Fast Globally Convergent Reconstruction in Emission Tomography Using COSEM, an Incremental EM Algorithm

Ing-Tsung Hsiao[1], Anand Rangarajan[2],
Parmeshwar Khurd[3] and Gene Gindi[3]*
[1]School of Medical Technology, Chang-Gung University, 333 Tao-yuan, Taiwan.
[2]Dept. of CISE, University of Florida, Gainesville, FL, USA.
[3]Depts. of Radiology and Electrical Engineering, SUNY at Stony Brook, Stony Brook, NY, USA.
ihsiao@mail.cgu.edu.tw, *anand@cise.ufl.edu*, {*pkhurd,gindi*}*@clio.mil.sunysb.edu*

Abstract

We present globally convergent incremental EM algorithms for reconstruction in emission tomography, COSEM-ML for maximum likelihood and COSEM-MAP for maximum *a posteriori* reconstruction. The COSEM (Complete data Ordered Subsets Expectation Maximization) algorithms use ordered subsets (OS) for fast convergence, but unlike other globally convergent OS-based ML and MAP algorithms such as RAMLA (Browne and De Pierro, 1996), BSREM (De Pierro and Yamagishi, 2001) and modified BSREM and relaxed OS-SPS (Ahn and Fessler, 2003), COSEM does not require a user-specified object-dependent relaxation schedule. For the ML case, the COSEM-ML algorithm was independently derived previously (Gunawardana, 2001), but our theoretical approach differs. We present convergence proofs for COSEM-ML and COSEM-MAP and we also demonstrate COSEM in SPECT simulations. The monotonicity of COSEM remains an open question. At early iterations, COSEM-ML is typically slower than RAMLA and COSEM-MAP is typically slower than optimized BSREM. For COSEM, the usual speed increase with subset number is slower than that typically observed for OS-type algorithms. We discuss how COSEM may be modified to overcome these limitations.

Index Terms

Image reconstruction, maximum likelihood reconstruction, maximum *a posteriori* reconstruction, ordered subsets, emission tomography.

This work was supported by NIH-NIBIB EB02629. * Contact information: G. Gindi, email:gindi@clio.mil.sunysb.edu

I. INTRODUCTION

Statistical reconstruction is useful in emission computed tomography (ECT) due to its ability to accurately model noise, the imaging physics, and incorporate prior knowledge about the object. Statistical reconstruction approaches typically use iterative algorithms that optimize objective functions based on maximum likelihood (ML) or maximum *a posteriori* (MAP, a.k.a. penalized likelihood) principles. A main drawback of statistical reconstruction algorithms is that they are slow when used in clinical studies, especially in comparison to the commonly used filtered-backprojection (FBP) algorithm. One way to address this drawback is to devise convergent ML and MAP algorithms that require few iterations to approach the fixed-point solution.

The slow convergence of statistical reconstruction algorithms has received much attention. In [1], an ordered subsets OSEM algorithm achieved considerable speedup by using only a subset of the projection data per sub-iteration. However, OSEM was heuristically motivated and was not convergent. An alternative algorithm [2], termed row-action maximum likelihood algorithm (RAMLA), used a relaxation parameter at each iteration (relaxation schedule) within a modified version of OSEM to attain convergence. The relaxation schedule had to satisfy certain properties as a prerequisite for convergence to the true ML solution. In practice, this relaxation schedule must be determined by trial and error to ensure good speed. More recently in [3], RAMLA was extended to the MAP case. The new algorithm, termed BSREM (Block Sequential Regularized EM), continued to use a user-determined relaxation schedule. In [4], a modified BSREM algorithm was presented that converged under more general conditions than BSREM. Another convergent MAP algorithm, relaxed OS-SPS (OS-Separable Paraboloidal Surrogates), was also presented in [4]. Both algorithms in [4] again used a user-determined relaxation schedule. The COSEM algorithms presented here avoid the use of relaxation schedules while maintaining convergence, but are not quite as "fast" as their relaxation-based counterparts. This issue is discussed in Sec. IX.

COSEM is a form of incremental EM algorithm [5]. We independently introduced our COSEM-ML and COSEM-MAP versions for emission tomography in [6], [7], but learned shortly after [8] that a COSEM-ML algorithm for emission tomography had been independently derived, using an approach quite different than our own, in [9]. No experimental results were presented in [9]. A further discussion of the distinctions of the work in [9] and our work appears in Sec IX.

In Section II, we state the ECT problem, and in Section III introduce the notion of a complete data energy function whose minimization also minimizes the (minus) log likelihood via COSEM-ML, an algorithm discussed in Section IV. In Section V, we add a prior to the complete data energy and formulate a COSEM-MAP algorithm. Global convergence proofs for COSEM-MAP and COSEM-ML are presented in Sections VI and VII. In Section VIII, we show simulation results, and conclude with a discussion in Section IX.

II. STATEMENT OF THE PROBLEM

We consider the object to be an N-dim lexicographically ordered vector $\mathbf{f}$ with elements $f_j, j = 1, \ldots, N$. We model image formation by a simple Poisson model $\mathbf{g} \sim Poisson(H\mathbf{f})$, where H is the system matrix whose element H_{ij} is proportional to the probability of receiving a count in detector bin i that emanates from pixel j, and $\mathbf{g}$ is the integer-valued random data vector (sinogram) whose elements are $g_i,\ i = 1, \ldots, M$.

The ML problem is written as the minimization of an objective function

$$\widehat{\mathbf{f}} = \arg\min_{\mathbf{f} \geq \mathbf{0}} E_{\mathrm{inc-ML}}(\mathbf{f}) \tag{1}$$

where the ML objective correponding to the Poisson likelihood is given by

$$\begin{aligned} E_{\mathrm{inc-ML}}(\mathbf{f}) &= -\log \Pr(\mathbf{g}|\mathbf{f}) \\ &= \sum_{ij} H_{ij} f_j - \sum_i g_i \log \sum_j H_{ij} f_j. \end{aligned} \tag{2}$$

We have used the subscript "inc" to anticipate the use of the observed or "incomplete" data, as opposed to the "complete" data that we utilize later.

The MAP problem may be similarly stated as the minimization

$$\widehat{\mathbf{f}} = \arg\min_{\mathbf{f} \geq \mathbf{0}} E_{\mathrm{inc-MAP}}(\mathbf{f}) \tag{3}$$

where the MAP objective is given by

$$E_{\mathrm{inc-MAP}}(\mathbf{f}) = E_{\mathrm{inc-ML}}(\mathbf{f}) + E_{\mathrm{prior}}(\mathbf{f}) \tag{4}$$

where $E_{\mathrm{prior}}(\mathbf{f})$ denotes the objective function corresponding to the prior distribution on the object. We consider quadratic priors (regularizers) of the form

$$E_{\mathrm{prior}}(\mathbf{f}) = -\log\ p(\mathbf{f}) = \beta \sum_j \sum_{j' \in \mathcal{N}(j)} w_{jj'} (f_j - f_{j'})^2 \tag{5}$$

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 $w_{jj'}$ are positive.

Our aim is to derive fast algorithms for the minimizations in (1) and (3) and to prove their convergence to the ML or MAP solutions. We note that (1) and (3) express the natural constraint $\mathbf{f} \geq \mathbf{0}$. However, we shall use a slightly modified constraint $\mathbf{f} > \mathbf{0}$. The reason for this is that the COSEM-MAP and COSEM-ML algorithms will allow an initially positive estimate of f_j to approach vanishingly close to zero as iterations proceed, but never to equal zero. The well known EM-ML algorithm for ECT has this same behavior.

III. Complete data objective for the ML Problem

As a step towards deriving COSEM-ML, we consider an unconventional approach towards solving the minimization of $E_{\mathrm{inc-ML}}$. This approach involves minimizing a "complete data objective" defined as:

$$\begin{aligned} E_{\mathrm{cmp-ML}}(\mathcal{C}, \mathbf{f}, \nu) = &-\sum_{ij} \mathcal{C}_{ij} \log H_{ij} f_j + \sum_{ij} H_{ij} f_j + \\ &\sum_{ij} \mathcal{C}_{ij} \log \mathcal{C}_{ij} + \sum_i \nu_i (\sum_j \mathcal{C}_{ij} - g_i) \end{aligned} \tag{6}$$

where the subscript "cmp" means "complete". Here, $\mathcal{C}_{ij}$ (to be discussed in more detail shortly) is the complete data, roughly analogous to the complete data as used in statistical derivations [10], [11] of ML-EM for ECT. It turns out that $\mathcal{C}_{ij}$ is real and positive, and that it obeys the constraint $\sum_j \mathcal{C}_{ij} = g_i$. In (6), this constraint is imposed via Lagrange parameters ν_i. It will be quite convenient to freely intermix notations $E_{\mathrm{cmp-ML}}(\mathcal{C}, \mathbf{f}, \nu)$ with $E_{\mathrm{cmp-ML}}(\mathcal{C}, \mathbf{f})$. The latter notation means that the constraint implicitly holds and does not appear in the objective in a Lagrange form. In the remainder of this section, our goal is to show that joint (on $\mathcal{C}$ and $\mathbf{f}$) minimization of (6) yields a solution for $\mathbf{f}$ that is also the solution to the ML problem (1). We derive (6) from (2) using convexity arguments and change of variables transformations.

