Alternating Direction Method of Multiplier for Tomography With Nonlocal Regularizers

Se Young Chun*, *Member, IEEE*, Yuni K. Dewaraja, *Member, IEEE*, and Jeffrey A. Fessler, *Fellow, IEEE*

Abstract—The ordered subset expectation maximization (OSEM) algorithm approximates the gradient of a likelihood function using a subset of projections instead of using all projections so that fast image reconstruction is possible for emission and transmission tomography such as SPECT, PET, and CT. However, OSEM does not significantly accelerate reconstruction with computationally expensive regularizers such as patch-based nonlocal (NL) regularizers, because the regularizer gradient is evaluated for every subset. We propose to use variable splitting to separate the likelihood term and the regularizer term for penalized emission tomographic image reconstruction problem and to optimize it using the alternating direction method of multiplier (ADMM). We also propose a fast algorithm to optimize the ADMM parameter based on convergence rate analysis. This new scheme enables more sub-iterations related to the likelihood term. We evaluated our ADMM for 3-D SPECT image reconstruction with a patch-based NL regularizer that uses the Fair potential function. Our proposed ADMM improved the speed of convergence substantially compared to other existing methods such as gradient descent, EM, and OSEM using De Pierro's approach, and the limited-memory Broyden–Fletcher–Goldfarb–Shanno algorithm.

Index Terms—Alternating direction method of multiplier, emission tomography, nonlocal (NL) regularizer, ordered-subset expectation maximization (OSEM).

I. INTRODUCTION

INCORPORATING noise models in tomographic image reconstruction can improve image quality. However, unlike analytical image reconstruction methods such as filtered back-projection (FBP), statistical image reconstruction methods such as the expectation-maximization (EM) algorithm [1], often require gradient-based iterative algorithms. Since the gradient of a likelihood function should be evaluated at each iteration, these algorithms (including EM) are undesirably slow.

*S. Y. Chun was with the Department of Electrical Engineering and Computer Science and Department of Radiology, University of Michigan, Ann Arbor, MI, USA. He is now with the School of Electrical and Computer Engineering, Ulsan National Institute of Science and Technology (UNIST), Ulsan 689-798, Republic of Korea (e-mail: sychun@unist.ac.kr).

Y. K. Dewaraja is with the Department of Radiology, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: yuni@umich.edu).

J. A. Fessler is with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: fessler@umich.edu).

Statistical image reconstruction for emission tomography started to be used widely in clinics and in commercial scanners after the fast algorithm called ordered-subset expectation-maximization (OSEM) was developed [2]. The main idea was to speed up gradient computation by approximating it using a subset of projections instead of using all projections (ordered-subset or OS approximation). This approximation has been used for unregularized emission tomographic image reconstruction [2] and regularized emission and transmission tomographic image reconstruction with simple quadratic or edge-preserving regularizers [3]. Since the computation cost for these regularizers is fairly low compared to that for the likelihood term, OS algorithms that approximate the gradient of the likelihood term often speed up penalized likelihood (PL) image reconstruction, too.

Recently, nonlocal (NL) regularizers have been proposed that improve image quality substantially compared to conventional regularizers such as quadratic or edge-preserving functions in image deconvolution [4], emission image reconstruction with convex functions [5], [6], and magnetic resonance imaging (MRI) image reconstruction with nonconvex functions [7]. NL regularizers have been extended to use high-resolution CT or MRI side information for emission and super-resolution image reconstruction for further improvement of image quality [8]–[11]. For emission tomography problems such as [6], [9]–[11], various optimization algorithms were used for image reconstruction such as gradient descent (GD) [10], EM (or OSEM) algorithm based on optimization transfer using De Pierro's lemma [6], the EM algorithm using one-step late approach [11], and the quasi-Newton algorithm called the limited-memory Broyden–Fletcher–Goldfarb–Shanno with a Box constraint (L-BFGS-B) [9]. They showed promising improvement of image quality, but their algorithms were undesirably slow since the computation cost of the NL regularizers can be comparable to or even higher than that of the likelihood. Therefore, the OS approximation does not significantly accelerate the convergence rate of existing PL image reconstruction algorithms with these NL regularizers.

In this paper, we propose to use variable splitting to separate the likelihood term and the regularizer term for PL image reconstruction problem and to optimize it using the alternating direction method of multipliers (ADMM) [12]. We also propose a fast algorithm to optimize the ADMM parameter based on convergence rate analysis that was extended from the work of Ghadimi *et al.* with quadratic data fitting and regularizer terms [13]. There are existing methods that use variable splitting for the data fidelity term and the regularizer term [4], [14]–[16]. These previous methods address nonsmooth regularizers such

as total variation (TV). A sub-problem with nonsmooth regularizers can be solved quickly by using efficient proximal operators such as shrinkage. Our proposed variable splitting has different motivation. We divide the original optimization into a few sub-problems and we solve the sub-problem related to the likelihood term using OS approximation and we update the sub-problem related to the NL regularizer less often.

We evaluated our new ADMM for 3-D SPECT image reconstruction with a patch-based convex NL regularizer that uses the Fair potential [6]. Our simulation using the XCAT phantom [17] shows that our proposed ADMM for image reconstruction with NL regularizers accelerated convergence substantially compared to existing methods such as GD, EM, and OSEM using De Pierro's approach [3], and the L-BFGS-B algorithm [18]. This paper is organized as follows. Section II reviews statistical image reconstruction in emission tomography, OS approximation, and various NL regularizers. Section III proposes an efficient method for NL regularized image reconstruction by using variable splitting and ADMM. Section IV proposes an analytical method to select automatically a suitable ADMM parameter for fast convergence rate. Lastly, Section V presents 3-D SPECT simulation results with the XCAT phantom [17] for an application in I-131 radioimmunotherapy (RIT) [19].

II. IMAGE RECONSTRUCTION WITH NL REGULARIZERS

A. Statistical Image Reconstruction for Tomography

The usual form of statistical image reconstruction is to perform the following optimization with respect to an image $\boldsymbol{f}$:

$$\hat{\boldsymbol{f}} \triangleq \arg\min_{\boldsymbol{f} \geq \boldsymbol{0}} \mathrm{L}(\boldsymbol{y}|\boldsymbol{f}) + \beta \mathrm{R}(\boldsymbol{f}) \tag{1}$$

where $\boldsymbol{y}$ is a measured sinogram data, L denotes a negative log-likelihood function, β is a regularization parameter, and R is a regularizer.

For emission tomography, the negative Poisson log-likelihood is

$$\mathrm{L}(\boldsymbol{y}|\boldsymbol{f}) = \sum_{l \in \Omega} \bar{y}_l(\boldsymbol{f}) - y_l \log \bar{y}_l(\boldsymbol{f}) \tag{2}$$

where y_l is the lth element of the measurement $\boldsymbol{y}$, Ω is the set of indexes of all measurements, and $\bar{y}_l(\boldsymbol{f}) \triangleq \sum_j a_{lj} f_j + s_l$ where a_{lj} is the element of the system matrix $\boldsymbol{A}$ at the lth row and the jth column, f_j is the jth element of the image vector $\boldsymbol{f}$, and s_l is a random and scatter component for the lth measurement. We focus on SPECT imaging where we incorporate an attenuation map and a depth-dependent point spread function model including penetration tails [20] in the system matrix $\boldsymbol{A}$. We assume known s_l; in practice, this scatter component can be estimated by using a triple energy window (TEW) method or by Monte Carlo methods [21].

