Alternating Direction Method of Multiplier for Emission Tomography with Non-Local Regularizers

Se Young Chun, Yuni K. Dewaraja, and Jeffrey A. Fessler

***Abstract*—The ordered subset expectation maximization (OSEM) algorithm provides a fast image reconstruction method for emission and transmission tomography such as SPECT, PET, and CT by approximating the gradient of a likelihood function using a subset of projections instead of using all projections. However, for computationally expensive regularizers such as patch-based non-local (NL) regularizers, OSEM does not help much to improve the speed of reconstruction because one evaluates the regularizer gradient 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). This new scheme allows us to run more sub-iterations for the optimization related to the likelihood term. We evaluated our ADMM for 3D SPECT image reconstruction with the patch-based NL regularizer that uses the Fair potential. 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 (L-BFGS-B) algorithm.**

***Index Terms*—ordered-subset expectation-maximization, non-local regularizer, emission tomography, alternating direction method of multiplier**

I. Introduction

Statistical image reconstruction methods such as the expectation-maximization (EM) algorithm can improve quality of images for emission tomography such as PET and SPECT as compared to the analytical image reconstruction such as the filtered back-projection (FBP) [1]. It started to be used widely in clinics and in commercial PET and SPECT scanners after the fast algorithm called ordered-subset expectation-maximization (OSEM) was developed [2]. By approximating the gradient of a likelihood function using the subset of projections instead of using all projections, OSEM algorithm performed faster image reconstruction. This approximation has been used for unregularized emission tomographic image reconstruction [2] and regularized emission and transmission tomographic image reconstruction using simple quadratic or edge-preserving regularizers. Since the computation cost for these regularizers is fairly low compared to that for the likelihood term, the OSEM algorithm could often also speed up penalized likelihood (PL) image reconstruction.

Recently, patch-based non-local (NL) regularizers have been proposed that improve image quality compared to other conventional regularizers such as quadratic or edge-preserving functions in general image processing [3], PET reconstruction [4], and MRI reconstruction [5]. The same principle has been used for emission image reconstruction or super resolution using high resolution CT or MRI side information [6]–[9]. For emission tomography problems such as [4], [7]–[9], many optimization algorithms were used for image reconstruction such as the gradient descent (GD) [8], the EM (or OSEM) algorithm from the optimization transfer using De Pierro's lemma [4], the EM algorithm using one-step late approach [9], and the quasi-Newton algorithm called the Limited-memory Broyden-Fletcher-Goldfarb-Shanno with a box constraint (L-BFGS-B) [7]. Since the computation cost of the NL regularizers is very high compared to that of the likelihood, the OS does not help much to improve the convergence rate of PL image reconstruction.

In this paper, 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 multipliers (ADMM). This new scheme allows us to run more sub-iterations for the optimization related to the likelihood term. There have been some methods to use a variable splitting for the data fidelity term and the regularizer term [3], [10]–[12]. However, these previous methods split the variable to deal with non-smooth regularizers such as the total variation and to solve the sub-problem related to the regularizers using efficient methods such as shrinkage. Our proposed variable splitting has different motivation. We divide the original optimization into a few sub problems and we update the sub problem related to the NL regularizer less often.

We evaluated our new ADMM for 3D SPECT image reconstruction with a patch-based NL regularizer that uses the Fair potential [4]. Our XCAT phantom-based simulation [13] shows that our proposed ADMM improved the speed of convergence substantially compared to existing methods such as GD, EM and OSEM using De Pierro's approach, and the L-BFGS-B algorithm.

II. Method

A. Statistical image reconstruction for emission tomography

Statistical image reconstruction methods for emission tomography yield better image quality than non-iterative algorithms. The usual form of statistical image reconstruction is

This work was supported in part by NIH grant 2RO1 EB001994.

Se Young Chun was with the University of Michigan, Department of EECS and Radiology, Ann Arbor, MI 48109-2122, USA, and is now with Ulsan National Institute of Science and Technology (UNIST), Department of ECE, Ulsan, South Korea. (e-mail: delight@umich.edu, sychun@unist.ac.kr).