We begin with (2) and selectively replace terms involving $\log \sum_j H_{ij} f_j$. Since $-\log(\cdot)$ is convex, we have from Jensen's inequality [12] $-\log(\sum_j \alpha_j x_j) \leq -\sum_j \alpha_j \log x_j$ where $\alpha_j \geq 0, \forall j$, $\sum_j \alpha_j = 1$ and $x_j > 0, \forall j$. Then we may write

$$-\log \sum_j H_{ij} f_j \leq -\sum_j \gamma_{ij} \log \frac{H_{ij} f_j}{\gamma_{ij}} \tag{7}$$

where $\gamma_{ij} \geq 0, \forall ij$ and $\sum_j \gamma_{ij} = 1, \forall j$ and with equality occurring at $\gamma_{ij} = \frac{H_{ij} f_j}{\sum_{j'} H_{ij'} f_{j'}}$. With (7) in place, we may define a new objective function, containing γ as an independent variable, as

$$E(\gamma, \mathbf{f}) = \sum_{ij} H_{ij} f_j - \sum_{ij} g_i \gamma_{ij} \log \frac{H_{ij} f_j}{\gamma_{ij}} \tag{8}$$

with the constraints $\sum_j \gamma_{ij} = 1, \forall i$ and $\gamma_{ij} \geq 0, \forall ij$. We now show that the new objective (8) when minimized solely w.r.t. γ while satisfying its constraints attains its minimum at $\gamma^\star_{ij} = \frac{H_{ij} f_j}{\sum_{j'} H_{ij'} f_{j'}}$. If the objective function $E(\gamma, \mathbf{f})$ attained its minimum at a point $\check{\gamma} \neq \gamma^\star_{ij}$, then $E(\check{\gamma}, \mathbf{f}) < E(\gamma^\star, \mathbf{f})$. But this violates the inequality in (7). (The equality $E(\check{\gamma}, \mathbf{f}) = E(\gamma^\star, \mathbf{f})$ is allowed since $\gamma^\star$ would still be a point at which $E(\gamma, \mathbf{f})$ attained its minimum). At the minimum, we have

$$E(\gamma^\star, \mathbf{f}) = E_{\mathrm{inc-ML}}(\mathbf{f}). \tag{9}$$

If we define $\mathcal{C}_{ij} \stackrel{\mathrm{def}}{=} g_i \gamma_{ij}$ and rewrite (8), we get a further transformation of (2):

$$E(\mathcal{C}, \mathbf{f}) = \sum_{ij} H_{ij} f_j - \sum_{ij} \mathcal{C}_{ij} \log \frac{g_i H_{ij} f_j}{\mathcal{C}_{ij}}. \tag{10}$$

with the former constraints modified to $\sum_j \mathcal{C}_{ij} = g_i, \forall i$ and $\mathcal{C}_{ij} \geq 0, \forall ij$. Using the new constraints, the objective in (10) is modified to

$$\begin{aligned} E(\mathcal{C}, \mathbf{f}) = \sum_{ij} H_{ij} f_j - \sum_{ij} \mathcal{C}_{ij} \log H_{ij} f_j + \sum_{ij} \mathcal{C}_{ij} \log \mathcal{C}_{ij} \\ - \sum_i g_i \log g_i \end{aligned} \tag{11}$$

where we have used the constraint $\sum_j \mathcal{C}_{ij} = g_i$ to modify the term $\sum_{ij} \mathcal{C}_{ij} \log g_i$ to $\sum_i g_i \log g_i$. Dropping terms in (11) which are independent of *both* $\mathbf{f}$ and $\mathcal{C}$ and using a Lagrange parameter ν to express the constraints $\sum_j \mathcal{C}_{ij} = g_i, \forall i$, we get (6).

If we define $\mathcal{C}_{ij}^{sol}(\mathbf{f}) = g_i \gamma_{ij}^{\star}$, then it follows that

$$\mathcal{C}_{ij}^{\mathrm{sol}}(\mathbf{f}) \quad = \quad g_i \frac{H_{ij} f_j}{\sum_n H_{in} f_n} \ \forall i, \ \forall j. \tag{12}$$

Then when (11) is minimized w.r.t. $\mathcal{C}$ while satisfying its constraints, it attains its minimum at $\mathcal{C}^{sol}(\mathbf{f})$ and at the minimum, we have the "touching condition"

$$E_{\mathrm{cmp-ML}}(\mathcal{C}^{\mathrm{sol}}(\mathbf{f}), \mathbf{f}) = E_{\mathrm{inc-ML}}(\mathbf{f}) \tag{13}$$

which follows from (9). Therefore, joint minimization of $E_{\mathrm{cmp-ML}}(\mathcal{C}, \mathbf{f})$ will yield a fixed point $(\mathcal{C}^*, \mathbf{f}^*)$, where $\mathbf{f}^* = \hat{\mathbf{f}}$ is the solution sought in (1), i.e. the fixed point of (1) is preserved.

This approach of using Jensen's inequality to derive a new complete data objective is similar to the pioneering work of De Pierro [13]. The complete data objective function is also very similar to new objective functions first derived in Appendix B of [14] and identical to the new complete data objective functions derived in [15] and in [16].

The material in this section simply motivates the use of a complete data energy for the ML case. In later sections, an algorithm, COSEM-ML, for minimizing (6) and convergence proof are given. We shall also offer a MAP version of the complete data objective, and show an appropriate algorithm, COSEM-MAP, and convergence proof for it.

IV. COSEM-ML ALGORITHM

To incorporate subsets, we rewrite (6) using subset notation. Assume that we have L non-overlapping subsets with the data in each subset S_l denoted as $\{g_i, \forall i \in S_l\}$ with $l \in \{1, \ldots, L\}$. We also have a corresponding division of $\mathcal{C}$ denoted by $\{\mathcal{C}_{ij}, \forall i \in S_l, \forall j\}$. Then (6) can be re-written using this subset notation as

$$\begin{aligned} E_{\mathrm{cmp-ML}}(\mathcal{C}, \mathbf{f}, \nu) = & -\sum_{l=1}^{L} \sum_{i \in S_l} \sum_{j} \mathcal{C}_{ij} \log H_{ij} f_j + \sum_{ij} H_{ij} f_j \\ & + \sum_{l=1}^{L} \sum_{i \in S_l} \sum_{j} \mathcal{C}_{ij} \log \mathcal{C}_{ij} - \sum_{l=1}^{L} \sum_{i \in S_l} \sum_{j=1}^{N} \mathcal{C}_{ij} + \sum_i \nu_i (\sum_j \mathcal{C}_{ij} - g_i). \end{aligned} \tag{14}$$

An extra term, $\sum_{l=1}^{L} \sum_{i \in S_l} \sum_{j=1}^{N} \mathcal{C}_{ij}$, has been added in (14). It does not change the minima since this term equals an additive constant $\sum_{l=1}^{L} \sum_{i \in S_l} g_i$ provided the constraints are satisfied.

We now embark upon an ordered subsets minimization strategy. As with standard OS-EM-like approaches, the iterations are divided into subiterations using an outer/inner loop structure. In the outer k loop, we assume that all subsets have been updated. In each inner loop (k, l) subiteration, we update $\{\mathcal{C}_{ij}, \forall i \in S_l, \forall j\}$ and $\{f_j, \forall j\}$. The update equations can be obtained by directly differentiating (14) w.r.t. $\mathcal{C}_{ij}$ and f_j and setting the result to zero while satisfying the constraints on $\mathcal{C}$. Adding iteration superscripts to the resulting relations result in a grouped coordinate descent algorithm.

We first differentiate w.r.t $\mathcal{C}_{ij}, \ j = 1, \cdots, N; \ i \in S_l$ to get

$$\frac{\partial E_{\mathrm{cmp-ML}}(\mathcal{C}, \mathbf{f})}{\partial \mathcal{C}_{ij}} = -\log(H_{ij} f_j) + \ln \mathcal{C}_{ij} + \nu_i \tag{15}$$

Setting (15) to zero, we get $\mathcal{C}_{ij} = \exp(-\nu_i) H_{ij} f_j$. By enforcing the constraint $\sum_j \mathcal{C}_{ij} = g_i$, we may solve for the term $\exp(-\nu_i)$ and using this solution, we obtain the update:

$$\mathcal{C}_{ij} = g_i \frac{H_{ij} f_j}{\sum_n H_{in} f_n} \tag{16}$$

We differentiate w.r.t. $f_j,\ j = 1, \cdots, N$ to get:

$$\frac{\partial E_{\mathrm{cmp-ML}}(\mathcal{C}, \mathbf{f})}{\partial f_j} = -\sum_{m=1}^{L} \sum_{i \in S_m} \mathcal{C}_{ij} \frac{1}{f_j} + \sum_i H_{ij} \tag{17}$$

Here, we have replaced the subset index l in (14) by m since we will be using l as a sub-iteration index later. Equating (17) to zero, we get the update:

$$f_j = \frac{\sum_{m=1}^{L} \sum_{i \in S_m} \mathcal{C}_{ij}}{\sum_i H_{ij}},\ \forall j \tag{18}$$

The set of update equations with the appropriate iteration indices are summarized below

$$\begin{aligned}
\mathcal{C}_{ij}^{(k,l)} &= g_i \frac{H_{ij} f_j^{(k,l-1)}}{\sum_n H_{in} f_n^{(k,l-1)}},\ \forall i \in S_l,\ \forall j && (19)\\
\mathcal{C}_{ij}^{(k,l)} &= \mathcal{C}_{ij}^{(k,l-1)},\ \forall i \notin S_l,\ \forall j && (20)\\
f_j^{(k,l)} &= \frac{\sum_{m=1}^{L} \sum_{i \in S_m} \mathcal{C}_{ij}^{(k,l)}}{\sum_i H_{ij}},\ \forall j. && (21)
\end{aligned}$$

We clarify the notation used. The symbols $\mathcal{C}_{ij}^{(k,l)}$ and $f_j^{(k,l)}$ denote the updates of $\mathcal{C}_{ij}$ and f_j at outer iteration k and subset iteration l. The subtlety here is that at iteration (k,l), we only update $\mathcal{C}_{ij}$ over all $i \in S_l$ and all $j \in \{1, \ldots, N\}$ as shown in (19). However, the update of the $f_j^{(k,l)}$ in (21) requires the summation over all the complete data $\mathcal{C}_{ij}^{(k,l)}, \forall i \in S_m, \forall m \in \{1, \ldots, L\}, \forall j \in \{1, \ldots, N\}$. Due to this, we *define* the the "copy" operation $\mathcal{C}_{ij}^{(k,l)} = \mathcal{C}_{ij}^{(k,l-1)}, \forall i \notin S_l, \forall j$ as shown in (20).

