A NEW CONVERGENT MAP RECONSTRUCTION ALGORITHM FOR EMISSION TOMOGRAPHY USING ORDERED SUBSETS AND SEPARABLE SURROGATES

Ing-Tsung Hsiao

School of Medical Technology
Chang-Gung University
Kwei-Shan
333 Tao-yuan, Taiwan

Anand Rangarajan† *and Gene Gindi*‡

†Dept. of CISE
Univ.of Florida
‡Depts. of Radiology and Electrical Engineering
SUNY Stony Brook

ABSTRACT

We investigate a new, fast and provably convergent MAP reconstruction algorithm for emission tomography. The new algorithm, termed C-OSEM has its origin in the alternating algorithm derivation of the well known EM algorithm for emission tomography. In this re-derivation, the complete data explicitly enters the objective function as an unknown variable. While the entire complete data gets updated in each iteration of EM, in C-OSEM the complete data is updated only along ordered subsets. C-OSEM has a straightforward extension to the MAP case especially when using convex, smoothing priors. Unlike RAMLA and BSREM, C-OSEM does not require relaxation parameters to be set at each iteration. We derive the MAP C-OSEM algorithm using the separable surrogate method and anecdotally compare performance with MAP EM and BSREM.

1. INTRODUCTION

Statistical reconstruction has become increasingly popular in emission tomography due to its ability to accurately model noise and the imaging physics and impose positivity constraints on the reconstruction. 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 Poisson. 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 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. If the log-priors are convex and well behaved, (and since the Poisson likelihood is also convex), there exist many statistical MAP reconstruction algorithms that can determine the global optimum of the log-posterior MAP cost function. The main drawback of statistical reconstruction algorithms is that they are slow when used for clinical studies especially in comparison to conventional reconstruction algorithms such as filtered backprojection. This is true regardless of whether likelihood- or MAP-based approaches are being considered.

The slow convergence of statistical reconstruction algorithms has recently received much attention. Recently, an ordered subsets EM (OS-EM) algorithm *et al.*[1], which uses only a subset of the projection data per backprojection, has become quite popular in emission tomography. This lead to an order of magnitude speedup over conventional EM. However, OS-EM is heuristically motivated and can and sometimes does converge to limit cycles since there is no proof of convergence. An alternative algorithm, termed row-action maximum likelihood algorithm (RAMLA) was proposed in [2]. 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. Recently in [3], 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.

We have recently introduced [4] a new convergent OS-EM like reconstruction algorithm for emission tomography. The algorithm, termed C-OSEM has more in common with OS-EM than with RAMLA because there are no relaxation parameters involved. To derive this algorithm, we began by recasting the EM-ML algorithm as an alternating minimization algorithm [5]. The alternating minimization is w.r.t. the complete data and the reconstruction. We then carried over the basic idea of using ordered subsets to C-OSEM by updating the reconstruction after each subset. This leads to a faster convergence to the ML solution than EM-ML. Since only the nature of the alternation has been altered and since each step lowers the ML cost function, C-OSEM is guaranteed to converge.

In this work, we extend C-OSEM to the MAP case. The extension is fairly straightforward. We use convex smoothing priors which help guarantee that the log-posterior is unimodal. Since the smoothing prior creates correlations between the reconstruction voxels, we have used the method of separable surrogates to decouple the reconstruction variables in the prior. Once this is accomplished, the extension of C-OSEM to the case of separable surrogate priors is straightforward. We derive an alternating "OS-EM like" algorithm which we continue to call C-OSEM to emphasize the importance of the complete data. No relaxation parameters need be specified for the MAP case as well: C-OSEM is essentially a parameter free, monotonic, and fast algorithm.

