

An overview of fast convergent ordered-subsets reconstruction methods for emission tomography based on the incremental EM algorithm

Ing-Tsung Hsiao^a * Parmeshwar Khurd^b Anand Rangarajan^c and Gene Gindi^b

^aDept. of Medical Imaging and Radiological Sciences, Chang Gung University, 333 Tao-yuan, Taiwan

^bDepts. of Radiology and Electrical Engineering, SUNY at Stony Brook, Stony Brook, NY, USA.

^cDept. of CISE, University of Florida, Gainesville, FL, USA.

Statistical reconstruction has become popular in emission computed tomography but suffers slow convergence (to the MAP or ML solution). Methods proposed to address this problem include the fast but non-convergent OSEM and the convergent RAMLA [1] for the ML case, and the convergent BSREM [2], relaxed OS-SPS and modified BSREM [3] for the MAP case. The convergent algorithms required a user-determined relaxation schedule. We proposed fast convergent OS reconstruction algorithms for both ML and MAP cases, called COSEM (Complete-data OSEM), which avoid the use of a relaxation schedule while maintaining convergence. COSEM is a form of incremental EM algorithm. Here, we provide a derivation of our COSEM algorithms and demonstrate COSEM using simulations. At early iterations, COSEM-ML is typically slower than RAMLA, and COSEM-MAP is typically slower than optimized BSREM while remaining much faster than conventional MAP-EM. We discuss how COSEM may be modified to overcome these limitations.

1. Introduction

In emission tomography, the reconstruction of vast quantities of noisy, low-count data in practical times remains a challenge. The convenient FBP (Filtered backprojection) reconstruction algorithm is widely used. However, FBP and its analytical relatives do not always lend themselves to specialized acquisition geometries. In addition, it has been shown that statistical reconstructions have outperformed FBP [4].

Statistical reconstruction is useful in emission computed tomography due to its ability to accurately model noise, imaging physics, and to incorporate prior knowledge about the object. Statistical reconstruction approaches use iterative algorithms that optimize objective functions based on maximum likelihood (ML) or maximum *a posteriori* (MAP, a.k.a. penalized likelihood) principles. An additional advantage of MAP and ML solutions is that their statistical properties such as covariance, mean and local

point spread function can be predicted using theoretical expressions [5]. A drawback of statistical reconstruction algorithms is that they are slow. It is a challenge to devise convergent ML and MAP algorithms that require few iterations to approach the fixed-point solution and to reduce the number of the algorithm-specific parameters that must be hand-tuned. Earlier attempts to address the speed problem included preconditioned conjugate-gradient [7] and coordinate descent methods [8], but each of these had practical drawbacks.

In [9], an ordered subsets expectation maximization (OSEM) algorithm achieved an order of magnitude speedup over conventional (non-OS) EM by using only a subset of the projection data per sub-iteration. The OSEM algorithm is fast, parallelizable and preserves positivity. However, these early OS methods did not converge to the ML fixed-point solution. An algorithm [1], termed row-action maximum likelihood algorithm (RAMLA), used a relaxation schedule, i.e. a relaxation parameter at each iteration of an OSEM-

*Contact information: I.T. Hsiao ihhsiao@mail.cgu.edu.tw

like update, to attain convergence. The relaxation schedule had to satisfy certain properties as a prerequisite for convergence. In practice, this relaxation schedule must be determined by trial and error to ensure good speed. More recently, RAMLA was extended to the MAP case in BSREM (Block Sequential Modified EM) [2], modified BSREM [3], and relaxed OS-SPS (OS Separable Paraboloidal Surrogates) [3]. These MAP approaches also required a user-determined relaxation schedule. A practical solution to the problem of determining a relaxation schedule remains open.