We can streamline (19) - (21) by the introduction of a "bookkeeping" scheme. We note that the summation $\sum_{m=1}^{L} \sum_{i \in S_m} \mathcal{C}_{ij}^{(k,l)}$ can be divided into two disjoint subsets

$$\begin{aligned}
\sum_{m=1}^{L} \sum_{i \in S_m} \mathcal{C}_{ij}^{(k,l)} &= \sum_{i \in S_l} \mathcal{C}_{ij}^{(k,l)} + \sum_{i \notin S_l} \mathcal{C}_{ij}^{(k,l)} && (22)\\
&= \sum_{i \in S_l} (\mathcal{C}_{ij}^{(k,l)} - \mathcal{C}_{ij}^{(k,l-1)}) + \sum_{i \in S_l} \mathcal{C}_{ij}^{(k,l-1)} + \sum_{i \notin S_l} \mathcal{C}_{ij}^{(k,l-1)}
\end{aligned}$$

since $\mathcal{C}_{ij}^{(k,l)} = \mathcal{C}_{ij}^{(k,l-1)}, \forall i \notin S_l, \forall j$. We define $B_j^{(k,l)} \stackrel{\text{def}}{=} \sum_i \mathcal{C}_{ij}^{(k,l)} = \sum_{m=1}^{L} \sum_{i \in S_m} \mathcal{C}_{ij}^{(k,l)}$. Combining the last two terms $\sum_{i \in S_l} \mathcal{C}_{ij}^{(k,l-1)}$ and $\sum_{i \notin S_l} \mathcal{C}_{ij}^{(k,l-1)}$, we get

$$B_j^{(k,l)} = \sum_{i \in S_l} (\mathcal{C}_{ij}^{(k,l)} - \mathcal{C}_{ij}^{(k,l-1)}) + B_j^{(k,l-1)} \tag{23}$$

We define $\mathcal{C}_{ij}^{(k+1,0)} \stackrel{\text{def}}{=} \mathcal{C}_{ij}^{(k,L)}$ and $B_j^{(k+1,0)} \stackrel{\text{def}}{=} B_j^{(k,L)}$. Similarly, we define $\mathbf{f}^{(k+1,0)} \stackrel{\text{def}}{=} \mathbf{f}^{(k,L)}$. Define $D_j \stackrel{\text{def}}{=} \sum_i H_{ij}$ as the sensitivity at voxel j.

With these notations and definitions, (19)-(21) becomes our COSEM-ML algorithm which is summarized below.

- **The COSEM-ML Algorithm**
- Initialize $\{f_j^{(0,0)} = f_j^{\text{init}}, \forall j \in \{1,\dots,N\}\}$ where $f_j^{\text{init}} \in (0,\infty), \forall j$
- Initialize $\{\mathcal{C}_{ij}^{(0,0)}, \forall i \in S_m, \forall m \in \{1,\dots,L\}$ and $\forall j \in \{1,\dots,N\}\}$ by $\mathcal{C}_{ij}^{(0,0)} = g_i \frac{H_{ij} f_j^{(0,0)}}{\sum_n H_{in} f_n^{(0,0)}}$
- $B_j^{(0,0)} = \sum_{m=1}^{L} \sum_{i \in S_m} \mathcal{C}_{ij}^{(0,0)}, \forall j$
- **Begin** *k-loop* $[k \in \{0,1,\dots\}]$

  $\mathcal{C}_{ij}^{(k,0)} = \mathcal{C}_{ij}^{(k-1,L)}, \forall ij$ and $k > 0$.
  $B_j^{(k,0)} = B_j^{(k-1,L)}, \forall j$ and $k > 0$.
  $f_j^{(k,0)} = f_j^{(k-1,L)}, \forall j$ and $k > 0$.
  Begin *l-loop* $[l \in \{1,\dots,L\}]$
  * $\mathcal{C}_{ij}^{(k,l)} = g_i \frac{H_{ij} f_j^{(k,l-1)}}{\sum_n H_{in} f_n^{(k,l-1)}}, \forall i \in S_l, \forall j$.
  * $\mathcal{C}_{ij}^{(k,l)} = \mathcal{C}_{ij}^{(k,l-1)}, \forall i \notin S_l, \forall j$
  * $B_j^{(k,l)} = \sum_{i \in S_l} (\mathcal{C}_{ij}^{(k,l)} - \mathcal{C}_{ij}^{(k,l-1)}) + B_j^{(k,l-1)}, \forall j$
  * $f_j^{(k,l)} = \frac{B_j^{(k,l)}}{D_j}, \forall j \in \{1,\dots,N\}$.

  End *l-loop*
- **End** *k-loop*

Note that the $f_j^{(k,l)}$ updates preserve positivity. Also, the COSEM-ML updates are parallel in the voxel space. The implementation of $B_j^{(k,l)}$ reduces, relative to (19)-(21), the overall computation per (k,l) subiteration to that of OSEM. If one knows *a priori* that some f_j are zero, then these can be fixed to zero and eliminated from the problem. We note that for $g_i = 0$, $f_j = 0$ and $H_{ij} = 0$, the corresponding $\mathcal{C}_{ij}$ will remain zero throughout their update. Therefore, we eliminate these $\mathcal{C}_{ij}$'s from the problem.

V. COSEM-MAP Algorithm

We extend (14) to MAP by adding the prior (5):

$$E_{\text{cmp-MAP}}(\mathcal{C}, \mathbf{f}, \nu) = \sum_{ij} H_{ij} f_j + \sum_{l=1}^{L} \sum_{i \in S_l} \sum_{j=1}^{N} \mathcal{C}_{ij} \log \frac{\mathcal{C}_{ij}}{H_{ij} f_j} - \sum_{l=1}^{L} \sum_{i \in S_l} \sum_{j=1}^{N} \mathcal{C}_{ij} + \beta \sum_{jj'} w_{jj'} (f_j - f_{j'})^2 + \sum_i \nu_i (\sum_j \mathcal{C}_{ij} - g_i) \quad (24)$$

We now derive a constrained grouped coordinate descent algorithm for (24).

Equations (24) and (14) are the same objectives w.r.t. $\mathcal{C}$ alone. Therefore, the $\mathcal{C}$-update for MAP is the same as (19)-(20) for the ML case. The prior (5) introduces a coupling between source voxels. Since we seek, as in COSEM-ML, a positivity-preserving parallel-update of all voxels, this coupling presents a problem. To solve this problem, we use the method of separable surrogates [17]. Separable surrogate priors are separable in $\mathbf{f}$ and depend on the current estimate $\mathbf{f}^{(k,l-1)}$. Therefore, they have the general form $E_{\text{s-prior}}(f_j; \mathbf{f}^{(k,l-1)})$ where “s” stands for surrogate. To maintain convergence, we need [17]-[18] to satisfy the following conditions: (A) $E_{\rm prior}(\mathbf{f}) \leq \sum_j E_{\rm s-prior}(f_j; \mathbf{f}^{(k,l-1)})$ and (B) $E_{\rm prior}(\mathbf{f}^{(k,l-1)}) = \sum_j E_{\rm s-prior}(f_j^{(k,l-1)}; \mathbf{f}^{(k,l-1)})$.

We construct the surrogate to have the properties (A) and (B). At step (k,l), assuming that we have $\mathbf{f}^{(k,l-1)}$, we observe that [17], [19]

$$(f_j - f_{j'})^2 \leq \frac{1}{2}\Big[(2f_j - f_j^{(k,l-1)} - f_{j'}^{(k,l-1)})^2 + (2f_{j'} - f_j^{(k,l-1)} - f_{j'}^{(k,l-1)})^2\Big] \tag{25}$$

Then the update for $\mathbf{f}$ at step (k,l) may be obtained by the minimization of the following objective:

$$E_{\rm cmp-MAP-fs}(\mathbf{f}; \mathbf{f}^{(k,l-1)}, \mathcal{C}^{(k,l)}) \stackrel{\rm def}{=} -\sum_{m=1}^{L}\sum_{i\in S_m}\sum_{j=1}^{N} \mathcal{C}_{ij}^{(k,l)} \log f_j + \sum_{ij} H_{ij} f_j + \frac{\beta}{2}\sum_{jj'} w_{jj'}\Big[(2f_j - f_j^{(k,l-1)} - f_{j'}^{(k,l-1)})^2 + (2f_{j'} - f_j^{(k,l-1)} - f_{j'}^{(k,l-1)})^2\Big]. \tag{26}$$

The subscript "cmp-MAP-fs" and the arguments of $E_{\rm cmp-MAP-fs}$ indicate that it is obtained from (24) by restricting the argument to $\mathbf{f}$ ($\mathcal{C} = \mathcal{C}^{(k,l)}$ is constant) and that the prior is now a surrogate at $\mathbf{f}^{(k,l-1)}$. Noting that (26) is separable w.r.t. f_j, the objective w.r.t. each f_j can be written as:

$$E_{\rm cmp-MAP-fs}(f_j; \mathbf{f}^{(k,l-1)}, \mathcal{C}^{(k,l)}) = -\sum_i \mathcal{C}_{ij}^{(k,l)} \log f_j + \sum_i H_{ij} f_j + \frac{\beta}{2}\sum_{j'} v_{jj'}(2f_j - f_j^{(k,l-1)} - f_{j'}^{(k,l-1)})^2 \tag{27}$$

where $v_{jj'} \stackrel{\rm def}{=} (w_{jj'} + w_{j'j})$. Setting the derivative of (27) to zero and solving for the positive root of the resulting quadratic equation, we obtain:

$$f_j^{(k,l)} = \frac{1}{8\beta\sum_{j'} v_{jj'}}\left(2\beta\sum_{j'} v_{jj'}(f_j^{(k,l-1)} + f_{j'}^{(k,l-1)}) - D_j + \sqrt{[2\beta\sum_{j'} v_{jj'}(f_j^{(k,l-1)} + f_{j'}^{(k,l-1)}) - D_j]^2 + 16\beta B_j^{(k,l)}\sum_{j'} v_{jj'}}\right). \tag{28}$$

We now collect our results and summarize the COSEM-MAP algorithm below:

- **The COSEM-MAP Algorithm**
- Initialize $\{f_j^{(0,0)} = f_j^{\text{init}}, \forall j \in \{1,\ldots,N\}\}$ where $f_j^{\text{init}} \in (0,\infty), \forall j$
- Initialize $\{\mathcal{C}_{ij}^{(0,0)}, \forall i \in S_l, \forall l \in \{1,\ldots,L\}$ and $\forall j \in \{1,\ldots,N\}\}$ by $\mathcal{C}_{ij}^{(0,0)} = g_i \frac{H_{ij} f_j^{(0,0)}}{\sum_n H_{in} f_n^{(0,0)}}$
- $B_j^{(0,0)} = \sum_{m=1}^{L} \sum_{i \in S_m} \mathcal{C}_{ij}^{(0,0)}, \forall j$
- **Begin** *k-loop* $[k \in \{0,1,\ldots\}]$

  $\mathcal{C}_{ij}^{(k,0)} = \mathcal{C}_{ij}^{(k-1,L)}, \forall ij \text{ and } k > 0.$

  $B_j^{(k,0)} = B_j^{(k-1,L)}, \forall j \text{ and } k > 0.$

  $f_j^{(k,0)} = f_j^{(k-1,L)}, \forall j \text{ and } k > 0.$

  Begin *l-loop* $[l \in \{1,\ldots,L\}]$

  * $\mathcal{C}_{ij}^{(k,l)} = g_i \frac{H_{ij} f_j^{(k,l-1)}}{\sum_n H_{in} f_n^{(k,l-1)}}, \forall i \in S_l, \forall j$.
  * $\mathcal{C}_{ij}^{(k,l)} = \mathcal{C}_{ij}^{(k,l-1)}, \forall i \notin S_l, \forall j$
  * $B_j^{(k,l)} = \sum_{i \in S_l} (\mathcal{C}_{ij}^{(k,l)} - \mathcal{C}_{ij}^{(k,l-1)}) + B_j^{(k,l-1)}, \forall j$
  * $f_j^{(k,l)} = \frac{1}{8\beta \sum_{j'} v_{jj'}} \Big[2\beta \sum_{j'} v_{jj'} (f_j^{(k,l-1)} + f_{j'}^{(k,l-1)}) - D_j + \sqrt{[2\beta \sum_{j'} v_{jj'} (f_j^{(k,l-1)} + f_{j'}^{(k,l-1)}) - D_j]^2 + 16\beta B_j^{(k,l)} \sum_{j'} v_{jj'}} \Big]$
    $, \forall j \in \{1,\ldots,N\}$

  End *l-loop*
- **End** *k-loop*

Note that the $f_j^{(k,l)}$ update preserves positivity and that the COSEM-MAP updates are parallel in the voxel space. The overall computation per (k,l) subiteration is close to that of OSEM. As in the COSEM-ML case, certain f_j and $\mathcal{C}_{ij}$ can be eliminated from the problem. We note that the closed form update for $f_j^{(k,l)}$ can be replaced by a 1-D optimization if the prior is non-quadratic but convex.

VI. Convergence Proof for COSEM-MAP

We now prove that the minimization of $E_{\text{cmp-MAP}}(\mathcal{C}, \mathbf{f}, \nu)$ by the COSEM-MAP algorithm yields a solution $(\mathcal{C}^*, \mathbf{f}^*)$, and that $\mathbf{f}^* = \hat{\mathbf{f}}$ is the unique solution of (3).

A. Strict convexity of the $E_{\text{inc-MAP}}$ objective

According to ([20], Lemma 1) and ([21], Theorem 1), our objective $E_{\text{inc-MAP}}(\mathbf{f})$ is strictly convex over the set $\mathcal{D}$ under the condition that $\mathbf{g}^T H \mathbf{1} \neq 0$, where $\mathbf{1}$ is a vector of $1's$. The definition of the set $\mathcal{D}$ is the set of $\mathbf{f} \geq \mathbf{0}$ such that $[H\mathbf{f}]_i > 0$ or $g_i = 0 \ \forall i$. Since $\mathbf{f} > \mathbf{0}$ in our case and $H_{ij} \geq 0$, we are in a subset of $\mathcal{D}$. The strict convexity of $E_{\text{inc-MAP}}$ will always apply in practice. This is because [4] the condition $\mathbf{g}^T H \mathbf{1} \neq 0$ is equivalent to $H^T \mathbf{g} \neq \mathbf{0}$, i.e. the backprojection of the data is a non-zero image. One could invent an H such that $H^T \mathbf{g} = \mathbf{0}$ for non-zero $\mathbf{g}$, but this H would be highly unrealistic. The strict convexity guarantees that $\hat{\mathbf{f}}$ in (3) is unique.

B. "Touching" property of $E_{\mathrm{cmp-MAP}}$ and $E_{\mathrm{inc-MAP}}$

In Sec. III, we showed (13) that the incomplete and the complete versions of the ML objective "touch" at $\mathcal{C} = \mathcal{C}^{\mathrm{sol}}(\mathbf{f})$, where $\mathcal{C}^{\mathrm{sol}}$ is as given in (12). If we add a prior to $E_{\mathrm{cmp-ML}}$ and to $E_{\mathrm{inc-ML}}$, then it is easy to show that this "touching" property holds for the MAP versions:

$$E_{\mathrm{cmp-MAP}}(\mathcal{C}^{\mathrm{sol}}(\mathbf{f}), \mathbf{f}) = E_{\mathrm{inc-MAP}}(\mathbf{f}) \tag{29}$$

C. Convexity of the $E_{\mathrm{cmp-MAP}}(\mathcal{C}, \mathbf{f})$ objective

We will show that $E_{\mathrm{cmp-MAP}}(\mathcal{C}, \mathbf{f})$ is convex in both $\mathcal{C}$ and $\mathbf{f}$. Since the prior is convex w.r.t. $\mathbf{f}$, we need only show that the remaining expression is convex w.r.t. $(\mathcal{C}, \mathbf{f})$. The remaining expression, which is in fact $E_{\mathrm{cmp-ML}}$, involves the summation of the terms $\mathcal{C}_{ij} \log \frac{\mathcal{C}_{ij}}{H_{ij} f_j} - \mathcal{C}_{ij} + H_{ij} f_j$. Each such term is of the form of a Kullback-Leibler (KL) divergence $\phi(x, y) \stackrel{\mathrm{def}}{=} x \log \frac{x}{y} - x + y$, where we identify x with $\mathcal{C}_{ij}$ and y with $H_{ij} f_j$. Since the KL divergence is convex [22], each of the terms is convex w.r.t. $\mathcal{C}_{ij}$ and $H_{ij} f_j$. Since $H_{ij} > 0$ is a constant (we needn't consider $H_{ij} = 0$ since the corresponding $\mathcal{C}_{ij}$'s were eliminated), and since the sum of convex functions is convex, $E_{\mathrm{cmp-ML}}$ is convex w.r.t. $\mathcal{C}$ and $\mathbf{f}$. Thus, $E_{\mathrm{cmp-MAP}}$ is convex w.r.t. $\mathcal{C}$ and $\mathbf{f}$.

D. The change in $E_{\mathrm{cmp-MAP}}$ at each COSEM-MAP iteration

We define

$$\begin{aligned} \Delta E^{(k,l)}_{\mathrm{cmp-MAP}} \stackrel{\mathrm{def}}{=} {} & E_{\mathrm{cmp-MAP}}(\mathcal{C}^{(k,l-1)}, \mathbf{f}^{(k,l-1)}) \\ & - E_{\mathrm{cmp-MAP}}(\mathcal{C}^{(k,l)}, \mathbf{f}^{(k,l)}) \end{aligned} \tag{30}$$

In the Appendix, we show that the COSEM-MAP updates are descent steps in (24), i.e. $\Delta E^{(k,l)}_{\mathrm{cmp-MAP}} \geq 0$, with equality occurring if and only if $\mathcal{C}^{(k,l)}_{ij} = \mathcal{C}^{(k,l-1)}_{ij}$, $\forall i \in S_l$, $\forall j$ *and* $f^{(k,l)}_j = f^{(k,l-1)}_j$, $\forall j$.

The change in the complete data objective function between steps $(k-1, L)$ and (k, L) is

$$\begin{aligned} \Delta E^{(k)}_{\mathrm{cmp-MAP}} & \stackrel{\mathrm{def}}{=} & & E_{\mathrm{cmp-MAP}}(\mathcal{C}^{(k-1,L)}, \mathbf{f}^{(k-1,L)}) - \\ & & & E_{\mathrm{cmp-MAP}}(\mathcal{C}^{(k,L)}, \mathbf{f}^{(k,L)}), \;\; k > 0 \\ & = & & \sum_{l=1}^{L} \Delta E^{(k,l)}_{\mathrm{cmp-MAP}} \geq 0 \end{aligned} \tag{31}$$

$\Delta E^{(k)}_{\mathrm{cmp-MAP}}$ becomes equal to zero if and only if $\mathcal{C}^{(k,l)}_{ij} = \mathcal{C}^{(k,l-1)}_{ij}$, $\forall i \in S_l$, $\forall l \in \{1, \ldots, L\}$, $\forall j$ and $f^{(k,l)}_j = f^{(k,l-1)}_j$, $\forall j$, $\forall l \in \{1, \ldots, L\}$. Define k^* as the iteration at which $\Delta E^{(k^*)}_{\mathrm{cmp-MAP}} = 0$. Then $\mathcal{C}^{(k^*,1)} = \mathcal{C}^{(k^*,2)} = \cdots = \mathcal{C}^{(k^*,L)} \stackrel{\mathrm{def}}{=} \mathcal{C}^*$. We similarly get $\mathbf{f}^{(k^*,1)} = \mathbf{f}^{(k^*,2)} = \cdots = \mathbf{f}^{(k^*,L)} \stackrel{\mathrm{def}}{=} \mathbf{f}^*$. We automatically have $\mathbf{f}^{(k^*+1,0)} = \mathbf{f}^{(k^*,L)}$ and $\mathcal{C}^{(k^*+1,0)} = \mathcal{C}^{(k^*,L)}$ which means that nothing can change once we have reached k^*.

