Accelerated Cross-Reference Maximum Likelihood Estimates for Positron Emission Tomography

Henry Horng-Shing Lu[1*] and Wen-Jie Tseng[1]

November 1997

Abstract

The state of art of positron emission tomography (PET) takes into account the accidental coincidence events and attenuation. The maximum likelihood estimator can handle this kind of random variation in the reconstruction of a PET image. However, the reconstruction is ill-posed and needs regularization. The boundary information, either from an expert or from the other medical modality of the same object, like the X-ray CT scan, MRI, and so forth, can be used to regularize the reconstruction. We have investigated new, efficient and robust approaches to extract the related but incomplete boundary information. Fast algorithms adapted from the expectation/conditional maximization (ECM) and space alternating generalized expectation maximization (SAGE) algorithms are proposed to accelerate the computation. The method of generalized approximate cross validation (GACV) is adjusted to select the penalty parameter from observed data quickly. The Monte Carlo studies demonstrate the improvement.

Key Words: Maximum likelihood estimate, EM algorithm, regularization, expectation/conditional maximization (ECM), space alternating generalized expectation maximization (SAGE), generalized approximate cross validation (GACV).

[1]Address for the corresponding author: Dr. Henry Horng-Shing Lu, Institute of Statistics, College of Science, National Chiao Tung University, 1001 Ta Hsueh Road, Hsinchu 30050, Taiwan. Email: hslu@stat.nctu.edu.tw. Tel: 886-3-5731870. Fax: 886-3-5728745. Mr. Wen-Jie Tseng was a graduate student in the Institute of Statistics.

1 Introduction

The maximum likelihood estimate (MLE) with expectation maximization (EM) algorithm has been investigated for positron emission tomography (PET), without or with accidental coincidence (AC) events and attenuation (for instance, Shepp and Vardi [1], Lange and Carson [2], Politte and Snyder [3], Fessler, Clinthorne, and Rofers [4]). Due to the ill-posedness of reconstruction based on indirect random observations in detectors outside human bodies, the estimator without regularization has edge and noise artifacts as shown in Snyder, Miller, and Politte [5].

A variety of regularization methods have been studied in literature, like the early stopping rule in Veklerov and Llacer [6], the method of sieves in Snyder and Miller [7], Chang and Hsiung [8], the Bayesian approach with different kinds of priors in Hebert and Leahy [9], Green [10], Herman, Pierro and Gai [11], and so forth. In order to have better reconstruction, the regularization methods shall not based solely on mathematical reasons. Geman and Geman [12], Ouyang, Wong, Johnson, Hu, and Chen [13], have proposed to use the prior information about edges and boundaries to have a better reconstruction through Bayesian setups. However, there are many parameters in the prior information needed to be decided in advance and the computation of maximum a posteriori (MAP) is quite demanding. We are hence motivated to search for a more efficient approach.

Efficient methods to equip the MLE with the related but incomplete boundary information with only one penalty parameter are investigated. These are particular forms of penalized MLE that have the penalty terms derived from the boundary information. Speedy approaches that combine the boundary information at the same time of reconstruction are discussed. The solution is proved to be uniquely existent and is termed as the accelerated cross-reference maximum likelihood estimate (ACRMLE). New algorithms adapted from the modified expectation/conditional maximization (ECM) [14] and space alternating generalized expectation maximization (SAGE) algorithms [15, 16] are developed in this article. These algorithms are shown to be convergent and have fast convergence speeds. Simulation studies reveal that an even faster convergence speed is achieved if one alternates the order of iterations. The penalty parameter can be selected by users or data-driven methods. The automatic selection of the penalty parameter can be found out quickly by applying the method of generalized approximate cross validation (GACV) [17].

