A Globally Convergent Regularized Ordered-Subset EM Algorithm for List-Mode Reconstruction

Parmeshwar Khurd [1], *Student Member, IEEE*, Ing-Tsung Hsiao, *Member, IEEE*[2], Anand Rangarajan, *Member, IEEE*[3] and Gene Gindi, *Member, IEEE.*[1]
[1] Department of Electrical & Computer Engineering, SUNY Stony Brook, NY, USA
[2] School of Medical Technology, Chang Gung University, Kwei-Shan, Tao-Yuan 333, Taiwan
[3] Department of Computer & Information Science and Engineering, University of Florida, Gainesville, FL, USA

***Abstract*— List-mode (LM) acquisition allows collection of data attributes at higher levels of precision than is possible with binned (i.e. histogram-mode) data. Hence it is particularly attractive for low-count data in emission tomography. A LM likelihood and convergent EM algorithm for LM reconstruction was presented in (Parra et al., TMI, v17, 1998). Faster ordered subset (OS) reconstruction algorithms for LM 3-D PET were presented in (Reader et al., Phys. Med. Bio., v43, 1998). However, these OS algorithms are not globally convergent and they also do not include regularization using convex priors which can be beneficial in emission tomographic reconstruction. LM-OSEM algorithms incorporating regularization via inter-iteration filtering were presented in (Levkovitz et al., TMI, v20, 2001), but these are again not globally convergent. Convergent preconditioned conjugate gradient algorithms for spatio-temporal list-mode reconstruction incorporating regularization were presented in (Nichols, et al., TMI, v21, 2002), but these do not use OS for speed-up. In this work, we present a globally convergent and regularized ordered-subset algorithm for LM reconstruction. Our algorithm is derived using an incremental EM approach. We investigated the speed-up of our LM OS algorithm (vs. a non-OS version) for a SPECT simulation, and found that the speed-up was somewhat less than that enjoyed by other OS-type algorithms.**

***Index Terms*— List-mode Reconstruction, Emission Tomography.**

I. Introduction

List-mode (LM) acquisition allows collection of data attributes at higher levels of precision than is possible with binned (i.e. histogram-mode) data [1]. Hence it is particularly attractive for low-count data in emission tomography where the sinogram is large, but sparsely occupied. Another advantage of list-mode acquisition is that event-by-event motion correction can be performed [2], [3], [4], [5], [6]. A particular example is 3D PET [7], where the sinogram consists of a large number of very finely measured lines of response, most of which do not get any counts in low-count imaging conditions. The lists can be more abstract. For example, in [8], each list event comprises the list of signals recorded by photomultiplier tubes (PMT's) in a modular gamma camera along with the camera ID. Reconstruction with this sort of data has been reported in [8].

An Expectation Maximization (EM) algorithm for LM reconstruction was presented in [9]. Faster ordered subset (OS) reconstruction algorithms for LM 3D PET were presented in [7]. However, the algorithms in [7] are not globally convergent and they also do not include regularization using convex priors which can be beneficial in emission tomographic reconstruction. LM-OSEM algorithms incorporating regularization via inter-iteration filtering were presented in [10], but these are again not globally convergent. Convergent preconditioned conjugate gradient algorithms for spatio-temporal list-mode reconstruction incorporating regularization were presented in [11], but these do not use OS for speed-up. Also, note that [11] differs considerably from [9], [7], [10], [8], since it is concerned with dynamic reconstruction wherein the mean photon intensities are reconstructed as a function of time using continuous temporal basis functions. We will not be taking this approach in our paper. Instead, we will attempt to reconstruct a static emission object from detected photon attributes. In this work, we present a globally convergent and regularized ordered-subset algorithm for LM reconstruction. Our algorithm is derived using the incremental EM approach in [12] and [13]. We investigate the speed-up of our LM OS algorithm (vs. a non-OS version) for a SPECT simulation.

Before moving on, we shall make a list of some acronyms that will be frequently used in this paper: LM = List Mode, OS = Ordered Subset, ML = Maximum Likelihood, EM = Expectation Maximization and MAP = Maximum *a Posteriori*. We may also lump these acronyms as needed, to form new acronyms. For example, the LM-EM-ML algorithm stands for the List Mode - Expectation Maximization - Maximum Likelihood algorithm. We can summarize the paper as follows: We show how to extend a derivation of a LM-EM-ML algorithm to a globally convergent LM-OS-EM-MAP algorithm.

There is some latitude in choosing subsets for list-mode reconstruction and this choice is heavily dependent on the modality under consideration. For example, in 3D PET, the first subset may comprise the first 1000 events, the second the next 1000 and so on. On the other hand, in other applications, such as the SPECT reconstruction discussed in section VI, we typically form subsets depending on the angle at which an event occurred.

Each subset corresponds to a sub-iteration and one interpretation of the word "ordering" in OS algorithms is the

* (telephone: 631-444-2539, email:gindi@clio.mil.sunysb.edu)

permutation of the subsets within a main iteration. One can decide upon a particular ordering and different convergence rates may be obtained by choosing different orderings. It has also been found [14] that changing this ordering from one main iteration to another according to specific schemes may lead to fast convergence.

II. The Parra-Barrett Model for EM-ML

In this section, we summarize in our notation the list-mode EM-ML algorithm presented in Parra and Barrett [9]. We shall end up taking a rather different route than the one in [9] in deriving our algorithm, but our summary of the derivation of the algorithm in [9] will prove useful. Please refer to [9] for additional details if needed.

We consider the preset-count(event) form of the LM likelihood. It is shown in [9] that both the preset-time and the preset-count forms of the LM likelihoods lead to the same ML solution. Let E events be detected in time T and let the measured attributes be $\mathbf{A}_1, \mathbf{A}_2, \cdots, \mathbf{A}_E$. An example of an attribute vector (for SPECT) is $\mathbf{A}_{17}$ = (position, gantry angle, energy) for the 17^{th} count. Yet another attribute vector is the vector of PMT signals and camera ID reported in [8]. To deal with this variety in the choice of attribute vectors, it is desirable to express the list-mode likelihood and the list-mode reconstruction algorithms in a form that are independent of the nature of the attribute vector.