B. Ordered-Subset Approximation

Iterative image reconstruction algorithms for (1) usually require calculating the gradient of $L(\boldsymbol{y}|\boldsymbol{f})$ and $\mathrm{R}(\boldsymbol{f})$ at every iteration. The gradient is evaluated at $\boldsymbol{f}^{(n)}$ where $\boldsymbol{f}^{(n)}$ is an estimate of $\boldsymbol{f}$ at the nth iteration. These algorithms include GD, EM [1],

and L-BFGS-B [18]. Calculating the gradient of typical $\mathrm{R}(\boldsymbol{f})$ such as quadratic or edge-preserving regularizers is very fast. Evaluating the gradient of $L(\boldsymbol{y}|\boldsymbol{f})$ is much slower since this requires one forward projection and one back projection of $\boldsymbol{A}$ for each iteration.

OS methods [2] approximate the gradient of $L(\boldsymbol{y}|\boldsymbol{f})$ at $\boldsymbol{f}^{(n)}$ with the gradient of $L_k(\boldsymbol{y}|\boldsymbol{f})$ at $\boldsymbol{f}^{(n+k/K)}$ where

$$L_k(\boldsymbol{y}|\boldsymbol{f}) \triangleq K \sum_{l \in \Omega_k} \bar{y}_l(\boldsymbol{f}) - y_l \log \bar{y}_l(\boldsymbol{f}) \tag{3}$$

Ω_k are mutually exclusive, $\Omega = \cup_k \Omega_k$, and $k = 1, \cdots, K$. Evaluating the OS approximated gradient in (3) is about K times faster than calculating the original gradient in (2). In this way, OSEM achieves faster image reconstruction. Note that OS methods with a fixed $K > 1$ do not guarantee convergence, but yield approximate PL images.

C. Nonlocal Regularizers

Recently, many researchers have formed high-quality images in many image reconstruction problems by replacing $\mathrm{R}(\boldsymbol{f})$ in (1) with a NL regularizer [4]–[7]. A typical NL regularizer looks like

$$\mathrm{R}(\boldsymbol{f}) \triangleq \sum_{i,j \in S_i} p_{ij} \left(\|\boldsymbol{N}_i \boldsymbol{f} - \boldsymbol{N}_j \boldsymbol{f}\|^2 \right) \tag{4}$$

where $p_{ij}(t)$ is a function of a scalar variable t, $\|\cdot\|$ is the ℓ_2 norm, S_i is the search neighborhood around the ith voxel (usually the set of all voxels within a fixed ℓ_∞ distance from the ith voxel), and $\boldsymbol{N}_i$ is an operator on the image $\boldsymbol{f}$ such that $\boldsymbol{N}_i \boldsymbol{f}$ is a vector of image intensities of all voxels within a fixed ℓ_∞ distance from the ith voxel (cube-shaped patch).

A typical choice for the function p_{ij} is [4], [5]

$$p_{ij}(t) = \frac{w_{i,j}(\tilde{\boldsymbol{f}})}{2N_{\mathrm{f}}} t \tag{5}$$

where N_{f} is the number of voxels in the patch $\boldsymbol{N}_i \boldsymbol{f}$ (assuming that the patch size is the same for all i), a weighting function is

$$w_{i,j}(\boldsymbol{f}) = \exp\left(-\frac{\|\boldsymbol{N}_i \boldsymbol{f} - \boldsymbol{N}_j \boldsymbol{f}\|^2}{2N_{\mathrm{f}} \sigma_{\mathrm{f}}^2}\right) \tag{6}$$

and σ_{f} is a design parameter. For the image $\tilde{\boldsymbol{f}}$, Lou *et al.* used an initial image from any analytical image reconstruction (e.g., FBP) [5] and Zhang *et al.* used an estimated image from the previous iteration $\boldsymbol{f}^{(n-1)}$ so that $p_{ij}(t)$ changes over iterations [4].

Yang *et al.* used a few nonconvex potentials including the Welsh potential [22]

$$p_{ij}(t) = \sigma_{\mathrm{f}}^2 \left(1 - \exp\left(-\frac{t}{2N_f \sigma_{\mathrm{f}}^2}\right)\right) \tag{7}$$

and showed that using (7) is equivalent to using (5) with an estimated value $\boldsymbol{f}^{(n-1)}$ for $\tilde{\boldsymbol{f}}$ at the nth iteration [7]. Wang *et al.* used the Fair potential [23], [24]

$$p_{ij}(t) = \sigma_{\mathrm{f}}^2 \left(\sqrt{\frac{t}{\sigma_{\mathrm{f}}^2 N_{\mathrm{f}}}} + \log\left(1 + \sqrt{\frac{t}{\sigma_{\mathrm{f}}^2 N_{\mathrm{f}}}}\right)\right). \tag{8}$$

Both (7) and (8) do not depend on an initial image and (8) is convex while (7) is nonconvex. It has been reported that nonconvex functions yielded better image quality than a convex function [7].

One can also design NL regularizers that incorporate high-resolution side information such as CT or MR images [9]–[11] to further improve image quality. One way to incorporate high-resolution side information into the NL regularizer is to use the following type of NL regularizer [10]:

$$p_{ij}(t) = \frac{w_{i,j}\left(\boldsymbol{f}^{(n-1)}\right) v_{i,j}(\boldsymbol{g})}{2N_{\mathrm{f}}} t \tag{9}$$

where

$$v_{i,j}(\boldsymbol{g}) = \exp\left(-\frac{\|\boldsymbol{M}_i\boldsymbol{g} - \boldsymbol{M}_j\boldsymbol{g}\|^2}{2N_{\mathrm{a}}\sigma_{\mathrm{a}}^2}\right) \tag{10}$$

$\boldsymbol{g}$ is a high-resolution image such as MR or CT, $\boldsymbol{M}_i$ is an operator on the image $\boldsymbol{g}$ such that $\boldsymbol{M}_i\boldsymbol{g}$ is a vector of image intensities in a patch around the ith voxel, N_{a} is the number of voxels in the patch $\boldsymbol{M}_i\boldsymbol{g}$, and σ_{a} is another design parameter. Another NL regularizer incorporating high resolution side information is [9]

$$p_{ij}(t) = \frac{w_{i,j}\left(\boldsymbol{f}^{(n-1)}\right) + v_{i,j}(\boldsymbol{g})}{4N_{\mathrm{f}}} t. \tag{11}$$

All of these NL regularizers are computationally expensive due to the calculation of (6) at each iteration. The gradients of both $L(\boldsymbol{y}|\boldsymbol{f})$ and $\mathrm{R}(\boldsymbol{f})$ should be evaluated at each iteration for optimization. Using OS approximations of the gradient of $L(\boldsymbol{y}|\boldsymbol{f})$ does not improve the speed of convergence much since one can not use OS approximations for $\mathrm{R}(\boldsymbol{f})$ so the gradient of $\mathrm{R}(\boldsymbol{f})$ must be evaluated at each *sub*-iteration.

