Parallelizable Bayesian Tomography Algorithms with Rapid, Guaranteed Convergence *

Jun Zheng[†], Suhail Saquib[◊], Ken Sauer[‡] and Charles Bouman[*]

[†]Delphi Delco Electronics Systems, Kokomo, IN 46904-9005

[◊]Polaroid Corporation, Cambridge, MA 02139

[‡]Department of Electrical Engineering, University of Notre Dame

Notre Dame, IN 46556, (219) 631-6999

[*]School of Electrical Engineering, Purdue University

West Lafayette, IN 47907-0501, (317) 494-0340

March 8, 2000

Abstract [1]

Bayesian tomographic reconstruction algorithms generally require the efficient optimization of a functional of many variables. In this setting as well as in many other optimization tasks, functional substitution (FS) has been widely applied to simplify each step of the iterative process. The function to be minimized is replaced locally by an approximation having a more easily manipulated form, e.g. quadratic, but which maintains sufficient similarity to descend the true functional while computing only the substitute. In this paper, we provide two new applications of FS methods in iterative coordinate descent for Bayesian tomography. The first is a modification of our coordinate descent algorithm with 1-D Newton-Raphson approximations to an alternative quadratic which allows convergence to be proven easily. In simulations we find essentially no difference in convergence speed between the two techniques. We also present a new algorithm which exploits the FS method to allow parallel updates of arbitrary sets of pixels using computations similar to iterative coordinate descent. The theoretical potential speed up of parallel implementations is nearly linear with the number of processors if communication costs are neglected.

EDICS number IP 2.3 (under new EDICS: TOMO, NUXR)

Correspondence to: *Ken Sauer*

Dept. of Electrical Engineering

University of Notre Dame

Notre Dame, IN 46556

sauer@nd.edu

office: (219) 631-6999

[1]Permission to publish this abstract separately is granted.

1 Introduction

With the choice of convex potential functions for Markov random field (MRF) style *a priori* image models, both maximum likelihood (ML) and maximum *a posteriori* probability (MAP) tomographic image reconstructions may be formulated as large scale convex optimization problems. Many approaches to this optimization have been proposed, among which popular alternatives have been variants of expectation-maximization (EM) [1], an approach derived from indirect optimization through the introduction of the notion of an unobservable *complete* data set whose expectation forms the algorithmic basis for ML. Unfortunately, for most common image models, EM has no simple, closed form for the MAP problem, though several modifications have been proposed which allow inclusion of the *a priori* density term[2, 3, 4]. Probably a more serious limitation of EM is its slow convergence for typical tomographic problems.

It is perhaps more natural to look at the numerical side of statistical tomographic estimation as a problem of direct optimization in the pixel domain, which brings to mind gradient-based techniques. The similarity of EM to gradient descent has often been noted [5, 6], and helps explain their common poor performance for tomographic reconstruction. Improvements such as preconditioned gradient or preconditioned conjugate gradient may have significantly more promise in terms of speed. But realistic positivity constraints remain more difficult to incorporate effectively than in, for example, EM.

This high-dimensional optimization can be greatly simplified by viewing the problem as a sequence of low-dimensional problems. Many previous techniques such as the algebraic reconstruction technique (ART) are known as "row-action," since they attempt to solve a sequence of problems, each corresponding to a subset of the data and therefore rows of the projection matrix. Much has been published of late concerning the ordered-subsets version of EM (OS-EM) [7], which rotates among subsets of projection data, resolving the EM-type problem for each subset in a fashion reminiscent of ART. But while OS-EM improves the speed of initial descent toward the functional minimum, it does not in general converge without the addition of temporally varying auxiliary parameters whose schedule apparently cancels gains in convergence speed [8]. The difference between the true MAP estimate and the results achieved with common implementations of OS-EM may be substantial [9].

ML or MAP optimization may also be solved sequentially among pixels, via methods known as

"column-action," due to the correspondence between a pixel and a single column of the transform matrix. Among the conceptually simplest is a technique we call iterative coordinate descent (ICD) [10, 11], which views the ML or MAP estimation problem, be it transmission or emission tomography or any other convex formulation, as a direct optimization task. ICD solves these problems by sequentially minimizing the objective functions posed by ML or MAP estimation, using greedy pixel-wise updates in the style of iterated conditional modes (ICM) [12] and Gauss-Seidel iterations [13]. Because ICD has fast convergence at high spatial frequencies, it performs very well in standard tomographic problems where the filtered backprojection (FBP) image serves as a good initial condition [14]. A variant using local Newton-Raphson style approximations of the likelihood function, called ICD/Newton-Raphson (ICD/NR), simplifies update computation with convergence speed in iteration counts which is very similar to direct ICD[10]. One may also solve the EM formulation pixel-sequentially, preserving the provable convergence of EM while substantially improving its speed, as in the space-alternating EM algorithm (SAGE) of Fessler and Hero [15]. Column-action methods can easily be made to converge reliably to the unique global minimum of the ML or MAP functional.

Here we present two improvements to current forms of ICD: (1) global convergence of the approximate greedy descent algorithm follows from the introduction of a new local quadratic approximation of the log-likelihood and (2) we derive a scheme for parallel updates of arbitrary sets of pixels while maintaining convergence properties. We approach these developments by discussing a technique exploited by many proposed algorithms in this field, which we call functional substitution (FS). The term FS is here understood to include the many choices for simplifying potentially expensive minimization by temporarily replacing the true convex function to be optimized by a simpler one which maintains its most important local properties, such as low-order derivatives. ICD/NR is a simple example of these for tomographic estimation, using a local one-dimensional quadratic [11]. EM also can be viewed as an FS method which solves the ML estimation by replacing the likelihood function with its expectation over the realizations of the "complete" data set. De Pierro [4] used the functional substitution idea in his extension of the EM framework directly to the MAP emission problem. For transmission tomography, EM has no simple closed form even for ML, leading Lange to apply a substitution similar in form to that of De Pierro to solve the transmission problem, again with decoupled parallel updates [16].

The ICD/NR algorithm has been experimentally demonstrated to converge very rapidly com-

pared to EM algorithms, but is thus far not guaranteed theoretically to converge to the unique global MAP solution. We therefore first present a modified ICD algorithm which we call ICD functional substitution (ICD/FS). The new algorithm locally approximates the exact log-likelihood function with an alternative quadratic function to the Newton-Raphson choice. Like ICD/NR, ICD/FS easily incorporates non-negativity constraints and non-Gaussian prior distributions for the MAP reconstruction problem. However, ICD/FS is guaranteed theoretically globally convergent for both the emission and transmission reconstruction problem when the log-prior distribution is strictly convex. The experimental convergence speeds of the two are identical.

Many low-cost high-speed computational engines are parallel configurations of medium speed processors. In clinical application of statistical reconstruction, it may be of benefit to implement algorithms specifically tailored for such parallel architectures. As a second example of the utility of the FS idea in optimization for tomographic reconstruction, we present a FS method which allows us to update arbitrary subsets of pixels in parallel, while maintaining any provable convergence possessed by the form of ICD algorithm chosen. We apply the parallel updates here using ICD/FS at each pixel. For practical numbers of parallel processors, the algorithm maintains the convergence advantages of ICD and yields linear speed-up with the number of processors in terms of iteration counts.

2   Functional Substitution in Tomographic Reconstruction

For the emission problem, the log-likelihood may be computed as

$$\log \mathcal{P}(Y = y | X = x) \quad = \quad \sum_{i=1}^{M} \left(-A_{i*}x - r_i + y_i \log\{A_{i*}x + r_i\} - \log(y_i!) \right) \tag{1}$$

where x is the unknown image, y is the projection data, M is the number of projections, A_{i*} is the ith row of the projection matrix A, y_i are the observed photon counts for projection i, and r_i are additive terms usually due to background noise or random coincidences in the case of positron emission tomography (PET). In the transmission case, we have

$$\log \mathcal{P}(Y = y | X = x) \quad = \quad \sum_{i=1}^{M} \left(-y_T \exp^{-A_{i*}x} + y_i (\log y_T - A_{i*}x) - \log(y_i!) \right) \tag{2}$$

where y_T is the photon dosage per ray [11].

Both log-likelihood functions have the form

$$-\log \mathcal{P}(Y = y | X = x) = \sum_{i=1}^{M} G_i(A_{i*}x), \tag{3}$$

in which the $\{G_i\}$ and therefore their sum are convex. The common form leads to similar methods of solving these two problems. In Bayesian formulations, we denote the *a priori* density function for X as $\mathcal{P}(X = x)$, with MAP tomographic reconstruction resulting in the numerical optimization problem of

$$\hat{x}_{MAP} = \arg\min_x \{-\log\mathcal{P}(Y = y|X = x) - \log\mathcal{P}(X = x)\}.$$

In this paper we deal only with those models for which $\log\mathcal{P}(X = x)$ is a concave function of x, making the entire optimization problem convex.

Several researchers have resorted to the direct optimization of the Poisson-modeled problem. Gradient ascent methods may be applied directly to the log-likelihood function, with several of these techniques having been derived from the similarity of EM to gradient ascent [17, 6, 18]. The improved preconditioned conjugated gradient approach of Mumcuoğlu *et al.* has been explicitly formulated for the MAP problem [19] as well. The ICD method [11] is a direct pixel-wise optimization of the MAP functional, typically allowing more rapid convergence than at least the unaccelerated gradient-type algorithms.

Though direct optimization is tractable for this problem, the celebrated EM algorithm [20] has become very widely known and applied to its solution. Based on the notion of indirect optimization through a set of unobservable "complete" data, it relies on a concept which is generally important to this problem, that of a tractable substitute function in place of the true log-likelihood. The EM algorithm replaces the log-likelihood function $\log\mathcal{P}(Y = y|X = x)$ with the substitute function Q in [1],

$$Q(x; x^m) = E[\log p(z|x)|x^m, y]$$

where z are the emission counts for each pixel/detector combination. As is the case with all FS methods, the substitute functional form is determined by the most recent estimate x^m. In EM, m indexes iterations, each of which updates all pixels in parallel. The substitute function matches the actual concave objective in its first derivatives at x^m, but has second derivatives of larger magnitude. Solving the optimization with $Q(x; x^m)$ at iteration m in place of $\log\mathcal{P}(Y = y|X = x)$ guarantees that at each step, the log-likelihood increases, since any increase in the former must result in an equal or greater increase in the latter. Most importantly operationally, it allows all pixel updates to be computed independently. The EM strategy does not result in a simple maximization step in transmission tomography, but Lange derived an alternative substitute functional resulting

in an approach called the Convex algorithm which similarly decouples parallel pixel updates [16].

The linkage of pixels through the prior generally makes the maximization step for Bayesian EM algorithms non-trivial, though several modifications have adapted EM to the MAP estimator [3, 2]. De Pierro proposed a modified EM algorithm for the MAP problem [4], observing that a substitution could be made for $\log \mathcal{P}(X = x)$ just as EM does for the log-likelihood, with the substitute functional having decoupled dependencies on pixel update values. The algorithm consists of replacing the log-prior function $\log \mathcal{P}(X = x)$ by an alternative cost function $C(x; x^m)$ at the $(m + 1)^{th}$ update, having the properties

$$C(x; x) = \log \mathcal{P}(X = x)$$

$$C(x; x^m) \leq \log \mathcal{P}(X = x),$$

and independent optimization at each pixel. This allows a more direct extension of EM to MAP estimation than previously known methods. This FS-method based EM algorithm has the advantage of guaranteed monotonic descent of the objective with simple computation, but retains the limited convergence speed of EM.