2. THEORY

We denote the object (the mean emission rate) by a vector $\mathbf{f}$ with elements $\{f_n; n = 1, \ldots, N\}$, and the projection data by $\mathbf{g} = \{g_m; m = 1, \ldots, M\}$. An estimate of $\mathbf{f}$ is expressed as $\hat{\mathbf{f}}$. The projection data $\mathbf{g}$ obey a Poisson distribution with mean $\bar{\mathbf{g}} = \mathcal{H}\mathbf{f}$, where $\mathcal{H}$ is the system matrix with element H_{mn} indicating the probability that detector bin m receives a count emitted from voxel n ($m = 1, \ldots, M; n = 1, \ldots, N$). The Poisson log-likelihood is then $\Phi(\mathbf{g}|\mathbf{f}) = \sum_m g_m \log \bar{g}_m - \bar{g}_m$ where $\bar{g}_m = \sum_n \mathcal{H}_{mn} f_n$. The log-likelihood can be optimized by any suitable algorithm. However, due to the coupling of the object pixels in the objective function, direct optimization of the likelihood is not straightforward. The EM algorithm [6] decouples the object pixels by introducing the complete data. This results in an update equation equation which guarantees positivity. However, EM-ML is slow to converge. The OSEM algorithm modifies the update schedule of EM-ML by first dividing the projection data $[1, \ldots, M]$ into L disjoint subsets ($S_l; l = 1, \ldots, L$), and updating the reconstruction subset by subset: for subset $l = 1, \ldots, L$

$$\hat{f}_n^{k,l+1} = \hat{f}_n^{k,l} \frac{\sum_{m \in S_l} \mathcal{H}_{mn} \frac{g_m}{\hat{g}_m^k}}{\sum_{m \in S_l} \mathcal{H}_{mn}} \quad (1)$$

for $n = 1, \ldots, N$, and where S_l indicates the lth subset projection data for $l = 1, \ldots, L$. Note that for each iteration k, there are L sub-iterations, and each update $\hat{f}_n^{k,l}$ at sub-iteration l uses only a subset of the projection, and serves as the initial estimate for next sub-iteration $\hat{f}_n^{k,l+1}$. Also $\hat{f}_n^{k,L} = \hat{f}_n^{k+1}$.

In [4] we used the alternating algorithm version of the ML-EM algorithm [5] to derive a new C-OSEM algorithm that was guaranteed to converge. First, we define the complete data as $\mathbf{C}$ with element $\{C_{mn}; m = 1, \ldots, M, n = 1, \ldots, N\}$ indicating the number of counts collected in bin m emitting from pixel n. Despite the fact that the complete data C_{mn} is an integer, the estimate of the complete data in the alternating algorithm is not, in general, an integer. This is also true in the EM algorithm where the estimate of the complete data is also used. The ML-EM estimation can then be re-derived as an alternating minimization [5] on the following objective function,

$$E(\mathbf{C}, \mathbf{f}) == \sum_{m=1}^{M} \sum_{n=1}^{N} (-C_{mn} \log H_{mn} f_n + H_{mn} f_n)$$
$$+ \sum_{m=1}^{M} \sum_{n=1}^{N} (C_{mn} \log C_{mn} - C_{mn}) + \sum_m \zeta_m \{\sum_n C_{mn} - g_m\} \quad (2)$$

One can obtain the update equation for C_{mn} for the first step of the alternation by minimizing (2) w.r.t. $\mathbf{C}$:

$$C_{mn} = g_m \frac{\mathcal{H}_{mn} f_n}{\sum_j \mathcal{H}_{mj} f_j} \quad (3)$$

This update (3) is exactly the E-step of the standard ML-EM algorithm[6]. Next we will optimize $E(\mathbf{C}, \mathbf{f})$ w.r.t. $\mathbf{f}$ with fixed $\mathbf{C}$,

$$\frac{\partial E(\mathbf{C}, \mathbf{f})}{\partial f_n} = 0 \rightarrow f_n = \frac{\sum_m C_{mn}}{\sum_m \mathcal{H}_{mn}} \quad (4)$$

Therefore, at iteration $k + 1$, the alternation becomes,

$$C_{mn}^{k+1} = g_m \frac{\mathcal{H}_{mn} \hat{f}_n^k}{\sum_j \mathcal{H}_{mj} \hat{f}_j^k} \quad , \forall m, n \quad (5)$$

$$\hat{f}_n^{k+1} = \frac{\sum_m C_{mn}^{k+1}}{\sum_m \mathcal{H}_{mn}} \quad , \forall n. \quad (6)$$

Note that (5) and (6) are the result of an alternating coordinate descent algorithm, and are identical to the ML-EM algorithm.