To overcome the need of relaxation schedules in these OS algorithms, we formulated relaxation-free COSEM (Complete-Data OSEM) algorithms [11,12]. COSEM, while derived in terms of an algebraic transformation [13], is a form of incremental EM algorithm [14]. We independently introduced COSEM-ML and COSEM-MAP in terms of algebraic transformations [11,12], but learned shortly after that a COSEM-ML algorithm for emission tomography had been independently derived [15] using incremental-EM approaches [14]. No experimental results were presented in [15]. We later extended COSEM-MAP to a list-mode version [16]. Despite the appeal of COSEM as a relaxation-free convergent algorithm, simulations show that it is not as fast as the aforementioned convergent relaxation-based approaches.

In Sections 2 and 3, we state the problem and derive the COSEM-ML and MAP algorithms. In Section 4, we compare COSEM-MAP speed with that of competitors, and in Section 5, discuss how COSEM can be sped up to close the speed gap with its competitors while maintaining its advantages.

2. Statement of the Problem

Define the object to be an N -dim lexicographically ordered vector \mathbf{f} with elements $f_j, j = 1, \dots, N$. We model image formation by a simple Poisson model $\mathbf{g} \sim \text{Poisson}(\mathcal{H}\mathbf{f})$, where \mathcal{H} is the system matrix whose element \mathcal{H}_{ij} indicates the probability of receiving a count in detector bin i from pixel j , and \mathbf{g} is the integer-valued random data vector (sinogram) with ele-

ments $g_i, i = 1, \dots, M$.

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

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{f} \geq \mathbf{0}} E_{\text{inc-ML}}(\mathbf{f}) \quad (1)$$

where the ML objective corresponding to the usual (incomplete data) Poisson likelihood is given by

$$E_{\text{inc-ML}}(\mathbf{f}) = \sum_{ij} \mathcal{H}_{ij} f_j - \sum_i g_i \log \sum_j \mathcal{H}_{ij} f_j. \quad (2)$$

The MAP problem may be similarly stated as the minimization of the ML objective plus a prior objective E_{prior} . We consider quadratic priors of the form

$$E_{\text{prior}}(\mathbf{f}) = \beta \sum_j \sum_{j' \in \mathcal{N}(j)} w_{jj'} (f_j - f_{j'})^2 \quad (3)$$

with smoothing parameter $\beta > 0$, and a neighborhood system $\mathcal{N}(j)$ with weights $w_{jj'}$. It will be convenient to define $E_{\text{inc-MAP}} \equiv E_{\text{inc-ML}} + E_{\text{prior}}$.

3. COSEM Objective and Algorithm

The direct minimization of $E_{\text{inc-ML}}$ is difficult. One way of carrying out the minimization is by the well-known ML-EM algorithm. The conventional EM derivation [6] 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 derived from an algebraic transformation of $E_{\text{inc-ML}}(\mathbf{f})$ in Eq.(2), from the use of Jensen's inequality [17]. We briefly summarize this step. From Jensen's inequality and the convexity of $-\log(\cdot)$, we have that $-\log \sum_j \mathcal{H}_{ij} f_j \leq -\sum_j \frac{C_{ij}}{g_i} \log \frac{g_i \mathcal{H}_{ij} f_j}{C_{ij}}$ where the new "complete data" $C_{ij} \geq 0, \forall ij$ and $\sum_j C_{ij} = g_i, \forall i$ and with equality occurring at $C_{ij} = \frac{g_i \mathcal{H}_{ij} f_j}{\sum_{j'} \mathcal{H}_{ij'} f_{j'}}$. This allows us to transform the original incomplete-data objective function in Eq.(2) to a new objective, $E_{\text{comp}}(\mathbf{f}, \mathbf{C})$ given by:

$$E_{\text{comp}}(\mathbf{f}, \mathbf{C}) = \sum_l \sum_{i \in S_l} \sum_j C_{ij} \log \frac{C_{ij}}{\mathcal{H}_{ij} f_j} + \sum_l \sum_{i \in S_l} \sum_j \mathcal{H}_{ij} f_j + \sum_l \sum_{i \in S_l} \gamma_i (\sum_j C_{ij} - g_i). \quad (4)$$