2 Model Setups and Backgrounds

Suppose the target image is partitioned into B pixels and there are D detector tubes. For each pixel, $b = 1,\ 2,\ \ldots,\ B$, and each tube, $d = 1,\ 2,\ \ldots,\ D$,

$$
\begin{aligned}
n_p^*(d) &\sim \text{Poisson}(\lambda^*(d) + \lambda_d^*(d)), && (1)\\
n_d^*(d) &\sim \text{Poisson}(\lambda_d^*(d)), && (2)\\
\lambda^*(d) &= \sum_{b=1}^{B} p(b,d)\lambda(b), && (3)\\
\lambda_p^*(d) &= \lambda^*(d) + \lambda_d^*(d), && (4)
\end{aligned}
$$

where $n_p^*(d)$ and $n_d^*(d)$ are the number of coincidences within the prompt and delayed windows respectively, "$\sim$" means "distributed as," $\lambda(b)$ is the emission intensity of the target image at pixel b, $p(b,d) = q(b,d) * s(b,d)$ is a transition probability, $\lambda_d^*(d)$ is the accidental coincidence rate, $n_p^*(d)$ and $n_d^*(d)$ are assumed to be statistically independent. The $q(b,d)$ is the probability of an annihilation in the bth pixel being detected by the dth detector pair, $s(b,d)$ is the survival probability that an emission photon pair emitted from box b traveling in tube d which has energy greater than the threshold level of detectors after attenuation. The $s(b,d)$ is determined by the line integral of attenuated function as equation (2) in Politte and Snyder [18]. Since $p(b,d)$ can be renormalized, it is assumed that $p(b,\cdot) = \sum_{d=1}^{D} p(b,d) = 1$ without loss of generality [1].

Based on the above model, the likelihood function becomes

$$L(\lambda, \lambda_d^*) = \prod_{d=1}^{D} e^{-\lambda_p^*(d)} \frac{\lambda_p^*(d)^{n_p^*(d)}}{n_p^*(d)!} \prod_{d=1}^{D} e^{-\lambda_d^*(d)} \frac{\lambda_d^*(d)^{n_d^*(d)}}{n_d^*(d)!}. \tag{5}$$

After deleting those terms that are functional independent of (λ, λ_d^*), the log-likelihood function, $l(\lambda, \lambda_d^*)$, becomes

$$\begin{aligned} l(\lambda, \lambda_d^*) &= \log(L(\lambda, \lambda_d^*)) \\ &= -\sum_{d=1}^{D}\sum_{b=1}^{B} p(b,d)\lambda(b) + \sum_{d=1}^{D} n_p^*(d)\log(\sum_{b=1}^{B} p(b,d)\lambda(b) + \lambda_d^*(d)) \\ &\quad -2\sum_{d=1}^{D} \lambda_d^*(d) + \sum_{d=1}^{D} n_d^*(d)\log\lambda_d^*(d). \end{aligned} \tag{6}$$

The iteration technique by the EM algorithm [19], was applied to find out the MLE in this situation [1, 2]. The EM algorithm is shown to be convergent and

$$l(\lambda^{new}, \lambda_d^{*new}) \geq l(\lambda^{old}, \lambda_d^{*old}). \tag{7}$$

However, the MLE-EM has noise and edge effects due to the ill-posed nature of inverse problems [5]. That is, the random Poisson noises are amplified after the inversion from indirect observations in the detector ring to the intensity in the target slice. This can be reduced by a proper regularization. Nevertheless, since the information about sharp edges in the target image is smeared out in the outside observations, the reconstructions of sharp edges remain extremely difficult. Therefore, one needs to borrow the strength from the other related information about edges that is not contained in the observations. This kind of anatomic boundary information may come from an expert or the other medical modality, such as MRI, X-ray CT scan, and so on [13]. However, the anatomic boundaries may be different from the functional boundaries. Caution must be made when one employs the anatomic boundaries. Assume the boundary information is available before the MLE reconstruction, then there is a fast approach that incorporate the boundary information, regularization, and MLE in one step. If the boundary information is available after an expert observes a primary reconstruction, then there are two- or three- step approaches to make use of the boundary information to improve the MLE reconstruction. They are explored in the next section. Because it is an efficient method to incorporate the related but incomplete boundary information, it is termed as an accelerated cross-reference MLE (ACRMLE).

