets in Fig. 5(a)), while OS-Nes05 algorithm remained stable with more than 100 subsets. (Results not shown here.) Based on our observations, we believe that OS-Nes05 is more stable than OS-Nes83. We can intuitively understand this behavior, since OS-Nes05 method uses accumulated momentum that is less prone to local inexactness, while OS-Nes83 uses the difference between two previous iterates as momentum which may be very inaccurate in OS-type methods. However, we need theoretical justification to better understand the behavior of OS in Nesterov's algorithms, and we leave it as a future work.

We also combined a non-uniform (NU) approach [8] with OS-SQS-Nes05 to investigate the net resulting acceleration. In Fig. 5(b), we obtained some acceleration when including NU, but the algorithm reached a larger limit-cycle than the case without NU. OS-Nes05 with 24 subsets showed promising acceleration (with a slightly larger limit-cycle), but the algorithm with 48 subsets reached a quite large limit-cycle after initial acceleration. Further refinement of NU method is needed to reach a relatively small limit-cycle while achieving noticeable acceleration for large M.

Fig. 6 presents the initial filtered back projection (FBP) image $x^{(0)}$, the converged image $\hat{x}$, and reconstructed images at 12th iteration from four different algorithms for comparison. Both SQS-Nes05 and OS48-NUSQS at 12 iteration are still far from the converged image. The proposed algorithms OS48-SQS-Nes05 and OS24-NUSQS-Nes05 reach low RMSD level after 10 iterations in Fig. 5(b), and their reconstructed images at 12 iteration are very close to the converged image. The

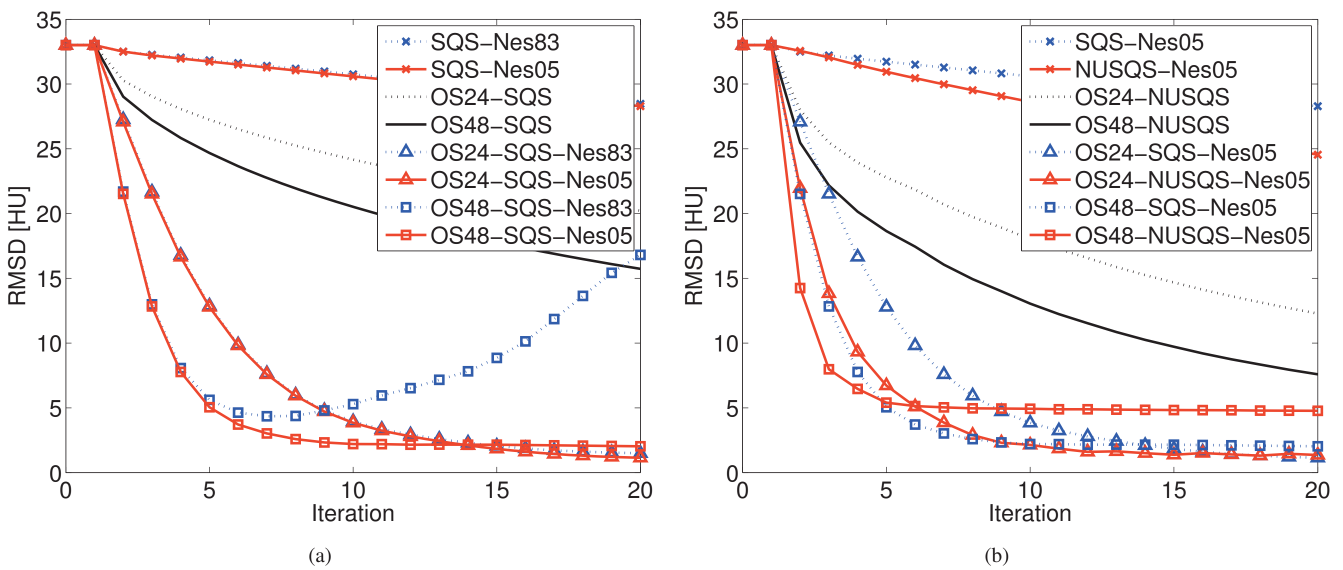

Fig. 5. Plots of RMSD in (9) versus iterations for various proposed Nesterov's algorithms with OS-NUSQS. (There are no changes in RMSD during the first iteration, since we count the precomputation of D as one iteration.)