Note that the sum over detector bins \sum_i in $E_{\text{inc-ML}}$ in Eq.(2) is replaced in Eq.(4) by a sum over subsets $\sum_l \sum_{i \in S_l}$ where $l = 1, \dots, L$ indexes subset number and S_l is the set of detector bins belonging to the l th subset. Here, C_{ij} is the complete data, roughly analogous to the complete data as used in statistical derivations of EM-ML. From its Jensen's inequality origin, we know that C_{ij} is real and positive, and that it obeys the constraints $\sum_j C_{ij} = g_i$ expressed in terms of a Lagrange parameter γ_i . Differentiating Eq.(4) w.r.t. \mathbf{C} , setting the result to zero and solving for the Lagrange parameter vector $\boldsymbol{\gamma}$ by imposing the constraints $\sum_j C_{ij} = g_i$, we get

$$C_{ij}^{\text{sol}}(\mathbf{f}) = g_i \frac{\mathcal{H}_{ij} f_j}{\sum_n \mathcal{H}_{in} f_n} \quad \forall i, \forall j. \quad (5)$$

Inserting the just established solution (5) into (4), we can write the identity

$$E_{\text{comp}}(\mathbf{C}^{\text{sol}}(\mathbf{f}), \mathbf{f}) = \sum_{ij} \mathcal{H}_{ij} f_j - \sum_i g_i \log \sum_j \mathcal{H}_{ij} f_j$$

+terms independent of $\mathbf{f} = E_{\text{inc-ML}}(\mathbf{f}) \quad (6)$

where $E_{\text{inc-ML}}(\mathbf{f})$ is the desired objective function in Eq.(2). We have shown that optimizing only over \mathbf{C} (while keeping \mathbf{f} fixed) in (4) results in Eq.(2). By optimizing $E_{\text{comp}}(\mathbf{f}, \mathbf{C})$ to obtain solutions $\hat{\mathbf{f}}, \hat{\mathbf{C}}$, we *also* obtain the solution $\hat{\mathbf{f}}$ of $E_{\text{inc-ML}}$ [17].

One may maximize $E_{\text{comp}}(\mathbf{f}, \mathbf{C})$ by a form of grouped coordinate ascent on subsets of \mathbf{C} and \mathbf{f} , which leads to a convergent and fast OS algorithm. Above, we showed that by first minimizing Eq.(4) w.r.t. \mathbf{C} , we obtained Eq.(2). We now show that minimizing Eq.(4) w.r.t. *subsets* of \mathbf{C} results in a convergent ordered subsets EM algorithm (COSEM). First, differentiate Eq.(4) w.r.t. a subset of \mathbf{C} and set the result to zero. Solve for the subset S_l of \mathbf{C} in terms of \mathbf{f} and the Lagrange multiplier γ_i where $i \in S_l$. Then, as before, we can eliminate γ_i *but only for* $i \in S_l$ by enforcing the constraints $\sum_j C_{ij} = g_i$, and obtain the update equation for the S_l subset of the complete data \mathbf{C} . We then solve for \mathbf{f} by setting the derivative of Eq.(4) w.r.t. \mathbf{f} to zero. We then repeat the above procedure for the next subset and so on. One can obtain the final update for all subsets of \mathbf{C} and \mathbf{f} as shown below. The update, in

fact, closely resembles that of OSEM itself, but the slight modifications introduced by COSEM-ML ensure convergence to the ML solution. It is worth examining the updates of both algorithms, and these are summarized below:

$$C_{ij}^{(k,l)} = g_i \frac{\mathcal{H}_{ij} f_j^{(k,l-1)}}{\sum_{j'} \mathcal{H}_{ij'} f_{j'}^{(k,l-1)}} \quad \forall i \in S_l \quad \forall j \quad (7)$$