2.1. Separable Surrogates for Convex Smoothing Priors

In this work, we are mainly interested in extending our previous C-OSEM algorithm to the MAP case. The MAP cost function used in our work is

$$\hat{\mathbf{f}} = \arg\min_{\mathbf{f} > 0} (-\Phi(\mathbf{g}|\mathbf{f}) + \lambda R(\mathbf{f})) \quad (7)$$

where $\lambda > 0$ is a smoothing parameter. A general expression of the log-prior term is

$$R(\mathbf{f}) = \sum_n \psi([D\mathbf{f}]_n) = \sum_n \psi(\sum_{n'} D_{nn'} f_{n'}) \quad (8)$$

where D is a matrix of (usually) nearest neighbor connections, and we further assume that the function $\psi(\cdot)$ is convex. Such a convex prior is readily amenable to a separable surrogate function [7] treatment as demonstrated below. Essentially, it was shown that

$$\psi(\sum_{n'} D_{nn'} f_{n'}) \le \sum_{n'} \gamma_{nn'} \psi([\frac{D_{nn'}}{\gamma_{nn'}} (f_{n'} - f_{n'}^{\text{old}}) + \sum_l D_{nl} f_l^{\text{old}}]). \quad (9)$$

The basic idea behind using separable surrogates is that rather than minimize $R(f) = \sum_n \psi([D\mathbf{f}]_n)$, we minimize

$$\sum_n R_n(f_n; \mathbf{f}^{\text{old}}) = \sum_n \sum_{n'} \gamma_{nn'} \psi([\frac{D_{nn'}}{\gamma_{nn'}}(f_{n'} - f_{n'}^{\text{old}}) + \sum_l D_{nl} f_l^{\text{old}}]). \quad (10)$$

Since the surrogate functions are always bounded from below [from (9)] by the original coupled log-prior, any descent step over the sum of the surrogates $\sum_n R_n(f_n; \mathbf{f}^{\text{old}})$ is guaranteed to be a descent step in the original prior cost function $\sum_n \psi([D\mathbf{f}]_n)$ provided the starting configuration $\mathbf{f}^{\text{old}}$ is the same for both cost functions.

We now specialize to convex, quadratic, smoothing priors. This prior is written as $R(\mathbf{f}) = \sum_n \sum_{n' \in \mathcal{N}(n)} w_{nn'}(f_n - f_{n'})^2$ Note that this form of prior corresponds to $D_{nn} = 1$, $D_{nn'} = -1$. Using (9) and specializing to the above choice of D, we can write the separable surrogate form of the prior as

$$R_{\text{s}}(\mathbf{f}) = \sum_n \frac{1}{2} \sum_{n' \in \mathcal{N}(n)} w_{nn'}([2f_n - f_n^{\text{old}} - f_{n'}^{\text{old}}]^2 + [-2f_{n'} + f_n^{\text{old}} + f_{n'}^{\text{old}}]^2) \quad (11)$$

An alternating update equation can be derived by taking derivatives w.r.t. f_n in a new posterior consisting of the complete data term $E(\mathbf{C}, \mathbf{f})$ from (2) and the surrogate prior $R_s(\mathbf{f})$ from (11). The update equation, which guarantees positivity and is free of any relaxation parameters, is obtained by differentiating the new posterior $E(\mathbf{C}, \mathbf{f}) + R_s(\mathbf{f})$ w.r.t. $\mathbf{f}$ and solving for f_n.

$$\frac{\partial E(\mathbf{C}, \mathbf{f}) + \lambda R_s(\mathbf{f})}{\partial f_n} = -\frac{1}{f_n} \sum_m C_{mn} + \sum_m H_{mn} + 4\lambda f_n \sum_{n' \in \mathcal{N}(n)} w_{nn'} - 2\lambda \sum_{n' \in \mathcal{N}(n)} w_{nn'}(f_n^{\text{old}} + f_{n'}^{\text{old}}) = 0 \quad (12)$$

The final update is $f_n = \frac{-b + \sqrt{b^2 + 4ac}}{2a}$ where $a = 4\lambda \sum_{n'} w_{nn'}$, $b = -2\lambda \sum_{n'} w_{nn'}(f_n^{\text{old}} + f_{n'}^{\text{old}})$ and $c = -\sum_m C_{mn}$. Note that the update requires knowledge of an estimate of the complete data $\mathbf{C}$. If the complete data is globally updated after an $\mathbf{f}$ update, an EM algorithm results. However, if the complete data is updated using ordered subsets, the MAP C-OSEM algorithm results.