III. ALTERNATING DIRECTION METHOD OF MULTIPLIERS

A. ADMM for NL Regularization

To benefit from an OS approximation for $L(\boldsymbol{y}|\boldsymbol{f})$, while avoiding heavy computation of the gradient of $\mathrm{R}(\boldsymbol{f})$ at each sub-iteration, we split the variable for the likelihood term and the regularizer term by replacing (1) with the following equivalent constrained optimization problem:

$$\hat{\boldsymbol{f}} \triangleq \arg\min_{\boldsymbol{f}\geq 0, \boldsymbol{u}} L(\boldsymbol{y}|\boldsymbol{f}) + \beta\mathrm{R}(\boldsymbol{u}), \text{ subject to } \boldsymbol{u} = \boldsymbol{f}. \tag{12}$$

The augmented Lagrangian for (12) is

$$L(\boldsymbol{y}|\boldsymbol{f}) + \beta\mathrm{R}(\boldsymbol{u}) + \frac{\mu}{2}\|\boldsymbol{f} - \boldsymbol{u} - \boldsymbol{d}\|^2, \text{ subject to } \boldsymbol{f} \geq 0 \tag{13}$$

where μ is a scalar value (design parameter) and $\boldsymbol{d}$ is a Lagrangian multiplier vector. We need to find a saddle point of the augmented Lagrangian (13).

We solve (13) using the ADMM algorithm [25], [26] as follows:

$$
\begin{aligned}
&For\ n = 1, 2, \cdots \\
&\boldsymbol{u}^{(n)} \in \arg\min_{\boldsymbol{u}} \frac{\mu}{2}\left\|\boldsymbol{u} - \boldsymbol{f}^{(n-1)} + \boldsymbol{d}^{(n-1)}\right\|^2 + \beta\mathrm{R}(\boldsymbol{u}) \quad (14) \\
&\boldsymbol{f}^{(n)} = \arg\min_{\boldsymbol{f}\geq 0} L(\boldsymbol{y}|\boldsymbol{f}) + \frac{\mu}{2}\left\|\boldsymbol{f} - \boldsymbol{u}^{(n)} - \boldsymbol{d}^{(n-1)}\right\|^2 \quad (15) \\
&\boldsymbol{d}^{(n)} = \boldsymbol{d}^{(n-1)} - \left(\boldsymbol{f}^{(n)} - \boldsymbol{u}^{(n)}\right) \quad (16) \\
&End.
\end{aligned}
$$

For convex R, this ADMM algorithm is guaranteed to converge for any $\mu > 0$ [26]. We can solve the sub-problems of (14) and (15) using any existing method. One need not solve the sub-problems exactly to guarantee the convergence of the ADMM algorithm [26, Th. 8].

B. Optimization for the Sub-Problem (14)

We used the GD algorithm to solve (14) as follows:

$$\boldsymbol{u}^{(n+1)} = \boldsymbol{u}^{(n)} - \alpha^{(n)}\nabla\Phi^{(n)}\left(\boldsymbol{u}^{(n)}\right) \tag{17}$$

where α is a step size and

$$\Phi^{(n)}(\boldsymbol{u}) \triangleq \frac{\mu}{2}\left\|\boldsymbol{u} - \boldsymbol{f}^{(n)} + \boldsymbol{d}^{(n)}\right\|^2 + \beta\mathrm{R}(\boldsymbol{u}). \tag{18}$$

We plug (17) into (14) to determine the step size as follows:

$$\alpha^{(n)} = \arg\min_{\alpha} \phi^{(n)}(\alpha) \tag{19}$$

where

$$\phi^{(n)}(\alpha) \triangleq \Phi^{(n)}\left(\boldsymbol{u}^{(n)} - \alpha\nabla\Phi^{(n)}\left(\boldsymbol{u}^{(n)}\right)\right). \tag{20}$$

The gradient of $\Phi^{(n)}(\boldsymbol{u})$ is

$$\nabla\Phi^{(n)}(\boldsymbol{u}) = \mu\left(\boldsymbol{u} - \boldsymbol{f}^{(n)} + \boldsymbol{d}^{(n)}\right) + \beta\nabla\mathrm{R}(\boldsymbol{u}) \tag{21}$$

where

$$
\begin{aligned}
\nabla\mathrm{R}(\boldsymbol{u}) = \sum_{i,j \in S_i} 2\dot{p}_{ij}\left(\|\boldsymbol{N}_i\boldsymbol{u} - \boldsymbol{N}_j\boldsymbol{u}\|^2\right) \\
\cdot(\boldsymbol{N}_i - \boldsymbol{N}_j)'(\boldsymbol{N}_i - \boldsymbol{N}_j)\boldsymbol{u} \quad (22)
\end{aligned}
$$

and $\dot{p}_{ij}(t)$ is the derivative of $p_{ij}(t)$. Since solving (19) is an intermediate step of solving (14), we do not need to find an exact α value to minimize (19). We chose to use one step of Newton's method for (19) as follows [27]:

$$\alpha^{(n)} = -\frac{\dot{\phi}^{(n)}(0)}{\ddot{\phi}^{(n)}(0)} \tag{23}$$

where $\dot{\phi}^{(n)}(\alpha)$ and $\ddot{\phi}^{(n)}(\alpha)$ are the first and second derivatives of $\phi^{(n)}(\alpha)$ with respect to α

$$\dot{\phi}^{(n)}(0) = -\left\|\nabla\Phi_{\boldsymbol{u}}^{(n)}\right\|^2 \tag{24}$$

where $\nabla \Phi_{\boldsymbol{u}}^{(n)} \triangleq \nabla \Phi^{(n)}(\boldsymbol{u}^{(n)})$ and

$$\ddot{\phi}^{(n)}(0) \approx \nabla \Phi_{\boldsymbol{u}}^{(n)\prime} \left(\mu \nabla \Phi_{\boldsymbol{u}}^{(n)} + \nabla \mathrm{R} \left(\nabla \Phi_{\boldsymbol{u}}^{(n)} \right) \right). \quad (25)$$

We approximate $\ddot{\phi}^{(n)}(\alpha)$ by excluding the second derivative of $p(t)$ as suggested in [28, p. 683].

C. Optimization for the Sub-Problem (15)

One can solve (15) using any statistical image reconstruction algorithms with a slight modification for a shifted quadratic regularizer. OS approximation can be usually used to speed up convergence rate. Using this splitting, one may focus more computational resources on solving the sub-problem (15) instead of solving the sub-problem (14) so that one may achieve faster *overall* convergence rate.