$$C_{ij}^{(k,l)} = C_{ij}^{(k,l-1)} \quad \forall i \notin S_l \quad \forall j \quad (8)$$

$$\tilde{f}_j^{(k,l)} = \frac{\sum_i C_{ij}^{(k,l)}}{\sum_i \mathcal{H}_{ij}} \quad \forall j \quad (9)$$

$$\check{f}_j^{(k,l)} = \frac{\sum_{i \in S_l} C_{ij}^{(k,l)}}{\sum_{i \in S_l} \mathcal{H}_{ij}} \quad \forall j \quad (10)$$

The updates Eq.(7) and Eq.(10) when combined, form the familiar OSEM update, while the updates Eq.(7), Eq.(8) and Eq.(9) constitute COSEM-ML [11]. The notation $f_j^{(k,l)}$ indicates the update of the j th voxel at the l th subiteration (corresponding to use of the l th subset) of iteration k . When all L subsets have been updated, k is incremented. A first look at details of these updates appears to indicate that COSEM-ML involves many more steps per iteration k than does OSEM, but in [17], we show that the amount of computation is about the same. The updates for COSEM-ML preserve positivity.

We are primarily interested in the MAP case, and note that a COSEM-MAP version is obtainable by adding a prior $E_{\text{prior}}(\mathbf{f})$ to Eq.(4). The updates for \mathbf{C} in the COSEM-MAP algorithm remain the same as in Eq.(7) and Eq.(8). When we attempt to derive an update analogous to Eq.(9), the prior introduces a coupling between voxels, and this causes difficulty in deriving a closed-form and parallel update for \mathbf{f} . To solve this, we use the method of separable surrogates [21] replacing Eq.(3) by

$$E_{\text{priorSS}}(\mathbf{f}) = \frac{\beta}{2} \sum_j \sum_{j' \in \mathcal{N}(j)} w_{jj'} \left[(2f_j - f_j^{(k)} - f_{j'}^{(k)})^2 + (2f_{j'} - f_j^{(k)} - f_{j'}^{(k)})^2 \right] \quad (11)$$

at iteration k . Since the resulting MAP objective function including Eq.(4) and Eq.(11) 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 the separable surrogate enhanced complete data MAP objective, Eqs.(4) plus (11), w.r.t. f_j to zero, and defining $v_{jj'} \equiv w_{jj'} + w_{j'j}$, and $D_j \equiv \sum_i H_{ij}$, we obtain the final COSEM-MAP update for \mathbf{f} [12]:

$$f_j^{(k,l)} = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \quad (12)$$

where $a = 4\beta \sum_{j'} v_{jj'}$, $b = -2\beta \sum_{j'} v_{jj'} (f_j^{(k,l-1)} + f_{j'}^{(k,l-1)}) + D_j$, and $c = -\sum_i C_{ij}^{(k,l)}$.

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.

4. Simulation Results

To assess COSEM-MAP speed we display the normalized objective difference (NOD), versus iteration number k defined as:

$$\text{NOD}(k) = \frac{E_{\text{inc-MAP}}(\mathbf{f}^k) - E_{\text{inc-MAP}}(\mathbf{f}^*)}{E_{\text{inc-MAP}}(\mathbf{f}^0) - E_{\text{inc-MAP}}(\mathbf{f}^*)} \quad (13)$$

where $\mathbf{f}^k \equiv \mathbf{f}^{(k,0)}$ and \mathbf{f}^0 is the initial estimate obtained via FBP. We define \mathbf{f}^* to be the true MAP solution at $k = \infty$, estimated from 5000 iterations of EM-MAP.

A sinogram (64 angles by 96 bins) with 300K counts was simulated using a 2D 64x64 phantom (pixel = 5.6 mm), shown in Fig. 1(a). The phantom consists of a background and hot and cold lesions with a contrast ratio of 1:4:8 (cold:background:hot). Uniform attenuation (H_2O at 140 KeV) and depth-dependent blur for a parallel collimator were modeled.

