Accelerated coordinate descent methods for Bayesian reconstruction using ordered subsets of projection data

Lee, Soo-Jin

Soo-Jin Lee, "Accelerated coordinate descent methods for Bayesian reconstruction using ordered subsets of projection data," Proc. SPIE 4121, Mathematical Modeling, Estimation, and Imaging, (4 October 2000); doi: 10.1117/12.402437

SPIE.

Event: International Symposium on Optical Science and Technology, 2000, San Diego, CA, United States

Accelerated Coordinate Descent Methods for Bayesian Reconstruction Using Ordered Subsets of Projection Data

Soo-Jin Lee

Department of Electronic Engineering
Paichai University
439-6 Doma 2-Dong, Seo-Ku
Taejon, 302-735, Korea

ABSTRACT

The ordered subsets (OS) algorithm[1] has enjoyed considerable interest for accelerating the well-known EM reconstruction algorithm for emission tomography and has recently found widespread use in clinical practice. This is primarily due to the fact that, while retaining the advantages of EM, the OS-EM algorithm can be easily implemented by slightly modifying the existing EM algorithm. The OS algorithm has also been applied[1] with the one-step-late (OSL) algorithm,[2] which provides maximum *a posteriori* estimation based on Gibbs priors. Unfortunately, however, the OSL approach is known to be unstable when the smoothing parameter that weights the prior relative to the likelihood is relatively large. In this work, we note that the OS principle can be applied to any algorithm that involves calculation of a sum over projection indices, and show that it can also be applied to a generalized EM algorithm with useful quadratic priors. In this case, the algorithm is given in the form of iterated conditional modes (ICM), which is essentially a coordinate-wise descent method, and provides a number of important advantages. We also show that, by scaling the smoothing parameter in a principled way, the degree of smoothness in reconstructed images, which appears to vary depending on the number of subsets, can be efficiently matched for different numbers of subsets. Our experimental results indicate that the OS-ICM algorithm along with the method of scaling the smoothing parameter provides robust results as well as a substantial acceleration.

Keywords: emission tomography, statistical reconstruction, ordered subsets, Gibbs priors, hyperparameters

1. INTRODUCTION

Statistical image reconstruction methods, such as maximum likelihood (ML) and maximum *a posteriori* (MAP) approaches, have played an important role in emission tomography since they accurately model the Poisson noise associated with gamma-ray projection data. A significant advance for these methods was made with the introduction of the ML approach using the famous expectation maximization (EM) algorithm. Since the ML-EM algorithm provides a number of potential benefits, its application has led to the introduction of many related reconstruction algorithms and its value is now well established. Although the application of EM in clinical practice in early days was of a difficult problem due to high per-iteration costs and large numbers of iterations, with the dramatically improved speed of computers, it is now inevitably becoming more practical and is indeed commonly available from nuclear medicine equipment suppliers. However, it is important to point out that, while the main limitation in applying EM due to processing time is becoming less of a problem, there is also continuing need for more complex models to characterize the actual physics involved more accurately, which requires more computational load.

Recently, there has been increasing emphasis on accelerating iterative reconstruction methods which can produce an image with much less iterations than the ML-EM algorithm. Some accelerated methods demonstrate a possibility of permitting reconstruction in a clinically acceptable time with a few iterations. For example, the OS-EM algorithm proposed by Hudson and Larkin[1] has been successful in accelerating the existing ML-EM algorithm and is continually increasing in its popularity. This is presumably due to the fact that, while retaining the advantages of ML-EM such as accurate modeling of any type of system, it provides an order-of-magnitude acceleration over ML-EM. In fact a number of other accelerated algorithms for finding ML or MAP estimation for emission tomography have also been introduced in the literature[3–6]; some of these are based on direct maximization[6] of the objective function rather

Further author information:
Email: sjlee@mail.paichai.ac.kr; Telephone: +82-42-520-5711; Fax: +82-42-520-5663

Mathematical Modeling, Estimation, and Imaging, David C. Wilson, Hemant D. Tagare, Fred L. Bookstein, Françoise J. Prêteux, Edward R. Dougherty, Editors, Proceedings of SPIE Vol. 4121 (2000)

than on the EM algorithm to avoid the slow convergence problem in the traditional EM algorithm. Nevertheless, the OS-EM algorithm has been more extensively used than others and have recently found widespread popularity in clinical practice presumably due to the fact that it is easily implemented by slightly modifying the existing EM algorithm. In this paper, we are concerned exclusively with the OS approach.

In order to accelerate the EM algorithm, which uses the estimation of all projections and calculates the ratios between estimated and measured values for all projections in the backprojection process, the OS-EM algorithm first subdivides projection data into several subsets (or blocks) and then progressively processes each subset of projections by calculating projection and backprojection in each iteration. Therefore, the standard EM algorithm corresponds to a special case of OS-EM when a single subset includes all projection data. As the number of subsets increases, the OS procedure accelerates convergence by a factor proportional to the number of subsets. Although the OS-EM algorithm provides good-quality reconstructions after only a few iterations, beyond a critical number of iterations, the noise artifact of ML solutions is magnified due to the inherent instability problem of ML-EM. One can alleviate the instability of OS-EM by applying the OS principle to regularized EM in the context of a Bayesian MAP framework, which allows the incorporation of suitable prior models to regularize the ill-posed nature of the tomographic inversion problem. A significant body of work for Bayesian tomographic reconstruction has been developed over the last decade. Early methods focused on the regularization of the unstable ML-EM algorithm. Later, the Bayesian approach was also viewed as a means of incorporating actual, known information regarding the local spatial character of the source. A host of different formulations to model the Bayesian priors have been proposed in the literature[2,7–13]; some of these implicitly model the underlying radionuclide density as globally smooth, and others extend the smoothness model by allowing for spatial discontinuities. While the former is associated with a smoothing penalty that is a quadratic function of nearby pixel, the latter uses a *non*quadratic penalty function.

Hudson and Larkin[1] applied the ordered subsets principle to the Green's one-step-late (OSL) algorithm[2] using a nonquadratic log-cosh penalty function and demonstrated that the instability of OS-EM could be reduced. Unfortunately, however, this method seems to be less than ideal for clinical practice due to the following reasons: (i) Although the OSL approach provides an easy way of implementing complex nonquadratic priors as well as simple quadratic priors, it is known to be unstable when the smoothing parameter that weights the prior relative to the likelihood is relatively large. (ii) The use of nonquadratic priors requires estimation of additional "hyperparameters" to adjust penalty functions for edge preservation. With regards to point (i), a number of other optimization algorithms exist, which provide a stable solution for nonquadratic priors as well as quadratic priors. However, in spite of good performance under properly chosen optimization algorithms and hyperparameters, any enthusiasm for using nonquadratic priors is usually tempered by the fact that most methods for estimating the hyperparameters involved in such priors are characterized by severe computational cost, which is prohibitive in clinical practice.

We note here that, by relaxing the requirement of preserving sharp edges and using instead useful quadratic priors which offer good performance even in the presence of spatial discontinuities, the above two points may be improved and OS algorithms may become more practical. In fact, the use of quadratic priors provides remarkable advantages; solutions are easier to compute and hyperparameter calculations become less of a problem.

In this work, we focus on the development of accelerated MAP-EM algorithms which may be useful in clinical practice. The remainder of this paper formulates ordered subsets MAP-EM algorithms, develops a method of calculating and scaling the smoothing parameter, and presents our experimental results.