We modified De Pierro's EM algorithm [3] for (15) by considering a simple shifted quadratic regularizer. The surrogate function for the likelihood term in (15) is

$$\mathrm{L}(\boldsymbol{y}|\boldsymbol{f}) \leq \sum_j \tilde{Q}_j \left(f_j | \boldsymbol{f}^{(n-1)} \right) \quad (26)$$

where $\tilde{Q}_j(f_j|\boldsymbol{f}^{(n-1)})$ is a surrogate function that can be found in [3] and [6]. We use a surrogate function $Q_j(f_j|\boldsymbol{f}^{(n-1)})$ that is equivalent to $\tilde{Q}_j(f_j|\boldsymbol{f}^{(n-1)})$ by omitting the terms that are independent of f_j as follows:

$$Q_j \left(f_j | \boldsymbol{f}^{(n-1)} \right) = a_j f_j - e_j \left(\boldsymbol{f}^{(n-1)} \right) f_j^{(n-1)} \log(f_j) \quad (27)$$

where

$$a_j = \sum_l a_{lj}$$

and

$$e_j(\boldsymbol{f}) = \sum_l a_{lj} \frac{y_l}{\bar{y}_l(\boldsymbol{f})}. \quad (28)$$

Note that $\tilde{Q}_j(f_j|\boldsymbol{f}^{(n-1)})$ and $Q_j(f_j|\boldsymbol{f}^{(n-1)})$ are minimized at the same f_j.

The regularizer in (15) is separable in the image domain

$$\frac{\mu}{2} \left\| \boldsymbol{f} - \boldsymbol{u}^{(n)} - \boldsymbol{d}^{(n-1)} \right\|^2 = \sum_j \frac{\mu}{2} \left(f_j - u_j^{(n)} - d_j^{(n-1)} \right)^2. \quad (29)$$

Therefore, one must minimize the following surrogate function to solve the sub-problem in (15):

$$Q_j \left(f_j | \boldsymbol{f}^{(n-1)} \right) + \frac{\mu}{2} \left(f_j - u_j^{(n)} - d_j^{(n-1)} \right)^2 \quad (30)$$

for all j. Differentiating (30) with respect to f_j and setting it to be zero leads to the following second-order polynomial with respect to f_j

$$\mu f_j^2 + \left(a_j - \mu \left(u_j^{(n)} + d_j^{(n-1)} \right) \right) f_j$$
$$- e_j \left(\boldsymbol{f}^{(n-1)} \right) f_j^{(n-1)} = 0. \quad (31)$$

The nonnegative root of (31) is the minimizer of (30). This root always exists because $e_j(\boldsymbol{f}^{(n-1)})$ and $f_j^{(n-1)}$ are nonnegative [29].

An OS approximation for this modified De Pierro's algorithm can be easily done by substituting the most time-consuming part (28) in (15) at each iteration with the following approximate term

$$e_j(\boldsymbol{f}) \approx K \sum_{l \in \Omega_k} a_{lj} \frac{y_l}{\bar{y}_l(\boldsymbol{f})}. \quad (32)$$

This new term (32) requires about K times less computation than the term (28) does. Since calculating (28) dominates the overall calculation of (31) for each iteration, OS approximation for (15) substantially reduces the computation time per update.

In this Section, we proposed to use variable splitting and ADMM for efficient computation of NL regularized image reconstruction in (14), (15), and (16). We also described detailed algorithms for each sub-problem: gradient descent with one step of Newton's method for (14) and modified De Pierro's OSEM algorithm for (15). Even though our proposed ADMM algorithm for convex R guarantees to converge for any μ, the parameter μ in (13) affects convergence rate. In the next Section, we propose a method to optimize μ automatically.

IV. PARAMETER SELECTION FOR ADMM

A. Ideal ADMM Update and Approximation

Ghadimi *et al.* optimized μ in ADMM for ℓ_2-regularized minimization with a quadratic data fitting term [13]. We generalize and extend their analysis to optimize μ for our case of having NL regularizers with the negative Poisson likelihood term.

Let us derive the "ideal" ADMM update for (14), (15), and (16). Using the gradient of (14) with respect to $\boldsymbol{u}$, the first update for (14) at the nth iteration is

$$\boldsymbol{u}^{(n)} = \left[\boldsymbol{I} + \frac{\beta}{\mu} \boldsymbol{R} \left(\boldsymbol{u}^{(n-1)} \right) \right]^{-1} \left(\boldsymbol{f}^{(n-1)} - \boldsymbol{d}^{(n-1)} \right) \quad (33)$$

where the approximate Hessian of (4) is

$$\boldsymbol{R}(\boldsymbol{u}) \triangleq \sum_{i,j \in S_i} 2\dot{p}_{ij} \left(\|(\boldsymbol{N}_i - \boldsymbol{N}_j)\boldsymbol{u}\|^2 \right) (\boldsymbol{N}_i - \boldsymbol{N}_j)'(\boldsymbol{N}_i - \boldsymbol{N}_j).$$

We used the GD algorithm in (17) since (33) is impractical. Similarly, using the gradient of (15) with respect to $\boldsymbol{f}$ and ignoring the non-negative constraint for $\boldsymbol{f}$, "ideally" the update for (15) at the nth iteration is

$$\boldsymbol{f}^{(n)} = \left[\boldsymbol{A}'\boldsymbol{W}^{(n-1)}\boldsymbol{A} + \mu \boldsymbol{I} \right]^{-1}$$
$$\times \left(\boldsymbol{A}'\boldsymbol{W}^{(n-1)}(\boldsymbol{y} - \boldsymbol{s}) + \mu \boldsymbol{u}^{(n)} + \mu \boldsymbol{d}^{(n-1)} \right) \quad (34)$$

where

$$\boldsymbol{W}^{(n-1)} \triangleq \boldsymbol{D} \left(\frac{1}{\bar{y}_l \left(\boldsymbol{f}^{(n-1)} \right)} \right).$$

Because (34) is impractical, we used De Pierro's EM algorithm in (31). We can use (16) itself as the third "ideal" ADMM update at the nth iteration.

To facilitate the analysis using these ideal ADMM updates (33), (34), and (16), we need to fix $\boldsymbol{R}(\boldsymbol{u}^{(n-1)})$ and $\boldsymbol{W}^{(n-1)}$ for all iterations or for all n. We conjecture that we can set $\boldsymbol{u}^{(n-1)} = \boldsymbol{f}^{(n-1)} = \hat{\boldsymbol{f}}^{\mathrm{init}}$ for $\boldsymbol{R}(\boldsymbol{u}^{(n-1)})$ and $\boldsymbol{W}^{(n-1)}$ so that we can fix

$$\boldsymbol{R} \triangleq \boldsymbol{R}(\hat{\boldsymbol{f}}^{\mathrm{init}}) \tag{35}$$

and

$$\boldsymbol{W} \triangleq \boldsymbol{D}\left(\frac{1}{\bar{y}_l(\hat{\boldsymbol{f}}^{\mathrm{init}})}\right). \tag{36}$$

We used these approximations *only* for selecting μ. The initial image $\hat{\boldsymbol{f}}^{\mathrm{init}}$ can be anything like the FBP image or the reconstructed image using the OSEM with a few iterations. For the results in Section V, we used the reconstructed image using unregularized OSEM with five iterations (six subsets). Please see Appendix A for technical details for the approximation of (35) and (36).

B. Nearly Optimal ADMM Parameter

To study the convergence rate, we need to derive the update equation of $\boldsymbol{f}^{(n)}$ in terms of $\boldsymbol{f}^{(n-1)}$. Here, we generalize the approach of [13] for our "ideal" ADMM update equations for (14)–(16).