Yuni K. Dewaraja is with the University of Michigan, Department of Radiology, Ann Arbor, MI 48109-2122, USA. (e-mail: yuni@umich.edu).

Jeffrey A. Fessler is with the University of Michigan, Department of EECS, Ann Arbor, MI 48109-2122, USA. (e-mail: fessler@umich.edu).

to perform the following constrained optimization with respect to an image $\boldsymbol{f}$:

$$\hat{\boldsymbol{f}} \triangleq \underset{\boldsymbol{f}\geq\boldsymbol{0}}{\operatorname{argmin}} \mathrm{L}(\boldsymbol{y}|\boldsymbol{f}) \tag{1}$$

where $\boldsymbol{y}$ is a measured sinogram data and L denotes a negative Poisson log-likelihood function. The negative Poisson log-likelihood for emission tomography is defined as follows:

$$\mathrm{L}(\boldsymbol{y}|\boldsymbol{f}) = \sum_i \bar{y}_i(\boldsymbol{f}) - y_i \log \bar{y}_i(\boldsymbol{f}) \tag{2}$$

where y_i is the ith element of the measurement $\boldsymbol{y}$ and

$$\bar{y}_i(\boldsymbol{f}) \triangleq [\boldsymbol{A}\boldsymbol{f}]_i + s_i$$

where $\boldsymbol{A}$ denotes the system model and s_i is a scatter component for the ith measurement.

For SPECT imaging, we can incorporate an attenuation map and a depth-dependent point spread function model including penetration tails [14] in the system matrix $\boldsymbol{A}$. In our simulation, we assumed known s_i, but in practice, this scatter component can be estimated by using a triple energy window (TEW) method or by Monte Carlo methods [15].

Unregularized image reconstruction in (1) is *ill-posed*. In this case, converged recontructed images are very noisy. There are usually three approaches to deal with this noise: First of all, one can stop iteration before convergence. However, more iteration may be necessary for recovering high-frequency information (*e.g.* details) of image. Secondly, one can use a post-reconstruction filter (*e.g.*, Gaussian filter) to reduce noise. Lastly, one can add a regularizer to (1) (*e.g.*, quadratic roughness penalty, non-local regularizer). When using non-local regularizers for 3D images, the computation-complexity is usually very high.

B. Non-local regularizer

Recently, NL regularizers have been shown to yield high-quality images in many image reconstruction problems [3]–[5], [16]. A NL regularizer R can be added to (1) as follows:

$$\hat{\boldsymbol{f}} \triangleq \underset{\boldsymbol{f}\geq\boldsymbol{0}}{\operatorname{argmin}} \mathrm{L}(\boldsymbol{y}|\boldsymbol{f}) + \beta \mathrm{R}(\boldsymbol{f}) \tag{3}$$

where β is a regularization parameter and

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

$\|\cdot\|$ is the L_2 norm, $\boldsymbol{N}_i$ is an operator on the image $\boldsymbol{f}$ such that $\boldsymbol{N}_i\boldsymbol{f}$ is a vector of image intensities that are on the cube-shaped patch around the ith voxel, and p is any potential function. A typical choice for the function p is [3], [16]

$$p(t) = \exp\left(-\frac{\|\boldsymbol{N}_i\tilde{\boldsymbol{f}} - \boldsymbol{N}_j\tilde{\boldsymbol{f}}\|^2}{2\sigma_{\mathrm{f}}^2 N_{\mathrm{f}}}\right)\frac{t}{2N_{\mathrm{f}}} \tag{5}$$

where $\tilde{\boldsymbol{f}}$ is an initial image from any analytical image reconstruction (*e.g.*, filtered back projection) [16] or an estimated image from the previous iteration $\boldsymbol{f}^{(n)}$ [3]. In this case, p depends on i and j. Yang *et al*. proposed to use a few non-convex potentials including the Welsh potential [17]

$$p(t) = \sigma_{\mathrm{f}}^2\left(1 - \exp\left(-\frac{t}{2\sigma_{\mathrm{f}}^2 N_{\mathrm{f}}}\right)\right). \tag{6}$$