2. FORMULATION OF MAP-EM

The MAP approach in the context of a Bayesian framework is to estimate the underlying source field $\mathbf{f}$ by maximizing the posterior probability, given as

$$\Pr(\mathbf{F}=\mathbf{f} \mid \mathbf{G}=\mathbf{g}) = \frac{\Pr(\mathbf{G}=\mathbf{g}|\mathbf{F}=\mathbf{f})\Pr(\mathbf{F}=\mathbf{f})}{\Pr(\mathbf{G}=\mathbf{g})}, \tag{1}$$

where $\mathbf{f}$ and $\mathbf{g}$ are the 2-D vector fields for the source intensities and projection data, respectively, and $\mathbf{F}$ and $\mathbf{G}$ are the associated random fields. In (1) $\Pr(\mathbf{G}=\mathbf{g}|\mathbf{F}=\mathbf{f})$ is the likelihood which is Poisson distributed, and $\Pr(\mathbf{F}=\mathbf{f})$ is the pdf of the prior.

In Bayesian approaches the prior is usually modeled as a Gibbs distribution since it provides mathematically powerful machinery to model a class of priors that specify local spatial correlations of the underlying source. The

Gibbs distribution is given by

$$\Pr(\mathbf{F}=\mathbf{f}) = \frac{1}{Z}\exp[-\lambda E_P(\mathbf{f})],$$

where $E_P(\mathbf{f})$ is the sum of energies of individual "cliques" which are defined by adjacent pixels in a local neighborhood of a pixel,[14,15] λ the positive hyperparameter that weights the prior relative to the likelihood term, Z a normalization of no concern here. To impose smoothness constraints on the underlying object, it is common to choose energy functions which penalize pixel configurations with large intensity gradient within each clique. In this work, we use the energy function derived from 2-D splines[16,17]:

$$E_P(\mathbf{f}) = (1-\tau)\sum_{ij}\left[f_h^2(i,j)+f_v^2(i,j)\right] + \tau\sum_{ij}\left[f_{hh}^2(i,j)+2f_{hv}^2(i,j)+f_{vv}^2(i,j)\right], \tag{2}$$

where $f_v(i,j)$ and $f_h(i,j)$ are the discrete first partial derivatives of the source distribution in the vertical and horizontal directions, respectively, $f_{vv}(i,j)$ and $f_{hh}(i,j)$ the discrete second partial derivatives in the vertical and horizontal directions, respectively, and $f_{hv}(i,j)$ the second partial cross derivative. Our choices for discretization of the derivatives in this work are:

$$\begin{aligned}
f_h(i,j) &\stackrel{\text{def}}{=} f_{i,j+1} - f_{i,j}\\
f_v(i,j) &\stackrel{\text{def}}{=} f_{i+1,j} - f_{i,j}\\
f_{hh}(i,j) &\stackrel{\text{def}}{=} f_{i,j+1} - 2f_{i,j} + f_{i,j-1}\\
f_{vv}(i,j) &\stackrel{\text{def}}{=} f_{i+1,j} - 2f_{i,j} + f_{i-1,j}\\
f_{hv}(i,j) &\stackrel{\text{def}}{=} f_{i+1,j+1} - f_{i+1,j} - f_{i,j+1} + f_{i,j}.
\end{aligned}$$

In (2) $\tau \in [0,1]$ is the weight parameter; $\tau = 0$ leads to interpolants having zeroth-order continuity and $\tau = 1$ leads to interpolants of first-order continuity. According the physical interpretations[18,19] in elasticity theory, the two extreme cases ($\tau = 0$ and $\tau = 1$) correspond to the deflection energy of a "membrane" and the deflection bending energy of a "thin plate", respectively. For the tomographic reconstruction problem, it has been shown[20] that a linear combination of the membrane and the thin-plate priors using $\tau = 0.5$ compromises the two extreme models well by reducing errors due to oversmoothing and overshooting inherent in the first- and second-order models, respectively.

Given the posterior distribution in (1), maximizing the posterior distribution is equivalent to maximizing log of the posterior probability and the MAP estimation reduces to

$$\hat{\mathbf{f}} = \arg\max_{\mathbf{f}}[\log\Pr(\mathbf{G}=\mathbf{g}|\mathbf{F}=\mathbf{f}) + \log\Pr(\mathbf{F}=\mathbf{f})]. \tag{3}$$

One possible way of solving this problem may be to use Green's one-step-late (OSL) algorithm, which is not derivable directly from (3) but is obtained by simply modifying the EM algorithm as follows:

$$\hat{f}_{ij}^{k+1} = \frac{\hat{f}_{ij}^k}{\sum_{t\theta}\mathcal{H}_{t\theta,ij} + \lambda\left.\frac{\partial E_P(\mathbf{f})}{\partial f_{ij}}\right|_{f_{ij}=\hat{f}_{ij}^k}} \sum_{t\theta}\frac{\mathcal{H}_{t\theta,ij}\,g_{t\theta}}{\sum_{mn}\mathcal{H}_{t\theta,mn}\hat{f}_{mn}^k}, \tag{4}$$

where $\hat{f}_{ij}^k$ is the object estimate at location (i,j) and iteration k, $g_{t\theta}$ the number of detected counts, $\bar{g}_{t\theta}$ the expected number of counts for a particular source $\mathbf{f}$ and a forward projection matrix $\mathcal{H}_{t\theta,ij}$. As noted by Green,[2] if λ is not too large, the OSL algorithm converges to the MAP solution. However, it is also known that, since the partial derivative term in the denominator might be negative, the pixel positivity is no longer guaranteed and the total number of counts is not strictly conserved. Therefore, the OSL algorithm often reveals unstable results even with a moderate value of the smoothing parameter.

In this work, we utilize the incomplete/complete data formulation of the generalized EM (GEM) approach.[9] In this case the maximization step of the MAP-EM algorithm results in the minimization of the following objective function:

$$M(\mathbf{f}|\hat{\mathbf{f}}) \stackrel{\text{def}}{=} -Q(\mathbf{f}|\hat{\mathbf{f}}) + \lambda E_P(\mathbf{f}) \tag{5}$$

where

$$Q(\mathbf{f} \mid \hat{\mathbf{f}}) = \sum_{t\theta} \sum_{ij} \left[g_{t\theta} \frac{\mathcal{H}_{t\theta,ij} \hat{f}_{ij}}{\sum_{mn} \mathcal{H}_{t\theta,mn} \hat{f}_{mn}} \log(f_{ij}) - \mathcal{H}_{t\theta,ij} f_{ij} \right]. \tag{6}$$

Since our prior model is concave, to find the solution that minimizes the above objective function, we can simply set $\frac{\partial M(\mathbf{f}|\hat{\mathbf{f}})}{\partial f_{ij}} = 0$ and solve for f_{ij}. In this case, the update procedure basically follows the method of iterated conditional modes (ICM), which is essentially a coordinate-wise descent method, and does not require step sizes necessary for gradient descent methods. ICM minimizes the energy function by a coordinate-wise descent procedure performed by minimizing the energy function with respect to f_{ij} while keeping all other source intensities fixed. Then a new location is chosen and the same method is repeated. After a full sweep of the lattice, the procedure is repeated until convergence criteria are met. The ICM update equation for our prior model is given by

$$\hat{f}_{ij}^{k+1} = \frac{-(X_1 - 2\lambda X_3) + \left[(X_1 - 2\lambda X_3)^2 + 8\lambda(4 + 16\tau) X_2 \right]^{\frac{1}{2}}}{4\lambda(4 + 16\tau)}, \tag{7}$$