First, rearrange (34) with (36) and use it for (16). Then, (16) becomes

$$\boldsymbol{d}^{(n)} = \frac{1}{\mu} \boldsymbol{A}'\boldsymbol{W}\left[\boldsymbol{A}\boldsymbol{f}^{(n)} + \boldsymbol{s} - \boldsymbol{y}\right] \tag{37}$$

and this also works for $n - 1$. Secondly, we use (37) with $n - 1$ for (14) with (35) to obtain

$$\boldsymbol{u}^{(n)} = \left[\boldsymbol{I} + \frac{\beta}{\mu}\boldsymbol{R}\right]^{-1}\left(\boldsymbol{I} - \frac{1}{\mu}\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}\right)\boldsymbol{f}^{(n-1)} + \boldsymbol{c} \tag{38}$$

where $\boldsymbol{c}$ is a constant vector. Lastly, combining (37) (with $n-1$), (38), and (34) [with (36)] yields

$$\boldsymbol{f}^{(n)} = \boldsymbol{S}_\mu \boldsymbol{f}^{(n-1)} + \tilde{\boldsymbol{c}} \tag{39}$$

where $\tilde{\boldsymbol{c}}$ is a constant vector and

$$\boldsymbol{S}_\mu \triangleq [\boldsymbol{H} + \mu\boldsymbol{I}]^{-1}[\beta\boldsymbol{R} + \mu\boldsymbol{I}]^{-1}(\beta\boldsymbol{R}\boldsymbol{H} + \mu^2\boldsymbol{I}) \tag{40}$$

where $\boldsymbol{H} \triangleq \boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}$.

With the goal of approximately optimizing the convergence rate, we choose the ADMM parameter μ as follows:

$$\mu^* = \arg\min_\mu \rho(\boldsymbol{S}_\mu) \tag{41}$$

where $\rho(\boldsymbol{S}_\mu)$ is the spectral radius of the matrix $\boldsymbol{S}_\mu$, which is the maximum eigen value of the matrix $\boldsymbol{S}_\mu$. However, it is very challenging to find eigen values for large matrices such as $\boldsymbol{S}_\mu$ and it is also infeasible to find the matrix $\boldsymbol{S}_\mu$ itself due to the inversion of large matrices such as $\boldsymbol{H} + \mu\boldsymbol{I}$ and $\beta\boldsymbol{R} + \mu\boldsymbol{I}$.

To make the problem (41) tractable, we approximately optimize the ADMM parameter μ for the center of the image and use that parameter everywhere else. We approximate $\boldsymbol{H}$ and $\boldsymbol{R}$ as being locally circulant around that local area; a similar assumption often is used when analyzing spatial resolution properties [30], [31]. Then, the eigen value of the matrix $\boldsymbol{S}_\mu$ for the pth eigen vector (or the pth discrete Fourier basis) becomes

$$\lambda_p(\boldsymbol{S}_\mu) = \frac{\beta r_p h_p + \mu^2}{(h_p + \mu)(\beta r_p + \mu)} \tag{42}$$

where r_p and h_p are the eigen values of $\boldsymbol{R}$ and $\boldsymbol{H}$ for the corresponding pth eigen vector, respectively. Then, our choice for the ADMM parameter μ is

$$\mu^o = \arg\min_{\mu \in \Lambda} \max_p \lambda_p(\boldsymbol{S}_\mu). \tag{43}$$

Unlike the case in [13], it is not easy to find the analytical solution for (43). Instead, we can quickly solve the problem (43). We find r_p and h_p for all p using the fast Fourier transform (FFT) of $\boldsymbol{R}\mathbf{1}_c$ and $\boldsymbol{H}\mathbf{1}_c$, respectively, where $\mathbf{1}_c$ is a unit vector where the element of $\mathbf{1}_c$ is one at the center voxel of the image and zero otherwise. Then, we calculated $\max_p \lambda_p(\boldsymbol{S}_\mu)$ for a finite set of μ values on $\Lambda = [0, \max_p \sqrt{\beta r_p h_p}]$, and then obtain μ^o from $\max_p \lambda_p(\boldsymbol{S}_\mu)$. Since $\lambda_p(\boldsymbol{S}_\mu)$ is a nondecreasing function of μ for $\mu > \sqrt{\beta r_p h_p}$, μ^o must be less than or equal to $\max_p \sqrt{\beta r_p h_p}$. The computational cost for evaluating $\max_p \lambda_p(\boldsymbol{S}_\mu)$ is linear in the number of voxels. This procedure is fast since it is parallelizable. One can speed up this procedure using coarse-to-fine approach. One could also use the Golden section search [28, p. 397] to find μ^o to minimize $\max_p \lambda_p(\boldsymbol{S}_\mu)$.

V. RESULT

A. Simulation Setup

We simulated the Siemens Symbia Truepoint 3-D SPECT system with high energy collimators (parallel hexagonal collimators with a septal thickness of 2 mm, a hole diameter of 4 mm, and a hole length of 59.7 mm) with a nonuniform attenuation map, depth-dependent collimator-detector response [20], and scatter component (128×21, 4.8^2 mm^2 pixel size). The system resolution at 10 cm was 13.4 mm and 60 views were collected around $360°$. We used the XCAT phantom [17] to generate the true SPECT image with activity distributions realistic for I-131 RIT. The dimension of the SPECT image was $128 \times 128 \times 21$, 4.8^3 mm^3 voxel size. Three spherical lesions were placed within the XCAT phantom with volumes 176 cc, 32 cc, and 9 cc as shown in the bottom-right figure of Fig. 4(a) or (b). Poisson noise was added after scaling the projections to the count-level corresponding to day 2 post-therapy in I-131 RIT (about 600 K total counts per slice with about 300 K scatter counts per slice).

We set the common regularization parameters for all optimization methods as follows: $\beta = 2^{-13}$, $\sigma_f = 2^{1.5}$, the patch size $3 \times 3 \times 3$, and the search neighborhood size $7 \times 7 \times 7$. Six subsets were used for OSEM and ADMM. ADMM separates the likelihood update and the regularizer update by split-
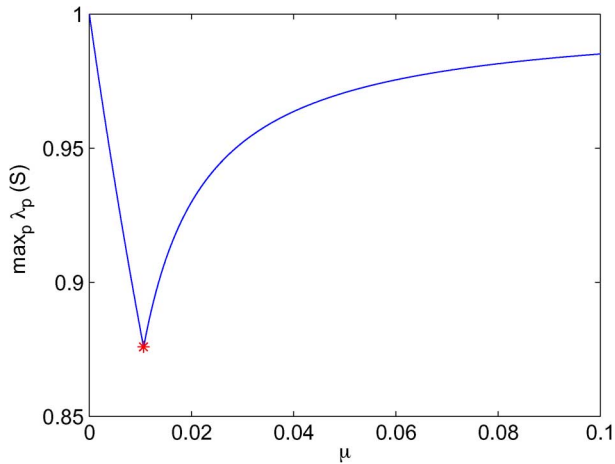
Fig. 1. Plot of $\max_p \lambda_p(\boldsymbol{S}_\mu)$ for $\mu = 0, 0.0001, 0.0002, \cdots, 0.1$. Red star ($*$) denotes our choice of μ^o.
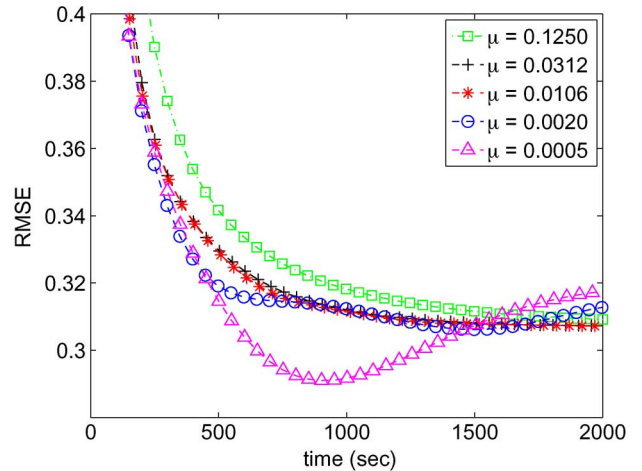

