Ordered subsets algorithms for transmission tomography

H Erdoğan and J A Fessler

4415 EECS Building, 1301 Beal Avenue, University of Michigan, Ann Arbor, MI 48109-2122, USA

Abstract. The ordered subsets EM (OSEM) algorithm has enjoyed considerable interest for emission image reconstruction due to its acceleration of the original EM algorithm and ease of programming. The transmission EM reconstruction algorithm converges very slowly and is not used in practice. In this paper, we introduce a simultaneous update algorithm called separable paraboloidal surrogates (SPS) that converges much faster than the transmission EM algorithm. Furthermore, unlike the 'convex algorithm' for transmission tomography, the proposed algorithm is monotonic even with nonzero background counts. We demonstrate that the ordered subsets principle can also be applied to the new SPS algorithm for transmission tomography to accelerate 'convergence', albeit with similar sacrifice of global convergence properties as for OSEM. We implemented and evaluated this ordered subsets transmission (OSTR) algorithm. The results indicate that the OSTR algorithm speeds up the increase in the objective function by roughly the number of subsets in the early iterates when compared to the ordinary SPS algorithm. We compute mean square errors and segmentation errors for different methods and show that OSTR is superior to OSEM applied to the logarithm of the transmission data. However, penalized-likelihood reconstructions yield the best quality images among all other methods tested.

1. Introduction

Attenuation is an important factor that should be corrected for in emission computed tomography. In modern PET and SPECT systems, transmission scans are performed in addition to emission scans to correct for the effects of attenuation. Statistical methods can be used to reconstruct attenuation maps, from which one can calculate attenuation correction factors (ACFs) to yield quantitatively accurate emission images.

Many algorithms exist for maximum likelihood (ML) and penalized likelihood (PL) transmission image reconstruction problems. Most of the recent ones (Sauer and Bouman 1993, Fessler *et al* 1997) are based on direct maximization of the objective function rather than on the famous expectation maximization (EM) algorithm (Lange and Carson 1984) due to the fact that the EM algorithm for transmission reconstruction converges very slowly (Ollinger 1994).

Recently, ordered subsets EM (OSEM) (Hudson and Larkin 1994) for the emission problem has been used extensively in emission image reconstruction, primarily because of the following factors.

- OSEM provides order-of-magnitude acceleration over EM in ML problems.
- The reconstructed images look good after only a few iterations.
- OSEM is implemented by slightly modifying the well known EM algorithm.
- OSEM is easily implemented with any type of system model.

Although the images look good, the resolution and variance properties of OSEM are unclear. In addition it does not converge and may cycle. Due to its popularity, OSEM has even been applied to transmission data after taking its logarithm. In the results section, we show that this approach yields lower quality images than the ordered subsets transmission (OSTR) algorithm that we introduce in this paper.

The ordered subsets principle can be applied to any algorithm which involves a sum over sinogram indices. The sums over all the sinogram indices are replaced by sums over subsets of the data and an ordered subsets version of the algorithm is obtained. However, it seems best to apply this idea to algorithms which update the parameters simultaneously at each iteration rather than to sequential update algorithms. Simultaneous update algorithms take smaller steps in the update direction than sequential update algorithms due to the requirement of a separable surrogate function which has higher curvature than a nonseparable one. Sequential update algorithms such as coordinate descent tend to update high frequencies faster (Sauer and Bouman 1993). When only a subset of the data is used, as in ordered subsets, there is no point in making high frequency details converge. For the algorithms that use only a portion of the data at each iteration such as ART, underrelaxation along the update direction helps the algorithm to converge (Browne and De Pierro 1996).

We introduce a new simultaneous update algorithm called the separable paraboloidal surrogates (SPS) algorithm in this paper. A paraboloidal surrogate (Erdoğan and Fessler 1998b) is a quadratic function that is designed to lie above the negative log-likelihood. Using convexity (Fessler *et al* 1997), we obtain a separable quadratic function that lies above this paraboloid. Another separable surrogate can be obtained for the penalty part by using De Pierro's methods (De Pierro 1993, 1995). The global separable surrogate function can be minimized by a simple simultaneous update.

The SPS algorithm has three advantages as compared to previous simultaneous update algorithms such as the transmission EM algorithm (Lange and Carson 1984) and Lange's convex algorithm (Lange and Fessler 1995): (1) it requires many fewer flops per iteration than the transmission EM algorithm and is comparable to the convex algorithm, (2) SPS is derived for the PL problem which is a more general form than the ML problem, (3) SPS is guaranteed to be monotonic, even with nonzero background events.

The ordered subsets principle has been applied to other transmission ML algorithms. Manglos *et al* (1995) applied the ordered subsets idea to the transmission EM method for the ML problem. Although ordered subsets accelerates the original transmission EM algorithm, it still converges slowly. Nuyts *et al* (1998) tested an ordered subsets version of an approximate simultaneous update algorithm they developed. Their algorithm disregards background counts (such as random coincidences in PET) and the convergence properties are unknown. Kamphuis and Beekman (1998) applied the ordered subsets principle to Lange's convex algorithm to accelerate ML transmission image reconstruction, also ignoring the background counts.

In this paper, we apply the ordered subsets principle to the SPS algorithm for both ML and PL transmission tomography problems. We show that ordered subsets accelerates the initial speed of the original SPS algorithm. However, OSTR is not guaranteed to be monotonic and does not converge to the true optimum for a number of subsets greater than one. Browne and De Pierro (1996) developed a new algorithm called RAMLA which is similar to OSEM with a relaxation parameter incorporated into the algorithm. For a certain class of relaxation parameters, they prove that RAMLA converges to the true ML solution for emission tomography. It might be possible to obtain a convergent version of OSTR by similar means. However, our results show that, even without relaxation, the PL images reconstructed with OSTR are very similar to the ones obtained by convergent algorithms.

In the rest of the paper, we first introduce the problem and the OSTR algorithm for a general penalized-likelihood (PL) objective. Then, we present results on real PET transmission data with ML and PL reconstructions. We analyse the algorithms in terms of their mean squared error. We also perform hard segmentation on the reconstructed images to analyse their tissue classification performance.

2. The problem

For transmission scans, it is realistic to assume the following statistical model if the raw (prompt) measurements $\{y_i\}$ are available:

$$y_i \sim \text{Poisson}\{b_i \, e^{-[\mathbf{A}\mu]_i} + r_i\} \qquad i = 1, \dots, N \tag{1}$$

where N is the number of measured rays, μ_j is the average linear attenuation coefficient in voxel j for $j = 1, \dots, p$, and p denotes the number of voxels. The notation $[\mathbf{A}\mu]_i = \sum_{j=1}^{p} a_{ij}\mu_j$ represents the line integral of the attenuation map μ, and $\mathbf{A} = \{a_{ij}\}$ is the $N \times p$ system matrix. We assume that $\{b_i\}$, $\{r_i\}$ and $\{a_{ij}\}$ are known nonnegative constants, where r_i is the mean number of background events, b_i is the blank scan count and y_i represents the number of coincident transmission events counted by the ith detector pair. Although we adopt PET terminology throughout, the algorithm is also applicable to SPECT and x-ray CT.

