Convergence Proofs for the COSEM-ML and COSEM-MAP Algorithms

Anand Rangarajan[1], Parmeshwar Khurd[3],
Ing-Tsung Hsiao[2] and Gene Gindi[3]

December 2003

Tech. Rep. MIPL-03-1

[1]Dept. of CISE
University of Florida
Gainesville, FL, USA
anand@cise.ufl.edu

[2]School of Medical Technology
Chang-Gung University
333 Tao-yuan, Taiwan
ihsiao@mail.cgu.edu.tw

[3]Depts. of Radiology and Elec. Eng.
SUNY at Stony Brook
Stony Brook, NY, USA
{pkhurd,gindi}@clio.mil.sunysb.edu

1 Introduction

Statistical reconstruction has become increasingly popular in emission computed tomography (ECT) due to its ability to accurately model noise and the imaging physics. In addition, information regarding the object can be incorporated using Bayesian priors. In emission tomography, a Poisson log-likelihood projection data model is widely used because the photon noise is independently Poisson at each detector bin. Given the Poisson likelihood, the maximum likelihood (ML) principle is typically invoked as an optimization criterion for statistical reconstruction. This leads to a variety of iterative ML algorithms of which the expectation-maximization (EM) algorithm is perhaps the most well known. When Bayesian priors are considered, a maximum *a posteriori* (MAP) principle is frequently used as an optimization criterion. There are many log-priors that are convex and well behaved, which when added to the convex Poisson log-likelihood form a strictly convex log-posterior objective [1]. There exist many statistical MAP reconstruction algorithms that can determine the global optimum of such log-posterior MAP objectives. The main drawback of statistical reconstruction algorithms is that they are slow when used for clinical studies, especially in comparison to the commonly used filtered backprojection algorithm. This is true regardless of whether likelihood- or MAP-based approaches are being considered.

The slow convergence of statistical reconstruction algorithms has received much attention. Recently, an ordered subsets EM (OS-EM) algorithm [2], which uses only a subset of the projection data per sub-iteration, has become quite popular in ECT. This lead to an order of magnitude speedup over conventional (non-OS) EM. However, OS-EM is heuristically motivated and lacks a proof of convergence. An alternative algorithm, termed row-action maximum likelihood algorithm (RAMLA) was proposed in [3]. RAMLA used a strong under relaxation parameter within a modified version of OS-EM to prove convergence. The relaxation parameter has to satisfy certain properties to guarantee convergence to the true ML solution. In practice, this relaxation schedule must be determined by trial and error to ensure good speed. Recently in [4], the authors have extended their approach to the MAP case as well. The new MAP reconstruction algorithm, termed BSREM, continues to use a relaxation parameter to guide convergence to the MAP solution. In [5], a modified BSREM algorithm was presented that converged under more general conditions than BSREM. Another convergent algorithm, termed the OS-SPS algorithm, was also presented in [5]. Both the algorithms in [5] again used a user-determined relaxation schedule to ensure convergence.

In this technical report, we consider new fast iterative algorithms for optimizing the ML and MAP objectives for ECT. These new algorithms are called COSEM-ML and COSEM-MAP. (COSEM = Complete data Ordered Subset Expectation Maximization.) They make use of a notion of a complete data objective as well as ordered subsets. These algorithms were introduced in [6, 7]. Unlike other recent fast methods, they do not need a user-determined relaxation schedule. Here, we present convergence proofs for both algorithms.

2 Statement of the Problem

We will first mention some conventions used in this report. We chose plain letters to denote scalar variables and scalar-valued functions. Bold and calligraphic quantities represent vectors or matrices or vector-valued functions. The distinction between random and non-random quantities will be made clear in context. 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

Variable	Matrix/vector notation	Index notation
Projection data instance	$\mathbf{g}$	$\{g_i\}$
Source distribution instance	$\mathbf{f}$	$\{f_j\}$
Projection matrix	H	$\{H_{ij}\}$
Continuous-valued complete data	$\mathcal{C}$	$\{\mathcal{C}_{ij}\}$
Source distribution iterate	$\mathbf{f}^{(k,l)}$	$\{f_j^{(k,l)}\}$
Complete data iterate	$\mathcal{C}^{(k,l)}$	$\{\mathcal{C}_{ij}^{(k,l)}\}$
Complete data closed form solution	$\mathcal{C}^{\text{sol}}(\mathbf{f})$	$\mathcal{C}_{ij}^{\text{sol}}(\mathbf{f})$
Source distribution fixed point	$\hat{\mathbf{f}}$	$\{\hat{f}_j\}$
Complete data fixed point	$\hat{\mathcal{C}}$	$\{\hat{\mathcal{C}}_{ij}\}$
Auxiliary variable	P	$\{P_{ij}\}$
Sum of complete data	B	$\{B_j\}$
Sum of projection matrix	D	$\{D_j\}$
Lagrange parameter	μ	μ
Lagrange parameter	ν	$\{\nu_i\}$
Lagrange parameter	η	$\{\eta_{ij}\}$
Prior parameter	β	β
Prior weight	W	$\{w_{jj'}\}$
Number of sinogram elements	M	M
Number of pixels/voxels	N	N

Table 1: Notation and symbols used

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 $g_i,\ i = 1, \ldots, M$. Other physical quantities used in the remainder of this report are summarized in Table 1. We note that in the context of MAP formulations, $\mathbf{f}$ is technically a random vector. The incomplete data Poisson likelihood for emission tomography can be written as

$$\Pr(\mathbf{g}|\mathbf{f}) = \prod_i \frac{e^{-\sum_j H_{ij} f_j} (\sum_j H_{ij} f_j)^{g_i}}{g_i!}. \tag{1}$$

Here, we assumed that the Poisson mean is $\bar{\mathbf{g}} = H\mathbf{f}$ and ignore an affine term $\bar{\mathbf{s}}$, e.g. $\bar{\mathbf{g}} = H\mathbf{f} + \bar{\mathbf{s}}$, that is often added to account for scatter or randoms. As discussed in Sec. 8, it is easy to extend the proof to include the affine term. We will consider quadratic priors (regularizers) of the form

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

where $p(\mathbf{f})$ is the probability density function of $\mathbf{f}$, a.k.a. the prior. 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 $w_{jj'} \in W$ are positive. For a 2-D problem, a typical neighborhood $\mathcal{N}(j)$ comprises the eight nearest neighbors of j.

The emission tomography maximum likelihood (ML) problem is written as the minimization of an objective function

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

where the ML objective is given by

$$E_{\text{inc}-\text{ML}}(\mathbf{f}) = -\log \Pr(\mathbf{g}|\mathbf{f}) = \sum_{ij} H_{ij} f_j - \sum_i g_i \log \sum_j H_{ij} f_j. \tag{4}$$

In (4), we have ignored terms independent of $\mathbf{f}$. We have used the subscript "inc" to anticipate the use of the observed or "incomplete" data, as opposed to the "complete" data that are used in EM algorithms [8]. The maximum *a posteriori* (MAP) problem may be similarly stated as the minimization

$$\widehat{\mathbf{f}} = \arg\min_{\mathbf{f}\geq\mathbf{0}} E_{\text{inc}-\text{MAP}}(\mathbf{f}) \tag{5}$$

where the MAP objective is given by

$$E_{\text{inc}-\text{MAP}}(\mathbf{f}) = \sum_{ij} H_{ij} f_j - \sum_i g_i \log \sum_j H_{ij} f_j + \beta \sum_{jj'} w_{jj'} (f_j - f_{j'})^2 = E_{\text{inc}-\text{ML}}(\mathbf{f}) + E_{\text{prior}}(\mathbf{f}) \tag{6}$$

We have again ignored terms independent of $\mathbf{f}$. Our aim is to derive fast algorithms for the minimizations in (3) and (5) and to prove their convergence to the ML or MAP solutions. We note that (3) and (5) 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 allow an initial positive estimate of f_j to approach zero as iterations proceed, but never to equal zero. The well known EM-ML algorithm for ECT has this same behavior.

3 Complete data objective for the ML Problem

The direct minimization of $E_{\text{inc}-\text{ML}}$ is difficult. A popular alternative way of carrying out the minimization is by the well-known ML-EM algorithm for ECT. The conventional derivation [9, 10] is statistical in nature, but there exists an alternate means for deriving this algorithm via a minimization of a so-called complete data objective function. This objective function is defined as follows:

$$\begin{aligned} E_{\text{cmp}-\text{ML}}(\mathcal{C}, \mathbf{f}, \nu) \quad = \quad & -\sum_{ij} \mathcal{C}_{ij} \log H_{ij} f_j + \sum_{ij} H_{ij} f_j + \sum_{ij} \mathcal{C}_{ij} \log \mathcal{C}_{ij} \\ & + \sum_i \nu_i (\sum_j \mathcal{C}_{ij} - g_i) \end{aligned} \tag{7}$$

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 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 (7), this constraint is imposed via Lagrange parameters ν_i. Note that in (7), we do not include extra Lagrange terms to impose positivity on $\mathcal{C}$ and $\mathbf{f}$. This is because the COSEM-ML algorithm naturally imposes this constraint, as will be seen in Sec. 4. We will later show that minimizing $E_{\text{cmp}-\text{ML}}$ via our COSEM-ML algorithm yields the solution to the ML problem. We will freely intermix notations $E_{\text{cmp-ML}}(\mathcal{C}, \mathbf{f}, \nu)$ with $E_{\text{cmp-ML}}(\mathcal{C}, \mathbf{f})$. The latter notation means that the constraint implicitly holds and does not appear in the objective in a Lagrange form.

To motivate the complete data objective, in this section, we directly derive (7) using convexity arguments and change of variables transformations. A brief sketch of the approach follows. Using Jensen's inequality [11], we show that the original incomplete data negative log-likelihood lies below a new complete data objective function and touches it for a specific choice of the new complete data variable. Consequently, the new complete data objective when minimized solely w.r.t. the complete data variable (and *not* the object $\mathbf{f}$) attains its minimum at the original incomplete data negative log-likelihood. This approach of using Jensen's inequality to derive a new complete data objective is strikingly similar to the pioneering work of de Pierro [12]. The resulting complete data objective function is also very similar to new objective functions first derived in [13] (Appendix B) and identical to the new complete data objective functions derived in [14] and in [15]. In the latter, a constructive derivation of the new complete data objective is notably absent. Finally, the constrained complete data objective function derived here is the same as (7).