Wang *et al*. proposed to use the Fair potential [18], [19]

$$p(t) = \sigma_{\mathrm{f}}^2\left(\sqrt{\frac{t}{\sigma_{\mathrm{f}}^2 N_{\mathrm{f}}}} + \log\left(1 + \sqrt{\frac{t}{\sigma_{\mathrm{f}}^2 N_{\mathrm{f}}}}\right)\right). \tag{7}$$

Note that both (6) and (7) do not depend on an initial image and (7) is convex while (6) is non-convex. It has been reported that non-convex functions yielded better image quality than a convex function [5].

We can also design NL regularizers that can incorporate high-resolution side information such as CT or MR images [7]–[9] for better image-quality. In this paper, we focus on the Fair potential in (7), but the proposed algorithm can be applied to any regularizer.

C. Alternating direction method of multipliers

We split the variable for the likelihood term and the regularizer term by replacing (3) with the following equivalent constrained optimization problem:

$$\hat{\boldsymbol{f}} \triangleq \underset{\boldsymbol{f}\geq\boldsymbol{0},\boldsymbol{u}}{\operatorname{argmin}} \mathrm{L}(\boldsymbol{y}|\boldsymbol{f}) + \beta \mathrm{R}(\boldsymbol{u}),\ \text{sub. to}\ \boldsymbol{u} = \boldsymbol{f}. \tag{8}$$

By adding the augmented Lagrangian term, (8) becomes

$$\mathrm{L}(\boldsymbol{y}|\boldsymbol{f}) + \beta \mathrm{R}(\boldsymbol{u}) + \frac{\mu}{2}\|\boldsymbol{f} - \boldsymbol{u} - \boldsymbol{d}\|^2 \tag{9}$$

where μ is a scalar value (design parameter) and $\boldsymbol{d}$ is a Lagrangian multiplier vector.

We can solve this optimization problem (9) by using the ADMM algorithm [20], [21] as follows:

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

$$\boldsymbol{u}^{(n+1)} \in \underset{\boldsymbol{u}}{\operatorname{argmin}} \frac{\mu}{2}\|\boldsymbol{u} - \boldsymbol{f}^{(n)} + \boldsymbol{d}^{(n)}\|^2 + \beta \mathrm{R}(\boldsymbol{u}) \tag{10}$$

$$\boldsymbol{f}^{(n+1)} \in \underset{\boldsymbol{f}\geq\boldsymbol{0}}{\operatorname{argmin}} \mathrm{L}(\boldsymbol{y}|\boldsymbol{f}) + \frac{\mu}{2}\|\boldsymbol{f} - \boldsymbol{u}^{(n+1)} - \boldsymbol{d}^{(n)}\|^2 \tag{11}$$

$$\boldsymbol{d}^{(n+1)} = \boldsymbol{d}^{(n)} - (\boldsymbol{f}^{(n+1)} - \boldsymbol{u}^{(n+1)}) \tag{12}$$

End

where $\boldsymbol{f}^{(n)}$ is an estimated vector value $\boldsymbol{f}$ at the nth iteration. We can solve the sub-problems of (10) and (11) using existing methods.

We used the GD algorithm to solve (10) as follows:

$$\boldsymbol{u}^{(n+1)} = \boldsymbol{u}^{(n)} - \alpha \nabla \Phi^{(n)}(\boldsymbol{u}^{(n)}) \tag{13}$$

where α is a step size and

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

We plug (13) into (10) to determine the step size as follows:

$$\alpha \in \underset{\alpha}{\operatorname{argmin}}\, \phi^{(n)}(\alpha) \tag{14}$$

where $\phi^{(n)}(\alpha) \triangleq \Phi^{(n)}(\boldsymbol{u}^{(n)} - \alpha \nabla \Phi^{(n)}(\boldsymbol{u}^{(n)}))$,