where

$$\begin{aligned}
X_1 &\stackrel{\text{def}}{=} \sum_{t\theta} \mathcal{H}_{t\theta,ij}, \\
X_2 &\stackrel{\text{def}}{=} \hat{f}_{ij}^k \sum_{t\theta} \frac{\mathcal{H}_{t\theta,ij} g_{t\theta}}{\sum_{mn} \mathcal{H}_{t\theta,mn} \hat{f}_{mn}^k}, \\
X_3 &\stackrel{\text{def}}{=} (1-\tau) \left(\hat{f}_{i,j+1}^k + \hat{f}_{i+1,j}^k + \hat{f}_{i,j-1}^{k+1} + \hat{f}_{i-1,j}^{k+1} \right) + \tau \left\{ 8 \left(\hat{f}_{i-1,j}^{k+1} + \hat{f}_{i,j-1}^{k+1} + \hat{f}_{i,j+1}^k + \hat{f}_{i+1,j}^k \right) \right. \\
&\quad \left. -2 \left(\hat{f}_{i-1,j-1}^{k+1} + \hat{f}_{i-1,j+1}^{k+1} + \hat{f}_{i+1,j-1}^k + \hat{f}_{i+1,j+1}^k \right) - \left(\hat{f}_{i-2,j}^{k+1} + \hat{f}_{i,j-2}^{k+1} + \hat{f}_{i,j+2}^k + \hat{f}_{i+2,j}^k \right) \right\}
\end{aligned}$$

Note that an ICM uses a raster-scan update in which each pixel is replaced as soon as it is updated. The superscripting in the expressions for X_2 and X_3 reflects this.

3. CONSTRUCTION OF ORDERED SUBSETS ICM ALGORITHMS

3.1. Ordered Subsets

To accelerate the convergence of our algorithms, we apply the OS idea,[1] which was used for accelerating the standard EM algorithm, to the ICM update equation in (7). Notice that X_2/X_1 corresponds the following well-known EM update equation:

$$\hat{f}_{ij}^{k+1} = \frac{\hat{f}_{ij}^k}{\sum_{t\theta} \mathcal{H}_{t\theta,ij}} \sum_{t\theta} \frac{\mathcal{H}_{t\theta,ij} g_{t\theta}}{\sum_{mn} \mathcal{H}_{t\theta,mn} \hat{f}_{mn}^k},$$

where $g_{t\theta}$ and $\sum_{mn} \mathcal{H}_{t\theta,mn} \hat{f}_{mn}^k$ are the observed projection measurements and the mean of estimated counts, respectively. By defining the ratio between the observed and measured projections as $\hat{r}_{t\theta}^k \stackrel{\text{def}}{=} g_{t\theta} / \sum_{mn} \mathcal{H}_{t\theta,mn} \hat{f}_{mn}^k$, we may rewrite the update equation as

$$\hat{f}_{ij}^{k+1} = \frac{\hat{f}_{ij}^k}{\sum_{t\theta} \mathcal{H}_{t\theta,ij}} \sum_{t\theta} \mathcal{H}_{t\theta,ij} \hat{r}_{t\theta}^k.$$

While the standard EM algorithm described above uses the estimation of all projections ($\sum_{mn} \mathcal{H}_{t\theta,mn} \hat{f}_{mn}^k$) and calculates the ratios ($\hat{r}_{t\theta}^k$) between estimated and measured values for all projections in the backprojection process ($\sum_{t\theta} \mathcal{H}_{t\theta,ij} \hat{r}_{t\theta}^k$), the OS-EM algorithm groups projection data into an ordered sequence of subsets or blocks and progressively processes each subset of projections by calculating projection and backprojection in each iteration. In this case the OS-EM algorithm provides an order-of-magnitude acceleration over EM. In fact the ordered subsets principle may be applied to any algorithm that involves calculation of a sum over projection indices; an OS version of the algorithm can be obtained by replacing the sums over all the projection indices with sums over a subset of the data. Subsets are usually chosen in a balanced way so that pixel activity contributes equally to any subset. It was also reported[1] that it is best to order the subsets so that two adjacent indices for projection angles (θ in our case) in a given subset corresponds to actual angles with maximum angular distance.

Let $S_1, S_2, \cdots, S_N$ denote the subsets in the order selected. Starting with some initial estimate $\hat{f} > 0$, the outline for ordered subsets applied to our ICM algorithm (referred to hereafter as OS-ICM) is as follows:

for each iteration $k = 1, \cdots, K$

for each subset $n = 1, \cdots, N$

project: $\bar{g}_{t\theta} := \sum_{ij} \mathcal{H}_{t\theta,ij} \hat{f}_{ij}, \; \forall (t,\theta) \in S_n$

$\hat{r}_{t\theta} := g_{t\theta} / \bar{g}_{t\theta}, \; \forall (t,\theta) \in S_n$

backproject: $X_1 := \sum_{t\theta \in S_n} \mathcal{H}_{t\theta,ij}, \quad X_2 := \sum_{t\theta \in S_n} \mathcal{H}_{t\theta,ij} \hat{r}_{t\theta}$

$$X_3 := (1-\tau)\left(\hat{f}_{i,j+1} + \hat{f}_{i+1,j} + \hat{f}_{i,j-1} + \hat{f}_{i-1,j}\right) + \tau\left\{8\left(\hat{f}_{i-1,j} + \hat{f}_{i,j-1} + \hat{f}_{i,j+1} + \hat{f}_{i+1,j}\right)\right.$$
$$\left. -2\left(\hat{f}_{i-1,j-1} + \hat{f}_{i-1,j+1} + \hat{f}_{i+1,j-1} + \hat{f}_{i+1,j+1}\right) - \left(\hat{f}_{i-2,j} + \hat{f}_{i,j-2} + \hat{f}_{i,j+2} + \hat{f}_{i+2,j}\right)\right\}$$

$$\hat{f}_{ij} := \frac{-(X_1 - 2\lambda X_3) + \left[(X_1 - 2\lambda X_3)^2 + 8\lambda(4+16\tau)X_2\right]^{\frac{1}{2}}}{4\lambda(4+16\tau)}$$

end

end

Table 1. Outline for the OS-ICM algorithm.

3.2. Calculating The Smoothing Parameter

The variety of approaches used to attack the problem of hyperparameter estimation in emission tomography include regularization methods and estimation-theoretic methods based on maximum likelihood (see our references[21,22] for a summary of work in this area). All of these methods are characterized by severe computational cost, yet this is the only principled way to attack the problem as stated. In this work, we use a rather different approach in which (noiseless) training exemplars[23] are used in place of noisy projection data for ML hyperparameter estimation. Training exemplars have been used in a variety of problems for MRF (Markov random field) parameter estimation.[24,25] Of course, in any application, adequate training samples must be available. In emission tomography, for example, one brain scan acquired under a very specific medical protocol looks quite similar to others acquired under the same protocol. Therefore, researchers utilize this fact to construct elaborate hardware and software "phantoms" that mimic, to varying degrees, the typical distribution of radionuclide within the relevant patient anatomy. Image formation from the phantom or set of phantoms is then simulated or physically acquired, and comparison of reconstructed image quality to the known phantom(s) then serves as a test of the reconstruction. (We have explored the use of autoradiography[26] as a source of realistic phantoms[13,23,20] and we include one of such phantoms along with a synthetic phantom, as shown in Fig. 2, in this paper. More details on the use of autoradiograph based phantoms for emission tomography can be found in our references.[27,28]) Hyperparameters calculated for a phantom may be easily related to an intensity- scaled or magnified version of the phantom, and this enables one to quickly scale hyperparameter values with regard to count level and patient size, two parameters of importance in phantom studies.