3 Provably Convergent Coordinate Descent by FS

3.1 ICD/FS Algorithm Description

The ICD method directly, sequentially optimizes the MAP cost function with respect to each pixel (i.e. coordinate of x_j) of the image with the remainder of x fixed. Previous results have shown that its convergence in tomographic problems is significantly faster than EM and gradient descent [14, 11]. Let x^n be the image at the nth update. The ICD update of the pixel x_j is computed by solving the MAP equation

$$x_j^{n+1} = \arg \min_{x_j \geq 0} \{F_j(x_j) + P_j(x_j)\} \tag{4}$$

where $F_j(x_j)$ and $P_j(x_j)$ represent the contribution of the likelihood and prior terms, respectively, to the objective function expressed in terms of only x_j. F_j and P_j are functions also of all other pixels in x^n, but since we deal primarily with optimization in one variable with all others fixed, we have suppressed x^n as an argument in the interest of economy of notation. For the present discussion, we let the index n increase with each pixel visit, making j, the index of the updated pixel, implicitly a function of n. The particular form of $F_j(x_j)$ varies for the emission and transmission tomography

cases of equations (1) and (2). But in both cases $F_j(x_j)$ is a convex function on $I\!\!R^+$ and its derivative $f_j(x_j) = \frac{dF_j(x_j)}{dx_j}$ is strictly concave on $I\!\!R^+$.

Since (4) updates with respect to a single x_j at each step, the N-dimensional optimization problem changes to 1-dimensional. This simplifies the MAP problem, making enforcement of positivity constraints trivial, but requires fast implementations of the sequential 1-D problems. Unfortunately, exact optimization of (4), requiring repeated evaluation of the derivatives of $F_j(x_j)$, may be computationally expensive.

An FS method can solve this problem simply by replacing the true log-likelihood function in 1-D at each iteration with a quadratic functional form. This kind of quadratic approximation is easy to optimize with any of several numerical methods. The most common quadratic form is Newton-Raphson type, which was proposed as the ICD/NR algorithm in [10]. It locally approximates the function with its second order Taylor series, replacing $F_j(x_j)$ with

$$\tilde{F}_j(x_j) = \theta_1(x_j - x_j^n) + \frac{1}{2}\theta_2(x_j - x_j^n)^2,$$

where

$$\theta_1 = f_j(x_j^n), \tag{5}$$

$$\theta_2 = f_j'(x_j^n) \equiv \left. \frac{df_j(x_j)}{dx_j} \right|_{x_j = x_j^n}. \tag{6}$$

$P_j(x_j)$ remains unchanged, which may create a nonlinear problem. Still, since the log-likelihood term typically dominates the computation, the quadratic approximation greatly simplifies the update. This NR-type approximation in the ICD algorithm has demonstrated quite good experimental convergence and easily incorporates non-negativity constraints. However, since in general Newton-Raphson steps are not guaranteed to monotonically approach a fixed point, a theoretical proof of convergence for the resulting iterations may be difficult.

Quadratic approximation algorithms may be made monotonically convergent in convex problems by locally replacing the Hessian matrix, or θ_2 from the 1-D version in (6), by certain estimates more liberal in their magnitude, causing a more conservative update. The quadratic lower bound algorithm in one dimension, for example, uses the lower bound on the second derivative of a concave function over an interval of interest [21]. The algorithm we shall call ICD/FS consists of replacing θ_2 with its average on the interval between 0 and x_j^n. The update equations for ICD/FS are given

as follows.

$$\breve{\theta}_2 = \begin{cases} \frac{f_j(x_j^n) - f_j(0)}{x_j^n} & \text{if } x_j^n > 0 \\ f_j'(0) & \text{if } x_j^n = 0 \end{cases} \tag{7}$$

$$\breve{F}_j(x_j) = \theta_1(x_j - x_j^n) + \frac{1}{2}\breve{\theta}_2(x_j - x_j^n)^2 \tag{8}$$

$$x_j^{n+1} = \arg\min_{x_j \geq 0} \left\{ \breve{F}_j(x_j) + P_j(x_j) \right\} \tag{9}$$

Figure 1 illustrates the difference between ICD/NR and ICD/FS. The plots are the first derivatives of NR-type and FS-type approximations of the original log-likelihood. As mentioned earlier, both use the same first derivative for the substitute functional at x_j^n but ICD/FS takes as its second derivative the average rate of change in the first derivative over the interval from 0 to x_j^n. Since the second derivative is monotonic decreasing, $\breve{\theta}_2$ is greater than or equal to θ_2, making the ICD/FS update more conservative. The updated optimal pixels are the zeros through root-finding operations, given the current state of all other pixels in x^n.

The ICD/FS algorithm can be applied in both emission and transmission tomography problems. The only difference between these two cases is the specific computation of the values for $f_j(x_j^n)$, $f_j(0)$, and $f_j'(0)$ in (5) and (7). For the emission case, these values are given by

$$f_j(x_j^n) = \sum_{i=1}^{M} A_{ij}\left(1 - \frac{y_i}{p_i^n}\right) \tag{10}$$

$$f_j(0) = \sum_{i=1}^{M} A_{ij}\left(1 - \frac{y_i}{p_i^n - A_{ij}x_j^n}\right) \tag{11}$$

$$f_j'(0)\Big|_{x_j^n=0} = \sum_{i=1}^{M} y_i \left(\frac{A_{ij}}{p_i^n}\right)^2 \tag{12}$$

where A_{ij} is the contribution of the jth pixel to the ith projection, and $p_i^n = A_{i*}x^n + r_i$ is the ith projection of the reconstruction at iteration n. Note that p_i^n may be efficiently updated by $p_i^{n+1} = p_i^n + A_{ij}(x_j^{n+1} - x_j^n)$, with computation reduced by the sparse structure of A.

For the transmission case, the update values are given by

$$f_j(x_j^n) = \sum_{i=1}^{M} A_{ij}\left(y_i - y_T e^{-p_i^n}\right) \tag{13}$$

$$f_j(0) = \sum_{i=1}^{M} A_{ij}\left(y_i - y_T e^{-p_i^n} e^{A_{ij}x_j^n}\right) \tag{14}$$

$$f_j'(0)\Big|_{x_j^n=0} = \sum_{i=1}^{M} A_{ij}^2 y_T e^{-p_i^n} \tag{15}$$

8

where $p_i^n = A_{i*}x^n$, i.e. r_i is assumed zero.

ICD/FS has nearly the same computational requirements as ICD/NR since it generally requires the computation of two first derivatives in place of the first and second derivatives required for ICD/NR. This computational cost includes approximately twice the number of multiplies and adds per iteration as gradient descent [11], but approximately equal numbers of additions and accesses to entries of the transform matrix A. The exponentiations required for the transmission problem can be efficiently implemented via table look-ups.

3.2   Global Convergence of ICD/FS

In order to prove the global convergence of this new ICD/FS algorithm, we simply verify that it meets the assumptions and necessary conditions of the global convergence proof presented by Fessler and Hero in [15] for convergence of SAGE under positivity constraints. Since this proof requires continuity of the log-likelihood on $I\!\!R^{N+}$, we must assume that the background noise is greater than zero, i.e. $r_i > 0$ in emission case (1). We discuss alternative methods for the case $r_i = 0$ later in this section.

Most of these conditions are either the same as for [15], or may be simply verified[2]. We will demonstrate the critical Condition 1, which states that the change in the substitute function is an upper bound on the change in the true functional to be minimized. By the construction of the function $\breve{f}_j(x_j) = \frac{d\breve{F}_j(x_j)}{dx_j}$, we know that $\breve{f}_j(0) = f_j(0)$, and $\breve{f}_j(x_j^n) = f_j(x_j^n)$. Since for both the emission and transmission case, $f_j(x_j)$ is a concave function and $\breve{f}_j(x_j)$ is a linear function, it follows that

$$f_j(x_j) \begin{cases} \geq \breve{f}_j(x_j) & 0 \leq x_j < x_j^n \\ \leq \breve{f}_j(x_j) & x_j > x_j^n \end{cases}$$

Integration of $f_j(x_j)$ and $\breve{f}_j(x_j)$ results in the inequality

$$F_j(x_j^n) - F_j(x_j) \geq \breve{F}_j(x_j^n) - \breve{F}_j(x_j) \ .$$

Defining the functions $\Phi_j(x_j) = F_j(x_j) + P_j(x_j)$ and $\breve{\Phi}_j(x_j) = \breve{F}_j(x_j) + P_j(x_j)$, both implicitly functions of x^n, then results in the following lemma.

Lemma: Let $F_j(x_j) + P_j(x_j)$ be convex, and $F_j(x_j)$ be continuously differentiable on $I\!\!R^+$. Furthermore, let $f_j(x_j) = \frac{dF_j(x_j)}{dx_j}$ be concave and continuous on $I\!\!R^+$. Then for all $x_j \in I\!\!R^+$

$$\Phi_j(x_j^n) - \Phi_j(x_j) \geq \breve{\Phi}_j(x_j^n) - \breve{\Phi}_j(x_j) \ . \tag{16}$$

---

[2]Using $\breve{\Phi}_j(t; x^n)$ to make explicit the dependence of $\breve{\Phi}$ in (16) on the current state, x^n, continuous differentiability of $\breve{\Phi}_j(t; x)$ as a function of $(t; x)$ on $I\!\!R^{(N+1)+}$ is also necessary in Condition 2 of [15].

This guarantees that the decrease in the substitute function will result in an equal or greater decrease in the exact function. Based on this lemma and the conditions proved in [15], the global convergence of the ICD/FS algorithm follows.

3.3 ICD/FS with Zero Background Emission Noise

As mentioned previously, the emission case when $r_i = 0$ is special, since in this case the log-likelihood term may tend to $-\infty$ on $I\!\!R^+$. This occurs in the unusual case in which x_j is the only nonzero pixel on a projection which has a nonzero photon count. In this case, $A_{i*}x = A_{ij}x_j$ and the log-likelihood functions have terms of the form $\log x_j$ which tend to $-\infty$ as $x_j \to 0$. There are a number of possible strategies for handling this case numerically.

Strategy 1: A very simple method is to set r_i to a small number such as $r_i = \frac{1}{100M}$. This guarantees that the expected number of additional photons due to this adjustment summed over all projections is much less than 1. In practice, such a small perturbation to the model should not have a significant effect on the resulting reconstruction. This strategy also has the added benefit of making the algorithm more robust to floating point round-off error.