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

$$\nabla \mathrm{R}(\boldsymbol{u}) = \sum_{i,j\in\Omega_i} (\boldsymbol{N}_i - \boldsymbol{N}_j)' 2\dot{p}\left(\|\boldsymbol{N}_i\boldsymbol{u} - \boldsymbol{N}_j\boldsymbol{u}\|^2\right)(\boldsymbol{N}_i - \boldsymbol{N}_j)\boldsymbol{u},$$

and $\dot{p}(t)$ is the first order derivative of $p(t)$. Since solving (14) is an intermediate step of solving (10), we do not need to find

an exact α value to minimize (14). We chose to use one step of Newton's method for (14) as follows [22]:

$$\alpha = -\frac{\dot{\phi}^{(n)}(0)}{\ddot{\phi}^{(n)}(0)} \tag{15}$$

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

$$\dot{\phi}^{(n)}(0) = -\|\nabla\Phi^{(n)}(\boldsymbol{u}^{(n)})\|^2$$

and

$$\ddot{\phi}^{(n)}(0) \approx \nabla\Phi^{(n)}(\boldsymbol{u}^{(n)})' \left(\mu\nabla\Phi^{(n)} + \nabla \mathrm{R}(\nabla\Phi^{(n)})\right).$$

We approximate $\ddot{\phi}^{(n)}(\alpha)$ by not using the second derivative of $p(t)$ as suggested in [23, p. 683].

Equation (11) can be solved using De Pierro's EM algorithm [24] and OS approximation can be used to speed up the convergence rate. Whereas Wang *et al*. used De Pierro's algorithm with the surrogate function of their NL regularizer [4], we use De Pierro's algorithm with a shifted quadratic regularizer, which requires far less computation. One can find a similar formula for this sub-problem of (11) without a Lagrangian multiplier vector in [12].

III. Result

We simulated a 3D SPECT-CT system with the non-uniform attenuation map, collimator-detector response, and scatter component (128×21, $4.8^2\mathrm{mm}^2$ pixel size). We used the XCAT phantom [13] to generate the true SPECT image. The dimension of the SPECT image was $128 \times 128 \times 21$, $4.8^3\mathrm{mm}^3$ voxel size. We set the regularization parameters as follows: $\beta = 2^{-13}$, $\sigma_{\mathrm{f}} = 2^{1.5}$, the patch size $3 \times 3 \times 3$, the search neighborhood size $7 \times 7 \times 7$, five past estimated images for hessian approximation (L-BFGS-B), and $\mu = 2^{-7}$ (ADMM). Six subsets were used for OSEM and ADMM and 12 threads were used for computation (Intel Xeon 2.67GHz). A uniform initial image was used for all methods. We measured a normalized root mean square error (RMSE) for estimated images at all (outer) iterations and the definition of the RMSE is

$$\mathrm{RMSE} = \frac{\|\hat{\boldsymbol{f}} - \boldsymbol{f}_{\mathrm{TRUE}}\|}{\|\boldsymbol{f}_{\mathrm{TRUE}}\|}. \tag{16}$$

Fig. 1 shows the plots of RMSE values versus computation time for different methods: GD, EM and OSEM using De Pierro's lemma, L-BFGS-B [25], and proposed ADMM. OSEM does not show much speed up as compared to EM due to computationally expensive NL regularizer calculation for all sub-iterations. ADMM separates the likelihood update and the regularizer update by splitting and runs more sub-iterations for the likelihood update (2 outer-iterations $\times 6$ subsets) than for the regularizer update (1 outer-iteration). These simulation results illustrate that repeated likelihood updates are more important for fast convergence than regularizer updates.

Fig. 2 shows estimated images of different methods at 500 seconds and the true image. At this early time, ADMM yielded the best contrast recovery among all other methods. Fig. 3 shows estimated images of different methods at 1000 seconds and the true image. In this case, L-BFGS-B achieved contrast recovery similar to ADMM. Both ADMM and L-BFGS-B yielded better contrast recovery than other methods.

