A Unified Noise Analysis for Iterative Image Estimation

Jinyi Qi

Department of Nuclear Medicine and Functional Imaging, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.

E-mail: jqi@lbl.gov

Abstract. Iterative image estimation methods have been widely used in emission tomography. Accurate estimation of the uncertainty of the reconstructed images is essential for quantitative applications. While a theoretical approach has been developed to analyze the noise propagation from iteration to iteration, the current results are limited to only a few iterative algorithms that have an explicit multiplicative update equation. This paper presents a theoretical noise analysis that is applicable to a wide range of preconditioned gradient-type algorithms. One advantage is that the proposed method does not require an explicit expression of the preconditioner and hence it is applicable to algorithms that involve line searches. By deriving a fixed-point expression from the iteration-based results, we show that the iteration-based noise analysis is consistent with fixed-point analysis. Examples in emission tomography and transmission tomography are shown. The results are validated using Monte Carlo simulations.

1. Introduction

In emission tomography (e.g., positron emission tomography [PET] and single photon emission computed tomography [SPECT]) data are not exactly line-integrals of the tracer distribution. Hence the images obtained by the filtered backprojection method have limited resolution and may contain artifacts (e.g., radial elongation in PET). Iterative image estimation methods have been developed to improve the image quality by accurately modeling the system response and noise distribution. It is essential to estimate the uncertainty of the reconstructed images for quantitative applications [Huesman, 1984, Mazoyer et al, 1986, Carson et al, 1993, Llacer et al, 1993, Maitra and O'Sullivan, 1998]. For the linear Landweber algorithm, Abbey and Barrett [Abbey and Barrett, 1995] presented exact expressions of the expectation and covariance. However, theoretical analysis of nonlinear iterative methods has been difficult. Below we briefly review some progress that has been made in this direction.

Barrett et al. [Barrett et al, 1994] derived approximate formulae for the mean and covariance of the maximum likelihood (ML) expectation maximization (EM)

reconstruction [Shepp and Vardi, 1982, Lange and Carson, 1984] as a function of the iteration number. The same approach was extended to one-step-late maximum *a posteriori* (OSL-MAP) EM algorithm [Green, 1990] with gamma and Gaussian priors by Wang and Gindi [Wang and Gindi, 1997] and more recently to unregularized block iterative algorithms by Soares et al. [Soares et al, 2000]. 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 [Hudson and Larkin, 1994]. However, the current results are limited to a few iterative algorithms with explicit multiplicative update equations, and the whole analysis in [Barrett et al, 1994] must be repeated for a slightly changed algorithm. In addition, the current method is not applicable to gradient-type algorithms that involve line searches.

An alternative approach was proposed by Fessler and Rogers [Fessler, 1996, Fessler and Rogers, 1996], 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 [Qi and Leahy, 1999, Qi and Leahy, 2000] extended this approach by deriving simplified expressions for the local impulse response function and covariance using Fourier transform. Stayman and Fessler [Stayman and Fessler, 2000] used a similar approximation in designing penalty functions for uniform resolution. Other extensions of the fixed-point analysis include [Hsiao and Gindi, 2000, Qi and Huesman, 2001, Fessler and Yendiki, 2002, Xing and Gindi, 2002]. 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.

The above two approaches are complementary. We would expect that if we iterate the algorithm until it converges, the iteration-based results in [Barrett et al, 1994, Wang and Gindi, 1997] should match with the fixed-point results in [Fessler, 1996]. This was shown in [Barrett et al, 1994] for the ML case. However, for MAP reconstruction considered in [Wang and Gindi, 1997] and [Fessler, 1996], some discrepancy exists (see Section 4.2).

Building on existing work, here we derive unified formulae for calculating the mean and variance of the image estimated using iterative algorithms. The iteration-based results are applicable to a wide range of preconditioned gradient-type algorithms, which includes ML-EM and MAP-EM as special cases. They are extendible to ordered subsets methods [Hudson and Larkin, 1994, Browne and De Pierro, 1996]. In addition to emission tomography, the proposed method is also applicable to iterative transmission reconstruction and image restoration.

This paper is organized as follows. Section 2 presents the theoretical analysis of noise propagation for the preconditioned gradient algorithms. The relation with fixed-point analysis is discussed in Section 3. In Section 4 we show some examples of the application of the theoretical analysis to the emission tomography problem with comparisons to previous results. The cause of the discrepancy between the results in

[Wang and Gindi, 1997] and [Fessler, 1996] is examined. An application to transmission tomography is shown in Section 5. In Section 6 we compare the theoretical results with Monte Carlo simulation results. Finally, the conclusions are presented in Section 7.

2. A Unified Approach to Noise Analysis

In image estimation problems, the image is often estimated through maximizing an objective function

$$\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}\geq 0} \left[L(\boldsymbol{y}|\boldsymbol{x}) - U(\boldsymbol{x})\right], \tag{1}$$

where $\boldsymbol{x}$ is the tracer distribution, $\boldsymbol{y}$ is the noisy data, $L(\boldsymbol{y}|\boldsymbol{x})$ is the log likelihood function (or data match term), and $U(\boldsymbol{x})$ is the prior function (or penalty term). When $U(\boldsymbol{x}) = 0$, (1) reduces to the ML estimate.

A preconditioned gradient ascent algorithm that solves (1) can be written as

$$\hat{\boldsymbol{x}}^{k+1} = \hat{\boldsymbol{x}}^k + \alpha \boldsymbol{C}^k(\hat{\boldsymbol{x}}^k)\left[\nabla_x L(\boldsymbol{y}|\boldsymbol{x}) - \nabla_x U(\boldsymbol{x})\right]_{\boldsymbol{x}=\hat{\boldsymbol{x}}^k}, \tag{2}$$

where $\alpha > 0$ is a fixed step size, $\boldsymbol{C}^k(\hat{\boldsymbol{x}}^k)$ is a positive definite matrix (preconditioner), and ∇_x denotes gradient operation with respect to $\boldsymbol{x}$. Note that the preconditioner $\boldsymbol{C}^k(\hat{\boldsymbol{x}}^k)$ can be a function of current image estimate. Both the ML-EM and MAP-EM in emission tomography are special cases of (2). For simplicity, we will write $\nabla_x L(\boldsymbol{y}|\boldsymbol{x})|_{\boldsymbol{x}=\hat{\boldsymbol{x}}^k}$ as $\nabla_x L(\boldsymbol{y}|\hat{\boldsymbol{x}}^k)$, and $\nabla_x U(\boldsymbol{x})|_{\boldsymbol{x}=\hat{\boldsymbol{x}}^k}$ as $\nabla_x U(\hat{\boldsymbol{x}}^k)$ in the rest of the paper.

To analyze the noise, we write

$$\hat{\boldsymbol{x}}^k = \bar{\hat{\boldsymbol{x}}}^k + \epsilon^k \tag{3}$$

$$\boldsymbol{y} = \bar{\boldsymbol{y}} + \boldsymbol{n} \tag{4}$$

where $\bar{\hat{\boldsymbol{x}}}^k$ and $\bar{\boldsymbol{y}}$ are the expectations of the image estimate $\hat{\boldsymbol{x}}^k$ and data $\boldsymbol{y}$, respectively, and ϵ^k and $\boldsymbol{n}$ are zero-mean noise vectors. Similar to [Barrett et al, 1994, Wang and Gindi, 1997, Fessler, 1996, Soares et al, 2000], we assume that the noise level in $\hat{\boldsymbol{x}}^k$ is low and hence the first-order Taylor series approximation can be used. Thus, we have

$$\nabla_x L(\boldsymbol{y}|\hat{\boldsymbol{x}}^k) \approx \nabla_x L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^k) + \nabla_{xy} L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^k)\boldsymbol{n} + \nabla_{xx} L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^k)\epsilon^k, \tag{5}$$

$$\nabla_x U(\hat{\boldsymbol{x}}^k) \approx \nabla_x U(\bar{\hat{\boldsymbol{x}}}^k) + \nabla_{xx} U(\bar{\hat{\boldsymbol{x}}}^k)\epsilon^k, \tag{6}$$

$$\boldsymbol{C}^k(\hat{\boldsymbol{x}}^k) \approx \boldsymbol{C}^k(\bar{\hat{\boldsymbol{x}}}^k) + \boldsymbol{C}^k_x(\epsilon^k; \bar{\hat{\boldsymbol{x}}}^k), \tag{7}$$

where the (j,l)th element of the operator ∇_{xx} is $\frac{\partial^2}{\partial x_j \partial x_l}$, the (j,l)th element of the operator ∇_{xy} is $\frac{\partial^2}{\partial x_j \partial y_l}$, and the (j,l)th element of $\boldsymbol{C}^k_x(\epsilon^k; \boldsymbol{x})$ is $\sum_m \epsilon^k_m \frac{\partial}{\partial x_m} C^k_{j,l}(\boldsymbol{x})$. Note that (5) may require $\boldsymbol{n}$ to be small if $\nabla_x L(\boldsymbol{y}|\boldsymbol{x})$ is not a linear function of $\boldsymbol{y}$. Fortunately, for Poisson and Gaussian distributions, which are commonly used in image estimation, $\nabla_x L(\boldsymbol{y}|\boldsymbol{x})$ is linear with respect to $\boldsymbol{y}$. Hence the following results are applicable to low-count situations in emission tomography with proper regularization.

