submitted to *Physics in Medicine and Biology*

An Accelerated Convergent Ordered Subsets Algorithm for Emission Tomography

Ing-Tsung Hsiao[†]§, **Anand Rangarajan**[♭], **Parmeshwar Khurd**[‡], **Gene Gindi**[‡]

[†]School of Medical Technology, Chang Gung University, Kwei-Shan, Tao-Yuan 333, Taiwan

[♭]Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL 32611, USA

[‡]Departments of Electrical & Computer Engineering and Radiology, SUNY Stony Brook, Stony Brook, NY 11784, USA

Abstract. We propose an algorithm, E-COSEM (Enhanced Complete-Data Ordered Subsets Expectation-Maximization) for fast maximum likelihood (ML) reconstruction in emission tomography. E-COSEM is founded on an incremental EM approach. Unlike the familiar OSEM (ordered subsets EM) algorithm which is not convergent, we show that E-COSEM converges to the ML solution. Alternatives to OSEM include RAMLA, and for the related maximum *a posteriori* (MAP) problem, the BSREM and OS-SPS algorithms. These are fast and convergent, but require a judicious choice of a user-specified relaxation schedule. E-COSEM itself uses a sequence of iteration-dependent parameters (very roughly akin to relaxation parameters) to control a tradeoff between a greedy, fast but non-convergent update and a slower but convergent update. These parameters are computed automatically at each iteration and require no user specification. For the ML case, our simulations show that E-COSEM is nearly as fast as RAMLA.

§ E-mail:ihsiao@mail.cgu.edu.tw (Ing-Tsung Hsiao)

1. Introduction

Since the introduction of the OSEM algorithm (Hudson and Larkin 1994) for emission tomography (ET), there has been considerable interest in further developing iterative OS (ordered subset) type methods for ET. Like the original EM-ML (expectation maximization - maximum likelihood) algorithm and unlike analytical methods, OS methods can easily incorporate system models for attenuation, detector response and system geometry. Unlike EM-ML, OS methods are typically quite fast, requiring only a few iterations for a reconstruction deemed acceptable by many users. This speed aspect has inspired the use of OS-type methods for fully 3D PET reconstructions (Matej *et al* 2001). The OSEM algorithm in itself does not maximize likelihood, and can lead to limit cycles in the iterative object estimates.

Thus interest has been focused in deriving provably convergent versions of fast OS methods. In Browne and De Pierro (1996), an OS method (RAMLA) for ML ET reconstruction was proposed along with a convergence proof. In RAMLA, the iterative updates are controlled by a user-specified relaxation schedule to ensure convergence. For a given object and noise level, the user must experiment with parameters of the relaxation schedule in order to attain the potential speed of RAMLA. This is a serious inconvenience. There has also been interest in OS-type algorithms to quickly maximize penalized likelihood (a.k.a. MAP or maximum *a posteriori*) methods. In De Pierro and Yamagishi (2001), and Ahn and Fessler (2003), convergent BSREM and OS-SPS algorithms were proposed for fast MAP reconstruction. Again, these MAP algorithms require a user-specified relaxation schedule.

As mentioned above, the main problem with RAMLA, OS-SPS and BSREM is that there is no easy way to determine relaxation schedules that lead to fast algorithms while simultaneously satisfying theoretical criteria to ensure convergence. In this paper, we build on previous work (Hsiao *et al* 2002a, Hsiao *et al* 2002b) to derive a convergent OS-type ML algorithm, termed E-COSEM-ML (Enhanced Complete Data OSEM-ML), that is fast (nearly as fast as optimized RAMLA), but avoids the problem of user-specified relaxation schedules. E-COSEM-ML is founded on an incremental EM approach (Neal and Hinton 1998, Gunawardana 2001) that is fundamentally different from the approaches in Browne and De Pierro (1996), De Pierro and Yamagishi (2001), and Ahn and Fessler (2003).

Our work in this paper is focused on ML algorithms for ET. Since such algorithms are rarely carried to convergence, then one might question the need for provably convergent versions and use instead an expedient empirical method such as OSEM. Furthermore, even early-terminated convergent ML methods have 3 user-specified parameters: (1) initial condition, (2) stopping rule and (3) relaxation schedule. We note that relaxation schedules are used in Browne and De Pierro (1996), but not in other approaches. For our E-COSEM algorithm, our goal is to eliminate (3) while maintaining a fast convergent algorithm. With this accomplished, E-COSEM-ML may be extended to a MAP case (as discussed in Sec. IV), thus removing the problem of stopping rule. For the remainder of the text, we shall use the acronym E-COSEM to imply the E-COSEM-ML version.

In section II, we develop the theory for E-COSEM-ML, and show some comparisons to EM-ML and RAMLA in section III. A discussion is included in section IV.

2. Theory

Let vector $\mathbf{f}$ with lexicographically ordered components $f_j, j = 1, \ldots, N$ denote the mean object activity at voxel j, vector $\mathbf{g}$ with lexicographically ordered components $g_i, i = 1, \ldots, M$ denote the measured counts at detector bin i. A bin could be a detector element in SPECT or a detector pair in PET. Denote the system matrix by $\mathcal{H}$ with elements $\mathcal{H}_{ij}$. The system matrix element $\mathcal{H}_{ij}$ is proportional to the probability of receiving a count in bin i originating from voxel j. With the usual assumption of independent detector counts and Poisson noise, $g_i \sim Poisson([\mathcal{H}\mathbf{f}]_i)$ so that the mean sinogram is $\bar{\mathbf{g}} = \mathcal{H}\mathbf{f}$. (The derivation is easily extended to include a Poisson case with $\bar{\mathbf{g}} = \mathcal{H}\mathbf{f} + \bar{\mathbf{s}}$ where $\bar{\mathbf{s}}$ is a vector with components $\bar{s}_i, i = 1, ..., M$ representing mean scatter or, for PET, randoms.) The maximum likelihood (ML) reconstruction, the goal of this work, is then obtained as

$$\mathbf{f}^* = \arg\min_{\mathbf{f}\geq 0} E_{\text{incomplete}}(\mathbf{f}) \tag{1}$$