Reconstructions were performed using EM-MAP, BSREM [2] and COSEM-MAP. (Note that EM-MAP is simply COSEM-MAP at $L = 1$ subset). We chose $L = 8$ subsets for BSREM and COSEM-MAP, and for all reconstructions set $\beta = 0.06$. The relaxation schedule for BSREM was chosen as: $\alpha^k = \frac{\alpha_0}{\max_{\{j,l\}} \sum_{i \in S_l} \mathcal{H}_{ij+k}}$ with $\alpha_0 = 3.2$. This schedule obeys the constraints

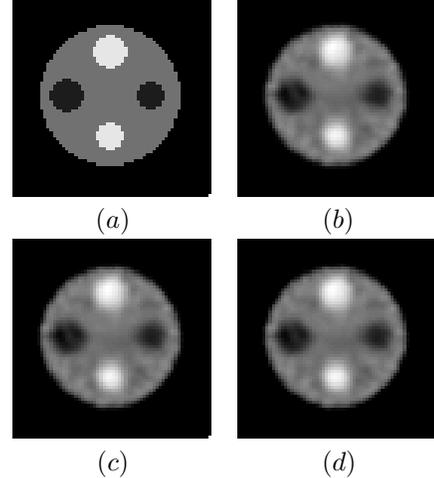


Figure 1. The 2D 64x64 phantom is shown in (a), while the anecdotal reconstructions are displayed in (b) EM-MAP, (c) COSEM-MAP, and (d) BSREM.

in [2]. The schedule was chosen to attain rapid convergence at early iterations. Anecdotal reconstructions for each algorithm are shown in Fig. 1 at iteration 30.

We plot NOD vs. k for each reconstruction in Fig. 2. 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 $\text{NOD}(k)$ (not shown here) for COSEM-MAP 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 competing OS-type algorithms.

5. Discussion

We have derived new convergent complete data ordered subsets algorithms for ML (COSEM-ML) and MAP (COSEM-MAP) reconstruction. It is straightforward to include randoms or scatter via an affine term $\bar{\mathbf{s}}$, corresponding to $\mathbf{g} \sim \text{Poisson}(\mathcal{H}\mathbf{f} + \bar{\mathbf{s}})$, in the algorithms. Detailed convergence proofs are presented in [17].

It is unknown whether COSEM converges monotonically. (By monotonic, we mean that the objective decreases at each outer loop iteration k .) In our many simulations with clinically realistic imaging parameters, we have not yet observed

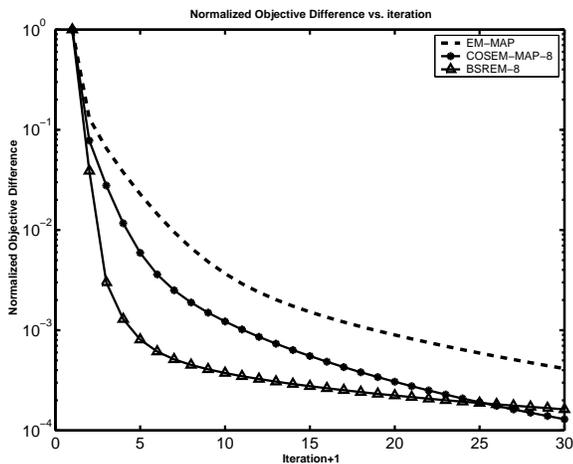


Figure 2. Normalized objective difference for EM-MAP, COSEM-MAP, and BSREM at $L=8$.

non-monotonicity.

Though the simulations in Sec. 4 are not extensive, the results are quite typical, and the speed gap relative to relaxation-based methods cannot be closed by simply choosing a better L or initial condition. One might consider speeding up COSEM while maintaining its desirable properties. In separate publications [22,18,23], we have demonstrated that this is possible with an “enhanced” COSEM (ECOSEM) algorithm for MAP and ML versions.