3 The ACRMLE

3.1 A One-Step Approach

The prior boundary information is going to be used to regularize the MLE. Because the boundaries separate regions with different properties, the image intensities within every regions are expected to be quite similar. Therefore, the variances of intensities within every regions shall be kept as small as possible when the log-likelihood is maximized. A penalty term about the local intensity variation within every regions is combined with the log-likelihood function to incorporate the boundary information in the MLE. Since the boundary information may be incomplete or partly incorrect, a penalty parameter shall be put into to perform the fine tune. The one-step ACRMLE (ACRMLE1) becomes

$$(\hat{\lambda}_{ACRMLE1}, \hat{\lambda}^*_{d,ACRMLE1}) = \underset{\lambda \geq 0, \lambda^*_d \geq 0}{argmax} \quad \Phi(\lambda, \lambda^*_d), \tag{8}$$

where

$$\Phi(\lambda, \lambda^*_d) = l(\lambda, \lambda^*_d) - \alpha \left\{ \sum_{j=1}^{m} \sum_{b=1}^{n_j} [\lambda(jb) - \bar{\lambda}_j]^2 + \|\lambda^*_d - n^*_d\|^2 \right\}, \alpha > 0. \tag{9}$$

The parameter α is a positive penalty parameter, and $\|\cdot\|$ is a 2-norm. If $\alpha \to 0$, then $\left(\hat{\lambda}_{ACRMLE1}, \hat{\lambda}^*_{d,ACRMLE1}\right) \to \left(\hat{\lambda}_{MLE}, \hat{\lambda}^*_{d,MLE}\right)$. On the other hand, if $\alpha \to \infty$, then $\left(\hat{\lambda}_{ACRMLE1}(jb), \hat{\lambda}^*_{d,ACRMLE1}\right) \to \left(\bar{\lambda}_j, n^*_d\right)$. The penalty parameter, α, balance the trade-off between the log-likelihood and the penalty term. Meanwhile, the notation m is the number of regions, n_j is the number of pixels in the jth region, $\lambda(jb)$ is the emission intensity at the bth pixel in the jth region, and $\bar{\lambda}_j = \frac{1}{n_j}\sum_{k=1}^{n_j}\lambda(jk)$. The first penalty term, $\sum_{j=1}^{m}\sum_{b=1}^{n_j}[\lambda(jb) - \bar{\lambda}_j]^2$, describe the local intensity variation within every regions. The second penalty term, $||\lambda_d^* - n_d^*||^2$, is a mathematical device that ensure the estimate of λ_d^* is always unique. The proof of the unique existence of solution is shown in the next Theorem. Two possible numerical approaches to find out the unique solution are discussed in the followed two subsections.

Theorem 3.1 *Suppose the observations are nontrivial, that is, $n_p^*(d) \neq 0$, for some $d = 1, 2, \ldots, D$. Then the maximum point of $\Phi(\lambda, \lambda_d^*)$ in (9) exists uniquely.*

Proof: This is proved by checking that the Hessian matrix of $\Phi(\lambda, \lambda_d^*)$ in (9) is negative definite under the assumption.

3.2 The Modified ECM Algorithm

The solution of ACRMLE1 is difficult to obtain by setting the first derivatives of object function $\Phi(\cdot)$ to be equal to zeros or applying the modified EM algorithm [10]. Therefore, the expectation/conditional maximization (ECM) algorithm [14] is extended to solve the optimization problem. This is the modified ECM algorithm. The idea of ECM algorithm is to simplify the maximum step through conditionals. Firstly, the E-step now becomes

$$
\begin{aligned}
& Q_1(\lambda^{new}, \lambda_d^{*new} \mid \lambda^{old}, \lambda_d^{*old}) \\
= \; & Q(\lambda^{new}, \lambda_d^{*new} \mid \lambda^{old}, \lambda_d^{*old}) - \alpha \left\{ \sum_{j=1}^{m}\sum_{b=1}^{n_j}[\lambda^{new}(jb) - \bar{\lambda}_j^{new}]^2 + ||\lambda_d^{*new} - n_d^*||^2 \right\} \\
= \; & -\sum_{d=1}^{D}\sum_{j=1}^{m}\sum_{b=1}^{n_j} p(jb, d)\lambda^{new}(jb) \\
& + \sum_{d=1}^{D}\sum_{j=1}^{m}\sum_{b=1}^{n_j} \log\{p(jb, d)\lambda^{new}(jb)\} n_p^*(d) \frac{p(jb, d)\lambda^{old}(jb)}{\sum_{j'=1}^{m}\sum_{b'=1}^{n_{j'}} p(j'b', d)\lambda^{old}(j'b') + \lambda_d^{*old}(d)}
\end{aligned}
$$