From the COSEM-MAP $\mathcal{C}$ update (19)-(20), we get:

$$\begin{aligned} \mathcal{C}^*_{ij} = \mathcal{C}^{(k^*,l)}_{ij} = g_i \frac{H_{ij} f^{(k^*,l-1)}_j}{\sum_n H_{in} f^{(k^*,l-1)}_n} \; \forall i, \forall j, \forall l \in \{1, \cdots, L\} \\ = g_i \frac{H_{ij} f^*_j}{\sum_n H_{in} f^*_n} \end{aligned} \tag{32}$$

Therefore, from (12), we get

$$\mathcal{C}^* = \mathcal{C}^{\text{sol}}(\mathbf{f}^*) \tag{33}$$

E. Convergence to the fixed point of $E_{\text{inc}-\text{MAP}}$

Define Γ as the set (possibly singleton) of global minima of $E_{\text{cmp}-\text{MAP}}(\mathcal{C}, \mathbf{f})$. We may argue that $(\mathcal{C}^*, \mathbf{f}^*) \in \Gamma$ as follows: $\Delta E^{(k,l)}_{\text{cmp}-\text{MAP}} = 0$ is equivalent to the fact that $\mathcal{C}^{(k,l)}_{ij}$ and $f^{(k,l)}_j$ are no longer changing. We note that an algorithm can lead to a repeating $(\mathcal{C}^{(k,l)}_{ij}, f^{(k,l)}_j)$ that is *not* an element of Γ. But this cannot happen because the COSEM-MAP algorithm is a grouped coordinate descent algorithm, and because $E_{\text{cmp}-\text{MAP}}$ is convex. The COSEM-MAP algorithm would continue to descend along coordinate directions whenever possible. Hence, if one approaches a condition where $(\mathcal{C}^{(k,l)}_{ij}, f^{(k,l)}_j)$ are repeating, then these are in Γ. Hence, $(\mathcal{C}^*, \mathbf{f}^*) \in \Gamma$.

We now show that $\mathbf{f}^*$ is a global minimum of $E_{\text{inc}-\text{MAP}}(\mathbf{f})$. We prove this by contradiction. Assume an $\check{\mathbf{f}}$ such that $E_{\text{inc}-\text{MAP}}(\mathbf{f}^*) > E_{\text{inc}-\text{MAP}}(\check{\mathbf{f}})$. Since by (33), $\mathcal{C}^* = \mathcal{C}^{\text{sol}}(\mathbf{f}^*)$, we have by the touching property (29) in Sec. VI-B:

$$\begin{aligned} & E_{\text{cmp}-\text{MAP}}(\mathcal{C}^{\text{sol}}(\mathbf{f}^*), \mathbf{f}^*) = E_{\text{inc}-\text{MAP}}(\mathbf{f}^*) \\ & > E_{\text{inc}-\text{MAP}}(\check{\mathbf{f}}) = E_{\text{cmp}-\text{MAP}}(\mathcal{C}^{\text{sol}}(\check{\mathbf{f}}), \check{\mathbf{f}}), \end{aligned} \tag{34}$$

but this is a contradiction because $(\mathcal{C}^{\text{sol}}(\mathbf{f}^*), \mathbf{f}^*) \in \Gamma$ is a global minimum of $E_{\text{cmp}-\text{MAP}}$. In fact, $\mathbf{f}^*$ is the *unique* global minimum $\hat{\mathbf{f}}$ of $E_{\text{inc}-\text{MAP}}(\mathbf{f})$, because $E_{\text{inc}-\text{MAP}}(\mathbf{f})$ is strictly convex, as we established in Sec. VI-A.

This completes the proof, in that we have established that $\mathbf{f}^*$, obtained from the COSEM-MAP algorithm, is indeed equal to $\hat{\mathbf{f}}$ in (3), which is the result we have sought.

VII. Convergence Proof for COSEM-ML

This proof differs from that in Sec. VI because we do not know whether the ML problem has a *unique* solution. In this section, we prove that the minimization of $E_{\text{cmp}-\text{ML}}(\mathcal{C}, \mathbf{f}, \nu)$ by the COSEM-ML algorithm yields a solution $(\mathcal{C}^*, \mathbf{f}^*)$, and that $\mathbf{f}^* = \hat{\mathbf{f}}$ is a solution to the ML problem in (1).

The convexity of of $E_{\text{cmp}-\text{ML}}$ follows from the results in Sec. VI-C.

We define $\Delta E^{(k,l)}_{\text{cmp}-\text{ML}} \stackrel{\text{def}}{=} E_{\text{cmp}-\text{ML}}(\mathcal{C}^{(k,l-1)}, \mathbf{f}^{(k,l-1)}) - E_{\text{cmp}-\text{ML}}(\mathcal{C}^{(k,l)}, \mathbf{f}^{(k,l)})$. We first show that the COSEM-ML updates are descent steps in (14), $\Delta E^{(k,l)}_{\text{cmp}-\text{ML}} \geq 0$, and that $\Delta E^{(k,l)}_{\text{cmp}-\text{ML}} = 0$ if and only if $\mathcal{C}^{(k,l)}_{ij} = \mathcal{C}^{(k,l-1)}_{ij}, \forall i \in S_l, \forall j$ *and* $f^{(k,l)}_j = f^{(k,l-1)}_j, \forall j$.

We may write (43) and (45) - (48) of the Appendix in a form appropriate for the ML problem by simply dropping any terms related to the prior. In particular (48) becomes

$$\Delta E^{(k,l)}_{\text{cmp}-\text{ML}} \geq \sum_{i \in S_l} \sum_{j=1}^{N} \mathcal{C}^{(k,l-1)}_{ij} \log \frac{\mathcal{C}^{(k,l-1)}_{ij}}{\mathcal{C}^{(k,l)}_{ij}} + \sum_{m=1}^{L} \sum_{i \in S_m} \sum_{j=1}^{N} \left[\mathcal{C}^{(k,l)}_{ij} \log \frac{f^{(k,l)}_j}{f^{(k,l-1)}_j} + H_{ij}(f^{(k,l-1)}_j - f^{(k,l)}_j) \right]. \tag{35}$$

From the Appendix , the first term on the right of (35) can be expressed as a KL divergence which equals 0 only if $\mathcal{C}_{ij}^{(k,l)} = \mathcal{C}_{ij}^{(k,l-1)}$, $\forall i \in S_l$, $\forall j$. We observe that the update (21) for $\mathbf{f}$ in COSEM-ML is chosen such that the following objective is minimized:

$$E_{\mathrm{cmp-ML-f}}(\mathbf{f};\mathcal{C}^{(k,l)}) \quad = \quad -\sum_{m=1}^{L}\sum_{i\in S_m}\sum_{j=1}^{N}\mathcal{C}_{ij}^{(k,l)}\log f_j + \sum_{ij} H_{ij} f_j. \tag{36}$$

Apart from the terms forming the KL divergence, the remaining terms on the right of (35) can be seen to be

$$\begin{aligned} &E_{\mathrm{cmp-ML-f}}(\mathbf{f}^{(k,l-1)};\mathcal{C}^{(k,l)}) - E_{\mathrm{cmp-ML-f}}(\mathbf{f}^{(k,l)};\mathcal{C}^{(k,l)}) \\ &= \sum_{m=1}^{L}\sum_{i\in S_m}\sum_{j=1}^{N}\left[\mathcal{C}_{ij}^{(k,l)}\log\frac{f_j^{(k,l)}}{f_j^{(k,l-1)}} + H_{ij}(f_j^{(k,l-1)} - f_j^{(k,l)})\right] \qquad \geq 0 \end{aligned} \tag{37}$$

with equality occurring if and only if $\mathbf{f}^{(k,l)} = \mathbf{f}^{(k,l-1)}$. This is due to the fact that the update equation in (21) is guaranteed to minimize $E_{\mathrm{cmp-ML-f}}(\mathbf{f};\mathcal{C}^{(k,l)})$. Consequently, $\Delta E_{\mathrm{cmp-ML}}^{(k,l)} \geq 0$ with equality occurring if and only if $\mathcal{C}_{ij}^{(k,l)} = \mathcal{C}_{ij}^{(k,l-1)}$, $\forall i \in S_l$, $\forall j$ *and* $f_j^{(k,l)} = f_j^{(k,l-1)}$, $\forall j$ which was our desired result.

The change in the complete data objective function between steps $(k-1, L)$ and (k, L) is

$$\begin{aligned} \Delta E_{\mathrm{cmp-ML}}^{(k)} \quad &\stackrel{\mathrm{def}}{=} \quad E_{\mathrm{cmp-ML}}(\mathcal{C}^{(k-1,L)}, \mathbf{f}^{(k-1,L)}) - \\ &\qquad E_{\mathrm{cmp-ML}}(\mathcal{C}^{(k,L)}, \mathbf{f}^{(k,L)}), \;\; k > 0 \\ &= \quad \sum_{l=1}^{L}\Delta E_{\mathrm{cmp-ML}}^{(k,l)} \geq 0 \end{aligned} \tag{38}$$