For most PET systems, the delayed coincidences are pre-subtracted from true (prompt) coincidences by the device hardware in an attempt to remove the background counts. The subtracted data are no longer Poisson (Fessler 1994, Yavuz and Fessler 1997), but a difference of two Poisson random variables:

$$y_i^s \sim \text{Poisson}\{b_i \, e^{-[\mathbf{A}\mu]_i} + r_i\} - \text{Poisson}\{r_i\}. \tag{2}$$

In this case, the model (1) is not accurate. Yavuz and Fessler (1997) showed that an accurate model is achieved by adding a sinogram which is a good estimate of twice the mean background events (r_i) to the subtracted data and assuming that this resultant random vector has the distribution:

$$y_i^s + 2r_i \sim \text{Poisson}\{b_i \, e^{-[\mathbf{A}\mu]_i} + 2r_i\} \qquad i = 1, \dots, N.$$

This 'shifted Poisson' model matches the mean and variance of the data and yields more accurate images than the PWLS method (Yavuz and Fessler 1997). For the rest of the paper, we focus on the model (1). However extension to the shifted Poisson model can easily be done by replacing y_i by $y_i^s + 2r_i$ and r_i by $2r_i$.

The negative log-likelihood function for the independent transmission data is:

$$-L(\mu) = \sum_{i=1}^{N} h_i([\mathbf{A}\mu]_i) \tag{3}$$

where $h_i(l) = b_i \, e^{-l} + r_i - y_i \log(b_i \, e^{-l} + r_i)$, ignoring the constant terms. Directly minimizing $-L(\mu)$ (ML method) results in a very noisy estimate $\hat{\mu}$. Segmentation of the attenuation map is commonly performed to reduce noise afterwards. Alternatively, penalized-likelihood (PL) (or MAP) methods regularize the problem and reduce the noise by adding a roughness penalty to the objective function as follows:

$$\hat{\mu} = \underset{\mu \geqslant 0}{\text{argmin}} \, \Phi(\mu) \qquad \Phi(\mu) = -L(\mu) + \beta R(\mu).$$

For simplicity we focus here on a roughness penalty R of this form:

$$R(\mu) = \frac{1}{2} \sum_{j=1}^{p} \sum_{k \in \mathcal{N}_j} w_{jk} \psi(\mu_j - \mu_k)$$

where $\mathcal{N}_j$ represents a neighbourhood of voxel j and ψ is a symmetric and convex function that penalizes neighbouring pixel differences. The method easily generalizes to other forms of penalty functions.

In the following discussion, we use the PL formulation to derive the new algorithm. Setting $\beta = 0$ in the following discussion yields the ML estimator.

3. The SPS algorithm

In this section, we describe a new simultaneous update algorithm called the separable paraboloidal surrogates (SPS) algorithm.

3.1. The likelihood part

3.1.1. Nonseparable paraboloidal surrogate. We presented the paraboloidal surrogates algorithm for transmission tomography previously (Erdoğan and Fessler 1998a, b). We first find a one-dimensional surrogate parabola $q_i(l; l_i^n)$ that is tangent to the marginal negative log-likelihood function $h_i(l)$ at the current iterate $l_i^n = [\mathbf{A}\mu^n]_i$ and lies above it for all $l > 0$. Then, we sum up these parabolas as in (3) to obtain an overall (nonseparable) paraboloidal surrogate function for the negative log-likelihood as follows:

$$Q_1(\mu; \mu^n) \triangleq \sum_{i=1}^{N} q_i([\mathbf{A}\mu]_i; l_i^n) \geqslant -L(\mu) \qquad \forall \mu \geqslant 0$$

where

$$q_i(l; l_i^n) \triangleq h_i(l_i^n) + \dot{h}_i(l_i^n)(l - l_i^n) + \tfrac{1}{2}c_i(l_i^n)(l - l_i^n)^2.$$

The optimum curvature that provides the fastest convergence rate while preserving monotonicity was shown to be (Erdoğan and Fessler 1998b)

$$c_i(l_i^n) = \begin{cases} \left[2\dfrac{h_i(0) - h_i(l_i^n) + \dot{h}_i(l_i^n)(l_i^n)}{(l_i^n)^2} \right]_+ & l_i^n > 0 \\[2ex] [\ddot{h}_i(0)]_+ & l_i^n = 0 \end{cases} \tag{4}$$

$$= \begin{cases} \left[(2/(l_i^n)^2) \left\{ b_i(1 - \mathrm{e}^{-l_i^n}) - y_i \log \dfrac{b_i + r_i}{\bar{y}_i^n} + l_i^n b_i \, \mathrm{e}^{-l_i^n} \left(\dfrac{y_i}{\bar{y}_i^n} - 1 \right) \right\} \right]_+ & l_i^n > 0 \\[3ex] \left[b_i \left(1 - \dfrac{y_i r_i}{(b_i + r_i)^2} \right) \right]_+ & l_i^n = 0 \end{cases}$$

where $\bar{y}_i^n = b_i \, \mathrm{e}^{-l_i^n} + r_i$. This surrogate function $Q_1(\mu; \mu^n)$ and each $q_i(l; l_i^n)$ are naturally convex. Previously, we used coordinate descent to minimize this function (Erdoğan and Fessler 1998b). That approach leads to a very fast and monotonic algorithm. However, the computational advantages only exist if the system matrix is precomputed and column accessible (Fessler 1992). For implementations in which the system matrix is not precomputed (e.g. software that uses projector/backprojector subroutines which compute the a_{ij} on the fly), algorithms that update all the parameters simultaneously are preferable since column access to the system matrix is not needed. Moreover, simultaneous update algorithms parallelize more readily than sequential updates. A simultaneous update algorithm can be obtained from the above paraboloidal surrogate by finding another separable surrogate function that lies above it for all possible feasible parameter values μ. The additive separability of the surrogate enables decoupling of the optimization problem for each parameter, so each parameter μ_j can be updated independently from the others. To obtain this separable function we use the convexity tricks employed by De Pierro (1993, 1995).