The basic notion is simple: Given a known, noiseless training object denoted by a vector $\mathbf{f}$, hyperparameters $\boldsymbol{\lambda}$, and a prior probability model $\Pr(\mathbf{f}|\boldsymbol{\lambda})$, compute the ML estimate $\hat{\boldsymbol{\lambda}} = \arg\max_{\boldsymbol{\lambda}} \Pr(\mathbf{f}|\boldsymbol{\lambda})$, and use $\hat{\boldsymbol{\lambda}}$ in subsequent reconstructions. (Note that here, $\mathbf{f}$ is a single realization of a random field modeled by the prior.) The ML problem tends to be robust since there are only one or two hyperparameters while the vector $\mathbf{f}$ may have thousands of components. In particular, for the quadratic penalties considered in this work, there exists a direct analytical solution[29] to the ML estimate of the hyperparameter (or the smoothing parameter) λ, which is given by

$$\hat{\lambda} = \frac{L}{2E_P(\mathbf{f})}, \tag{8}$$

where L is the number of object pixels and $E_P(\mathbf{f})$ is the prior energy function defined in (2). Note that L should be set equal to the number of pixels of the object proper, not the zero-valued background pixels, lest the smoothness of the background bias the hyperparameter estimates.

For OS algorithms, the hyperparameter estimated via (8) may not be directly applicable since the projections and backprojections, which are associated with pixel updates, are performed for only the elements of a single block. According to our experiments (see below), when the smoothing parameter is fixed, the overall degree of smoothness in reconstructed images varies depending on the number of subsets used in the OS algorithm. This indicates that,

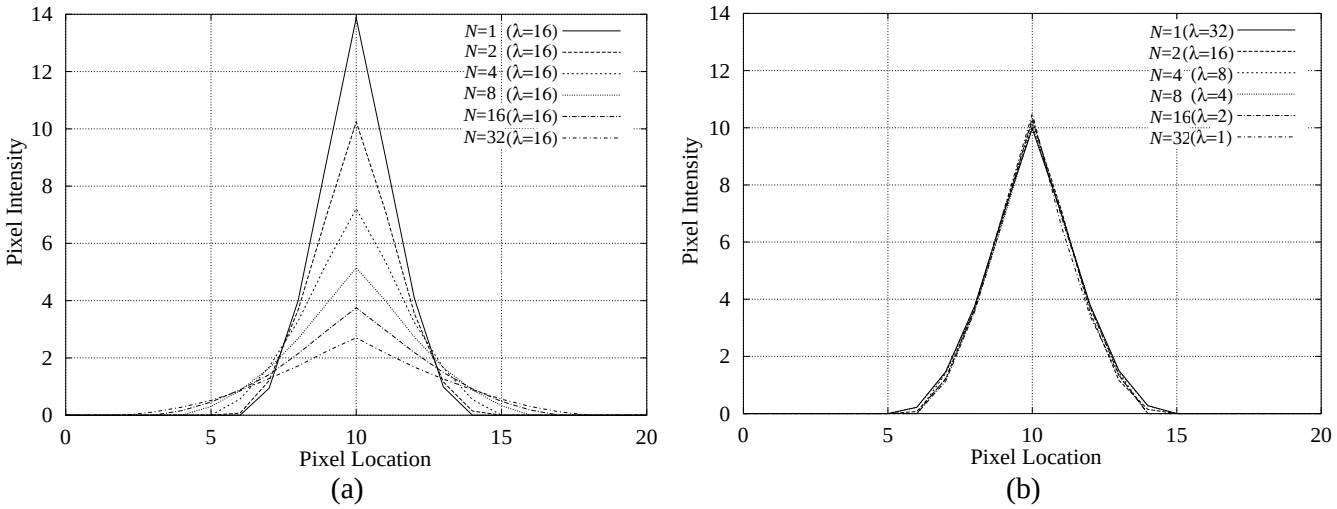

Figure 1. Profile plots for point-source reconstructions from noiseless projections: (a) Reconstructions without scaling λ for different subset numbers; (b) Reconstructions with scaled λ to match the reconstruction for $N = 2$ and $\lambda = 16$.

to achieve approximately the same degree of smoothness for different numbers of subsets, the smoothing parameter should be scaled properly. The relationship between the number of subsets and the scale factor for the smoothing parameter λ may be intuitively explained by observing the OSL algorithm described in (4). Note that the derivative energy function $\frac{\partial E_P(\mathbf{f})}{\partial f_{ij}}$ in (4) acts as a penalty term to enforce the conditions required by the prior; the larger the magnitude of $\frac{\partial E_P(\mathbf{f})}{\partial f_{ij}}$, the smaller the estimate of $\hat{f}_{ij}$ becomes, thereby suppressing pixel intensity differences. Therefore, the magnitude of $\frac{\partial E_P(\mathbf{f})}{\partial f_{ij}}$ specifies the "strength" of smoothing. To find the strength of smoothing for a given value of λ, we may rewrite (4) as

$$\left.\frac{\partial E_P(\mathbf{f})}{\partial f_{ij}}\right|_{f_{ij}=\hat{f}_{ij}^k} = \frac{1}{\lambda}\frac{\hat{f}_{ij}^k}{\hat{f}_{ij}^{k+1}}\sum_{t\theta} A_{t\theta,ij}, \tag{9}$$

where

$$A_{t\theta,ij} \stackrel{\text{def}}{=} \frac{\mathcal{H}_{t\theta,ij}g_{t\theta}}{\sum_{mn}\mathcal{H}_{t\theta,mn}\hat{f}_{mn}^k} - \mathcal{H}_{t\theta,ij}.$$

For the OS version, the summation term in (9), $\sum_{t\theta} A_{t\theta,ij}$, is replaced by $\sum_{t\theta\in S_n} A_{t\theta,ij}$, where S_n $(n = 1, \cdots, N)$ is the nth subset. In this case, since $\hat{f}_{ij}^{k+1} \approx \hat{f}_{ij}^k$ for convergence, the strength of smoothing can be expressed as

$$\left.\frac{\partial E_P(\mathbf{f})}{\partial f_{ij}}\right|_{f_{ij}=\hat{f}_{ij}^k} \approx \frac{1}{\lambda}\sum_{t\theta\in S_n} A_{t\theta,ij}. \tag{10}$$

Note that, when all the N subsets are perfectly balanced, $\sum_{t\theta\in S_n} A_{t\theta,ij} = \frac{1}{N}\sum_{t\theta} A_{t\theta,ij}$. Therefore, in order for the image reconstructed by the OS algorithm with $N > 1$ subsets to have approximately the same degree of smoothness as the case for $N = 1$, the smoothing parameter λ calculated for $N = 1$ should be scaled as follows:

$$\lambda_N = \frac{\lambda}{N}, \tag{11}$$

where λ_N is the scaled smoothing parameter for $N > 1$ subsets. In general, the relationship between the two scaled smoothing parameters, λ_N and λ_M for N and M subsets, respectively, is given by

$$\lambda_N = \frac{M}{N}\lambda_M. \tag{12}$$

The above relationship also holds for the ICM algorithm in Table 1.

Figure 1 shows the profile plots for point-source reconstructions from noiseless projections using the OS-ICM algorithm with $\tau = 0$, which qualitatively validates the relationship in (12). Figure 1(a) indicates that, for fixed λ, increasing the number of subsets N also increases the degree of smoothness in the reconstruction. Figure 1(b) shows as an example that, in order to retain the same degree of smoothness in the reconstruction as the case for $N = 2$ and $\lambda = 16$ in (a), the smoothing parameter for each subset number should be scaled as follows: $\lambda_1 = 32$, $\lambda_2 = 16$, $\lambda_4 = 8$, $\lambda_8 = 4$, $\lambda_{16} = 2$, and $\lambda_{32} = 1$, where the subscript in each λ indicates the number of subsets.