In EM approaches, the objective in (1) is referred to as the "incomplete-data negative log-likelihood". This ML objective is given by

$$E_{\text{incomplete}}(\mathbf{f}) = \sum_{ij} \mathcal{H}_{ij} f_j - \sum_i g_i \log \sum_j \mathcal{H}_{ij} f_j \tag{2}$$

To motivate our E-COSEM algorithm for optimizing (2), we first show a novel means of deriving the familiar EM-ML algorithm for ET. (This is but one of five ways to derive EM-ML!). It turns out (Rangarajan *et al* 2000, Hsiao *et al* 2002a) that EM-ML may be derived by optimizing the following "complete-data" objective:

$$\begin{aligned} E_{\text{complete}}(\mathbf{f}, \mathbf{C}, \boldsymbol{\lambda}) &= -\sum_{ij} C_{ij} \log \mathcal{H}_{ij} f_j + \sum_{ij} \mathcal{H}_{ij} f_j \\ &+ \sum_{ij} C_{ij} \log C_{ij} + \sum_i \lambda_i (\sum_j C_{ij} - g_i). \end{aligned} \tag{3}$$

In (3), $\mathbf{C}$ (with entry C_{ij}) is identified with (but not equal to) the complete data as used in conventional statistical derivations of EM-ML. Conventionally, C_{ij} is identified as the integer number of photons leaving j and detected at i, but in (3), C_{ij} is a positive analog quantity explained below. Also, $\boldsymbol{\lambda}$ is a Lagrange parameter vector with component λ_i which expresses the constraint $\sum_j C_{ij} = g_i$. In Rangarajan *et al* (2000) and Hsiao *et al* (2002a), we showed that a coordinate descent on $\mathbf{C}$ and $\mathbf{f}$ while imposing the constraint $\sum_j C_{ij} = g_i$ leads to update equations for $\mathbf{C}$ and $\mathbf{f}$ which, when combined, yield the familiar EM-ML update. To see this, we first write the fixed point solution for $\mathbf{C}$ in terms of $\mathbf{f}$ and $\mathbf{g}$. This is obtained from the first order Karush-Kuhn-Tucker condition

(Karush 1939, Kuhn and Tucker 1950) by setting the first derivative of $E_{\text{complete}}(\mathbf{f}, \mathbf{C}, \boldsymbol{\lambda})$ w.r.t. $\mathbf{C}$ to zero and solving for $\mathbf{C}$ to get

$$C_{ij} = \mathcal{H}_{ij} f_j \exp\{-\lambda_i - 1\}. \tag{4}$$

The Lagrange parameter vector $\boldsymbol{\lambda}$ can be eliminated by using the constraint $\sum_j C_{ij} = g_i$ in (4) to get

$$C_{ij} = g_i \frac{\mathcal{H}_{ij} f_j}{\sum_{j'} \mathcal{H}_{ij'} f_{j'}},\ \forall i,\ \forall j. \tag{5}$$

The fixed point solution for $\mathbf{f}$ in terms of $\mathbf{C}$ and $\mathbf{g}$ is obtained by setting the first derivative of $E_{\text{complete}}(\mathbf{f}, \mathbf{C}, \boldsymbol{\lambda})$ w.r.t. $\mathbf{f}$ to zero and solving for $\mathbf{f}$ to get

$$f_j = \frac{\sum_i C_{ij}}{\sum_i \mathcal{H}_{ij}},\ \forall j. \tag{6}$$

Equations (5) and (6) specify a fixed-point solution for $\mathbf{C}$ in terms of $\mathbf{f}$ and a fixed-point solution for $\mathbf{f}$ in terms of $\mathbf{C}$, respectively. This pair of mutually dependent fixed-point solutions can be converted into a single $\mathbf{f}$ updating algorithm. Introducing a sequence of iterations indexed by an integer $k \in \{1, 2, \ldots\}$, we convert the above fixed point solutions for $\mathbf{C}$ and $\mathbf{f}$ in (5) and (6), respectively, to get

$$f_j^{(k+1)} = \frac{f_j^{(k)}}{\sum_i \mathcal{H}_{ij}} \sum_i \frac{\mathcal{H}_{ij} g_i}{\sum_{j'} \mathcal{H}_{ij'} f_{j'}^{(k)}},\ \forall j \tag{7}$$

where $f_j^{(k)}$ is the estimate at voxel j and at iteration k. This is the familiar EM-ML algorithm for emission tomography but derived via the new complete data objective function (3).

The above derivation of the ET EM-ML algorithm from a complete data objective is not new. To our knowledge, albeit for a problem in mixtures decomposition rather than ET, the first example of recasting an incomplete negative log-likelihood objective function as a constrained objective is (Hathaway 1986) wherein the mixtures (incomplete) likelihood is recast using a complete data objective function. In contrast, the first place, to our knowledge where the ET EM-ML algorithm was derived using a complete data objective was in (Lee 1994) (Appendix B) albeit using a complete data variable which is closely related yet different from the one used here. The complete data objective function used here therefore has a lineage in Hathaway (1986), Lee (1994), Rangarajan *et al* (1996), Neal and Hinton (1998), Rangarajan *et al* (2000), Hsiao *et al* (2002a) and Rangarajan *et al* (2003). Note that no connection to the world of ordered subsets has been established thus far.