Fig. 2. RMSEs of estimated images over time using ADMM with different μ values. Atomatically selected value ($\mu = 0.0106$, red star *) yielded relatively fast convergence rate of ADMM compared to other choices of μ.

ting and we chose to run more sub-iterations for the likelihood update (2 outer-iterations ×6 subsets) than for the regularizer update (one outer-iteration). We used six threads for computation (Intel Core i7 2.8 GHz) for all methods and they used the same compiled ANSI C99 code to evaluate the cost function and the gradient of the cost function. We measured the normalized root mean square error (RMSE) over the whole image at each (outer) iteration

$$\text{RMSE} = \frac{\|\hat{\boldsymbol{f}} - \boldsymbol{f}_{\text{TRUE}}\|}{\|\boldsymbol{f}_{\text{TRUE}}\|}. \tag{44}$$

B. Parameter Selection

There are parameters to tune for each optimization method. We selected the step size for GD and the number of past estimated images for hessian approximation of L-BFGS-B. We chose those values to yield the fastest convergence rate empirically: the step size for GD was 0.04 and the number of past estimated images for L-BFGS-B was 5. In our experiment, not shown here, the step size for GD was critical for convergence. GD diverged with too large step size and GD converged slowly with too small step size. However, the number of past images used for L-BFGS-B did not affect convergence rate much.

We selected the ADMM parameter μ using (43). We first obtained $\max_p \lambda_p(\boldsymbol{S}_\mu)$ for $\mu = 0, 0.0001, 0.0002, \cdots, 0.2125$ as shown in Fig. 1 where $\max_p \sqrt{\beta r_p h_p} = 0.2125$. Based on this plot, we chose the ADMM parameter μ to be 0.0106, which is the red star mark ($*$) in Fig. 1.

We evaluated the ADMM convergence rate as a function of μ empirically. Fig. 2 shows that our choice of ADMM parameter $\mu = 0.0106$ achieved reasonably fast convergence compared to other choices of μ. Too large μ such as $\mu = 0.1250$ (green square mark) yielded slower convergence rate and too small μ value such as $\mu = 0.0005$ (magenta triangle mark) resulted in fluctuating tails. Due to the approximations such as local shift invariance and fixing some matrices like $\boldsymbol{R}$ and $\boldsymbol{W}$, we can not claim that our choice of μ is optimal, but Fig. 2 suggests that our choice is adequate for fast convergence of ADMM.
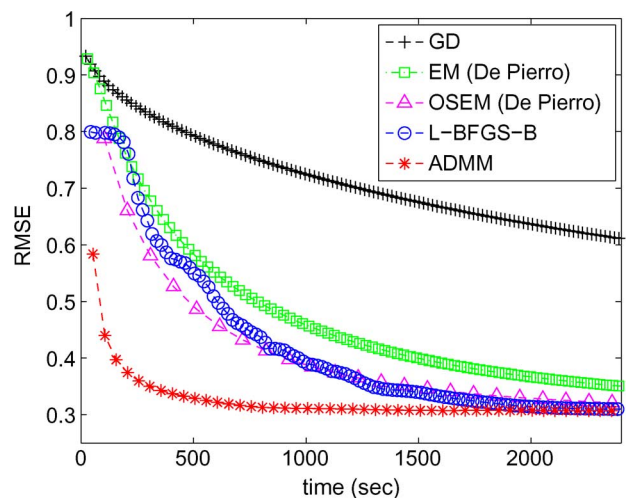

Fig. 3. RMSEs of estimated images using different algorithms versus time for NL regularization with the Fair potential. Proposed ADMM showed faster convergence rate than other methods.

C. Simulation Results for NL Regularization

We reconstructed images using different optimization algorithms for the cost function (1) with the (convex) Fair potential (8).

Fig. 3 shows the plots of RMSE values versus computation time for different methods: GD, EM, and OSEM using De Pierro's lemma, L-BFGS-B, and proposed ADMM. EM yielded faster convergence rate than GD with fixed step size. OSEM does improve convergence speed as compared to EM, but it provides little acceleration due to computationally expensive NL regularizer calculation for all sub-iterations. L-BFGS-B yielded similar convergence rate to OSEM with six subsets. Our proposed ADMM substantially improved convergence speed over all other methods. Other methods did not reach the minimum RMSE before 2000 s, but ADMM achieved the minimum RMSE before 1000 s. These simulation results illustrate that repeated likelihood updates are more important for fast convergence than regularizer updates.
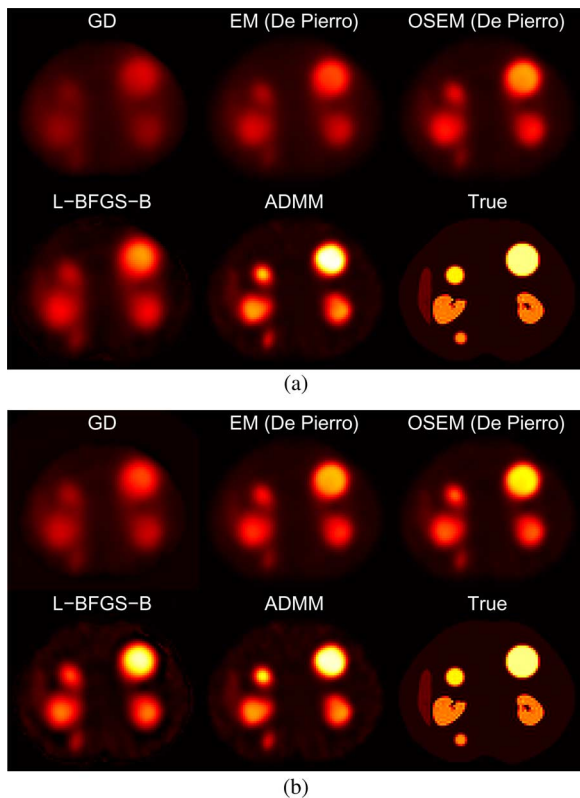
Fig. 4. Estimated images of different methods at 500 and 1000 s and the true image for the case of the Fair potential. ADMM yielded the best contrast recovery among all other methods. (a) 500 s. (b) 1000 s.

Fig. 4(a) and (b) shows reconstructed images by different methods at 500 and 1000 s and the true image for NL regularization with the Fair potential. At this early time of 500 s, ADMM yielded the best contrast recovery among all other methods. As time goes by (at 1000 s), other methods also started to yield similar images to that of ADMM since all optimization methods minimize the almost same cost function. However, they may not be exactly the same due to the OS approximation.