$\Delta E_{\mathrm{cmp-ML}}^{(k)}$ becomes equal to zero if and only if $\mathcal{C}_{ij}^{(k,l)} = \mathcal{C}_{ij}^{(k,l-1)}$, $\forall i \in S_l$, $\forall l \in \{1,\ldots,L\}$, $\forall j$ and $f_j^{(k,l)} = f_j^{(k,l-1)}$, $\forall j$, $\forall l \in \{1,\cdots,L\}$. As in Sec. VI-D, define k^* as the iteration at which $\Delta E_{\mathrm{cmp-ML}}^{(k^*)} = 0$. Then $\mathcal{C}^{(k^*,1)} = \mathcal{C}^{(k^*,2)} = \cdots = \mathcal{C}^{(k^*,L)} \stackrel{\mathrm{def}}{=} \mathcal{C}^*$. We similarly get $\mathbf{f}^{(k^*,1)} = \mathbf{f}^{(k^*,2)} = \cdots = \mathbf{f}^{(k^*,L)} \stackrel{\mathrm{def}}{=} \mathbf{f}^*$. We automatically have $\mathbf{f}^{(k^*+1,0)} = \mathbf{f}^{(k^*,L)}$ and $\mathcal{C}^{(k^*+1,0)} = \mathcal{C}^{(k^*,L)}$ which means that nothing can change once we have reached k^*. From the COSEM-ML $\mathcal{C}$ update (19)-(20), we once again get (32). Therefore, from (12), we get

$$\mathcal{C}^* = \mathcal{C}^{\mathrm{sol}}(\mathbf{f}^*) \tag{39}$$

Define Γ as the set of global minima of $E_{\mathrm{cmp-ML}}(\mathcal{C},\mathbf{f})$. We may argue that $(\mathcal{C}^*,\mathbf{f}^*) \in \Gamma$. The argument follows that at the start of Section VI-E if we simply replace $E_{\mathrm{cmp-MAP}}$ by $E_{\mathrm{cmp-ML}}$ and COSEM-MAP by COSEM-ML. We now show that $\mathbf{f}^*$ is a global minimum of $E_{\mathrm{inc-ML}}(\mathbf{f})$. We prove this by contradiction. Assume an $\check{\mathbf{f}}$ such that $E_{\mathrm{inc-ML}}(\mathbf{f}^*) > E_{\mathrm{inc-ML}}(\check{\mathbf{f}})$. Since by (39), $\mathcal{C}^* = \mathcal{C}^{\mathrm{sol}}(\mathbf{f}^*)$, we have by the touching property in (13) that:

$$\begin{aligned} &E_{\mathrm{cmp-ML}}(\mathcal{C}^{\mathrm{sol}}(\mathbf{f}^*),\mathbf{f}^*) = E_{\mathrm{inc-ML}}(\mathbf{f}^*) \\ &> E_{\mathrm{inc-ML}}(\check{\mathbf{f}}) = E_{\mathrm{cmp-ML}}(\mathcal{C}^{\mathrm{sol}}(\check{\mathbf{f}}),\check{\mathbf{f}}), \end{aligned} \tag{40}$$

but this is a contradiction because $(\mathcal{C}^{\mathrm{sol}}(\mathbf{f}^*),\mathbf{f}^*) \in \Gamma$ is a global minimum of $E_{\mathrm{cmp-ML}}$ (though it may not be the *unique* global minimum).

This completes the proof, in that we have established that $\mathbf{f}^*$, obtained from the COSEM-ML algorithm, is indeed equal to $\hat{\mathbf{f}}$ in (1), which is the result we have sought.

VIII. Simulation Results

In this section, we anecdotally explore 2D SPECT COSEM-MAP and COSEM-ML. We illustrate the relative speeds by displaying the normalized objective difference, NOD, versus iteration number k defined as:

$$\mathrm{NOD}(k) = \frac{E_{\mathrm{inc-MAP}}(\mathbf{f}^k) - E_{\mathrm{inc-MAP}}(\mathbf{f}^*)}{E_{\mathrm{inc-MAP}}(\mathbf{f}^0) - E_{\mathrm{inc-MAP}}(\mathbf{f}^*)} \tag{41}$$

where $\mathbf{f}^k \equiv \mathbf{f}^{(k,0)}$. The quantity $\mathbf{f}^0$ is the initial estimate and we define $\mathbf{f}^*$ to be the true MAP or ML solution at $k = \infty$, estimated here from 5000 iterations of EM-MAP or EM-ML. The $\mathrm{NOD}(k)$ curve for any of the algorithms descends from unity at $k = 0$ towards zero as $k \to \infty$.

A noisy sinogram with 300K counts was simulated using a 2D 64x64 phantom with a pixel size of 5.6 mm, shown in Fig. 1(a). The phantom consists of a disk background, two hot lesions, and two cold lesions with contrast ratio of 1:4:8 (cold:background:hot). The sinogram had dimensions of 64 equispaced angles by 96 detector bins, with a parallel-ray collimation geometry. Uniform attenuation (H_2O at 140 KeV) and depth-dependent blur characteristic of a SPECT collimator were modeled, but no other physical or background events were simulated. Subsets were defined by projection angle, so that e.g. L=16 corresponded to angles 1,17,33,49 for the first subset, angles 2,18,34,50 for the second subset, and so on. We did not optimize the ordering of the subsets.

For MAP, reconstructions were performed using EM-MAP, BSREM [3] and COSEM-MAP. (Note that EM-MAP is simply COSEM-MAP at $L = 1$ subset). For $\mathbf{f}^0$, we used an apodized FBP reconstruction with a Hamming filter and a cutoff at 0.7 of the Nyquist frequency. We chose $L = 8$ subsets for BSREM and COSEM-MAP, and for all three reconstructions, the smoothing parameter was set to $\beta = 0.06$. The relaxation schedule for BSREM was chosen as:

$$\alpha^k = \frac{\alpha_0}{max_{\{j,l\}} \sum_{i \in S_l} H_{ij} + k} \tag{42}$$

with $\alpha_0 = 3.2$. This schedule obeys the constraints in [3]. The schedule and α_0 were chosen to attain fairly rapid convergence at early iterations. Anecdotal reconstructions for each algorithm are shown in Fig. 1 at iteration 30, at which point, they are visually quite similar.

We plot NOD vs. k for each reconstruction in Fig. 2. We note that concept of "speed comparison" is hard to define. For BSREM, a particular choice of relaxation schedule and subset number L can lead to fast convergence at the first few iterations and very slow convergence thereafter. A different choice yields slower convergence at early iterations, but a lower value of NOD at higher iterations. For Fig. 2, we chose $L = 8$ for both BSREM and COSEM-MAP and (42) with $\alpha_0 = 3.2$ for BSREM to demonstrate a reasonably representative speed difference. At early iterations, the speed of COSEM-MAP lies between that of BSREM and EM-MAP, and by iteration 25, the NOD's of COSEM-MAP and BSREM cross each other. We also plot in Fig. 3 $\mathrm{NOD(k)}$ for COSEM-MAP under the same physical effects, β and $\mathbf{f}^0$, but with varying numbers of subsets (L= 4, 8, 16, 32, 64). After an initial speedup at $L = 4$, further speedup with increasing L is modest and is not as good as that observed with conventional OS-type algorithms.

For ML, we compared the performance of EM-ML, RAMLA and COSEM-ML. The object, imaging conditions and initial estimate $\mathbf{f}^0$ were the same as in the MAP case. For RAMLA, we used the relaxation schedule (42), which is consistent with [2]. For a representative speed comparison, we chose $L = 8$ for both RAMLA and COSEM-ML, and $\alpha_0 = 3.9$ for RAMLA. Fig. 4 shows that the speed of COSEM-ML lies between that of RAMLA and EM-ML. We also plotted $\mathrm{NOD}(k)$ for COSEM-ML for various subset numbers ($L = 4, 8, 16, 32, 64$) in Fig. 5. COSEM-ML also showed a slow speed increase with L.

IX. Discusssion

We have derived new convergent complete data ordered subsets algorithms for ML reconstruction (COSEM-ML) and MAP reconstruction (COSEM-MAP) in emission tomography. It is straightforward to include randoms or scatter via an affine term $\bar{\mathbf{s}}$, corresponding to $\mathbf{g} \sim Poisson(H\mathbf{f} + \bar{\mathbf{s}})$, in the COSEM-ML and COSEM-MAP algorithms. We can show that we only need to modify the complete data update equations so that $\bar{\mathbf{s}}$ is added to the forward projection in the denominator.

So far, it is unknown whether COSEM converges monotonically. (By monotonic, we mean that the objective decreases at each outer loop iteration k.) For the case of ML estimation of the parameters of a mixture model using an incremental EM algorithm, a mild non-monotonic behavior was experimentally observed [9]. However, the mixture problem is non-convex, whereas for our simpler convex ECT problem, one might hope for monotonicity. In our many simulations using COSEM-MAP and COSEM-ML with clinically realistic imaging parameters, we have not yet observed non-monotonicity. Even under various extreme conditions (including low counts and high subset number) that we thought might induce non-monotonic behavior, we have always observed the reconstructions to be monotonic. Further work is needed regarding this issue.

The COSEM-ML algorithm for emission tomography was independently derived by Gunawardana [9], though no experimental results were shown. The work in [9] was in turn related to the work in [5]. We discuss a few differences between the work here and that in [9]. In [9], the complete data variable is $\Pr(\mathbf{C} = \mathbf{c})$, with $\mathbf{C}$ an integer valued random variable as used in statistical derivations [10], [11], and the complete data energy function is of the form $E(\Pr(\mathbf{C} = \mathbf{c}), \mathbf{f})$. Hence, the complete data is a probability measure over all possible (allowed) configurations of the complete data random variable $\mathbf{C}$. The complete data energy function in [9] consists of a Kullback-Leibler divergence between a known distribution and the unknown $\Pr(\mathbf{C} = \mathbf{c})$. We have shown in this paper that it is unnecessary to consider a complete data variable which is a probability measure over all possible configurations of $\mathbf{C}$. It turns out that for ECT, a continuous-valued complete data variable $\mathcal{C}$ is sufficient for our purposes and results in a valid incremental EM algorithm. For the ECT problem, our representation of complete data results in more accessible and transparent derivations. We also provide an extension of the COSEM-ML convergence proof to the MAP case using a separable surrogates approach.