Let the object be an N-dimensional vector $\mathbf{f}$, where N is the number of voxels. Then the LM likelihood is:

$$\Pr(\mathbf{A}_1, \cdots, \mathbf{A}_E, T|\mathbf{f}, E) = \Pr(T|\mathbf{f}, E)\Pi_{i=1}^{E} \Pr(\mathbf{A}_i|\mathbf{f}) \quad (1)$$

Here, we have used the fact that the events are independent which implies the independence of the measured attribute vectors for each event. The term $\Pr(\mathbf{A}_i|\mathbf{f})$ can be expanded as:

$$\Pr(\mathbf{A}_i|\mathbf{f}) = \sum_{j=1}^{N} \Pr(\mathbf{A}_i|j) \Pr(j|\mathbf{f}) \quad (2)$$

where $\Pr(\mathbf{A}_i|j)$ denotes the probability that the event i occurs due to an emission at voxel j and $\Pr(j|\mathbf{f})$ is the probability that an emission occurs at voxel j and gets detected. If we denote the sensitivity function of the imaging system at pixel j (i.e. a term proportional to the probability of whether a photon emission at pixel j gets detected at all by the imaging system) by D_j, then:

$$\Pr(j|\mathbf{f}) = \frac{f_j D_j}{\sum_{j'=1}^{N} f_{j'} D_{j'}} \quad (3)$$

The term $\Pr(\mathbf{A}_i|j)$ depends upon the attributes measured. If we use 2D SPECT attributes angle $\theta(i)$ and bin $t(i)$ for event i, then:

$$\Pr(\mathbf{A}_i|j) = \frac{\mathcal{H}_{\{\theta(i),t(i)\},j}}{D_j} \quad (4)$$

where $\mathcal{H}$ is the system matrix and $\mathcal{H}_{\{\theta(i),t(i)\},j}$ is proportional to the probability that a photon from j resulted in list event $\mathbf{A}_i$.

From physical considerations [9], the inter-arrival times for the events are exponentially distributed. We will assume that the mean detected activity per unit time is $\lambda = \sum_{j=1}^{N} f_j D_j$. Then T is the sum of E independent exponentially distributed random variables and hence $\Pr(T|\mathbf{f}, E)$ is given by the following Erlang (Gamma$(E, 1/\lambda)$) density:

$$\Pr(T|\mathbf{f}, E) \;=\; \frac{1}{(E-1)!}\lambda^E T^{E-1} \exp(-\lambda T) \quad (5)$$

The ML estimate of $\mathbf{f}$ can be found by maximizing the LM likelihood in (1). Using (1)-(5) and some algebra, we may write this negative *log*-likelihood (ignoring constants) as:

$$\mathcal{E}_L(\mathbf{f}) = T\sum_{j=1}^{N} f_j D_j - \sum_{i=1}^{E} \ln\Big[\sum_{j=1}^{N} f_j D_j \Pr(\mathbf{A}_i|j)\Big] \quad (6)$$

Direct minimization of (6) while enforcing a positivity constraint on $\mathbf{f}$ is difficult and one can instead derive simple closed-form updates by using a general statistical procedure, the EM algorithm [15]. The EM algorithm is commonly used for reconstruction from binned-mode data in emission tomography [16], [17]. The first step in deriving an EM algorithm is the specification of complete data. By contrast, $\mathbf{A}_1, \cdots, \mathbf{A}_E, T$ is the "incomplete data" and (6) is the incomplete data log-likelihood. There are some subtleties in choosing the complete data and one usually chooses a set of complete data, which when observed, would make the ML estimation of the unknown parameters, i.e. $\mathbf{f}$, rather easy. We define complete data vectors $\mathbf{z}_i$ which indicate the voxel j that caused the event i. The binary components of the vector $\mathbf{z}_i$ are given by $z_{ij}, j = 1, \cdots, N$ and an element z_{ij} is given by:

$$z_{ij} = \begin{cases} 1, & \text{if an emission at voxel j caused event i} \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

The LM complete data likelihood is given by:

$$\begin{aligned} &\Pr(\mathbf{A}_1, \mathbf{z}_1, \cdots, \mathbf{A}_E, \mathbf{z}_E, T|E, \mathbf{f}) \quad (8) \\ &= \Pr(T|\mathbf{f}, E)\Pi_{i=1}^{E} \Pr(\mathbf{A}_i|\mathbf{z}_i, \mathbf{f}) \Pr(\mathbf{z}_i|\mathbf{f}) \end{aligned}$$

Once again, we have used the independence of the detected events in deriving this expression. Note that

$$\Pr(\mathbf{A}_i|\mathbf{z}_i, \mathbf{f}) = \sum_{j=1}^{N} z_{ij} \Pr(\mathbf{A}_i|j) = \prod_{j=1}^{N} \Pr(\mathbf{A}_i|j)^{z_{ij}} \quad (9)$$

and

$$\Pr(\mathbf{z}_i|\mathbf{f}) = \sum_{j=1}^{N} z_{ij} \Pr(j|\mathbf{f}) = \prod_{j=1}^{N} \Pr(j|\mathbf{f})^{z_{ij}} \quad (10)$$

The equalities in (9) and (10) follow from the binary nature of z_{ij} as defined in (7). We may then derive the following expression for the LM complete data *log*-likelihood:

$$\begin{aligned} &\mathcal{E}_{CL}(\mathbf{A}_1, \mathbf{z}_1, \cdots, \mathbf{A}_E, \mathbf{z}_E, T|E, \mathbf{f}) \quad (11) \\ = \quad & \ln \Pr(T|E, \mathbf{f}) + \sum_{i=1}^{E}\sum_{j=1}^{N} z_{ij}\Big(\ln \Pr(j|\mathbf{f}) + \ln \Pr(\mathbf{A}_i|j)\Big) \end{aligned}$$