The Jensen's inequality-based approach to deriving the complete data constrained objective function is as follows. We begin with the original negative log-likelihood objective function in (4). In this objective, we selectively replace terms involving $\log \sum_j H_{ij} f_j$.

$$E_{\text{inc-ML}}(\mathbf{f}) = \sum_{ij} H_{ij} f_j - \sum_i g_i \log \sum_j H_{ij} f_j. \tag{8}$$

The terms involving $-\log \sum_j H_{ij} f_j$ are the ones that will be transformed in order to obtain the new objective. Since $-\log(\cdot)$ is a convex function, we have from Jensen's inequality [11]

$$-\log(\sum_j \alpha_j x_j) \le -\sum_j \alpha_j \log x_j \tag{9}$$

where $\alpha_j \ge 0, \forall j$, $\sum_j \alpha_j = 1$ and $x_j > 0, \forall j$. Using (9), we may write

$$-\log \sum_j H_{ij} f_j \le -\sum_j \gamma_{ij} \log \frac{H_{ij} f_j}{\gamma_{ij}} \tag{10}$$

where $\gamma_{ij} \ge 0, \forall ij$ 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 this inequality in place, we may define a new objective function containing γ as an independent variable as

$$E(\gamma, \mathbf{f}) = \sum_{ij} H_{ij} f_j - \sum_{ij} g_i \gamma_{ij} \log \frac{H_{ij} f_j}{\gamma_{ij}} \tag{11}$$

with the constraints

$$\sum_j \gamma_{ij} = 1, \forall i \,\text{and}\, \gamma_{ij} \ge 0, \forall ij. \tag{12}$$

It is easy to show that the new objective function in (11) when minimized solely w.r.t. γ while satisfying the constraints in (12) 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 $\breve{\gamma} \neq \gamma^{\star}_{ij}$ where $\gamma^{*}_{ij} = \frac{H_{ij} f_j}{\sum_{j'} H_{ij'} f_{j'}}$, then $E(\breve{\gamma}, \mathbf{f}) < E(\gamma^{\star}, \mathbf{f})$. But this violates the inequality in (10). (The equality $E(\breve{\gamma}, \mathbf{f}) = E(\gamma^{\star}, \mathbf{f})$ is allowed since γ^{*} would still be a point at which $E(\gamma, \mathbf{f})$ attained its minimum). At the minimum, we have

$$E(\gamma^{\star}, \mathbf{f}) = E_{\mathrm{inc-ML}}(\mathbf{f}). \tag{13}$$

The new variable γ is not quite the complete data variable used in the traditional EM algorithm. However, if we define $\mathcal{C}_{ij} \stackrel{\mathrm{def}}{=} g_i \gamma_{ij}$ and rewrite the new objective function in (11) using the new variable $\mathcal{C}$, we get

$$E(\mathcal{C}, \mathbf{f}) = \sum_{ij} H_{ij} f_j - \sum_{ij} \mathcal{C}_{ij} \log \frac{g_i H_{ij} f_j}{\mathcal{C}_{ij}}. \tag{14}$$

The constraints in (12) get modified to

$$\sum_{j} \mathcal{C}_{ij} = g_i, \ \forall i \, \text{and} \, \mathcal{C}_{ij} \geq 0, \ \forall ij. \tag{15}$$

Using the new constraints, the new objective function in (14) is modified to

$$E(\mathcal{C}, \mathbf{f}) = \sum_{ij} H_{ij} f_j - \sum_{ij} \mathcal{C}_{ij} \log H_{ij} f_j + \sum_{ij} \mathcal{C}_{ij} \log \mathcal{C}_{ij} - \sum_{i} g_i \log \, g_i \tag{16}$$

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 (16) 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

$$\begin{aligned} E_{\mathrm{cmp-ML}}(\mathcal{C}, \mathbf{f}, \nu) \quad = \quad & -\sum_{ij} \mathcal{C}_{ij} \log \, H_{ij} f_j + \sum_{ij} H_{ij} f_j + \sum_{ij} \mathcal{C}_{ij} \log \, \mathcal{C}_{ij} \\ & + \sum_{i} \nu_i (\sum_{j} \mathcal{C}_{ij} - g_i) \end{aligned} \tag{17}$$

which is the same as (7). As before, we have introduced a Lagrange parameter ν to express the constraint $\sum_j \mathcal{C}_{ij} = g_i, \ \forall i$. A closely related complete data objective was first presented in [13] (Appendix B) and subsequently modified into the present form in [14].

We will now see that the new constrained optimization problem is designed within the fixed-point preservation constraint. Differentiating (7) w.r.t. $\mathcal{C}$ and setting the result to zero, we obtain a solution for $\mathcal{C}$ in terms of $\mathbf{f}$, which we call $\mathcal{C}^{\mathrm{sol}}(\mathbf{f})$.

$$\mathcal{C}^{sol}_{ij}(\mathbf{f}) = H_{ij} f_j e^{-\nu_i - 1}. \tag{18}$$

Summing this solution for $\mathcal{C}^{\mathrm{sol}}$ over all j in order to enforce the constraint $\sum_j \mathcal{C}_{ij} = g_i$, we get

$$\begin{aligned} \sum_{j} \mathcal{C}^{\mathrm{sol}}_{ij}(\mathbf{f}) \quad &= \quad \sum_{j} H_{ij} f_j e^{-\nu_i - 1} = g_i \\ & \Rightarrow \ \mathcal{C}^{\mathrm{sol}}_{ij}(\mathbf{f}) = g_i \frac{H_{ij} f_j}{\sum_n H_{in} f_n} \ \forall i, \ \forall j. \end{aligned} \tag{19}$$

Name	Objective function	Constraints	Optimization problem
Negative log-likelihood	$E_{\text{inc-ML}}(\mathbf{f}) = \sum H_{ij} f_j - \sum_i g_i \log \sum_j H_{ij} f_j$	$\{f_j > 0, \forall j\}$	$\hat{\mathbf{f}} = \arg\min_{\mathbf{f}} E_{\text{inc-ML}}(\mathbf{f})$
Complete data negative log-likelihood	$E_{\text{cmp-ML}}(\mathcal{C}, \mathbf{f}) = \sum_{ij} H_{ij} f_j + \sum_{ij} \mathcal{C}_{ij} \log \frac{\mathcal{C}_{ij}}{H_{ij} f_j}$	$\{\mathcal{C}_{ij} > 0, f_j > 0, \forall ij\}$ $\sum_j \mathcal{C}_{ij} = g_i$	$(\hat{\mathcal{C}}, \hat{\mathbf{f}}) = \arg\min_{(\mathcal{C},\mathbf{f})} E_{\text{cmp-ML}}(\mathcal{C}, \mathbf{f})$
Incomplete data MAP objective	$E_{\text{inc-MAP}}(\mathbf{f}) = \sum_{ij} H_{ij} f_j - \sum_i g_i \log \sum_j H_{ij} f_j + \sum_i g_i \log\, g_i - \sum_i g_i + \beta \sum_{jj'} w_{jj'} (f_j - f_{j'})^2$	$\{f_j > 0, \forall j\}$	$\hat{\mathbf{f}} = \arg\min_{\mathbf{f}} E_{\text{inc-MAP}}(\mathbf{f})$
Complete data MAP objective	$E_{\text{cmp-MAP}}(\mathcal{C}, \mathbf{f}) = \sum_{ij} H_{ij} f_j + \sum_{ij} \mathcal{C}_{ij} \log \frac{\mathcal{C}_{ij}}{H_{ij} f_j} - \sum_{ij} \mathcal{C}_{ij} + \beta \sum_{jj'} w_{jj'} (f_j - f_{j'})^2$	$\{\mathcal{C}_{ij} > 0, f_j > 0, \forall ij\}$ $\sum_j \mathcal{C}_{ij} = g_i$	$(\hat{\mathcal{C}}, \hat{\mathbf{f}}) = \arg\min_{(\mathcal{C},\mathbf{f})} E_{\text{cmp-MAP}}(\mathcal{C}, \mathbf{f})$
Separable surrogate MAP objective	$E_{\text{s-MAP}}(\mathbf{f}; \mathbf{f}^{(k,l-1)}, \mathcal{C}^{(k,l)}) = -\sum_{ij} \mathcal{C}_{ij}^{(k,l)} \log\, f_j + \sum_{ij} H_{ij} f_j + \frac{\beta}{2} \sum_{jj'} w_{jj'} (2f_j - f_j^{(k,l-1)} - f_{j'}^{(k,l-1)})^2 + \frac{\beta}{2} \sum_{jj'} w_{jj'} (2f_{j'} - f_j^{(k,l-1)} - f_{j'}^{(k,l-1)})^2$	$\{f_j > 0, \forall j\}$	$\mathbf{f}^{(k,l)} = \arg\min_{\mathbf{f}} E_{\text{s-MAP}}(\mathbf{f}; \mathbf{f}^{(k,l-1)}, \mathcal{C}^{(k,l)})$

Table 2: Objective functions, constraints and optimization problems

Inserting the just established solution (19) into (17), we can write the identity

$$E_{\text{cmp-ML}}(\mathcal{C}^{\text{sol}}(\mathbf{f}), \mathbf{f}) = \sum_{ij} H_{ij} f_j - \sum_i g_i \log \sum_j H_{ij} f_j + \text{terms independent of } \mathbf{f} = E_{\text{inc-ML}}(\mathbf{f}) \tag{20}$$

where $E_{\text{inc-ML}}(\mathbf{f})$ is the desired negative log-likelihood objective function in (4) which we sought to minimize in the first place. Note that in (20), we eliminate ν in the argument of $E_{\text{cmp-ML}}$ because the constraint holds at $\mathcal{C}^{\text{sol}}(\mathbf{f})$. Thus, we have shown that minimizing $E_{\text{cmp-ML}}(\mathcal{C}, \mathbf{f})$ w.r.t. $\mathcal{C}$ yields a solution $\mathcal{C}^{\text{sol}}(\mathbf{f})$ and that the minimum value of $E_{\text{cmp-ML}}(\mathcal{C}, \mathbf{f})$ thus obtained equals $E_{\text{inc-ML}}(\mathbf{f})$. Therefore, joint minimization of $E_{\text{cmp-ML}}(\mathcal{C}, \mathbf{f})$ will yield a fixed point $(\mathcal{C}^*, \mathbf{f}^*)$, where $\mathbf{f}^* = \hat{\mathbf{f}}$ is the solution sought in (3), i.e. the fixed point of (3) is preserved.

4 COSEM-ML Algorithm

The material in this section is closely related to recent work in [16], which was in turn based on the work in [17]. However, the present work was independently derived, with the connection to [16] realized later [18]. The principal differences between the work in [16] and the material presented here are (i) we derive a new complete data ordered subsets objective function which was not the goal of [16], (ii) we provide an extension to the MAP case (in Sec. 5) using a separable surrogates approach while the work in [16] is restricted to ML, (iii) we provided experimental results (in [6, 7, 19]), whereas the work in [16] presents none and (iv) we provide a mathematical "book-keeping" scheme that reduces memory requirements and enhances speed so that the COSEM algorithms are about as fast as OSEM. The objective-function approach in (i) leads to advantages. For instance, in [19], we showed how a faster "enhanced" COSEM-ML algorithm could be derived based on $E_{\text{cmp-ML}}$.