The principal goal in this paper is to derive fast and provably convergent ordered subsets-like algorithms for emission tomography. To facilitate this and to introduce notation, we first review the familiar ordered subsets EM (OSEM) algorithm for ET (Hudson and Larkin 1994). For clarity, we use a SPECT context where subsets are associated with projection angles, but our argument applies to any subset scheme. Assume that the set of projection angles is subdivided into L subsets with the projection data in each subset denoted as $\{g_i, \forall i \in S_l\}$ with $l \in \{1, \ldots, L\}$. The OSEM ET algorithm is written as

$$f_j^{(k,l)} = \frac{f_j^{(k,l-1)}}{\sum_{i \in S_l} \mathcal{H}_{ij}} \sum_{i \in S_l} \frac{g_i \mathcal{H}_{ij}}{\sum_{j'} \mathcal{H}_{ij'} f_{j'}^{(k,l-1)}}, \forall j \tag{8}$$

where $f_j^{(k,l)}$ denotes the estimate at voxel j at iteration (k, l). In the outer k loop, we assume that all subsets have been updated, whereas in the inner $l \in \{1, \ldots, L\}$ loop, each subiteration l corresponds to an update using the *limited* backprojection $\sum_{i \in S_l} g_i \frac{\mathcal{H}_{ij}}{\sum_{j'} \mathcal{H}_{ij'} f_{j'}^{(k,l-1)}}$. The outer/inner loop structure of the iterations is required because the projection data from only a limited number of projection angles are backprojected at each "subset" subiteration (k, l). This is repeated over all the projection angles after which the inner l loop is terminated and the next outer loop iteration k initiated. In each inner loop (k, l) subiteration, we update $\{f_j, \forall j\}$.

The above OSEM algorithm is now rewritten using the complete data notation. We do this in order to firmly establish the close relationship between our *convergent* ordered subsets algorithms and OSEM. Rewriting (8), we get

$$f_j^{(k,l)} = \frac{\sum_{i \in S_l} C_{ij}^{(k,l)}}{\sum_{i \in S_l} \mathcal{H}_{ij}} \tag{9}$$

where $\mathbf{C}^{(k,l)}$ is the estimate of the *entire set* of the complete data at iteration (k, l). Corresponding to the division of the incomplete data into subsets (e.g. projection angles for SPECT), we also have a division of the complete data $\mathbf{C}$ which is denoted by $\{C_{ij}, \forall i \in S_l, \forall j\}$ with $l \in \{1, \ldots, L\}$. The update of the complete data is

$$C_{ij}^{(k,l)} = g_i \frac{\mathcal{H}_{ij} f_j^{(k,l-1)}}{\sum_{j'} \mathcal{H}_{ij'} f_{j'}^{(k,l-1)}}, \forall i \in S_l, \forall j \tag{10}$$

$$C_{ij}^{(k,l)} = C_{ij}^{(k,l-1)}, \forall i \notin S_l, \forall j. \tag{11}$$

Equations (10) and (11) partition the update of the complete data into subsets. At subiteration (k, l), we only update the complete data $\{C_{ij}, \forall i \in S_l, \forall j\}$ over a subset S_l of the projection angles. The remaining complete data $\{C_{ij}, \forall i \notin S_l, \forall j\}$ are not updated. Instead, as shown in (11), the remaining complete data retain their values from the previous subiteration.

We now revisit the complete data objective function in (3) and rewrite it using ordered subsets notation. After rewriting the complete data objective, we can derive a subset version of the update equations for $\mathbf{C}$ and $\mathbf{f}$, following a similar strategy as used in obtaining (4), (5) and (6). The only difference is that, in the sequence of updates, only a limited subset of the complete data is updated. This is immediately followed by an $\mathbf{f}$ update. We rewrite (3) using ordered subsets notation.

$$E_{\text{complete}}(\mathbf{f}, \mathbf{C}, \boldsymbol{\lambda}) = \sum_{l=1}^{L} \sum_{i \in S_l} \sum_{j=1}^{N} C_{ij} \log \frac{C_{ij}}{\mathcal{H}_{ij} f_j}$$

$$+\sum_{ij}\mathcal{H}_{ij}f_j+\sum_{l=1}^{L}\sum_{i\in S_l}\lambda_i(\sum_j C_{ij}-g_i) \tag{12}$$

where the M bins have been divided into L subsets $S_l, l=1,\ldots,L$ in a manner similar to OSEM approaches. We repeat the strategy used to obtain (7), but instead of attempting a descent on all of $\mathbf{C}$, we do this one subset at a time. At each (k,l), we first update only those $C_{ij}, i \in S_l$ followed by an update of all f_j. Doing this leads to the COSEM updates reported in (Hsiao *et al* 2002a):

$$C_{ij}^{(k,l)} = g_i\frac{\mathcal{H}_{ij}f_j^{(k,l-1)}}{\sum_{j'}\mathcal{H}_{ij'}f_{j'}^{(k,l-1)}},\ \forall i\in S_l,\forall j \tag{13}$$

$$C_{ij}^{(k,l)} = C_{ij}^{(k,l-1)},\ \forall i\notin S_l,\forall j \tag{14}$$

$$f_j^{(k,l)} = \frac{\sum_i C_{ij}^{(k,l)}}{\sum_i\mathcal{H}_{ij}} \tag{15}$$

where $f_j^{(k,l)}$ and $C_{ij}^{(k,l)}$ denote the estimate of f_j and C_{ij} at iteration (k,l). Note that $C_{ij}^{(k,0)} = C_{ij}^{(k-1,L)}, \forall ij$. Despite the fact that only a subset of the complete data is updated at iteration (k,l), we carry along the rest of the complete data as shown in (14). This permits the summation in the numerator of (15) to be over all i.