The E-step in the general EM algorithm [15] now requires the computation of the expectation of the complete data log-likelihood w.r.t. the conditional distribution of the complete data given the incomplete data and the current estimate of the unknown parameters, i.e. the conditional distribution $\Pr(\mathbf{z}_1, \cdots, \mathbf{z}_E | \mathbf{A}_1, \cdots, \mathbf{A}_E, T, \hat{\mathbf{f}}^{(k)})$. Since the events are independent, we may write:

$$\Pr(\mathbf{z}_1, \cdots, \mathbf{z}_E | \mathbf{A}_1, \cdots, \mathbf{A}_E, T, \hat{\mathbf{f}}^{(k)}) = \prod_{i=1}^{E} \Pr(\mathbf{z}_i | \mathbf{A}_i, \hat{\mathbf{f}}^{(k)}) \tag{12}$$

We will use the notation $\Pr(j|\mathbf{A}_i, \hat{\mathbf{f}}^{(k)}) = \Pr(z_{i1} = 0, z_{i2} = 0, \cdots, z_{ij} = 1, \cdots, z_{iN} = 0 | \mathbf{A}_i, \hat{\mathbf{f}}^{(k)})$ to denote the probability that the i^{th} detected event was caused by an emission at voxel j given that the underlying object distribution was $\hat{\mathbf{f}}^{(k)}$. The terms z_{ij} appear in a linear and uncoupled fashion in the complete data log-likelihood in (11) and hence applying the E-step is equivalent to computing the expected value of the terms z_{ij}, yielding the following iterative update for z_{ij}:

$$\begin{aligned}
\bar{z}_{ij}^{(k+1)} &= \Pr(j|\mathbf{A}_i, \hat{\mathbf{f}}^{(k)}) \qquad (13) \\
&= \frac{\Pr(j|\hat{\mathbf{f}}^{(k)}) \Pr(\mathbf{A}_i|j)}{\sum_{j'} \Pr(j'|\hat{\mathbf{f}}^{(k)}) \Pr(\mathbf{A}_i|j')} \\
&= \frac{D_j \hat{f}_j^{(k)} \Pr(\mathbf{A}_i|j)}{\sum_{j'} D_{j'} \hat{f}_{j'}^{(k)} \Pr(\mathbf{A}_i|j')} \;\; \forall i \, \forall j
\end{aligned}$$

where $\bar{z}_{ij}^{(k+1)}$ denotes the expected value of z_{ij} at iteration $k+1$ w.r.t the conditional distribution $\Pr(\mathbf{z}_1, \cdots, \mathbf{z}_E | \mathbf{A}_1, \cdots, \mathbf{A}_E, T, \hat{\mathbf{f}}^{(k)})$.

For EM, the M-step requires maximizing w.r.t the unknown parameters, i.e. $\mathbf{f}$, the expectation of the complete data log-likelihood computed previously in the E-step. This results in the following iterative update:

$$\hat{f}_j^{(k+1)} = \frac{\sum_i \bar{z}_{ij}^{(k+1)}}{T D_j} \;\; \forall j \tag{14}$$

Thus, (13) and (14) constitute the LM-EM-ML algorithm of [9].

III. Re-derivation of LM-EM-ML via Complete Data Energy

We will now re-derive the LM-EM-ML algorithm in the previous section via a totally different approach. This alternative approach will allow us to derive new algorithms in section V that use the OS notion for speedup. The new approach makes use of a "complete data energy" objective $\mathcal{E}_C(\mathbf{z}, \mathbf{f})$ that is a function of $\mathbf{f}$ as well as of $\mathbf{z}$, a variable analogous, as explained below, to the complete data as used conventionally. The key point is that joint minimization on $\mathbf{z}$ and $\mathbf{f}$ yields a solution for $\mathbf{f}$ that is also the solution to $\mathcal{E}_L(\mathbf{f})$. We will show that the LM-EM-ML algorithm specified by (13) and (14) can also be derived by a constrained alternating joint minimization of a complete data energy function. The complete data energy is:

$$\begin{aligned}
\mathcal{E}_C(\mathbf{z}, \mathbf{f}) = -\sum_{i=1}^{E} \sum_{j=1}^{N} z_{ij} \Big(\ln(D_j f_j) \\
+ \ln(\Pr(\mathbf{A}_i|j)) \Big) + \sum_{i=1}^{E} \sum_{j=1}^{N} z_{ij} (\ln z_{ij} - 1) + \\
T \sum_{j=1}^{N} D_j f_j + \sum_{i=1}^{E} \mu_i (-1 + \sum_{j=1}^{N} z_{ij})
\end{aligned} \tag{15}$$

In this minimization, $\mu_i, i = 1, \cdots, E$ are Lagrange multipliers that enforce the constraint $\sum_j z_{ij} = 1$ and $z_{ij} \in [0, 1], i = 1, \cdots, E, j = 1, \cdots, N$ are treated as continuous variables rather than binary variables. The vector $\mathbf{z}$ is the concatenation of the vectors $\mathbf{z}_i, i = 1, \cdots, E$. Though z_{ij} is termed "complete data" here, it is not the same as $\bar{z}_{ij}$ in (13) or the binary z_{ij} in the previous section. Instead, the new z_{ij} is a continuous (within $[0, 1]$) variable that attains the expected value of the (old) binary complete data z_{ij} in the previous section when $\mathcal{E}_C(\mathbf{z}, \mathbf{f})$ is minimized w.r.t. the new z_{ij}. Note that the complete data objective function in (15) differs from the complete data log-likelihood in (11). The LM-EM-ML algorithm can now be derived by a grouped coordinate descent on (15), where we alternately update $\mathbf{z}_i, i = 1, \cdots, E$ and $\mathbf{f}$, while holding the other fixed.