Since fast algorithms in the ordered subsets vein are the main goal of this paper, we rewrite (17) using subsets of projection angles notation. (This choice of subsets is appropriate for SPECT, and we use it for illustrative purposes. For PET, a different subset definition might be more appropriate. At any rate, the math that follows does not depend on choice of subsets.) 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 $l \in \{1, \ldots, L\}$. Corresponding to this division of the incomplete data, we also have a division of the continuous-valued complete data $\mathcal{C}$ which is denoted by $\{\mathcal{C}_{ij}, \forall i \in S_l, \forall j\}$ with $l \in \{1, \ldots, L\}$ as before. The complete data objective function in (17) can be re-written using this subset notation as

$$E_{\text{cmp-ML}}(\mathcal{C}, \mathbf{f}, \nu) = -\sum_{l=1}^{L} \sum_{i \in S_l} \sum_{j} \mathcal{C}_{ij} \log H_{ij} f_j + \sum_{ij} H_{ij} f_j + \sum_{l=1}^{L} \sum_{i \in S_l} \sum_{j} \mathcal{C}_{ij} \log \mathcal{C}_{ij} + \sum_{i} \nu_i (\sum_{j} \mathcal{C}_{ij} - g_i). \tag{21}$$

We now embark upon an ordered subsets minimization strategy. As with standard OS-EM and 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, whereas in the inner $l \in \{1, \ldots, L\}$ loop, each subiteration l corresponds to the update of the complete data $\{\mathcal{C}_{ij}, \forall i \in S_l, \forall j\}$. 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 (21) 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. To perform grouped coordinate descent w.r.t $\mathcal{C}_{ij},\ j = 1, \cdots, N\ i \in S_l$, we need the following expression:

$$\frac{\partial E_{\text{cmp-ML}}(\mathcal{C}, \mathbf{f})}{\partial \mathcal{C}_{ij}} = -\log(H_{ij} f_j) + \ln \mathcal{C}_{ij} + 1 + \nu_i \tag{22}$$

Equating this expression to zero, we get an update for $\mathcal{C}_{ij}$ in terms of the Lagrange multipliers:

$$\mathcal{C}_{ij} = \exp(-\nu_i - 1) H_{ij} f_j \tag{23}$$

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

$$\exp(-\nu_i - 1) = \frac{g_i}{\sum_j H_{ij} f_j} \tag{24}$$

Substituting (24) in (23), we obtain the update:

$$\mathcal{C}_{ij} = g_i \frac{H_{ij} f_j}{\sum_n H_{in} f_n} \tag{25}$$

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

$$\frac{\partial E_{\text{cmp-ML}}(\mathcal{C}, \mathbf{f})}{\partial f_j} = -\sum_{m=1}^{L} \sum_{i \in S_m} \mathcal{C}_{ij} \frac{1}{f_j} + \sum_i H_{ij} \tag{26}$$

Here, we have replaced the subset index l in (21) by m since we will be using l as a sub-iteration index later. Equating this expression to zero, we get the update:

$$f_j = \frac{\sum_{m=1}^{L} \sum_{i \in S_m} \mathcal{C}_{ij}}{\sum_i H_{ij}}, \forall j \tag{27}$$

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

$$\mathcal{C}_{ij}^{(k,l)} = g_i \frac{H_{ij} f_j^{(k,l-1)}}{\sum_n H_{in} f_n^{(k,l-1)}}, \forall i \in S_l, \forall j \tag{28}$$

$$\mathcal{C}_{ij}^{(k,l)} = \mathcal{C}_{ij}^{(k,l-1)}, \forall i \notin S_l, \forall j \tag{29}$$

$$f_j^{(k,l)} = \frac{\sum_{m=1}^{L} \sum_{i \in S_m} \mathcal{C}_{ij}^{(k,l)}}{\sum_i H_{ij}}, \forall j. \tag{30}$$

We clarify the notation used. The symbol $\mathcal{C}_{ij}^{(k,l)}$ denotes the update of the continuous-valued complete data variable at detector bin index i and voxel j at outer iteration k and subset iteration l. The subtlety here is that at iteration (k, l), we only update the complete data $\mathcal{C}_{ij}$ over all detector bins in subset S_l and all voxels $j \in \{1, \ldots, N\}$ as shown in (28). However, the update of the source distribution $f_j^{(k,l)}$ in (30) 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 $\mathcal{C}_{ij}^{(k,l)} = \mathcal{C}_{ij}^{(k,l-1)}, \forall i \notin S_l, \forall j$ in the complete data update as shown in (29).

This implies that the summation $\sum_{m=1}^{L} \sum_{i \in S_m} \mathcal{C}_{ij}^{(k,l)}$ can be divided into two disjoint subsets

$$\begin{aligned}
\sum_{m=1}^{L} \sum_{i \in S_m} \mathcal{C}_{ij}^{(k,l)} &= \sum_{i \in S_l} \mathcal{C}_{ij}^{(k,l)} + \sum_{i \notin S_l} \mathcal{C}_{ij}^{(k,l)} \\
&= \sum_{i \in S_l} (\mathcal{C}_{ij}^{(k,l)} - \mathcal{C}_{ij}^{(k,l-1)}) + \sum_{i \in S_l} \mathcal{C}_{ij}^{(k,l-1)} + \sum_{i \notin S_l} \mathcal{C}_{ij}^{(k,l)} \\
&= \sum_{i \in S_l} (\mathcal{C}_{ij}^{(k,l)} - \mathcal{C}_{ij}^{(k,l-1)}) + \sum_{i \in S_l} \mathcal{C}_{ij}^{(k,l-1)} + \sum_{i \notin S_l} \mathcal{C}_{ij}^{(k,l-1)}
\end{aligned} \tag{31}$$

since $\mathcal{C}_{ij}^{(k,l)} = \mathcal{C}_{ij}^{(k,l-1)}, \forall i \notin S_l, \forall j$. Combining the two terms $\sum_{i \in S_l} \mathcal{C}_{ij}^{(k,l-1)}$ and $\sum_{i \notin S_l} \mathcal{C}_{ij}^{(k,l-1)}$ together into one term $\sum_i \mathcal{C}_{ij}^{(k,l-1)}$, we get

$$B_j^{(k,l)} \stackrel{\text{def}}{=} \sum_{m=1}^{L} \sum_{i \in S_m} \mathcal{C}_{ij}^{(k,l)} = \sum_{i \in S_l} (\mathcal{C}_{ij}^{(k,l)} - \mathcal{C}_{ij}^{(k,l-1)}) + \sum_i \mathcal{C}_{ij}^{(k,l-1)}$$
$$= \sum_{i \in S_l} (\mathcal{C}_{ij}^{(k,l)} - \mathcal{C}_{ij}^{(k,l-1)}) + B_j^{(k,l-1)} \tag{32}$$

where $B_j^{(k,l-1)} \stackrel{\text{def}}{=} \sum_i \mathcal{C}_{ij}^{(k,l-1)}$ and $B_j^{(k,l)} \stackrel{\text{def}}{=} \sum_i \mathcal{C}_{ij}^{(k,l)}$. The definition and implementation of $B_j^{(k,l)}$ is quite important in our overall COSEM scheme. From (32), we may write

$$B_j^{(k,l)} = \sum_{i \in S_l} (\mathcal{C}_{ij}^{(k,l)} - \mathcal{C}_{ij}^{(k,l-1)}) + B_j^{(k,l-1)} \tag{33}$$

with the understanding that $B_j^{(k+1,0)} = B_j^{(k,L)}$. After all the complete data are updated over all L subsets, we end up at $\mathcal{C}_{ij}^{(k,L)}$ with the corresponding $B_j^{(k,L)} = \sum_i \mathcal{C}_{ij}^{(k,L)}$. After all L subsets are updated at iteration k, the k counter is incremented and we begin at $(k+1,0)$. We define $\mathcal{C}_{ij}^{(k+1,0)} \stackrel{\text{def}}{=} \mathcal{C}_{ij}^{(k,L)}$ and $B_j^{(k+1,0)} \stackrel{\text{def}}{=} B_j^{(k,L)}$. The iteration $(k=0, l=0)$ is a special case. At $(k=0, l=0)$, all $\mathcal{C}$ are initialized to $\mathcal{C}_{ij}^{(0,0)} = 0$ and therefore $B_j^{(0,0)} = 0$. Similarly, the $\mathbf{f}$ update is carried over from k to $k+1$. The update $\mathbf{f}^{(k+1,0)} \stackrel{\text{def}}{=} \mathbf{f}^{(k,L)}$. With this change in notation, the update equations in (28) and (30) are re-written as

$$\mathcal{C}_{ij}^{(k,l)} = g_i \frac{H_{ij} f_j^{(k,l-1)}}{\sum_n H_{in} f_n^{(k,l-1)}}, \forall i \in S_l, \forall j, \tag{34}$$
$$\mathcal{C}_{ij}^{(k,l)} = \mathcal{C}_{ij}^{(k,l-1)}, \forall i \notin S_l, \forall j \tag{35}$$
$$B_j^{(k,l)} = \sum_{i \in S_l} (\mathcal{C}_{ij}^{(k,l)} - \mathcal{C}_{ij}^{(k,l-1)}) + B_j^{(k,l-1)}, \forall j \tag{36}$$
$$f_j^{(k,l)} = \frac{B_j^{(k,l)}}{D_j}, \forall j, \tag{37}$$

where $D_j \stackrel{\text{def}}{=} \sum_i H_{ij}$ is the voxel sensitivity at voxel j. Note that (37) imposes positivity on f_j.

The double loop algorithm is summarized below.

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

- **End** *k-loop*

We note that for zero bins $g_i = 0$, the corresponding $\mathcal{C}_{ij}$ will equal zero and would remain zero throughout its update. Therefore, we eliminate these $\mathcal{C}_{ij}$'s from the problem. Also, if one knows *a priori* that some f_j are zero, then these are fixed to zero and eliminated from the problem. The $\mathcal{C}_{ij}$ corresponding to these $f_j = 0$ will also be zero and are thus eliminated from the problem.

5 COSEM-MAP Algorithm

Consider the complete data MAP objective function, obtained by adding the prior to (21) and using the subset scheme

$$E_{\text{cmp}-\text{MAP}}(\mathcal{C}, \mathbf{f}, \nu) = \sum_{ij} H_{ij} f_j + \sum_{l=1}^{L} \sum_{i \in S_l} \sum_{j=1}^{N} \mathcal{C}_{ij} \log \frac{\mathcal{C}_{ij}}{H_{ij} f_j} - \sum_{l=1}^{L} \sum_{i \in S_l} \sum_{j=1}^{N} \mathcal{C}_{ij} + \beta \sum_{jj'} w_{jj'} (f_j - f_{j'})^2 + \sum_i \nu_i (\sum_j \mathcal{C}_{ij} - g_i) \tag{38}$$

An extra term $(-\sum_{l=1}^{L} \sum_{i \in S_l} \sum_{j=1}^{N} \mathcal{C}_{ij})$ has also been added which does not change the minima of the new complete data MAP objective function provided the constraints $\sum_{j=1}^{N} \mathcal{C}_{ij} = g_i, \forall i \,\text{and}\, \mathcal{C}_{ij} > 0, \, \forall ij$ are satisfied. Thus $\sum_{l=1}^{L} \sum_{i \in S_l} \sum_{j=1}^{N} \mathcal{C}_{ij} = \sum_{l=1}^{L} \sum_{i \in S_l} g_i$, which is just an additive constant in (38). This minor modification simplifies the convergence proof in Sec. 6. Note that in (38), we do not include extra Lagrange terms to impose positivity on $\mathcal{C}$ and $\mathbf{f}$. This is because the COSEM-MAP algorithm naturally imposes this constraint, as will be seen in Sec. 5. We now proceed to derive a constrained grouped coordinate descent algorithm for (38). In Sec. 5.1, we consider the $\mathcal{C}$ update and in Sec. 5.2 the $\mathbf{f}$ update.