Substituting (5)-(7) into (2) and dropping all second-order noise terms results in

$$\begin{aligned}\bar{\hat{\boldsymbol{x}}}^{k+1} + \epsilon^{k+1} \approx\ & \bar{\hat{\boldsymbol{x}}}^{k} + \alpha \boldsymbol{C}^k(\bar{\hat{\boldsymbol{x}}}^k)\left[\nabla_x L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^k) - \nabla_x U(\bar{\hat{\boldsymbol{x}}}^k)\right] \\ & + \epsilon^k + \alpha \boldsymbol{C}^k(\bar{\hat{\boldsymbol{x}}}^k)\left[\nabla_{xy} L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^k)\boldsymbol{n} + \nabla_{xx} L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^k)\epsilon^k - \nabla_{xx} U(\bar{\hat{\boldsymbol{x}}}^k)\epsilon^k\right] \\ & + \alpha \boldsymbol{C}_x^k(\epsilon^k; \bar{\hat{\boldsymbol{x}}}^k)\left[\nabla_x L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^k) - \nabla_x U(\bar{\hat{\boldsymbol{x}}}^k)\right]. \end{aligned} \tag{8}$$

Separating the signal from the noise, we have

$$\bar{\hat{\boldsymbol{x}}}^{k+1} \approx \bar{\hat{\boldsymbol{x}}}^{k} + \alpha \boldsymbol{C}^k(\bar{\hat{\boldsymbol{x}}}^k)\left[\nabla_x L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^k) - \nabla_x U(\bar{\hat{\boldsymbol{x}}}^k)\right] \tag{9}$$

$$\begin{aligned}\epsilon^{k+1} \approx\ & \epsilon^k + \alpha \boldsymbol{C}^k(\bar{\hat{\boldsymbol{x}}}^k)\left[\nabla_{xx} L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^k) - \nabla_{xx} U(\bar{\hat{\boldsymbol{x}}}^k)\right]\epsilon^k \\ & + \alpha \boldsymbol{M}\left[\nabla_x L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^k) - \nabla_x U(\bar{\hat{\boldsymbol{x}}}^k); \bar{\hat{\boldsymbol{x}}}^k\right]\epsilon^k \\ & + \alpha \boldsymbol{C}^k(\bar{\hat{\boldsymbol{x}}}^k)\nabla_{xy} L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^k)\boldsymbol{n}, \end{aligned} \tag{10}$$

where the (j,l)th element of $\boldsymbol{M}[\boldsymbol{g};\boldsymbol{x}]$ is $\sum_m g_m \partial C^k_{j,m}(\boldsymbol{x})/\partial x_l$.

Equation (9) shows that under the low-noise assumption, the expectation of the reconstructed image is equal to the reconstruction of the expectation of the data. Equation (10) is a linear update equation for the noise vector, which can be written in the form of

$$\epsilon^{k+1} \approx [\boldsymbol{I} - \boldsymbol{A}^k]\epsilon^k + \boldsymbol{B}^k\boldsymbol{n} \equiv \boldsymbol{V}^{k+1}\boldsymbol{n}, \tag{11}$$

where

$$\boldsymbol{A}^k = -\alpha \boldsymbol{C}^k(\bar{\hat{\boldsymbol{x}}}^k)\left[\nabla_{xx} L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^k) - \nabla_{xx} U(\bar{\hat{\boldsymbol{x}}}^k)\right] - \alpha \boldsymbol{M}\left[\nabla_x L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^k) - \nabla_x U(\bar{\hat{\boldsymbol{x}}}^k); \bar{\hat{\boldsymbol{x}}}^k\right], \tag{12}$$

$$\boldsymbol{B}^k = \alpha \boldsymbol{C}^k(\bar{\hat{\boldsymbol{x}}}^k)\nabla_{xy} L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^k), \tag{13}$$

and

$$\boldsymbol{V}^{k+1} = [\boldsymbol{I} - \boldsymbol{A}^k]\boldsymbol{V}^k + \boldsymbol{B}^k. \tag{14}$$

The covariance of the reconstructed image is then

$$\Sigma_{\hat{\boldsymbol{x}}^k} = \boldsymbol{V}^k \Sigma_y \left[\boldsymbol{V}^k\right]', \tag{15}$$

where Σ_y is the covariance matrix of the data.

The evaluation of (12) depends on the log likelihood function, the prior term, as well as the structure of the preconditioner. In the following sections, we will study (12) for linear preconditioners and derive a simplified expression for other complicated preconditioners.

2.1. Algorithms with Linear Preconditioners

For the algorithms of which

$$\boldsymbol{C}^k(\hat{\boldsymbol{x}}^k) = \operatorname{diag}\left[\hat{\boldsymbol{x}}^k\right]\boldsymbol{D}^k, \tag{16}$$

where $\operatorname{diag}[\boldsymbol{x}]$ denotes a diagonal matrix with the (j,j)th element being x_j and $\boldsymbol{D}^k$ is a positive definite matrix, $\boldsymbol{M}[\boldsymbol{g};\hat{\boldsymbol{x}}^k]$ reduces to

$$\boldsymbol{M}[\boldsymbol{g};\hat{\boldsymbol{x}}^k] = \operatorname{diag}\left[\boldsymbol{D}^k\boldsymbol{g}\right].$$

The ML-EM algorithm and the MAP-EM algorithm with an independent gamma prior [Lange et al, 1987] are examples of this type of algorithm.

Using (2.1), equation (12) can be simplified to

$$\begin{aligned}\boldsymbol{A}^k = & -\alpha \left\{ \operatorname{diag}\left[\bar{\hat{\boldsymbol{x}}}^k\right] \boldsymbol{D}^k \left[\nabla_{xx} L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^k) - \nabla_{xx} U(\bar{\hat{\boldsymbol{x}}}^k)\right] \right. \\ & \left. + \operatorname{diag}\left(\boldsymbol{D}^k \left[\nabla_x L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^k) - \nabla_x U(\bar{\hat{\boldsymbol{x}}}^k)\right]\right)\right\}. \end{aligned} \tag{17}$$

2.2. Simplification for Other Preconditioners

For a preconditioner that does not have the form of (16), the computation of $\boldsymbol{M}[\boldsymbol{g}; \hat{\boldsymbol{x}}^k]$ can be difficult. To simplify calculation, we note that the overall contribution of

$$\boldsymbol{M}\left[\nabla_x L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^k) - \nabla_x U(\bar{\hat{\boldsymbol{x}}}^k); \bar{\hat{\boldsymbol{x}}}^k\right] \epsilon^k$$

is quite small. At initial iterations ϵ^k is very small (when starting from a uniform image), so ϵ^{k+1} is dominated by $\boldsymbol{B}^k\boldsymbol{n}$; and at later iterations $\nabla_x L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^k) - \nabla_x U(\bar{\hat{\boldsymbol{x}}}^k)$ approaches zero (if the algorithm increases the objective function). Therefore, we will use the following approximation

$$\boldsymbol{M}\left[\nabla_x L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^k) - \nabla_x U(\bar{\hat{\boldsymbol{x}}}^k); \bar{\hat{\boldsymbol{x}}}^k\right] \epsilon^k \approx 0. \tag{18}$$

Equation (18) is satisfied when $\boldsymbol{C}^k(\hat{\boldsymbol{x}}^k)$ is independent of $\hat{\boldsymbol{x}}^k$, or $\nabla_x L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^k) - \nabla_x U(\bar{\hat{\boldsymbol{x}}}^k) \approx 0$. For the ML-EM algorithm the latter is equivalent to "approximation 2" used in [Barrett et al, 1994] that assumes the projection of the reconstructed image is equal to the noise-free projection. For regularized algorithms it takes into account the effect of the prior function. As demonstrated in the Monte Carlo simulations [Wilson et al, 1994, Wang and Gindi, 1997], this approximation is accurate enough for most situations, especially for regularized image reconstruction where the objective function is better conditioned and hence gradient-type algorithms converge faster.