4. EXPERIMENTAL RESULTS

We performed 2-D simulation studies with projection data from 128×128 digital phantoms, with 128 projection angles and 128 radial bins at each projection. Figure 2 shows the two phantoms, A and B, used in our experiments. Phantom A is a 2-D Hoffman brain phantom with activities of 4:1:0 in grey matter, white matter, and CSF, respectively, and phantom B was derived from a digitized rhesus monkey autoradiograph which contains a variety of realistic edge structures. The total number of detector counts for phantoms A and B were approximately 300K and 500K, respectively.

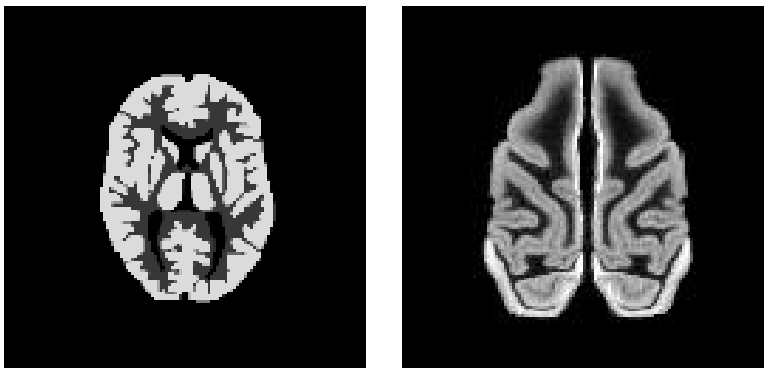

Figure 2. Phantoms used in the experiments: (a) Hoffman brain phantom A; (b) Autoradiograph phantom B.

To compare qualitatively the performance of our linearly combined quadratic priors summarized in (2), we chose three different values of τ (0, 0.5, 1). For each value of τ, we calculated the smoothing parameter λ using the ML estimate in (8). In this case, the estimated values of λ for phantom A were 3.87, 1.94, 1.30 for $\tau = 0$, 0.5, 1, respectively, and those for phantom B were 3.97, 2.80, 2.15 for $\tau = 0$, 0.5, 1, respectively. Figure 3 shows anecdotal reconstruction results for the four different algorithms – ML-EM and MAP-EM using the three different values of τ. For MAP-EM, we used the ICM algorithm. Several qualitative observations may be noted here; while the membrane prior ($\tau = 0$) incurs bias error around edge regions due to its oversmoothing behavior[30] (see Figs. 3 (b) and (f)), the TP prior ($\tau = 1$) seems to exhibit overshoots around sharp edges[17,20] (see Figs. 3 (d) and (h)). On the other hand, the linearly combined quadratic prior using $\tau = 0.5$ reveals a good compromise between the two extreme cases as shown in Figs. 3 (c) and (g).

To examine the performance of ordered subsets MAP-EM algorithms, we applied the ordered subsets principle to both OSL and ICM algorithms described in (4) and (7), respectively. In all these cases, the weight factor τ was set to 0.5. The variants of ordered subsets used in the simulation are distinguished by the OS level (number of subsets). Levels considered were 1, 2, 4, 8, 16, and 32. For a given level, the subsets of projections were ordered so as to maintain maximum distance between the projections in each subset. This ordering scheme is based on the requirement in the theoretical assumptions underpinning the OS algorithm that the sum of counts in projections forming the subsets be equal.[1]

For a given OS level, the estimated smoothing parameter via (8) was scaled (divided by the number of subsets) using (12); $\lambda_1 = 2.8$, $\lambda_2 = 1.4$, $\lambda_4 = 0.7$, $\lambda_8 = 0.35$, $\lambda_{16} = 0.175$, and $\lambda_{32} = 0.0875$. Figure 4 shows qualitative effects of scaling the smoothing parameter to match the degree of smoothness for different OS levels. The number of iterations for each level was determined based on the order-of-magnitude acceleration of OS-EM; the number of iterations for each level was adjusted so that effective iterations match 160 iterations for level 1, where the iteration

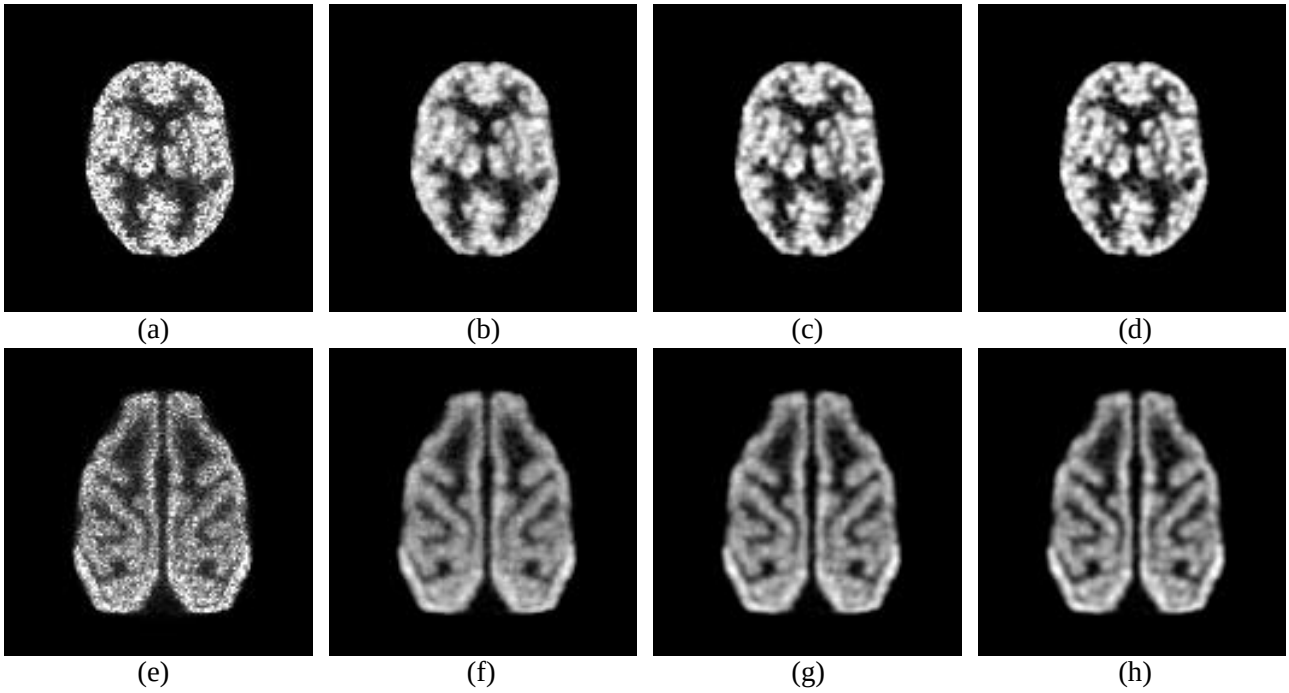

Figure 3. Anecdotal reconstructions for phantoms A ((a)-(d)) and B ((e)-(h)) using ML-EM and MAP-EM: (a)(e) ML-EM (32 iterations); (b)(f) MAP-EM ($\tau = 0$); (c)(g) MAP-EM ($\tau = 0.5$); (d)(h) MAP-EM ($\tau = 1$).