Strategy 2: Modify the algorithm so that in the case when $f_j(0) = -\infty$, the function is computed at $f_j(\epsilon)$ where ϵ is chosen to assure that $f_j(\epsilon) < 0$. One such choice is $\epsilon = \min(\gamma, \frac{x_j^n}{2})$, where $\gamma = \min_j \left\{ \left(p^{-1}\sigma^{-p} + \sum_{i=1}^{M} A_{ij} \right)^{-1} \right\}$. In this case, $\breve{\theta}_2$ is given by

$$\begin{aligned} \breve{\theta}_2 &= \frac{f_j(x_j^n) - f_j(\epsilon)}{x_j^n - \epsilon} \\ &= \sum_{i=1}^{M} \frac{A_{ij}^2 y_i}{(p_i^n)^2 - p_i^n A_{ij}(x_j^n - \epsilon)} \end{aligned}$$

and the update equation is still given by

$$x_j^{n+1} = \arg \min_{x_j \geq 0}\{\breve{F}_j(x_j) + P(x_j)\}$$

3.4 ICD/FS Numerical Results

Numerical results include three data sets, two from emission and one from industrial transmission tomography. The synthetic phantom emission data used for Fig. 2 are 128×128 projections with approximately 3 million total counts, while the single photon emission computed tomography (SPECT) sestamibi heart perfusion data of Figure 3 form a single slice of dimension 120×128 from a 3D set. Both reconstructions are computed at a resolution of 128×128 pixels. The transmission data used in Fig. 4 are a 192×256 subset of a high-resolution industrial CT scan, with reconstruction

10

at 256×256. We solve MAP reconstructions using for comparison several other methods relying on functional substitution for the design of updates, all with guaranteed convergence. We include iterations of the Lange's convex method [16] for the transmission reconstruction, and both Fessler and Hero's SAGE-3 [15] and De Pierro's method [4] in the emission problems. In trials with Gaussian priors, plots show also results from the preconditioned conjugate gradient (PCG) approach [19] with a line search for optimal step size at each iteration. As discussed in [11], all the algorithms compared require the equivalent of a forward and backprojection at each iteration and thus have the same order of computational cost. The ICD techniques require approximately the same number of additions and twice as many multiplies and divides per iteration as EM. All the sequential algorithms typically include a root-finding step at each pixel, which may raise cost somewhat for non-Gaussian prior models.

The *a priori* image model here consists of two choices of p for the generalized Gaussian Markov random field (GGMRF) [22] prior model with the prior log density function of $P(x) = \sum_{j,k} \frac{b_{jk}}{p\sigma^p} |x_k - x_j|^p$, where b_{jk} is the coefficient linking pixels j and k, σ is a scale (temperature) parameter, and $1 \leq p \leq 2$ is a parameter which controls the smoothness of the reconstruction. The parameter σ for all six cases and the dosage parameter y_T for the transmission data were estimated directly from the noisy projection data via the maximum-likelihood methods described in [23]. For $1 < p < 2$, $P(x)$ has unbounded positive second derivatives at points where neighboring pixels are equal, violating the twice-differentiability called for in Condition 2 in the convergence proof of [15], but as shown in the Appendix, this case still carries through the convergence proof without modification.

Except for one case noted below, all iterative reconstructions begin with a filtered backprojection image with negative-valued pixels set to zero or, for the generalized EM algorithm, a small positive value to avoid potential problems with multiplicative corrections. Since our goal is minimization of the MAP objective for a convex problem, the final image is independent of this initial condition and no "stopping rule" for early termination is considered. The ICD algorithms are normally run until the objective function's value is converged in, e.g., the first 8-10 decimal places and beyond visible changes in the image, though our plots do not necessarily show values all the way to the termination point.

Plots in Figs. 5, 6 show the convergence of the ICD/NR and ICD/FS in the emission cases described above. These trials employ sequential pixel updates in raster ordering for the ICD methods and SAGE-3, not necessarily the fastest converging scan pattern. The plots show that the

convergence properties of ICD/NR and ICD/FS are virtually identical, in spite of the potentially more conservative updates of ICD/FS. But this is not surprising, since the log-likelihood function is generally close to quadratic and the values of θ_2 and $\breve{\theta}_2$ are therefore generally very close. In fact, we have previously found there to be little difference in pixel values between ICD/NR updates and exact greedy pixel-wise minimization of $F_j(x_j) + P_j(x_j)$ [10]. Although a proof for the global convergence of ICD/NR has not yet been found, its convergence appears consistently rapid. In practical applications, the greater conceptual simplicity of ICD/NR may in fact still make it the more desirable of the two. The performance of SAGE-3 is close to that of ICD in all cases, though in the head phantom plots we see somewhat slower asymptotic convergence, presumably due to the fact that the substitute function of SAGE, based on the expectation derived from EM, is more conservative than those of the ICD quadratic approximations. DePierro's method, using the EM substitute functional for all pixels simultaneously, is significantly slower to converge.

The PCG results for $p = 2$, which are intermediate in convergence speed, require additional comment. The algorithm in [19] uses a one-sided quadratic penalty for pixels violating the positivity constraint, with parameters governing behavior at the boundary to be set heuristically. We found the convergence behavior of PCG to be potentially sensitive to both these parameters' values and the initial condition chosen for optimization. In each featured result, we experimented with a wide range of values for each of the parameters, including those suggested in [19], and both the FBP and uniform images as initial conditions. In each case we present the best results achieved with (1) violation of the positivity constraint limited to a magnitude of 0.01 of the maximum in the reconstruction, and (2) PCG objective function value close to that of the other, strictly constrained methods. For the SPECT phantom, PCG converged fastest with the uniform initial image and the heuristic parameters set to values similar to those in [19]. The plot reaches a value higher than the other methods due only to the better fit in likelihood achievable with the slightly negative-valued pixels. (The added penalty is not included in the computation of the objective function in the plot.) The non-negativity penalty proved more problematic with the head phantom, where it is active for nearly all the area outside the phantom. The penalty is encountered in the first iterations regardless of which initial condition is chosen, and a weighting of the penalty adequate to satisfy our two requirements above leads to very slow convergence. To achieve a reasonable result here, we initialized with the FBP image and used a weighting of the penalty which increased by a factor of 2.5 to its final value over the first 20 iterations. The asymptotic value in the plot again exceeds

12

that of the other methods due to the permitted negativity.

In Figure 7 we decompose the information for ICD/FS in Figure 6 concerning the gain in the log *a posteriori* density into the gain in the log-likelihood and the log prior terms. With the standard FBP as the initial condition, we see contribution of both terms in the same range, though far larger change in log-likelihood than in log prior. This difference is exaggerated when we use a second, much poorer initial condition, in this example a uniform image equal to the average of the FBP. This initiation leaves the likelihood much farther from its (MAP) optimal value; on the other hand, the zero-penalty uniform image represents the *most* likely (*a priori*) choice under the MRF model, and is significantly closer to the MAP image than is the FBP. Its log prior term, of course, decreases toward the MAP estimate's value.

Similar overall convergence comparisons for transmission data are visible in Figure 8, where ICD/FS and ICD/NR are indistinguishable. The plots for the two *a priori* models are very similar, due to the dominance of the log-likelihood term in the functional of this high signal-to-noise ratio reconstruction. The pattern in which pixels should be scanned in the sequential methods is not always clear. The analyis in [14] showed potential advantage in alternating raster scans between two orthogonal spatial directions, but we have not found experimentally that this holds consistently. A random ordering of the scan was recently reported in [24] to improve speed. We have applied a pseudo-random scan which visits each pixel once in each sweep, similarly to the choice of Bowsher, et al. and found that in the transmission reconstruction its convergence is significantly faster than the regular scan, as shown in Figure 8. Later in the paper we find that a regular, decimated pattern for parallel updates may achieve similar speed-up. As in the emission head phantom data, we found that PCG performed best with a gradually increasing weighting for the non-negativity enforcing function for the transmission reconstruction, The weighting was increased five-fold over the first 20 iterations to the value recommended in [19], with the image after 40 iterations again having negative pixels which raised its asymptotic objective value slightly above that of the others. In this case its performance was competitive with that of ICD with a regular scan but still significantly slower than the random scan ICD. Lange's Convex algorithm is slower to converge to the MAP reconstructions, similarly to De Pierro's generalized EM for emission. Here again, the use of a substitute functional for all pixels simultaneously appears to limit the rate of convergence.

These results represent solution of the exact Poisson-modeled transmission likelihood. In practice, with photon counts in the range present in this data, it is unlikely we would see appreciable

benefit from solving the exact problem, and would more likely use a single global quadratic approximation, to be optimized sequentially as in [14].

4 Parallel Computation of Pixel Updates via FS

In spite of the rapid convergence of sequential optimization techniques in statistical tomography, the computation of the same sort of greedy updates in parallel is an important generalization, due to the greater economy of parallel computation using multiple processors of modest cost. Parallel implementation of these algorithms in future systems seems likely, and one would hope for the flexibility to assign a segment of the image to each processor [25]. Alternatively, whether or not computation is in parallel, updates of groups of pixels may be advantageous in terms of certain numerical operation counts, as Fessler argued in a derivation of a transmission algorithm similar to the one below which was done concurrently with ours [26]. Parallelism is automatic with gradient-type algorithms. However, accelerated techniques such as conjugate gradient make incorporation of nonnegativity constraints more complicated. We will see that the FS approach can also be applied to allow computation similar to that of the ICD sequential framework for parallel updates of arbitrary sets of pixels in either emission or transmission reconstructions.

In real applications, parallel computation based on shared memory involves issues of communications and timing which may have appreciable effects on the overall speed of the parallel iterations. The results below yield the algorithm to be applied at each pixel involved in the parallel updates, but do not address such architecture-dependent implementation factors, which we leave for future research.

4.1 FS Method Applied in Parallel MAP

De Pierro's substitute function $C(x; x^m)$, discussed in Section 2, allows complete parallel updates for Bayesian tomographic emission reconstructions by replacing the *log-prior* in the cost function with a sum of terms each involving only one pixel update value. The following result amounts to applying the same idea to the more complex *log-likelihood* function.

In the MAP reconstruction problem, with degrees of parallelism which are likely to be practical in our applications, and Markov random field (MRF) *a priori* image models, it is easy to choose parallel updates which are not coupled through $\log \mathcal{P}(X = x)$. We seek a substitute to only the log-likelihood function, $\log \mathcal{P}(Y = y | X = x)$, since through this function large numbers of pixel pairs are coupled.

14

Suppose that we consider the parallel update of a collection of pixels whose indices form the set S during update $n + 1$, with the remainder of the image fixed at x^n. This is conceptually similar to the potential EM update of collections of pixels under SAGE [27]. The set S may be chosen arbitrarily, but here it is intended to be any collection of pixels to be updated in parallel, possibly by separate computing elements. As the analysis below indicates, they would likely be chosen to maximize the feasible distances among pixels in S. A complete iteration consists of a sequence of such sets covering the entire image. We may view the log-likelihood form shown in (3) at one step as a function of only $\{x_j; j \in S\}$. If we define

$$W_{S,i} = \sum_{j \in S} A_{ij},$$

then we may express the dependence on $\{x_j; j \in S\}$ by using the convex function $G_i(\cdot)$, showing dependence on changes in the values of pixels in S. Let

$$\Delta_j^S = \begin{cases} x_j - x_j^n & j \in S \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$
\begin{aligned}
-\log \mathcal{P}(Y = y | X = x^n + \Delta^S) &= \sum_i G_i \left(A_{i*}(x^n + \Delta^S) \right) && (17) \\
&= \sum_i G_i \left(A_{i*}x^n + \sum_{j \in S} \frac{A_{ij} W_{S,i}}{W_{S,i}} (x_j - x_j^n) \right) \\
&= \sum_i G_i \left(\sum_{j \in S} \frac{A_{ij}}{W_{S,i}} \left[A_{i*}x^n + W_{S,i}(x_j - x_j^n) \right] \right).
\end{aligned}
$$

From the definition of $W_{S,i}$ above, it is apparent that $\sum_{j \in S} \frac{A_{ij}}{W_{S,i}} = 1$ also in the limit as $W_{S,i} \to 0$. Applying Jensen's inequality results in the expression

$$-\log \mathcal{P}(Y = y | X = x^n + \Delta^S) \le \sum_{j \in S} \underbrace{\sum_i \frac{A_{ij}}{W_{S,i}} G_i \left(A_{i*}x^n + W_{S,i}(x_j - x_j^n) \right)}_{F_j^S(x_j)}. \qquad (18)$$

This applies to common likelihood functions for tomographic problems. As in the previous section, we will suppress the argument showing the dependence of functions such as F_j^S on x^n. Note that (18) is a summation over S, each term involving only one x_j, which allows for simple optimization.