Fig. 5(a)–(c) shows recovery coefficients (RCs) of different methods over time when using the Fair potential. The larger the lesion is, the faster it approached to the achievable RC value. Note that we did not optimize NL regularizer parameters such as σ_f and one may achieve better RC for smaller lesions like 9 cc than that in Fig. 5(c).

VI. DISCUSSION

We developed a new algorithm for tomography with computationally expensive NL regularizers using ADMM. We also proposed a method to determine automatically a suitable μ value for fast convergence rate. By combining with the OS approach, our proposed ADMM approached convergence much faster than existing methods such as GD, EM, and OSEM using the De Pierro lemma, and L-BFGS-B. Since it seems more important to update the likelihood part frequently, our ADMM yielded faster convergence. Since the cost function has both the likelihood term and the regularization term, increasing the number of iterations for the likelihood term did not always further accelerate convergence. We chose a good combination of iterations
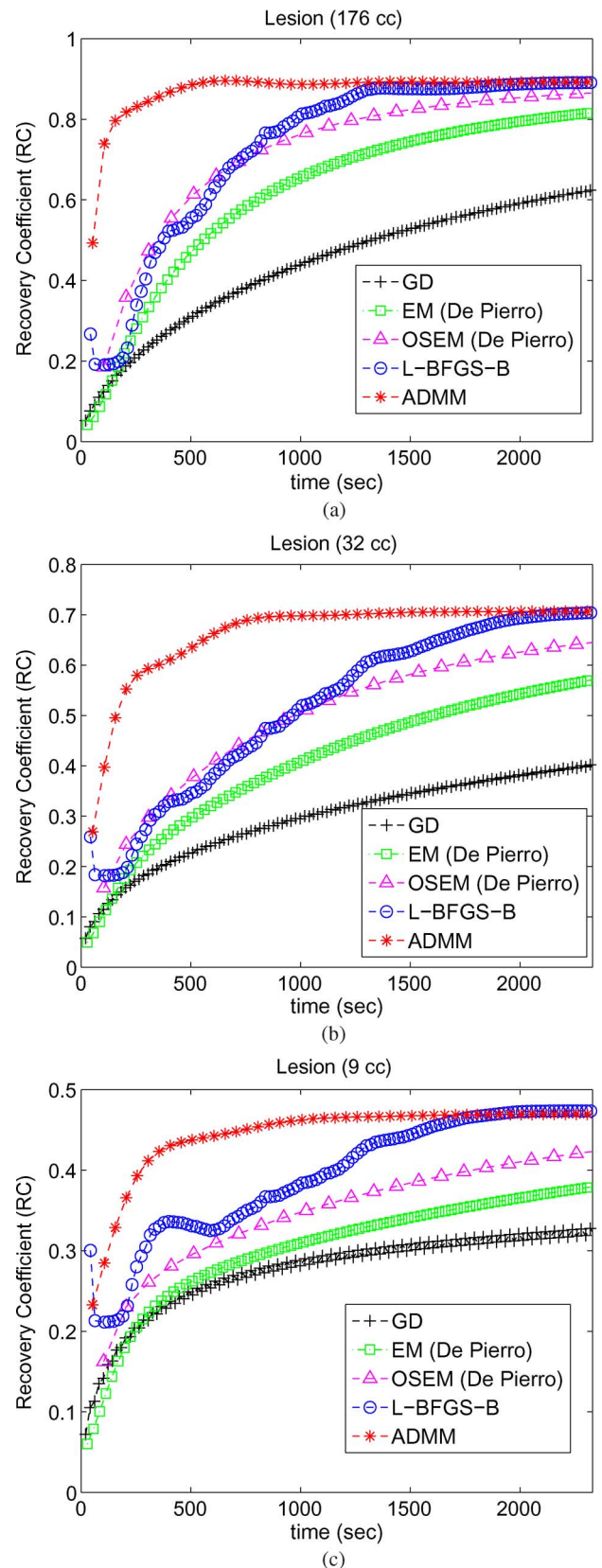

Fig. 5. Recovery coefficients for different size lesions over time when using the Fair potential. ADMM yielded the best recovery coefficients among all other methods. (a) 176 cc. (b) 32 cc. (c) 9 cc.

for both terms empirically. Similarly, one could use an approximate gradient for the NL regularizer to reduce computation, but

in our simulation not shown here, our ADMM outperformed this approximation.

In this paper, we minimized the same NL regularized cost function (the Fair potential) as Wang *et al.* did [6]. Whereas Wang *et al.* used De Pierro's algorithm with the surrogate function of their NL regularizer, we use De Pierro's algorithm with a shifted quadratic regularizer, which requires far less computation. Zhang *et al.* also applied a splitting approach to the optimization problem with NL regularizers [4]. However, their motivation for splitting was to apply shrinkage operator to the nonsmooth potential function such as TV. In addition, our way of splitting in (12) was different from theirs. Xu *et al.* used the same type of splitting as ours for the case of using nonsmooth regularizer [16]. They used a similar formula for the sub-problem of (15) except a Lagrangian multiplier vector, but their motivation was to deal with nonsmooth regularizer rather than to deal with computation-intensive NL regularizer.

Our approach to optimizing μ was based on 'ideal' updates. In other words, we assumed fully converged images for sub-iterations of (14) and (15). Nevertheless, our optimized μ worked well even for sub-iterations that did not converge fully. This may be because a few sub-iterations yielded good approximations of fully converged images for the sub-problems (14) and (15).

The proposed method worked well for SPECT image reconstruction with the patch-based convex Fair potential function [6]. Our proposed method can be easily extended to other computationally expensive NL regularizers [4], [5], [7] and NL regularizers that use high-resolution side information [9]–[11] for both emission and transmission tomography. Even though ADMM works only with convex cases *theoretically*, our proposed ADMM can be a practical method for many nonconvex NL regularizers. Improving image quality with proper regularizers and appropriate regularization parameter selection using our fast ADMM algorithm will be an important and interesting future work.

APPENDIX A
APPROXIMATION FOR IDEAL ADMM UPDATE

For $\boldsymbol{R}$, even though $\hat{\boldsymbol{f}}^{\text{init}}$ and $\boldsymbol{u}^{(n-1)}$ differ substantially, the patch selection operator $\boldsymbol{N}_i - \boldsymbol{N}_j$ and the ℓ_2 norm can make this difference much smaller. Therefore, it seems reasonable to use the approximation (35). For $\boldsymbol{W}$, this type of approximation in (36) has been used in different gradient-based analysis such as mean-variance analysis [32], spatial resolution analysis [30], [31], and noise analysis [33]. Even though there may be non-negligible difference between $\hat{\boldsymbol{f}}^{\text{init}}$ and $\boldsymbol{f}^{(n-1)}$, the approximation $\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} \approx \boldsymbol{A}'\boldsymbol{W}^{(n-1)}\boldsymbol{A}$ still holds fairly well. Thus, it also seems reasonable to use the approximation (36).

ACKNOWLEDGMENT

REFERENCES

[1] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Stat. Soc. Ser. B* vol. 39, no. 1, pp. 1–38, 1977 [Online]. Available: http://www.jstor.org/stable/info/2984875?seq=1

[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. 601–609, Dec. 1994.