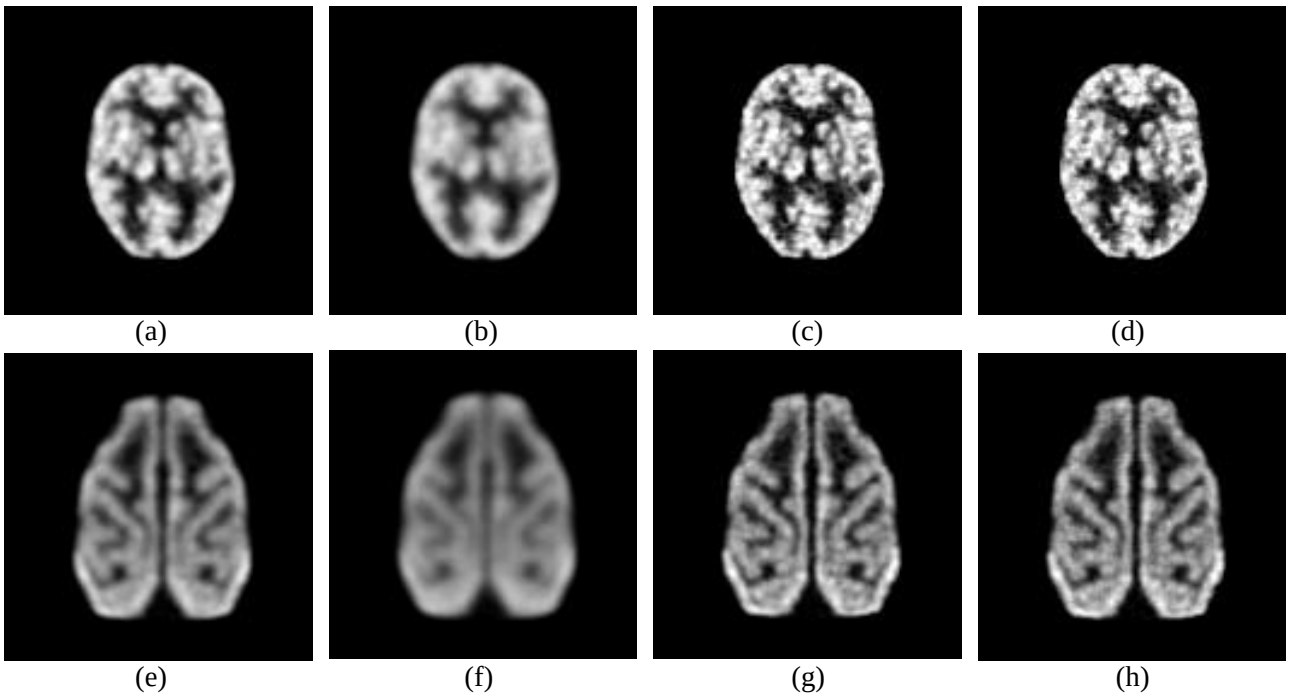

Figure 4. Qualitative effects of scaling the smoothing parameter to match the degree of smoothness for different OS levels (4 and 16): (a)(e) level 4, $\lambda = \lambda_1$; (b)(f) level 16, $\lambda = \lambda_1$; (c)(g) level 4, $\lambda = \lambda_4 = \frac{1}{4}\lambda_1$; (d)(h) level 16, $\lambda = \lambda_{16} = \frac{1}{16}\lambda_1$.

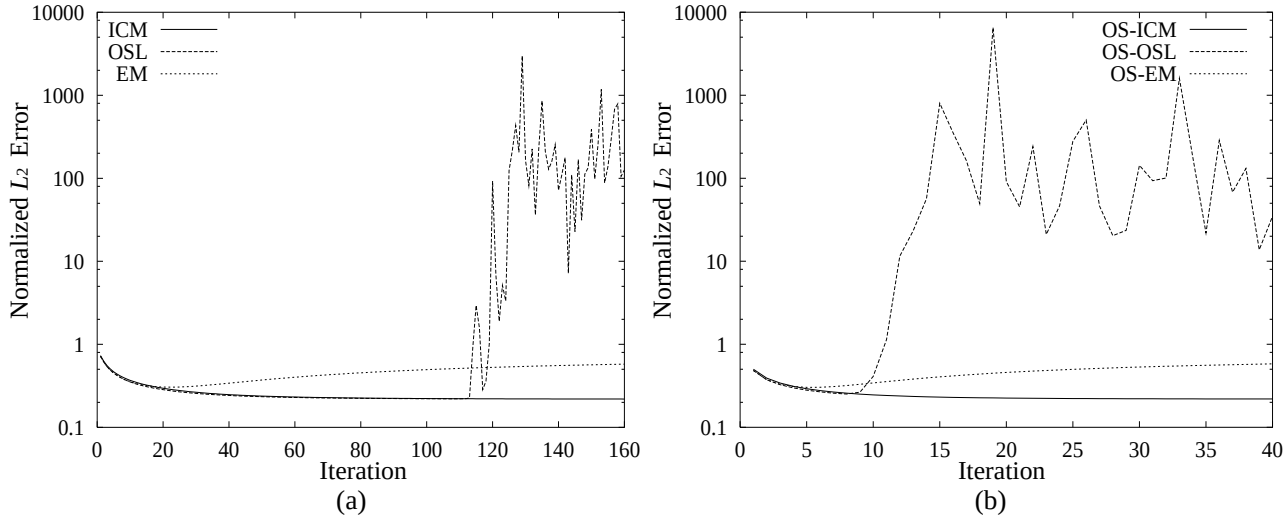

Figure 5. Normalized L_2 error versus iterations for OSL and ICM methods: (a) Results for non-OS algorithms (or OS level 1); (b) Results for OS algorithms (OS level 4).

number for level 1 was empirically chosen by examining its convergence behavior. For example, the number of iterations used for level 8 was 20. Figs. 4(a)(e) and (b)(f) are the results for levels 4 and 16, respectively, where each smoothing parameter estimated for level 1 in Figs. 3 (c) and (g) was directly used without scaling. In these cases, it is clearly shown that, as the number of subsets increases, the overall degree of smoothness also increases. On the other hand, comparison of Figs. 3 (c) and (g) to Figs. 4 (c)(d) and (g)(h), respectively, shows that, the degree of smoothness for different OS levels can be matched by scaling the smoothing parameter based on the relationship described in (12); the smoothing parameters for levels 4 and 16 in Fig. 4 (c)(g) and (d)(h), respectively, are 1/4 and 1/16 of those in (a)(e) and (b)(f), respectively.

Below, we examine convergence behavior of the OS algorithms using phantom B. In all our simulations for the convergence test, we set $\tau = 0.5$ and scaled the smoothing parameter for each OS level to match the degree of smoothness. Figure 5 shows normalized L_2 error (scaled by the L_2 norm of the true image) versus iterations for the OSL and ICM methods for levels 1 and 4. (The EM results are also included as references.) Compared to the ICM algorithm, the OSL algorithm badly diverged at both levels 1 and 4. (Note that the ordinate in Fig. 5 is $\log_{10}$ scaled.) In fact, according to our simulations, while the ICM algorithm yielded stable results for a large range of the smoothing parameter, the OSL algorithm in this case diverged at $\lambda \geq 1.70$ for level 1. It was also observed that level 4 diverged much earlier due to the acceleration by OS (see Fig. 5(b)).