We define

$$p_{S,i}(x_j) = A_{i*}x^n + W_{S,i}(x_j - x_j^n) + r_i,$$

letting $r_i = 0$ for the transmission case. Substituting into (18) for the standard Poisson models for transmission and emission tomography, the right-hand side of the inequality, which we refer to as the substitute function $F_j^S(x_j)$, takes the form

$$F_j^S(x_j) = \sum_i \frac{A_{ij}}{W_{S,i}} \left[p_{S,i}(x_j) - y_i \log(p_{S,i}(x_j)) \right] \tag{19}$$

for the emission case and

$$F_j^S(x_j) = \sum_i \frac{A_{ij}}{W_{S,i}} \left[y_T \exp\{-p_{S,i}(x_j)\} + y_i p_{S,i}(x_j) \right] \tag{20}$$

for the transmission case.

The substitute log-likelihood function satisfies

$$-\log \mathcal{P}(Y = y | X = x^n) = \sum_{j \in S} F_j^S(x_j^n)$$

and

$$-\log \mathcal{P}(Y = y | X = x^n + \Delta^S) \leq \sum_{j \in S} F_j^S(x_j),$$

similarly to De Pierro's $C(x; x^m)$, and the MAP reconstruction problem with parallel computation is converted into the following optimization problem with respect to the single pixel x_j

$$x_j^{n+1} = \arg \min_{x_j \geq 0} \{ F_j^S(x_j) + P_j(x_j) \}.$$

Any numerical method may then be applied to minimization of the substitute functional. Provided it preserves the monotonicity discussed above and satisfies the Kuhn-Tucker conditions for constrained optimization, guaranteed global convergence is preserved by virtue of the inequality of (18).

Since all pixel updates are decoupled by the substitute functional, ICD methods already developed are natural for this parallel MAP reconstruction problem, locally approximating the F_j^S terms in the fashion of ICD/NR or ICD/FS. Given the similarity in performance of the two, and the results for ICD/NR published in [25], we limit our numerical results here to the provably convergent optimization by parallel ICD/FS. This maintains guaranteed global convergence in the results which follow. The parallel algorithm can still be expressed as in (8-9), but the algorithm is applied to the substitute function for parallel optimization. This leaves θ_1 unchanged from (5) and $\check{\theta}_2$ calculated as in (7), except that derivatives calculated at $x_j = 0$ are those of the substitute functional

with scaling of the second derivative. With "f" once again corresponding to first derivatives of the substitute log-likelihood functions, the resulting forms are

$$f_j^S(x_j) \quad = \quad f_j(x_j) = \sum_i A_{ij} \left(1 - \frac{y_i}{p_i^n} \right) \tag{21}$$

$$f_j^S(0) \quad = \quad \sum_i A_{ij} \left(1 - \frac{y_i}{p_i^n - W_{S,i}x_j^n} \right) \tag{22}$$

$$\left. \frac{df_j^S(x_j)}{dx_j} \right|_{x_j^n, x_j = 0} \quad = \quad \sum_i \left(\frac{A_{ij} W_{S,i}}{(p_i^n)^2} \right) \tag{23}$$

for the emission case and

$$f_j^S(x_j) \quad = \quad f_j(x_j) = \sum_i A_{ij} \left(y_i - y_T \exp[-p_i^n] \right) \tag{24}$$

$$f_j^S(0) \quad = \quad \sum_i A_{ij} \left(y_i - y_T \exp[-p_i^n + W_{S,i}x_j^n] \right) \tag{25}$$

$$\left. \frac{df_j^S(x_j)}{dx_j} \right|_{x_j^n, x_j = 0} \quad = \quad \sum_i A_{ij} W_{S,i} y_T \exp[-p_i^n] \tag{26}$$

for transmission.

The approximate degree of local conservatism resulting from parallel update computation may be seen through a global quadratic approximation of the log-likelihood. The log-likelihood functions in both emission and transmission cases can be approximated by a second-order Taylor series expansion in x

$$\log \mathcal{P}(Y = y | X = x) \approx -1/2(p - Ax)^t D(p - Ax) + c(y), \tag{27}$$

with $p_i = y_i - r_i$ (emission) or $p_i = \log(y_T/y_i)$ (transmission), $c(y)$ a constant relative to x, and D a diagonal matrix with entries being the photon counts $\{y_i\}$ (transmission) or $\{y_i^{-1}\}$ (emission)[10]. This approximation is quite accurate for most common transmission problems, and in both cases allows better understanding of optimization techniques and their convergence behavior.

Under this approximation, we may apply the result of (18) to obtain the following.

$$\sum_i G_i \left(A_{i*}x \right) \leq \sum_{j \in S} \sum_i \frac{A_{ij} D_{ii}}{W_{S,i}} (e_i^n - W_{S,i} \Delta_j^{n+1})^2. \tag{28}$$

where

$$e_i^n \quad = \quad p_i - \sum_j A_{ij} x_j^n$$

$$\Delta_j^{n+1} \quad = \quad x_j - x_j^n$$

17

Here e_i^n is the error state vector in the projection data after update n, and Δ_j^{n+1} is the current change in pixel j. Minimization as a function of Δ_j^{n+1} yields

$$\Delta_j^{n+1} = \frac{\sum_i D_{ii} A_{ij} e_i^n}{\sum_i D_{ii} A_{ij} W_{S,i}} \tag{29}$$

This is the same form as the updates derived for ML estimation by ICD under a global quadratic approximation in [14], except that this formulation calls for under-relaxation of the greedy updates by the factor

$$\frac{\sum_i D_{ii} A_{ij}^2}{\sum_i D_{ii} A_{ij} W_{S,i}},$$

which reduces to the local update of [14] when S contains only one pixel. A plot of this factor as a function of the number of pixels updated in parallel from 128×128 uniformly spaced projections is shown in Figure 9 for the center pixel of a 128×128 reconstruction. Computations for this plot assume the set S is a uniform rectangular sampling of the entire image. The ratio $A_{ij}/W_{S,i}$ is a measure of the linkage of pixel j to others in S through common intersection with projection i. The degree of under-relaxation therefore increases as pixel j moves closer to other members of S, so the set would normally be chosen in a uniform pattern spaced as widely as possible.

The prescribed computation of θ_1 and $\breve{\theta}_2$ can always be made in the transmission problem, but a condition, discussed in Section 3, must be considered for the emission case. For any measurements having $y_i > 0$, the log-likelihood is not defined for projection values less than $-r_i$. With scaling of the change in x_j for each i by $W_{S,i}/A_{ij}$ as required by the parallel update algorithm, we are not guaranteed that the derivative of the substitute function can be evaluated at $x_j = 0$ as directed in the default version of ICD/FS, even if the $\{r_i\}$ are all positive. This case is illustrated in Figure 10. If this occurs, we follow a procedure similar to Strategy 2 in Section 3.3. If the derivative is not evaluable at zero, we need only find a value for x_j between 0 and x_j^n at which the derivative of $F_j^S(x)$ is defined and negative. Reference to Figures 10 and 1 verifies graphically that use of the derivative at this point will maintain monotonicity in descent of the substitute functional for the two possible cases illustrated. Therefore we define

$$\begin{aligned} \Xi_j &= \max_i \left\{ [x_j : p_{S,i}(x_j) = -r_i] \right\} \\ &= \max_i \left[x_j^n - \frac{p_i^n}{W_{S,i}} \right] \end{aligned} \tag{30}$$

the point at which the substitute log-likelihood "blows up" as we approach from the right. (Provided the algorithm has been properly initialized and monotonic descent has been followed, this must

always be a value less than x_j^n.) The second value of the first derivative can then be computed at $\Xi_j^n + \epsilon$, with $0 < \epsilon < x_j^n - \Xi_j^n$ and $f_j^S(\epsilon) < 0$. One may compute ϵ off-line as in Section 3.3 for the substitute functional, and use it or $(\Xi_j^n + x_j^n)/2$, whichever is smaller. For high degrees of parallelism and consequent large values of $W_{S,i}$, we find the pre-computed ϵ to be excessively conservative and simply use $(\Xi_j^n + x_j^n)/2$. Should the derivative be positive at the chosen point, we repeat by recursively halving the distance to Ξ_j^n, which has substantial worst-case computational cost due to re-evaluation of the derivative, but does not arise often enough in practice to substantially affect cost. This procedure preserves guaranteed monotonicity of descent and therefore provable convergence. The projections may be scanned in advance of the computation of the derivatives, or the condition may be checked projection-by-projection. The relative efficiency of the two will depend on the degree of parallelism. In our simulations, this case rarely arose with fewer than 256 parallel updates.

4.2   Parallel Computation Numerical Results

In this section, we solve the same MAP tomography reconstruction problems presented in Section 3 but by parallel computation with the ICD/FS method. The GGMRF prior model again features two choices of the parameter p.

The parallel computation assumes that after each processor has updated its respective pixel, the state of the projection error vector e^n can be updated and shared among all processors. Our results do not include consideration of the costs of inter-processor communication and management of joint access to common memory. We show results for a number of processors (N) ranging from 1 to 2^{12} for the emission trials and up to 2^{14} for transmission. The members of each set S form a uniform rectangular sub-sampling of the image, with successive choices of S shifted in a raster scan to cover the entire image without repetition. This pattern allocates a square block of pixels to each processor. The largest number of parallel updates corresponds in each case to 2×2 pixel blocks. The plots of Figures 11, 12 and 13 show that in terms of iteration counts, there is little change in convergence rates as the degree of parallelism increases. Thus for at least up to 256 pixel updates in parallel in most cases, we achieve essentially linear speed-up in terms of iteration count. The fraction of this gain which is realizable under constraints of hardware implementation remains to be investigated.

The purely sequential updates follow a raster scan of the image in the simulations of Figure 11, 12 and 13, while the parallel updates are spatially distributed in a uniform pattern, corresponding

to a square block of pixels for each processor. Among these decimated patterns, the scan is again regular. The plots suggest a dependence on the order in which pixels are visited as well as the degree of parallelism in computation, just as randomization of the scan pattern showed improved performance in the results of Section 3.4. For modest degrees of parallelism, there is supra-linear speed-up due to the decimated pattern, which shows a potentially useful distribution of updates for sequential optimization as well. The factors $W_{S,i}$ are minimized by maximizing distances among members of S, suggesting that the best pattern of parallel updates would be the one maximizing the mean distance among S, similarly to the decoupling of pixels through spatial pattern selection in [28]. Thus alternatives such as approximately hexagonal patterns may improve on our rectangular placement when the number of processors is large. But the optimal pattern remains an open issue, just as in the purely sequential case.