$$
\begin{aligned}
&-2\sum_{d=1}^{D}\lambda_d^{*new}(d)\\
&+\sum_{d=1}^{D}\log(\lambda_d^{*new}(d))n_p^*(d)\frac{\lambda_d^{*old}(d)}{\sum_{j'=1}^{m}\sum_{b'=1}^{n_{j'}}p(j'b',d)\lambda^{old}(j'b')+\lambda_d^{*old}(d)}\\
&+\sum_{d=1}^{D}n_d^*(d)\log(\lambda_d^{*new}(d))\\
&-\alpha\left\{\sum_{j=1}^{m}\sum_{b=1}^{n_j}[\lambda^{new}(jb)-\bar{\lambda}_j^{new}]^2+\|\lambda_d^{*new}-n_d^*\|^2\right\}. \qquad (10)
\end{aligned}
$$

It is possible to find out the maximizers of (10) for all λ_d^{*new} in M-step as the solutions of

$$\frac{\partial Q_1(\lambda^{new},\lambda_d^{*new}\mid\lambda^{old},\lambda_d^{*old})}{\partial\lambda_d^{*new}(d)}=0,\ \ for\ d\ =1,\ 2,,\ldots,\ D. \qquad (11)$$

After calculation, one has

$$\mathcal{A}_1[\lambda_d^{*new}(d)]^2+\mathcal{B}_1\lambda_d^{*new}(d)-\mathcal{C}_1=0, \qquad (12)$$

where

$$
\begin{aligned}
\mathcal{A}_1 &= \alpha, && (13)\\
\mathcal{B}_1 &= 1-n_d^*(d)\alpha, && (14)\\
\mathcal{C}_1 &= \frac{1}{2}\left[\frac{n_p^*(d)\lambda_d^{*old}(d)}{\sum_{j'=1}^{m}\sum_{b'=1}^{n_{j'}}\lambda^{old}(j'b')p(j'b',d)+\lambda_d^{*old}(d)}+n_d^*(d)\right]. && (15)
\end{aligned}
$$

There are two possible roots of (12) but only one root is nonnegative, which is

$$\lambda_d^{*new}(d)=\frac{-\mathcal{B}_1+\sqrt{\mathcal{B}_1^2+4\mathcal{A}_1\mathcal{C}_1}}{2\mathcal{A}_1}. \qquad (16)$$

However, the M-step in finding the maximizers of (10) for all λ^{new} simultaneously is intractable. Instead, the CM-step is to find out λ^{new} that maximizes (10) iteratively. One possible way of CM-step is to maximize (10) for one parameter each time conditional on the other parameters being fixed. After simplification, the solution is the root of

$$\mathcal{A}[\lambda^{new}(jb)]^2 + \mathcal{B}\lambda^{new}(jb) - \mathcal{C} = 0, \tag{17}$$

where

$$\begin{aligned}
\mathcal{A} &= 2\alpha\left(\frac{n_j - 1}{n_j}\right), &(18)\\
\mathcal{B} &= 1 - 2\alpha\left[\frac{1}{n_j}\left(\sum_{s=1}^{b-1}\lambda^{new}(js) + \sum_{t=b+1}^{n_j}\lambda^{old}(jt)\right)\right], &(19)\\
\mathcal{C} &= \lambda^{old}(jb)\sum_{d=1}^{D}\frac{n_p^*(d)p(jb,d)}{\sum_{j'=1}^{m}\sum_{b'=1}^{n_{j'}}\lambda^{old}(j'b')p(j'b',d) + \lambda_d^{*old}(d)}. &(20)
\end{aligned}$$

Again, there are two possible roots of (17) and only one root is nonnegative, which is

$$\lambda^{new}(jb) = \frac{-\mathcal{B} + \sqrt{\mathcal{B}^2 + 4\mathcal{A}\mathcal{C}}}{2\mathcal{A}}. \tag{21}$$

The M-step in finding λ_d^{*new} in (12) is also a case of the CM-step because the conditionals do not affect the maximization points. Therefore, the iterative method is a modified ECM algorithm which is summarized as follows.