2.2. The MAP C-OSEM Algorithm

We now describe our MAP C-OSEM algorithm. We apply the idea of ordered subsets $S_l, l = 1, \ldots, L$ in the update of $\mathbf{C}$ above. The ordered subsets approach is applied to the subsets of the complete data $\mathbf{C}$. Now (2) is optimized with two loops, the outer loop indexed by k, and the inner loop by l. At iteration k and the inner loop sub-iteration l, (2) is first minimized w.r.t. C_{mn} for given subset $m \in S_l$ while keeping $\{C_{mn}; m \notin S_l\}$ and $\mathbf{f}$ fixed. Note that the prior does not contain $\mathbf{C}$, so this step is identical to the likelihood case. This optimization leads to the following update equation,

$$C_{mn}^{(k,l)} = g_m \frac{\mathcal{H}_{mn} f_n^{(k,l-1)}}{\sum_j \mathcal{H}_{mj} f_j^{(k,l-1)}} \forall n, \text{and } \{m \in S_l\}. \quad (13)$$

Next, we keep C_{mn} fixed (for all m, n), and optimize the posterior w.r.t. $\mathbf{f}$. However, instead of using only the lth subset ($\{m \in S_l\}$), all data are used to get the C-OSEM update equation for $\mathbf{f}$. But this is precisely the update equation derived in the previous section. The main difference from conventional EM is that the complete data is more frequently updated using ordered subsets as in (13). The alternation continues until all sub-iterations $l = 1, \ldots, L$ are completed, and then advances to the next outer iteration $k+1$. The resulting C-OSEM algorithm can be summarized as the following pseudo-code in Fig.1.

C-OSEM Algoritm

Initial conditions $\{\hat{\mathbf{f}}^{(0,0)}\}$: $k = 0$

$C_{mn}^{(0,l)} = g_m \frac{\mathcal{H}_{mn} \hat{f}_n^{(0,0)}}{\sum_j \mathcal{H}_{mj} \hat{f}_j^{(0,0)}} \quad \forall n$, and $m \in S_l; l = 1, .., L$

$B_n^{(0,0)} = \sum_{l=1}^{L} \sum_{m \in S_l} C_{mn}^{(0,l)} \quad \forall n$

Begin A: $k - loop$**:**

Begin B: $l - loop$: $l = 1, .., L$,

$C_{mn}^{(k,l)} = g_m \frac{\mathcal{H}_{mn} \hat{f}_n^{(k,l-1)}}{\sum_j \mathcal{H}_{mj} \hat{f}_j^{(k,l-1)}} \ \forall m \in S_l$

$B_n^{(k,l)} = B_n^{(k,l-1)} + \sum_{m \in S_l} [C_{mn}^{(k,l)} - C_{mn}^{(k-1,l)}]$

$\hat{f}_n^{k,l} = \frac{-b + \sqrt{b^2 + 4ac}}{2a}$ where $a = 4\lambda \sum_{n' \in \mathcal{N}(n)} w_{nn'}$, $b = -2\lambda \sum_{n' \in \mathcal{N}(n)} w_{nn'}(f_n^{(k,l-1)} + f_{n'}^{(k,l-1)})$, $c = -B_n^{(k,l)}$

End B

$\hat{\mathbf{f}}^{k+1,0} = \hat{\mathbf{f}}^{k,L}$

$B_n^{(k+1,0)} = B_n^{(k,L)} \quad \forall n$

$k = k + 1$.

End A

Fig. 1. The pseudocode for the C-OSEM Algorithm.

Please note that in the C-OSEM algorithm, only the complete data w.r.t. the subset $m \in S_l$ is updated at a time. Consequently, the summation $\sum_{l=1}^{L}$ above does not need to be explicitly carried out. Instead, we define a new array $B_n = \sum_{l=1}^{L} \sum_{m \in S_l} C_{mn}$ to keep track of the partial changes made to $\mathbf{C}$. In the inner l sub-iteration, after updating $C_{mn}^{(k,l)}$, we update $B_n^{(k,l)} = B_n^{(k,l-1)} + \sum_{m \in S_l} [C_{mn}^{(k,l)} - C_{mn}^{(k-1,l)}]$. Initial conditions are $B_n^{(0,0)} = \sum_{l=1}^{L} \sum_{m \in S_l} C_{mn}^{(0,l)}$. This results in an efficient scheme for keeping track of the changes to $\mathbf{C}$. The restriction of the complete data update to a subset of the complete data has no effect on the convergence proof. All the complete data continue to affect the $\mathbf{f}$ update (via $B_n^{(k,l)}$). Convergence can be shown by first showing that both sides of the alternation reduce the cost function. Since each step reduces the MAP cost function (which is bounded from below), convergence is guaranteed.