Figure 6 shows the convergence behavior of the OS-ICM algorithm for six different OS levels (1, 2, 4, 8, 16, and 32) initialized with a uniform image, where (a) shows L_2 error versus iterations and (b) shows objective function decrease versus iterations. The order-of-magnitude acceleration can be seen at some iterations for OS levels 1 through 16; two iterations of level 16 decrease both the L_2 error and the objective almost as much as 32 iterations of level 1, 4 iterations of level 8, and 8 iterations of level 4. While there is a speed-up in using more subsets, the order-of-magnitude acceleration does not hold for relatively large number of subsets at low iterations. For example, two iterations of level 16 decrease the objective more than one iteration of level 32. In addition, convergent level 32 reveals relatively large L_2 error compared to other convergent lower levels. Therefore, increasing the number of subsets over 16 does not seem to improve the overall performance of the algorithm.

To evaluate the performance of algorithms more quantitatively, we generated 50 noise trials for each phantom by adding independent realizations of Poisson noise to the noiseless projection data, and calculated mean and standard deviation (STD) of normalized L_2 error (ϵ) over 50 noise trials, which are defined as

$$\bar{\epsilon} \stackrel{\text{def}}{=} \frac{1}{M} \sum_{m=1}^{M} \epsilon^m \quad \text{and} \quad \sigma \stackrel{\text{def}}{=} \sqrt{\frac{1}{M-1} \sum_{m=1}^{M} (\epsilon^m - \bar{\epsilon})^2},$$

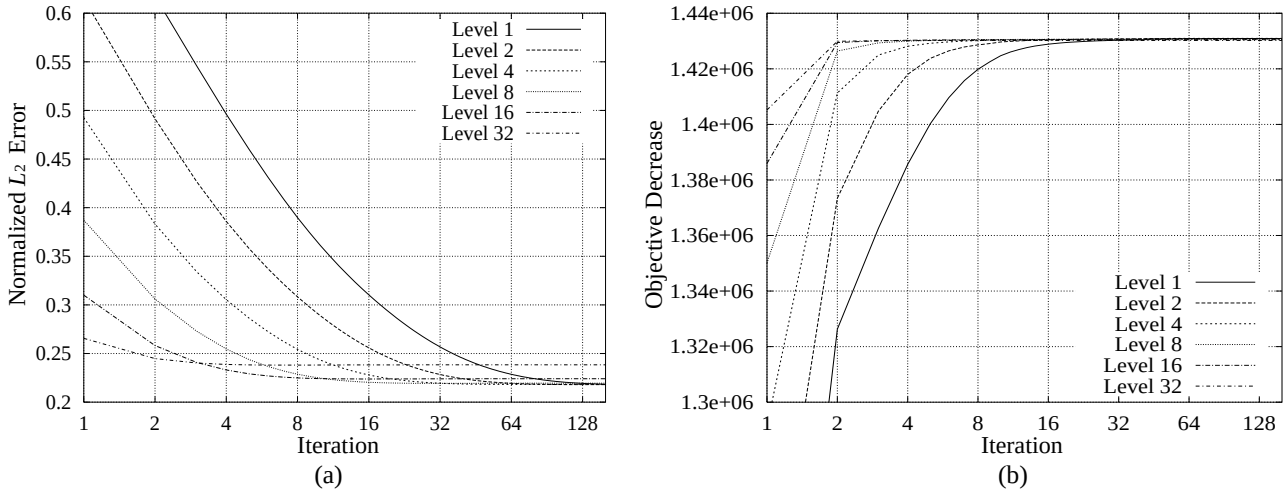

Figure 6. Convergence behavior of OS-ICM algorithms for 6 different OS levels: (a) Normalized L_2 error versus iterations; (b) Objective function decrease versus iterations.

respectively, where the normalized L_2 error of the m^{th} reconstruction $\hat{\mathbf{f}}^m$ is given by

$$\epsilon^m = \frac{\| \mathbf{f} - \hat{\mathbf{f}}^m \|}{\| \mathbf{f} \|}.$$

Table 2. Quantitative performance of OS-ICM and OS-EM in CPU s and normalized L_2 error.

Method	No of subsets	No of iterations	Total CPU s	CPU s per iteration	Mean (STD) of L_2 error for phantom A	Mean (STD) of L_2 error for phantom B
OS-ICM	1	160	46.7	0.29	0.2564 (0.0031)	0.2161 (0.0049)
	2	80	25.3	0.32	0.2564 (0.0031)	0.2162 (0.0047)
	4	40	15.4	0.39	0.2568 (0.0032)	0.2167 (0.0053)
	8	20	11.5	0.58	0.2579 (0.0032)	0.2179 (0.0061)
	16	10	9.4	0.94	0.2625 (0.0035)	0.2224 (0.0086)
	32	5	5.9	1.17	0.2780 (0.0045)	0.2376 (0.0154)
OS-EM	1	32	9.2	0.29	0.3443 (0.0043)	0.3111 (0.0100)
	2	16	4.9	0.31	0.3449 (0.0045)	0.3111 (0.0103)
	4	8	2.9	0.37	0.3465 (0.0045)	0.3121 (0.0112)
	8	4	2.2	0.54	0.3492 (0.0045)	0.3145 (0.0124)
	16	2	1.7	0.87	0.3550 (0.0051)	0.3200 (0.0151)
	32	1	1.0	1.0	0.3756 (0.0065)	0.3361 (0.0226)

Table 2 summarizes the quantitative performance of OS-ICM and OS-EM for the six different OS levels. Notice that both mean and STD of ϵ increase as the number of subsets N increases for both phantoms A and B. This shows that OS algorithms do not converge to the true minimum when $N > 1$. However, the results for OS levels up to 8 appear to be comparable to the convergent non-OS (level 1) algorithm in terms of both mean and STD of the L_2 error. In fact, the difference of the mean values of ϵ between levels 1 and 8 is less than 0.59% for phantom A and 0.83% for phantom B. The tabulated CPU seconds are from our C code run on a Pentium III 866MHz processor. Unlike the results in Fig. 6, the order-of-magnitude acceleration does not hold for actual CPU time in both OS-ICM and OS-EM. According to our simulation results shown in Table 2, increasing the number of subsets tends to slow down computational speed per iteration. In addition, due to the additional calculation of X_3 and the final update

equation which is slightly more complex than the calculation of X_2/X_1 for EM, CPU times per iteration for OS-ICM are slightly larger than those for OS-EM. However, for reasonable numbers of subsets, such as 8 or less, the increment of CPU time per iteration is negligible, and the OS-ICM algorithm indeed provides a substantial acceleration as well as comparable quality of reconstructions to level 1 in terms of normalized L_2 error.

5. CONCLUSION

We introduced an improved method to apply the ordered subsets principle to the Bayesian reconstruction problem for emission tomography. Although the OS principle can be applied to the OSL algorithm so that OS-EM can be extended to Bayesian MAP approaches, our method applied to a version of the ICM method provides more robust results since the pixel positivity is guaranteed and convergence is not affected by a large value of the smoothing parameter.

One could also consider applying the OS algorithm to other MAP-EM algorithms which offer better performance in edge preservation. However, the use of such algorithms requires calculation of additional hyperparameters, and the question of how these hyperparameters are affected by the number of subsets remains mostly moot. On the other hand, the smoothing parameter involved in our OS-ICM algorithms can be efficiently adjusted using the scaling method described in the paper so that the overall degree of smoothness for different numbers of subsets can be matched.

We note that, for moderate numbers of subsets, the CPU time per iteration for OS-ICM is indistinguishable from that for OS-EM. In addition, OS-ICM not only provides much better quality of reconstructions than OS-EM, but also allows easy implementation with any type of system just like OS-EM. We expect that, as OS-EM inevitably becomes practical, so will this form of accelerated ICM algorithm in clinical studies.