With (18), equation (12) reduces to

$$\boldsymbol{A}^k = -\alpha \boldsymbol{C}^k(\bar{\hat{\boldsymbol{x}}}^k)\left[\nabla_{xx} L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^k) - \nabla_{xx} U(\bar{\hat{\boldsymbol{x}}}^k)\right]. \tag{19}$$

2.3. Generalization to Data-Dependent Preconditioners

The above analysis has assumed that the preconditioner is conditionally independent of data $\boldsymbol{y}$ given the current estimate $\boldsymbol{x}^k$. Here we extend the noise analysis to preconditioners that are explicitly dependent on data. We rewrite the update equation in (2) as

$$\hat{\boldsymbol{x}}^{k+1} = \hat{\boldsymbol{x}}^k + \boldsymbol{C}^k(\hat{\boldsymbol{x}}^k, \boldsymbol{y})\left[\nabla_x L(\boldsymbol{y}|\hat{\boldsymbol{x}}^k) - \nabla_x U(\hat{\boldsymbol{x}}^k)\right]. \tag{20}$$

Equation (20) includes the gradient-type algorithms that use data-dependent preconditioners [Lange et al, 1987] and/or determine the step sizes using line searches [Lange, 1990]. Note that the variable step size has been lumped into $\boldsymbol{C}^k(\hat{\boldsymbol{x}}^k, \boldsymbol{y})$.

Using the first-order Taylor series approximation, the data-dependent preconditioner can be approximated by

$$\boldsymbol{C}^k(\hat{\boldsymbol{x}}^k, \boldsymbol{y}) \approx \boldsymbol{C}^k(\bar{\hat{\boldsymbol{x}}}^k, \bar{\boldsymbol{y}}) + \boldsymbol{C}^k_x(\epsilon^k; \bar{\hat{\boldsymbol{x}}}^k, \bar{\boldsymbol{y}}) + \boldsymbol{C}^k_y(\boldsymbol{n}; \bar{\hat{\boldsymbol{x}}}^k, \bar{\boldsymbol{y}}), \tag{21}$$

where the (j,l)th element of $\boldsymbol{C}_x^k(\epsilon^k;\boldsymbol{x},\boldsymbol{y})$ is $\sum_m \epsilon_m^k \partial C_{j,l}^k(\boldsymbol{x},\boldsymbol{y})/\partial x_m$ and the (j,l)th element of $\boldsymbol{C}_y^k(\boldsymbol{n};\boldsymbol{x},\boldsymbol{y})$ is $\sum_m n_m \partial C_{j,l}^k(\boldsymbol{x},\boldsymbol{y})/\partial y_m$. Note this approximation requires that ϵ^k and $\boldsymbol{n}$ are within the linear range of $\boldsymbol{C}^k(\hat{\boldsymbol{x}}^k,\boldsymbol{y})$.

With this result, the new noise propagation equation becomes

$$\begin{aligned}\epsilon^{k+1} \approx \epsilon^k &+ \boldsymbol{C}(\bar{\hat{\boldsymbol{x}}}^k,\bar{\boldsymbol{y}})\left[\nabla_{xx}L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^k) - \nabla_{xx}U(\bar{\hat{\boldsymbol{x}}}^k)\right]\epsilon^k \\ &+ \boldsymbol{C}_x^k(\epsilon^k;\bar{\hat{\boldsymbol{x}}}^k,\bar{\boldsymbol{y}})\left[\nabla_{x}L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^k) - \nabla_{x}U(\bar{\hat{\boldsymbol{x}}}^k)\right] \\ &+ \boldsymbol{C}(\bar{\hat{\boldsymbol{x}}}^k,\bar{\boldsymbol{y}})\nabla_{xy}L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^k)\boldsymbol{n} \\ &+ \boldsymbol{C}_y^k(\boldsymbol{n};\bar{\hat{\boldsymbol{x}}}^k,\bar{\boldsymbol{y}})\left[\nabla_{x}L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^k) - \nabla_{x}U(\bar{\hat{\boldsymbol{x}}}^k)\right].\end{aligned} \tag{22}$$

For data-dependent preconditioners, the computation of $\boldsymbol{C}_x^k(\epsilon^k;\bar{\hat{\boldsymbol{x}}}^k,\bar{\boldsymbol{y}})$ and $\boldsymbol{C}_y^k(\boldsymbol{n};\bar{\hat{\boldsymbol{x}}}^k,\bar{\boldsymbol{y}})$ can be very difficult. If the algorithm involves line searches, explicit expressions may not even exist. To simplify the result, we assume

$$\boldsymbol{C}_x^k(\epsilon^k;\bar{\hat{\boldsymbol{x}}}^k,\bar{\boldsymbol{y}})\left[\nabla_{x}L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^k) - \nabla_{x}U(\bar{\hat{\boldsymbol{x}}}^k)\right] \approx 0, \tag{23}$$

$$\boldsymbol{C}_y^k(\boldsymbol{n};\bar{\hat{\boldsymbol{x}}}^k,\bar{\boldsymbol{y}})\left[\nabla_{x}L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^k) - \nabla_{x}U(\bar{\hat{\boldsymbol{x}}}^k)\right] \approx 0. \tag{24}$$

Both assumptions are satisfied when $\nabla_x L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^k) - \nabla_x U(\bar{\hat{\boldsymbol{x}}}^k) \approx 0$ (equivalent to approximation 2 in [Barrett et al, 1994]).

With the above approximations, (22) reduces to

$$\begin{aligned}\epsilon^{k+1} \approx \epsilon^k &+ \boldsymbol{C}(\bar{\hat{\boldsymbol{x}}}^k,\bar{\boldsymbol{y}})\left[\nabla_{xx}L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^k) - \nabla_{xx}U(\bar{\hat{\boldsymbol{x}}}^k)\right]\epsilon^k \\ &+ \boldsymbol{C}(\bar{\hat{\boldsymbol{x}}}^k,\bar{\boldsymbol{y}})\nabla_{xy}L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^k)\boldsymbol{n},\end{aligned} \tag{25}$$

and we get

$$\boldsymbol{A}^k = -\boldsymbol{C}(\bar{\hat{\boldsymbol{x}}}^k,\bar{\boldsymbol{y}})\left[\nabla_{xx}L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^k) - \nabla_{xx}U(\bar{\hat{\boldsymbol{x}}}^k)\right], \tag{26}$$

$$\boldsymbol{B}^k = \boldsymbol{C}(\bar{\hat{\boldsymbol{x}}}^k,\bar{\boldsymbol{y}})\nabla_{xy}L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^k), \tag{27}$$

which are essentially the same as (19) and (13).

2.4. Remarks

Equations (11)-(14) [also (26) and (27)] are the main results in this paper. They are valid for a wide range of preconditioned gradient-type algorithms. Using the proper likelihood and prior functions, most results in [Barrett et al, 1994, Wang and Gindi, 1997, Soares et al, 2000] can be directly obtained from these expressions (see Section 4). One advantage of our method is that the theoretical analysis does not require an explicit expression of the preconditioner and hence is applicable to algorithms that use line searches.

Moreover, the theoretical analysis is also applicable to the algorithms that use a different objective function at each iteration, such as the ordered subsets gradient methods [Hudson and Larkin, 1994, Byrne, 1997]. While such algorithms may not converge, the iteration-based noise analysis is valid as long as the noise in the reconstruction is small.

The fundamental assumption of the above analysis is that the noise in the image is within the linear range of the gradient of the objective function and the preconditioner so that the first-order Taylor series expansion can be used. As a result, the accuracy of these expressions depends on the objective function and the algorithm. For algorithms optimizing a quadratic objective function using a linear preconditioner, the above assumption is always satisfied if there is no nonnegativity constraint, in which case the theoretical results are exact. As the objective function deviates from quadratic form and the preconditioner becomes nonlinear, the accuracy of the theoretical prediction drops. In emission tomography with Poisson likelihood function we found that the above assumption is often satisfied as long as the noise is small compared to the expected image.

Another issue that requires the noise in reconstruction to be small is the nonnegativity constraint. We can see that once we use the first order Taylor series expansion and reach equation (8), the nonnegativity of $\bar{\hat{\boldsymbol{x}}}^{k+1} + \epsilon^{k+1}$ is only satisfied when the noise term in (10) is small compared to the mean reconstruction. Monte Carlo simulations in [Wilson et al, 1994] show that this condition is almost always true for clinically useful images. In practice one should check the above condition by comparing the estimated variance with the reconstructed image. For the rare cases where the condition is not met (this often occurs at regions that have extremely low activity [Fessler, 1996, Qi and Leahy, 1999]), a method similar to that developed in [Qi and Leahy, 2000] may be used to compensate for the nonnegativity constraint.