5.1 Complete data $\mathcal{C}$ update for the MAP case

The objective function w.r.t. $\mathcal{C}$ alone can be written from (38). We write down only the terms that are dependent on $\mathcal{C}$ to get (in ordered subsets notation)

$$E_{\text{cmp}-\text{MAP}}(\mathcal{C}) = -\sum_{l=1}^{L} \sum_{i \in S_l} \sum_j \mathcal{C}_{ij} \log H_{ij} f_j + \sum_{l=1}^{L} \sum_{i \in S_l} \sum_j \mathcal{C}_{ij} \log \mathcal{C}_{ij} - \sum_{l=1}^{L} \sum_{i \in S_l} \sum_{j=1}^{N} \mathcal{C}_{ij} + \sum_i \nu_i (\sum_j \mathcal{C}_{ij} - g_i) \tag{39}$$

Except for the (constant) additional term $\sum_{l=1}^{L} \sum_{i \in S_l} \sum_{j=1}^{N} \mathcal{C}_{ij}$, this objective function, when examining only the terms dependent on $\mathcal{C}$ alone, is identical to the earlier COSEM-ML case (21). Consequently, the update equation for $\mathcal{C}$ remains (for subset S_l) at (k,l)

$$\mathcal{C}_{ij}^{(k,l)} = g_i \frac{H_{ij} f_j^{(k,l-1)}}{\sum_n H_{in} f_n^{(k,l-1)}}, \forall i \in S_l, \forall j. \tag{40}$$

$$\mathcal{C}_{ij}^{(k,l)} = \mathcal{C}_{ij}^{(k,l-1)}, \forall i \notin S_l, \forall j. \tag{41}$$

5.2 Object f update for the MAP case

The prior in (6) $\left(\beta \sum_{jj'} w_{jj'} (f_j - f_{j'})^2\right)$ introduces a coupling between different source voxels f_j. Since we seek a closed-form parallel-update of all voxels, this coupling presents a problem; it is difficult to derive a parallel update for all voxels while respecting the positivity constraint and still guaranteeing convergence. While a positivity-preserving sequential update or a parallel, positivity-preserving gradient descent-based update are possible options, we instead use the method of separable surrogates. For a general convex prior $E_{\text{cnvx}-\text{prior}}(\mathbf{f})$, the separable surrogates approach decouples the prior and allows for separate, parallel 1-D optimization of all source voxels f_j. Since this optimization is 1-D, imposition of the positivity constraint is trivial. For the quadratic prior $E_{\text{prior}}(\mathbf{f})$ considered in (2), the 1-D optimization becomes a closed form expression that implicitly preserves positivity.

Assume that we are step (k,l) where k is the outer loop iteration index and l the inner subsets loop iteration index. (This notation is identical to the notation used in the earlier COSEM-ML algorithm.) Separable surrogate priors depend on the current estimate $\mathbf{f}^{(k,l-1)}$, and are also separable in $\mathbf{f}$. Therefore, they have the general form $E_{\text{s}-\text{prior}}(f_j; \mathbf{f}^{(k,l-1)})$ where "s" stands for surrogate. In order to maintain convergence, we need [20, 21, 22, 23] to satisfy the following conditions:

$$E_{\text{cnvx}-\text{prior}}(\mathbf{f}) \leq \sum_j E_{\text{s}-\text{prior}}(f_j; \mathbf{f}^{(k,l-1)}) \tag{42}$$

and

$$E_{\text{cnvx}-\text{prior}}(\mathbf{f}^{(k,l-1)}) = \sum_j E_{\text{s}-\text{prior}}(f_j^{(k,l-1)}; \mathbf{f}^{(k,l-1)}) \tag{43}$$

Instead of descending on the original complete data MAP objective function, we descend on the combination of the complete data negative log-likelihood and surrogate prior to get an update $\mathbf{f}^{(k,l)}$

$$\mathbf{f}^{(k,l)} = \arg\min_{\mathbf{f}>\mathbf{0}} - \sum_{m=1}^{L} \sum_{i \in S_m} \sum_{j=1}^{N} \mathcal{C}_{ij}^{(k,l)} \log H_{ij} f_j + \sum_{ij} H_{ij} f_j + \sum_j E_{\text{s}-\text{prior}}(f_j; \mathbf{f}^{(k,l-1)}) \tag{44}$$

with $E_{\text{s}-\text{prior}}(f_j; \mathbf{f}^{(k,l-1)})$ yet to be specified. In (44), we have included only terms that are dependent on $\mathbf{f}$. An important and subtle point is that the summation over all subsets is denoted as $\sum_{m=1}^{L} \sum_{i \in S_m}$ and not $\sum_{l=1}^{L} \sum_{i \in S_l}$. At subiteration l, we have $\mathcal{C}^{(k,l)}$ and the objective function w.r.t. $\mathbf{f}$ requires a summation over all subsets which is denoted by $\sum_{m=1}^{L} \sum_{i \in S_m}$. Please note that the surrogate prior is constructed anew at each step (k,l) which is why the update $\mathcal{C}_{ij}^{(k,l)}$ appears in (44). Since the surrogate prior objective is always constructed anew at each step to be above the original prior objective, a descent step in $\mathbf{f}$ taken w.r.t. the separable surrogate objective in (44) is guaranteed to be a descent step in $\mathbf{f}$ w.r.t. the original complete data MAP objective.

We now specialize $E_{\rm cnvx-prior}(\mathbf{f})$ to the quadratic prior of (2) repeated here:

$$E_{\rm prior}(\mathbf{f}) = \beta \sum_{jj'} w_{jj'}(f_j - f_{j'})^2. \tag{45}$$

We construct the surrogate to have the properties in (42) and (43). At step (k,l), assuming that we have $\mathbf{f}^{(k,l-1)}$, we observe that [20, 21]

$$(f_j - f_{j'})^2 \le \frac{1}{2}\left[(2f_j - f_j^{(k,l-1)} - f_{j'}^{(k,l-1)})^2 + (2f_{j'} - f_j^{(k,l-1)} - f_{j'}^{(k,l-1)})^2\right] \tag{46}$$

which leads to the following separable surrogate objective function w.r.t. $\mathbf{f}$ at time step (k,l)

$$\begin{aligned} E_{\rm s-MAP}(\mathbf{f};\mathbf{f}^{(k,l-1)},\mathcal{C}^{(k,l)}) \;=\; & -\sum_{m=1}^{L}\sum_{i\in S_m}\sum_{j=1}^{N}\mathcal{C}_{ij}^{(k,l)}\log\, f_j + \sum_{ij} H_{ij} f_j \\ & + \frac{\beta}{2}\sum_{jj'} w_{jj'}\left[(2f_j - f_j^{(k,l-1)} - f_{j'}^{(k,l-1)})^2 + (2f_{j'} - f_j^{(k,l-1)} - f_{j'}^{(k,l-1)})^2\right]. \end{aligned} \tag{47}$$

Having specified the separable surrogate objective function w.r.t. $\mathbf{f}$ which depends on both $\mathcal{C}^{(k,l)}$ and $\mathbf{f}^{(k,l-1)}$, we now derive the update for $\mathbf{f}^{(k,l)}$. The objective function in (47) is separable w.r.t. f_j and the objective function w.r.t. each f_j can be written as:

$$E_{\rm s-MAP}^{(k,l)}(f_j;\mathbf{f}^{(k,l-1)},\mathcal{C}^{(k,l)}) = -\sum_i \mathcal{C}_{ij}^{(k,l)}\log\, f_j + \sum_i H_{ij} f_j + \frac{\beta}{2}\sum_{j'} v_{jj'}(2f_j - f_j^{(k,l-1)} - f_{j'}^{(k,l-1)})^2 \tag{48}$$

where $v_{jj'} \stackrel{\rm def}{=} (w_{jj'} + w_{j'j})$. Since this objective function is a simple, convex 1-D objective w.r.t. f_j, we can trivially find that $f_j^{(k,l)}$ which minimizes it. Setting the derivative of $E_{\rm s-MAP}^{(k,l)}(f_j;\mathbf{f}^{(k,l-1)},\mathcal{C}^{(k,l)})$ w.r.t. f_j to zero, we get

$$-\frac{B_j^{(k,l)}}{f_j^{(k,l)}} + D_j + 2\beta\sum_{j'} v_{jj'}(2f_j^{(k,l)} - f_j^{(k,l-1)} - f_{j'}^{(k,l-1)}) = 0 \tag{49}$$

where $B_j^{(k,l)} = \sum_i \mathcal{C}_{ij}^{(k,l)} = \sum_{m=1}^{L}\sum_{i\in S_m}\mathcal{C}_{ij}^{(k,l)}$ as before, with $D_j = \sum_i H_{ij}$ being the sensitivity as before. We now solve for $f_j^{(k,l)}$ in (49) to get

$$f_j^{(k,l)} = \frac{2\beta\sum_{j'} v_{jj'}(f_j^{(k,l-1)} + f_{j'}^{(k,l-1)}) - D_j + \sqrt{[2\beta\sum_{j'} v_{jj'}(f_j^{(k,l-1)} + f_{j'}^{(k,l-1)}) - D_j]^2 + 16\beta B_j^{(k,l)}\sum_{j'} v_{jj'}}}{8\beta\sum_{j'} v_{jj'}}. \tag{50}$$

Note that this update $f_j^{(k,l)}$ is guaranteed to be positive since we take only the positive root when solving the quadratic equation. The negative root leads to a spurious negative solution.

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

As before, the zero bins $g_i = 0$ and the zero voxels, i.e. the voxels f_j known to be zero *a priori*, are eliminated from the problem. The corresponding $\mathcal{C}_{ij}$ and B_j may also be removed.

6 Convergence Proof for COSEM-MAP

In this section, we 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 to the MAP problem as stated in (5).

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

According to ([1], Lemma 1) and ([24], 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 [5] 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 (5) is unique.

6.2 "Touching" property of $E_{\text{cmp-MAP}}$ and $E_{\text{inc-MAP}}$

In Sec. 3, we showed (20) that the incomplete and the complete versions of the ML objective "touch" at $\mathcal{C} = \mathcal{C}^{\text{sol}}(\mathbf{f})$.

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

where as in (19)

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

If we add a prior to $E_{\text{cmp-ML}}$ and to $E_{\text{inc-ML}}$, then it is easy to show (by the same steps used in deriving (20)) that this "touching" property holds for the MAP versions:

$$E_{\text{cmp-MAP}}(\mathcal{C}^{\text{sol}}(\mathbf{f}), \mathbf{f}) = E_{\text{inc-MAP}}(\mathbf{f}) \tag{51}$$

6.3 $E_{\text{cmp-MAP}}$ is bounded from below by $E_{\text{inc-MAP}}$

In this section, we demonstrate an interesting property on the boundedness of $E_{\text{cmp-MAP}}$ by $E_{\text{inc-MAP}}$, but this result is not needed in our proof.

We begin by re-writing the incomplete data penalized negative log-likelihood objective function in (6) using the ordered subset notation and adding an extra entropy term $\sum_i g_i \log g_i - \sum_i g_i$ (which is independent of $\mathbf{f}$) for reasons that will shortly become obvious:

$$E_{\text{inc-MAP}}(\mathbf{f}) = \sum_{ij} H_{ij} f_j - \sum_{m=1}^{L} \sum_{i \in S_m} g_i \log \sum_{j=1}^{N} H_{ij} f_j + \sum_i g_i \log g_i - \sum_i g_i + \beta \sum_{jj'} w_{jj'} (f_j - f_{j'})^2 \tag{52}$$

The additional entropy term does not affect the fixed points.

First, we show that the complete data MAP objective function in (38) is bounded from below by $E_{\text{inc-MAP}}(\mathbf{f})$. Using the fact that $x \log \frac{x}{y} - x + y \geq 0$ for $x, y > 0$, we get

$$E_{\text{cmp-MAP}}(\mathcal{C}, \mathbf{f}) - E_{\text{inc-MAP}}(\mathbf{f}) = \sum_{ij} \mathcal{C}_{ij} \log \frac{\mathcal{C}_{ij}}{g_i \frac{H_{ij} f_j}{\sum_n H_{in} f_n}} - \sum_{ij} \mathcal{C}_{ij} + \sum_{ij} g_i \frac{H_{ij} f_j}{\sum_n H_{in} f_n} \geq 0 \tag{53}$$

with equality occurring only at $\mathcal{C}_{ij} = g_i \frac{H_{ij} f_j}{\sum_n H_{in} f_n}, \forall ij$. We have used the constraint $\sum_j \mathcal{C}_{ij} = g_i, \forall i$. The extra entropy term $\sum_i g_i \log g_i$ is useful here. Without the extra entropy term, we could not have easily obtained (53). Therefore, we have that $E_{\text{cmp-MAP}}(\mathcal{C}, \mathbf{f}) \geq E_{\text{inc-MAP}}(\mathbf{f})$. Consequently, the complete data MAP objective function in (38) is bounded from below by the original incomplete data MAP objective function in (6).