In ECOSEM, the basic strategy is to apply an automatically computed parameter that controls a trade-off between a COSEM update and a faster OSEM update at each sub-iteration, while maintaining convergence. Simulations results for ECOSEM-ML [18] show its speed to approach that of optimized RAMLA.

In sum, we have presented COSEM-ML and COSEM-MAP for emission tomography based on our notion of a complete data energy. The COSEM algorithms needs no user-specified relaxation schedule as do competitors. While the early-iteration speed of COSEM is slower than that of competitors, we are developing faster enhanced ECOSEM versions competitive in speed with RAMLA and BSREM.

6. Acknowledgment

This work was supported in part by NSC93-2320-

B-182-028 NSC Taiwan, CMRPD32013 and CM-RPD34005 from Chang Gung Research Fund, Taiwan, and by NIH-NIBIB RO1-EB02629.

REFERENCES

1. J. Browne, A. De Pierro, *IEEE Trans. Med. Imaging* 15 (5) (1996) 687.
2. A. De Pierro, M. Yamagishi, *IEEE Trans. Med. Imaging* 20 (4) (2001) 280.
3. S. Ahn, J. A. Fessler, *IEEE Trans. Med. Imaging* 22 (5) (2003) 613.
4. J. Qi, R. Huesman, *IEEE Trans. Med. Imaging* 20 (8) (2001) 815.
5. J. A. Fessler, *IEEE Trans. Image Processing* 5 (3) (1996) 493.
6. L. Shepp, Y. Vardi, *IEEE Trans. Med. Imaging MI-1* (1982) 113.
7. E. U. Mumcuoglu et al., *IEEE Trans. Med. Imaging* 13 (4) (1994) 687.
8. C. A. Bouman, K. Sauer, *IEEE Trans. Image Processing* 5 (3) (1996) 480.
9. H. M. Hudson, R. S. Larkin, *IEEE Trans. Med. Imaging* 13 (1994) 601.
10. E. Tanaka, H. Kudo, *Phys. Med. Biol.* 48 (10) (2003) 1405.
11. I. T. Hsiao et al., in: Sonka and Hanson (eds.), *Proc. SPIE*, Vol. 4684, (2002), 10.
12. I. T. Hsiao et al., *Proc IEEE Intl Symp on Biomedical Imaging*, Washington DC, 2002, 409.
13. E. Mjolsness, C. Garrett, *Neural Networks* 3 (1990) 651–669.
14. R. Neal, G. E. Hinton, in: M. I. Jordan (Ed.), *Learning in Graphical Models*, Dordrecht: Kluwer, 1998, pp. 355–368.
15. A. J. R. Gunawardana, Ph.D. Thesis, Johns Hopkins University, 2001.
16. P. Khurd et al., *IEEE Trans. Nuc. Sci.* 51 (2004) 719.
17. A. Rangarajan et al., *Tech. Rep. MIPL-03-1*, SUNY Stony Brook, Dept Radiology, Medical Image Processing Laboratory (Dec. 2003).
18. I.T. Hsiao et al., *Phys. Med. Biol.* 49 (11) (2004) 2145.
19. M. Lee, Ph.D. Thesis, Yale University, New Haven, CT (1994).
20. A. Rangarajan, et al, in: *Maximum Entropy and Bayesian Methods*, Kluwer Academic Publishers, Dordrecht, Netherlands, 1996, pp. 117.
21. A. De Pierro, *IEEE Trans. Med. Imaging* 14 (1) (1995) 132.
22. I.T. Hsiao et al., *Proc VII Intl Conf. on Fully 3D Recon. in Rad. and Nucl. Med.*, 2003.
23. I.T. Hsiao et al., *Proc 2004 IEEE Nuc. Sce. Sym. Med. Img. Conf.*, Rome, Italy, 2004.