

# Noise Propagation in Iterative Reconstruction Algorithms with Line Searches

Jinyi Qi, *Member, IEEE*

**Abstract**—In this paper we analyze the propagation of noise in iterative image reconstruction algorithms. We derive theoretical expressions for the general form of preconditioned gradient algorithms with line searches. The results are applicable to a wide range of iterative reconstruction problems, such as emission tomography, transmission tomography, and image restoration. A unique contribution of this paper comparing to our previous work [1] is that the line search is explicitly modeled and we do not use the approximation that the gradient of the objective function is zero. As a result, the error in the estimate of noise at early iterations is significantly reduced.

## I. INTRODUCTION

Iterative image estimation methods have been developed to improve image quality through accurately modeling the system response and noise distribution. It is essential to estimate the uncertainty of the reconstructed images for quantitative applications.

Barrett *et al.* [2] derived approximate formulae for the mean and covariance of the maximum likelihood (ML) expectation maximization (EM) reconstruction [3], [4] as a function of the iteration number. The same approach was extended to two maximum *a posteriori* (MAP) EM algorithms [5], [6] by Wang and Gindi [7] and to block iterative algorithms [8], [9] by Soares *et al* [10]. This iteration-based approach is attractive for algorithms that are terminated before convergence, as is common practice for the EM algorithm and its ordered-subsets variants [11].

An alternative approach was proposed by Fessler and Rogers [12], [13], who analyzed the mean, variance, and spatial resolution at the fixed point of the objective function. The resolution and noise properties are computed at the fixed point using partial derivatives and truncated Taylor series approximations. Qi and Leahy [14], [15] extended this approach by deriving simplified expressions for the local impulse response function and covariance using Fourier transforms. Stayman and Fessler [16] used a similar approximation in designing penalty functions for uniform resolution. Other extensions of the fixed point analysis include [17], [18], [19], [20], [21]. While these results are independent of the particular optimizing algorithm used, they require that the algorithm be iterated to effective convergence. Hence, they are not applicable to the images obtained at early iterations.

Building on the existing work, we [1] have derived unified formulae for calculating the mean and variance of the image es-

This work is supported in part by the National Institute of Biomedical Imaging and Bioengineering under grant nos. R01 EB00194 and R01 EB00363, and by the Director, Office of Science, Office of Biological and Environmental Research, Medical Sciences Division, of the U.S. Department of Energy under contract no. DE-AC03-76SF00098.

J. Qi is with the Department of Nuclear Medicine and Functional Imaging, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. (telephone: 510-486-4695, e-mail: jq@lbl.gov)

timate using preconditioned gradient algorithms, which includes ML-EM and MAP-EM as special cases. We have also shown that the iteration-based result is consistent with the fixed-point based results. While the results in [1] can be applied to algorithms with line search by using the approximation that the gradient of the objective function is zero, this results in large error at early iterations. In this paper, we extend the work in [1] by explicitly modeling the line search in the noise propagation. We only use the low noise approximation and hence improve the accuracy of the theoretical results at early iterations.

## II. THEORY

In image estimation problems, the image is often estimated through maximizing an objective function  $\Phi(\mathbf{y}, \mathbf{x})$

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} \Phi(\mathbf{y}, \mathbf{x}), \quad (1)$$

where  $\mathbf{x}$  is the unknown image and  $\mathbf{y}$  is the measurement.

A general form of iterative algorithms that solves (1) can be written as

$$\hat{\mathbf{x}}^{k+1} = \hat{\mathbf{x}}^k + \alpha \mathbf{a}^k \quad (2)$$

where  $\mathbf{a}^k$  is a search direction and  $\alpha$  is the step size that is either preselected, or determined through a one dimensional line search.

To analyze the noise propagation, we rewrite (2) as

$$\bar{\mathbf{x}}^{k+1} + \epsilon^{k+1} = \bar{\mathbf{x}}^k + \epsilon^k + (\bar{\alpha} + \tilde{\alpha})(\bar{\mathbf{a}}^k + \tilde{\mathbf{a}}^k) \quad (3)$$

$$= \bar{\mathbf{x}}^k + \bar{\alpha} \bar{\mathbf{a}}^k + \epsilon^k + \bar{\alpha} \tilde{\mathbf{a}}^k + \tilde{\alpha} \bar{\mathbf{a}}^k + \tilde{\alpha} \tilde{\mathbf{a}}^k. \quad (4)$$

where  $\bar{\mathbf{x}}^k$ ,  $\bar{\alpha}$ ,  $\bar{\mathbf{a}}^k$  are the expectations of the image estimate  $\hat{\mathbf{x}}^k$ , step size  $\alpha$ , and search direction  $\mathbf{a}^k$ , respectively, and  $\epsilon^k$ ,  $\tilde{\alpha}$ ,  $\tilde{\mathbf{a}}^k$  are zero mean noise vectors. In the rest of this paper, we will assume that the noise in each term is small compared to the corresponding expectation.

Taking the expectation on both side of (4), we have

$$\bar{\mathbf{x}}^{k+1} = \bar{\mathbf{x}}^k + \bar{\alpha} \bar{\mathbf{a}}^k + \langle \tilde{\alpha} \tilde{\mathbf{a}}^k \rangle, \quad (5)$$

where  $\langle \cdot \rangle$  denote the expectation.