For book-keeping purposes, we keep track of the value of $B_j^{(k,l)} \stackrel{\text{def}}{=} \sum_i C_{ij}^{(k,l)}$ and the values of

$$A_j^{(k,l)} \stackrel{\text{def}}{=} \sum_{i\in S_l} C_{ij}^{(k,l)},\ \forall S_l, l=1,\ldots,L,\ \forall j. \tag{16}$$

We also define the quantities $D_j \stackrel{\text{def}}{=} \sum_i\mathcal{H}_{ij},\ \forall j$ and $T_j^{(l)} \stackrel{\text{def}}{=} \sum_{i\in S_l}\mathcal{H}_{ij},\ \forall j$. The values of $B_j^{(k,l)}$ can be recursively updated according to the following equation:

$$B_j^{(k,l)} = A_j^{(k,l)} - A_j^{(k-1,l)} + B_j^{(k,l-1)},\ \forall j \tag{17}$$

with initial condition $B_j^{(k,0)} = B_j^{(k-1,L)}$. Then (15) can be re-expressed as

$$\tilde{f}_j^{(k,l)} = \frac{B_j^{(k,l)}}{D_j},\ \forall j \tag{18}$$

where D_j is the system sensitivity at voxel j, and we append a tilde to $f_j^{(k,l)}$ to denote $\tilde{f}_j^{(k,l)}$ as the COSEM estimate at voxel j and at iteration (k,l). The OSEM algorithm (9) in this new notation becomes

$$\breve{f}_j^{(k,l)} = \frac{\sum_{i\in S_l} C_{ij}^{(k,l)}}{\sum_{i\in S_l}\mathcal{H}_{ij}} = \frac{A_j^{(k,l)}}{T_j^{(l)}} \tag{19}$$

where $\breve{f}_j^{(k,l)}$ denotes the OSEM estimate at voxel j and at iteration (k,l).

It can be shown (Hsiao *et al* 2002a, Gunawardana 2001, Neal and Hinton 1998, Rangarajan *et al* 2003) that COSEM monotonically decreases the complete data objective (12) and that $\mathbf{f}^{(k,l)}$ asymptotically approaches the ML solution $\mathbf{f}^*$. (Please see the Appendix for an outline of the proof.) The decrease in incomplete-data negative log-likelihood with k is not guaranteed to be monotonic, but in practice, we have always observed it to be so. In contrast, OSEM does not actually minimize any objective function. In particular, OSEM cannot be shown to asymptotically minimize the incomplete-data negative log-likelihood and consequently does not converge to a true ML solution. In our convergent ordered subsets approach, there are two different but closely related objective functions. The first objective function is the familiar incomplete-data negative log-likelihood that appears in (2) and the second is the complete data objective function (12). The sequence of updates in (13), (14), (16), (17) and (18) decrease the complete data objective function in (12) and not the incomplete data negative log likelihood in (2). However, it can be shown (Neal and Hinton 1998, Gunawardana 2001, Rangarajan *et al* 2003) that the above sequence of updates asymptotically reaches the minimum of the original incomplete data negative log-likelihood. Therefore, despite the fact that we minimize an objective function which is different from, albeit very closely related to, the incomplete-data negative log-likelihood, asymptotically, we do minimize the incomplete-data negative log-likelihood.

From (17), (18) and (19), we see that the computational expense of each COSEM iteration is about the same as for OSEM. In particular, the numerator in (18) requires only an incremental computation, the first term in (17), since the second and third terms in (17) needn't be recomputed.

We found that COSEM, while much faster than EM-ML in maximizing likelihood, was "slower" than OSEM. By "slower", we mean that a plot of incomplete data log-likelihood vs iteration for COSEM and OSEM shows OSEM increasing the incomplete data log-likelihood towards its asymptotic maximum value faster than COSEM in the first few iterations. Since OSEM does not lead to a true ML solution, it is not guaranteed to attain the correct asymptotic value of log-likelihood. Nevertheless, with these results, we are motivated to speed up COSEM by making it "resemble" OSEM more closely while still retaining its convergence properties. This heuristic impulse is justified below.

The basic idea is as follows: We seek a compromise between the faster OSEM update in (19) and the slower COSEM update in (18), while still guaranteeing convergence. This compromise is achieved by using a linear combination of the OSEM and COSEM updates. The linear combination introduces a tradeoff between the faster OSEM and the slower, but convergent, COSEM with a relaxation parameter $\alpha \in [0, 1]$ expressing the degree of tradeoff. The relaxation parameter α is **not** a user-specified free parameter. Instead, it will be *automatically* chosen to guarantee convergence while keeping the tradeoff as close to OSEM as possible.

To do this, we modify the COSEM update (18) to the following E-COSEM form:

$$f_j^{(k,l)} = \alpha^{(k,l)} \breve{f}_j^{(k,l)} + (1 - \alpha^{(k,l)}) \tilde{f}_j^{(k,l)} \tag{20}$$

where $\alpha^{(k,l)}$ is an iteration-dependent parameter estimated so that it speeds up COSEM. In (20), the OSEM update $\breve{f}_j^{(k,l)}$ and the COSEM update $\tilde{f}_j^{(k,l)}$ are as in (19) and (18), respectively. The update for $\mathbf{C}$, (13) and (14), remains the same. Note that

$$f_j^{(k,l)}|_{\alpha^{(k,l)}=1} = \breve{f}_j^{(k,l)} \tag{21}$$

and

$$f_j^{(k,l)}|_{\alpha^{(k,l)}=0} = \tilde{f}_j^{(k,l)}. \tag{22}$$

Thus α controls the trade-off between OSEM and COSEM. The replacement of (18) by (20) entails two challenges: choose $\alpha^{(k,l)}$ to attain speed, and guarantee that the resulting E-COSEM update is still convergent.