3.1.2. Separable surrogate. Lange (Lange and Fessler 1995) applied De Pierro's ideas (De Pierro 1993, 1995) to transmission tomography to obtain a separable function that is tangent to the negative log-likelihood and lies above it everywhere when it is convex. It can be based on rewriting the sum

$$[\mathbf{A}\mu]_i = \sum_{j=1}^{p} a_{ij}\mu_j = \sum_{j=1}^{p} \alpha_{ij}\left\{\frac{a_{ij}}{\alpha_{ij}}(\mu_j - \mu_j^n) + [\mathbf{A}\mu^n]_i\right\} \tag{5}$$

where

$$\sum_{j=1}^{p} \alpha_{ij} = 1 \qquad \forall i \text{ and } \alpha_{ij} \geqslant 0. \tag{6}$$

Using the convexity of q_i, we can show that:

$$q_i([\mathbf{A}\mu]_i; l_i^n) \leqslant \sum_{j=1}^{p} \alpha_{ij}q_i\left(\frac{a_{ij}}{\alpha_{ij}}(\mu_j - \mu_j^n) + [\mathbf{A}\mu^n]_i; l_i^n\right). \tag{7}$$

The form of the right hand side of (7) ensures that the function value and gradient of the left hand side are equal to those of the right hand side at the current iterate $\mu = \mu^n$. In other words the curves are tangent at the current iterate. One possible choice for α_{ij} that has been used by De Pierro (1993) and Lange and Fessler (1995) is:

$$\alpha_{ij} = \frac{a_{ij}\mu_j^n}{[\mathbf{A}\mu^n]_i}. \tag{8}$$

We call this choice of α_{ij} the 'multiplicative' form. Using the inequality (7) with these α_{ij}, we obtain a separable surrogate function for $Q_1(\mu; \mu^n)$ as follows:

$$Q_2^{\mathrm{M}}(\mu; \mu^n) = \sum_{i=1}^{N}\sum_{j=1}^{p} \alpha_{ij}q_i\left(\frac{[\mathbf{A}\mu^n]_i\mu_j}{\mu_j^n}; l_i^n\right).$$

This is the separable surrogate obtained using the 'multiplicative' form (8), hence we use *M* in the superscript.

Another possible set of α_{ij} is given by Fessler *et al* (1997):

$$\alpha_{ij} = \frac{a_{ij}}{\gamma_i}$$

where $\gamma_i = \sum_{k=1}^{p} a_{ik}$ is the projection of an image of all ones. We call this choice the 'additive' form, which results in a separable surrogate as follows:

$$Q_2(\mu; \mu^n) = \sum_{i=1}^{N}\sum_{j=1}^{p} \frac{a_{ij}}{\gamma_i}q_i(\gamma_i(\mu_j - \mu_j^n) + [\mathbf{A}\mu^n]_i; l_i^n).$$

The function $Q_2(\mu; \mu^n)$ is separable in j and quadratic, so that the exact minimization is reduced to minimization of p 1D parabolas each of which depend on one pixel value μ_j only.

The separable surrogate obtained from the multiplicative form has some problems with convergence speed due to the fact that the curvature is inversely proportional to the current iterate value μ_j^n:

$$\frac{\partial^2}{\partial^2 \mu_j}Q_2^{\mathrm{M}}(\mu_j^n; \mu_j^n) = \frac{1}{\mu_j^n}\sum_{i=1}^{N} a_{ij}[\mathbf{A}\mu^n]_i c_i(l_i^n).$$

The surrogate parabola becomes infinitely thinner when μ_j^n approaches zero and slows down the convergence for zero regions in the image. The convergence rates of the two algorithms based on multiplicative and additive forms (PL problem with optimum curvature (4)) are shown in figure 1. This figure reveals that the additive form yields a faster algorithm than the multiplicative form does. Hence, we focus on the additive form for the rest of the paper.
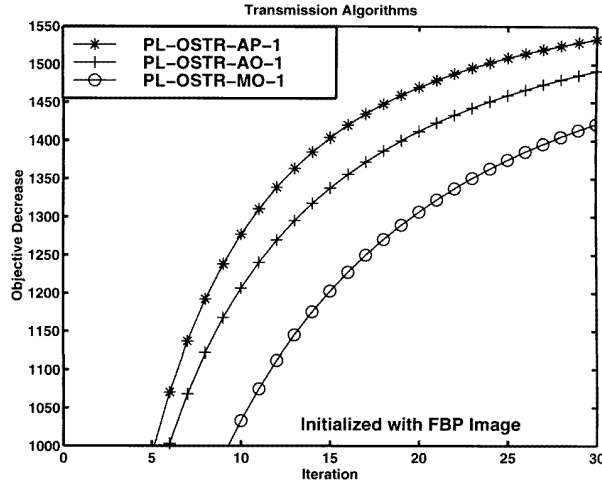
Figure 1. Comparison of additive form with optimum curvature (AO), with precomputed curvature (AP) and multiplicative form with optimum curvature (MO) SPS algorithms for PL image reconstruction.

3.2. The penalty part

Section 3.1 derived separable surrogate functions for the log-likelihood function. A similar separable surrogate is needed for the penalty part $R(\mu)$ to obtain a simultaneous update for the PL objective function. We exploit the convexity of the potential function $\psi(t)$ to obtain the surrogate. For completeness, we repeat the arguments of De Pierro (1995) and Lange and Fessler (1995):

$$
\begin{aligned}
\psi(\mu_j - \mu_k) \quad &= \psi\left(\tfrac{1}{2}[2\mu_j - \mu_j^n - \mu_k^n] + \tfrac{1}{2}[-2\mu_k + \mu_j^n + \mu_k^n]\right) \\
&\leqslant \hat{\psi}_{jk}(\mu; \mu^n) \triangleq \tfrac{1}{2}\psi(2\mu_j - \mu_j^n - \mu_k^n) + \tfrac{1}{2}\psi(2\mu_k - \mu_j^n - \mu_k^n).
\end{aligned} \tag{9}
$$

Using this inequality, one obtains the following separable surrogate function for the penalty:

$$
S(\mu; \mu^n) \triangleq \frac{1}{2}\sum_{j=1}^{p}\sum_{k\in\mathcal{N}_j} w_{jk}\hat{\psi}_{jk}(\mu; \mu^n) \geqslant R(\mu) \qquad \forall \mu \in \mathbb{R}. \tag{10}
$$

One can verify that this surrogate function is tangent to $R(\mu)$ at the current iterate and lies above it for all μ values. Furthermore, the curvature of the surrogate at the current iterate μ^n is exactly twice the curvature of the original penalty function at the same point.

3.3. The SPS algorithm

We designed separable surrogate functions for both the likelihood and the penalty parts in the preceding sections. By combining those, we define the global surrogate function

$$
\phi(\mu; \mu^n) \triangleq Q_2(\mu; \mu^n) + \beta S(\mu; \mu^n)
$$

which satisfies $\phi(\mu; \mu^n) \geqslant -L(\mu) + \beta R(\mu) = \Phi(\mu)$, $\forall \mu \geqslant 0$, and is tangent to $\Phi(\mu)$ at current iterate μ^n, i.e.

$$
\Phi(\mu^n) = \phi(\mu^n; \mu^n) \qquad \nabla\Phi(\mu^n) = \nabla\phi(\mu^n).
$$

We minimize (or decrease) the function $\phi(\mu; \mu^n)$ at each iteration and repeat the procedure iteratively:

$$
\mu^{n+1} = \underset{\mu \geqslant 0}{\operatorname{argmin}}\, \phi(\mu; \mu^n).
$$