Subtracting (5) from (4), we get the following noise propagation equation

$$\epsilon^{k+1} = \epsilon^k + \bar{\alpha} \tilde{\mathbf{a}}^k + \tilde{\alpha} \bar{\mathbf{a}}^k + \tilde{\alpha} \tilde{\mathbf{a}}^k - \langle \tilde{\alpha} \tilde{\mathbf{a}}^k \rangle. \quad (6)$$

Assuming  $\tilde{\alpha} \ll \bar{\alpha}$  and  $\tilde{\mathbf{a}} \ll \bar{\mathbf{a}}$ , we can drop the second-order noise terms in (5) and (6).

To compute these equations, we need knowledge of the search direction  $\mathbf{a}^k$  and step size  $\alpha$ . In the following, we focus on preconditioned gradient algorithms.

### A. Search Direction $\mathbf{a}^k$

In preconditioned gradient algorithms, the search direction  $\mathbf{a}^k$  is formed by

$$\mathbf{a}^k = \mathbf{C}^k(\hat{\mathbf{x}}^k, \mathbf{y}) \nabla_x \Phi(\mathbf{y}, \hat{\mathbf{x}}^k), \quad (7)$$

where  $\mathbf{C}^k(\hat{\mathbf{x}}^k, \mathbf{y})$  is a positive definite matrix (called preconditioner) and  $\nabla_x \Phi(\mathbf{y}, \hat{\mathbf{x}}^k)$  denotes the gradient vector of the objective function at  $\hat{\mathbf{x}}^k$ . Here we allow the preconditioner  $\mathbf{C}^k(\hat{\mathbf{x}}^k, \mathbf{y})$  to vary from iteration to iteration.

When the noise in  $\mathbf{y}$  and  $\hat{\mathbf{x}}^k$  is small, we can approximate the gradient vector and preconditioner using the first order Taylor series expansion

$$\begin{aligned} \nabla_x \Phi(\mathbf{y}, \hat{\mathbf{x}}^k) &\approx \nabla_x \Phi(\bar{\mathbf{y}}, \bar{\hat{\mathbf{x}}}^k) \\ &\quad + \nabla_{xy} \Phi(\bar{\mathbf{y}}, \bar{\hat{\mathbf{x}}}^k) \mathbf{n} + \nabla_{xx} \Phi(\bar{\mathbf{y}}, \bar{\hat{\mathbf{x}}}^k) \epsilon^k, \quad (8) \\ \mathbf{C}^k(\hat{\mathbf{x}}^k, \mathbf{y}) &\approx \mathbf{C}^k(\bar{\hat{\mathbf{x}}}^k, \bar{\mathbf{y}}) \\ &\quad + \mathbf{C}_x^k(\epsilon^k, \bar{\hat{\mathbf{x}}}^k, \bar{\mathbf{y}}) + \mathbf{C}_y^k(\mathbf{n}; \bar{\hat{\mathbf{x}}}^k, \bar{\mathbf{y}}), \quad (9) \end{aligned}$$

where  $\bar{\mathbf{y}}$  is the expectation of the data  $\mathbf{y}$ ,  $\mathbf{n} \equiv \mathbf{y} - \bar{\mathbf{y}}$  is the noise, the  $(j, l)$ th element of the operator  $\nabla_{xx}$  is  $\frac{\partial^2}{\partial x_j \partial x_l}$ , the  $(j, l)$ th element of the operator  $\nabla_{xy}$  is  $\frac{\partial^2}{\partial x_j \partial y_l}$ , the  $(j, l)$ th element of  $\mathbf{C}_x^k(\epsilon^k; \mathbf{x}, \mathbf{y})$  is  $\sum_m \epsilon_m^k \partial C_{j,l}^k(\mathbf{x}, \mathbf{y}) / \partial x_m$ , and the  $(j, l)$ th element of  $\mathbf{C}_y^k(\mathbf{n}; \mathbf{x}, \mathbf{y})$  is  $\sum_m n_m \partial C_{j,l}^k(\mathbf{x}, \mathbf{y}) / \partial y_m$ .

From the above results,  $\mathbf{a}$  can be approximated by

$$\begin{aligned} \mathbf{a} &\approx \bar{\mathbf{a}} + \mathbf{C}^k(\bar{\hat{\mathbf{x}}}^k, \bar{\mathbf{y}}) \nabla_x \Phi(\bar{\mathbf{y}}, \bar{\hat{\mathbf{x}}}^k) \epsilon^k \\ &\quad + \mathbf{M}_x \left[ \nabla_x \Phi(\bar{\mathbf{y}}, \bar{\hat{\mathbf{x}}}^k); \bar{\hat{\mathbf{x}}}^k, \bar{\mathbf{y}} \right] \epsilon^k \\ &\quad + \mathbf{C}^k(\bar{\hat{\mathbf{x}}}^k, \bar{\mathbf{y}}) \nabla_{xy} \Phi(\bar{\mathbf{y}}, \bar{\hat{\mathbf{x}}}^k) \mathbf{n} \\ &\quad + \mathbf{M}_y \left[ \nabla_x \Phi(\bar{\mathbf{y}}, \bar{\hat{\mathbf{x}}}^k); \bar{\hat{\mathbf{x}}}^k, \bar{\mathbf{y}} \right] \mathbf{n} \quad (10) \end{aligned}$$