As far as convergence, we first note that the $\mathbf{C}^{(k,l)}$ update, (13) and (14), remain unchanged for E-COSEM, so that optimization w.r.t. $\mathbf{C}$ remains unchanged. It remains to be shown that with suitable choice of $\alpha^{(k,l)}$, (20) guarantees that the complete data objective (12) will still decrease (or remain constant). It turns out, as explained below, that such a decrease in (12) will guarantee a convergence of E-COSEM $\mathbf{f}^{(k,l)}$ to $\mathbf{f}^*$, although $\mathbf{f}^{(k,l)}$ will not converge monotonically to the minimum of the incomplete-data negative log-likelihood. To show the decrease in (12), we first observe that the objective in (12) with $\mathbf{C}$ held fixed at $\mathbf{C}^{(k,l)}$ can be written as

$$E^{(k,l)}(\mathbf{f}) = \sum_j E_j^{(k,l)}(f_j) \tag{23}$$

where $E_j^{(k,l)}(f_j)$ is given by

$$E_j^{(k,l)}(f_j) = D_j(-\frac{B_j^{(k,l)}}{D_j} \log f_j + f_j). \tag{24}$$

In (24), we have dropped terms independent of $\mathbf{f}$. The initial value of each $E^{(k,l)}(\mathbf{f})$ is $E^{(k,l)}(\mathbf{f}^{(k,l-1)})$ with the understanding that $\mathbf{f}^{(k,0)} = \mathbf{f}^{(k-1,L)}$. To attain convergence of (12) it is sufficient for the update $\mathbf{f}^{(k,l)}$ to satisfy

$$E^{(k,l)}(\mathbf{f}^{(k,l)}) \leq E^{(k,l)}(\mathbf{f}^{(k,l-1)}). \tag{25}$$

Using (24) and (18), we may write

$$E_j^{(k,l)}(f_j) = D_j(-\tilde{f}_j^{(k,l)} \log f_j + f_j) \tag{26}$$

where $\tilde{f}_j^{(k,l)}$ is the COSEM update in (18).

To make (20) look like the speedy OSEM, we want α close to unity, but at the same time, we need guarantee a decrease in $E^{(k,l)}(\mathbf{f})$. This strategy is attained by our α update, which is given by

$$\hat{\alpha}^{(k,l)} = \max_{\alpha^{(k,l)} \in [0,1]} \alpha^{(k,l)} \tag{27}$$

subject to

$$\sum_j E_j^{(k,l)}(f_j^{(k,l)}) < \sum_j E_j^{(k,l)}(f_j^{(k,l-1)}) \tag{28}$$

where $f_j^{(k,l)}$ is as defined in (20). Furthermore, note that (Rangarajan *et al* 2003) COSEM (i.e. $\alpha = 0$) guarantees $\sum_j E_j^{(k,l)}(f_j^{(k,l)}) \leq \sum_j E_j^{(k,l)}(f_j^{(k,l-1)})$, so that if the search in (27) and (28) fails, we set $\alpha = 0$. This implies that α will inevitably be driven to zero, and that E-COSEM will approach COSEM. Since COSEM converges, so does ECOSEM. (See Appendix and (Rangarajan *et al* 2003).)

The final E-COSEM $\mathbf{f}$ update equation then is

$$f_j^{(k,l)} = \hat{\alpha}^{(k,l)} \breve{f}_j^{(k,l)} + (1 - \hat{\alpha}^{(k,l)}) \tilde{f}_j^{(k,l)}$$
$$= \hat{\alpha}^{(k,l)} \frac{\sum_{i \in S_l} C_{ij}^{(k,l)}}{\sum_{i \in S_l} \mathcal{H}_{ij}} + (1 - \hat{\alpha}^{(k,l)}) \frac{\sum_i C_{ij}^{(k,l)}}{\sum_i \mathcal{H}_{ij}}$$
$$= \hat{\alpha}^{(k,l)} \frac{A_j^{(k,l)}}{T_j^{(l)}} + (1 - \hat{\alpha}^{(k,l)}) \frac{B_j^{(k,l)}}{D_j}. \quad (29)$$

Note that this final form (29) of the iterative update incorporates the optimum value $\hat{\alpha}^{(k,l)}$, whereas the intermediate form (20) leaves the value of $\alpha^{(k,l)}$ unspecified. The search strategy for α adopted here is to start with $\alpha^{(k,l)} = 1$ corresponding to OSEM, and then repeatedly decrease $\alpha^{(k,l)}$ by a factor of 0.9 until the inequality in (28) is satisfied, or if $\alpha^{(k,l)}$ becomes zero, whichever occurs first. Note that because of our geometric update of α, it cannot exactly equal zero. In our implementation, we limit our α update to a finite number of steps so that it ends up at a small value (0.0096) near zero. In the limit of an infinite number of α-update steps, α approaches 0. This α update, together with (13), (14) and (29) define our E-COSEM algorithm. Thus our "enhancement" parameter α is computed automatically and is *not* a user-specified parameter.

3. Results

We anecdotally explore 2D SPECT reconstruction speed using the EM-ML, OSEM, COSEM, E-COSEM and RAMLA algorithms. We first generated noisy sinograms using a 2D 64×64 phantom, shown in figure 1(a), consisting of a disk background, two hot lesions and two cold lesions with contrast ratio of 1:4:8 (cold:background:hot). The projection data had dimensions of 64 angles by 96 detector bins. Poisson noise (300 K counts), uniform attenuation, and depth-dependent blur were simulated. We used an initial condition $\mathbf{f}^{(0,0)}$ = a constant.

The image was reconstructed using the EM-ML (7), OSEM (19), COSEM (13, 14, 18), E-COSEM (13, 14, 29) and RAMLA (14, 15, 16 in (Browne and De Pierro 1996)) algorithms using $L = 32$ subsets for all algorithms except EM-ML. For RAMLA, the relaxation schedule (Browne and De Pierro 1996) was chosen as

$$\beta^k = \frac{\beta_0}{\max_l T_j^{(l)} + k + 1} \quad (30)$$

with $\beta_0 = 2.90$. This choice adheres to the necessary conditions in (Browne and De Pierro 1996), and through extensive numerical exploration, β_0 was chosen to yield as fast a convergence as we could attain.