5 Conclusion

Bayesian tomography can be viewed as optimization of a functional, whether by sequential or parallel update methods. It is natural to use functional substitution with an approximation allowing simpler optimization of each pixel value individually. This paper shows that a newly modified ICD algorithm and its parallel implementation, derived by FS methods using local quadratic approximations, yield provable global convergence and attractive experimental convergence results. The simple coordinate descent methods using quadratic approximations have convergence at least as fast as any other known method, with comparable per-iteration computational cost. It also appears that there is no fundamental barrier to applying ICD techniques to tomographic reconstruction with parallel updates in which blocks of pixels are assigned to each processor, and that high degrees of parallelism are possible with no loss of convergence speed per iteration. Though ICD/FS allows us to guarantee convergence, its performance in iteration counts is nearly identical to that of ICD/NR. This observation, plus the fact that neither an experimental nor a theoretic counter-example to the global convergence of ICD/NR for convex tomographic problems has appeared, make the greater conceptual simplicity of ICD/NR and its slight computational advantage still quite attractive.

6 Appendix: Convergence under Non-Gaussian GGMRF Model

Among the conditions assumed for the functionals to be optimized in [15] is that each $\breve{\Phi}_j(x_j)$ be twice differentiable on $I\!\!R^+$. When neighboring pixels have the same value under the GGMRF model with $p < 2$, this condition is violated in our example, since the second derivative of the

negative log-prior, while always positive, is not bounded. However, the key property of the Hessian subsequently used in the convergence proof is a *lower* bound on its eigenvalues. The positivity of this lower bound, required by Condition 5 of the proof, is guaranteed in our case as in that of [15] by the strict convexity of the functions $\breve{F}_j(x_j)$ and of the log prior term $P_j(x_j)$ in addition to the closedness of the set composed of all possible interim states occupied by x between the initial and final estimates.

These properties are applied in Lemma 2 of [15], where one needs to show, in our problem's notation, that there exists a constant $C > 0$ such that for every n,

$$C(x_j^n - x_j^{n+1})^2 \leq \breve{\Phi}_j(x_j^n) - \breve{\Phi}_j(x_j^{n+1}).$$

This guarantees that as $\breve{\Phi}_j(x_j^n) - \breve{\Phi}_j(x_j^{n+1}) \to 0$, we also have $|x_j^n - x_j^{n+1}| \to 0$. The proof relies on there being a lower bound on the remainder term from a Taylor series expansion of the functional at x^{n+1}, with the expansion evaluated at x^n. Let $H_j(x_j) = \frac{d^2 x_j}{dx_j^2}\breve{\Phi}_j(x_j)$. We need only that

$$\int_0^1 (1-t)H_j((1-t)x_j^{n+1} + tx_j^n)dt \geq C.$$

The log-likelihood portion of this integral is non-negative. Substituting the actual form of the log prior for pixel j, this part of the integral becomes

$$\sum_k b_{jk}\frac{(p-1)}{\sigma^p}\int_0^1 (1-t)|x_k - (1-t)x_j^{n+1} - tx_j^n|^{p-2}dt.$$

The integrand has a strictly positive lower bound on any bounded set with $p > 1$ for the GGMRF, satisfying the inequality, and the integral is well-behaved with the exception of the case $x_k = x_j^{n+1} = x_j^n$, when it goes to $+\infty$. But even in this case the product of the integral and $(x_j^n - x_j^{n+1})^2$ goes to zero, allowing the first equation above to be used as required. Thus the proof of [15] may be applied here with no modification save relaxation of the stated requirement from twice-differentiability to continuous differentiability.

References

[1] L. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. on Medical Imaging*, vol. MI-1, no. 2, pp. 113–122, October 1982.

[2] P. J. Green, "Bayesian reconstruction from emission tomography data using a modified EM algorithm," *IEEE Trans. on Medical Imaging*, vol. 9, no. 1, pp. 84–93, March 1990.

[3] T. Hebert and R. Leahy, "A generalized EM algorithm for 3-d Bayesian reconstruction from Poisson data using Gibbs priors," *IEEE Trans. on Medical Imaging*, vol. 8, no. 2, pp. 194–202, June 1989.

[4] A. D. Pierro, "A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography," *IEEE Trans. on Medical Imaging*, vol. 14, no. 1, pp. 132–137, 1995.

[5] R. Lewitt and G. Muehllehner, "Accelerated iterative reconstruction for positron emission tomography based on the EM algorithm for maximum likelihood estimation," *IEEE Trans. on Medical Imaging*, vol. MI-5, pp. 16–22, March 1986.

[6] L. Kaufman, "Implementing and accelerating the EM algorithm for positron emission tomography," *IEEE Trans. on Medical Imaging*, vol. MI-6, no. 1, pp. 37–51, 1987.

[7] H. Hudson and R. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Trans. on Medical Imaging*, vol. 13, no. 4, pp. 601–609, December 1994.

[8] J. Browne and A. De Pierro, "A row-action alternative to the EM algorithm for maximizing likelihoods in emission tomography," *IEEE Trans. on Medical Imaging*, vol. 15, no. 5, pp. 687–699, October 1996.

[9] J. M. Ollinger, "Maximum likelihood reconstruction in fully 3D PET via the SAGE algorithm," *Proc. of IEEE Nucl. Sci. Symp. and Med. Imaging Conf.*, November 3-9 1996, Anaheim, California.

[10] C. A. Bouman and K. Sauer, "Fast numerical methods for emission and transmission tomographic reconstruction," *Proceedings of the Twenty-seventh Annual Conference on Information Sciences and Systems*, March 1993, The Johns Hopkins University, Baltimore, MD, pp. 611–616.

[11] C. A. Bouman and K. Sauer, "A unified approach to statistical tomography using coordinate descent optimization," *IEEE Trans. on Image Processing*, vol. 5, no. 3, pp. 480–492, March 1996.

[12] J. Besag, "On the statistical analysis of dirty pictures," *Journal of the Royal Statistical Society B*, vol. 48, no. 3, pp. 259–302, 1986.

[13] D. Young, *Iterative Solution of Large Linear Systems*. New York: Academic Press, 1971.

[14] K. Sauer and C. A. Bouman, "A local update strategy for iterative reconstruction from projections," *IEEE Trans. on Signal Processing*, vol. 41, no. 2, February 1993.

[15] J. Fessler and A. Hero, "Penalized maximum-likelihood image reconstruction using space-alternating generalized EM algorithms," *IEEE Trans. on Image Processing*, vol. 4, no. 10, pp. 1417–1429, October 1995.

[16] K. Lange, "An overview of Bayesian methods in image reconstruction," *Proc. of the SPIE Conference on Digital Image Synthesis and Inverse Optics*, vol. SPIE-1351, 1990, San Diego, CA, pp. 270–287.

[17] Y. Vardi, L. Shepp, and L. Kaufman, "A statistical model for poisson emission tomography," *Journal of the American Statistical Association*, vol. 80, no. 389, pp. 8–20, March 1985.

[18] M. B. K. Lange and R. Little, "A theoretical study of some maximum likelihood algorithms for emission and transmission tomography," *IEEE Trans. on Medical Imaging*, vol. MI-6, no. 2, pp. 106–114, June 1987.

[19] E. Ü. Mumcuoğlu, R. Leahy, S. R. Cherry, and Z. Zhou, "Fast gradient-based methods for Bayesian reconstruction of transmission and emission pet images," *IEEE Trans. on Medical Imaging*, vol. 13, no. 4, pp. 687–701, December 1994.

[20] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society B*, vol. 39, no. 1, pp. 1–38, 1977.

[21] D. Boehning and B. G. Lindsay, "Monotonicity of quadratic-approximation algorithms," *Ann. Inst. Statist. Math.*, vol. 40, no. 4, pp. 641–663, 1988.

[22] C. A. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving map estimation," *IEEE Trans. on Image Processing*, vol. 2, pp. 296–310, July 1993.

[23] S. S. Saquib, C. A. Bouman, and K. Sauer, "ML parameter estimation for Markov random fields, with applications to Bayesian tomography," *IEEE Trans. on Image Processing*, vol. 7, pp. 1029–1044, July 1998.

[24] J. Bowsher, M. Smith, J. Peter, and R. Jaszczak, "A comparison of OSEM and ICD for iterative reconstruction of SPECT brain images," *Journal of Nuclear Medicine*, vol. 39, p. 79P, 1998.

[25] K. D. Sauer, S. Borman, and C. A. Bouman, "Parallel computation of sequential pixel updates in statistical tomographic reconstruction," *Proc. of IEEE Int'l Conf. on Image Proc.*, vol. 2, Oct. 22-25 1995, Washington, DC, pp. 93–96.

[26] J. Fessler, E. Ficaro, N. Clinthorne, and K. Lange, "Grouped-coordinate ascent algorithms for penalized-likelihood transmission image reconstruction," *IEEE Trans. on Medical Imaging*, vol. 16, no. 2, pp. 166–175, April 1997.

[27] J. Fessler and A. Hero, "Space-alternating generalized expectation-maximization algorithms," *IEEE Trans. on Acoust. Speech and Signal Proc.*, vol. 42, no. 10, pp. 2664–2677, October 1994.

[28] G. T. Herman, D. Odhner, K. Toennies, and S. Zenios, "A parallelized algorithm for image reconstruction from noisy projections," in *Large Scale Numerical Optimization, Proceedings in Applied Mathematics* (T. Coleman and Y. Li, eds.), vol. 46, pp. 3–21, SIAM Publications, 1990.
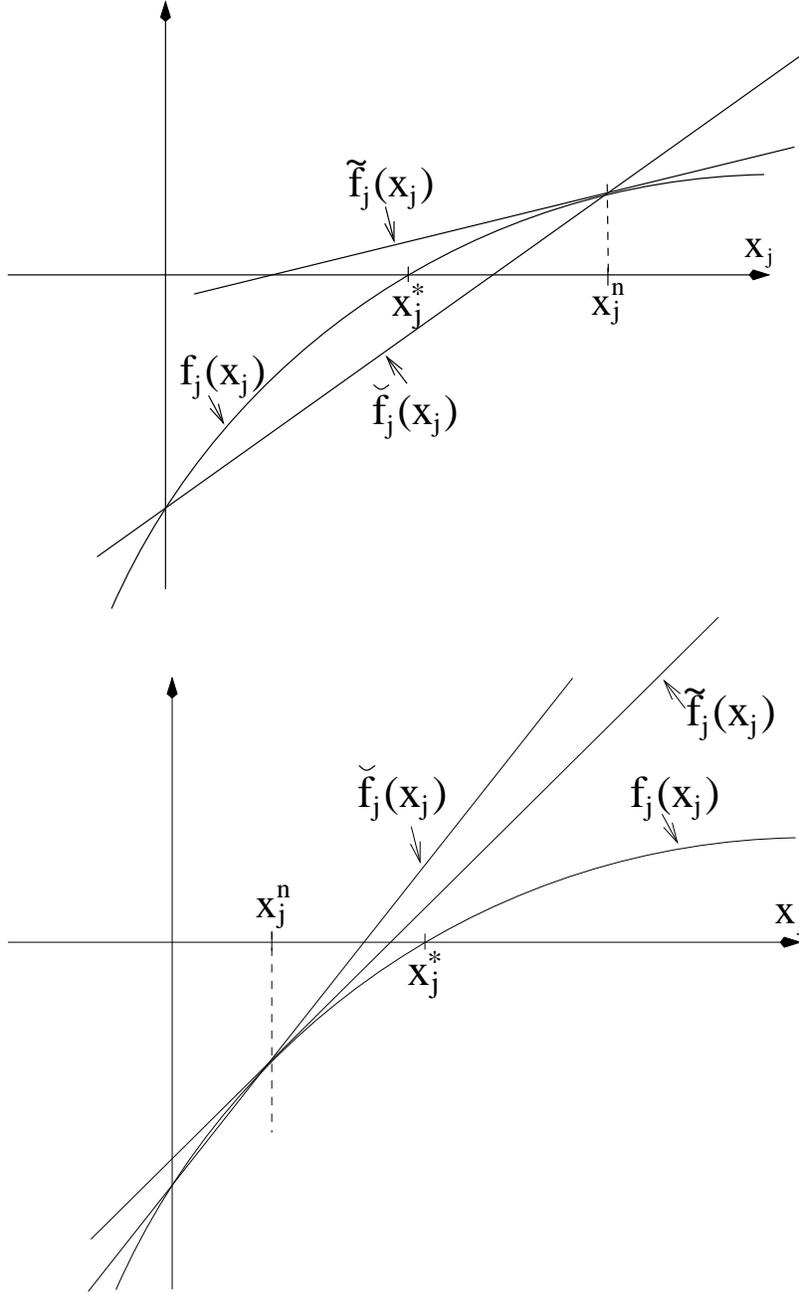
Figure 1: Comparison of ICD/NR and ICD/FS in two possible relations between x_j^n and the currently optimal x_j, marked as x_j^*. Upper sketch shows the derivative of the log-likelihood ($f_j(x_j)$) and the two linear approximations to it for the two techniques when $x_j^n > x_j^*$, and the lower shows $x_j^n < x_j^*$. The function $\tilde{f}_j$ is the approximation of ICD/NR and $\breve{f}_j$ that of ICD/FS. The areas of the regions between $f_j(x_j)$ and $\breve{f}_j(x_j)$ from x_j^n to the zero crossing of $\breve{f}_j(x_j)$ are equal to $(F_j(x_j^n) - F_j(x_j^{n+1})) - (\breve{F}_j(x_j^n) - \breve{F}_j(x_j^{n+1}))$.