We call this algorithm the separable paraboloidal surrogates (SPS) algorithm. One can show (Erdoğan and Fessler 1998b) that decreasing the surrogate function $\phi(\mu; \mu^n)$ also decreases the original objective function $\Phi(\mu)$. Hence, this algorithm is intrinsically monotonic. The minimization of $\phi(\mu; \mu^n)$ is easy. Due to the additive separability, the update for each parameter only involves the parameter itself and μ^n. When a quadratic penalty is used, i.e. $\psi(t) = t^2/2$ and the nonnegativity constraint is ignored, the maximization can be done exactly in a single step via Newton's algorithm as follows:

$$\mu^{n+1} = \mu^n - \mathbf{D}^{-1} \nabla' \Phi(\mu^n) \tag{11}$$

where $\nabla' \Phi(\mu^n)$ is the column gradient of Φ at μ^n and $\mathbf{D}$ is a $p \times p$ diagonal matrix with diagonal entries

$$D_{jj} = d_j^n + 2\beta \sum_k w_{jk} \qquad \text{for } j = 1 \ldots p.$$

The factor 2 in the denominator comes from the curvature of the separable surrogate $S(\mu; \mu^n)$ in (10). The denominator terms d_j^n are:

$$d_j^n = \sum_{i=1}^{N} a_{ij} \gamma_i c_i(l_i^n). \tag{12}$$

For transmission tomography, it is advantageous to use edge-preserving nonquadratic penalties, such as (Lange 1990):

$$\psi(t) = \delta^2 [|t/\delta| - \log(1 + |t/\delta|)] \tag{13}$$

where $\delta > 0$ is predetermined. We used this penalty in our PL reconstruction results.

In the nonquadratic penalty case, exact minimization of $\phi(\mu; \mu^n)$ is not easy, but one can monotonically decrease the surrogate objective by doing one or more of the following Newton–Raphson type subiteration(s):

$$\hat{\mu}_j := \left[\hat{\mu}_j - \frac{\frac{\partial}{\partial \mu_j} \phi(\hat{\mu}; \mu^n)}{d_j^n + 2\beta \sum_{k \in \mathcal{N}_j} w_{jk} \omega_\psi(\hat{\mu}_j - \mu_k^n)} \right]_+ \tag{14}$$

where $\omega_\psi(t) = \dot{\psi}(t)/t$. The detailed explanation of the $\omega_\psi(t)$ function can be found in Erdoğan and Fessler (1998b) and Fessler (1997). The partial derivative of the surrogate ϕ with respect to μ_j can be found as:

$$\frac{\partial}{\partial \mu_j} \phi(\hat{\mu}; \mu^n) = \sum_{i=1}^{N} a_{ij} \dot{h}_i(l_i^n) + d_j^n(\hat{\mu}_j - \mu_j^n) + \beta \sum_{k \in \mathcal{N}_j} w_{jk} \dot{\psi}(\hat{\mu}_j - \mu_k^n) \tag{15}$$

where $\dot{h}_i(l) = (y_i/(b_i e^{-l} + r_i) - 1) b_i e^{-l}$.

Next, we apply the ordered subsets idea to the simultaneous update algorithm developed above.

3.4. Ordered subsets

The ordered subsets principle can be used with any algorithm that involves sums over sinogram indices. The SPS algorithm (14) contains sums over sinogram indices in computing the denominator d_j^n terms (12) and the gradient terms $\frac{\partial}{\partial \mu_j} \phi$ (15). We apply the ordered subsets idea to this algorithm.

Ordered subsets methods group projection data into an ordered sequence of subsets or blocks and processes each block at once. These blocks are usually chosen so that the projections within one block correspond to projections of the image with downsampled projection angles.

It was reported (Hudson and Larkin 1994) that it is best to order the subsets such that the projections corresponding to angles with maximum angular distance from previously used angles are chosen at each step. This accelerates convergence as compared to random or sequential ordering of the subsets. This is due to the fact that the rows of the system matrix corresponding to subsets are chosen to be as orthogonal as possible to previously used subsets.

Let M be the number of subsets chosen in the projection domain. Let $S_1, \ldots, S_M$ denote the subsets in the order selected. At step m the following objective function corresponding to the subset S_m should be minimized (or decreased):

$$\Phi_m(\mu) = M \left\{ \sum_{i \in S_m} h_i([\mathbf{A}\mu]_i) \right\} + \beta R(\mu). \tag{16}$$

The scaling of the negative log-likelihood function ensures that effective β value is independent of the number of subsets. Note that the original objective function can be written in terms of the objective functions (16) as follows:

$$\Phi(\mu) = \sum_{m=1}^{M} \frac{1}{M} \Phi_m(\mu). \tag{17}$$

The success of the ordered subsets methods depends on the following approximation:

$$\Phi(\mu) \approx \Phi_m(\mu) \tag{18}$$

which should be reasonable if the subsets are chosen by subsampling the projection angles.

One iteration is completed when the algorithm cycles through all the projections by using all the subsets. An update performed using a single subset is called a subiteration. The modification of the SPS algorithm to incorporate the ordered subsets idea is relatively easy. We call the resulting algorithm the ordered subsets transmission (OSTR) algorithm. The algorithm outline is shown in table 1.

Table 1. OSTR algorithm outline.

Precompute d_j if possible
for each iteration $n = 1, \ldots, n_{\text{iter}}$
    for each subset $m = 1, \ldots, M$

$$\hat{l}_i = \sum_{j=1}^{p} a_{ij} \hat{\mu}_j \qquad \dot{h}_i = \left(\frac{y_i}{b_i \, e^{-\hat{l}_i} + r_i} - 1 \right) b_i \, e^{-\hat{l}_i} \qquad \forall i \in S_m \tag{19}$$

$$\mu^{\text{old}} = \hat{\mu}$$

    for $j = 1, \ldots, p$

$$\dot{L}_j = M \sum_{i \in S_m} a_{ij} \dot{h}_i \tag{20}$$

$$d_j = M \sum_{i \in S_m} a_{ij} \gamma_i c_i(\hat{l}_i) \tag{21}$$

$$\hat{\mu}_j := \left[\hat{\mu}_j - \frac{\dot{L}_j + \beta \sum_k w_{jk} \dot{\psi}(\hat{\mu}_j - \mu_k^{\text{old}})}{d_j + 2\beta \sum_k w_{jk} \omega_\psi(\hat{\mu}_j - \mu_k^{\text{old}})} \right]_+ \tag{22}$$

    end
    end
end

The form of the update (22) requires the gradient and curvature associated with the penalty term to be computed for each subset. Although the contribution of that computation is only about 4–5% in SPS, it might be costly for a large number of subsets since it is repeated for each subset. Other approaches might be possible such as to consider the penalty function as one of the subsets and update it separately at the end of each cycle (De Pierro and Yamagishi 1998). It might also be possible to break down the penalty term similar to the likelihood part to reduce computation at each iteration. The choice we made in this paper follows naturally from the approximation (16) of the PL objective function. Further investigation is required to reduce this computation.