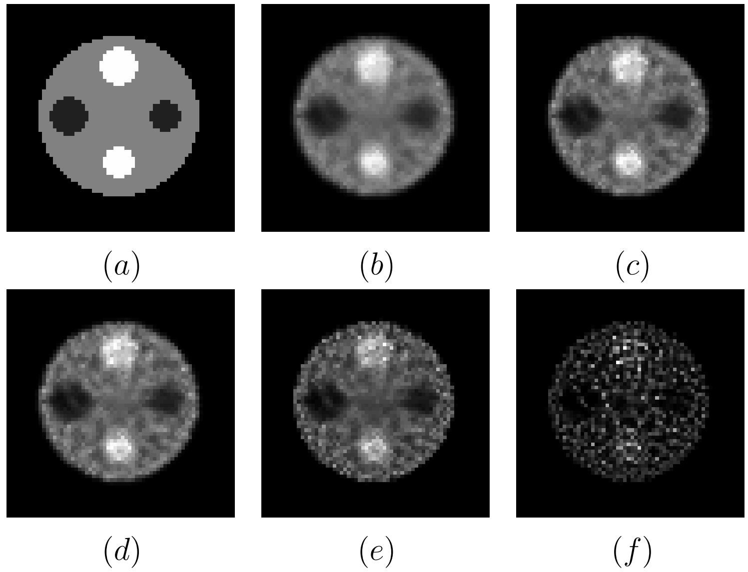

Figure 1. The 2D 64x64 phantom is shown in (a), while the anecdotal reconstructions are displayed in (b) EM-ML, (c) COSEM, (d) E-COSEM, (e) RAMLA, and (f) OSEM. There are 32 subsets used in the OS-type reconstructions, and all are run to 20 iterations.

Anecdotal reconstructions for each algorithm at iteration 20, using the above physical effects, and $L = 32$ subsets, are shown in figure 1(b) EM-ML, 1(c) COSEM, 1(d) E-COSEM, 1(e) RAMLA, and 1(f) OSEM. By 20 iterations, COSEM, E-COSEM, and RAMLA have converged nearly to the same value of the incomplete-data log-likelihood, and the reconstructions look similar. The EM-ML algorithm indeed has not converged as much and looks somewhat smoother. The OSEM algorithm in figure 1(f) yields a somewhat different reconstruction.

We plot the incomplete-data log-likelihood vs. log iteration of each reconstruction in figure 2 for EM-ML ($--$), COSEM ($+$), E-COSEM ($\triangleright$), and RAMLA ($\circ$). As expected, all methods show relative order-of-magnitude acceleration over EM-ML. The speed of COSEM lies between that of RAMLA and EM-ML, and the speed of E-COSEM lies between that of COSEM and RAMLA. Note that ECOSEM is considerably faster than COSEM. In figure 2, the curve for OSEM is not plotted. However, it almost exactly overlaps that for RAMLA.

We also plot the incomplete-data log-likelihood vs. log-iteration for E-COSEM, under the same physical effects, but with varying numbers of subsets, in figure 3, using 1, 4, 8, 16, 32 and 64 subsets. One can see the speed-up of E-COSEM using larger numbers of subsets. Note that the speed-up continues through 32 subsets, but slows down as the number of subset increases to 64.

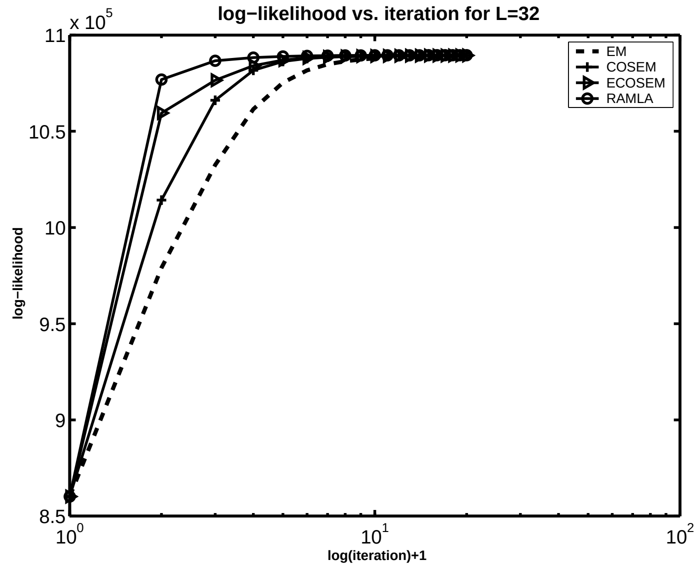

Figure 2. This shows the plots of log-likelihood vs. log-iteration for the reconstructions using EM-ML ($--$), COSEM($+$), E-COSEM ($\triangleright$) and RAMLA ($\circ$).