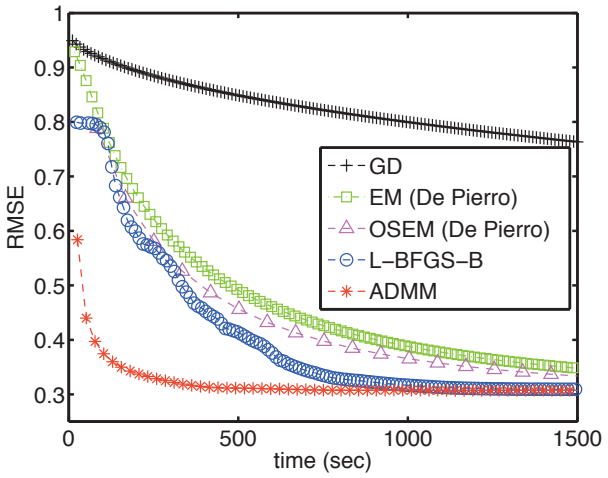

Fig. 1. RMSEs of estimated images using different algorithms over time. Proposed ADMM showed faster convergence rate than other previous methods.

IV. Discussion

We developed a new algorithm for emission tomography with computationally expensive NL regularizers using the ADMM. By combining with the OS approach, our proposed ADMM approached convergence much faster than existing methods such as GD, EM - De Pierro, OSEM - De Pierro, and L-BFGS-B. Since it seems more important to update the likelihood part frequently, our ADMM yielded faster convergence. Comparing our new method with other algorithms such as preconditioned conjugate gradient can be an interesting future work.

We demonstrated that our proposed method worked well for SPECT image reconstruction with the patch-based Fair potential function [4]. Our proposed method can be easily extended to other computationally expensive NL regularizers [3], [5], [16] and NL regularizers that use high-resolution side information [7]–[9] for both emission and transmission tomography.

ADMM requires a good μ value. Even though the theory of ADMM says that the algorithm will converge for any μ value, the choice of μ affects the convergence rate. Our future research will be an investigation on how to choose a good μ value.

Acknowledgment

The authors would like to thank Dr. Sathish Ramani for helpful discussion on ADMM.

References

[1] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Stat. Soc. Ser. B*, vol. 39, no. 1, pp. 1–38, 1977.

[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.

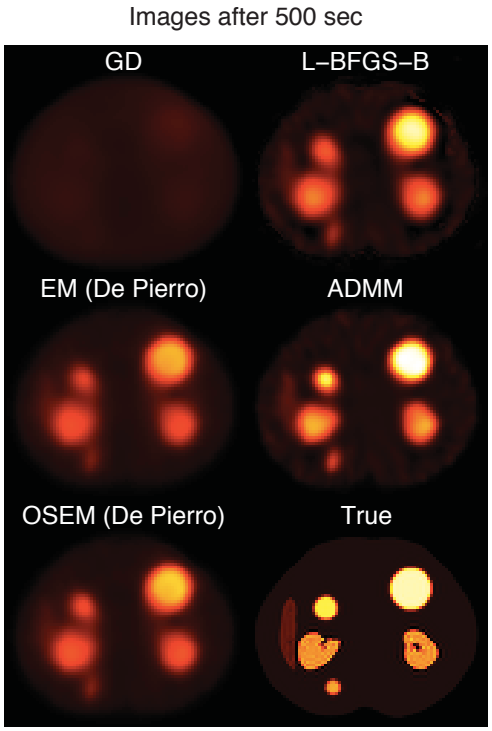

Fig. 2. Estimated images of different methods at 500 seconds and the true image. ADMM yielded the best contrast recovery among all other methods.