where

$$\bar{\mathbf{a}} = \mathbf{C}^k(\bar{\hat{\mathbf{x}}}^k, \bar{\mathbf{y}}) \nabla_x \Phi(\bar{\mathbf{y}}, \bar{\hat{\mathbf{x}}}^k), \quad (11)$$

the  $(j, l)$ th element of  $\mathbf{M}_x[\mathbf{g}; \mathbf{x}, \mathbf{y}]$  is

$$\sum_m g_m \partial C_{j,m}^k(\mathbf{x}, \mathbf{y}) / \partial x_l,$$

and the  $(j, l)$ th element of  $\mathbf{M}_y[\mathbf{g}; \mathbf{x}, \mathbf{y}]$  is

$$\sum_m g_m \partial C_{j,m}^k(\mathbf{x}, \mathbf{y}) / \partial y_l.$$

### B. Gradient algorithms with preselected step size

There are many iterative algorithms use predetermined step size. For examples, ML-EM [3], [4] and MAP-EM [5], [6] (and their ordered subsets variants [11]) use a constant step size of one; RAMLA [22] and BSREM [23] use predetermined variable step sizes. For such algorithms, there is no noise in  $\alpha$ , i.e.,  $\tilde{\alpha} \equiv 0$ .

Substituting (10) into (6), we get the following noise propagation equation

$$\epsilon^{k+1} \approx [\mathbf{I} - \mathbf{A}^k] \epsilon^k + \mathbf{B}^k \mathbf{n} \equiv \mathbf{V}^{k+1} \mathbf{n}, \quad (12)$$

where

$$\begin{aligned} \mathbf{A}^k &= -\alpha \left\{ \mathbf{C}^k(\bar{\hat{\mathbf{x}}}^k, \bar{\mathbf{y}}) \nabla_{xx} \Phi(\bar{\mathbf{y}}, \bar{\hat{\mathbf{x}}}^k) \right. \\ &\quad \left. + \mathbf{M}_x \left[ \nabla_x \Phi(\bar{\mathbf{y}}, \bar{\hat{\mathbf{x}}}^k); \bar{\hat{\mathbf{x}}}^k, \bar{\mathbf{y}} \right] \right\}, \quad (13) \end{aligned}$$

$$\begin{aligned} \mathbf{B}^k &= \alpha \left\{ \mathbf{C}^k(\bar{\hat{\mathbf{x}}}^k, \bar{\mathbf{y}}) \nabla_{xy} \Phi(\bar{\mathbf{y}}, \bar{\hat{\mathbf{x}}}^k) \right. \\ &\quad \left. + \mathbf{M}_y \left[ \nabla_x \Phi(\bar{\mathbf{y}}, \bar{\hat{\mathbf{x}}}^k); \bar{\hat{\mathbf{x}}}^k, \bar{\mathbf{y}} \right] \right\}, \quad (14) \end{aligned}$$

and

$$\mathbf{V}^{k+1} = [\mathbf{I} - \mathbf{A}^k] \mathbf{V}^k + \mathbf{B}^k. \quad (15)$$

### C. Gradient algorithms with line searches

For gradient algorithms with line searches, the step size  $\alpha$  is implicitly defined by

$$\frac{\partial}{\partial \alpha} \Phi(\mathbf{y}, \hat{\mathbf{x}}^k + \alpha \mathbf{a}^k) = 0, \quad (16)$$

or equivalently,

$$[\nabla_x \Phi(\mathbf{y}, \hat{\mathbf{x}}^k + \alpha \mathbf{a}^k)]' \mathbf{a}^k = 0, \quad (17)$$

where  $'$  denotes matrix (or vector) transpose.

We can analyze  $\alpha$  using the implicit function theorem [12]. Explicitly writing  $\alpha$  as a function of  $\mathbf{y}, \mathbf{x}$  and  $\mathbf{a}$ , we have

$$[\nabla_x \Phi(\mathbf{y}, \mathbf{x} + \alpha(\mathbf{y}, \mathbf{x}, \mathbf{a}))]' \mathbf{a} = 0. \quad (18)$$

Taking the partial derivative of (18) with respect to  $\mathbf{y}, \mathbf{x}, \mathbf{a}$ , and using the chain rule, we get

$$\begin{aligned} 0 &= \mathbf{a}' \nabla_{xx} \Phi(\mathbf{y}, \mathbf{x} + \alpha(\mathbf{y}, \mathbf{x}, \mathbf{a})) \mathbf{a} \nabla_y \alpha(\mathbf{y}, \mathbf{x}, \mathbf{a}) \\ &\quad + \mathbf{a}' \nabla_{xy} \Phi(\mathbf{y}, \mathbf{x} + \alpha(\mathbf{y}, \mathbf{x}, \mathbf{a})) \mathbf{a} \quad (19) \end{aligned}$$