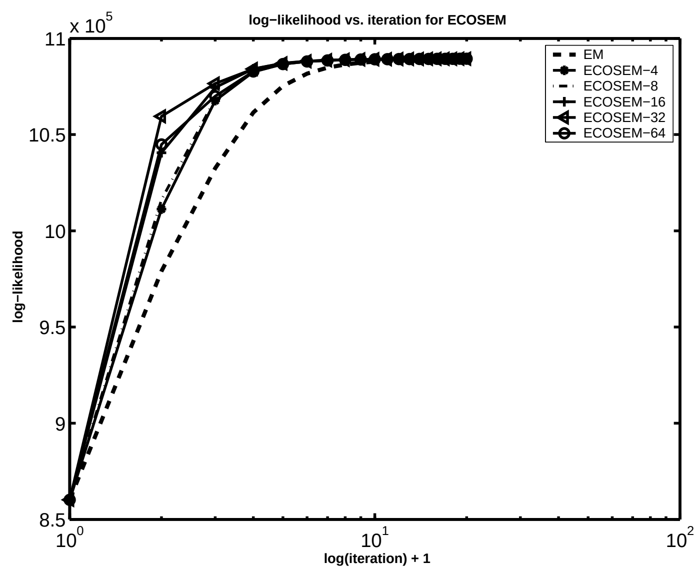

Figure 3. This shows the plots of incomplete data log-likelihood vs. iteration for the E-COSEM reconstruction with various subset numbers of 1, 4, 8, 16, 32 and 64.

4. Discussion

We have enhanced our COSEM-ML algorithm to a faster E-COSEM-ML version that uses the estimated parameter $\hat{\alpha}^{(k,l)}$, specified in equations (27) and (28). The resulting speed is close to that of RAMLA, but E-COSEM avoids the need for a user-specified relaxation schedule. A single iteration (one outer-loop k iteration) of COSEM has the same computational complexity as EM-ML. But for E-COSEM, the extra step of

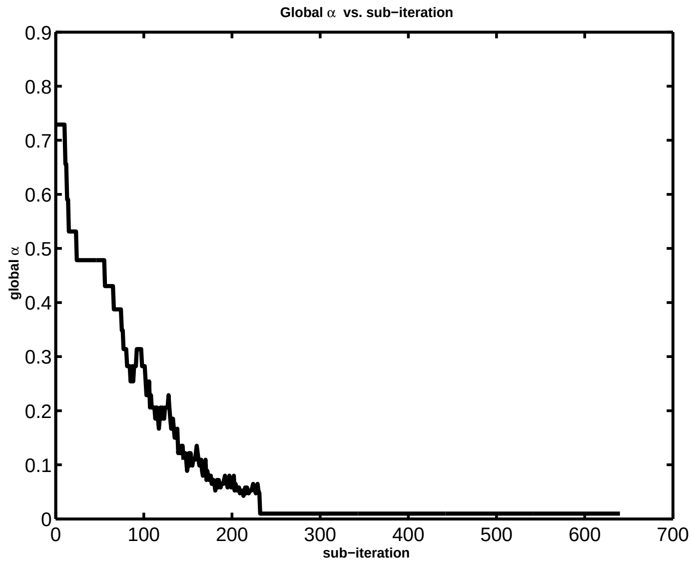

Figure 4. This shows the plots of global α vs. sub-iteration E-COSEM. Here sub-iteration index is defined as $L*(k-1)+l$, $L=32, k=1,..,20$, and $l=1,..,L$. The value of α decreased to the lower limit, 0.0096 (it becomes COSEM), after sub-iteration 250 ($k=8$).

searching for a suitable value of α can add to the computational burden per iteration relative to EM-ML. We currently see an empirical increase of roughly 20% per iteration. However, our crude α-search step can be modified using closed-form Taylor series approximations so that the extra computation time is negligible.

By adjusting $\alpha^{(k,l)}$, we have captured the intuitive notion of using the fast OSEM algorithm in the early stages of a reconstruction, and then switching to a slower algorithm that ensures convergence. To show that E-COSEM indeed captures this notion, we plot $\hat{\alpha}^{(k,l)}$ vs. subiteration in figure 4. (Since $L=32$, every 32 subiterations correspond to a single outer-loop k iteration.) As seen in figure 4, $\hat{\alpha}^{(k,l)}$ begins near unity, so that E-COSEM resembles OSEM, then eventually drops to nearly zero, so that E-COSEM becomes equivalent to COSEM. (The limiting value of $\hat{\alpha}^{(k,l)}$ was 0.0096 due to the finite number of steps in the α update.) Though not shown, ECOSEM at lower values of L will result in $\alpha=1$ (i.e. pure OSEM) for the first few outer loop iterations. Thus our intuition is borne out.

In section I, we motivated the development of E-COSEM-ML as a stepping stone towards an E-COSEM-MAP algorithm. We have already extended (Hsiao *et al* 2002b) COSEM-ML ($\alpha=0$) to the MAP case, by deriving a convenient update using a separable surrogate form of a prior. The speed of COSEM-MAP was slightly slower than that of optimized BSREM. It appears plausible that by introducing a speed-enhancing $\hat{\alpha}^{(k,l)}$ factor to COSEM-MAP, we can derive an E-COSEM-MAP algorithm competitive with those in (Ahn and Fessler 2003, De Pierro and Yamagishi 2001), but without the inconvenience of a user-specified relaxation schedule.

We note also that there is considerable latitude in the choice of update schemes like

(20) that incorporate α. We are exploring alternatives to our simple linear combination rule in (20) to see if faster algorithms are possible.

Acknowledgments

This work was supported by Grant R01-EB02629 from NIH-NIBIB, and Grants NSC91-2320-B-182-036 and NSC92-2614-B-182-037 from NSC, Taiwan, and CMRPD32013 from the Research Fund of Chang Gung Memorial Hospital, Taiwan.

Appendix

In this appendix, we provide an outline of the COSEM-ML proof, referring the reader to (Rangarajan *et al* 2003) for details. This reference is available online. We will utilize the notation $E_{\text{complete}}(\mathbf{f}, \mathbf{C})$ to indicate the objective function

$$E_{\text{complete}}(\mathbf{f}, \mathbf{C}) = \sum_{l=1}^{L} \sum_{i \in S_l} \sum_{j=1}^{N} C_{ij} \log \frac{C_{ij}}{\mathcal{H}_{ij} f_j} + \sum_{ij} \mathcal{H}_{ij} f_j - \sum_{ij} C_{ij} \tag{31}$$

This is equation (12) without the Lagrange term but with an extra term, $-\sum_{ij} C_{ij}$, added, that does not affect the fixed point.