3. Relation to the Fixed-Point Analysis

We can obtain the fixed-point expression of the noise directly from the iteration-based result. Let us assume that the step size and preconditioner are properly chosen so that (2) converges [Luenberger, 1984]. Setting (10) to the fixed-point and using the fixed-point condition $\nabla_x L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^\infty) - \nabla_x U(\bar{\hat{\boldsymbol{x}}}^\infty) = 0$, we get

$$\begin{aligned}\epsilon^\infty \approx \epsilon^\infty &+ \alpha \boldsymbol{C}^k(\bar{\hat{\boldsymbol{x}}}^\infty)\left[\nabla_{xx} L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^\infty) - \nabla_{xx} U(\bar{\hat{\boldsymbol{x}}}^\infty)\right]\epsilon^\infty \\ &+ \alpha \boldsymbol{C}^k(\bar{\hat{\boldsymbol{x}}}^\infty)\nabla_{xy} L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^\infty)\boldsymbol{n},\end{aligned} \tag{28}$$

i.e.,

$$\epsilon^\infty \approx \left[-\nabla_{xx} L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^\infty) + \nabla_{xx} U(\bar{\hat{\boldsymbol{x}}}^\infty)\right]^{-1} \nabla_{xy} L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^\infty)\boldsymbol{n}, \tag{29}$$

and the covariance of the reconstructed noise is

$$\begin{aligned}\Sigma_\infty \approx &\left[-\nabla_{xx} L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^\infty) + \nabla_{xx} U(\bar{\hat{\boldsymbol{x}}}^\infty)\right]^{-1} \nabla_{xy} L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^\infty) \\ &\Sigma_y [\nabla_{xy} L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^\infty)]' \left[-\nabla_{xx} L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^\infty) + \nabla_{xx} U(\bar{\hat{\boldsymbol{x}}}^\infty)\right]^{-1}.\end{aligned} \tag{30}$$

Equation (30) is exactly the same as the fixed-point result obtained in [Fessler, 1996]. Therefore, it shows that if the algorithm is iterated to convergence, the iteration-based noise analysis and the fixed-point noise analysis are equivalent. While the fixed-point expression (30) is obtained from the gradient algorithm of (2), (30) is valid for all

convergent algorithms, provided that the solution of (1) is unique. This property is partly demonstrated by the disappearance of the step size α and the preconditioner $\boldsymbol{C}^k(\hat{\boldsymbol{x}}^k)$.

4. Applications: Emission Tomography

In emission tomography, data are modeled as independent Poisson random variables with the expectation $\bar{\boldsymbol{y}}$ related to the unknown tracer distribution $\boldsymbol{x}$ through an affine transform

$$\bar{\boldsymbol{y}} = \boldsymbol{P}\boldsymbol{x} + \boldsymbol{r}, \tag{31}$$

where $\boldsymbol{P}$ is the detection probability matrix with the (i,j)th element equal to the probability of detecting an event from the jth voxel at the ith measurement, and $\boldsymbol{r}$ accounts for the presence of scattered and random events in the data. To be consistent with [Barrett et al, 1994, Wang and Gindi, 1997, Soares et al, 2000], we set $\boldsymbol{r} = 0$.

The log likelihood function of the independent Poisson distribution is

$$L(\boldsymbol{y}|\boldsymbol{x}) = \sum_i \left(y_i \log \bar{y}_i - \bar{y}_i - \log y_i!\right), \tag{32}$$

from which one can derive

$$\nabla_x L(\boldsymbol{y}|\boldsymbol{x}) = \boldsymbol{P}'\mathrm{diag}\left[\boldsymbol{P}\boldsymbol{x}\right]^{-1}\boldsymbol{y} - \boldsymbol{s}, \tag{33}$$

$$\nabla_{xx} L(\boldsymbol{y}|\boldsymbol{x}) = -\boldsymbol{P}'\mathrm{diag}\left[\boldsymbol{P}\boldsymbol{x}\right]^{-2}\mathrm{diag}\left[\boldsymbol{y}\right]\boldsymbol{P}, \tag{34}$$

$$\nabla_{xy} L(\boldsymbol{y}|\boldsymbol{x}) = \boldsymbol{P}'\mathrm{diag}\left[\boldsymbol{P}\boldsymbol{x}\right]^{-1}, \tag{35}$$

where $\boldsymbol{s} = \boldsymbol{P}'\mathbf{1}$ and $\mathbf{1}$ is a vector with all elements equal to unity.

4.1. ML-EM

For the ML-EM algorithm, we have $\alpha\boldsymbol{C}^k(\boldsymbol{x}^k) = \mathrm{diag}\left[\boldsymbol{x}^k\right]\mathrm{diag}\left[\boldsymbol{s}^{-1}\right]$ and $U(\boldsymbol{x}) = 0$. Substituting these expressions and (33)-(35) into (17) and (13), we get

$$\begin{aligned}\boldsymbol{A}^k = {} & \mathrm{diag}\left[\bar{\hat{\boldsymbol{x}}}^k\right]\mathrm{diag}\left[\boldsymbol{s}^{-1}\right]\boldsymbol{P}'\mathrm{diag}\left[\boldsymbol{P}\bar{\hat{\boldsymbol{x}}}^k\right]^{-2}\mathrm{diag}\left[\bar{\boldsymbol{y}}\right]\boldsymbol{P} \\ & - \mathrm{diag}\left[\bar{\hat{\boldsymbol{x}}}^{k+1}\right]\mathrm{diag}\left[\bar{\hat{\boldsymbol{x}}}^k\right]^{-1} + \boldsymbol{I},\end{aligned} \tag{36}$$

$$\boldsymbol{B}^k = \mathrm{diag}\left[\bar{\hat{\boldsymbol{x}}}^k\right]\mathrm{diag}\left[\boldsymbol{s}^{-1}\right]\boldsymbol{P}'\mathrm{diag}\left[\boldsymbol{P}\bar{\hat{\boldsymbol{x}}}^k\right]^{-1}. \tag{37}$$

These are exactly the same results as those derived in [Barrett et al, 1994] using approximation 1 only, although the noise in [Barrett et al, 1994] is expressed in the logarithm of the reconstructed image.

When approximation (18) is used, equation (36) reduces to

$$\boldsymbol{A} = -\mathrm{diag}\left[\bar{\hat{\boldsymbol{x}}}^k\right]\mathrm{diag}\left[\boldsymbol{s}^{-1}\right]\boldsymbol{P}'\mathrm{diag}\left[\boldsymbol{P}\bar{\hat{\boldsymbol{x}}}^k\right]^{-1}\boldsymbol{P},$$

which is the same as equation (28) in [Barrett et al, 1994] with some change of variables (note that under approximation 2, $\bar{\hat{\boldsymbol{x}}}^k = \bar{\hat{\boldsymbol{x}}}^{k+1}$).

4.2. One-Step-Late MAP-EM

For MAP-EM algorithms with a Gibbs prior that is defined on neighboring voxels, the exact solution of the M-step in the EM algorithm is difficult to solve. Green [Green, 1990] suggested a one-step-late (OSL) strategy to solve the problem. The OSL-MAP-EM algorithm can be written in the form of (2) with

$$\boldsymbol{C}^k(\boldsymbol{x}^k) = \text{diag}\left[\boldsymbol{x}^k\right] \text{diag}\left[\boldsymbol{s} + \nabla_x U(\hat{\boldsymbol{x}}^k)\right]^{-1} \tag{38}$$

and $\alpha = 1$. The convergence of the algorithm can be achieved if the step size is determined by a one-dimensional line search [Lange, 1990]. Wang and Gindi [Wang and Gindi, 1997] studied the noise propagation of OSL-MAP-EM algorithm ($\alpha = 1$) with gamma and Gaussian priors. Here we derive more generalized results.