$$\begin{aligned} 0 &= \mathbf{a}' \nabla_{xx} \Phi(\mathbf{y}, \mathbf{x} + \alpha(\mathbf{y}, \mathbf{x}, \mathbf{a})) \mathbf{a} \nabla_x \alpha(\mathbf{y}, \mathbf{x}, \mathbf{a}) \\ &\quad + \mathbf{a}' \nabla_{xx} \Phi(\mathbf{y}, \mathbf{x} + \alpha(\mathbf{y}, \mathbf{x}, \mathbf{a})) \mathbf{a} \quad (20) \end{aligned}$$

$$\begin{aligned} 0 &= \alpha(\mathbf{y}, \mathbf{x}, \mathbf{a}) \mathbf{a}' \nabla_{xx} \Phi(\mathbf{y}, \mathbf{x} + \alpha(\mathbf{y}, \mathbf{x}, \mathbf{a})) \mathbf{a} \\ &\quad + \mathbf{a}' \nabla_{xx} \Phi(\mathbf{y}, \mathbf{x} + \alpha(\mathbf{y}, \mathbf{x}, \mathbf{a})) \mathbf{a} \nabla_a \alpha(\mathbf{y}, \mathbf{x}, \mathbf{a}) \\ &\quad + [\nabla_x \Phi(\mathbf{y}, \mathbf{x} + \alpha(\mathbf{y}, \mathbf{x}, \mathbf{a}))]' \quad (21) \end{aligned}$$

Solving the above equations, we obtain the gradient vectors of  $\alpha$  with respect to  $\mathbf{y}, \mathbf{x}$ , and  $\mathbf{a}$

$$\nabla_y \alpha(\mathbf{y}, \mathbf{x}, \mathbf{a}) = - \frac{\mathbf{a}' \nabla_{xy} \Phi(\mathbf{y}, \mathbf{x} + \alpha(\mathbf{y}, \mathbf{x}, \mathbf{a})) \mathbf{a}}{\mathbf{a}' \nabla_{xx} \Phi(\mathbf{y}, \mathbf{x} + \alpha(\mathbf{y}, \mathbf{x}, \mathbf{a})) \mathbf{a}} \mathbf{a} \quad (22)$$

$$\nabla_x \alpha(\mathbf{y}, \mathbf{x}, \mathbf{a}) = - \frac{\mathbf{a}' \nabla_{xx} \Phi(\mathbf{y}, \mathbf{x} + \alpha(\mathbf{y}, \mathbf{x}, \mathbf{a})) \mathbf{a}}{\mathbf{a}' \nabla_{xx} \Phi(\mathbf{y}, \mathbf{x} + \alpha(\mathbf{y}, \mathbf{x}, \mathbf{a})) \mathbf{a}} \mathbf{a} \quad (23)$$

$$\begin{aligned} \nabla_a \alpha(\mathbf{y}, \mathbf{x}, \mathbf{a}) &= - \frac{\alpha(\mathbf{y}, \mathbf{x}, \mathbf{a}) \mathbf{a}' \nabla_{xx} \Phi(\mathbf{y}, \mathbf{x} + \alpha(\mathbf{y}, \mathbf{x}, \mathbf{a})) \mathbf{a}}{\mathbf{a}' \nabla_{xx} \Phi(\mathbf{y}, \mathbf{x} + \alpha(\mathbf{y}, \mathbf{x}, \mathbf{a})) \mathbf{a}} \mathbf{a} \\ &\quad - \frac{[\nabla_x \Phi(\mathbf{y}, \mathbf{x} + \alpha(\mathbf{y}, \mathbf{x}, \mathbf{a}))]' }{\mathbf{a}' \nabla_{xx} \Phi(\mathbf{y}, \mathbf{x} + \alpha(\mathbf{y}, \mathbf{x}, \mathbf{a})) \mathbf{a}} \quad (24) \end{aligned}$$

Assuming the noise is small, we can approximate  $\alpha$  as

$$\begin{aligned} \alpha &\approx \bar{\alpha} + \nabla_y \alpha(\bar{\mathbf{y}}, \bar{\hat{\mathbf{x}}}^k, \bar{\mathbf{a}}^k) \mathbf{n} + \nabla_x \alpha(\bar{\mathbf{y}}, \bar{\hat{\mathbf{x}}}^k, \bar{\mathbf{a}}^k) \epsilon^k \\ &\quad + \nabla_a \alpha(\bar{\mathbf{y}}, \bar{\hat{\mathbf{x}}}^k, \bar{\mathbf{a}}^k) \tilde{\mathbf{a}}^k, \quad (25) \end{aligned}$$

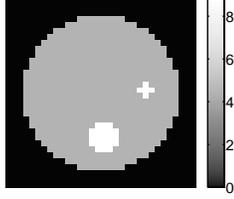


Fig. 1. The phantom image.

where  $\bar{\alpha}$  satisfies

$$[\nabla_x \Phi(\bar{\mathbf{y}}, \bar{\mathbf{x}}^k + \bar{\alpha} \bar{\mathbf{a}}^k)]' \bar{\mathbf{a}}^k = 0. \quad (26)$$