The OSTR algorithm reduces to the SPS algorithm (14) when $M = 1$. Since the projections and backprojections are performed for only the elements of a single block, processing of each block in an OSTR algorithm with M subsets (OSTR-M) roughly takes $1/M$ of time that it would take for one iteration of the SPS algorithm for the ML problem. For the PL problem, actually it would take more than $1/M$ of the time since the CPU time required for computing the gradient and curvatures of the penalty surrogate at each full iteration is multiplied by the number of subsets. Yet, one hopes that processing of one block increases the objective function as much as one iteration of the original algorithm. That is, the objective increase for M iterations of OSTR-1 should be close to that increase for one full iteration of OSTR-M. This intuition is verified in the initial iterations and for up to a reasonable number of subsets in the results section.

3.5. Precomputed denominator OSTR

We obtained the OSTR algorithm above from a monotonic simultaneous update algorithm. However, the monotonicity is destroyed by doing ordered subsets iterations. So, the algorithm is monotonic only when one subset is used, which is equivalent to SPS.

Since the monotonicity is destroyed at the end anyway, we can remove the condition that the surrogate paraboloid $Q_1(\mu; \mu^n)$ lie above the original objective function and obtain a yet faster algorithm (Erdoğan and Fessler 1998b). Our aim is to precompute the denominator terms d_j^n before iterating and save computation by not updating them. This 'precomputed curvature' idea was introduced by Fessler *et al* (1997) and Erdoğan and Fessler (1998b) for algorithms that used all the data at once unlike the OS algorithms. We can generalize this idea to ordered subsets easily. First, we notice that we can replace the curvature $c_i(l_i^n)$ terms with the Newton curvatures $\ddot{h}_i(l_i^n)$ in (21) and obtain a new algorithm which is no longer guaranteed to be monotonic (even for single subset). We notice that the second derivative of h_i does not change very rapidly and the projections remain very close to the values $l_i^* \triangleq \log(b_i/y_i - r_i)$ during the iterations (which is actually the minimum for $h_i(l)$ over l). So, as a second approximation, we replace $\ddot{h}_i(l_i^n)$ with $\ddot{h}_i(l_i^*)$†. The third approximation is to replace M times the sum of the curvatures $\ddot{h}_i(l_i^*)$ over the subset S_m in (21) with the sum over all sinogram indices $\{1, \ldots, N\}$. This is an accurate approximation if the projections l_i^* vary slowly with respect to the projection angle and each subset is chosen by subsampling the projection angles. So, we can precompute and fix the denominator terms d_j^n by:

$$d_j^n = M \sum_{i \in S_m} a_{ij} \gamma_i \ddot{h}_i(l_i^n)$$

$$\approx d_j^* \triangleq \sum_{i=1}^N a_{ij} \gamma_i \ddot{h}_i \left(\log \left(\frac{b_i}{y_i - r_i} \right) \right) = \sum_{i=1}^N a_{ij} \gamma_i \frac{(y_i - r_i)^2}{y_i}. \tag{23}$$

† l_i^* can only be evaluated when $y_i > r_i$. Otherwise, $h_i(l)$ is a convex function which is monotonically decreasing as $l \to \infty$. When $r_i \geq y_i$, $\lim_{l \to \infty} \ddot{h}_i(l) = 0$, so we replace $\ddot{h}_i(l_i^*)$ with zero or a very small number in that case.

This approximation further reduces CPU time. The minimization step is similar to (14) but the gradient terms in (20) are computed using just a subset of the data. We also found that doing more than one subiteration (14) does not improve 'convergence' and costs a lot computationally in the ordered subsets case. So, we only perform one subiteration to improve 'convergence' and CPU time.

The algorithm looks very simple for the ML problem. The updates are done as follows in ML reconstruction using the fast precomputed denominator:

$$\text{Precompute and store: } d_j^* = \sum_{i=1}^{N} a_{ij} \gamma_i (y_i - r_i)^2 / y_i, \text{ where } \gamma_i = \sum_j a_{ij}$$

for each subset   Compute: $\hat{l}_i, \dot{h}_i$ as in (19) in table 1

Update:

$$\mu_j := \left[\mu_j - \frac{M \sum_{i \in S_m} a_{ij} \dot{h}_i}{d_j^*} \right]_+ . \tag{24}$$

end

This ML-OSTR algorithm is very easy to implement using any kind of system matrix. Precomputation of the d_j^n denominator terms is applicable to the PL problem as well. Figure 1 shows that PL-OSTR with precomputed denominators converges faster than PL-OSTR with optimum curvature. We used this precomputed denominator approach for the results presented next.

4. Phantom data results

We acquired a 15 hour blank scan (b_i) and a 12 min transmission scan data (y_i) using a Siemens/CTI ECAT EXACT 921 PET scanner with rotating rod sources for transmission scans. The phantom used was an anthropomorphic thorax phantom (Data Spectrum, Chapel Hill, NC). Delayed coincidence sinograms were collected separately in each scan. The blank and transmission scan delayed-coincidence sinograms were shown to be numerically close (Yavuz and Fessler 1997), so we used a time-scaled version of blank scan delayed coincidences as the r_i factors with no other processing. The projection space was 160 radial bins and 192 angles, and the reconstructed images were 128×128 with 4.2 mm pixels. The system matrix a_{ij} was computed by using 3.375 mm wide strip integrals with 3.375 mm spacing, which roughly approximates the system geometry.

4.1. Reconstructions

The attenuation map was reconstructed for both ML and PL methods using the OSTR algorithm with 1, 2, 4, 8, 16 and 32 subsets. In all the reconstructions, a precomputed denominator (23) was used. Figure 2 shows the objective function decrease for the ML reconstructions initialized with a uniform image. The order-of-magnitude acceleration can be seen by the horizontal dashed lines in this plot for initial iterations. One iteration of ML-OSTR-16 decreases the objective almost as much as 16 iterations of ML-OSTR-1 and four iterations of ML-OSTR-4 for initial iterations. Although, when $M > 1$, the algorithm does not converge to the true ML solution, in practice one would only do a few iterations using ML-OSTR-M. In the ML problem, exact maximization is not desired since the ML image is extremely noisy.