While (38) is not a linear preconditioner in the form of (16), an explicit expression of $\boldsymbol{M}[\boldsymbol{g};\boldsymbol{x}]$ can still be derived. After some calculation, we get

$$\boldsymbol{M}[\boldsymbol{g};\hat{\boldsymbol{x}}^k] = \text{diag}\left[\frac{\boldsymbol{g}}{\boldsymbol{s} + \nabla_x U(\hat{\boldsymbol{x}}^k)}\right]\left[\boldsymbol{I} - \boldsymbol{C}^k(\boldsymbol{x}^k)\nabla_{xx}U(\bar{\hat{\boldsymbol{x}}}^k)\right]. \tag{39}$$

Here we use the component notation [Barrett et al, 1994] in which $\boldsymbol{a}/\boldsymbol{b}$ denotes a vector whose nth component is a_n/b_n, where $\boldsymbol{a}$ and $\boldsymbol{b}$ are two vectors. Substituting these expressions into (12) and (13), we obtain the following results for the OSL-MAP-EM algorithm

$$\begin{aligned}\boldsymbol{A}^k = {} & \text{diag}\left[\frac{\bar{\hat{\boldsymbol{x}}}^k}{\boldsymbol{s} + \nabla_x U(\bar{\hat{\boldsymbol{x}}}^k)}\right]\boldsymbol{P}'\text{diag}\left[\boldsymbol{P}\bar{\hat{\boldsymbol{x}}}^k\right]^{-2}\text{diag}\left[\bar{\boldsymbol{y}}\right]\boldsymbol{P} \\ & + \text{diag}\left[\frac{\bar{\hat{\boldsymbol{x}}}^{k+1}}{\boldsymbol{s} + \nabla_x U(\bar{\hat{\boldsymbol{x}}}^k)}\right]\nabla_{xx}U(\bar{\hat{\boldsymbol{x}}}^k) - \text{diag}\left[\frac{\bar{\hat{\boldsymbol{x}}}^{k+1}}{\bar{\hat{\boldsymbol{x}}}^k}\right] + \boldsymbol{I},\end{aligned} \tag{40}$$

$$\boldsymbol{B}^k = \text{diag}\left[\frac{\bar{\hat{\boldsymbol{x}}}^k}{\boldsymbol{s} + \nabla_x U(\bar{\hat{\boldsymbol{x}}}^k)}\right]\boldsymbol{P}'\text{diag}\left[\boldsymbol{P}\bar{\hat{\boldsymbol{x}}}^k\right]^{-1}. \tag{41}$$

Equations (40) and (41) are independent of any particular form of prior. For a given prior, we just need to substitute the corresponding $\nabla_x U(\bar{\hat{\boldsymbol{x}}}^k)$ and $\nabla_{xx}U(\bar{\hat{\boldsymbol{x}}}^k)$ into (40) and (41).

To make a direct comparison with the results in [Wang and Gindi, 1997], we focus on the Gaussian prior, of which $U(\boldsymbol{x})$ can be written as

$$U(\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{x} - \boldsymbol{m})'\boldsymbol{K}^{-1}(\boldsymbol{x} - \boldsymbol{m}), \tag{42}$$

where $\boldsymbol{m}$ is the mean vector and $\boldsymbol{K}$ is the covariance matrix.

The corresponding expressions of $\boldsymbol{A}^k$ and $\boldsymbol{B}^k$ are

$$\begin{aligned}\boldsymbol{A}^k = {} & \text{diag}\left[\frac{\bar{\hat{\boldsymbol{x}}}^k}{\boldsymbol{s} + \boldsymbol{K}^{-1}(\bar{\hat{\boldsymbol{x}}}^k - \boldsymbol{m})}\right]\boldsymbol{P}'\text{diag}\left[\boldsymbol{P}\bar{\hat{\boldsymbol{x}}}^k\right]^{-2}\text{diag}\left[\bar{\boldsymbol{y}}\right]\boldsymbol{P} \\ & + \text{diag}\left[\frac{\bar{\hat{\boldsymbol{x}}}^{k+1}}{\boldsymbol{s} + \boldsymbol{K}^{-1}(\bar{\hat{\boldsymbol{x}}}^k - \boldsymbol{m})}\right]\boldsymbol{K}^{-1} - \text{diag}\left[\frac{\bar{\hat{\boldsymbol{x}}}^{k+1}}{\bar{\hat{\boldsymbol{x}}}^k}\right] + \boldsymbol{I}\end{aligned} \tag{43}$$

$$\boldsymbol{B}^k = \text{diag}\left[\frac{\bar{\hat{\boldsymbol{x}}}^k}{\boldsymbol{s} + \boldsymbol{K}^{-1}(\bar{\hat{\boldsymbol{x}}}^k - \boldsymbol{m})}\right] \boldsymbol{P}'\text{diag}\left[\boldsymbol{P}\bar{\hat{\boldsymbol{x}}}^k\right]^{-1}. \tag{44}$$

The above equations only depend on the first-order Taylor series approximation. Hence they are more accurate than those in [Wang and Gindi, 1997] where $\nabla_x L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^k) \approx 0$ is assumed. Because $\nabla_x L(\bar{\boldsymbol{y}}|\bar{\hat{\boldsymbol{x}}}^k) \approx 0$ is the fixed-point condition of the ML estimate, the approximation error can be quite large at early iterations or with strong priors. In comparison, (43) and (44) do not have such problem (see simulation results in Section 6).

Another problem with using the fixed-point condition of the ML estimate is that the result in [Wang and Gindi, 1997] is not consistent with the fixed-point analysis in [Fessler, 1996]. One has to assume the gradient of the prior $\boldsymbol{K}^{-1}(\bar{\hat{\boldsymbol{x}}}^k - \boldsymbol{m})$ in [Wang and Gindi, 1997] is zero in order to reach the correct fixed-point expression. This can easily be understood: if the gradient of the likelihood term is assumed to be zero, the gradient of the prior term will have to be zero as well to satisfy the fixed point condition of the MAP estimation. Using approximation (18) can avoid such inconsistency problems.

4.3. Ordered Subsets Algorithms

As we have pointed out in Section 2.4, the noise analysis in this paper is applicable to ordered subsets gradient methods, such as the OSEM [Hudson and Larkin, 1994], RBI-EM [Byrne, 1997, Byrne, 1998], RBI-MAP [Lalush and Tsui, 1998, Lalush et al, 2000], and RAMLA [Browne and De Pierro, 1996] algorithms. For example, the rescaled block-iterative (RBI) EM algorithm [Byrne, 1998, Soares et al, 2000] can be written as

$$\hat{\boldsymbol{x}}^{k+1} = \hat{\boldsymbol{x}}^k + \boldsymbol{C}^k(\hat{\boldsymbol{x}}^k)\nabla_x L^k(\boldsymbol{y}|\hat{\boldsymbol{x}}^k), \tag{45}$$

where $\boldsymbol{C}^k(\hat{\boldsymbol{x}}^k) = \text{diag}\left[\hat{\boldsymbol{x}}^k\right]\text{diag}\left[r_m\boldsymbol{s}\right]^{-1}$, $L^k(\boldsymbol{y}|\hat{\boldsymbol{x}}^k)$ is the likelihood function of the mth subset of the data, m is the index of the subset used in the kth iteration, and r_m is a constant defined in [Soares et al, 2000] [equation (18)].

Using equations (13) and (17), we get

$$\begin{aligned}\boldsymbol{A}^k &= \text{diag}\left[\bar{\hat{\boldsymbol{x}}}^k\right]\text{diag}\left[r_m\boldsymbol{s}\right]^{-1}\boldsymbol{P}'_m\text{diag}[\boldsymbol{P}\bar{\hat{\boldsymbol{x}}}^k]^{-2}\text{diag}\left[\bar{\boldsymbol{y}}\right]\boldsymbol{P} \\ &\quad - \text{diag}\left[r_m\boldsymbol{s}\right]^{-1}\text{diag}\left[\boldsymbol{P}'_m\text{diag}\left[\boldsymbol{P}\bar{\hat{\boldsymbol{x}}}^k\right]^{-1}\boldsymbol{y}\right] + r_m^{-1}\boldsymbol{I},\end{aligned} \tag{46}$$

$$\boldsymbol{B}^k = \text{diag}\left[\bar{\hat{\boldsymbol{x}}}^k\right]\text{diag}\left[\boldsymbol{s}^{-1}\right]\boldsymbol{P}'_m\text{diag}\left[\boldsymbol{P}\bar{\hat{\boldsymbol{x}}}^k\right]^{-1}, \tag{47}$$

which are the same as those in [Soares et al, 2000] [(45) and (46)] with changes of variables. $\boldsymbol{P}_m$ ($= \boldsymbol{H}_m$ in [Soares et al, 2000]) is the projection matrix for the mth subset.