[3] A. R. De Pierro, "A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography," *IEEE Trans. Med. Imag.*, vol. 14, no. 1, pp. 132–137, Mar. 1995.

[4] X. Zhang, M. Burger, X. Bresson, and S. Osher, "Bregmanized nonlocal regularization for deconvolution and sparse reconstruction," *SIAM J. Imag. Sci.*, vol. 3, no. 3, pp. 253–276, 2010.

[5] Y. Lou, X. Zhang, S. Osher, and A. Bertozzi, "Image recovery via non-local operators," *J. Sci. Comput.*, vol. 42, no. 2, pp. 185–197, Feb. 2010.

[6] G. Wang and J. Qi, "Penalized likelihood PET image reconstruction using patch-based edge-preserving regularization," *IEEE Trans. Med. Imag.*, vol. 31, no. 12, pp. 2194–2204, Dec. 2012.

[7] Z. Yang and M. Jacob, "Nonlocal regularization of inverse problems: A unified variational framework," *IEEE Trans. Image Process.*, vol. 22, no. 8, pp. 3192–3203, Aug. 2013.

[8] F. Rousseau, "A non-local approach for image super-resolution using intermodality priors," *Med. Image Anal.*, vol. 14, no. 4, pp. 594–605, Aug. 2010.

[9] S. Y. Chun, J. A. Fessler, and Y. K. Dewaraja, "Non-local means methods using CT side information for I-131 SPECT image reconstruction," in *Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf.*, 2012, pp. 3362–3366.

[10] K. Vunckx, A. Atre, K. Baete, A. Reilhac, C. Deroose, K. V. Laere, and J. Nuyts, "Evaluation of three MRI-based anatomical priors for quantitative pet brain imaging," *IEEE Trans. Med. Imag.*, vol. 31, no. 3, pp. 599–612, Mar. 2012.

[11] V.-G. Nguyen and S.-J. Lee, "Incorporating anatomical side information into PET reconstruction using nonlocal regularization," *IEEE Trans. Image Process.*, vol. 22, no. 10, pp. 3961–3973, Oct. 2013.

[12] S. Y. Chun, Y. K. Dewaraja, and J. A. Fessler, "Alternating direction method of multiplier for emission tomography with non-local regularizers," in *Proc. Int. Meet. Fully 3-D Image Recon. Rad. Nucl. Med*, 2013, pp. 62–65 [Online]. Available: proc/13/web/chun-13-adm.pdf

[13] E. Ghadimi, A. Teixeira, I. Shames, and M. Johansson, "On the optimal step-size selection for the alternating direction method of multipliers," in *Proc. IFAC Workshop Estimat. Control Netw. Syst.*, 2012 [Online]. Available: www.ee.kth.se/mikaelj/papers/gtsj12.pdf

[14] T. Goldstein and S. Osher, "The split Bregman method for L1-regularized problems," *SIAM J. Imag. Sci.*, vol. 2, no. 2, pp. 323–343, 2009.

[15] M. A. T. Figueiredo and J. M. Bioucas-Dias, "Restoration of Poissonian images using alternating direction optimization," *IEEE Trans. Image Process.*, vol. 19, no. 12, pp. 3133–3145, Dec. 2010.

[16] J. Xu, S. Chen, and B. M. W. Tsui, "Total variation penalized maximum-likelihood image reconstruction for a stationary small animal SPECT system," in *Proc. Int. Meet. Fully 3-D Image Recon. Rad. Nucl. Med*, 2011, pp. 225–228.

[17] W. P. Segars, M. Mahesh, T. J. Beck, E. C. Frey, and B. M. W. Tsui, "Realistic CT simulation using the 4D XCAT phantom," *Med. Phys.*, vol. 35, no. 8, pp. 3800–3808, Aug. 2008.

[18] J. L. Morales and J. Nocedal, "Remark on "algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound constrained optimization"," *ACM Trans. Math. Softw.*, vol. 38, no. 1, pp. 7:1–7:4, Nov. 2011.

[19] Y. K. Dewaraja, M. J. Schipper, P. L. Roberson, S. J. Wilderman, H. Amro, D. D. Regan, K. F. Koral, M. S. Kaminski, and A. M. Avram, "131I-tositumomab radioimmunotherapy: Initial tumor dose-response results using 3-dimensional dosimetry including radiobiologic modeling," *J. Nucl. Med.*, vol. 51, no. 7, pp. 1155–1162, Jul. 2010.

[20] S. Chun, J. Fessler, and Y. Dewaraja, "Correction for collimator-detector response in SPECT using point spread function template," *IEEE Trans. Med. Imag.*, vol. 32, no. 2, pp. 295–305, Feb. 2013.

[21] Y. K. Dewaraja, M. Ljungberg, and J. A. Fessler, "3-D Monte Carlo-based scatter compensation in quantitative I-131 SPECT reconstruction," *IEEE Trans. Nucl. Sci.*, vol. 53, no. 1, pp. 181–188, Feb. 2006.

[22] M. Rivera and J. L. Marroquin, "Efficient half-quadratic regularization with granularity control," *Imag Vision Comput.*, vol. 21, no. 4, pp. 345–357, Apr. 2003.

[23] R. C. Fair, "On the robust estimation of econometric models," *Ann. Econ. Social Measur.* vol. 2, pp. 667–677, Oct. 1974 [Online]. Available: http://fairmodel.econ.yale.edu/rayfair/pdf/1974D.HTM

[24] K. Lange, "Convergence of EM image reconstruction algorithms with Gibbs smoothing," *IEEE Trans. Med. Imag.*, vol. 9, no. 4, pp. 439–446, Dec. 1990.

[25] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2010.

[26] J. Eckstein and D. P. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Programm.*, vol. 55, no. 1–3, pp. 293–318, Apr. 1992.

[27] J. A. Fessler and S. D. Booth, "Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction," *IEEE Trans. Image Process.*, vol. 8, no. 5, pp. 688–699, May 1999.

[28] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical recipes in C*, 2nd ed. New York: Cambridge Univ. Press, 1992.

[29] J. A. Fessler and A. O. Hero, "Penalized maximum-likelihood image reconstruction using space-alternating generalized EM algorithms," *IEEE Trans. Image Process.*, vol. 4, no. 10, pp. 1417–1429, Oct. 1995.

[30] J. A. Fessler and W. L. Rogers, "Spatial resolution properties of penalized-likelihood image reconstruction methods: Space-invariant tomographs," *IEEE Trans. Image Process.*, vol. 5, no. 9, pp. 1346–1358, Sep. 1996.

[31] S. Y. Chun and J. A. Fessler, "Spatial resolution properties of motion-compensated image reconstruction methods," *IEEE Trans. Med. Imag.*, vol. 31, no. 7, pp. 1413–1425, Jul. 2012.

[32] J. A. Fessler, "Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): Applications to tomography," *IEEE Trans. Image Process.*, vol. 5, no. 3, pp. 493–506, Mar. 1996.

[33] S. Y. Chun and J. A. Fessler, "Noise properties of motion-compensated tomographic image reconstruction methods," *IEEE Trans. Med. Imag.*, vol. 32, no. 2, pp. 141–152, Feb. 2013.