a	b
c	d
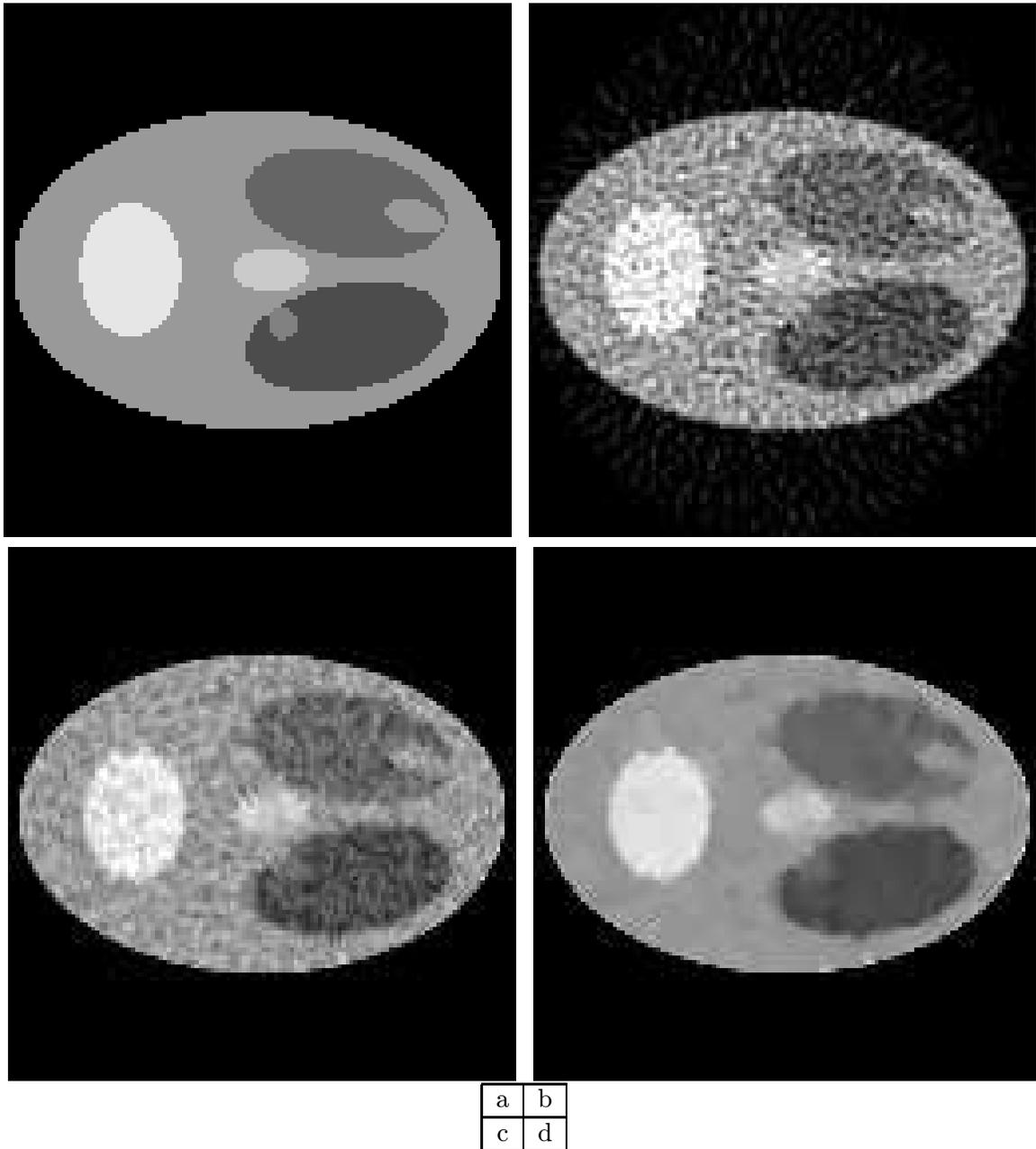
Figure 2: Emission tomographic simulation with synthetic head phantom in $200 \times 200mm$ field and display range is (0.0,4.0). Total photon counts $\approx 3 \times 10^6$. (a) Original head phantom; (b) FBP reconstruction at 128×128 pixels from 128×128 projections; (c) MAP estimate, GGMRF with p = 2.0; (d) MAP estimate, GGMRF with p = 1.1. ML parameter estimates were $\hat{\sigma} = 0.584$ in (c) and $\hat{\sigma} = 0.307$ in (d).
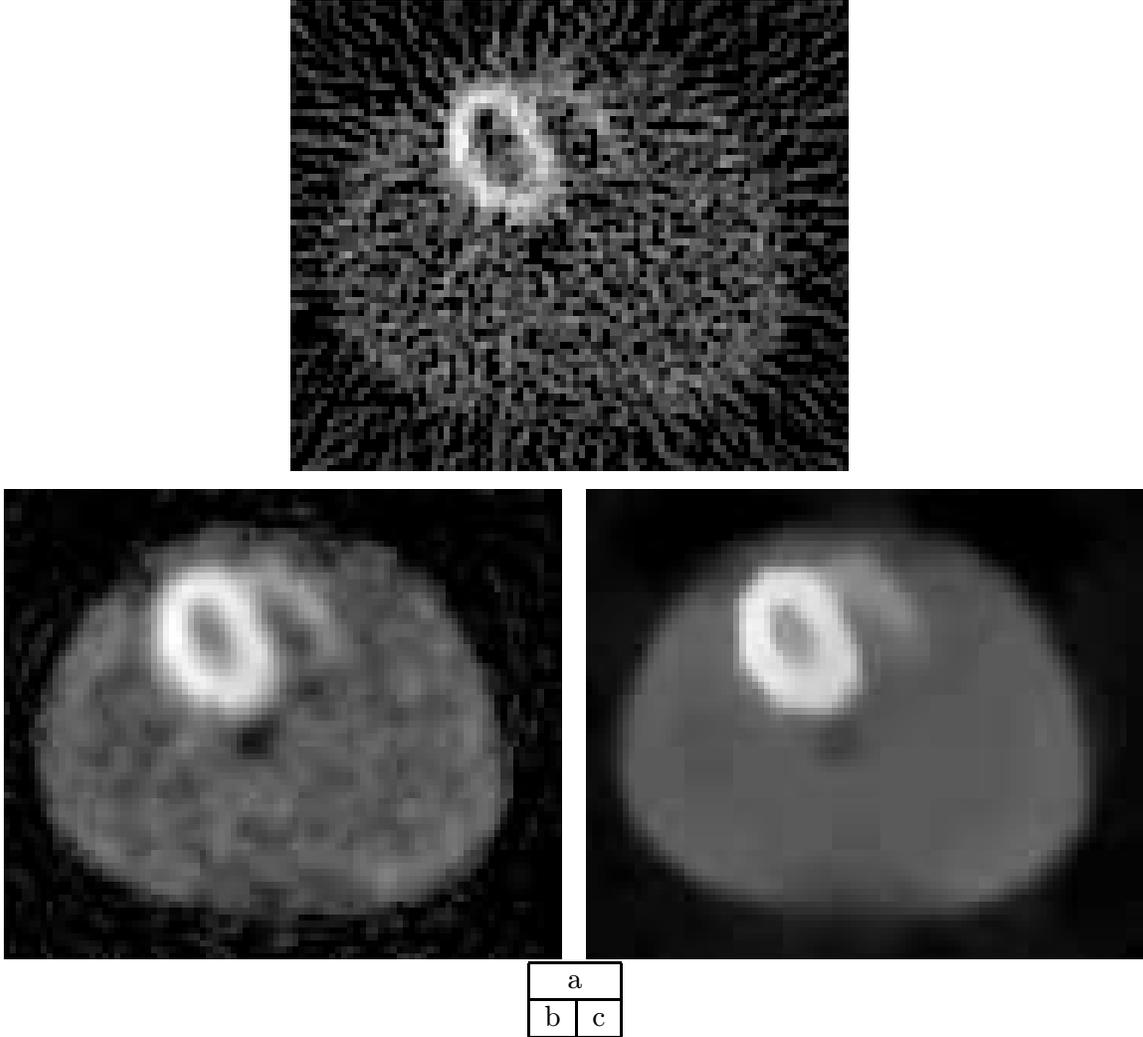
Figure 3: Reconstructions of human thorax from T99 sestamibi heart perfusion SPECT data. Reconstruction covers approximately $320 \times 256mm$, with range of display of (0.0,0.4) for gamma-corrected ($\gamma = 0.6$) emission intensities. Total photon counts $\approx 1.5 \times 10^5$. (a) Filtered back projection; (b) MAP estimate with GGMRF prior and $p = 2.0$; (c) MAP estimate with GGMRF and $p = 1.1$. ML parameter estimates were $\hat{\sigma} = 0.0283$ in (b) and $\hat{\sigma} = 0.0175$ in (c). (Data courtesy of Tin-Su Pan and Michael King, Univ. of Mass.)
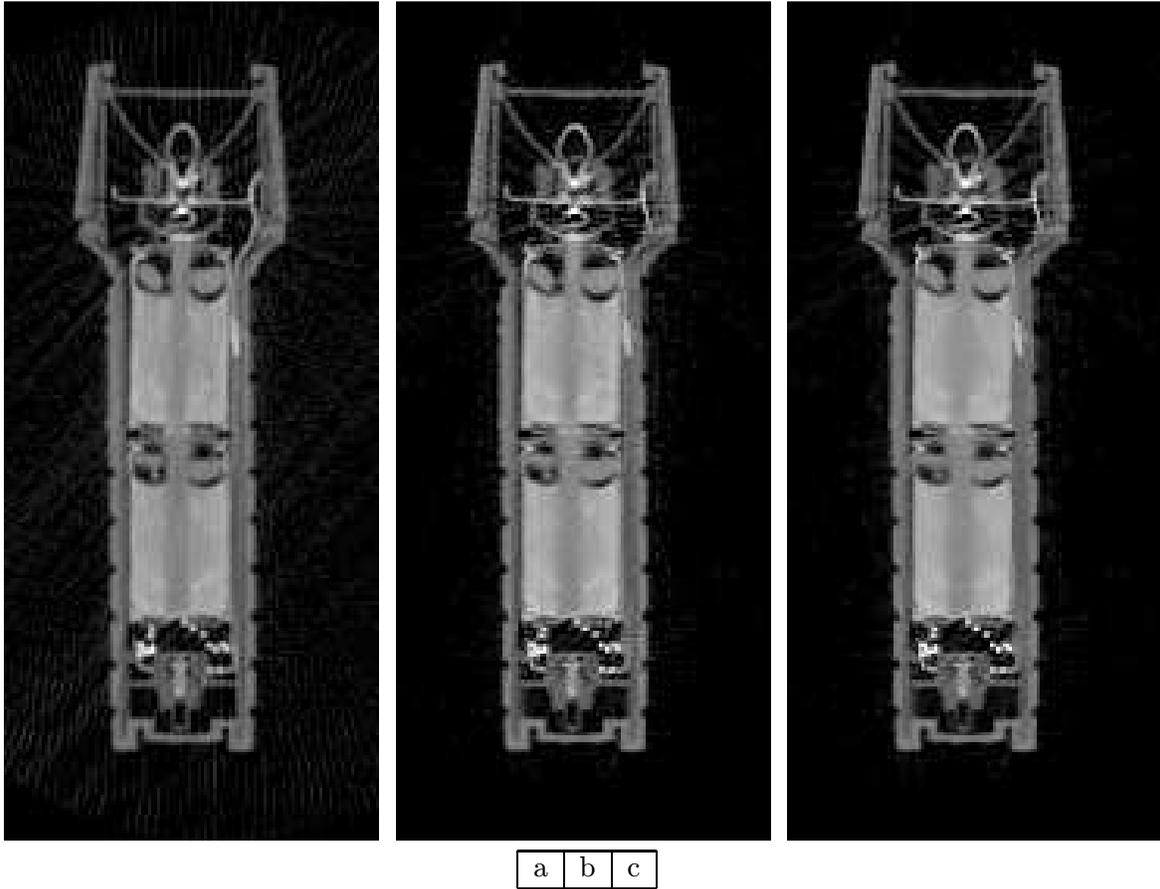
Figure 4: Reconstructions of flashlight cross section from transmission data. Region of reconstruction shown measures $120 \times 260mm$, with range of display $(0.0, 0.1)$ for gamma-corrected $(\gamma = 0.6)$ attenuation values. (a) Filtered backprojection; (b) MAP estimate with GGMRF prior and $p = 2.0$; (c) MAP estimate with GGMRF and $p = 1.1$. ML dosage estimate (y_T) is 2350, $\hat{\sigma} = 0.00866$ in (b) and $\hat{\sigma} = 0.00410$ in (c). (Data courtesy of Trent Neel, Wright-Patterson Air Force Base, and Nicolas Dussausoy, Aracor.)