The RBI-MAP algorithm in [Lalush et al, 2000] can be written as

$$\hat{\boldsymbol{x}}^{k+1} = \hat{\boldsymbol{x}}^k + \boldsymbol{C}^k(\hat{\boldsymbol{x}}^k)\left[\nabla_x L^k(\boldsymbol{y}|\hat{\boldsymbol{x}}^k) - \nabla_x U(\hat{\boldsymbol{x}}^k)\right],$$

where $\boldsymbol{C}^k(\hat{\boldsymbol{x}}^k) = t_m^{-1}\text{diag}\left[\hat{\boldsymbol{x}}^k\right]\text{diag}\left[\boldsymbol{s} + \nabla_x U(\hat{\boldsymbol{x}}^k)\right]^{-1}$ and t_m^{-1} is defined in [Lalush et al, 2000] [equation (8)].

Since $\boldsymbol{C}^k(\hat{\boldsymbol{x}}^k)$ cannot be written in the form of (16), we use (13) and (19). The final results are

$$\boldsymbol{A}^k = t_m^{-1}\mathrm{diag}\left[\frac{\bar{\hat{\boldsymbol{x}}}^k}{\boldsymbol{s}+\nabla_x U(\bar{\hat{\boldsymbol{x}}}^k)}\right]\left\{\boldsymbol{P}'_m\mathrm{diag}\left[\boldsymbol{P}\bar{\hat{\boldsymbol{x}}}^k\right]^{-1}\boldsymbol{P}+\nabla_{xx}U(\bar{\hat{\boldsymbol{x}}}^k)\right\} \tag{48}$$

$$\boldsymbol{B}^k = t_m^{-1}\mathrm{diag}\left[\frac{\bar{\hat{\boldsymbol{x}}}^k}{\boldsymbol{s}+\nabla_x U(\bar{\hat{\boldsymbol{x}}}^k)}\right]\boldsymbol{P}'_m\mathrm{diag}\left[\boldsymbol{P}\bar{\hat{\boldsymbol{x}}}^k\right]^{-1}. \tag{49}$$

Similar results can also be derived for the RAMLA [Browne and De Pierro, 1996] and BSREM algorithms [De Pierro and Yamagishi, 2001].

4.4. Algorithms with Inter-Update Linear Filtering

The noise analysis can also be extended to the algorithms with inter-update linear filtering [Silverman et al, 1990, Jacobson et al, 2000]. Inter-update filtering was introduced to reduce the noise in the ML-EM algorithm. With inter-update filtering, the update equation becomes

$$\hat{\boldsymbol{x}}^{k+1} = \boldsymbol{F}^k\hat{\boldsymbol{x}}^k + \boldsymbol{F}^k\boldsymbol{C}^k(\hat{\boldsymbol{x}}^k)\left[\nabla_x L^k(\boldsymbol{y}|\hat{\boldsymbol{x}}^k) - \nabla_x U(\hat{\boldsymbol{x}}^k)\right], \tag{50}$$

where $\boldsymbol{F}^k$ denotes the convolution matrix of the linear filter at the kth iteration.

Applying the same approach as shown in Section 2, we can obtain the following update equation for the noise vector

$$\epsilon^{k+1} \approx \boldsymbol{F}^k[\boldsymbol{I}-\boldsymbol{A}^k]\epsilon^k + \boldsymbol{F}^k\boldsymbol{B}^k\boldsymbol{n}, \tag{51}$$

where $\boldsymbol{A}^k$ and $\boldsymbol{B}^k$ are defined in (12) and (13), respectively.

5. Applications: Transmission Tomography

In transmission tomography the data can be modeled as independent Poisson random variables with the mean equal to

$$\bar{y}_i = b_i e^{-\sum_j P_{ij}x_j},$$

where b_i denotes the rate of emissions from the transmission source to the ith measurement and P_{ij} denotes the intersection length of the ith ray passing though the jth voxel.

For the Poisson likelihood function of transmission data, one can derive

$$\nabla_x L(\boldsymbol{y}|\boldsymbol{x}) = \boldsymbol{P}'(\boldsymbol{q}(\boldsymbol{x})-\boldsymbol{y}) \tag{52}$$

$$\nabla_{xx} L(\boldsymbol{y}|\boldsymbol{x}) = -\boldsymbol{P}'\mathrm{diag}\left[\boldsymbol{q}(\boldsymbol{x})\right]\boldsymbol{P}, \tag{53}$$

$$\nabla_{xy} L(\boldsymbol{y}|\boldsymbol{x}) = -\boldsymbol{P}', \tag{54}$$

where $\boldsymbol{q}(\boldsymbol{x})$ denotes a vector whose ith component is equal to $b_i e^{-\sum_j P_{ij}x_j}$.

A gradient-type algorithm for transmission image reconstruction can be written as [Lange et al, 1987]

$$\hat{\boldsymbol{x}}^{k+1} = \hat{\boldsymbol{x}}^k + \alpha^k\boldsymbol{C}^k(\hat{\boldsymbol{x}}^k,\boldsymbol{y})\nabla_x L(\boldsymbol{y}|\hat{\boldsymbol{x}}^k),$$

where $\boldsymbol{C}^k(\hat{\boldsymbol{x}}^k, \boldsymbol{y}) = \text{diag}\left[\boldsymbol{x}^k\right] \text{diag}\left[\boldsymbol{P}'\boldsymbol{y}\right]^{-1}$ and α^k is a step size.

Substituting the above results into (26) and (27), we get

$$\boldsymbol{A}^k = -\alpha^k \text{diag}\left[\bar{\hat{\boldsymbol{x}}}^k\right] \text{diag}\left[\boldsymbol{P}'\bar{\boldsymbol{y}}\right]^{-1} \boldsymbol{P}' \text{diag}\left[\boldsymbol{q}(\bar{\hat{\boldsymbol{x}}}^k)\right] \boldsymbol{P}, \tag{55}$$

$$\boldsymbol{B}^k = -\alpha^k \text{diag}\left[\bar{\hat{\boldsymbol{x}}}^k\right] \text{diag}\left[\boldsymbol{P}'\bar{\boldsymbol{y}}\right]^{-1} \boldsymbol{P}'. \tag{56}$$

Using similar methods, it is also possible to derive the noise propagation expressions for the SPS algorithms developed in [Erdogan and Fessler, 1999a, Erdogan and Fessler, 1999b].

6. Simulation Results

Using the OSL-MAP algorithms with and without line searches, we conducted computer-based Monte Carlo simulations to validate some theoretical results. A phantom similar to that in [Wang and Gindi, 1997] was used (Figure 1). The image has 32×32 pixels. The sinogram has 32 projection angles covering 180° 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. The OSL-MAP algorithms use a Gaussian prior with smoothing parameters of 0.1 and 1. The initial image is a constant uniform image.

Figure 1 shows the original phantom image, mean of 8,000 Monte Carlo reconstructions, and reconstruction of the noise-free data. The smoothing parameter is 0.1. As we expected, the noise-free reconstruction resembles the expectation of Monte Carlo reconstructions.

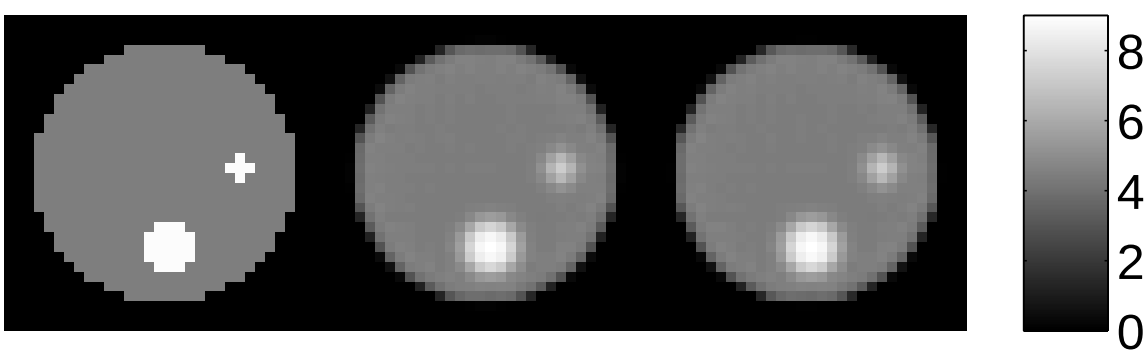

Figure 1. Phantom image (left), mean of 8,000 Monte Carlo reconstructions (middle), and noise-free reconstruction (right).

Figure 2 shows the comparison of the variance estimates that are computed using Monte Carlo reconstructions, our theoretical expression [(43) and (44)], and Wang and Gindi's result in [Wang and Gindi, 1997] for the OSL-MAP algorithm without a line search. Monte Carlo results were calculated using 8,000 independent reconstructions. The left column plots the variance as a function of iteration. The right column plots the relative root mean squared errors (RMS) of our results and Wang and Gindi's results,

when using the Monte Carlo results as the ground truth. The relative RMS is defined as

$$\mathrm{RMS} = \sqrt{\frac{1}{P}\sum_{i\in\mathrm{Disk}}\left[\frac{\mathrm{var}_i^{th}-\mathrm{var}_i^{MC}}{\mathrm{var}_i^{MC}}\right]^2}, \tag{57}$$

where P is the number of pixels in the disk region, var_i^{th} is the theoretical variance at pixel i, and var_i^{MC} is the Monte Carlo variance.