3. RESULTS

We now present anecdotal reconstructions using MAP-EM, BSREM and MAP C-OSEM to demonstrate the speed advantage of MAP C-OSEM and BSREM. To this end, we generated noisy (300K counts) sinograms using the 2D 64×64 phantom in Fig. 2. The phantom consists of a disk background, two hot lesions and two cold lesions with contrast ratio of 1:4:8 (cold:background:hot). Reconstructions are then performed using the above three algorithms. We used 4 subsets for BSREM and MAP C-OSEM with λ set to 0.8. The relaxation parameter in BSREM was chosen according to the following schedule: $\beta_k = \frac{1}{max_l \sum_{m \in S_l} H_{mn} + k + 1}$. The reconstructed images for each algorithm at different iterations are shown in Fig.2; 40 iterations for MAP EM, 20 iterations for MAP C-OSEM and 10 iterations for BSREM. The growth of the log-posterior and the decrease of the root mean squared error (RMSE) of the reconstructed images for each algorithm are shown in Fig. 3. As expected, the MAP C-OSEM and BSREM algorithms are much faster than the MAP EM algorithm with MAP C-OSEM being in between BSREM and MAP EM.

4. CONCLUSIONS

We have derived a new convergent C-OSEM algorithm for MAP reconstruction in emission tomography. Preliminary results indicate that the new algorithm is much faster than conventional EM while being slower than BSREM. Unlike BSREM, there are no relaxation parameters that need to be set at each iteration. It remains to be seen if further MAP C-OSEM speed enhancements are possible without compromising its convergence properties.

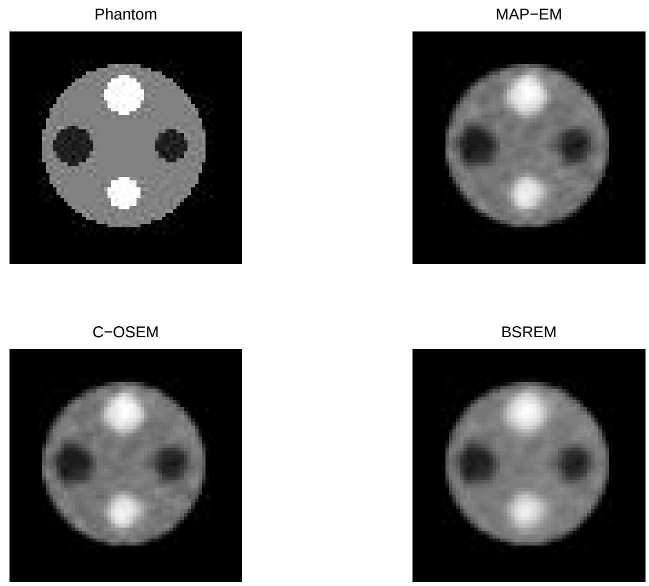

Fig. 2. **MAP Reconstructions:** Top left: Original phantom. Top right: MAP OS-EM reconstruction. Bottom left: MAP C-OSEM reconstruction. Bottom right: BSREM reconstruction. The reconstructed images are shown at different iterations: 40 iterations for MAP EM, 20 iterations for MAP C-OSEM and 10 iterations for BSREM.

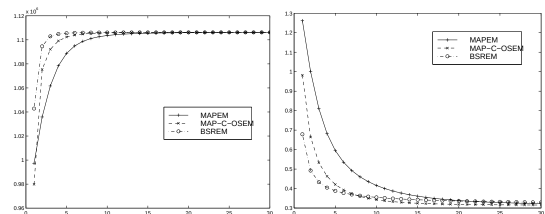

Fig. 3. **Log-posterior and RMSE plots:** Left: The log-posterior for MAP C-OSEM, BSREM and MAP OS-EM. Right: RMSE error for the same algorithms

5. REFERENCES

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

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

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

[4] I.-T. Hsiao, A. Rangarajan, and G. Gindi, "A provably convergent OS-EM like reconstruction algorithm for emission tomography," in *Proc. SPIE Medical Imaging Conference*, 2002.

[5] I. Csiszar and G. Tusnady, "Information geometry and alternating minimization procedures," *Statistics and Decisions*, pp. 205–237, 1984.

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

[7] 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.