We can also derive this boundedness condition in another way. From Sec. 3, we observe

$$\mathcal{C}^{\text{sol}}(\mathbf{f}) = \arg\min_{\mathcal{C}} E_{\text{cmp-MAP}}(\mathcal{C}, \mathbf{f}, \nu) \tag{54}$$

Then

$$E_{\text{cmp-MAP}}(\mathcal{C}, \mathbf{f}) \geq \min_{\mathcal{C}} E_{\text{cmp-MAP}}(\mathcal{C}, \mathbf{f}) = E_{\text{cmp-MAP}}(\mathcal{C}^{\text{sol}}(\mathbf{f}), \mathbf{f}) = E_{\text{cmp-MAP}}(\mathbf{f}) \tag{55}$$

where the first equality follows from (54) and the second from (51).

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

We will show that the complete data MAP objective function (38) is convex in both $\mathcal{C}$ and $\mathbf{f}$. In fact, we will show that $E_{\text{cmp}-\text{MAP}}(\mathcal{C}, \mathbf{f})$ is convex, whether or not the constraints on $\mathcal{C}$ hold. Since the prior is convex w.r.t. $\mathbf{f}$, we need only show that the complete data negative log-likelihood is convex w.r.t. $(\mathcal{C}, \mathbf{f})$. If each term $\mathcal{C}_{ij} \log \frac{\mathcal{C}_{ij}}{H_{ij} f_j} - \mathcal{C}_{ij} + H_{ij} f_j$ is convex w.r.t. both $\mathcal{C}_{ij}$ and $H_{ij} f_j$, then the complete data likelihood is convex w.r.t. $(\mathcal{C}, \mathbf{f})$ since it is a sum of such terms. Note that in order to show convexity w.r.t. $\mathcal{C}_{ij}$ and f_j, it is enough to show convexity w.r.t. $\mathcal{C}_{ij}$ and $H_{ij} f_j$. This is because H_{ij} is just a constant (independent of f_j) and hence, convexity w.r.t. both $\mathcal{C}_{ij}$ and $H_{ij} f_j$ implies convexity w.r.t. both $\mathcal{C}_{ij}$ and f_j. We need to show that

$$\phi(x, y) \stackrel{\text{def}}{=} x \log \frac{x}{y} - x + y \tag{56}$$

is convex w.r.t. x and y for $x, y > 0$. (Here, we associate x with $\mathcal{C}_{ij}$ and y with $H_{ij} f_j$.) For this, we need

$$\phi[\alpha x + (1-\alpha) z, \alpha y + (1-\alpha) w] - \alpha \phi(x, y) - (1-\alpha) \phi(z, w) \leq 0 \tag{57}$$

for $\alpha \in [0, 1]$ and $x, y, z, w > 0$. Substituting (56) in (57), we get

$$\begin{aligned}
&[\alpha x + (1-\alpha) z] \log \frac{\alpha x + (1-\alpha) z}{\alpha y + (1-\alpha) w} - \alpha x - (1-\alpha) z + \alpha y + (1-\alpha) w - \alpha [x \log \frac{x}{y} - x + y] \\
&-(1-\alpha)[z \log \frac{z}{w} - z + w] = \alpha x \log [\frac{\alpha x + (1-\alpha) z}{\alpha y + (1-\alpha) w} \cdot \frac{y}{x}] + (1-\alpha) z \log [\frac{\alpha x + (1-\alpha) z}{\alpha y + (1-\alpha) w} \cdot \frac{w}{z}] \\
&\leq \alpha x \left[\frac{\alpha x + (1-\alpha) z}{\alpha y + (1-\alpha) w} \cdot \frac{y}{x} - 1 \right] + (1-\alpha) z \left[\frac{\alpha x + (1-\alpha) z}{\alpha y + (1-\alpha) w} \cdot \frac{w}{z} - 1 \right] \\
&\leq 0
\end{aligned} \tag{58}$$

where we have used the fact that $\log x \leq x - 1$. A similar result on the convexity of the KL-divergence involving probabilities is given in [11]. We have shown that the complete data negative log- likelihood and hence the complete data MAP objective function is convex w.r.t. both $\mathcal{C}$ and $\mathbf{f}$.

6.5 The change in $E_{\text{cmp}-\text{MAP}}$ at each COSEM-MAP sub-iteration

We define $\Delta E^{(k,l)}_{\text{cmp}-\text{MAP}} \stackrel{\text{def}}{=} E_{\text{cmp}-\text{MAP}}(\mathcal{C}^{(k,l-1)}, \mathbf{f}^{(k,l-1)}) - E_{\text{cmp}-\text{MAP}}(\mathcal{C}^{(k,l)}, \mathbf{f}^{(k,l)})$ which is essentially the change in the objective (38) from step $(k, l-1)$ to step (k, l). In this section, we first show that the $\mathcal{C}^{(k,l)}$ and $\mathbf{f}^{(k,l)}$ COSEM-MAP updates are descent steps in the complete data MAP objective function (38), i.e. $\Delta E^{(k,l)}_{\text{cmp}-\text{MAP}} \geq 0$, and that useful conditions obtain when $\Delta E^{(k,l)}_{\text{cmp}-\text{MAP}} = 0$.

Using (38), we obtain

$$\begin{aligned}\Delta E_{\text{cmp-MAP}}^{(k,l)} &= \sum_{m=1}^{L}\sum_{i\in S_m}\sum_{j=1}^{N}\left[\mathcal{C}_{ij}^{(k,l-1)}\log\frac{\mathcal{C}_{ij}^{(k,l-1)}}{H_{ij}f_j^{(k,l-1)}}-\mathcal{C}_{ij}^{(k,l)}\log\frac{\mathcal{C}_{ij}^{(k,l)}}{H_{ij}f_j^{(k,l)}}\right]+\sum_{ij}H_{ij}(f_j^{(k,l-1)}-f_j^{(k,l)})\\ &+\beta\sum_{jj'}w_{jj'}\left[(f_j^{(k,l-1)}-f_{j'}^{(k,l-1)})^2-(f_j^{(k,l)}-f_{j'}^{(k,l)})^2\right]. \qquad (59)\end{aligned}$$

The basic inequality (46) that is used in the surrogate transformation of the prior is repeated here:

$$(f_j-f_{j'})^2\le\frac{1}{2}\left[(2f_j-f_j^{(k,l-1)}-f_{j'}^{(k,l-1)})^2+(2f_{j'}-f_j^{(k,l-1)}-f_{j'}^{(k,l-1)})^2\right]. \qquad (60)$$

From (60), we get that

$$-(f_j^{(k,l)}-f_{j'}^{(k,l)})^2\ge-\frac{1}{2}[(2f_j^{(k,l)}-f_j^{(k,l-1)}-f_{j'}^{(k,l-1)})^2+(2f_{j'}^{(k,l)}-f_j^{(k,l-1)}-f_{j'}^{(k,l-1)})^2]. \qquad (61)$$

Using (61),(59), we may write

$$\begin{aligned}\Delta E_{\text{cmp-MAP}}^{(k,l)} &\ge \sum_{m=1}^{L}\sum_{i\in S_m}\sum_{j=1}^{N}\left[\mathcal{C}_{ij}^{(k,l-1)}\log\frac{\mathcal{C}_{ij}^{(k,l-1)}}{H_{ij}f_j^{(k,l-1)}}-\mathcal{C}_{ij}^{(k,l)}\log\frac{\mathcal{C}_{ij}^{(k,l)}}{H_{ij}f_j^{(k,l)}}\right]+\sum_{ij}H_{ij}(f_j^{(k,l-1)}-f_j^{(k,l)})\\ &+\frac{\beta}{2}\sum_{jj'}w_{jj'}\left[(2f_j^{(k,l-1)}-f_j^{(k,l-1)}-f_{j'}^{(k,l-1)})^2-(2f_j^{(k,l)}-f_j^{(k,l-1)}-f_{j'}^{(k,l-1)})^2\right]\\ &+\frac{\beta}{2}\sum_{jj'}w_{jj'}\left[(2f_{j'}^{(k,l-1)}-f_j^{(k,l-1)}-f_{j'}^{(k,l-1)})^2-(2f_{j'}^{(k,l)}-f_j^{(k,l-1)}-f_{j'}^{(k,l-1)})^2\right]. \qquad (62)\end{aligned}$$

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, we get

$$\begin{aligned}\Delta E_{\text{cmp-MAP}}^{(k,l)} &\ge \sum_{i\in S_l}\sum_{j=1}^{N}\left[\mathcal{C}_{ij}^{(k,l-1)}\log\frac{\mathcal{C}_{ij}^{(k,l-1)}}{\mathcal{C}_{ij}^{(k,l)}}+(\mathcal{C}_{ij}^{(k,l-1)}-\mathcal{C}_{ij}^{(k,l)})\log\frac{\mathcal{C}_{ij}^{(k,l)}}{H_{ij}f_j^{(k,l-1)}}\right]\\ &+\sum_{m=1}^{L}\sum_{i\in S_m}\sum_{j=1}^{N}\left[\mathcal{C}_{ij}^{(k,l)}\log\frac{f_j^{(k,l)}}{f_j^{(k,l-1)}}+H_{ij}(f_j^{(k,l-1)}-f_j^{(k,l)})\right]\\ &+\frac{\beta}{2}\sum_{jj'}w_{jj'}\left[(2f_j^{(k,l-1)}-f_j^{(k,l-1)}-f_{j'}^{(k,l-1)})^2-(2f_j^{(k,l)}-f_j^{(k,l-1)}-f_{j'}^{(k,l-1)})^2\right]\\ &+\frac{\beta}{2}\sum_{jj'}w_{jj'}\left[(2f_{j'}^{(k,l-1)}-f_j^{(k,l-1)}-f_{j'}^{(k,l-1)})^2-(2f_{j'}^{(k,l)}-f_j^{(k,l-1)}-f_{j'}^{(k,l-1)})^2\right]. \qquad (63)\end{aligned}$$

From the update equation (40) for $\mathcal{C}^{(k,l)}$ which is $\mathcal{C}_{ij}^{(k,l)} = g_i \frac{H_{ij} f_j^{(k,l-1)}}{\sum_n H_{in} f_n^{(k,l-1)}}, \forall i \in S_l, \forall j$, we get

$$\sum_j (\mathcal{C}_{ij}^{(k,l-1)} - \mathcal{C}_{ij}^{(k,l)}) \log \frac{\mathcal{C}_{ij}^{(k,l)}}{H_{ij} f_j^{(k,l-1)}} = \sum_j (\mathcal{C}_{ij}^{(k,l-1)} - \mathcal{C}_{ij}^{(k,l)}) \log \frac{g_i}{\sum_n H_{in} f_n^{(k,l-1)}}. \tag{64}$$