To perform grouped coordinate descent w.r.t $\mathbf{z}_i$, we need the following expression:

$$\frac{\partial \mathcal{E}_C(\mathbf{z}, \mathbf{f})}{\partial z_{ij}} = -\Big(\ln(D_j f_j) + \ln(\Pr(\mathbf{A}_i|j)) \Big) + \ln z_{ij} + \mu_i \tag{16}$$

Equating this expression to zero, we get an update for z_{ij} in terms of the Lagrange multipliers:

$$\bar{z}_{ij}^{(k+1)} = \exp(-\mu_i) D_j f_j^{(k)} \Pr(\mathbf{A}_i|j) \tag{17}$$

The Lagrange multipliers may be eliminated from these expressions by enforcing the constraint $\sum_j z_{ij} = 1$. Enforcing the constraint yields:

$$\exp(-\mu_i) = \frac{1}{\sum_j D_j f_j^{(k)} \Pr(\mathbf{A}_i|j)} \tag{18}$$

Substituting (18) in (17), we obtain the update in (13). To perform grouped coordinate descent w.r.t $\mathbf{f}$, we need the following expression:

$$\frac{\partial \mathcal{E}_C(\mathbf{z}, \mathbf{f})}{\partial f_j} = -\sum_i z_{ij} \frac{1}{f_j} + T D_j \tag{19}$$

Equating this expression to zero, we get the update in (14).

Thus, the justification of (15) is that it yields the desired ML solution for $\mathbf{f}$ and that grouped coordinate descent on (15) re-generates the actual EM-ML equations (13) and (14). We naturally inherit all the convergence properties of the LM-EM-ML algorithm. For an extended discussion of our approach to re-deriving an EM algorithm via a complete data energy function as applied to conventional binned data, see [18], [13].

IV. LM-EM-MAP: Addition of Regularization

Though not done in [9], one could add regularization to this EM algorithm to derive a non-OS list mode MAP algorithm. We start with adding a regularization term to the negative log-likelihood in (6). The regularization term may assume various forms, but here, we will use a quadratic prior term $\beta \sum_{j' \in \mathcal{N}(j)} w_{jj'}(f_j - f_{j'})^2$. Here, $\beta > 0$ controls the amount of regularization and $w_{jj'}$ are neighborhood weights. The term $\mathcal{N}(j)$ is a local neighborhood about j. The weights are symmetrical $w_{jj'} = w_{j'j}$ and positive. For a 2-D problem, a typical neighborhood $\mathcal{N}(j)$ comprises the eight nearest neighbors of j. The new incomplete data objective function, replacing (6), is given by:

$$\mathcal{E}_P(\mathbf{f}) = T\sum_{j=1}^{N} f_j D_j - \sum_{i=1}^{E} \ln\big[\sum_{j=1}^{N} f_j D_j \Pr(\mathbf{A}_i|j)\big] + \beta \sum_j \sum_{j' \in \mathcal{N}(j)} w_{jj'}(f_j - f_{j'})^2 \quad (20)$$

The corresponding new complete data objective function, replacing (15), is given by:

$$\mathcal{E}_{CP}(\mathbf{z}, \mathbf{f}) = -\sum_{i=1}^{E}\sum_{j=1}^{N} z_{ij}\Big(\ln(D_j f_j) + \ln(\Pr(\mathbf{A}_i|j))\Big) + \sum_{i=1}^{E}\sum_{j=1}^{N} z_{ij}(\ln z_{ij} - 1) + T\sum_{j=1}^{N} D_j f_j + \beta \sum_j \sum_{j' \in \mathcal{N}(j)} w_{jj'}(f_j - f_{j'})^2 + \sum_{i=1}^{E}\mu_i(-1 + \sum_{j=1}^{N} z_{ij}) \quad (21)$$

Note that the updates in (13) and (14) are parallel, but the coupling between pixels f_j, f_j' introduced by the regularization term in (21) apparently destroys this parallel nature of the update. We can recover a parallel update that incorporates regularization by using the separable surrogates idea [19]. Here, we seek to derive a surrogate function $\mathcal{E}_{CP,surr}(\mathbf{z}, \mathbf{f}; \hat{\mathbf{f}}^{(k)})$ in order to replace $\mathcal{E}_{CP}(\mathbf{z}, \mathbf{f})$ at each iteration k. We choose surrogate objective functions that are easier to minimize than $\mathcal{E}_{CP}(\mathbf{z}, \mathbf{f})$ and whose iterative minimization leads to the same final solution. We desire that this surrogate function be separable in pixel space, i.e. $\mathcal{E}_{CP,surr}(\mathbf{z}, \mathbf{f}; \hat{\mathbf{f}}^{(k)}) = \sum_j \mathcal{E}_{CP,surr}(\mathbf{z}, f_j; \hat{\mathbf{f}}^{(k)})$ and hence lead to parallel updates. The surrogate function needs to satisfy three properties [20] in order to obtain a convergent algorithm:

1) $\mathcal{E}_{CP,surr}(\mathbf{z}, \hat{\mathbf{f}}^{(k)}; \hat{\mathbf{f}}^{(k)}) = \mathcal{E}_{CP}(\mathbf{z}, \hat{\mathbf{f}}^{(k)})$,
2) $\nabla_{\mathbf{f}}\mathcal{E}_{CP,surr}(\mathbf{z}, \mathbf{f}; \hat{\mathbf{f}}^{(k)}) = \nabla_{\mathbf{f}}\mathcal{E}_{CP}(\mathbf{z}, \mathbf{f})$ at $\mathbf{f} = \hat{\mathbf{f}}^{(k)}$ and
3) $\mathcal{E}_{CP,surr}(\mathbf{z}, \mathbf{f}; \hat{\mathbf{f}}^{(k)}) \geq \mathcal{E}_{CP}(\mathbf{z}, \mathbf{f})$.