The results in Sec. VIII show that while COSEM does not require a user-specified relaxation schedule, it is generally slower than its relaxation-based counterparts in early iterations. Though the simulations in Sec. VIII are not very extensive, it has been our experience that the results are quite typical, and that the speed gap relative to relaxation-based methods cannot be closed by simply choosing a better L or initial condition. It would, of course, be desirable to be able to speed up COSEM while maintaining its convergence property and lack of dependence on a relaxation schedule. In separate publications [23], [24], we have, in fact, demonstrated that this is possible (albeit so far for the ML case only) with an "enhanced" COSEM-ML (ECOSEM-ML) algorithm. The convergence proof in [24] relied upon the convergence proof of COSEM itself. This COSEM convergence proof (which had previously appeared only in our technical report [25]) has been a main goal of the present paper. In ECOSEM-ML, the basic strategy is to apply an automatically computed parameter that controls a trade-off between a COSEM-ML update and a faster OSEM update at each sub-iteration, while maintaining convergence. ECOSEM-ML is designed to monotonically decrease $E_{\mathrm{cmp-ML}}$. Simulations results for ECOSEM-ML, available in [24], show the speed of ECOSEM-ML to approach that of optimized RAMLA at early iterations. It turns out that the ECOSEM idea is extendable to the MAP case.

In sum, we have presented COSEM-ML and COSEM-MAP convergence proofs for ECT based on our notion of a complete data energy. The COSEM algorithms need no user-specified relaxation schedule as do RAMLA, the various versions of BSREM and relaxed OS-SPS. The question of COSEM monotonicity remains open. Simulations show that the early-iteration speed of COSEM is slower than that of other provably convergent, relaxation-based MAP and ML algorithms. However, faster enhanced versions of COSEM, competitive in speed with RAMLA and BSREM, appear possible.

APPENDIX

Using (24), we obtain

$$\Delta E^{(k,l)}_{\mathrm{cmp-MAP}} = \sum_{m=1}^{L} \sum_{i \in S_m} \sum_{j=1}^{N} \left[\mathcal{C}_{ij}^{(k,l-1)} \log \frac{\mathcal{C}_{ij}^{(k,l-1)}}{H_{ij} f_j^{(k,l-1)}} - \mathcal{C}_{ij}^{(k,l)} \log \frac{\mathcal{C}_{ij}^{(k,l)}}{H_{ij} f_j^{(k,l)}} \right] + \sum_{ij} H_{ij} (f_j^{(k,l-1)} - f_j^{(k,l)}) + \beta \sum_{jj'} w_{jj'} \left[(f_j^{(k,l-1)} - f_{j'}^{(k,l-1)})^2 - (f_j^{(k,l)} - f_{j'}^{(k,l)})^2 \right]. \tag{43}$$

Note that in the first term of (43), we have replaced the subset index l by m, as we did in (17). From (25), we get that

$$-(f_j^{(k,l)} - f_{j'}^{(k,l)})^2 \geq -\frac{1}{2} [(2 f_j^{(k,l)} - f_j^{(k,l-1)} - f_{j'}^{(k,l-1)})^2 + (2 f_{j'}^{(k,l)} - f_j^{(k,l-1)} - f_{j'}^{(k,l-1)})^2]. \tag{44}$$

Note that at step (k,l), only the values of $\{\mathcal{C}_{ij}^{(k,l)}, \forall i \in S_l, \forall j\}$ are changed with all other values of $\mathcal{C}$ held fixed. Using this fact and (43)-(44), we get

$$\begin{aligned} \Delta E_{\text{cmp-MAP}}^{(k,l)} \geq \sum_{i \in S_l} \sum_{j=1}^{N} \Bigg[\mathcal{C}_{ij}^{(k,l-1)} \log \frac{\mathcal{C}_{ij}^{(k,l-1)}}{\mathcal{C}_{ij}^{(k,l)}} + \\ (\mathcal{C}_{ij}^{(k,l-1)} - \mathcal{C}_{ij}^{(k,l)}) \log \frac{\mathcal{C}_{ij}^{(k,l)}}{H_{ij} f_j^{(k,l-1)}} \Bigg] \\ + \sum_{m=1}^{L} \sum_{i \in S_m} \sum_{j=1}^{N} \left[\mathcal{C}_{ij}^{(k,l)} \log \frac{f_j^{(k,l)}}{f_j^{(k,l-1)}} + H_{ij} (f_j^{(k,l-1)} - f_j^{(k,l)}) \right] \\ + \frac{\beta}{2} \sum_{jj'} w_{jj'} \Big[(2 f_j^{(k,l-1)} - f_j^{(k,l-1)} - f_{j'}^{(k,l-1)})^2 - \\ (2 f_j^{(k,l)} - f_j^{(k,l-1)} - f_{j'}^{(k,l-1)})^2 \Big] \\ + \frac{\beta}{2} \sum_{jj'} w_{jj'} \Big[(2 f_{j'}^{(k,l-1)} - f_j^{(k,l-1)} - f_{j'}^{(k,l-1)})^2 - \\ (2 f_{j'}^{(k,l)} - f_j^{(k,l-1)} - f_{j'}^{(k,l-1)})^2 \Big]. \end{aligned} \tag{45}$$

From the update equation (19) for $\mathcal{C}^{(k,l)}$ we get

$$\begin{aligned} \sum_j (\mathcal{C}_{ij}^{(k,l-1)} - \mathcal{C}_{ij}^{(k,l)}) \log \frac{\mathcal{C}_{ij}^{(k,l)}}{H_{ij} f_j^{(k,l-1)}} = \\ \sum_j (\mathcal{C}_{ij}^{(k,l-1)} - \mathcal{C}_{ij}^{(k,l)}) \log \frac{g_i}{\sum_n H_{in} f_n^{(k,l-1)}}. \end{aligned} \tag{46}$$

From the fact that the constraint $\sum_j \mathcal{C}_{ij} = g_i$ is always satisfied, we have the identity

$$\sum_{i \in S_l} \sum_{j=1}^{N} (\mathcal{C}_{ij}^{(k,l-1)} - \mathcal{C}_{ij}^{(k,l)}) \log \frac{g_i}{\sum_n H_{in} f_n^{(k,l-1)}} = 0. \tag{47}$$

Using (45), (46) and (47), we may write

$$\begin{aligned} \Delta E_{\text{cmp-MAP}}^{(k,l)} \geq \sum_{i \in S_l} \sum_{j=1}^{N} \mathcal{C}_{ij}^{(k,l-1)} \log \frac{\mathcal{C}_{ij}^{(k,l-1)}}{\mathcal{C}_{ij}^{(k,l)}} + \\ \sum_{m=1}^{L} \sum_{i \in S_m} \sum_{j=1}^{N} \left[\mathcal{C}_{ij}^{(k,l)} \log \frac{f_j^{(k,l)}}{f_j^{(k,l-1)}} + H_{ij} (f_j^{(k,l-1)} - f_j^{(k,l)}) \right] \\ + \frac{\beta}{2} \sum_{jj'} w_{jj'} \Big[(2 f_j^{(k,l-1)} - f_j^{(k,l-1)} - f_{j'}^{(k,l-1)})^2 - \\ (2 f_j^{(k,l)} - f_j^{(k,l-1)} - f_{j'}^{(k,l-1)})^2 \Big] \\ + \frac{\beta}{2} \sum_{jj'} w_{jj'} \Big[(2 f_{j'}^{(k,l-1)} - f_j^{(k,l-1)} - f_{j'}^{(k,l-1)})^2 - \\ (2 f_{j'}^{(k,l)} - f_j^{(k,l-1)} - f_{j'}^{(k,l-1)})^2 \Big]. \end{aligned} \tag{48}$$

The first term on the right of (48) can be expressed as a KL divergence of the form $x \log \frac{x}{y} - x + y$. To see this, note that $\sum_j \mathcal{C}_{ij}^{(k,l-1)} = \sum_j \mathcal{C}_{ij}^{(k,l)} = g_i$. Hence, we may write

$$\sum_{j=1}^{N} \mathcal{C}_{ij}^{(k,l-1)} \log \frac{\mathcal{C}_{ij}^{(k,l-1)}}{\mathcal{C}_{ij}^{(k,l)}} =$$

$$\sum_{j=1}^{N} [\mathcal{C}_{ij}^{(k,l-1)} \log \frac{\mathcal{C}_{ij}^{(k,l-1)}}{\mathcal{C}_{ij}^{(k,l)}} - \mathcal{C}_{ij}^{(k,l-1)} + \mathcal{C}_{ij}^{(k.l)}] \geq 0 \quad (49)$$

with equality occurring only if $\mathcal{C}_{ij}^{(k,l)} = \mathcal{C}_{ij}^{(k,l-1)}$, $\forall i \in S_l$, $\forall j$.

The remaining terms on the right in (48) can be seen from (26) to be

$$E_{\mathrm{cmp-MAP-fs}}(\mathbf{f}^{(k,l-1)}; \mathbf{f}^{(k,l-1)}, \mathcal{C}^{(k,l)}) -$$

$$E_{\mathrm{cmp-MAP-fs}}(\mathbf{f}^{(k,l)}; \mathbf{f}^{(k,l-1)}, \mathcal{C}^{(k,l)}) \geq 0 \quad (50)$$

with equality occurring if and only if $\mathbf{f}^{(k,l)} = \mathbf{f}^{(k,l-1)}$. This is due to the fact that the update equation in (28) is guaranteed to minimize $E_{\mathrm{cmp-MAP-fs}}(\mathbf{f}; \mathbf{f}^{(k,l-1)}, \mathcal{C}^{(k,l)})$. Consequently, from (48)-(50), we have reached our stated conclusion: the change in the complete data MAP objective function at step (k,l) is $\Delta E_{\mathrm{cmp-MAP}}^{(k,l)} \geq 0$ with equality occurring if and only if $\mathcal{C}_{ij}^{(k,l)} = \mathcal{C}_{ij}^{(k,l-1)}$, $\forall i \in S_l$, $\forall j$ *and* $f_j^{(k,l)} = f_j^{(k,l-1)}$, $\forall j$.

References

[1] H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Trans. Med. Imag.*, vol. 13, no. 4, pp. 601–609, Dec. 1994.

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

[3] A. R. De Pierro and M. E. B. Yamagishi, "Fast EM-like methods for maximum *a posteriori* estimates in emission tomography," *IEEE Trans. Med. Imag.*, vol. 20, no. 4, pp. 280–288, Apr. 2001.