Using (64), we may write

$$\begin{aligned} \Delta E_{\text{cmp-MAP}}^{(k,l)} \geq \textstyle\sum_{i \in S_l} & \sum_{j=1}^{N} \left[\mathcal{C}_{ij}^{(k,l-1)} \log \frac{\mathcal{C}_{ij}^{(k,l-1)}}{\mathcal{C}_{ij}^{(k,l)}} + (\mathcal{C}_{ij}^{(k,l-1)} - \mathcal{C}_{ij}^{(k,l)}) \log \frac{g_i}{\sum_n H_{in} f_n^{(k,l-1)}} \right] \\ & + \sum_{m=1}^{L} \sum_{i \in S_m} \sum_{j=1}^{N} \left[\mathcal{C}_{ij}^{(k,l)} \log \frac{f_j^{(k,l)}}{f_j^{(k,l-1)}} + H_{ij} (f_j^{(k,l-1)} - f_j^{(k,l)}) \right] \\ & + \frac{\beta}{2} \sum_{jj'} w_{jj'} \left[(2f_j^{(k,l-1)} - f_j^{(k,l-1)} - f_{j'}^{(k,l-1)})^2 - (2f_j^{(k,l)} - f_j^{(k,l-1)} - f_{j'}^{(k,l-1)})^2 \right] \\ & + \frac{\beta}{2} \sum_{jj'} w_{jj'} \left[(2f_{j'}^{(k,l-1)} - f_j^{(k,l-1)} - f_{j'}^{(k,l-1)})^2 - (2f_{j'}^{(k,l)} - f_j^{(k,l-1)} - f_{j'}^{(k,l-1)})^2 \right]. \end{aligned} \tag{65}$$

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

$$\sum_{i \in S_l} \sum_{j=1}^{N} (\mathcal{C}_{ij}^{(k,l-1)} - \mathcal{C}_{ij}^{(k,l)}) \log \frac{g_i}{\sum_n H_{in} f_n^{(k,l-1)}} = 0. \tag{66}$$

Using (65) and (66), we may write

$$\begin{aligned} \Delta E_{\text{cmp-MAP}}^{(k,l)} \geq & \sum_{i \in S_l} \sum_{j=1}^{N} \mathcal{C}_{ij}^{(k,l-1)} \log \frac{\mathcal{C}_{ij}^{(k,l-1)}}{\mathcal{C}_{ij}^{(k,l)}} + \sum_{m=1}^{L} \sum_{i \in S_m} \sum_{j=1}^{N} \left[\mathcal{C}_{ij}^{(k,l)} \log \frac{f_j^{(k,l)}}{f_j^{(k,l-1)}} + H_{ij} (f_j^{(k,l-1)} - f_j^{(k,l)}) \right] \\ & + \frac{\beta}{2} \sum_{jj'} w_{jj'} \left[(2f_j^{(k,l-1)} - f_j^{(k,l-1)} - f_{j'}^{(k,l-1)})^2 - (2f_j^{(k,l)} - f_j^{(k,l-1)} - f_{j'}^{(k,l-1)})^2 \right] \\ & + \frac{\beta}{2} \sum_{jj'} w_{jj'} \left[(2f_{j'}^{(k,l-1)} - f_j^{(k,l-1)} - f_{j'}^{(k,l-1)})^2 - (2f_{j'}^{(k,l)} - f_j^{(k,l-1)} - f_{j'}^{(k,l-1)})^2 \right]. \end{aligned} \tag{67}$$

We will now show that the first term on the right of (67) can be expressed as a non-negative I-divergence of the form $x \log \frac{x}{y} - x + y$. Since $\sum_j \mathcal{C}_{ij}^{(k,l-1)} = \sum_j \mathcal{C}_{ij}^{(k,l)} = g_i$, we may write

$$\sum_{j=1}^{N} \mathcal{C}_{ij}^{(k,l-1)} \log \frac{\mathcal{C}_{ij}^{(k,l-1)}}{\mathcal{C}_{ij}^{(k,l)}} = \sum_{j=1}^{N} [\mathcal{C}_{ij}^{(k,l-1)} \log \frac{\mathcal{C}_{ij}^{(k,l-1)}}{\mathcal{C}_{ij}^{(k,l)}} - \mathcal{C}_{ij}^{(k,l-1)} + \mathcal{C}_{ij}^{(k.l)}] \geq 0 \tag{68}$$

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

The update equation (50) for $\mathbf{f}$ in COSEM-MAP is chosen such that the objective function (47) is minimized.

For convenience, we repeat (47) here:

$$E_{\text{s-MAP}}(\mathbf{f};\mathbf{f}^{(k,l-1)},\mathcal{C}^{(k,l)}) = -\sum_{m=1}^{L}\sum_{i\in S_m}\sum_{j=1}^{N}\mathcal{C}_{ij}^{(k,l)}\log f_j + \sum_{ij} H_{ij} f_j + \frac{\beta}{2}\sum_{jj'} w_{jj'}\left[(2f_j - f_j^{(k,l-1)} - f_{j'}^{(k,l-1)})^2 + (2f_{j'} - f_j^{(k,l-1)} - f_{j'}^{(k,l-1)})^2\right]. \quad (69)$$

Apart from the terms forming the I-divergence, the remaining terms on the right of (67) can be seen to be

$$\begin{aligned} &E_{\text{s-MAP}}(\mathbf{f}^{(k,l-1)};\mathbf{f}^{(k,l-1)},\mathcal{C}^{(k,l)}) - E_{\text{s-MAP}}(\mathbf{f}^{(k,l)};\mathbf{f}^{(k,l-1)},\mathcal{C}^{(k,l)}) \\ &= \sum_{m=1}^{L}\sum_{i\in S_m}\sum_{j=1}^{N}\left[\mathcal{C}_{ij}^{(k,l)}\log\frac{f_j^{(k,l)}}{f_j^{(k,l-1)}} + H_{ij}(f_j^{(k,l-1)} - f_j^{(k,l)})\right] \\ &+\frac{\beta}{2}\sum_{jj'} w_{jj'}\left[(2f_j^{(k,l-1)} - f_j^{(k,l-1)} - f_{j'}^{(k,l-1)})^2 - (2f_j^{(k,l)} - f_j^{(k,l-1)} - f_{j'}^{(k,l-1)})^2\right] \\ &+\frac{\beta}{2}\sum_{jj'} w_{jj'}\left[(2f_{j'}^{(k,l-1)} - f_j^{(k,l-1)} - f_{j'}^{(k,l-1)})^2 - (2f_{j'}^{(k,l)} - f_j^{(k,l-1)} - f_{j'}^{(k,l-1)})^2\right] \geq 0 \end{aligned} \quad (70)$$

with equality occurring if $\mathbf{f}^{(k,l)} = \mathbf{f}^{(k,l-1)}$. This is due to the fact that the update equation in (50) is guaranteed to minimize $E_{\text{s-MAP}}(\mathbf{f};\mathbf{f}^{(k,l-1)},\mathcal{C}^{(k,l)})$. Consequently, the change in the complete data MAP objective function at step (k,l) is

$$\begin{aligned} \Delta E_{\text{cmp-MAP}}^{(k,l)} &\geq \sum_{i\in S_l}\sum_{j=1}^{N}[\mathcal{C}_{ij}^{(k,l-1)}\log\frac{\mathcal{C}_{ij}^{(k,l-1)}}{\mathcal{C}_{ij}^{(k,l)}} - \mathcal{C}_{ij}^{(k,l-1)} + \mathcal{C}_{ij}^{(k.l)}] \\ &\quad + E_{\text{s-MAP}}(\mathbf{f}^{(k,l-1)};\mathbf{f}^{(k,l-1)},\mathcal{C}^{(k,l)}) - E_{\text{s-MAP}}(\mathbf{f}^{(k,l)};\mathbf{f}^{(k,l-1)},\mathcal{C}^{(k,l)}) \\ &\geq 0 \end{aligned} \quad (71)$$

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* $f_j^{(k,l)} = f_j^{(k,l-1)}$, $\forall j$.

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

$$\begin{aligned} \Delta E_{\text{cmp-MAP}}^{(k)} &\stackrel{\text{def}}{=} E_{\text{cmp-MAP}}(\mathcal{C}^{(k-1,L)},\mathbf{f}^{(k-1,L)}) - E_{\text{cmp-MAP}}(\mathcal{C}^{(k,L)},\mathbf{f}^{(k,L)}),\ k>0 \\ &= \sum_{l=1}^{L}\Delta E_{\text{cmp-MAP}}^{(k,l)} \geq 0 \end{aligned} \quad (72)$$

$\Delta E_{\text{cmp-MAP}}^{(k)}$ 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 $f_j^{(k,l)} = f_j^{(k,l-1)}$, $\forall j$, $\forall l \in \{1,\ldots,L\}$.

Define k^* as the iteration at which $\Delta E_{\text{cmp-MAP}}^{(k^*)} = 0$. Then $\mathcal{C}^{(k^*,1)} = \mathcal{C}^{(k^*,2)} = \cdots = \mathcal{C}^{(k^*,L)} \stackrel{\text{def}}{=} \mathcal{C}^*$. We similarly get $\mathbf{f}^{(k^*,1)} = \mathbf{f}^{(k^*,2)} = \cdots = \mathbf{f}^{(k^*,L)} \stackrel{\text{def}}{=} \mathbf{f}^*$. 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 update (40), we get:

$$\mathcal{C}^*_{ij} = \mathcal{C}^{(k^*,l)}_{ij} = g_i \frac{H_{ij} f_j^{(k^*,l-1)}}{\sum_n H_{in} f_n^{(k^*,l-1)}} \, \forall i, \, \forall j, \, \forall l \in \{1, \cdots, L\} = g_i \frac{H_{ij} f^*_j}{\sum_n H_{in} f^*_n} \tag{73}$$

Therefore, from (19), we get

$$\mathcal{C}^* = \mathcal{C}^{\rm sol}(\mathbf{f}^*) \tag{74}$$

From the definition of $\mathcal{C}^{\rm sol}(\mathbf{f})$ in Sec. 3, we may see that (74) is equivalent to the two conditions $\nabla_{\mathcal{C}} E_{\rm cmp-MAP}(\mathcal{C}, \mathbf{f}, \nu)|_{(\mathcal{C}^*, \mathbf{f}^*)} = \mathbf{0}$ and $\nabla_\nu E_{\rm cmp-MAP}(\mathcal{C}, \mathbf{f}, \nu)|_{(\mathcal{C}^*, \mathbf{f}^*)} = \mathbf{0}$.

Now, we will show that $\nabla_{\mathbf{f}} E_{\rm cmp-MAP}(\mathcal{C}, \mathbf{f}, \nu)|_{(\mathcal{C}^*, \mathbf{f}^*)} = \mathbf{0}$. We need to show that

$$\frac{\partial E_{\rm cmp-MAP}(\mathcal{C}, \mathbf{f}, \nu)}{\partial f_j}|_{(\mathcal{C}^*, \mathbf{f}^*)} = -\frac{\sum_i \mathcal{C}^*_{ij}}{f^*_j} + D_j + 2\beta \sum_{j' \in \mathcal{N}(j)} v_{jj'} (f^*_j - f^*_{j'}) = 0 \;\; \forall j \tag{75}$$

Since the $\mathbf{f}$ updates are chosen to satisfy (49), we may write

$$-\frac{\sum_i \mathcal{C}^{(k^*,l)}_{ij}}{f_j^{(k^*,l)}} + D_j + 2\beta \sum_{j' \in \mathcal{N}(j)} v_{jj'} (2 f_j^{(k^*,l)} - f_j^{(k^*,l-1)} - f_{j'}^{(k^*,l-1)}) = 0 \;\; \forall j \tag{76}$$

Since $\mathbf{f}^{(k^*,1)} = \mathbf{f}^{(k^*,2)} = \cdots = \mathbf{f}^{(k^*,L)} = \mathbf{f}^*$, we may see that (75) follows from (76).