For convex priors, De Pierro [19] has developed an inequality that leads to separable surrogate functions with the desired properties. This inequality as it applies to our quadratic prior is given by:

$$(f_j - f_{j'})^2 \leq \frac{1}{2}(2f_j - \hat{f}_{j'}^{(k)} - \hat{f}_j^{(k)})^2 + \frac{1}{2}(2f_{j'} - \hat{f}_{j'}^{(k)} - \hat{f}_j^{(k)})^2 \quad (22)$$

A separable surrogate function may now be derived. It is given by:

$$\mathcal{E}_{CP,surr}(\mathbf{z}, \mathbf{f}; \hat{\mathbf{f}}^{(k)}) = -\sum_{i=1}^{E}\sum_{j=1}^{N} z_{ij}\Big(\ln(D_j f_j) + \ln(\Pr(\mathbf{A}_i|j))\Big) + \sum_{i=1}^{E}\sum_{j=1}^{N} z_{ij}(\ln z_{ij} - 1) + T\sum_{j=1}^{N} D_j f_j + \beta \sum_j \sum_{j' \in \mathcal{N}(j)} w_{jj'}(2f_j - \hat{f}_j^{(k)} - \hat{f}_{j'}^{(k)})^2 + \sum_{i=1}^{E}\mu_i(-1 + \sum_{j=1}^{N} z_{ij}) \quad (23)$$

In deriving (23), we used the fact that $w_{jj'} = w_{j'j}$. Note that this surrogate function is separable in pixel space and so we can separately optimize for each pixel f_j. The surrogate function for pixel f_j is given below:

$$\mathcal{E}_{CP,surr}(\mathbf{z}, f_j; \hat{\mathbf{f}}^{(k)}) = T f_j D_j - \sum_{i=1}^{E} z_{ij}\ln(D_j f_j) + \beta \sum_{j' \in \mathcal{N}(j)} w_{jj'}(2f_j - \hat{f}_j^{(k)} - \hat{f}_{j'}^{(k)})^2 \quad (24)$$

Differentiating this function leads to a quadratic equation resulting in the following M-update:

$$\hat{f}_j^{(k+1)} = \frac{-B + \sqrt{B^2 - 4AC}}{2A} \quad \forall j \quad (25)$$

where $A = 8\beta \sum_{j' \in \mathcal{N}(j)} w_{jj'}$, $B = TD_j - 4\beta \sum_{j' \in \mathcal{N}(j)} w_{jj'}(\hat{f}_j^{(k)} + \hat{f}_{j'}^{(k)})$, $C = -\sum_{i=1}^{E} \bar{z}_{ij}^{(k+1)}$. The update (13) remains unchanged. Thus, (13) and (25) comprise a LM-EM-MAP algorithm. We can establish [13], [21] that the fixed point of (20) is unique and that a constrained joint minimization of the complete data objective function (21) always yields a fixed point of the original incomplete data log-posterior in (20) . Since our LM-EM-MAP algorithm converges to a fixed point of (21), which is also a fixed point of (20), this algorithm converges to the MAP solution. Note that the update equations for f_j automatically preserve the positivity of f_j.

V. List-mode Ordered Subsets EM-MAP Algorithm

Now, we are in position to derive our final goal, an ordered subset regularized version of the LM-EM-ML algorithm in (13) and (14) by using the approach in [13] or [12], [22]. However, in order to more clearly explain the incorporation of the OS notion into the optimization, we will first deal with the *unregularized* case and derive an LM-OS-EM-ML algorithm.

We will now modify the method of minimization of the complete data energy function to a different form of grouped coordinate descent, which uses the notion of ordered-subsets to obtain faster convergence. Suppose that the detected events are divided into subsets $S_l, l = 1, \cdots, L$. We can re-state

(15) in an OS representation simply by replacing $\sum_{i=1}^{E}$ with $\sum_{l=1}^{L}\sum_{i\in S_l}$:

$$\mathcal{E}_C(\mathbf{z},\mathbf{f}) = -\sum_{l=1}^{L}\sum_{i\in S_l}\sum_{j=1}^{N} z_{ij}\Big(\ln(D_j f_j) + \ln(\Pr(\mathbf{A}_i|j))\Big) + \sum_{l=1}^{L}\sum_{i\in S_l}\sum_{j=1}^{N} z_{ij}(\ln z_{ij} - 1) + T\sum_{j=1}^{N} D_j f_j + \sum_{l=1}^{L}\sum_{i\in S_l}\mu_i(-1+\sum_{j=1}^{N} z_{ij}) \tag{26}$$

Now, we will alternately update $\mathbf{z}_i, i\in S_l$ and $\mathbf{f}$ in each (k,l) iteration. Using the expressions (16) and (19), and following similar steps as before, we can obtain the following incremental EM algorithm:

$$\bar{z}_{ij}^{(k,l)} = \frac{D_j \hat{f}_j^{(k,l-1)}\Pr(\mathbf{A}_i|j)}{\sum_{j'} D_{j'}\hat{f}_{j'}^{(k,l-1)}\Pr(\mathbf{A}_i|j')}\ \forall i\in S_l\,\forall j \tag{27}$$

$$\hat{f}_j^{(k,l)} = \frac{\sum_{i=1}^{E}\bar{z}_{ij}^{(k,l)}}{TD_j}\ \forall j \tag{28}$$

Details of a convergence proof for the LM-OS-EM-ML algorithm given by (27) and (28) can be found in [21].

A MAP (regularized) version of (27) and (28) can now be obtained by performing a constrained joint minimization on the complete data objective function (21) which incorporates a regularization term. However, now we re-express (21) in OS notation by replacing $\sum_{i=1}^{E}$ with $\sum_{l=1}^{L}\sum_{i\in S_l}$. We can once again derive a separable surrogate objective function using a modified version of (22) and proceed to use the same grouped coordinate descent strategy that was used to derive the updates in (27) and (28). Equation (22) would have to be modified by replacing $\hat{f}_j^{(k)}$ by $\hat{f}_j^{(k,l-1)}$. The update equations for z_{ij} will remain the same as in (27). We may write the modified form of the M-step (28) as follows:

$$\hat{f}_j^{(k,l)} = \frac{-B+\sqrt{B^2-4AC}}{2A}\ \forall j \tag{29}$$

where now $A = 8\beta\sum_{j'\in\mathcal{N}(j)} w_{jj'}$, $B = TD_j - 4\beta\sum_{j'\in\mathcal{N}(j)} w_{jj'}(\hat{f}_j^{(k,l-1)} + \hat{f}_{j'}^{(k,l-1)})$, $C = -\sum_{i=1}^{E}\bar{z}_{ij}^{(k,l)}$.

