A residual correction method for high-resolution PET reconstruction with application to on-the-fly Monte Carlo based model of positron range

Lin Fu and Jinyi Qi^{a)}

Department of Biomedical Engineering, University of California at Davis, GBSF 2303, Davis, California 95616

(Received 24 June 2009; revised 24 October 2009; accepted for publication 13 December 2009; published 21 January 2010)

Purpose: The quality of tomographic images is directly affected by the system model being used in image reconstruction. An accurate system matrix is desirable for high-resolution image reconstruction, but it often leads to high computation cost. In this work the authors present a maximum *a posteriori* reconstruction algorithm with residual correction to alleviate the tradeoff between the model accuracy and the computation efficiency in image reconstruction.

Methods: Unlike conventional iterative methods that assume that the system matrix is accurate, the proposed method reconstructs an image with a simplified system matrix and then removes the reconstruction artifacts through residual correction. Since the time-consuming forward and back projection operations using the accurate system matrix are not required in every iteration, image reconstruction time can be greatly reduced.

Results: The authors apply the new algorithm to high-resolution positron emission tomography reconstruction with an on-the-fly Monte Carlo (MC) based positron range model. Computer simulations show that the new method is an order of magnitude faster than the traditional MC-based method, whereas the visual quality and quantitative accuracy of the reconstructed images are much better than that obtained by using the simplified system matrix alone.

Conclusions: The residual correction method can reconstruct high-resolution images and is computationally efficient. © 2010 American Association of Physicists in Medicine.
[DOI: 10.1118/1.3284980]

Key words: residual correction, positron range modeling, on-the-fly Monte Carlo simulation, iterative image reconstruction

I. INTRODUCTION

Accurate system modeling is critical to the success of iterative image reconstruction in emission tomography. A number of studies have shown that accurate modeling of system response, such as spatial-variant geometric sensitivity and intercrystal penetration and scattering, can lead to superior image quality.^{1–11} In positron emission tomography (PET), highly accurate system matrices are important to take full advantage of high-resolution detector modules that are made of finely pixilated scintillator crystals. In particular, modeling of positron physics, such as positron range and photon non-collinearity, has attracted growing attention. As shown by several simulation studies, positron physics is becoming the major factor that limit the spatial resolution of future PET scanners.^{12–15}

While in theory, iterative reconstruction methods are amenable to arbitrary and complicated system models, practical implementations are often constrained by the associated high computing demands. It has been observed that using an accurate system matrix can drastically increase the computation time of image reconstruction.⁷ In general, as more system degrading (mostly blurring) factors are included, the system matrix becomes more ill-conditioned and has less degree of sparsity. The former leads to slower convergence rate, while the latter results in longer computation time per iteration. Both increase the total computational cost. For fast

image reconstruction, various approximations have been adopted to obtain a more sparse system matrix or a system matrix with a certain structure that allows fast forward and back projection. For example, many PET image reconstruction methods use a simple line integral model, which ignores the effects of positron range, non collinearity of photon pairs, variation in sensitivity along a line of response, and spatially variant inter-crystal penetration and scatter. While using a simpler system matrix leads to substantial savings in computation time, it impairs the potential for resolution recovery, especially for high-count PET data.

This paper presents a residual correction algorithm that can efficiently incorporate an accurate, but time-consuming, system model into iterative image reconstruction. Unlike conventional iterative methods which use a single system matrix in image reconstruction and assume it is accurate, the proposed method uses a pair of system matrices, one being accurate but time-consuming, such as a Monte Carlo (MC) based model, and the other being simplified but relatively fast, such as an ideal line integral model. First, an image is reconstructed using the simplified system matrix. Then, the reconstruction artifacts are removed through residual correction based on the error propagation from the system matrix to reconstructed image. When the pair of system matrices is properly constructed, the overall computation cost of the new

method can be much less than a conventional method that uses the accurate model alone. As an example, we apply the proposed method to maximum *a posteriori* (MAP) image reconstruction with an on-the-fly MC-based positron range model. Some preliminary results of this work have been presented at conferences.^{16,17} Here we include more details of theoretical analysis of the proposed algorithm and quantitative evaluations.

I.A. Motivation for Monte Carlo based positron range model

Among various image degrading factors in PET, positron range is probably the least understood one.¹³ Traditionally, positron range was not a major concern for lower-energy emitters such as ^{18}F ,¹⁸ but in a few recent simulation studies positron range has been identified as a limiting factor for new scanners with submillimeter spatial resolution and novel photon positioning methods.^{12,14,15} For long-ranged isotopes such as ^{82}Rb in cardiac imaging, positron range imposes even greater influence on overall system resolution.⁴ The blurring effect of positron range of several medically important isotopes has been measured by physical experiments,^{19,20} computed using analytical formulae,^{21,22} or simulated by MC software packages.^{13,23–25} In water medium, the blurring effect ranges from a few tenths of a millimeter for lower-energy emitters such as ^{18}F and ^{11}C , to several millimeters for higher-energy emitters, such as ^{68}Ga and ^{82}Rb .

A handful of works have attempted to model positron range and to remove its blurring effect through image processing. Most of these approaches use analytic space-invariant isotropic filters to deconvolve a blurring kernel.^{21,24,26,27} These approaches are computationally efficient and are valid for phantoms in which the activity is contained inside a homogeneous medium. However, it has been reported that in *in vivo* small animal studies, a non-negligible amount of positrons close to the skin/air boundary can escape from the animal, and therefore violates the space-invariant assumption, and using space-invariant deblurring filters can result in severe artifacts.²⁸ Boundaries between lung/soft tissue and tissue/bone may lead to similar artifacts for the same reason. To address the presence of inhomogeneous media, space-variant analytic filtering approach were proposed.^{22,29} These filters are based on object attenuation map from coregistered CT scans, but due to the complexity of positron migration at irregular biological interfaces, developing such filters remains difficult.