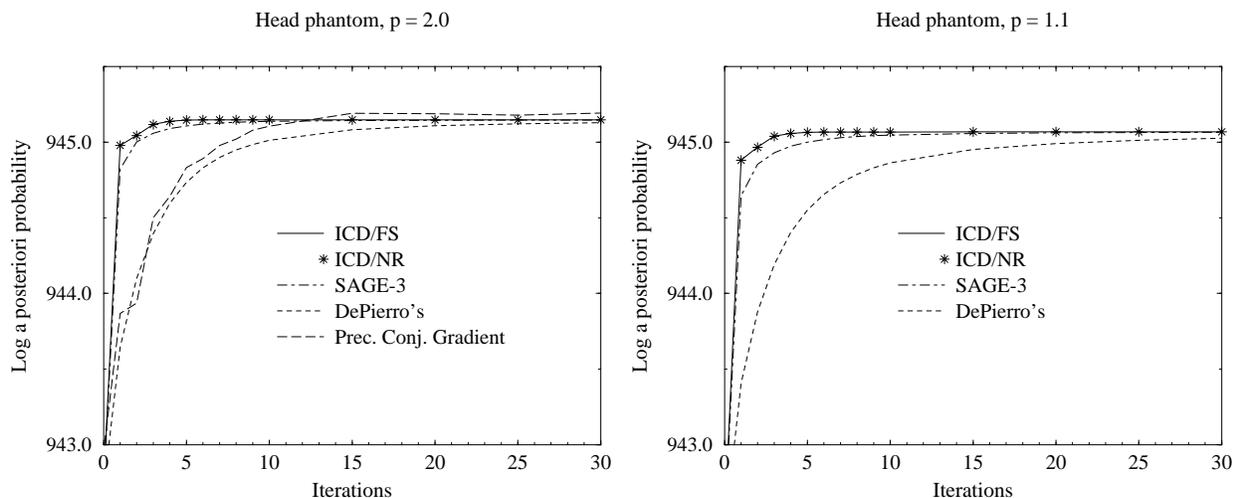
Figure 5: Convergence of objective function vs. iterations for the synthetic head phantom emission case with $p = 2.0$ (left) and $p = 1.1$ (right) GGMRF prior models, employing ICD/FS, ICD/NR, SAGE-3, PCG (for $p = 2.0$) and DePierro algorithms. All initialized with FBP image. Larger values of *a posteriori* probability density for PCG are due to violation of non-negativity constraint.
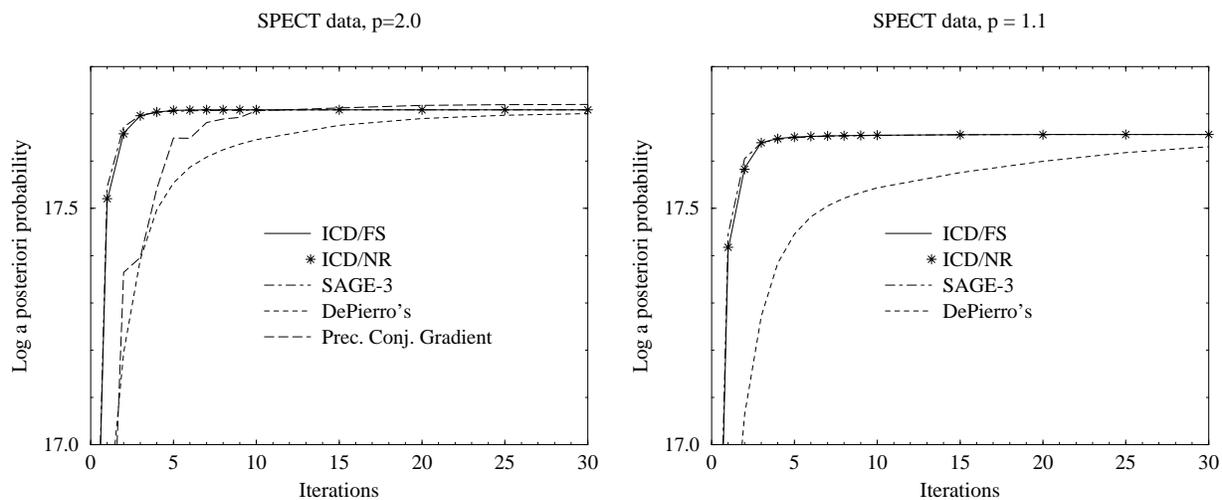
Figure 6: Convergence of objective function vs. iterations for the SPECT cardiac perfusion data with $p = 2.0$ (left) and $p = 1.1$ (right) GGMRF prior models, and ICD/FS, ICD/NR, SAGE-3, PCG (for $p = 2.0$) and DePierro algorithms. All were initialized with FBP image except PCG, for which a uniform initialization yielded faster convergence. Higher asymptotic value of objective for PCG is due to violation of non-negativity constraint.
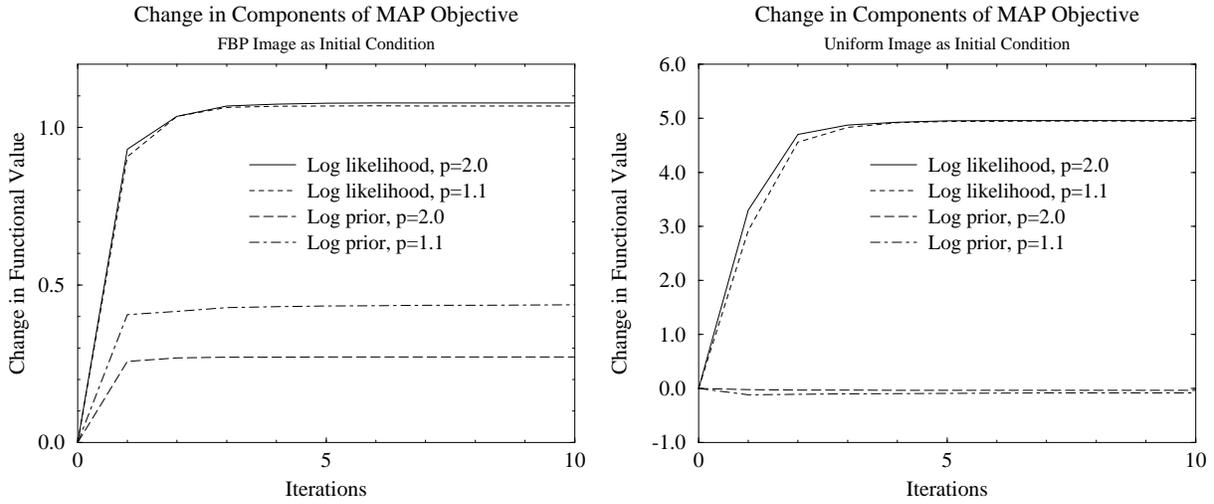
Figure 7: Gain in log *a posteriori* probability for SPECT case, separated into log-likelihood and log-prior terms of objective function, negatives of which are denoted F and P in text. A single iteration includes a sweep across the entire image. Left: filtered backprojection image as initial condition. Right: uniform image, with total mass equal to that of FBP, as initial condition.
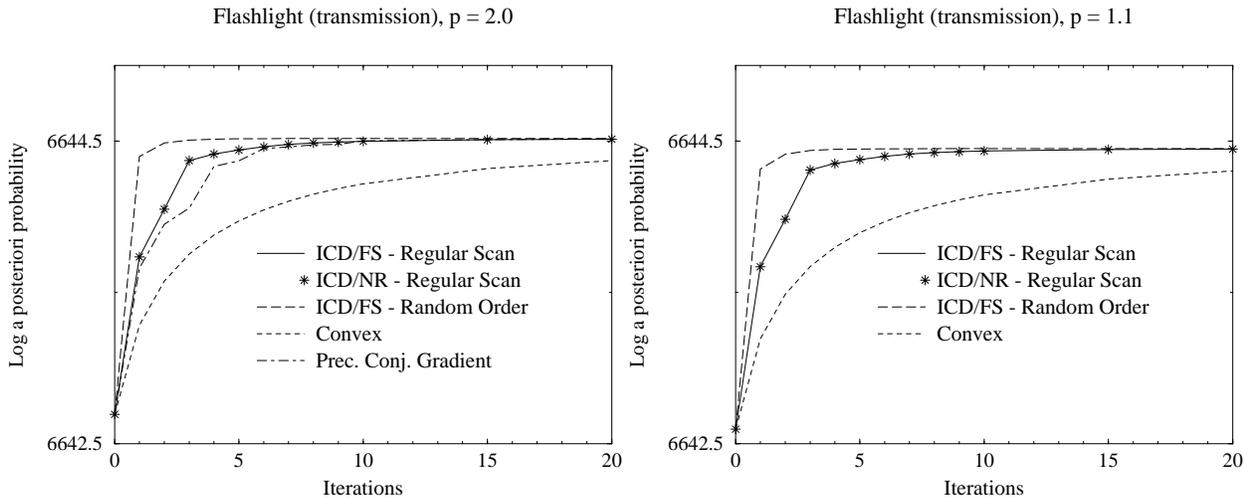