Thus (27) and (29) comprise our final goal, a convergent LM-OS-EM-MAP algorithm. Since (27) and (29) minimize (21), they also minimize (20). Hence, the LM-OS-EM-MAP algorithm converges to the MAP solution. More details on the convergence proof for this LM-OS-EM-MAP algorithm can be found in [21].

Now, we will describe efficient implementations of the LM-OS-EM-ML algorithm given by (27) and (28), and the LM-OS-EM-MAP given by (27) and (29). Our goal is to come up with implementations which have the same computational complexity per k iteration as the LM-EM-ML algorithm in (13) and (14). This is accomplished by recursively updating the term $B_j^{(k,l)} \equiv \sum_{i=1}^{E}\bar{z}_{ij}^{(k,l)}$ in each (k,l) iteration, so that projections and backprojections only over the events in the subset S_l are required. By considering the equations (14), (25), (28) and (29), we may observe that iterative updating of B_j is the crucial computation in the M-steps or modified M-steps of all algorithms considered in this paper. For example, we may re-write the term C in (29) in the LM-OS-EM-MAP algorithm as follows: $C = -B_j^{(k,l)}$. Other than C, the other terms in (29) can be quickly computed. An iterative update of B_j first requires a projection to update $\bar{z}_{ij}$ in the E-step, followed by a backprojection over all events. For our OS algorithms, we need to reduce this computation to a projection and backprojection only over a subset of events. Define $A_j^{(k,l)} \equiv \sum_{i\in S_l}\bar{z}_{ij}^{(k,l)}, \forall S_l, l=1,\cdots,L$. Then our recursive update scheme is given by:

$$B_j^{(k,l)} = A_j^{(k,l)} - A_j^{(k,l-1)} + B_j^{(k,l-1)}, \forall j \tag{30}$$

Thus, our primary computation reduces to iteratively update $A_j^{(k,l)}$ in each (k,l) iteration and this requires projections and backprojections only over the events in the subset S_l. Our recursive updates do imply some additional memory requirements ($L+1$ extra images need to be stored), but this is not a severe requirement.

VI. Simulations and Discussion

We used the object in Fig. 1 (left) and simulated 2D SPECT with attenuation effects. SPECT with conventional attributes is not an application that usually benefits significantly from list-mode data acquisition, but it has been shown to be advantageous in [23], [24]. Here, our aim is to merely demonstrate the efficacy of our algorithm. We have imitated [23] in choosing the conditions for our list-mode SPECT simulations. We collected 50000 LM events, $\theta(i), t(i), i=1,\cdots,50000$ over 128 angles and 1536 bins. Note that we are measuring the bin-position with tremendous accuracy and hence list-mode reconstruction is desirable. We followed the procedure given below in simulating list-mode data:

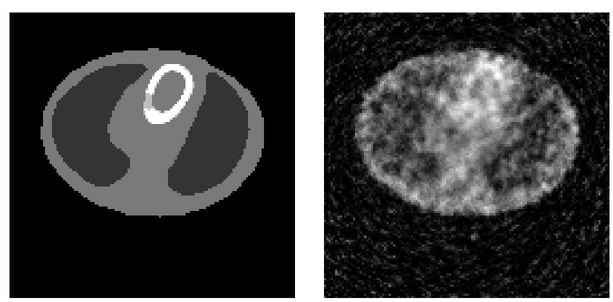

Fig. 1. 128×128 Object (left) with Reconstruction (right)

1) Generate noiseless (floating point) projection data.
2) Scale projection data so mean count level typically < 1 at each bin.
3) Add Poisson noise using this low mean. Typically, we realized 0 counts in a bin, with a few counts = 1 and even fewer = 2.
4) Re-organize counts into a list.

The value of the regularization parameter β was 500.

An important issue in the reconstruction process is the estimation of the pixel sensitivity functions D_j. (The image formed by $D_j, j = 1, \cdots, N$ is referred to as the efficiency image in [25].) It was shown in that [25] that errors in estimating D_j can significantly reduce image quality. Estimation of D_j is a difficult task and different strategies are mentioned in [25]. Here, we adopted a simple approach. We backprojected over all bins on a lower resolution (we used 192 bins instead of 1536) and then scaled the efficiency image appropriately.

A reconstruction is shown in Fig. 1 (right). A speed comparison of our LM-OS-EM-MAP algorithm in (27) and (29) with 1, 4, 8, 16, 32 and 64 subsets is shown in Fig. 2. In this plot, an iteration corresponds to a pass through all subsets. Subsets were chosen by angle. For example, for $L = 4$, the first subset consisted of events whose angle attribute index was $0, 4, 8, \cdots$, the second subset consisted of events whose angle attribute index was $1, 5, 9, \cdots$ and so on. The figure plots (on a log scale) the normalized energy difference, defined by $NED = \frac{\mathcal{E}_P(\hat{\mathbf{f}}^*) - \mathcal{E}_P(\hat{\mathbf{f}}^{(k)})}{\mathcal{E}_P(\hat{\mathbf{f}}^*) - \mathcal{E}_P(\hat{\mathbf{f}}^{(0)})}$ vs. iteration k. Here, $\hat{\mathbf{f}}^*$ is the convergent solution obtained by running 1000 iterations of our algorithm at $L = 64$, and $\hat{\mathbf{f}}^{(0)}$, the initial estimate, was a constant image scaled to the appropriate count level.