An alternative to analytic models is MC simulation, which is generally regarded as the “silver” standard in the modeling of emission imaging systems when experimental measurements of the system response are difficult or impossible to obtain. MC simulation is able to model complex structures of biological tissues, as long as sufficient physical details about the attenuation media are provided. Close agreement in the positron range profiles have been found between experimental data¹⁵ and results simulated by the GEANT4 software, a well-validated MC package (CERN, Geneva, Switzerland).³⁰ However, MC simulation is also as-

sociated with excessively long computation time. Various variance reduction methods for MC simulation have been developed,^{31,32} but it remains expensive to use MC simulation directly as the forward and back projectors in iterative reconstruction algorithms.^{33,34} One method to reduce the computational cost is to use a simplified, fast unmatched back projector,^{35–37} but an accurate MC-based forward projector is still required at every iteration. In iterative single photon emission computed tomography (SPECT) reconstruction, intermittent MC correction method was proposed to reduce the computation cost of MC-based scatter compensation. This method holds the MC-based scatter estimate constant for several iterations before it is updated, and thereby reduces the number of MC simulations that is required.^{38,39} However, this method is only applicable to MC-based additive correction factors such as scattered events.

In this work, we develop a positron range model based on GEANT4 and incorporate it into a factorized system matrix to be used in high-resolution MAP image reconstruction. We assume the attenuation property of the object can be obtained from a coregistered CT scan. During image reconstruction, the GEANT4 simulation is performed on-the-fly as a part of the forward projector. The high computation cost of the MC simulation is addressed by the proposed general methodology that employs a residual correction technique to be described in the next section.

II. THEORY

II.A. MAP image reconstruction

PET data $y \in \mathbb{R}^{M \times 1}$ are well modeled as a collection of independent Poisson random variables with the expectation

$$\bar{y} = Px + r,$$

where $P \in \mathbb{R}^{M \times N}$ is the system matrix, $x \in \mathbb{R}^{N \times 1}$ is the unknown tracer distribution, and $r \in \mathbb{R}^{M \times 1}$ accounts for the presence of scattered and random events in the data. The log-likelihood of the PET data is

$$L(y|x, r) = \Phi(y, \bar{y}) = \sum_{i=1}^M (y_i \log \bar{y}_i - \bar{y}_i - \log y_i!). \quad (1)$$

A MAP estimate is found by maximizing the log-posterior density function $\Psi(x|y, r)$

$$\hat{x}_{\text{MAP}} = \arg \max_{x \geq 0} \Psi(x|y, r),$$

$$\Psi(x|y, r) = L(y|x, r) - \beta U(x), \quad (2)$$

where $-\beta U(x)$ is the log-prior term that regularizes the unknown image and β is a hyperparameter that controls the resolution of the reconstructed image.

Equation (2) does not have a closed-form analytical solution and is often solved by iterative algorithms. The primary computation cost of iterative methods is in the forward and back projection operations, which are matrix-vector product calculations using P and P^T , respectively. The computation

cost can be largely reduced when the system matrix has a high degree of sparsity or has certain structure that allows fast computation.

II.B. Error propagation from the system matrix

Many practical algorithms use a simplified approximate system matrix to improve computation efficiency. Doing this inevitably degrades image quality. Let us define the error in the system matrix as $\Delta P = P_{\text{true}} - P_{\text{approx}}$, where P_{true} is the accurate system matrix and P_{approx} is a simplified system matrix, and define the corresponding artifact in an reconstructed image caused by ΔP as $\Delta \hat{x} = x^* - \hat{x}$, where x^* and $\hat{x}$ are the images reconstructed using P_{true} and P_{approx} , respectively. It has been shown that for any log-likelihood function that can be written as $L(y|x, r) = \Phi(y, \hat{y})$, $\Delta \hat{x}$ can be approximated by⁴⁰

$$\Delta \hat{x} \approx -[P_{\text{approx}}^T \nabla^2 \Phi(y, \hat{y}) P_{\text{approx}} - \beta \nabla^2 U(\hat{x})]^{-1} \cdot [P_{\text{approx}}^T \nabla^2 \Phi(y, \hat{y})](\Delta P) \hat{x}, \quad (3)$$

where “ T ” denotes matrix transpose, $\hat{y} = P_{\text{approx}} \hat{x} + r$, $\nabla^2 \Phi(y, \hat{y})$ is a symmetric matrix with the (i, j) th element being $(\partial^2 / \partial \hat{y}_i \partial \hat{y}_j) \Phi(y, \hat{y})$, and $\nabla^2 U(\hat{x})$ is a symmetric matrix with the (i, j) th element being $(\partial^2 / \partial \hat{x}_i \partial \hat{x}_j) U(\hat{x})$. For the independent Poisson log-likelihood function in Eq. (1), we have $\Phi(y, \hat{y}) = \sum_i (y_i \log \hat{y}_i - \hat{y}_i - \log \hat{y}_i!)$, and $\nabla^2 \Phi(y, \hat{y}) = \text{diag}(-y_i / \hat{y}_i^2)$. For the Gaussian log-likelihood function, we have $\Phi(y, \hat{y}) = -1/2 (y - \hat{y})^T W (y - \hat{y})$ and $\nabla^2 \Phi(y, \hat{y}) = -W$, where W is the inverse of the covariance matrix of data y .

II.C. Artifact correction

Equation (3) suggests that the reconstruction artifacts $\Delta \hat{x}$ caused by using a simplified system matrix can be approximately computed by solving the following linear equations of $\Delta \hat{x}$:

$$[P_{\text{approx}}^T \nabla^2 \Phi(y, \hat{y}) P_{\text{approx}} - \beta \nabla^2 U(\hat{x})] \Delta \hat{x} \approx -[P_{\text{approx}}^T \nabla^2 \Phi(y, \hat{y})](\Delta P) \hat{x}, \quad (4)$$

where $(\Delta P) \hat{x} = P_{\text{true}} \hat{x} - P_{\text{approx}} \hat{x}$ is computed as the difference between the forward projections using the accurate and simplified system matrices. Once $\Delta \hat{x}$ is solved, a corrected image is given by $\hat{x}^{\text{new}} = \hat{x} + \Delta \hat{x}$.