Combining (6), (10) and (25), we obtain

$$\epsilon^{k+1} \approx [\mathbf{I} - \mathbf{A}^k] \epsilon^k + \mathbf{B}^k \mathbf{n} \equiv \mathbf{V}^{k+1} \mathbf{n}, \quad (27)$$

where

$$\begin{aligned} \mathbf{A}^k &= -\bar{\alpha} \left[ \mathbf{I} + \bar{\mathbf{a}}^k \nabla_{\alpha} \alpha(\bar{\mathbf{y}}, \bar{\mathbf{x}}^k, \bar{\mathbf{a}}^k) \right] \\ &\quad \times \left\{ \mathbf{C}(\bar{\mathbf{x}}^k, \bar{\mathbf{y}}) \nabla_{xx} \Phi(\bar{\mathbf{y}}, \bar{\mathbf{x}}^k) \right. \\ &\quad \left. + \mathbf{M}_x \left[ \nabla_x \Phi(\bar{\mathbf{y}}, \bar{\mathbf{x}}^k); \bar{\mathbf{x}}^k, \bar{\mathbf{y}} \right] \right\} \\ &\quad - \bar{\mathbf{a}}^k \nabla_x \alpha(\bar{\mathbf{y}}, \bar{\mathbf{x}}^k, \bar{\mathbf{a}}^k), \\ \mathbf{B}^k &= \bar{\alpha} [\mathbf{I} + \bar{\mathbf{a}}^k \nabla_{\alpha} \alpha(\bar{\mathbf{y}}, \bar{\mathbf{x}}^k, \bar{\mathbf{a}}^k)] \\ &\quad \times \left\{ \mathbf{C}(\bar{\mathbf{x}}^k, \bar{\mathbf{y}}) \nabla_{xy} \Phi(\bar{\mathbf{y}}, \bar{\mathbf{x}}^k) \right. \\ &\quad \left. + \mathbf{M}_y \left[ \nabla_x \Phi(\bar{\mathbf{y}}, \bar{\mathbf{x}}^k); \bar{\mathbf{x}}^k, \bar{\mathbf{y}} \right] \right\} \\ &\quad + \bar{\mathbf{a}}^k \nabla_y \alpha(\bar{\mathbf{y}}, \bar{\mathbf{x}}^k, \bar{\mathbf{a}}^k) \end{aligned} \quad (28)$$

The covariance of the reconstructed image is then

$$\Sigma_{\hat{\mathbf{x}}^k} = \mathbf{V}^k \Sigma_y \left[ \mathbf{V}^k \right]', \quad (30)$$

where  $\Sigma_y$  is the covariance matrix of the data.

Substituting (11) and (26) into (5), we found that the expectation of the reconstruction is just the reconstruction of  $\bar{\mathbf{y}}$ . This is the same as those obtained in [2], [7], [12].

To simplify (28) and (29), we can use the following approximation (Approximation 2)

$$\nabla_x \Phi(\bar{\mathbf{y}}, \bar{\mathbf{x}}^k) \approx 0.$$

Then, (28) and (29) can be reduced to

$$\mathbf{A}^k = -\bar{\alpha} \mathbf{C}(\bar{\mathbf{x}}^k, \bar{\mathbf{y}}) \nabla_{xx} \Phi(\bar{\mathbf{y}}, \bar{\mathbf{x}}^k), \quad (31)$$

$$\mathbf{B}^k = \bar{\alpha} \mathbf{C}(\bar{\mathbf{x}}^k, \bar{\mathbf{y}}) \nabla_{xy} \Phi(\bar{\mathbf{y}}, \bar{\mathbf{x}}^k), \quad (32)$$

which are the same as those derived in [1].

### III. MONTE CARLO SIMULATIONS

We conducted Monte Carlo simulations to validate the theoretical results. The phantom contains a uniform disk and two hot spots (Fig. 1). It is represented using  $32 \times 32$  square pixels. The sinogram has 32 projection angles covering  $180^\circ$  and 32 detector bins per angle with a sampling distance of one pixel.

The total number of expected events in the sinogram is 80,000. Scattered and random events are not simulated. We use a MAP algorithm with a gamma prior. The prior probability of  $\mathbf{x}$  is

$$p(\mathbf{x}) = \prod_j \frac{1}{\Gamma(\beta)} (\beta/\alpha)^\beta x_j^{\beta-1} \exp(-\beta x_j/\alpha). \quad (33)$$

The mean and variance of  $x_j$  are  $\alpha$  and  $\alpha^2/\beta$ , respectively. In the simulation, we set  $\alpha$  equal to  $\sum_i y_i / \sum_j s_j$ , where  $s_j$  is the detection sensitivity at voxel  $j$ , and use  $\beta$  to control the noise in the reconstruction.

Combining with Poisson likelihood function, the objective function  $\Phi(\mathbf{y}, \mathbf{x})$  is

$$\begin{aligned} \Phi(\mathbf{y}, \mathbf{x}) &= \sum_i [y_i \log(\bar{y}_i) - \bar{y}_i] \\ &\quad + \sum_j [(\beta - 1) \log(x_j) - \beta x_j/\alpha]. \end{aligned} \quad (34)$$

At each iteration, the update direction is formed by

$$\mathbf{a}^k = \text{diag} \left[ \frac{\hat{x}_j^k}{s_j} \right] \nabla_x \Phi(\mathbf{y}, \hat{\mathbf{x}}^k), \quad (35)$$

and the step size  $\alpha$  is found by one dimensional Newton-Raphson line search. The initial image is a constant image equal to the mean of the prior.