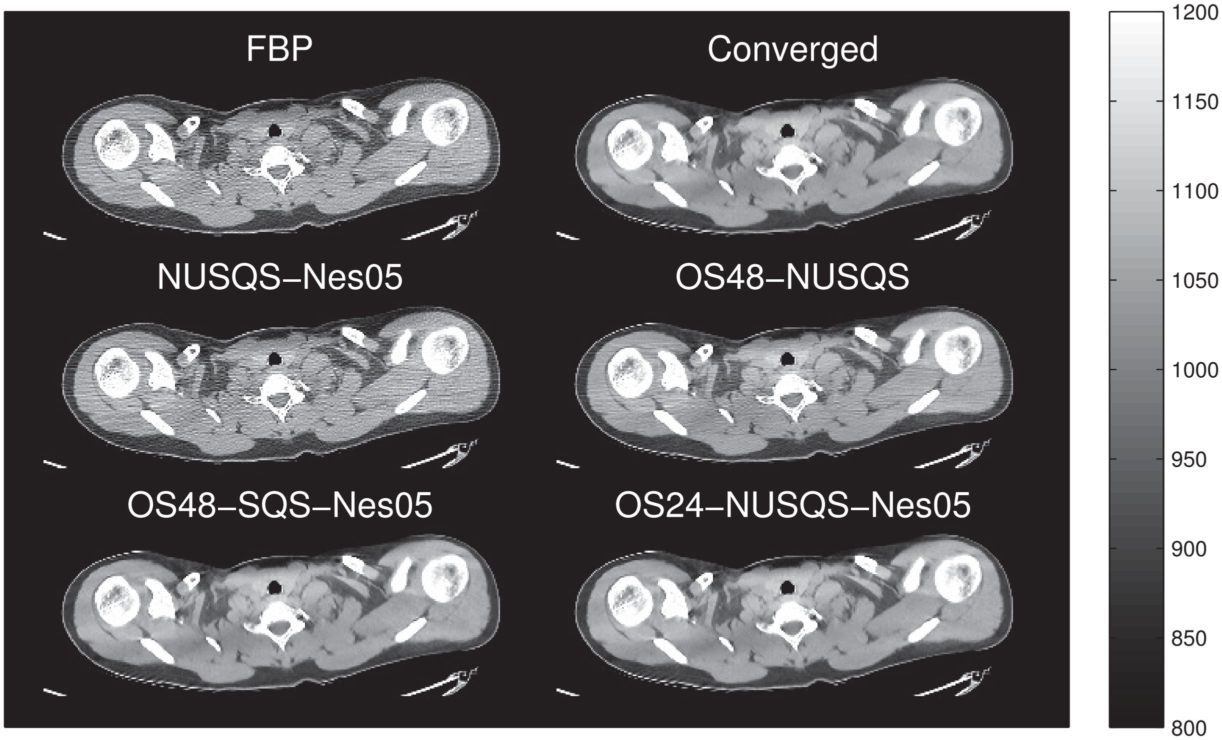

Fig. 6. Center slice of FBP image $x^{(0)}$, converged image $\hat{x}$, and reconstructed images at 12th iteration.

results confirm that the proposed combinations of OS and Nesterov's algorithms reach a decent image (close to $\hat{x}$) in few iterations.

V. Discussion

In this paper, we used a helical CT data set that corresponds to 984 projection views per turn with pitch 1.0. From the results, we were able to assess the behavior of OS-(NU)SQS-Nes empirically for this specific geometry. However, the number of subsets used for this geometry may not be optimal for other geometries. So, it is important to investigate the problem of selecting the appropriate number of subsets for a given geometry that would ensure fast convergence without encountering stability issues.

VI. Conclusion

We proposed two algorithms that combine Nesterov's methods with OS. The proposed algorithms provide dramatic acceleration in X-ray CT reconstruction with relatively small number of subsets. We found that the Nesterov's algorithm (2005) [3] is more stable with ordered subsets than the other choice [2] in our experiment. But, this should be examined on various other data sets, and we leave the theoretical justification as a future work.

Here, we investigated two specific methods [2], [3] for combining "momentum" terms with ordered subsets. There are many other possible ways to introduce momentum into OS methods and our future work aims at finding ways that are fast yet relatively stable for OS-type updates.

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] Y. Nesterov, "A method of solving a convex programming problem with convergence rate $O(1/k^2)$," *Soviet Math. Dokl.*, vol. 27, no. 2, pp. 372–76, 1983.
[3] Y. Nesterov, "Smooth minimization of non-smooth functions," *Mathematical Programming*, vol. 103, no. 1, pp. 127–52, May 2005.
[4] 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.
[5] D. Kim, S. Ramani, and J. A. Fessler, "Ordered subsets with momentum for accelerated X-ray CT image reconstruction," in *Proc. IEEE Conf. Acoust. Speech Sig. Proc.*, 2013, To appear.
[6] 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.
[7] H. Erdoğan and J. A. Fessler, "Ordered subsets algorithms for transmission tomography," *Phys. Med. Biol.*, vol. 44, no. 11, pp. 2835–51, Nov. 1999.
[8] 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. 33–6.
[9] K. Choi, J. Wang, L. Zhu, T-S. Suh, S. Boyd, and L. Xing, "Compressed sensing based cone-beam computed tomography reconstruction with a first-order method," *Med. Phys.*, vol. 37, no. 9, pp. 5113–25, Nov. 2010.
[10] J. H. Jorgensen, T. L. Jensen, P. C. Hansen, S. H. Jensen, E. Y. Sidky, and X. Pan, "Accelerated gradient methods for total-variation-based CT image reconstruction," in *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, 2011, pp. 435–8.
[11] S. Ramani and J. A. Fessler, "A splitting-based iterative algorithm for accelerated statistical X-ray CT reconstruction," *IEEE Trans. Med. Imag.*, vol. 31, no. 3, pp. 677–88, Mar. 2012.
[12] M. W. Jacobson and J. A. Fessler, "An expanded theoretical treatment of iteration-dependent majorize-minimize algorithms," *IEEE Trans. Im. Proc.*, vol. 16, no. 10, pp. 2411–22, Oct. 2007.
[13] P. Tseng, "On accelerated proximal gradient methods for convex-concave optimization," 2008.