There are two drawbacks with this direct correction approach. One is that a specialized optimization algorithm is needed for solving Eq. (4). The other is that $\hat{x}^{\text{new}}$ may contain negative values that violate the nonnegativity constraint defined in Eq. (2).

To overcome these drawbacks, we propose a different approach to remove the reconstruction artifacts. It has been shown that the derivative of the reconstructed image $\hat{x}$ with respect to the background term r is^{41,42}

$$\nabla_r \hat{x} = -[P_{\text{approx}}^T \nabla^2 \Phi(y, \hat{y}) P_{\text{approx}} - \beta \nabla^2 U(\hat{x})]^{-1} \cdot [P_{\text{approx}}^T \nabla^2 \Phi(y, \hat{y})]. \quad (5)$$

Comparing Eq. (3) with Eq. (5), we find that the error propagation expression in Eq. (3) can be rewritten as

$$\Delta \hat{x} \approx (\nabla_r \hat{x})(\Delta P) \hat{x}. \quad (6)$$

Then an error-corrected image can be obtained by using the first order Taylor approximation

$$\begin{aligned} \hat{x}^{\text{new}} &\approx \hat{x} + (\nabla_r \hat{x})(\Delta P) \hat{x} = \arg \max_{x \geq 0} \Psi_{\text{approx}}(x|y, r) \\ &\quad + \left[\frac{\partial}{\partial r} (\arg \max_{x \geq 0} \Psi_{\text{approx}}(x|y, r)) \right] (\Delta P) \hat{x} \\ &\approx \arg \max_{x \geq 0} \Psi_{\text{approx}}(x|y, r + (\Delta P) \hat{x}), \end{aligned} \quad (7)$$

where $\Psi_{\text{approx}}(\cdot) \equiv \Psi(\cdot)|_{P=P_{\text{approx}}}$ is the MAP objective function defined by the simplified system matrix P_{approx} . The above formula indicates that after precompensating the data for the error propagation, a more accurate image can be reconstructed using only the simplified system matrix P_{approx} . In this way, the whole reconstruction algorithm consists of two passes. The first pass is an initial reconstruction to obtain an approximate estimate $\hat{x}$, and the second pass is an artifact correction step to refine the estimate by reconstructing the precompensated data. The advantage of using Eq. (7) as compared to solving Eq. (4) directly is that no specialized optimization algorithm is required. Any existing convergent MAP optimization algorithm can be used in both passes.

It should be noted that the knowledge required for computing the data correction term is only a forward projection using P_{true} , so a “black-box” forward projector suffices this purpose and the values of the individual elements of the full matrix need not be explicitly computed. For instance, for factorized system matrices or on-the-fly MC-based forward projectors, the entries in the system matrix are never explicitly formed, but one can conveniently use such projectors to calculate the proposed data correction term. In addition, computing the correction term does not require a backprojection using P_{true}^T . When the forward and backprojection using P_{approx} is much more efficient than using P_{true} , substantial savings in the reconstruction time can be achieved compared with a conventional iterative reconstruction method that uses P_{true} in every forward and back projections.

II.D. Iterative artifact correction

Equation (7) is only an approximation because of the first-order Taylor approximation used. The refined image after correction may still contain residual errors. To further improve the image quality, we may apply the correction multiple times. This iterative correction procedure can be written as

$$\begin{aligned} \hat{x}^{(0)} &= \hat{x}, \\ \hat{x}^{(n+1)} &= \arg \max_{x \geq 0} \Psi_{\text{approx}}(x|y, r + (\Delta P) \hat{x}^{(n)}), \end{aligned} \quad (8)$$

where $\hat{x}^{(n)}$ is the image reconstructed after the n th update of the data correction term. In our implementation, the image reconstruction program just updates the correction term every a few iterations using the latest image estimate.

We analyzed the convergence property of the iterative correction scheme (see Appendix). It is found that as $n \rightarrow \infty$, $\hat{x}^{(n)}$ defined in Eq. (8) will actually converge to x^* (which is the accurate image reconstructed using P_{true}), for maximum likelihood reconstruction ($\beta=0$) under the following sufficient conditions: (i) The MAP solution x^* defined by P_{true} is unique; (ii) the optimization in Eq. (8) reaches convergence for each correction step; and (iii) the difference between P_{approx} and P_{true} can be modeled by an image domain filter B , i.e., $P_{\text{true}} = P_{\text{approx}} B$, and B has positive eigenvalues.

There is no guarantee of convergence in general cases. To improve the convergence property, a line search can be added to the algorithm such that

$$\hat{x}^{(n+1)} = (1 - \alpha^{(n)})\hat{x}^{(n)} + \alpha^{(n)}[\arg \max_{x \geq 0} \Psi_{\text{approx}}(x|y, r + (\Delta P)\hat{x}^{(n)})], \quad (9)$$

where $0 \leq \alpha^{(n)} \leq 1$ is a step size that is chosen to ensure $\Psi_{\text{true}}(\hat{x}^{(n+1)}) \geq \Psi_{\text{true}}(\hat{x}^{(n)})$, where $\Psi_{\text{true}}(\cdot) \equiv \Psi(\cdot)|_{P=P_{\text{true}}}$ is the MAP objective function defined using the accurate system matrix. As a result, the sequence $\{\hat{x}^{(n)}\}_{n=0}^{\infty}$ generated by Eq. (9) from a non-negative initial estimate $\hat{x}^{(0)}$ is guaranteed to converge. The line search step adds an insignificant amount of computation overhead because the time-consuming forward projection has already been calculated.