Figure 3 shows objective function decrease versus iterations for PL reconstructions ($\beta = 2^{10}$ and nonquadratic Lange penalty (13)). The iterations are initialized with an FBP
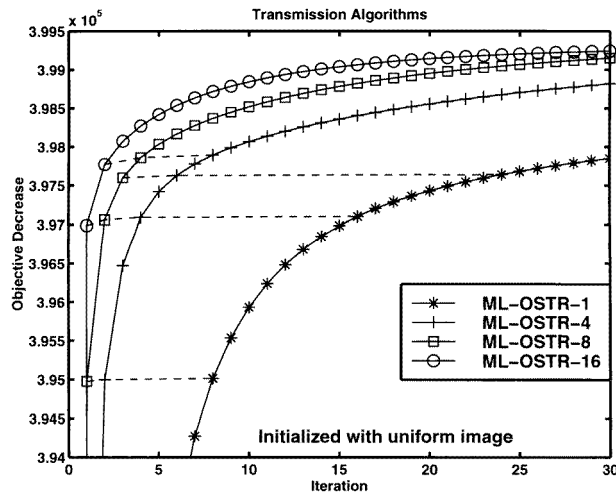
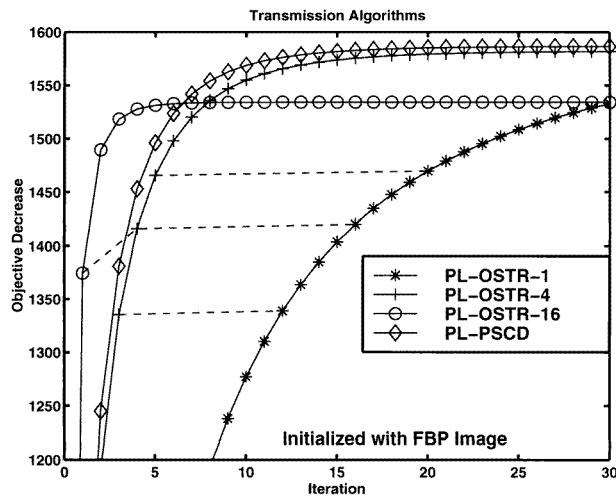**Figure 2.** Maximum likelihood $L(\mu^0) - L(\mu^n)$.

Figure 3. Penalized-likelihood $\Phi(\mu^0) - \Phi(\mu^n)$.

image. There is a speed-up in using more subsets, but as the number of subsets increase, the order-of-magnitude acceleration does not hold. For example, one iteration of PL-OSTR-16 decreases the objective more than one iteration of PL-OSTR-32 (not shown). So, more than 16 subsets did not seem to improve 'convergence' for this configuration and data. For comparison, the image is also reconstructed with the optimum curvature paraboloidal surrogates coordinate descent (PL-PSCD) method which is a fast monotonic algorithm (Erdoğan and Fessler 1998b). The CPU times for one iteration of PL-PSCD and one iteration of PL-OSTR-1 are similar. It is clearly seen that PL-OSTR-M algorithms do not converge to the true minimum when $M > 1$. To assure convergence, one could sequentially decrease the number of subsets with each iteration.

4.2. Mean squared and segmentation errors

The reconstructions were done using real data. We wished to find mean squared errors and segmentation errors on the reconstructed images. The true image of course was unknown. We acquired a long 14 hour scan of the thorax phantom which was almost noise free. We reconstructed the data with FBP with a sharp filter. Then, we performed a four level hard thresholding segmentation on this image with attenuation map parameters assumed to be average standard attenuation map values for air, soft tissue, lungs and bone. We obtained regions for each attenuation level. Then, we eroded these regions with a 5×5 window to obtain more conservative estimates of the regions and calculated the average value of the FBP image in these regions. These new values were assumed to be the true attenuation coefficient levels for the image (air $= 0$, lungs $= 0.035$, soft tissue (water) $= 0.093$, bone (Teflon) $= 0.164$ cm^{-1}). Then, the FBP image was segmented by thresholding using new levels to obtain the 'true' phantom image shown in figure 4.
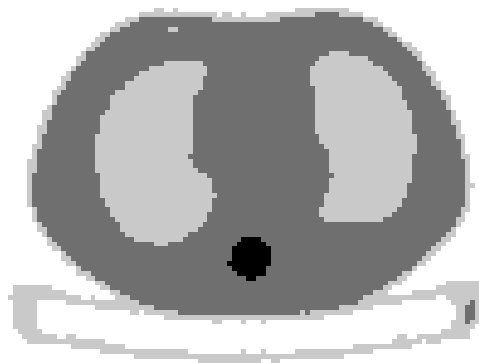

Figure 4. Image obtained by hard segmentation of the FBP image reconstructed from the 14 hour scan assumed as the true image.

We computed normalized mean squared errors (NMSE) for each reconstruction method by comparing to the true phantom image in figure 4. The reconstructed images were also hard segmented with the thresholds found above and we evaluated their segmentation performance by counting the number of misclassified pixels.

We also applied the emission ML-OSEM algorithm to the logarithm of the transmission data $-\log\{(y_i - r_i)/b_i\}$. Although there is no theoretical basis for this approach, it has nevertheless been used by many groups. Our results show that this approach is inferior to the ML-OSTR method and that it should be avoided.

Figure 5 shows NMSE versus iterations for ML-OSTR, ML-OSEM, PL-OSTR and PL-PSCD methods. Figure 6 shows the percentage of segmentation errors versus iterations for the same methods. These results show that ML-OSTR algorithms become noisy after a certain number of iterations and that the iterations should be stopped before convergence. For this transmission scan, the ML-OSTR-16 algorithm should be stopped at the third iteration for lowest NMSE. ML-OSEM applied to the logarithm of the transmission data is inferior in quality to all other methods we tried, regardless of number of subsets. PL reconstructions have better quality than ML reconstructions in terms of both lower mean squared errors and lower segmentation errors. Although the PL-OSTR-16 algorithm does not converge to the minimum of Φ in figure 3, remarkably it appears to be comparable to the convergent PL-PSCD algorithm in terms of NMSE and segmentation performance. In fact, the normalized mean
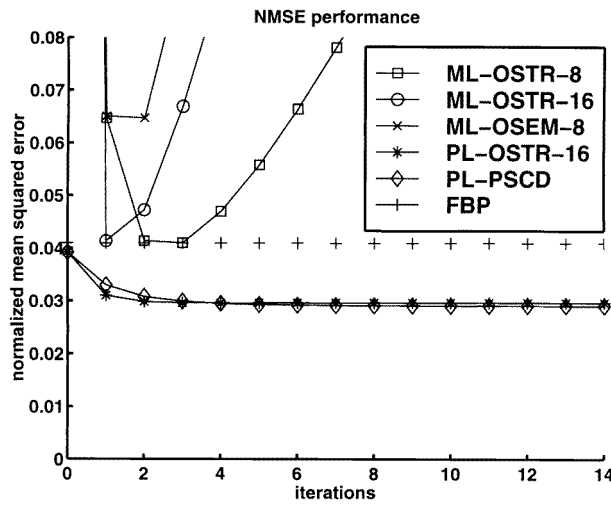
NMSE performance

Figure 5. Normalized mean squared errors versus iterations for various methods of reconstruction.
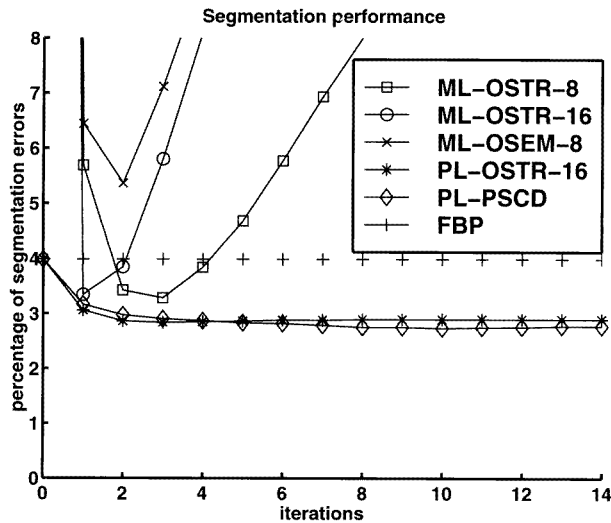
Segmentation performance

Figure 6. Segmentation errors versus iterations for various methods.

squared difference between images reconstructed by PL-PSCD and PL-OSTR-16 is less than 0.015% at iteration 30 of each algorithm.