Figure 8: Convergence of objective function vs. iterations for the transmission case with $p = 2.0$ and $p = 1.1$ GGMRF prior model, and ICD/FS, ICD/NR, Convex, and (for $p = 2.0$) preconditioned conjugate gradient (PCG) algorithms. ICD/FS is included with both regular and random scan patterns.

29
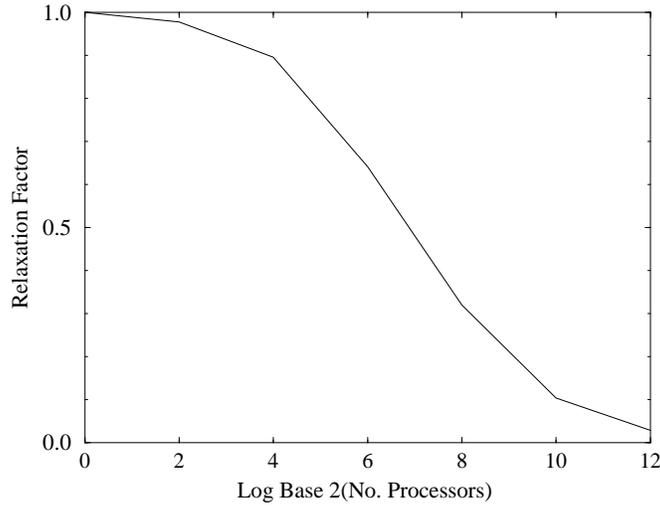
Figure 9: Value of the under-relaxation factor for the quadratic approximation of the log-likelihood. Pixels updated in parallel in all cases form a regular, rectangular sampling of the entire reconstruction. The number of processors ranges from one for the entire image on the left, to one for each 2×2 block of pixels on the right.
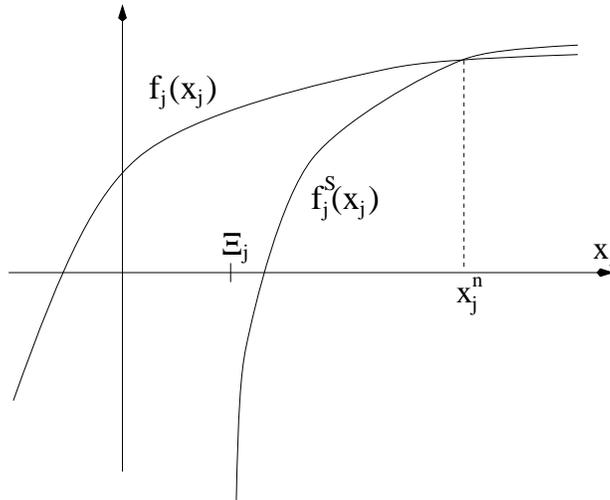

Figure 10: Illustration of an occurrence of $\Xi_j > 0$ under parallel updates. The transformation due to $W_{S,i}/A_{ij}$ is similar to a scaling of the log-likelihood's derivative $f_j(x_j)$ about the point $x_j = x_j^n$.
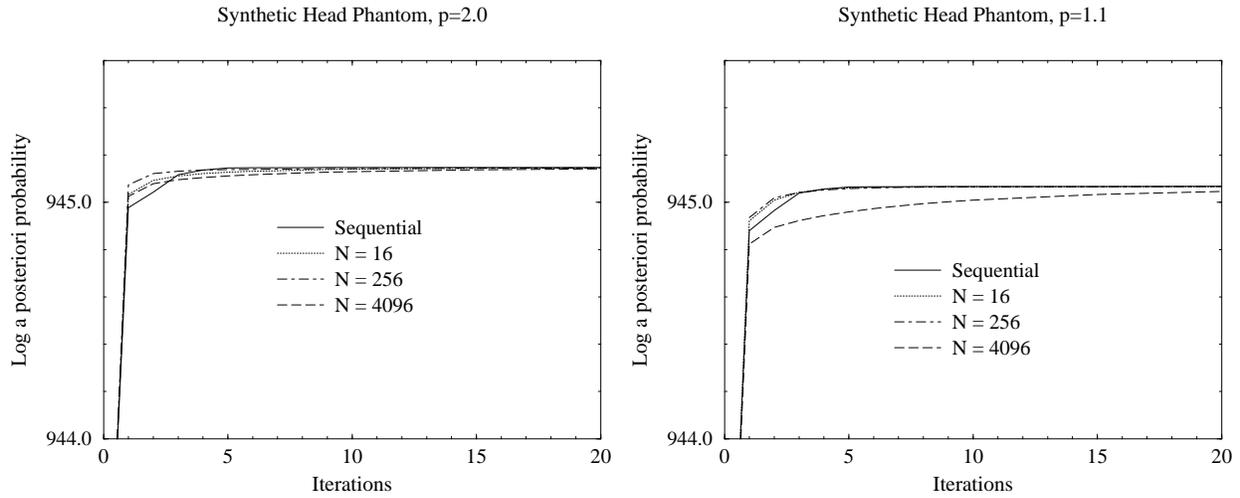
Figure 11: Convergence of objective function vs. iterations for the emission head phantom with $p = 2.0$ (left) and $p = 1.1$ (right) GGMRF prior model, employing ICD/FS algorithm in sequential form and varying degrees of parallelism. In this and subsequent plots, N is the number of pixels updated in parallel.
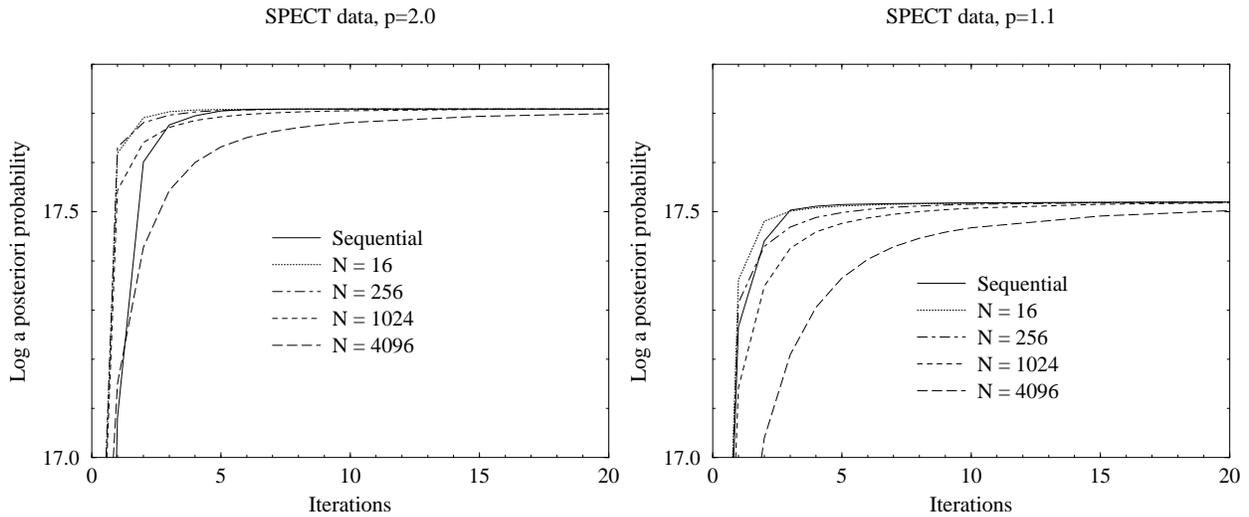

Figure 12: Convergence of objective function vs. iterations for the SPECT heart perfusion data with $p = 2.0$ (left) and $p = 1.1$ (right) GGMRF prior model, employing ICD/FS algorithm in sequential form and varying degrees of parallelism.
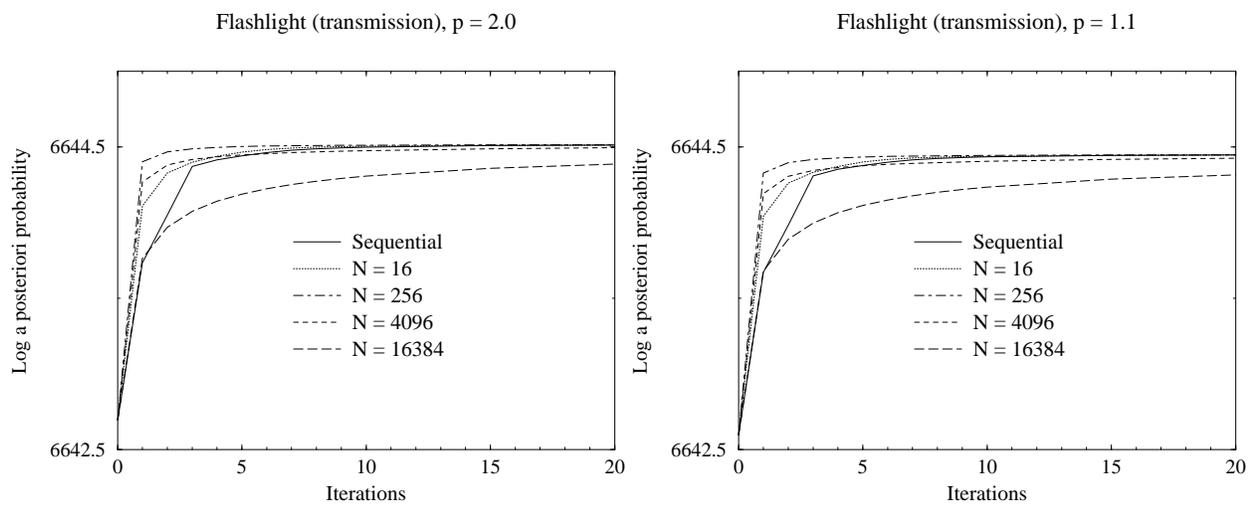
Figure 13: Convergence of objective function vs. iterations for the transmission reconstruction with $p = 2.0$ and $p = 1.1$ in GGMRF prior model, employing ICD/FS algorithm in sequential form and varying degrees of parallelism.