III. VALIDATION BY COMPUTER SIMULATIONS

III.A. Positron range modeling using Monte Carlo simulation

We developed a GEANT4-based simulation program to compute the distribution of positron annihilation points from a distribution of positron emitters in inhomogeneous media. Physical processes modeled in the simulation included poly-energetic positron decay, multiple elastic scattering, ionization, delta ray production, Bremsstrahlung, and annihilation with electrons. The isotope-dependent energy spectra of emitted positrons were calculated analytically based on their maximum (end point) energy. The tortuous trajectory of a positron was traced until it escaped from the simulation volume or annihilated with an electron.

The emission source and the attenuation media were defined on a common 64×64 voxel grid, with a voxel size of $0.5 \times 0.5 \times 10 \text{ mm}^3$, occupying a total simulation volume of $32 \times 32 \times 10 \text{ mm}^3$. The objects are constant along the axial direction of the scanner. The emission intensity and material composition of each voxel were assigned independently. The positron emitter was ^{18}F , a widely used radioisotope in PET. Figure 1(a) shows the 2D digital mouse chest phantom²⁸ used to define the positron source distribution. We added a small tumor in the center of the right lung. The relative emission intensities in the heart, tumor, torso background, and lungs regions were 250, 150, 90, and 10, respectively. The torso background and lungs were assigned with different attenuation properties with the density of the lungs tissue being about one third of that of the torso background. Elementwise

FIG. 1. (a) The mouse chest emission phantom used in the computer simulation (Ref. 28); (b) A positron range blurred phantom computed by a GEANT4 realization.

biological tissue composition was provided in the GEANT4 source code.

Each run of the MC simulation took about two minutes on a PC cluster with 64 CPUs, tracking the trajectories of 6 million positrons. The range-blurred image was produced by counting the number of positrons annihilated in each voxel. Figure 1(b) shows a range-blurred image of the emission phantom produced by a GEANT4 realization. Note that the tumor is nearly invisible because of the long positron range in the lung.

III.B. PET sinogram generation

We simulated a 2D PET scanner having 180 angular views over 180° and 128 radial samples per view. The field of view of the scanner had a diameter of 6.4 cm. Simulated PET sinograms were generated by forward-projecting the range-blurred images using a parallel beam geometric projector. The geometric projection was computed using a strip-integral model and took about 0.3 s on a single CPU. Computer generated Poisson noise was then added to the sinogram, with an expected total number of detected events of 6×10^6 (same count level as the number of positrons tracked in the Monte Carlo simulation). Figure 2 shows the PET sinogram before and after adding noise. We did not model the object attenuation, scattered and random events in the generation of the sinogram, as they are less significant in such a small phantom.

III.C. Image reconstruction

Three different reconstruction methods were compared. Method 1 is the conventional method that incorporates the aforementioned MC-based positron range model into forward and back projectors and perform the MC simulation

FIG. 2. Simulated PET sinograms. Left: Before adding Poisson noise; right: After adding Poisson noise.

on-the-fly in every iteration of the image reconstruction. This method is expected to produce the best results, but is extremely slow. Method 2 is the conventional simplified method that uses only the geometric projection matrix in forward and back projectors and does not model the positron range. Method 3 is the proposed residual correction method, i.e., Eq. (8), which uses the geometric projection matrix as P_{approx} and uses the MC-based positron range model for artifact correction. For method 3, we investigated applying residual corrections after every five, ten, or 20 image updates. We refer to these three versions of method 3 as methods 3a, 3b, and 3c, respectively.

MAP estimates were obtained using a preconditioned conjugate gradient (PCG) optimization algorithm with an EM-type preconditioner and a non-negativity constraint enforced by a bent line search.² We used a combination of the Poisson log-likelihood function and a quadratic pairwise smoothing prior that penalizes differences between 8 nearest neighboring pixels. A wide range of the regularization parameter, from no regularization ($\beta=0$) to oversmoothed image ($\beta=0.01$), was tested. All the three methods in the study used the same PCG algorithm, but differed in the system matrix being used and whether the residual correction was applied. All reconstructions started with a uniform image as the initial and the PCG algorithm was run for 40 iterations.

III.D. Region of interest activity quantification

The quantitative accuracy of different reconstruction methods was compared in terms of region of interest (ROI) quantification. A ROI was defined for the tumor in the phantom shown in Fig. 1(a). The ROI activity in a reconstructed image was computed by summing up all pixel values in the ROI. The bias and standard deviation of the ROI quantification were computed from 32 independent noisy realizations. Bias versus standard deviation curves were obtained by varying the regularization parameter β .

IV. RESULTS

IV.A. Reconstructed images

Figure 3 compares images reconstructed by different methods. No regularization was imposed in the noiseless cases ($\beta=0$), while both unregularized ($\beta=0$) and regularized ($\beta=10^{-4}$) results are shown for the noisy data. Note that the images reconstructed from the noiseless data are still subject to the statistical noise of the MC-based forward and back projectors (methods 1 and 3). In the noisy cases, regularized reconstruction produces less noise and smoother images. For both noiseless and noisy data, the proposed method (method 3) gives better visual contrast than method 2 that uses the simplified system model alone. The tumor spot inside the lung region can be easily identified by method 3, but is barely visible in the images produced by method 2. The proposed method also achieves similar visual quality as the conventional method that uses the MC-based model in every iteration (method 1).

FIG. 3. Sample images reconstructed by different methods.