Algorithm 3.1 *The Modified ECM Algorithm*

1. *Choose initial values* $\lambda^{old}(jb) > 0,\ j = 1,\ 2,\ \ldots,\ m;\quad b = 1,\ 2,\ \ldots,\ n_j.$
2. *Choose initial values* $\lambda_d^{*old}(d) > 0,\ d = 1,\ 2,\ \ldots,\ D.$
3. *Compute a new estimate* $\lambda^{new}(jb)$ *by* (21) *for* $j = 1,\ 2,\ \ldots,\ m;\ b = 1,\ 2,\ \ldots,\ n_j.$
4. *Compute a new estimate* $\lambda_d^{*new}(d)$ *by* (16) *for* $d = 1,\ 2,\ \ldots,\ D.$
5. *If* $\Phi(\lambda^{new}, \lambda_d^{*new}) - \Phi(\lambda^{old}, \lambda_d^{*old})$ *is smaller than a tolerance, then stop.*
   Otherwise, go to step 3 with λ^{old} *replaced by* λ^{new} *and* λ_d^{*old} *replaced by* λ_d^{*new}.

Based on the techniques developed in Meng and Rubin [14], the modified ECM can include other conditional-maximization routes besides the one that is demonstrated above in Algorithm 3.1. After generalizing their theorems, one can show the monotonic convergence of any modified ECM algorithm, including the Algorithm 3.1. Theorems 3.2 – 3.4 hold for any modified ECM algorithms and Corollary 3.1 works for Algorithm 3.1 in PET. Let the penalized log-likelihood be $\Phi(\theta) = l(\theta) - \alpha J(\theta)$, where $l(\theta)$ is the log-likelihood term, $J(\theta)$ is the penalty term, and $\alpha > 0$.

Theorem 3.2 *For any modified ECM algorithm, the penalized log-likelihood is always increasing. That is,* $\Phi(\theta^{new}) \geq \Phi(\theta^{old})$.

Proof: The notations of (3.1) and (3.2) of Meng and Rubin [14] are adapted. Let $Q_1(\theta) = Q(\theta) - \alpha J(\theta)$. For $s = 1, 2, \ldots, S$, the sth CM-step in the tth iteration of the modified ECM finds $\theta^{(t+s/S)}$ such that

$$\begin{aligned} &Q_1(\theta^{(t+s/S)}|\theta^{(t)}) \geq Q_1(\theta|\theta^{(t)}) \\ &\textit{for all } \theta \in \Theta_s(\theta^{(t+(s-1)/S)}) \equiv \left\{\theta \in \Theta : g_s(\theta) = g_s(\theta^{(t+(s-1)/S)})\right\}. \end{aligned} \tag{22}$$

It is noted that $\theta^{(t+(s-1)/S)} \in \Theta_s(\theta^{(t+(s-1)/S)})$. By induction, (22) implies $Q_1(\theta^{(t+1)}|\theta^{(t)}) \geq Q_1(\theta^{(t)}|\theta^{(t)})$. Hence $Q_1(\theta^{new}) \geq Q_1(\theta^{old})$. Appling (1), (2), (3) of Green [20], one has that $\Phi(\theta^{new}) \geq \Phi(\theta^{old})$.

Theorem 3.3 *Assume the penalty term* $J(\theta)$ *is a convex function and* $\alpha > 0$. *Consider the conditions of (6)–(10) in Wu [21] with the replacement of* $L(\theta)$ *and* $Q(\theta)$ *by* $l(\theta) - \alpha J(\theta)$ *and* $Q(\theta) - \alpha J(\theta)$ *correspondingly. Suppose that all the conditional maximizations in* (22) *of the modified ECM are unique. Then all limit points of any modified ECM sequences* $\{\theta^{(t)}, t \geq 0\}$ *are stationary points of* $l(\theta) - \alpha J(\theta)$ *if* G *is space filling at all* $\theta^{(t)}$, *where* $G = \{g_s(\theta); s = 1, 2, \ldots, S\}$.

Proof: The proof is similar to the proof of Theorem 2 of [14], except the replacement of $L(\theta)$ and $Q(\theta)$ by $l(\theta) - \alpha J(\theta)$ and $Q(\theta) - \alpha J(\theta)$ respectively. The definition of space filling is given in (3.5) of [14].