Figure 2 shows that our theoretical results match very well with the Monte Carlo results at all the iterations for both the smoothing parameters that were studied. In comparison, Wang and Gindi's results have relatively large RMS at initial iterations and with a large smoothing parameter ($\beta = 1$).

Figure 3 compares the theoretical predicted noise variance with Monte Carlo results for the OSL-MAP algorithm with line searches. The line search was restricted to satisfy the nonnegativity constraint in the image domain. We found that the step sizes obtained in a noise-free reconstruction are not good representations of those in noisy reconstructions because the nonnegativity constraint is much less active in a noise-free reconstruction since data are consistent. Therefore, step sizes and reconstructions obtained from a noisy data set are used in the theoretical expressions. The results are more relevant to the real situations where noise-free data are unavailable. The plots show that the maximum RMS is less than 18% at all iterations for both smoothing parameters. RMS drops below 10% after 5 iterations, indicating that the approximations (23) and (24) are reasonable.

In the above simulations the theoretical expressions are evaluated using the matrix form (14). Each iteration of (14) contains about 1024 forward projection and 2048 backprojection operations, where 1024 is the number of image pixels. Hence the computation cost for estimating the covariance matrix of the whole image can be quite high for three-dimensional images. Fortunately, in most quantitative applications we are only interested in a few regions of interest, which contain far fewer pixels. The computation cost for estimating the variance of one pixel and its correlations with all other pixels is about three times that of the reconstruction algorithm itself. Therefore, for a limited number of regions of interest, the computation time is manageable.

7. Conclusion

We have presented a unified approach to the theoretical analysis of noise propagation in iterative methods. The noise can be studied at each iteration as well as at the fixed-point. The results are applicable to a wide range of gradient-type algorithms, including the ordered subsets variants. We have shown that the iteration-based results are consistent with the fixed-point analysis if the algorithm converges. Examples in emission tomography and transmission tomography were shown. Our theoretical analysis is consistent with results of the existing literature. Using OSL-MAP algorithms as examples, we conducted Monte Carlo simulations showing that the proposed method can

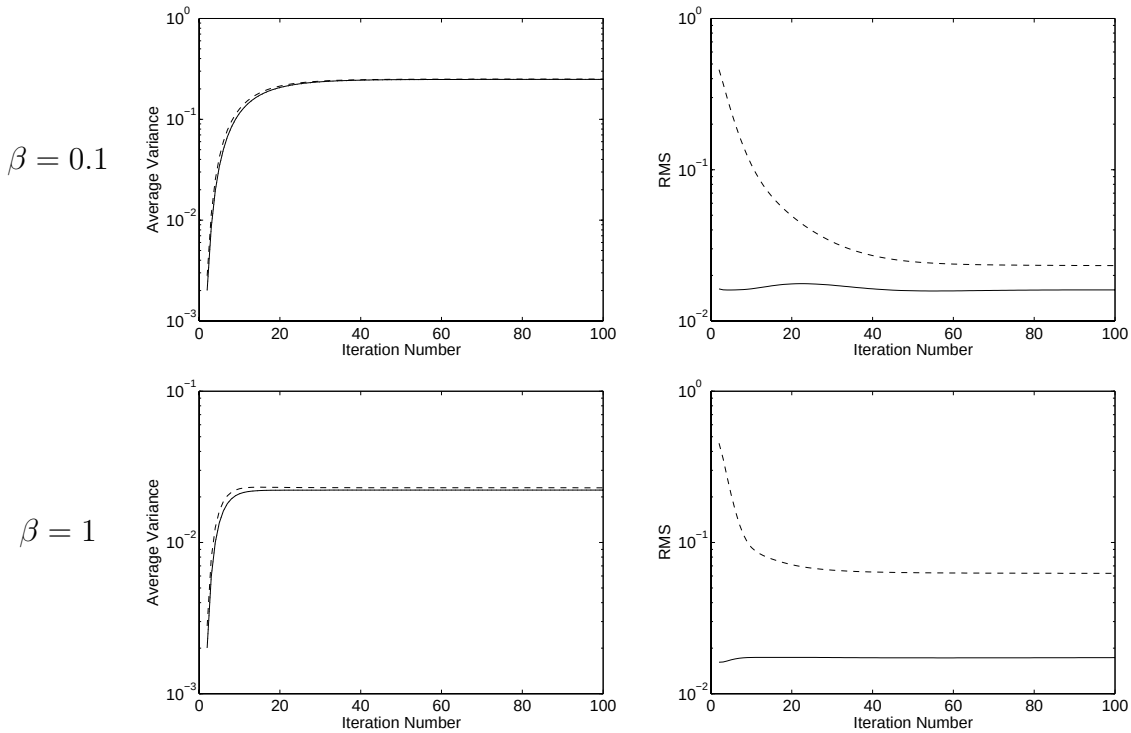

Figure 2. Plots of variances of OSL-MAP without line searches as function of iteration (left column) and the relative RMS of the theoretical predictions (right column). The legends are Monte Carlo results (dotted lines), our theoretical results (solid lines), and Wang and Gindi's results (dashed lines). In the left column the Monte Carlo results overlap with our theoretical results.

predict variance of the reconstruction with and without line searches and outperforms the results in the existing literature.

Acknowledgments

The author would like to thank Drs. Ronald Huesman and Grant Gullberg for helpful discussions and encouragement, Dr. Harrison Barrett for reading the manuscript, and Dr. D.N. Ghosh Roy for providing an interesting preprint, which in part motivated the writing of this paper. The author also thank Robert Smith for copyediting the manuscript.

This work is supported in part by the National Institute of Biomedical Imaging and Bioengineering under grant nos. R01 EB00363 and R01 EB00194, 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.

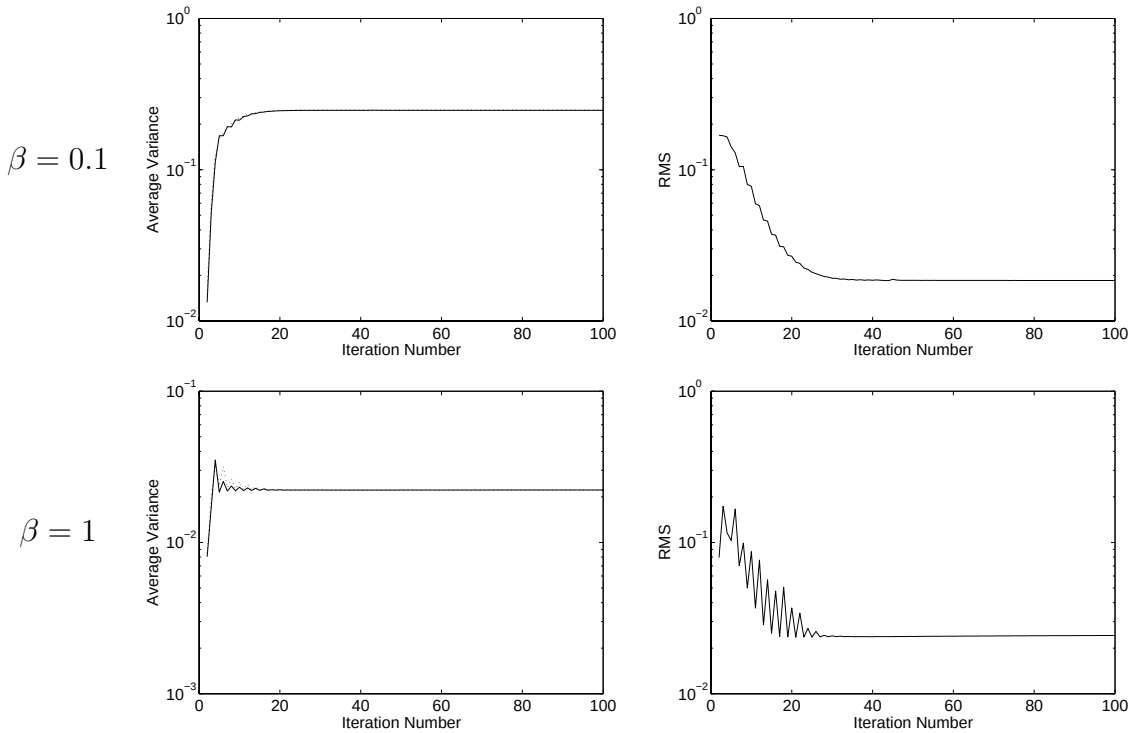

Figure 3. Plots of variances of OSL-MAP with line searches as function of iteration (left column) and the relative RMS of the theoretical predictions (right column). The legends are Monte Carlo results (dotted lines) and our theoretical results (solid lines).