IV.B. Convergence speed

A common way to compare the convergence rate of optimization algorithms is to plot the objective function value as a function of iteration. However, since each method in this study optimizes a different objective function, direct comparison of the individual objective function values is not meaningful. Instead, we use $\Psi_{\text{true}}(\cdot)$, the MAP objective function defined by the accurate system matrix that includes the MC-based positron range model, as the indicator of image optimality. Figure 4 plots $\Psi_{\text{true}}(\cdot)$ as a function of iteration number. All three versions of method 3 (updating the data correction term per five, ten, and 20 image updates) produce objective function values that are much greater than that achieved by method 2. The curves of the proposed method jump up immediately after each update of the data correction term, suggesting that the information from the accurate system model is effectively exploited in the subsequent iterations. Figure 4 also shows that updating the correction term more frequently results in faster increase in the objective function as a function of iteration, which is expected. Surprisingly, method 3a, which updates the correction term every five image updates, obtained even greater values of the objective function than those of method 1 at most iterations. This is likely due to the improved condition of the simplified system matrix. At iteration 40, method 3a achieves comparable objective function value as that achieved by method 1. Comparing the noiseless and noisy data, the curves for noisy reconstruction plateau at earlier iterations because of the effect of regularization and inconsistency in the data, but the relative performances among all methods remain similar.

FIG. 4. Plots of objective function values as a function of iteration number. Left: The noiseless case; right: The noisy case with $\beta=10^{-4}$.

IV.C. Computation cost

It is important to note that the computation cost per iteration is dramatically different for different methods. Figure 5 plots the objective function value as a function of computation time in seconds. The MC simulation of positron range using the GEANT4 software package is the most time-consuming part in methods 1 and 3. In method 1, the MC simulation was performed at least twice for each image update, one for the forward projection and one for the back projection, leading to a total number of 80 MC runs for 40 image updates. For the proposed method that performs data correction every five image updates (method 3a in Fig. 5), only seven MC runs were required for 40 image updates. Therefore, the proposed method is about 11 times faster than method 1 for reaching the same level of objective function value. For the other two versions of the new method which update the data correction term less frequently, the computation advantage is even greater (about 20 fold acceleration for

method 3b and about 40 fold acceleration for method 3c).

IV.D. ROI quantification result

Figure 6 compares the mean and standard deviation images of 32 independent noisy reconstructions by methods 2 and 3a. Method 1 is excluded from the comparison because it is very time-consuming to reconstruct all 32 noisy realizations with various regularization parameters. For the smallest two regularization parameters, namely, $\beta=0$ and $\beta=10^{-4}$, the mean images for method 3a show sharper boundaries and higher resolution than those of method 2. The proposed method also shows slightly higher standard deviation because the reconstruction problem is more ill-posed after positron range is included in the system matrix. As we increase the regularization parameter to $\beta=10^{-3}$ and $\beta=10^{-2}$, the reconstructed images are more dominated by the prior term and become smoother, thus less difference is observed between the two methods.

FIG. 5. Plots of objective function values as a function of computation time. Left: The noiseless case; right: The noisy case with $\beta=10^{-4}$. The line style definitions are the same as in Fig. 4.

FIG. 6. Mean and standard deviation images of methods 2 and 3a with different regularization parameters.

Figure 7 compares the bias versus standard deviation tradeoff characteristics of the two methods for quantifying the total activity inside the tumor ROI. The curves were generated by varying the regularization parameter. It shows that the proposed method significantly reduces the quantification bias with a moderate increase in the standard deviation. The minimum bias of method 2 is 70%, whereas the minimum bias of the proposed method is 20%. At the same bias levels,

FIG. 7. Comparison of normalized bias and standard deviation of the ROI quantification between methods 2 (conventional) and 3a (proposed).

the proposed method results in less standard deviation and offers better quantification results than the conventional method (method 2).

V. DISCUSSION

We aim to incorporate a highly accurate system model into iterative tomographic reconstruction while maintaining reasonable computation cost. For this purpose, a two-level algorithm is proposed. In the inner level, a simplified approximate system matrix is used for efficient computation, whereas in the outer level, an accurate system matrix is used to retain the accuracy of the final solution. The proposed method compensates for the error in the simplified system matrix by introducing a data correction term. Since the more accurate but time-consuming system matrix is used less frequently, the overall computation time is reduced. The proposed algorithm resembles the characteristics of many multiresolution and multigrid techniques that recursively operate at different resolutions and use the solution obtained at lower resolution to compute the solution at a higher resolution. Multiresolution methods are especially popular in solving partial differential equations based inverse problems such as optical diffusion tomography. They have also been applied to emission and transmission tomography problems.^{43,44}

Finding a proper approximate system matrix is the key in the proposed method. In our computer simulation, the approximate system matrix is constructed by excluding the positron range in the imaging model. Similar approaches were adopted to construct efficient unmatched back projec-

tors by ignoring the effect of in-object photon scatter or non-uniform object attenuation.^{35–37} However, there exists a wide variety of alternatives to construct the approximate system matrix being used in the proposed method. For example, an approximate system matrix may be obtained by reducing the number of nonzero elements to make it more sparse,^{7,45} or using fast forward and back projectors based on Fourier slice theorem⁴⁶ and Fourier rebinning.⁴⁷ The proposed method may also be used in combination with image reconstruction algorithms running on modern vector computing platforms such as graphical processing units,⁴⁸ where the simplified system model can be tailored to take advantage of special features of the hardware. In this way, the new method has the potential to combine the computational advantages of a variety of approximate system models and hardware platforms with the image quality offered by highly accurate system models.

Another advantage of the proposed method is that it can employ any convergent MAP reconstruction algorithm. The derivation of the proposed method is based on perturbation analysis of the MAP solution, which is independent of the optimization algorithm used to seek the solution. This allows us to choose the latest and fastest optimization algorithm for image reconstruction, including the convergent versions of subset-based algorithms.^{49–51}

A common intuition is that any approximations in the system matrix will always slow down the convergence to the true solution. However, as shown in Fig. 5, the convergence speed of the proposed method may actually outperform the conventional method that uses the accurate system model in every iteration. We show in the Appendix that the introduction of an approximate system matrix can be viewed as an implicit preconditioner in special cases. This is similar to the findings that a properly formed mismatched forward/back projector pair can improve convergence speed.^{35–37}

Empirically we have found that the proposed method works best when the difference between P_{true} and P_{approx} is small and the overall difference can be modeled entirely by an image domain filter. When the difference is too large, a single data correction term may not be sufficient to retain the accuracy of the final solution and the performance of proposed method may degrade.