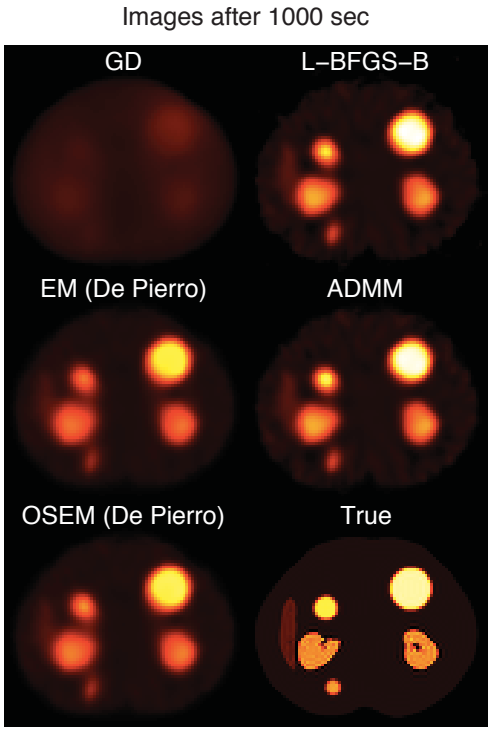

Fig. 3. Estimated images of different methods at 1000 seconds and the true image. L-BFGS-B achieved contrast recovery similar to ADMM. Both ADMM and L-BFGS-B yielded better contrast recovery than other methods.

[3] X. Zhang, M. Burger, X. Bresson, and S. Osher, "Bregmanized nonlocal regularization for deconvolution and sparse reconstruction," *SIAM J. Imaging Sci.*, vol. 3, no. 3, pp. 253–76, 2010.

[4] G. Wang and J. Qi, "Penalized likelihood pet image reconstruction using patch-based edge-preserving regularization," *Medical Imaging, IEEE Transactions on*, vol. 31, no. 12, pp. 2194 –2204, dec. 2012.

[5] Z. Yang and M. Jacob, "Nonlocal regularization of inverse problems: A unified variational framework," *IEEE Trans. Im. Proc.*, 2012.

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

[7] 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 Nuc. Sci. Symp. Med. Im. Conf.*, 2012, pp. 3362–6.

[8] 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. Imaging*, vol. 31, no. 3, pp. 599–612, 2012.

[9] V.-G. Nguyen and S.-J. Lee, "Anatomy-based PET image reconstruction using nonlocal regularization," in *Proc. SPIE 8313 Medical Imaging 2012: Phys. Med. Im.*, 2012.

[10] T. Goldstein and S. Osher, "The split Bregman method for L1-regularized problems," *SIAM J. Imaging Sci.*, vol. 2, no. 2, pp. 323–43, 2009.

[11] M. A. T. Figueiredo and J. M. Bioucas-Dias, "Restoration of Poissonian images using alternating direction optimization," *IEEE Trans. Im. Proc.*, vol. 19, no. 12, pp. 3133–45, Dec. 2010.

[12] 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. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, 2011, pp. 225–8.

[13] 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. 3800–8, Aug. 2008.

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

[15] Y. K. Dewaraja, M. Ljungberg, and J. A. Fessler, "3-D Monte Carlo-based scatter compensation in quantitative I-131 SPECT reconstruction," *IEEE Trans. Nuc. Sci.*, vol. 53, no. 1, pp. 181–8, Feb. 2006.

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

[17] M. Rivera and J. L. Marroquin, "Efficient half-quadratic regularization with granularity control," *Im. and Vision Computing*, vol. 21, no. 4, pp. 345–57, Apr. 2003.

[18] R. C. Fair, "On the robust estimation of econometric models," *Ann. Econ. Social Measurement*, vol. 2, pp. 667–77, Oct. 1974.

[19] K. Lange, "Convergence of EM image reconstruction algorithms with Gibbs smoothing," *IEEE Trans. Med. Imag.*, vol. 9, no. 4, pp. 439–46, Dec. 1990, corrections, T-MI, 10:2(288), June 1991.

[20] 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 in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.

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

[22] J. A. Fessler and S. D. Booth, "Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction," *IEEE Trans. Im. Proc.*, vol. 8, no. 5, pp. 688–99, May 1999.

[23] 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.

[24] 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. 132–7, Mar. 1995.

[25] 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.