References

[Abbey and Barrett, 1995] Abbey CK and Barrett HH. Linear iterative reconstruction algorithms: Study of observer performance. In *Proc. 14th International Conference on Information Processing in Medical Imaging*, pages 65–76. Kluwer Academic Publishers, 1995.

[Barrett et al, 1994] Barrett HH, Wilson DW, and Tsui BMW, (1994). Noise properties of the EM algorithm: I. theory. *Physics in Medicine and Biology*, 39:833–846.

[Browne and De Pierro, 1996] Browne J and De Pierro AR, (1996). A row-action alternative to the EM algorithm for maximizing likelihoods in emission tomography. *IEEE Transactions on Medical Imaging*, 15:687–699.

[Byrne, 1997] Byrne CL, (1997). Convergent block-iterative algorithms for image reconstruction from inconsistent data. *IEEE Transactions on Image Processing*, 6:1296–1304.

[Byrne, 1998] Byrne CL, (1998). Accelerating the EMML algorithm and related iterative algorithms by rescaled block-iterative (RBI) methods. *IEEE Transactions on Image Processing*, 7:100–109.

[Carson et al, 1993] Carson RE, Yan Y, Daube-Witherspoon ME, Freedman N, Bacharach SL, and Herscovitch P, (1993). An approximation formula for the variance of PET region-of-interest values. *IEEE Transactions on Medical Imaging*, 12(2):240–250.

[De Pierro and Yamagishi, 2001] De Pierro AR and Yamagishi MEB, (2001). Fast EM-like methods

for maximum a posteriori estimates in emission tomography. *IEEE Transactions on Medical Imaging*, 20:280–288.

[Erdogan and Fessler, 1999a] Erdogan H and Fessler JA, (1999). Monotonic algorithms for transmission tomography. *IEEE Transactions on Medical Imaging*, 18:801–814.

[Erdogan and Fessler, 1999b] Erdogan H and Fessler JA, (1999). Ordered subsets algorithms for transmission tomography. *Physics in Medicine and Biology*, 44:2835–2851.

[Fessler, 1996] Fessler J, (1996). Mean and variance of implicitely defined biased estimators (such as penalized maximum likelihood): Applications to tomography. *IEEE Transactions on Image Processing*, 5(3):493–506.

[Fessler and Rogers, 1996] Fessler JA and Rogers WL, (1996). Spatial resolution properties of penalized-likelihood image reconstruction: Spatial-invariant tomographs. *IEEE Transactions on Image Processing*, 9(5):1346–1358.

[Fessler and Yendiki, 2002] Fessler JA and Yendiki A. Channelized Hotelling observer performance for penalized-likelihood image reconstruction. In *Proceedings of IEEE Nuclear Science Symposium and Medical Imaging Conference*, page to appear, 2002.

[Green, 1990] Green P, (1990). Bayesian reconstructions from emission tomography data using a modified EM algorithm. *IEEE Transactions on Medical Imaging*, 9(1):84–93.

[Hsiao and Gindi, 2000] Hsiao IT and Gindi G. Noise propagation from attenuation correction into PET reconstructions. In *Proceedings of IEEE Nuclear Science Symposium and Medical Imaging Conference*, Lyon, France, 2000.

[Hudson and Larkin, 1994] Hudson HM and Larkin RS, (1994). Accelerated image reconstruction using ordered subsets of projection data. *IEEE Transactions on Medical Imaging*, 13(4):601–609.

[Huesman, 1984] Huesman R, (1984). A new fast algorithm for the evaluation of regions of interest and statistical uncertainty in computed tomography. *Physics in Medicine and Biology*, 29:543–552.

[Jacobson et al, 2000] Jacobson M, Levkovitz R, A Ben-Tal KT, Spinks T, Belluzzo D, Pagani E, Bettinardi V, Gilardi M, Zverovich A, and Mitra G, (2000). Enhanced 3D PET OSEM reconstruction using inter-update metz filtering. *Physics in Medicine and Biology*, 45:2417–2439.

[Lalush and Tsui, 1998] Lalush DS and Tsui BMW, (1998). Block-iterative techniques for fast 4D reconstruction using a priori motion models in gated cardiac SPECT. *Physics in Medicine and Biology*, 43:875–886.

[Lalush et al, 2000] Lalush D, Frey E, and Tsui B, (2000). Fast maximum entropy approximation in SPECT using the RBI-MAP algorithm. *IEEE Transactions on Medical Imaging*, 19(4):286–294.

[Lange, 1990] Lange K, (1990). Convergence of EM image reconstruction algorithms with Gibbs smoothing. *IEEE Transactions on Medical Imaging*, 9(4):439–446. Correction:10(2), June 1991, pp.228.

[Lange and Carson, 1984] Lange K and Carson R, (1984). EM reconstruction algorithms for emission and transmission tomography. *Journal of Computer Assisted Tomography*, 8(2):306–316.

[Lange et al, 1987] Lange K, Bahn M, and Little R, (1987). A theoretical study of some maximum likelihood algorithms for emission and transmission

tomography. *IEEE Transactions on Medical Imaging*, 6(2):106–114.

[Llacer et al, 1993] Llacer J, Veklerov E, Coakley K, Hoffman E, and Nunez J, (1993). Statistical analysis of maximum likelihood estimator images of human brain fdg studies. *IEEE Transactions on Medical Imaging*, 12(2):215–231.

[Luenberger, 1984] Luenberger D, (1984). *Linear and Nonlinear Programming.* Addison-Wesley Publishing Company, second edition.

[Maitra and O'Sullivan, 1998] Maitra R and O'Sullivan F, (1998). Variability assessment in PET and related generalized deconvolution methods. *Journal of American Statistical Association*, 93(444):1340–1355.

[Mazoyer et al, 1986] Mazoyer BM, Huesman RH, Budinger TF, and Knittel BL, (1986). Dynamic PET data analysis. *Journal of Computer Assisted Tomography*, 10(4):645–653.

[Qi and Huesman, 2001] Qi J and Huesman RH, (2001). Theoretical study of lesion detectability of MAP reconstruction using computer observers. *IEEE Transactions on Medical Imaging*, 20:815–822.

[Qi and Leahy, 1999] Qi J and Leahy RM, (1999). A theoretical study of the contrast recovery and variance of MAP reconstructions from PET data. *IEEE Transactions on Medical Imaging*, 18(4):293–305.

[Qi and Leahy, 2000] Qi J and Leahy RM, (2000). Resolution and noise properties of MAP reconstruction for fully 3D PET. *IEEE Transactions on Medical Imaging*, 19(5):493–506.

[Shepp and Vardi, 1982] Shepp L and Vardi Y, (1982). Maximum likelihood reconstruction for emission tomography. *IEEE Transactions on Medical Imaging*, 1(2):113–122.

[Silverman et al, 1990] Silverman B, Jones M, Wilson J, and Nychka D, (1990). A smoothed EM approach to indirect estimation problems, with particular reference to stereology and emission tomography. *Journal of Royal Statistical Society, Ser. B*, 52(2):271–324.

[Soares et al, 2000] Soares EJ, Byrne C, and Glick S, (2000). Noise characterization of block-iterative reconstruction algorithms: 1. theory. *IEEE Transactions on Medical Imaging*, 19(4):261–270.

[Stayman and Fessler, 2000] Stayman JW and Fessler JA, (2000). Regularization for uniform spatial resolution properties in penalized-likelihood image reconstruction. *IEEE Transactions on Medical Imaging*, 19:601–615.

[Wang and Gindi, 1997] Wang W and Gindi G, (1997). Noise analysis of MAP-EM algorithms for emission tomography. *Physics in Medicine and Biology*, 42:2215–2232.

[Wilson et al, 1994] Wilson DW, Tsui BMW, and Barrett HH, (1994). Noise properties of the EM algorithm: II. Monte Carlo simulations. *Physics in Medicine and Biology*, 39:847–872.

[Xing and Gindi, 2002] Xing Y and Gindi G. Rapid calculation of detectability in Bayesian SPECT. In *Proceedings of IEEE International Symposimum on Biomedical Imaging*, page CDROM, 2002.