VI. CONCLUSION

We proposed a residual correction MAP algorithm for high-resolution image reconstruction. The major advantage is that the proposed method does not require the accurate system model at every iteration. The new method reconstructs an initial image with an approximate system model and then removes the reconstruction artifacts by residual correction. We presented an application of the proposed method to incorporate an on-the-fly Monte Carlo based positron range model into PET image reconstruction. Computer simulation shows that images produced by the proposed method are similar to those obtained by the conventional method using the accurate MC model, but the new method is an order of magnitude faster. The method is especially efficient

when the computation cost of using the approximate system model is much less than the accurate system model. The error correction capability can be added to existing reconstruction algorithms with little modification. Future work includes detailed evaluation of the proposed method using phantom and real data.

ACKNOWLEDGMENTS

The authors would like to thank Matthew J. Middione for his help with the GEANT4 simulation. The authors also thank Michel S. Tohme for his assistance in the preparation of the manuscript. This work is supported by the National Institute of Biomedical Imaging and Bioengineering under Grant No. R01EB005322.

APPENDIX: CONVERGENCE PROPERTIES OF EQ. (8)

In this Appendix we study the convergence property of the proposed artifact correction algorithm. We limited the analysis to maximum likelihood image reconstruction (i.e., $\beta=0$). We will show that under the following conditions, the sequence $\hat{x}^{(n)}$ defined in Eq. (8) will converge to x^* (which is the accurate image reconstructed using P_{true})

- (i) the MAP solution x^* defined by P_{true} is unique, i.e., the null space of P_{true} is empty;
- (ii) the optimization in Eq. (8) reaches convergence for each correction step;
- (iii) the difference between P_{approx} and P_{true} can be modeled by an image domain filter B , i.e., $P_{\text{true}} = P_{\text{approx}}B$, and B has positive eigenvalues.

Proof:

Ignoring the nonnegativity constraint on the image, Eq. (8) can be written as

$$\begin{aligned}\hat{x}^{(n+1)} &= \arg \max_x \Phi(y, P_{\text{approx}}x + r + (P_{\text{true}} - P_{\text{approx}})\hat{x}^{(n)}), \\ \hat{x}^{(0)} &= \arg \max_x \Phi(y, P_{\text{approx}}x + r).\end{aligned}\quad (\text{A1})$$

Taking derivative of the above objective functions with respect to x , we obtain the normal equations that implicitly define $\hat{x}^{(n+1)}$ and $\hat{x}^{(0)}$, respectively [condition (ii)],

$$\begin{aligned}P_{\text{approx}}^T \nabla^{01} \Phi(y, P_{\text{approx}}\hat{x}^{(n+1)} + r + (P_{\text{true}} - P_{\text{approx}})\hat{x}^{(n)}) &= 0, \\ P_{\text{approx}}^T \nabla^{01} \Phi(y, P_{\text{approx}}\hat{x}^{(0)} + r) &= 0,\end{aligned}\quad (\text{A2})$$

where $\nabla^{01} \Phi(\cdot, \cdot)$ is the gradient of $\Phi(\cdot, \cdot)$ with respect to its second variable. Substituting condition (iii) $P_{\text{true}} = P_{\text{approx}}B$ into Eq. (A2) gives

$$\begin{aligned}P_{\text{approx}}^T \nabla^{01} \Phi(y, P_{\text{approx}}(\hat{x}^{(n+1)} + (B - I)\hat{x}^{(n)} + r)) &= 0, \\ P_{\text{approx}}^T \nabla^{01} \Phi(y, P_{\text{approx}}\hat{x}^{(0)} + r) &= 0.\end{aligned}\quad (\text{A3})$$

The combination of conditions (i) and (iii) implies that the null space of P_{approx} is also empty. In this case, since the log-likelihood function $\Phi(y, \hat{y})$ is a strictly concave

function with respect to $\hat{y}$, the normal equation $P_{\text{approx}}^T \nabla^{01} \Phi(y, P_{\text{approx}} x + r) = 0$ has a unique solution for x . Comparing the two sets of equations in Eq. (A3), we obtain a closed-form expression for $\hat{x}^{(n+1)}$

$$\hat{x}^{(n+1)} = \hat{x}^{(0)} - (B - I)\hat{x}^{(n)}. \quad (\text{A4})$$

This is a Jacobian iteration for solving the equation $Bx = \hat{x}^{(0)}$. The iteration will converge if the eigenvalues of $(B - I)$ are within the interval of $(-1, 1)$, or equivalently, the eigenvalues of B are in the interval of $(0, 2)$. Since a scaling factor can always be applied to scale the approximate system matrix P_{approx} such that the largest eigenvalue of B is less than two, this requirement is satisfied by B having positive eigenvalues [condition (iii)]. Therefore, the sequence $\hat{x}^{(n)}$ converges to

$$\hat{x}^{(\infty)} = \lim_{n \rightarrow \infty} \hat{x}^{(n)} = B^{-1} \hat{x}^{(0)}, \quad (\text{A5})$$

which is the solution x^* obtained by a conventional iterative reconstruction method that uses P_{true} in every iteration.