Figure 7 shows reconstructed images and their segmentations for FBP, ML-OSTR, ML-OSEM, PL-OSTR and PL-PSCD methods. Each image is the best among their kind. For example, to obtain the FBP image, we performed 20 different FBPs with Hanning windows with different cutoff frequencies and picked the one with lowest NMSE. The ML-OSTR image is obtained by 8 subsets at three iterations. The ML-OSEM image is obtained by 8 subsets at two iterations. We used the images obtained at the tenth iteration of PL-PSCD and the fourth iteration of PL-OSTR-16. The bars show the levels of NMSE and segmentation errors. We conclude that PL reconstruction images are much better than the images obtained using other methods.
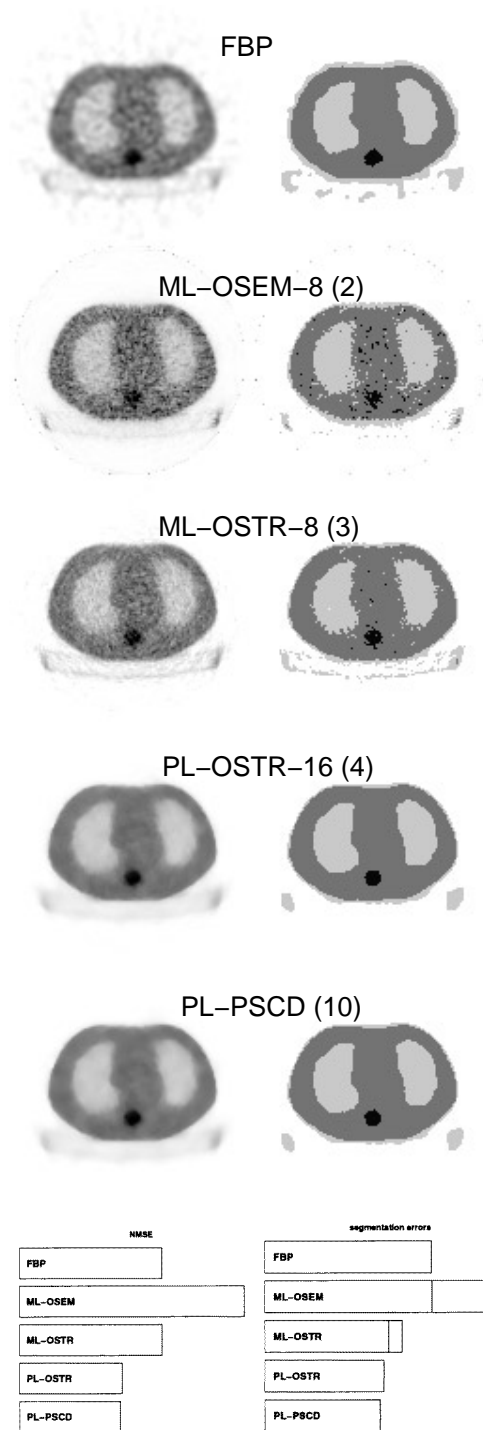
FBP

ML–OSEM–8 (2)

ML–OSTR–8 (3)

PL–OSTR–16 (4)

PL–PSCD (10)

NMSE

| FBP |
| ML-OSEM |
| ML-OSTR |
| PL-OSTR |
| PL-PSCD |

segmentation errors

| FBP |
| ML-OSEM |
| ML-OSTR |
| PL-OSTR |
| PL-PSCD |

Figure 7. Reconstructed (left) and segmented (right) attenuation map images using various methods. The bar plots show the relative NMSE and segmentation error levels. The middle lines in the right hand side bars for ML methods indicate the segmentation error reduction after median filtering.
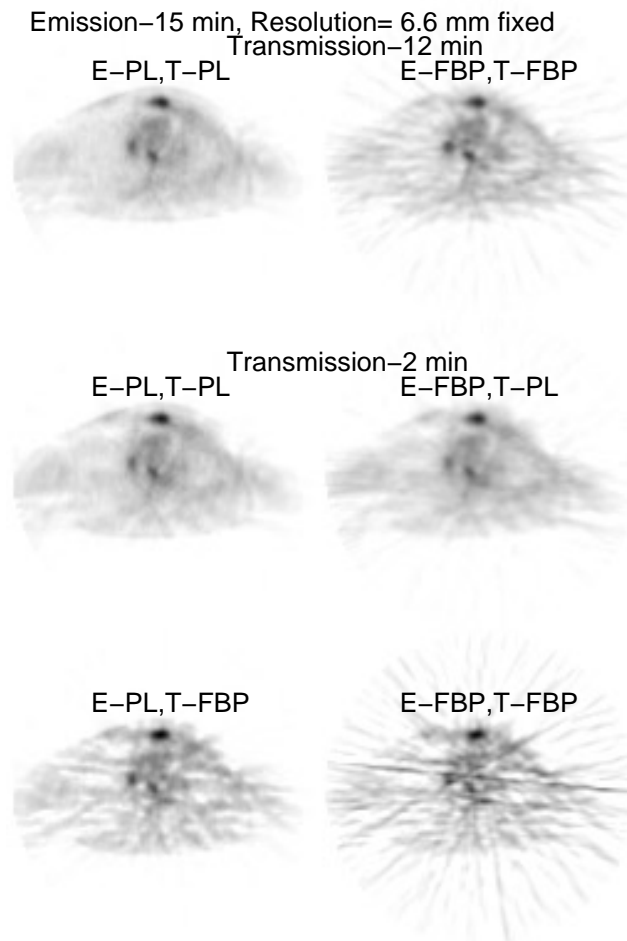
Emission–15 min, Resolution= 6.6 mm fixed
Transmission–12 min

E–PL,T–PL E–FBP,T–FBP

Transmission–2 min

E–PL,T–PL E–FBP,T–PL

E–PL,T–FBP E–FBP,T–FBP

Figure 8. FBP (denoted E-FBP) and quadratically penalized PL (denoted E-PL) emission image reconstructions with attenuation correction factors obtained using conventional (denoted T-FBP) and nonquadratic penalty PL (denoted T-PL) attenuation map reconstructions using the OSTR algorithm. The top row shows emission images obtained using ACFs from a 12 minute transmission scan. The bottom two rows show emission images reconstructed using the ACFs obtained from a 2 minute transmission scan.