Theorem 3.4 *Suppose that $l(\theta) - \alpha J(\theta)$ is unimodal in $\theta \in \Theta$ with θ^* being the only stationary point. Then any modified ECM sequence $\{\theta^{(t)}\}$ converges to the unique maximizer θ^* if G is space filling at all θ^*, and each CM maximization is unique.*

Proof: Use Theorem 3.3 and the uniqueness condition directly.

After the replacements of $L(\cdot)$ and $Q(\cdot)$ by $l(\cdot) - \alpha J(\cdot)$ and $Q(\cdot) - \alpha J(\cdot)$ correspondingly, all the results of Wu [21] and Meng, Rubin [14] can be applied accordingly. The above three theorems are the typical results. Based on these theorems, the next corollary for Algorithm 3.1 is shown.

Corollary 3.1 *The modified ECM sequence $\{\lambda^{(t)}, \lambda_d^{*(t)}\}$ obtained by Algorithm 3.1 increases the penalized log-likelihood, and converges to the unique maximizer, $(\hat{\lambda}_{ACRMLE1}, \hat{\lambda}^*_{d,ACRMLE1})$.*

Proof: By Theorem 3.1, $l(\theta) - \alpha J(\theta)$ is strong unimodal in (λ, λ_d^*) and $J(\cdot)$ is convex. It is easy to show that G is space filling using (3.6) of [14]. So, by Theorem 3.2, $\Phi(\lambda^{new}, \lambda_d^{*new}) \geq \Phi(\lambda^{old}, \lambda_d^{*old})$. According to Theorem 3.4, the modified ECM sequence $(\lambda^{(t)}, \lambda_d^{*(t)})$ obtained from Algorithm 3.1 converges to the unique maximizer, $(\hat{\lambda}_{ACRMLE1}, \hat{\lambda}^*_{d,ACRMLE1})$.

3.3 The Modified SAGE Algorithm

Fessler and Hero [15, 16] proposed another iterative methods to accelerate the convergence speed of the EM algorithm. They investigated a different approach in specifying the hidden-data space such that the resulting space alternating generalized expectation maximization algorithm (SAGE) has a faster convergence speed than the EM algorithm does. For instance, they found ML-SAGE-2 and ML-SAGE-3 algorithms that improve the convergence speeds for the constructions of PET images when the intensities of AC events, λ_d^*, are known. It is easy to modify their algorithms for reconstructions when λ_d^* are unknown. For instance, the ML-SAGE-3 algorithm is applied to get the ACRMLE1 as stated in the next Algorithm.

Algorithm 3.2 *The Modified SAGE-3 Algorithm*

1. *Choose initial values* $\lambda^{old}(jb) > 0,\ j = 1,\ 2,\ \ldots,\ m;\quad b = 1,\ 2,\ \ldots,\ n_j.$
2. *Choose initial values* $\lambda_d^{*old}(d) > 0,\ d = 1,\ 2,\ \ldots,\ D.$
3. $\bar{y}(d) = \sum_{j=1}^{m}\sum_{b=1}^{n_j} p(jb,d)\lambda^{old}(jb) + \lambda_d^{*old}(d),\ d = 1,\ 2,\ \ldots,\ D.$
4. $z(jb) = \min_{d:p(jb,d)\neq 0}\left\{\left(\frac{\bar{y}(d)}{p(jb,d)}\right) - \lambda^{old}(jb)\right\},\ j = 1,\ 2,\ \ldots,\ m;\quad b = 1,\ 2,\ \ldots,\ n_j.$
5. *For* $j = 1,\ 2,\ \ldots,\ m;$

$$\mathcal{A}_2 = 2\alpha(1 - \frac{1}{n_j})^2,$$

for $b = 1,\ 2,\ \ldots,\ n_j;$

$$e(jb) = \sum_{d=1}^{D}\frac{p(jb,d)n_p^*(d)}{\bar{y}(d)},$$

$$\mathcal{D}_2 = 1 - 2\alpha(\frac{1}{n_j})(1 - \frac{1}{n_j})\left(\sum_{s=1}^{b-1}\lambda^{new}(js) + \sum_{t=b+1}^{n_j}\lambda^{old}(jt)\right),$$