The above analysis applies to any convex log-likelihood function in the form of $\Phi(y, \hat{y})$. In particular, for Gaussian likelihood function, or weighted least-squares reconstruction, Eq. (A1) becomes

$$\hat{x}^{(n+1)} = \arg \max_x \{-\|y - r - (P_{\text{true}} - P_{\text{approx}})\hat{x}^{(n)} - P_{\text{approx}}x\|_W^2\}.$$

The corresponding normal equations are

$$(P_{\text{approx}}^T W P_{\text{approx}})\hat{x}^{(n+1)} = P_{\text{approx}}^T W(y - r - (P_{\text{true}} - P_{\text{approx}})\hat{x}^{(n)}).$$

Assuming $P_{\text{approx}}^T W P_{\text{approx}}$ is full rank, we have

$$\hat{x}^{(n+1)} = \hat{x}^{(n)} + (P_{\text{approx}}^T W P_{\text{approx}})^{-1} [P_{\text{approx}}^T W(y - r - P_{\text{true}}\hat{x}^{(n)})].$$

This is a generalized Landweber iteration with preconditioner $(P_{\text{approx}}^T W P_{\text{approx}})^{-1}$ and unmatched forward/back projector pair P_{true} and P_{approx}^T , which reveals that the iterative correction scheme can accelerate convergence through an implicitly formed preconditioner.

^{a)}Electronic mail: qi@ucdavis.edu

¹A. R. Formiconi, A. Pupi, and A. Passeri, "Compensation of spatial system response in SPECT with conjugate gradient reconstruction technique," *Phys. Med. Biol.* **34**(1), 69–84 (1989).

²J. Qi *et al.*, "High-resolution 3D Bayesian image reconstruction using the microPET small-animal scanner," *Phys. Med. Biol.* **43**, 1001–1013 (1998).

³V. Y. Panin *et al.*, "Fully 3D PET reconstruction with system matrix derived from point source measurements," *IEEE Trans. Med. Imaging* **25**(7), 907–921 (2006).

⁴A. Rahmim *et al.*, "Analytic system matrix resolution modeling in PET: An application to Rb-82 cardiac imaging," *Phys. Med. Biol.* **53**, 5947–5965 (2008).

⁵T. Frese *et al.*, "Quantitative comparison of FBP, EM and Bayesian reconstruction algorithms for the IndyPET scanner," *IEEE Trans. Med. Imaging* **22**(2), 258–276 (2003).

⁶A. Reader *et al.*, "EM algorithm system modeling by image-space techniques for PET reconstruction," *IEEE Trans. Nucl. Sci.* **50**, 1392–1397 (2003).

⁷F. van der Have *et al.*, "System calibration and statistical image reconstruction for ultra-high resolution stationary pinhole SPECT," *IEEE Trans. Med. Imaging* **27**, 960–971 (2008).

⁸V. Selivanov *et al.*, "Detector response models for statistical iterative image reconstruction in high resolution PET," *IEEE Trans. Nucl. Sci.* **47**(3), 1168–1175 (2000).

⁹K. Lee *et al.*, "Pragmatic fully 3D image reconstruction for the MiCES mouse imaging PET scanner," *Phys. Med. Biol.* **49**, 4563–4578 (2004).

¹⁰T. Yamaya *et al.*, "Transaxial system models for jPET-D4 image reconstruction," *Phys. Med. Biol.* **50**, 5339–5355 (2005).

¹¹G. Zeng *et al.*, "Three-dimensional iterative reconstruction algorithms with attenuation and geometric point response correction," *IEEE Trans. Nucl. Sci.* **38**, 693–702 (1991).

¹²J. R. Stickel and S. R. Cherry, "High-resolution PET detector design: Modeling components of intrinsic spatial resolution," *Phys. Med. Biol.* **50**, 179–195 (2005).

¹³C. S. Levin and E. J. Hoffman, "Calculation of positron range and its effect on the fundamental limit of positron emission tomography system spatial resolution," *Phys. Med. Biol.* **44**, 781–799 (1999).

¹⁴K. Lee *et al.*, "Impact of system design parameters on image figures of merit for a mouse PET scanner," *IEEE Trans. Nucl. Sci.* **51**(1), 27–33 (2004).

¹⁵A. Blanco, "Positron range effects on the spatial resolution of RPC-PET," in Proceedings of the IEEE Nuclear Science Symposium Conference, San Diego, CA, 2006, pp. 2570–2573.

¹⁶L. Fu and J. Qi, "A residual correction method for iterative reconstruction with inaccurate system model," in Proceedings of the IEEE International Symposium on Biomedical Imaging, Paris, France, 2008, pp. 1311–1314.

¹⁷L. Fu and J. Qi, "A novel iterative image reconstruction method for high-resolution PET Imaging with a Monte Carlo based positron range model," in Proceedings of the IEEE Nuclear Science Symposium Conference, Dresden, Germany, 2008, pp. 3609–3612.

¹⁸M. E. Phelps *et al.*, "Effect of positron range on spatial resolution," *J. Nucl. Med.* **16**(7), 649–652 (1975).

¹⁹R. Laforest, D. J. Rowland, and M. J. Welch, "MicroPET imaging with non-conventional isotopes," in Proceedings of the IEEE Nuclear Science Symposium Conference, San Diego, CA, 2001, pp. 1572–1576.

²⁰S. E. Derenzo, in *Positron Annihilation*, edited by R. R. Hasiguti and K. Fujiwara (The Japan Institute of Metals, Sendai, 1979), pp. 819–823.

²¹M. R. Palmer and G. L. Brownell, "Annihilation density distribution calculations for medically important positron emitters," *IEEE Trans. Med. Imaging* **11**(3), 373–378 (1992).

²²A. Alessio and L. MacDonald, "Spatially variant positron range modeling derived from CT for PET image reconstruction," in Proceedings of the IEEE Nuclear Science Symposium Conference, Dresden, Germany, 2008, pp. 3637–3640.

²³R. Harrison *et al.*, "Positron range and coincidence non-collinearity in SimSET," in Proceedings of the IEEE Nuclear Science Symposium Medical Imaging Conference, Seattle, WA, 1999, pp. 1265–1268.