Fig. 2 shows the comparison of the variance of the reconstruction computed using the Monte Carlo method and theoretical expressions as a function of iteration. Monte Carlo results were calculated using 8000 independent reconstructions. Fig. 3 shows the average relative root mean squared errors (RMSE) of the theoretical predictions, which is defined as

$$\text{RMSE} = \sqrt{\frac{1}{P} \sum_{i \in \text{Disk}} \left[ \frac{\text{var}_i^{\text{th}} - \text{var}_i^{\text{MC}}}{\text{var}_i^{\text{MC}}} \right]^2}, \quad (36)$$

where  $P$  is the number of pixels in the disk region,  $\text{var}_i^{\text{th}}$  and  $\text{var}_i^{\text{MC}}$  are the theoretical and Monte Carlo variances at pixel  $i$ , respectively.

The results show that in general the theoretical predictions match the Monte Carlo results. However, at early iterations, the theoretical predictions with Approximation 2 ('+'s) exhibit relatively large error because the condition of the approximation is not satisfied. In comparison, the error is significantly reduced by explicitly modeling the line search ('o's).

#### A. Effect of nonnegativity constraint

We conducted simulations to study whether the iteration-based results can model the nonnegativity constraint. We use a similar disk phantom, but with a cold spot (5% activity of the surround hot region). An OSL-MAP algorithm with a Gaussian prior [7] was used. The step size was constant. Fig. 4 shows the variance images of the Monte Carlo reconstructions and theoretical predictions with and without Approximation 2 at iteration 10, 100, and 2000. It shows that the theoretical results with Approximation 2 cannot predict the variance in the cold region.

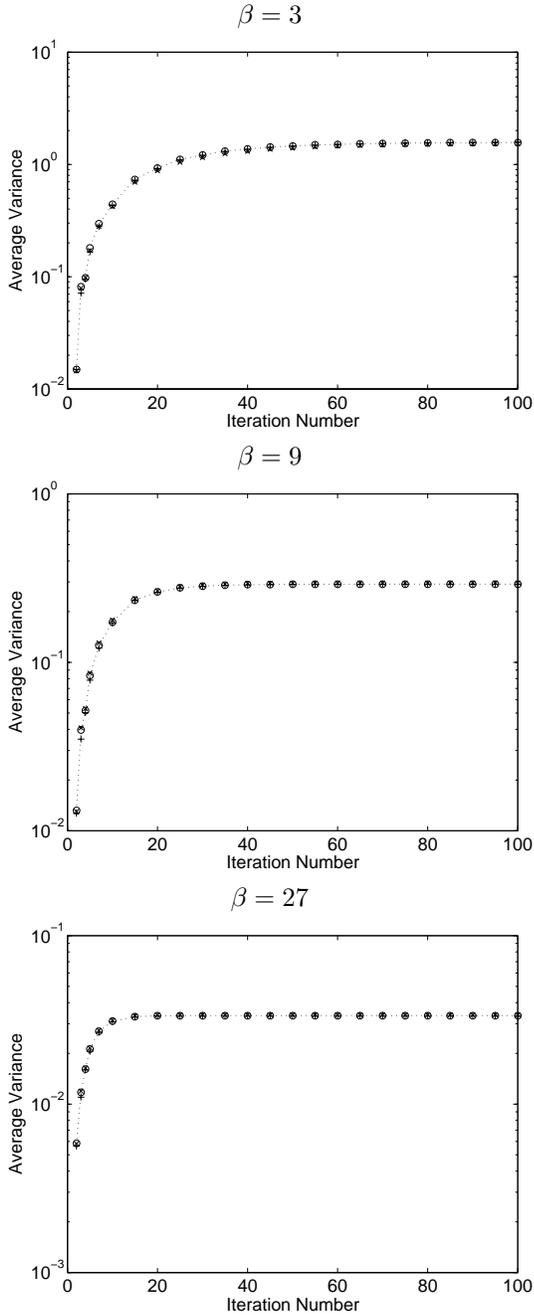


Fig. 2. Plots of variances of MAP reconstruction with gamma prior as function of iteration. The legends are Monte Carlo results ('x'), theoretical results with modeling of line search ('o'), and theoretical results with Approximation 2 ('+').

The theoretical results with (13)-(14) match the Monte Carlo results up to about 100 iterations, but eventually breaks down. These trends are more clearly shown in Fig. 5. Therefore, although the iteration-based results do not assume the gradient of the objective function is zero, the effect of nonnegativity constraint still cannot be modeled. This is mostly because the first order Taylor series expansion linearizes the solution around the noise-free reconstruction and hence cannot preserve the nonnegative property.