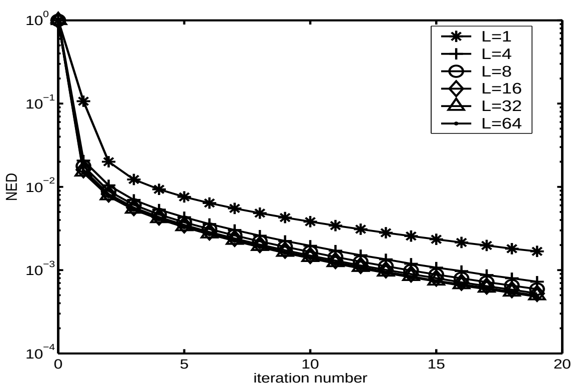

Fig. 2. Normalized Energy Distance (NED) vs. iteration for LM-OS-EM-MAP algorithm using different numbers of subsets. Note $L = 1$ corresponds to LM-EM-MAP.

The use of $L = 4$ subsets led to an approximate factor of 3 speed-up. However, the use of higher numbers of subsets ($L = 8, 16, 32, 64$) led to only slight further increases in speed. This behavior is unlike that of conventional OS-type algorithms where use of subsets leads to a typically greater speed-up. We have explored a new idea in [26], [27] for binned-mode data in which an automatically computed parameter controls a tradeoff between application of a fast traditional OS algorithm and a binned-mode version of our convergent OS algorithm. The resulting algorithm is *still* convergent, yet the OS- speedup factor is considerably enhanced. It appears likely that we can extend this idea to our LM-OS-EM-MAP algorithm to achieve greater OS speed-up.

We plan to conduct more realistic simulations modeling physical effects such as depth-dependent detector response and scatter. We also plan to explore different subset schemes and compare the speed of the resulting variants. We have used quadratic regularizers in our algorithms because they lead to simple closed form updates that automatically enforce positivity. One can use convex non-quadratic edge preserving regularizers, but this destroys the closed form updates of our LM-OS-EM-MAP algorithm. One needs to use a bent line search to perform 1-D optimizations during each sub-iteration, but this does not increase overall computation much.

The modified BSREM (Block Sequential Regularized EM) algorithm in [28] is a recent fast, globally convergent OS algorithm for MAP reconstruction. It was derived using a penalized binned-mode likelihood. One may derive a modified BSREM algorithm starting from a penalized LM likelihood. However, this algorithm requires a carefully chosen relaxation schedule in order to ensure convergence. An advantage of our LM-OS-EM-MAP algorithm is that no relaxation schedule is needed in order to obtain convergence.

Finally, we will mention an alternative approach to deriving the OS algorithms described in this paper. This approach is in the same spirit as in [7]. We may start with the final form of the iterative updates of our convergent OS-EM-ML [29], [21] and OS-EM-MAP [13], [21] algorithms for *binned*-mode data and then algebraically manipulate them into a list-mode representation to get our OS algorithms in the previous section. The algebraic manipulation involves replacing weighted sums over all sinogram bins, where the weights are the measured counts in each bin, by sums over all measured events.

Thus, following the spirit of the Parra-Barrett approach [9], we have derived a LM-OS-EM-MAP algorithm starting from the concept of a list-mode likelihood, albeit using the LM likelihood via a complete data objective formulation. One could take an alternate path and derive the LM-OS-EM-MAP algorithm beginning with a *binned*-mode likelihood and applying suitable transformations. Our approach is nevertheless valuable since it starts from first principles and is more flexible. We point to one example where the use of LM likelihood approaches can be of value: An important fact that distinguishes the list-mode derivation of the EM-type algorithms in this paper from the binned-mode derivations [16] is the binary nature of our complete data. EM algorithms that use binary complete data are commonly employed for parameter estimation in mixture models. Faster variants of EM algorithms [30] have been successfully used for such mixture models. In the future, we plan to investigate whether such fast variants can be developed for emission tomography. Such variants depend on the binary nature of the complete data and are naturally developed starting from a LM likelihood formulation. Other advances may similarly benefit from a LM likelihood formulation.

VII. Acknowledgements

This work was supported by NIH-NIBIB EB02629. We thank A. J. Reader, A. Lehovich and H. Barrett for useful discussions.

References

[1] H. H. Barrett, T. White, and L. Parra, "List-mode Likelihood," *J. Optical Soc. America A*, vol. 14, no. 11, pp. 2914–2923, Nov. 1997.

[2] J. Qi and R. Huesman, “List-mode Reconstruction for PET with Motion Compensation: a Simulation Study,” in *Conf. Rec. IEEE Int. Symp. Biomed. Imaging*. July 2002, pp. 413–416, IEEE.

[3] R. E. Carson, W. C. Barker, J.-S. Liow, S. Adler, and C. A. Johnson, “Design of a Motion-Compensation OSEM List-mode Algorithm for Resolution-Recovery Reconstruction for the HRRT,” in *Conf. Rec. IEEE Nuc. Sci. Symp. Med. Imaging Conf.* Oct. 2003, IEEE.

[4] R. R. Fulton, I. Nickel, L. Tellmann, S. R. Meikle, U. K. Pietrzyk, and H. Herzog, “Event-by-event Motion Compensation in 3D PET,” in *Conf. Rec. IEEE Nuc. Sci. Symp. Med. Imaging Conf.* Oct. 2003, IEEE.

[5] A. Rahmim, P. M. Bloomfield, S. Houle, M. W. Lenox, C. Michel, and V. Sossi, “Motion Correction in Histogram-Mode and List-Mode EM Reconstructions,” in *Conf. Rec. IEEE Nuc. Sci. Symp. Med. Imaging Conf.* Oct. 2003, IEEE.

[6] K. Thielemans and S. Mustafovic, “Image Reconstruction of Motion Corrected Sinograms,” in *Conf. Rec. IEEE Nuc. Sci. Symp. Med. Imaging Conf.* Oct. 2003, IEEE.

[7] A. J. Reader, K. Erlandsson, M. A. Flower, and R. J. Ott, “Fast Accurate Iterative Reconstruction for Low-statistics Positron Volume Imaging ,” *Physics in Medicine and Biology*, vol. 43, no. 4, pp. 823–834, April 1998.