[4] S. Ahn and J. Fessler, "Globally convergent ordered subsets algorithms for emission tomography using relaxed ordered subsets algorithms," *IEEE Trans. Med. Imag.*, vol. 22, no. 5, pp. 613–626, May 2003.

[5] R. Neal and G. Hinton, "A view of the EM algorithm that justifies incremental, sparse, and other variants," in *Learning in Graphical Models*, M. I. Jordan, Ed. Kluwer, 1998, pp. 355–368.

[6] 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*, vol. 4684. SPIE, Feb. 2002, pp. 10–19.

[7] ——, "A new convergent MAP reconstruction algorithm for emission tomography using ordered subsets and separable surrogates," in *Conf. Rec. IEEE Int. Symp. Biomed. Imaging*. IEEE, July 2002, pp. 409–412.

[8] J. Fessler, *Personal Communication*, 2002.

[9] A. J. R. Gunawardana, "The information geometry of EM variants for speech and image processing," Ph.D. dissertation, Johns Hopkins University, Baltimore, MD, 2001.

[10] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. Med. Imag.*, vol. 1, pp. 113–122, 1982.

[11] K. Lange and R. Carson, "EM reconstruction algorithms for emission and transmission tomography," *J. Comp. Assist. Tomography*, vol. 8, no. 2, pp. 306–316, April 1984.

[12] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 1991.

[13] A. R. De Pierro, "On the relation between the ISRA and the EM algorithm for positron emission tomography," *IEEE Trans. Med. Imag.*, vol. 12, no. 2, pp. 328–333, 1993.

[14] M. Lee, "Bayesian reconstruction in emission tomography using Gibbs priors," Ph.D. dissertation, Yale University, New Haven, CT, 1994.

[15] A. Rangarajan, S.-J. Lee, and G. Gindi, "Mechanical models as priors in Bayesian tomographic reconstruction," in *Maximum Entropy and Bayesian Methods*. Dordrecht, Netherlands: Kluwer Academic Publishers, 1996, pp. 117–124.

[16] A. Rangarajan, I.-T. Hsiao, and G. R. 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, 2000.

[17] A. De Pierro, "A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography," *IEEE. Trans. Med. Imag.*, vol. 14, pp. 132–137, March 1995.

[18] K. Lange, D. R. Hunter, and I. Yang, "Optimization transfer using surrogate objective functions," *Journal of Computational and Graphical Statistics*, vol. 9, pp. 1–59, 2000, (with discussion).

[19] A. De Pierro, "On the convergence of an EM-type algorithm for penalized likelihood estimation in emission tomography," *IEEE. Trans. Med. Imag.*, vol. 14, pp. 762–765, Dec. 1995.

[20] K. Lange, "Convergence of EM image reconstruction algorithms with Gibbs smoothing," *IEEE Trans. Med. Imag.*, vol. 9, no. 4, pp. 439–446, Dec. 1990 corrections, June 1991.

[21] J. Fessler, "Penalized weighted least-squares image reconstruction for positron emission tomography," *IEEE Trans. Med. Imag.*, vol. 13, no. 2, pp. 290–300, June 1994.

[22] I. T. Hsiao, A. Rangarajan, Y. Xing, and G. Gindi, "A smoothing prior with embedded positivity constraint for tomographic reconstruction," in *Conf. Rec. Int. Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*. Web URL: http://www.mil.sunysb.edu/mipl/publications.html, 2001, pp. 81–84.

[23] 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.

[24] I. T. Hsiao, A. Rangarajan, P. Khurd, and G. Gindi, "An accelerated convergent ordered subsets algorithm for emission tomography," *Phys. Med. Biol.*, Web URL: http://www.mil.sunysb.edu/mipl/publications.html, accepted 2004.

[25] A. Rangarajan, P. Khurd, I.-T. Hsiao, and G. Gindi, "Convergence Proofs for the COSEM-ML and COSEM-MAP Algorithms," 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, Tech. Rep. MIPL-03-1, Dec. 2003.

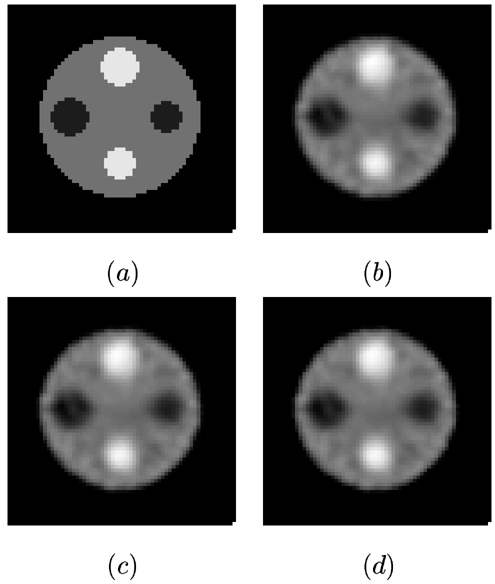

Fig. 1. The 2D 64x64 phantom is shown in (a), while the anecdotal reconstructions are displayed in (b) EM-MAP, (c) COSEM-MAP, and (d) BSREM. All reconstructions are displayed at $k = 30$ iterations.

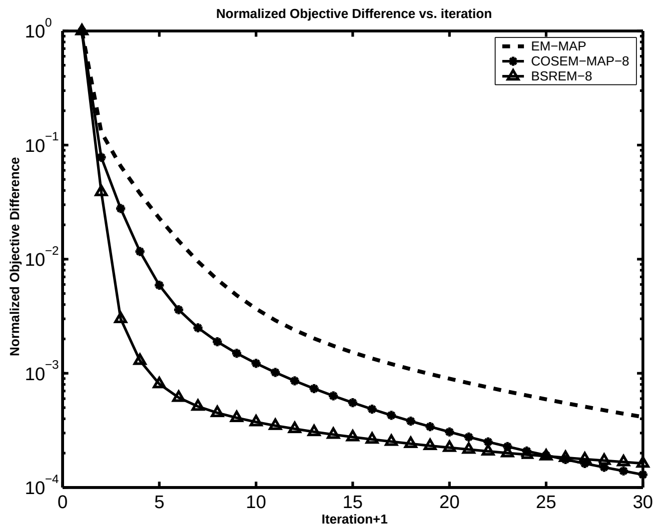

Fig. 2. Normalized objective difference for reconstructions using EM-MAP, COSEM-MAP, and BSREM at L=8.

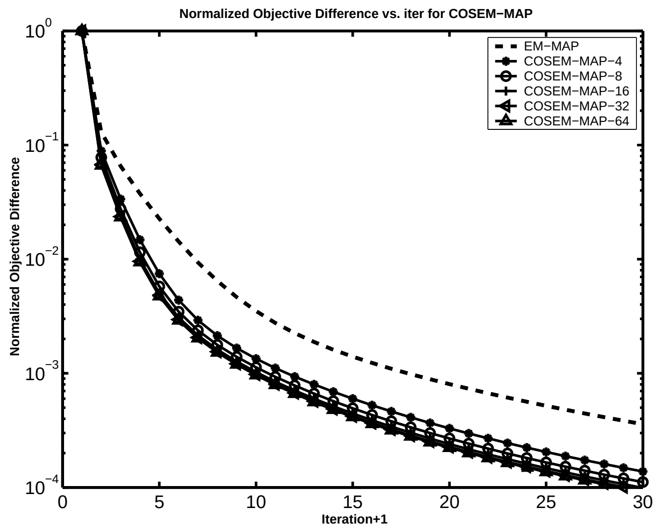

Fig. 3. Normalized objective difference plots of EM-MAP, and COSEM-MAP with various subsets, L=4, 8, 16, 32, and 64.

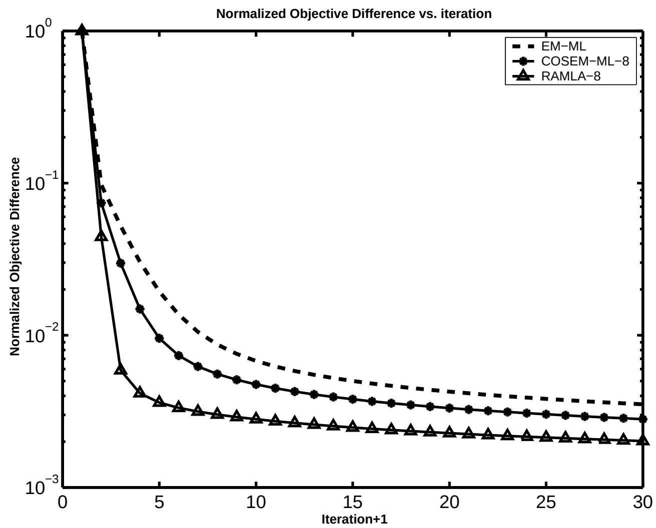

Fig. 4. Normalized objective difference for reconstructions using EM-ML, COSEM-ML, and RAMLA at L=8.

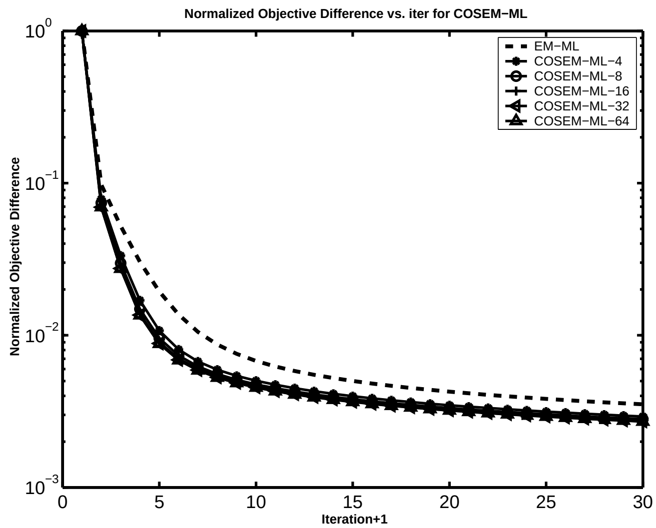

Fig. 5. Normalized objective difference plots of EM-ML, and COSEM-ML with various subsets, L=4, 8, 16, 32, and 64.