5. Patient data results

We applied the new OSTR algorithm to patient transmission and FDG emission data obtained from the ECAT EXACT 921 scanner. We reconstructed emission images using ACFs obtained from the transmission scan. ACFs were computed using two different methods: (1) conventional (or FBP reconstructed and reprojected) and (2) nonquadratic penalty PL-OSTR-16 reconstructions with precomputed denominators and five iterations. Attenuation maps were both post-smoothed axially with the same Gaussian shaped filter with 5 mm FWHM to reduce noise. Emission reconstructions were done with (1) FBP and (2) quadratic penalty PL-SAGE (Fessler and Hero 1994). The resolutions of the emission images were matched at 6.6 mm FWHM.

There was only a 12 minute transmission scan data available. The transmission randoms were pre-subtracted, so we used the shifted Poisson model (2) for the data. The randoms were assumed uniform and the percentages of randoms were estimated from total delayed counts which was available in the file header. To obtain the 2 minute transmission data, we thinned the 12 minute transmission data $y_i^s + 2r_i$ by generating binomial random variables with parameters $n = \max\{0, y_i^s + 2r_i\}$ and $p = 2/12$. Here, n is an approximately Poisson random variable with mean $\bar{n}$. This binomial thinning approach yields a new (approximately) Poisson random variable with a reduced mean of $p\bar{n}$. The 2 minute scan randoms level was adjusted to 2 minutes as well. We also used the shifted Poisson model and uniform randoms estimate for the emission data. Scatter and deadtime effects were ignored.

In figure 8, we show two emission reconstructions in the top row with ACFs obtained from the 12 minute transmission scan. The image obtained from the statistical method shows some improvement in the image quality over the conventional method. The bottom four images are emission images obtained from different combinations of image reconstruction methods with a 2 minute transmission scan. With the 2 minute transmission scan, the improvements in image quality are more apparent for the statistical method as compared to the conventional method as shown in figure 8. These images show that statistical image reconstruction in transmission scans is more important than that in emission scans especially for short transmission scan times.

6. Conclusion

We introduced a new ordered subsets algorithm for ML and PL image reconstruction in transmission tomography. Although the algorithm does not converge for a number of subsets greater than one, it seems to rapidly decrease the objective function value in the early iterations. The images reconstructed from real PET data with the ML method are worse in quality than images reconstructed with the PL method. However, ML-OSTR is superior to ML-OSEM applied to the logarithm of transmission data for these particular data. The new algorithm is easy to implement with any type of system model and does not require column access to the system matrix unlike sequential update algorithms such as coordinate descent. It is also easily parallelizable.

Kudo *et al* (1999) claim that for a general convex objective function, it is possible to obtain convergent ordered subsets algorithms by using appropriate relaxation schemes. The general form of Kudo *et al* (1999) includes the OSTR algorithm as a special case. So it might be possible to obtain convergent algorithms by incorporating a relaxation parameter into the OSTR algorithm.

We conclude that if an approximate minimum is acceptable due to practical time and programming constraints, then the OSTR algorithm offers faster convergence than prior methods. However, for guaranteed global convergence to the minimum, other methods must be used, such as that of Erdoğan and Fessler (1998b).

References

Browne J A and De Pierro A R 1996 A row-action alternative to the EM algorithm for maximizing likelihoods in emission tomography *IEEE Trans. Med. Imaging* **15** 687–99

De Pierro A R 1993 On the relation between the ISRA and the EM algorithm for positron emission tomography *IEEE Trans. Med. Imaging* **12** 328–33

——1995 A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography *IEEE Trans. Med. Imaging* **14** 132–7

De Pierro A R and Yamagishi M E B 1998 Fast EM-like methods for maximum 'a posteriori' estimates in emission tomography (Preprint)

Erdoğan H and Fessler J A 1998a Accelerated monotonic algorithms for transmission tomography *Proc. IEEE Int. Conf. on Image Processing (Chicago, 1998)* vol 2 pp 680–4

——1998b Fast monotonic algorithms for transmission tomography *IEEE Trans. Med. Imaging* at press

Fessler J A 1992 ASPIRE: a sparse, precomputed, iterative reconstruction library

——1994 Penalized weighted least-squares image reconstruction for positron emission tomography *IEEE Trans. Med. Imaging* **13** 290–300

——1997 Grouped coordinate descent algorithms for robust edge-preserving image restoration *Proc. SPIE* **3071** 184–94

Fessler J A, Ficaro E P, Clinthorne N H and Lange K 1997 Grouped-coordinate ascent algorithms for penalized-likelihood transmission image reconstruction *IEEE Trans. Med. Imaging* **16** 166–75

Fessler J A and Hero A O 1994 Space-alternating generalized expectation-maximization algorithm *IEEE Trans. Signal Process.* **42** 2664–77

Hudson H M and Larkin R S 1994 Accelerated image reconstruction using ordered subsets of projection data *IEEE Trans. Med. Imaging* **13** 601–9

Kamphuis C and Beekman F J 1998 Accelerated iterative transmission CT reconstruction using an ordered subsets convex algorithm *IEEE Trans. Med. Imaging* **17** 1101–5

Kudo H, Nakazawa H and Saito T 1999 Convergent block-iterative method for general convex cost functions *Proc. 1999 Int. Meeting on Fully Three Dimensional Image Reconstruction in Radiology and Nuclear Medicine (Amsterdam, 1999)* pp 247–50

Lange K 1990 Convergence of EM image reconstruction algorithms with Gibbs smoothing *IEEE Trans. Med. Imaging* **9** 439–46 (Corrections, June 1991)

Lange K and Carson R 1984 EM reconstruction algorithms for emission and transmission tomography *J. Comput. Assisted Tomogr.* **8** 306–16

Lange K and Fessler J A 1995 Globally convergent algorithms for maximum a posteriori transmission tomography *IEEE Trans. Image Processing* **4** 1430–8

Manglos S H, Gagne G M, Krol A, Thomas F D and Narayanaswamy R 1995 Transmission maximum-likelihood reconstruction with ordered subsets for cone beam CT *Phys. Med. Biol.* **40** 1225–41

Nuyts J, Man B D, Dupont P, Defrise M, Suetens P and Mortelmans L 1998 Iterative reconstruction for helical CT: A simulation study *Phys. Med. Biol.* **43** 729–37

Ollinger J M 1994 Maximum likelihood reconstruction of transmission images in emission computed tomography via the EM algorithm *IEEE Trans. Med. Imaging* **13** 89–101

Sauer K and Bouman C 1993 A local update strategy for iterative reconstruction from projections *IEEE Trans. Signal Process.* **41** 534–48

Yavuz M and Fessler J A 1997 New statistical models for randoms-precorrected PET scans *Information Processing in Medical Imaging (Lecture Notes in Computer Science)* vol 1230 ed J Duncan and G Gindi (Berlin: Springer) pp 190–203