²⁴M. Palmer, X. Zhu, and J. Parker, "Modeling and simulation of positron range effects for high resolution PET imaging," *IEEE Trans. Nucl. Sci.* **52**, 1391–1395 (2005).

²⁵A. Blanco *et al.*, "RPC-PET: A new very high resolution PET technology," *IEEE Trans. Nucl. Sci.* **53**(5), 2489–2494 (2006).

²⁶S. E. Derenzo, "Mathematical removal of positron range blurring in high resolution tomography," *IEEE Trans. Nucl. Sci.* **33**, 565–569 (1986).

²⁷S. Haber, S. Derenzo, and D. Uber, "Application of Mathematical removal of positron range blurring in positron emission tomography," *IEEE Trans. Nucl. Sci.* **37**(3), 1293–1299 (1990).

²⁸B. Bai *et al.*, "Positron range modeling for statistical PET image reconstruction," in Proceedings of the IEEE Nuclear Science Symposium Conference, Portland, OR, 2003, pp. 2501–2505.

²⁹B. Bai *et al.*, "Evaluation of MAP image reconstruction with positron range modeling for 3D PET," in Proceedings of the IEEE Nuclear Science Symposium Conference, Puerto Rico, 2005, pp. 2686–2689.

³⁰S. Agostinelli *et al.*, "Geant4—A simulation toolkit," *Nucl. Instrum. Methods Phys. Res. A* **506**(3), 250–303 (2003).

³¹R. Harrison *et al.*, "Acceleration of SimSET photon history generation," in Proceedings of the IEEE Nuclear Science Symposium Medical Imaging Conference, 2002, pp. 1835–1838.

³²H. de Jong, E. Slijpen, and F. J. Beekman, "Acceleration of Monte Carlo SPECT simulation using convolution-based forced detection," *IEEE Trans. Nucl. Sci.* **48**(1), 58–64 (2001).

- ³³C. Floyd, R. Jaszcak, and R. Coleman, "Inverse Monte Carlo: A unified reconstruction algorithm for SPECT," *IEEE Trans. Nucl. Sci.* **32**, 779–785 (1985).
- ³⁴D. Lazaro *et al.*, "Fully 3D Monte-Carlo reconstruction in SPECT: A feasibility study," *Phys. Med. Biol.* **50**, 3739–3754 (2005).
- ³⁵G. L. Zeng and G. T. Gullberg, "Unmatched projector/backprojector pairs in an iterative reconstruction algorithm," *IEEE Trans. Med. Imaging* **19**(5), 548–555 (2000).
- ³⁶F. J. Beekman, H. de Jong, and S. van Geloven, "Efficient fully 3-D iterative SPECT reconstruction with Monte Carlo-based scatter compensation," *IEEE Trans. Med. Imaging* **21**(8), 867–877 (2002).
- ³⁷C. Kamphuis *et al.*, "Dual matrix ordered subsets reconstruction for accelerated 3D scatter compensation in single-photon emission tomography," *Eur. J. Nucl. Med.* **25**, 8–18 (1997).
- ³⁸D. J. Kadrmas, *et al.*, "Fast implementations of reconstruction-based scatter compensation in fully 3D SPECT image reconstruction," *Phys. Med. Biol.* **43**(4), 857–873 (1998).
- ³⁹Y. 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.* **53**(1), 181–188 (2006).
- ⁴⁰J. Qi and R. H. Huesman, "Effect of errors in the system matrix on maximum a posteriori image reconstruction," *Phys. Med. Biol.* **50**(14), 3297–3312 (2005).
- ⁴¹J. Qi and R. H. Huesman, "Scatter correction for positron emission mammography," *Phys. Med. Biol.* **47**, 2759–2771 (2002).
- ⁴²J. A. Fessler and W. L. Rogers, "Spatial resolution properties of penalized-likelihood image reconstruction: Space-invariant tomographs," *IEEE Trans. Image Process.* **5**(9), 1346–1358 (1996).
- ⁴³S. Oh, C. A. Bouman, and K. J. Webb, "Multigrid tomographic inversion with variable resolution data and image spaces," *IEEE Trans. Image Process.* **15**(9), 2805–2819 (2006).
- ⁴⁴C. Bouman and K. Sauer, "Nonlinear multigrid methods of optimization in Bayesian tomographic image reconstruction," in *Proceedings of the SPIE Conference Neural and Stochastic Methods in Image and Signal Processing*, San Diego, CA, 1992, pp. 296–306.
- ⁴⁵D. R. Gilland *et al.*, "Approximate 3D iterative reconstruction for SPECT," *Med. Phys.* **24**(9), 1421–1429 (1997).
- ⁴⁶S. Matej, J. A. Fessler, and I. G. Kazantsev, "Iterative tomographic image reconstruction using Fourier-based forward and back-projectors," *IEEE Trans. Med. Imaging* **23**, 401–412 (2004).
- ⁴⁷S. Cho *et al.*, "Fast projectors for iterative 3D PET reconstruction," in *Proceedings of the IEEE Nuclear Science Symposium Medical Imaging Conference*, 2005, pp. 1933–1937.
- ⁴⁸F. Xu and K. Mueller, "Accelerating popular tomographic reconstruction algorithms on commodity PC graphics hardware," *IEEE Trans. Nucl. Sci.* **52**(3), 654–663 (2005).
- ⁴⁹J. Browne and A. D. Pierro, "A row-action alternative to the EM algorithm for maximizing likelihoods in emission tomography," *IEEE Trans. Med. Imaging* **15**, 687–699 (1996).
- ⁵⁰I. Hsiao *et al.*, "An accelerated convergent ordered subsets algorithm for emission tomography," *Phys. Med. Biol.* **49**(11), 2145–2156 (2004).
- ⁵¹P. Khurd *et al.*, "A globally convergent regularized ordered-subset EM algorithm for list-mode reconstruction," *IEEE Trans. Nucl. Sci.* **51**(3), 719–725 (2004).