Denote the C_{ij} in (4), i.e. the fixed point solution for C_{ij} in terms of $\mathbf{f}$, as $C_{ij}^{sol}(\mathbf{f})$. Then as proven in (Rangarajan *et al* 2003)

$$E_{\text{complete}}(\mathbf{f}, \mathbf{C}^{sol}(\mathbf{f})) = E_{\text{incomplete}}(\mathbf{f}) \tag{32}$$

We can show (Rangarajan *et al* 2003) that $E_{\text{complete}}(\mathbf{f}, \mathbf{C})$ is convex. To see this, observe that each term in (31) can be manipulated into a form $\phi(x, y) = x \log(\frac{x}{y}) - x + y$, where $x = C_{ij}$ and $y = \mathcal{H}_{ij} f_j$. With some algebra, it is easy to show that $\phi(x, y)$ is convex with respect to both x and y, and , hence, $E_{\text{complete}}(\mathbf{f}, \mathbf{C})$ is convex in C_{ij} and $\mathcal{H}_{ij} f_j$. Then since $\mathcal{H}_{ij}$ is a constant (independent of f_j), $E_{\text{complete}}(\mathbf{f}, \mathbf{C})$ is convex in C_{ij} and f_j.

Define $\Delta E_{\text{complete}}^{(k,l)} = E_{\text{complete}}(\mathbf{f}^{(k,l)}, \mathbf{C}^{(k,l)}) - E_{\text{complete}}(\mathbf{f}^{(k,l-1)}, \mathbf{C}^{(k,l-1)})$. Then in (Rangarajan *et al* 2003), we show that via the COSEM-ML algorithm, $\Delta E_{\text{complete}}^{(k,l)} \leq 0$. At $\Delta E_{\text{complete}}^{(k,l)} = 0$, for all subiterations l, we reach a fixed point $(\hat{\mathbf{f}}, \hat{\mathbf{C}})$. That is, all C_{ij}, f_j stop changing with k, l. Furthermore $\hat{\mathbf{C}} = \mathbf{C}^{sol}(\hat{\mathbf{f}})$.

Since E_{complete} is convex and since COSEM-ML is a form of grouped coordinate descent, it turns out that $(\hat{\mathbf{f}}, \hat{\mathbf{C}})$ is, in fact, an element of the set Γ of global minima of (12). This $\hat{\mathbf{f}}$ can also be shown to be a global minimum $\mathbf{f}^*$ of $E_{\text{incomplete}}(\mathbf{f})$ (Rangarajan *et al* 2003). We conclude $\hat{\mathbf{f}} = \mathbf{f}^*$.

References

Ahn S and Fessler J A 2003 Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms. *IEEE Trans. Med. Imag.* **22** 5 613-626.

Browne J and De Pierro A 1996 A row-action alternative to the EM algorithm for maximizing likelihoods in emission tomography *IEEE Trans. Med. Imag.* 15 5 687-699.

Gunawardana A 2001 *The Information Geometry of EM Variants for Speech and Image Processing* PhD thesis, Dept of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD 21218.

Hathaway R J 1986 Another interpretation of the EM algorithm for mixture distributions. *Stat. Prob. Letters* **4** 53-56

Hsiao I T, Rangarajan A, and Gindi G 2002a A provably convergent OS-EM like reconstruction algorithm for emission tomography *Proc. SPIE Medical Imaging Conference* **4684** 10-19

Hsiao I T, Rangarajan A, and Gindi G 2002b A new convergent MAP reconstruction algorithm for emission tomography using ordered subsets and separable surrogates. In *Conf. Rec. IEEE Int. Symp. Biomed. Imaging* 409-412

Hudson H M and Larkin R S 1994 Accelerated image reconstruction using ordered subsets of projection data. *IEEE Trans. Med. Imag.* **13** 4 601-609

Karush W 1939 *Minima of functions of several variables with inequalities as side conditions* PhD thesis, Dept. of Mathematics, Univ. of Chicago, Chicago, IL.

Kuhn H and Tucker A 1950 Nonlinear programming In *Proc. of the Second Berkeley Symposium on Mathematical Statistics and Probability* 481-492

Lee M 1994 *Bayesian Reconstruction In Emission Tomography Using Gibbs Priors.* PhD thesis, Yale University, New Haven, CT.

Matej S, Daube-Witherspoon M, and Karp J 2001 Performance of 3D RAMLA with smooth basis functions on fully 3D PET data In *Conf. Rec. Int. Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine* 193-196

Neal R and Hinton G 1998 A view of the EM algorithm that justifies incremental, sparse, and other variants. In M. I. Jordan, editor, *Learning in Graphical Models.* Kluwer,

De Pierro A R and Yamagishi M E 2001 Fast EM-like methods for maximum a posteriori estimates in emission tomography. *IEEE Trans. Med. Imag.* **20** 4 280-288

Rangarajan A, Hsiao I T, and Gindi G 2000 A Bayesian joint mixture framework for the integration of anatomical information in functional image reconstruction. *Journal of Mathematical Imaging and Vision* **12** 3 199-217.

Rangarajan A, Khurd P, Hsiao I T, and Gindi G 2003 Convergence proofs for the COSEM-ML and COSEM-MAP algorithms. Technical Report MIPL-03-1, Medical Image Processing Lab, Dept of Radiology, SUNY Stony Brook. Web address http://www.mil.sunysb.edu/mipl/publications.html

Rangarajan A, Lee S J, and Gindi G 1996 Mechanical models as priors in Bayesian tomographic reconstruction. In *Maximum Entropy and Bayesian Methods* 117-124