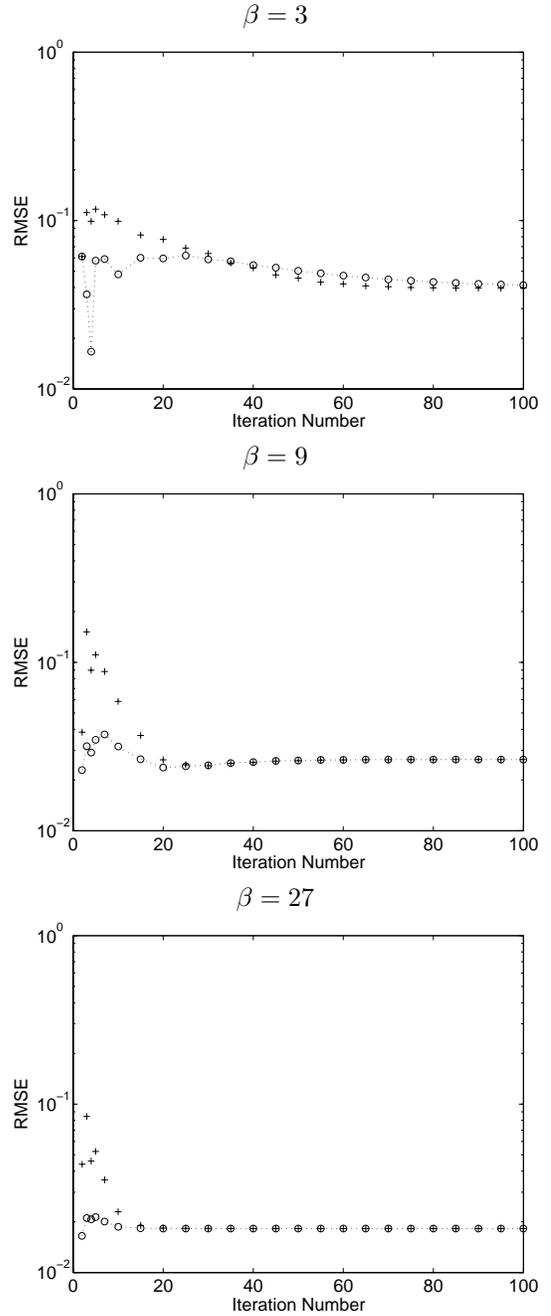


Fig. 3. Plots the relative mean squared errors of the theoretical predictions as function of iteration. The legends are the theoretical results with modeling of line search ('o'), and theoretical results with Approximation 2 ('+').

#### IV. CONCLUSIONS

We have theoretically analyzed the noise propagation in iterative reconstruction algorithms with line searches. The Monte Carlo simulation results show that by explicitly modeling the line search, we significantly reduce the error of the theoretical predictions at early iterations comparing to our previous results in [1].

#### REFERENCES

[1] J. Qi, "A unified noise analysis for iterative image estimation," *Phy. Med. Bio.*, vol. 48, pp. 3505–3519, 2003.

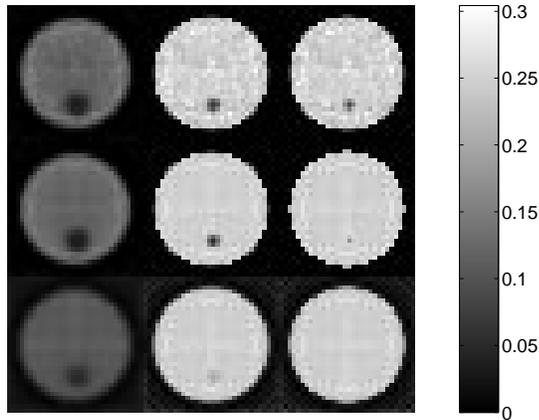


Fig. 4. Variance images at iteration 10 (left column), 100 (middle column), and 2000 (right column): Monte Carlo results (Top row), theoretical predictions with approximation 1 only (middle row), with approximation 1 and 2 (bottom row).

[2] H. H. Barrett, D. W. Wilson, and B. M. W. Tsui, "Noise properties of the EM algorithm: I. theory," *Phy. Med. Bio.*, vol. 39, pp. 833–846, 1994.

[3] L. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans Med Im*, vol. 1, pp. 113–122, 1982.

[4] K. Lange and R. Carson, "EM reconstruction algorithms for emission and transmission tomography," *Journal of Computer Assisted Tomography*, vol. 8, pp. 306–316, 1984.

[5] K. Lange, M. Bahn, and R. Little, "A theoretical study of some maximum likelihood algorithms for emission and transmission tomography," *IEEE Trans Med Im*, vol. 6, pp. 106–114, 1987.

[6] P. Green, "Bayesian reconstructions from emission tomography data using a modified EM algorithm," *IEEE Trans Med Im*, vol. 9, pp. 84–93, 1990.

[7] W. Wang and G. Gindi, "Noise analysis of MAP-EM algorithms for emission tomography," *Phy. Med. Bio.*, vol. 42, pp. 2215–2232, 1997.

[8] C. L. Byrne, "Convergent block-iterative algorithms for image reconstruction from inconsistent data," *IEEE Trans Im Proc*, vol. 6, pp. 1296–1304, 1997.

[9] C. L. Byrne, "Accelerating the EMLL algorithm and related iterative algorithms by rescaled block-iterative (RBI) methods," *IEEE Trans Im Proc*, vol. 7, pp. 100–109, 1998.

[10] E. J. Soares, C. Byrne, and S. Glick, "Noise characterization of block-iterative reconstruction algorithms: 1. theory," *IEEE Trans Med Im*, vol. 19, pp. 261–270, 2000.