[8] A. Lehovich and H. Barrett, “List-mode SPECT Reconstruction Using Empirical Likelihood,” in *Conf. Rec. Int. Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, 2003, pp. Mo–PM3–3.

[9] L. Parra and H. Barrett, “List-mode Likelihood: EM Algorithm and Image Quality Estimation Demonstrated on 2-D PET,” *IEEE Trans. Med. Imaging*, vol. 17, no. 2, pp. 228–235, April 1998.

[10] R. Levkovitz, D. Falikman, M. Zibulevsky, A Ben-Tal, and A. Nemirovski, “The Design and Implementation of COSEM, an Iterative Algorithm for Fully 3-D Listmode Data,” *IEEE Trans. Med. Imaging*, vol. 20, no. 7, pp. 633–642, July 2001.

[11] T. E. Nichols, J. Qi, E. Asma, and R. Leahy, “Spatiotemporal Reconstruction of List-mode PET Data ,” *IEEE Trans. Med. Imaging*, vol. 21, no. 4, pp. 396–404, April 2002.

[12] Asela Gunawardana, *The Information Geometry of EM Variants for Speech and Image Processing*, Ph.D. thesis, Dept of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD 21218, April 2001.

[13] I. T. Hsiao, A. Rangarajan, and G. Gindi, “A New Convergent MAP Reconstruction Algorithm for Emission Tomography Using Ordered Subsets and Separable Surrogates,” in *Conf. Rec. IEEE Int. Symp. Biomed. Imaging*. July 2002, pp. 409–412, IEEE.

[14] K. Mueller, R. Yagel, and J. F. Cornhill, “The Weighted-Distance Scheme: a Globally Optimizing Projection Ordering Method for ART ,” *IEEE Trans. Med. Imaging*, vol. 16, no. 2, pp. 223–230, April 1997.

[15] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum Likelihood from Incomplete Data via the EM Algorithm,” *Journal of the Royal Statistical Society*, vol. 39, no. 1, pp. 1–38, 1977.

[16] L. A. Shepp and Y. Vardi, “Maximum Likelihood Reconstruction for Emission Tomography,” *IEEE Trans. Med. Imaging*, vol. 1, pp. 113–122, 1982.

[17] K. Lange and R. Carson, “EM Reconstruction Algorithms for Emission and Transmission Tomography,” *Journal of Computer Assisted Tomgraphy*, vol. 8, no. 2, pp. 306–316, Apr. 1984.

[18] A. Rangarajan, I.-T. Hsiao, and G. Gindi, “A Bayesian Joint Mixture Framework for the Integration of Anatomical Information in Functional Image Reconstruction,” *Journal of Mathematical Imaging and Vision*, vol. 12, no. 3, pp. 199–217, June 2000.

[19] A. R. De Pierro, “A Modified Expectation Maximization Algorithm for Penalized Likelihood Estimation in Emission Tomography ,” *IEEE Trans. Med. Imaging*, vol. 14, no. 1, pp. 132–137, Mar. 1995.

[20] H. Erdogan and J. Fessler, “Monotonic Algorithms for Transmission Tomography,” *IEEE Trans. Med. Imaging*, vol. 18, no. 9, pp. 801–814, Sep 1999.

[21] A. Rangarajan, P. Khurd, I.-T. Hsiao, and G. Gindi, “Convergence Proofs for the COSEM-ML and COSEM-MAP Algorithms,” Tech. Rep. MIPL-03-1, Medical Image Processing Lab, Depts. of Radiology and Electrical Engineering, State University of New York at Stony Brook. Web URL: http://www.mil.sunysb.edu/mipl/publications.html, Dec. 2003.

[22] R. Neal and G. Hinton, “A View Of The EM Algorithm That Justifies Incremental, Sparse, And Other Variants,” *Learning in Graphical Models*, 1998.

[23] L. Bouwens, R. Van de Walle, H. Gifford, M. King, I. Lemahieu, and R.A. Dierckx, “LMIRA: List-mode Iterative Reconstruction Algorithm for SPECT,” *IEEE Trans. Nuclear Science*, vol. 20, no. 4, pp. 1364–1370, Aug 2001.

[24] L. Bouwens, H. Gifford, R. Van de Walle, M. King, I. Lemahieu, and R.A. Dierckx, “Resolution Recovery for List-mode Reconstruction in SPECT,” *Phys. Med. Biol.*, vol. 46, pp. 2239–2253, Aug 2001.

[25] J. Qi, “Propagation of Errors from the Efficiency Image in List Mode Reconstruction,” in *Conf. Rec. IEEE Nuc. Sci. Symp. Med. Imaging Conf.* Oct. 2003, IEEE.

[26] I. T. Hsiao, P. Khurd, A. Rangarajan, and G. Gindi, “An Accelerated Convergent Ordered Subset Algorithm for 3D Emission Tomography,” in *Conf. Rec. Int. Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, July 2003, pp. Th AM2–1.

[27] I. T. Hsiao, A. Rangarajan, P. Khurd, and G. Gindi, “An Accelerated Convergent Ordered Subsets Algorithm for Emission Tomography,” *Phys. Med. Biol.*, to appear, 2004.

[28] S. Ahn and J. Fessler, “Globally Convergent Ordered Subsets Algorithms for Emission Tomography Using Relaxed Ordered Subsets Algorithms,” *IEEE Trans. Med. Imaging*, vol. 22, no. 5, pp. 613–626, May 2003.

[29] I. T. Hsiao, A. Rangarajan, and G. Gindi, “A Provably Convergent OS-EM Like Reconstruction Algorithm for Emission Tomography,” in *Conf. Rec. SPIE Med. Imaging*. Feb. 2002, vol. 4684, pp. 10–19, SPIE.

[30] Y. Matsuyama, “The α-EM Algorithm : Surrogate Likelihood Maximization Using α-Logarithmic Measures,” *IEEE Trans. Information Theory*, vol. 94, no. 3, pp. 692–706, Mar. 2003.