ACKNOWLEDGMENTS

This work was supported in parts by Nuclear Energy Research Grant from Ministry of Science and Technology, Korea, and Grant HMP-98-E-1-0008 from Ministry of Health and Welfare, Korea.

REFERENCES

1. H.M. Hudson and R.S. Larkin, "Accelerated Image Reconstruction Using Ordered Subsets of Porjection Data", *IEEE Trans. Med. Imaging*, MI-13(4), pp. 601–609, Dec. 1994.
2. P.J. Green, "Bayesian Reconstructions from Emission Tomography Data Using a Modified EM Algorithm", *IEEE Trans. Med. Imaging*, MI-9(1), pp. 84–93, March 1990.
3. A.R. De Pierro, "A Modified Expectation Maximization Algorithm for Penalized Likelihood Estimation in Emission Tomography", *IEEE Trans. Med. Imaging*, MI-14(2), pp. 132–137, Mar. 1995.
4. J.A. Browne and A.R. De Pierro, "A Row-Action Alternative to the EM Algorithm for Maximizing Likelihoods in Emission Tomography", *IEEE Trans. Image Processing*, 15(5), pp. 687–699, 1996.
5. J.A. Fessler and A.O. Hero, "Penalized Maximum-Likelihood Image Reconstruction Using Space-Alternating Generalized EM Algorithms", *IEEE Trans. Image Processing*, 4(10), pp. 1417–1429, Oct. 1996.
6. C. Bouman and K. Sauer, "A Unified Approach to Statistical Tomography Using Coordinate Descent Optimization", *IEEE Trans. Image Processing*, 5(3), pp. 480–492, Mar. 1996.
7. S. Geman and D. E. McClure, "Statistical Methods for Tomographic Image Reconstruction", *Bulletin of the International Statistical Institute*, LII-4, pp. 5–21, 1987.
8. E. Levitan and G. T. Herman, "A Maximum *A Posteriori* Probability Expectation Maximization Algorithm for Image Reconstruction in Emission Tomography", *IEEE Trans. Med. Imaging*, MI-6, pp. 185–192, 1987.
9. T. Hebert and R. Leahy, "A Generalized EM Algorithm for 3-D Bayesian Reconstruction for Poisson Data Using Gibbs Priors", *IEEE Trans. Med. Imaging*, MI-8(2), pp. 194–202, June 1989.
10. V. E. Johnson, W. H. Wong, X. Hu, and C. Chen, "Bayesian Reconstruction of PET Images Using Gibbs priors", In D. A. Ortendahl and J. Llacer, editors, *Information Processing in Medical Imaging*, pp. 15–28, Wiley-Liss, 1989.
11. D. S. Lalush and B. M. W. Tsui, "Simulation Evaluation of Gibbs Prior Distributions for Use in Maximum *a posteriori* SPECT Reconstructions", *IEEE Trans. Med. Imaging*, MI-11, pp. 267–275, 1992.

12. G. Gindi, M. Lee, A. Rangarajan, and G. Zubal, "Bayesian Reconstruction of Functional Images Using Anatomical Information as Priors", *IEEE Trans. Med. Imaging*, MI-12, pp. 670–680, Dec. 1993.
13. S.J. Lee, A. Rangarajan, and G. Gindi, "Bayesian Image Reconstruction in SPECT Using Higher Order Mechanical Models as Priors", *IEEE Trans. Med. Imaging*, MI-14(4), pp. 669–680, Dec. 1995.
14. J. Besag, "Spatial Interaction and the Statistical Analysis of Lattice Systems", *Journal of the Royal Statistical Society,* Series B, 36, pp. 192–236, 1974.
15. S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images", *IEEE Trans. Patt. Anal. Mach. Intell.*, PAMI-6(6), pp. 721–741, November 1984.
16. W. E. L. Grimson, *From Surfaces to Images*, The MIT press, Cambridge, MA, 1981.
17. M. Gokmen and A.K. Jain, "$\lambda\tau$-Space Representation of Images and Generalized Edge Detector", *IEEE Trans. Patt. Anal. Mach. Intell.*, 19(6), pp. 545–563, Jun. 1997.
18. D. Terzopoulos, "Multilevel Computational Processes for Visual Surface Reconstruction", *Computer Vision, Graphics, and Image Processing*, 24, pp. 52–96, 1983.
19. D. Terzopoulos, "Regularization of Inverse Visual Problems Involving Discontinuities", *IEEE Trans. Patt. Anal. Mach. Intell.*, 8, pp. 413–424, July 1986.
20. S.J. Lee, Y. Choi, and G. Gindi, "Validation of New Gibbs Priors for Bayesian Tomographic Reconstruction Using Numerical Studies and Physically Acquired Data", *IEEE Trans. Nuclear Science*, NS-46(6), pp. 1254–1261, Dec. 1999.
21. Z. Zhou, R.M. Leahy, and J. Qi, "Approximate Maximum Likelihood Hyperparameter Estimation for Gibbs Priors", *IEEE Trans. Med. Imaging*, MI-6(6), pp. 844–861, June 1997.
22. S.S. Saquib, C.A. Bouman, and K. Sauer, "ML Parameter Estimation for Markov Random Fields, with Applications to Bayesian Tomography", *IEEE Trans. Image Processing*, 7(7), pp. 1029–1044, Jul. 1998.
23. S. J. Lee, G. R. Gindi, I. G. Zubal, and A. Rangarajan, "Using Ground-Truth Data to Design Priors in Bayesian SPECT Reconstruction", In Y. Bizais, C. Barillot, and R. D. Paola, editors, *Information Processing in Medical Imaging*, pp. 27–38, Kluwer Academic Publishers, 1995.
24. S. Geman and C. Graffigne, "Markov Random Fields Image Models and their application to Computer Vision", In A. M. Gleason, editor, *Proc. of the Intl. Congress of Mathematicians 1986*, pp. 1496–1517, American Mathematical Society, Providence, 1987.
25. J. Besag, "On the Statistical Analysis of Dirty Pictures", *Journal of the Royal Statistical Society,* Series B, 48(3), pp. 259–302, 1986.
26. R. C. Walovitch, T. C. Hill, S. T. Garrity, E. H. Cheesman, B. A. Burgess, D. H. O'Leary, A. D. Watson, M. V. Ganey, R. A. Morgan, and S. J. Williams, "Characterization of Technetium-99m-L,L-ECD for Brain Perfusion Imaging, Part 1: Pharmacology of Technetium-99m ECD in Nonhuman Primates", *Journal of Nuclear Medicine*, 30, pp. 1892–1901, 1989.
27. G. Gindi, D. Dougherty, I. Hsiao, and A. Rangarajan, "Autoradiograph Based Phantoms for Emission Tomography", In *Proc. SPIE Symposium on Medical Imaging - Image Processing*, pp. 403–414, 1997.
28. D. Dougherty, I.T. Hsiao, and G.R. Gindi, "Computerized Biological Brain Phantom for Evaluation of PET and SPECT Reconstruction", *IEEE Trans. Nuclear Science*, NS-45(3), pp. 1238–1243, 1998.
29. K. Lange, "An Overview of Bayesian Methods in Image Reconstruction", In *Digital Image Synthesis and Inverse Optics – SPIE*, volume 1351, pp. 270–287, 1990.
30. S.J. Lee, I.T. Hsiao, and G.R. Gindi, "Quantitative Effects of Using Thin-Plate Priors in Bayesian SPECT Reconstruction", In *Proc. SPIE Image Reconstruction and Restoration II*, pp. 252–263, Jul. 1997.