[11] H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Trans Med Im*, vol. 13, pp. 601–609, 1994.

[12] J. Fessler, "Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): Applications to tomography," *IEEE Trans Im Proc*, vol. 5, pp. 493–506, 1996.

[13] J. A. Fessler and W. L. Rogers, "Spatial resolution properties of penalized-likelihood image reconstruction: Spatial-invariant tomographs," *IEEE Trans Im Proc*, vol. 9, pp. 1346–1358, 1996.

[14] J. Qi and R. M. Leahy, "A theoretical study of the contrast recovery and variance of MAP reconstructions from PET data," *IEEE Trans Med Im*, vol. 18, pp. 293–305, 1999.

[15] J. Qi and R. M. Leahy, "Resolution and noise properties of MAP reconstruction for fully 3D PET," *IEEE Trans Med Im*, vol. 19, pp. 493–506, 2000.

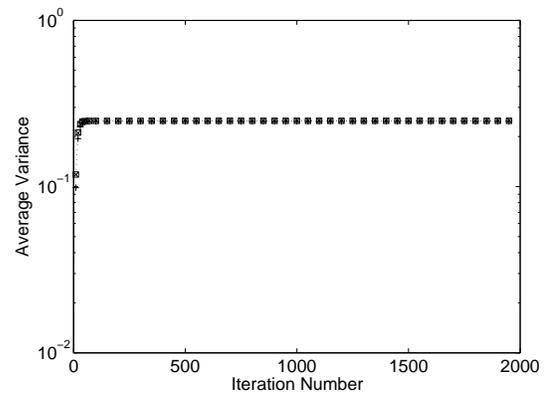
[16] J. W. Stayman and J. A. Fessler, "Regularization for uniform spatial resolution properties in penalized-likelihood image reconstruction," *IEEE Trans Med Im*, vol. 19, pp. 601–615, 2000.

[17] S. Mustafovic and K. Thielemans, "Comparison of unconventional inter-filtering methods to penalised-likelihood for space-invariant tomographs," in *Proc. IEEE NSS-MIC*, S. Metzler, Ed., pp. M11–75, 2002.

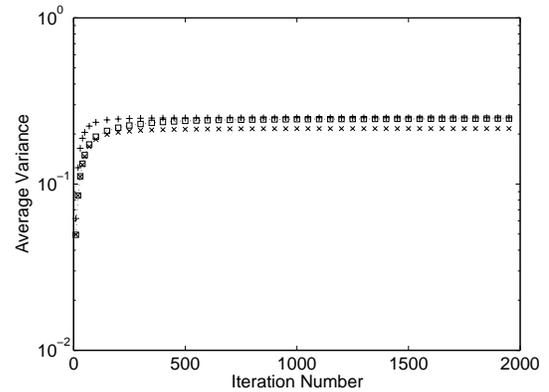
[18] J. Qi and R. H. Huesman, "Theoretical study of lesion detectability of MAP reconstruction using computer observers," *IEEE Trans Med Im*, vol. 20, pp. 815–822, 2001.

[19] I. T. Hsiao and G. Gindi, "Noise propagation from attenuation correction into PET reconstructions," *IEEE Trans. Nucl. Sci.*, vol. 49, pp. 90–97, 2002.

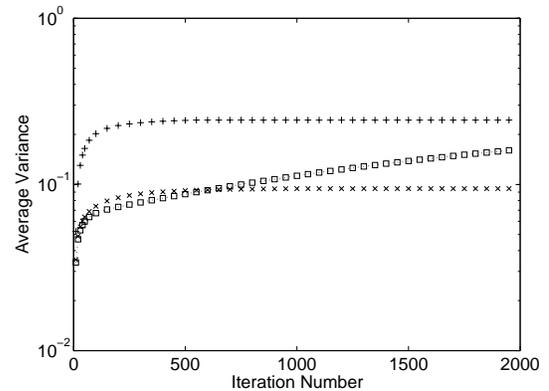
[20] J. A. Fessler and A. Yendiki, "Channelized Hotelling observer perfor-



(a)



(b)



(c)

Fig. 5. Plots of variances as function of iteration: (a) mean variances in the hot region; (b) mean variances in the cold region; (c) the minimum variances in the cold region. The legends are Monte Carlo results ('x'), theoretical results using (13)-(14) (squares), and theoretical results using (31)-(32) ('+').

mance for penalized-likelihood image reconstruction," in *Proc. IEEE NSS-MIC*, p. to appear, 2002.

[21] Y. Xing, I. T. Hsiao, and G. Gindi, "Rapid calculation of detectability in Bayesian single photon emission computed tomography," *Phy. Med. Bio.*, vol. 48, pp. 3755–3773, 2003.

[22] J. Browne and A. R. De Pierro, "A row-action alternative to the EM algorithm for maximizing likelihoods in emission tomography," *IEEE Trans Med Im*, vol. 15, pp. 687–699, 1996.

[23] A. R. De Pierro and M. E. B. Yamagishi, "Fast EM-like methods for maximum a posteriori estimates in emission tomography," *IEEE Trans Med Im*, vol. 20, pp. 280–288, 2001.