Thus, we have shown that the full gradient of $E_{\rm cmp-MAP}(\mathcal{C}, \mathbf{f}, \nu)$ at $(\mathcal{C}^*, \mathbf{f}^*)$ is zero, i.e. $\nabla E_{\rm cmp-MAP}(\mathcal{C}, \mathbf{f}, \nu)|_{(\mathcal{C}^*, \mathbf{f}^*, \nu)} = \mathbf{0}$.

6.6 Convergence to the fixed point of $E_{\rm inc-MAP}$

Define Γ as the set (possibly singleton) of global minima of $E_{\rm cmp-MAP}(\mathcal{C}, \mathbf{f})$. We may argue that $(\mathcal{C}^*, \mathbf{f}^*) \in \Gamma$ as follows: $\Delta E^{(k,l)}_{\rm cmp-MAP} = 0$ is equivalent to the fact that $\mathcal{C}^{(k,l)}_{ij}$ and $f_j^{(k,l)}$ are no longer changing. We note that an algorithm can lead to a repeating $(\mathcal{C}^{(k,l)}_{ij}, f_j^{(k,l)})$ that is *not* an element of Γ. But this cannot happen because the COSEM-MAP algorithm is a grouped coordinate descent algorithm, and because $E_{\rm cmp-MAP}$ is convex. The COSEM-MAP algorithm would continue to descend along coordinate directions whenever possible. Hence, if one approaches a condition where $(\mathcal{C}^{(k,l)}_{ij}, f_j^{(k,l)})$ 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 $E_{\rm inc-MAP}(\mathbf{f})$. We prove this by contradiction. Assume an $\check{\mathbf{f}}$ such that $E_{\rm inc-MAP}(\mathbf{f}^*) > E_{\rm inc-MAP}(\check{\mathbf{f}})$. Since $\mathcal{C}^* = \mathcal{C}^{\rm sol}(\mathbf{f}^*)$, we have by the touching property in Sec. 6.2 that $E_{\rm cmp-MAP}(\mathcal{C}^{\rm sol}(\mathbf{f}^*), \mathbf{f}^*) = E_{\rm inc-MAP}(\mathbf{f}^*)$. This implies

$$E_{\rm cmp-MAP}(\mathcal{C}^{\rm sol}(\mathbf{f}^*), \mathbf{f}^*) = E_{\rm inc-MAP}(\mathbf{f}^*) > E_{\rm inc-MAP}(\check{\mathbf{f}}) = E_{\rm cmp-MAP}(\mathcal{C}^{\rm sol}(\check{\mathbf{f}}), \check{\mathbf{f}}), \tag{77}$$

but this is a contradiction because $(\mathcal{C}^{\rm sol}(\mathbf{f}^*), \mathbf{f}^*) \in \Gamma$ is a global minimum of $E_{\rm cmp-MAP}$. In fact, $\mathbf{f}^*$ is the *unique* global minimum $\hat{\mathbf{f}}$ of $E_{\rm inc-MAP}(\mathbf{f})$, because $E_{\rm inc-MAP}(\mathbf{f})$ is strictly convex, as we established in Sec. 6.1.

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 (5), which is the result we have sought. Thus, minimizing the complete data objective (38) (in our case, by the COSEM-MAP algorithm) yields a solution $(\mathcal{C}^*, \mathbf{f}^*)$, where $\mathbf{f}^* = \hat{\mathbf{f}}$, our sought-after solution in (5).

6.7 Strict convexity of $E_{\text{cmp}-\text{MAP}}$

Though this result is not needed, we now demonstrate the interesting consequence that $E_{\text{cmp}-\text{MAP}}(\mathcal{C}, \mathbf{f}, \nu)$ is, in fact, strictly convex. First we show that $\forall(\mathcal{C}, \mathbf{f}) \in \Gamma, \mathcal{C} = \mathcal{C}^{\text{sol}}(\mathbf{f})$. This follows from the fact that $\forall(\mathcal{C}, \mathbf{f}) \in \Gamma$, $\nabla_{\mathcal{C}} E_{\text{cmp}-\text{MAP}}(\mathcal{C}, \mathbf{f}, \nu) = \mathbf{0}$ and $\nabla_{\nu} E_{\text{cmp}-\text{MAP}}(\mathcal{C}, \mathbf{f}, \nu) = \mathbf{0}$. Please refer to Sec. 3 for details.

Next we will show that Γ is a singleton set. Let $(\mathcal{C}', \mathbf{f}') \in \Gamma$ where $\mathbf{f}' \neq \mathbf{f}^*$. Then $\mathcal{C}' = \mathcal{C}^{\text{sol}}(\mathbf{f}')$. By the definition of Γ, $E_{\text{cmp}-\text{MAP}}(\mathcal{C}^{\text{sol}}(\mathbf{f}'), \mathbf{f}') = E_{\text{cmp}-\text{MAP}}(\mathcal{C}^{\text{sol}}(\mathbf{f}^*), \mathbf{f}^*)$. Hence by the touching property in Sec. 6.2, we get $E_{\text{inc}-\text{MAP}}(\mathbf{f}') = E_{\text{inc}-\text{MAP}}(\mathbf{f}^*)$ which contradicts the fact that $\mathbf{f}^*$ is the unique global minimum of $E_{\text{inc}-\text{MAP}}$. So, since $E_{\text{cmp}-\text{MAP}}$ has been shown to be convex (in Sec. 6.4) and it has a unique minimum, it is strictly convex.

7 Convergence Proof for COSEM-ML

This proof differs somewhat from the proof in Sec. 6 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 (3).

7.1 "Touching" property of $E_{\text{cmp}-\text{ML}}$ and $E_{\text{inc}-\text{ML}}$

In Sec. 3, we showed (20) that

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

where as in (19)

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

7.2 $E_{\text{cmp}-\text{ML}}$ is bounded from below by $E_{\text{inc}-\text{ML}}$

In this section, we mention an interesting property, but it is not needed in our proof. This follows from the results in Sec. 6.3. If we remove the prior from the expressions in Sec. 6.3, we easily obtain that (53) applies to the ML case. The proof then follows.

7.3 Convexity of the $E_{\text{cmp}-\text{ML}}(\mathcal{C}, \mathbf{f})$ objective

This follows from the results in Sec. 6.4. In particular, $E_{\text{cmp}-\text{MAP}} = E_{\text{cmp}-\text{ML}} + E_{\text{prior}}$. E_{prior} is convex and in Sec. 6.4, we showed that $E_{\text{cmp}-\text{ML}}$ was convex.

7.4 The change in $E_{\text{cmp}-\text{ML}}$ at each COSEM-ML sub-iteration

We define $\Delta E_{\text{cmp}-\text{ML}}^{(k,l)} \overset{\text{def}}{=} E_{\text{cmp}-\text{ML}}(\mathcal{C}^{(k,l-1)}, \mathbf{f}^{(k,l-1)}) - E_{\text{cmp}-\text{ML}}(\mathcal{C}^{(k,l)}, \mathbf{f}^{(k,l)})$ which is essentially the change in the objective (21) from step $(k, l-1)$ to step (k, l). In this section, we first show that the $\mathcal{C}^{(k,l)}$ and $\mathbf{f}^{(k,l)}$ COSEM-ML updates are descent steps in the complete data ML objective function (21), i.e. $\Delta E^{(k,l)}_{\text{cmp-ML}} \geq 0$, and that useful conditions obtain when $\Delta E^{(k,l)}_{\text{cmp-ML}} = 0$.

Using (21), we obtain

$$\Delta E^{(k,l)}_{\text{cmp-ML}} = \sum_{m=1}^{L} \sum_{i \in S_m} \sum_{j=1}^{N} \left[\mathcal{C}_{ij}^{(k,l-1)} \log \frac{\mathcal{C}_{ij}^{(k,l-1)}}{H_{ij} f_j^{(k,l-1)}} - \mathcal{C}_{ij}^{(k,l)} \log \frac{\mathcal{C}_{ij}^{(k,l)}}{H_{ij} f_j^{(k,l)}} \right] + \sum_{ij} H_{ij} (f_j^{(k,l-1)} - f_j^{(k,l)}). \quad (78)$$

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, we get

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

From the update equation (34) for $\mathcal{C}^{(k,l)}$ which is $\mathcal{C}_{ij}^{(k,l)} = g_i \frac{H_{ij} f_j^{(k,l-1)}}{\sum_n H_{in} f_n^{(k,l-1)}}, \forall i \in S_l, \forall j$, we get

$$\sum_j (\mathcal{C}_{ij}^{(k,l-1)} - \mathcal{C}_{ij}^{(k,l)}) \log \frac{\mathcal{C}_{ij}^{(k,l)}}{H_{ij} f_j^{(k,l-1)}} = \sum_j (\mathcal{C}_{ij}^{(k,l-1)} - \mathcal{C}_{ij}^{(k,l)}) \log \frac{g_i}{\sum_n H_{in} f_n^{(k,l-1)}}. \quad (79)$$

Using (79), we may write

$$\begin{aligned} \Delta E^{(k,l)}_{\text{cmp-ML}} \geq \textstyle\sum_{i \in S_l} & \sum_{j=1}^{N} \left[\mathcal{C}_{ij}^{(k,l-1)} \log \frac{\mathcal{C}_{ij}^{(k,l-1)}}{\mathcal{C}_{ij}^{(k,l)}} + (\mathcal{C}_{ij}^{(k,l-1)} - \mathcal{C}_{ij}^{(k,l)}) \log \frac{g_i}{\sum_n H_{in} f_n^{(k,l-1)}} \right] \\ & + \sum_{m=1}^{L} \sum_{i \in S_m} \sum_{j=1}^{N} \left[\mathcal{C}_{ij}^{(k,l)} \log \frac{f_j^{(k,l)}}{f_j^{(k,l-1)}} + H_{ij} (f_j^{(k,l-1)} - f_j^{(k,l)}) \right]. \end{aligned} \quad (80)$$

Also, from the fact that the constraint $\sum_j \mathcal{C}_{ij} = g_i$ is always satisfied, as in Sec. 6.5, we again have the identity

$$\sum_{i \in S_l} \sum_{j=1}^{N} (\mathcal{C}_{ij}^{(k,l-1)} - \mathcal{C}_{ij}^{(k,l)}) \log \frac{g_i}{\sum_n H_{in} f_n^{(k,l-1)}} = 0. \quad (81)$$

Using (80) and (81), we may write

$$\Delta E^{(k,l)}_{\text{cmp-ML}} \geq \sum_{i \in S_l} \sum_{j=1}^{N} \mathcal{C}_{ij}^{(k,l-1)} \log \frac{\mathcal{C}_{ij}^{(k,l-1)}}{\mathcal{C}_{ij}^{(k,l)}} + \sum_{m=1}^{L} \sum_{i \in S_m} \sum_{j=1}^{N} \left[\mathcal{C}_{ij}^{(k,l)} \log \frac{f_j^{(k,l)}}{f_j^{(k,l-1)}} + H_{ij} (f_j^{(k,l-1)} - f_j^{(k,l)}) \right]. \quad (82)$$

We will now show that the first term on the right of (82) can be expressed as a non-negative I-divergence of the form $x \log \frac{x}{y} - x + y$. Since $\sum_j \mathcal{C}_{ij}^{(k,l-1)} = \sum_j \mathcal{C}_{ij}^{(k,l)} = g_i$, we may write

$$\sum_{j=1}^{N} \mathcal{C}_{ij}^{(k,l-1)} \log \frac{\mathcal{C}_{ij}^{(k,l-1)}}{\mathcal{C}_{ij}^{(k,l)}} = \sum_{j=1}^{N} [\mathcal{C}_{ij}^{(k,l-1)} \log \frac{\mathcal{C}_{ij}^{(k,l-1)}}{\mathcal{C}_{ij}^{(k,l)}} - \mathcal{C}_{ij}^{(k,l-1)} + \mathcal{C}_{ij}^{(k.l)}] \geq 0 \tag{83}$$

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

We may observe that the update equation (30) for $\mathbf{f}$ in COSEM-ML is chosen such that the following objective function is minimized:

$$E_{\text{s-ML}}(\mathbf{f}; \mathcal{C}^{(k,l)}) \quad = \quad -\sum_{m=1}^{L} \sum_{i \in S_m} \sum_{j=1}^{N} \mathcal{C}_{ij}^{(k,l)} \log f_j + \sum_{ij} H_{ij} f_j. \tag{84}$$

Apart from the terms forming the I-divergence, the remaining terms on the right of (82) can be seen to be

$$\begin{aligned} & E_{\text{s-ML}}(\mathbf{f}^{(k,l-1)}; \mathcal{C}^{(k,l)}) - E_{\text{s-ML}}(\mathbf{f}^{(k,l)}; \mathcal{C}^{(k,l)}) \\ &= \sum_{m=1}^{L} \sum_{i \in S_m} \sum_{j=1}^{N} \left[\mathcal{C}_{ij}^{(k,l)} \log \frac{f_j^{(k,l)}}{f_j^{(k,l-1)}} + H_{ij}(f_j^{(k,l-1)} - f_j^{(k,l)}) \right] \quad > 0 \end{aligned} \tag{85}$$

with equality occurring if $\mathbf{f}^{(k,l)} = \mathbf{f}^{(k,l-1)}$. This is due to the fact that the update equation in (30) is guaranteed to minimize $E_{\text{s-ML}}(\mathbf{f}; \mathcal{C}^{(k,l)})$. Consequently, the change in the complete data ML objective function at step (k, l) is

$$\begin{aligned} \Delta E_{\text{complete ML}}^{(k,l)} \quad &\geq \quad \sum_{i \in S_l} \sum_{j=1}^{N} [\mathcal{C}_{ij}^{(k,l-1)} \log \frac{\mathcal{C}_{ij}^{(k,l-1)}}{\mathcal{C}_{ij}^{(k,l)}} - \mathcal{C}_{ij}^{(k,l-1)} + \mathcal{C}_{ij}^{(k.l)}] \\ & \quad\quad + E_{\text{s-ML}}(\mathbf{f}^{(k,l-1)}; \mathcal{C}^{(k,l)}) - E_{\text{s-ML}}(\mathbf{f}^{(k,l)}; \mathcal{C}^{(k,l)}) \\ &\geq \quad 0 \end{aligned} \tag{86}$$

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* $f_j^{(k,l)} = f_j^{(k,l-1)}$, $\forall j$.

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

$$\begin{aligned} \Delta E_{\text{cmp-ML}}^{(k)} \quad &\stackrel{\text{def}}{=} \quad E_{\text{cmp-ML}}(\mathcal{C}^{(k-1,L)}, \mathbf{f}^{(k-1,L)}) - E_{\text{cmp-ML}}(\mathcal{C}^{(k,L)}, \mathbf{f}^{(k,L)}),\ k > 0 \\ &= \quad \sum_{l=1}^{L} \Delta E_{\text{cmp-ML}}^{(k,l)} \geq 0 \end{aligned} \tag{87}$$

$\Delta E_{\text{cmp-ML}}^{(k)}$ 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. 6.5, define k^* as the iteration at which $\Delta E_{\text{cmp-ML}}^{(k^*)} = 0$. Then $\mathcal{C}^{(k^*,1)} = \mathcal{C}^{(k^*,2)} = \cdots = \mathcal{C}^{(k^*,L)} \stackrel{\text{def}}{=} \mathcal{C}^*$. We similarly get $\mathbf{f}^{(k^*,1)} = \mathbf{f}^{(k^*,2)} = \cdots = \mathbf{f}^{(k^*,L)} \stackrel{\text{def}}{=} \mathbf{f}^*$. 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 update (34), we once again get:

$$\mathcal{C}_{ij}^{*}=\mathcal{C}_{ij}^{(k^*,l)}=g_i\frac{H_{ij}f_j^{(k^*,l-1)}}{\sum_n H_{in}f_n^{(k^*,l-1)}}\ \forall i,\ \forall j,\ \forall l\in\{1,\cdots,L\}=g_i\frac{H_{ij}f_j^{*}}{\sum_n H_{in}f_n^{*}} \tag{88}$$

Therefore, from (19), we get

$$\mathcal{C}^{*}=\mathcal{C}^{\mathrm{sol}}(\mathbf{f}^{*}) \tag{89}$$

From the definition of $\mathcal{C}^{\mathrm{sol}}(\mathbf{f})$ in Sec. 3, we may see that (74) is equivalent to the two conditions $\nabla_{\mathcal{C}}E_{\mathrm{cmp-ML}}(\mathcal{C},\mathbf{f},\nu)|_{(\mathcal{C}^*,\mathbf{f}^*)}=\mathbf{0}$ and $\nabla_{\nu}E_{\mathrm{cmp-ML}}(\mathcal{C},\mathbf{f},\nu)|_{(\mathcal{C}^*,\mathbf{f}^*)}=\mathbf{0}$.

Since the $\mathbf{f}$ updates are chosen to zero the expression for $\frac{\partial E_{\mathrm{cmp-ML}}(\mathcal{C},\mathbf{f},\nu)}{\partial f_j}$ in (26), we note that $\nabla_{\mathbf{f}}E_{\mathrm{cmp-ML}}(\mathcal{C},\mathbf{f},\nu)|_{(\mathcal{C}^*,\mathbf{f}^*)}=\mathbf{0}$.

Thus, we have shown that the full gradient of $E_{\mathrm{cmp-ML}}(\mathcal{C},\mathbf{f},\nu)$ at $(\mathcal{C}^*,\mathbf{f}^*)$ is zero, i.e. $\nabla E_{\mathrm{cmp-ML}}(\mathcal{C},\mathbf{f},\nu)|_{(\mathcal{C}^*,\mathbf{f}^*,\nu)}=\mathbf{0}$.

7.5 Convergence to a fixed point of $E_{\mathrm{inc-ML}}$

Define Γ as the set (possibly singleton) of global minima of $E_{\mathrm{cmp-ML}}(\mathcal{C},\mathbf{f})$. We may argue that $(\mathcal{C}^*,\mathbf{f}^*)\in\Gamma$ as follows: $\Delta E_{\mathrm{cmp-ML}}^{(k,l)}=0$ is equivalent to the fact that $\mathcal{C}_{ij}^{(k,l)}$ and $f_j^{(k,l)}$ are no longer changing. We note that an algorithm can lead to a repeating $(\mathcal{C}_{ij}^{(k,l)},f_j^{(k,l)})$ that is *not* an element of Γ. But this cannot happen because the COSEM-ML algorithm is a grouped coordinate descent algorithm, and because $E_{\mathrm{cmp-MAP}}$ is convex. The COSEM-ML algorithm would continue to descend along coordinate directions whenever possible. Hence, if one approaches a condition where $(\mathcal{C}_{ij}^{(k,l)},f_j^{(k,l)})$ 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 $E_{\mathrm{inc-ML}}(\mathbf{f})$. 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 $\mathcal{C}^*=\mathcal{C}^{\mathrm{sol}}(\mathbf{f}^*)$, we have by the touching property in Sec. 6.2 that $E_{\mathrm{cmp-ML}}(\mathcal{C}^{\mathrm{sol}}(\mathbf{f}^*),\mathbf{f}^*)=E_{\mathrm{inc-ML}}(\mathbf{f}^*)$. This implies

$$E_{\mathrm{cmp-ML}}(\mathcal{C}^{\mathrm{sol}}(\mathbf{f}^*),\mathbf{f}^*)=E_{\mathrm{inc-ML}}(\mathbf{f}^*)>E_{\mathrm{inc-ML}}(\check{\mathbf{f}})=E_{\mathrm{cmp-ML}}(\mathcal{C}^{\mathrm{sol}}(\check{\mathbf{f}}),\check{\mathbf{f}}), \tag{90}$$

but this is a contradiction because $(\mathcal{C}^{\mathrm{sol}}(\mathbf{f}^*),\mathbf{f}^*)\in\Gamma$ is a global minimum of $E_{\mathrm{cmp-ML}}$. Thus, minimizing the complete data objective (7) (in our case, by the COSEM-ML algorithm) yields a solution $(\mathcal{C}^*,\mathbf{f}^*)$, where $\mathbf{f}^*=\hat{\mathbf{f}}$, our sought-after solution in (3).

8 Conclusions

We have derived new convergent complete data ordered subsets algorithms for ML reconstruction (COSEM-ML) and MAP reconstruction (COSEM-MAP) in emission tomography. We have achieved our original goal, namely that COSEM-ML and COSEM-MAP converge to the fixed points of $E_{\mathrm{inc-ML}}$ and $E_{\mathrm{inc-MAP}}$ respectively. However, there is no proof that the COSEM-ML and COSEM-MAP algorithms optimize $E_{\mathrm{inc-ML}}(\mathbf{f})$ and $E_{\mathrm{inc-MAP}}(\mathbf{f})$ monotonically. In many simulations, we have always observed monotonic convergence, however.

It is straightforward to include randoms or scatter via an affine term $\bar{\mathbf{s}}$ (corresponding to $\bar{\mathbf{g}}=H\mathbf{f}+\bar{\mathbf{s}}$) 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.

We also note that the COSEM-ML and COSEM-MAP algorithms are in a form suitable for reconstruction from list-mode data. List-mode versions of the COSEM-ML and COSEM-MAP algorithms may be derived either by applying transformations as in [25] or by adopting the approach in [26].

For ML and MAP, preliminary results indicate that COSEM-ML and COSEM-MAP are much faster than ML-EM and the MAP-EM algorithm of [20]. However, COSEM-ML is somewhat slower than RAMLA [6], while COSEM-MAP is somewhat slower than BSREM [7]. Unlike RAMLA and BSREM, our COSEM algorithms do not need a user-specified relaxation schedule. We are exploring [19] whether COSEM speed enhancements are possible without compromising its convergence properties.

References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

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

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

[8] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood estimation from incomplete data via the EM algorithm," *J. Royal Statist. Soc. B*, vol. 39, pp. 1–38, 1977.

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

[10] 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.

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

[12] 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.

[13] M. Lee, *Bayesian Reconstruction In Emission Tomography Using Gibbs Priors*, Ph.D. thesis, Yale University, New Haven, CT, 1994.

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

[15] 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.

[16] A. J. R. Gunawardana, *The Information Geometry of EM Variants for Speech and Image Processing*, Ph.D. thesis, Johns Hopkins University, Baltimore, MD, 2001.

[17] 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. 1998, pp. 355–368, Kluwer.

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

[19] 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.

[20] 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.

[21] 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.

[22] A. L. Yuille and A. Rangarajan, "The Concave-Convex procedure," *Neural Computation*, vol. 15, no. 4, pp. 915–936, 2003.

[23] 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).

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

[25] A. J. Reader, K. Erlandsson, A. Flower, and R. J. Ott, "Fast accurate iterative reconstruction for low-statistics positron volume imaging," *Physics in Medicine and Biology*, vol. 43, no. 4, pp. 823–834, April 1998.

[26] L. Parra and H. Barrett, "List-mode likelihood: EM algorithm and image quality estimation demonstrated on 2-D PET," *IEEE Trans. Med. Imag.*, vol. 17, no. 2, pp. 228–235, April 1998.