$$\mathcal{B}_2 = \mathcal{A}_2 z(jb) + \mathcal{D}_2,$$

$$\mathcal{C}_2 = \left(\lambda^{old}(jb) + z(jb)\right)e(jb) - \mathcal{D}_2 z(jb),$$

$$\lambda^{new}(jb) = \max\left\{\frac{-\mathcal{B}_2 + \sqrt{\mathcal{B}_2^2 + 4\mathcal{A}_2\mathcal{C}_2}}{2\mathcal{A}_2}, 0\right\},$$

$$\bar{y}(d) \leftarrow \bar{y}(d) + \left(\lambda^{new}(jb) - \lambda^{old}(jb)\right)p(jb,d)\ ,\ d = 1,\ 2,\ \ldots,\ D,$$

end b;

end j.

6. *For* $d = 1,\ 2,\ \ldots,\ D;$

$$\mathcal{B}_1 = 1 - n_d^*(d)\alpha,$$

$$\mathcal{C}_1' = \frac{1}{2}\left[\frac{n_p^*(d)\lambda_d^{*old}(d)}{\bar{y}(d)} + n_d^*(d)\right],$$

$$\lambda_d^{*new}(d) = \frac{-\mathcal{B}_1 + \sqrt{\mathcal{B}_1^2 + 4\alpha\mathcal{C}_1'}}{2\alpha},$$
$$\bar{y}(d) \leftarrow \bar{y}(d) + \left(\lambda_d^{*new}(d) - \lambda_d^{*old}(d)\right),$$

end d.

7. *If* $\Phi(\lambda^{new}, \lambda_d^{*new}) - \Phi(\lambda^{old}, \lambda_d^{*old})$ *is smaller than a tolerance, then stop.*

   Otherwise, go to the step 3 with λ^{old} *replaced by* λ^{new}. *and* λ_d^{*old} *replaced by* λ_d^{*new}.

The modified SAGE-2 algorithm for the ACRMLE is similar to the above algorithm with the step 4 replaced by the following updating formula:

$$z(jb) = \min_{d:p(jb,d)\neq 0} \left(\frac{\lambda_d^{*old}(d)}{p(jb,d)}\right). \tag{23}$$

3.4 Two- and Three- Step Approaches

When the prior boundary information is not available in advance, the one-step ACRMLE reconstruction can not be applied. Then, there is a need to generate an initial reconstruction first. This can be generated by the MLE-EM or the WLSE-ART algorithm [22]. An expert can draw the boundaries of image according to one's knowledge after the initial image is observed. Once the correlated boundaries are suggested, one can use this information and the one-step approach to get the ACRMLE. Because there are two steps of reconstruction involved now, it is called a two-step ACRMLE (ACRMLE2).

Alternatively, similar to the approaches of [23] and [22], there are three-step approaches. That is, once the boundaries are suggested after the first step, a mean estimate is obtained from the local average of initial image intensities within every regions. This provides an intermediate step. The intermediate step, $\hat{\lambda}_{MEAN}$, is only a device for further cross-reference. Since the mean estimate may oversmooth the image when the target image has inhomogeneous regions, dedicated balance is needed. Thus, the three-step ACRMLE (ACRMLE3) becomes

$$(\hat{\lambda}_{ACRMLE3}, \hat{\lambda}^*_{d,ACRMLE3}) = \underset{\lambda \geq 0, \lambda^*_d \geq 0}{argmax} \quad \Phi^{'}(\lambda, \lambda^*_d), \tag{24}$$

where

$$\Phi^{'}(\lambda, \lambda^*_d) = l(\lambda, \lambda^*_d) - \alpha \left\{ \|\lambda - \hat{\lambda}_{MEAN}\|^2 + \|\lambda^*_d - n^*_d\|^2 \right\}, \alpha > 0. \tag{25}$$

The proof of the unique existence of ACRMLE3 for any mean estimate is similar to that of ACRMLE1 in Theorem 3.1. The modified EM algorithm can be applied to find out the ACRMLE3 in (24). The algorithm is similar to the Algorithm 3.1 after replacing $\lambda^{new}(jb)$ by $\frac{-\mathcal{B}^{'} + \sqrt{\mathcal{B}^{'2} + 4\mathcal{A}^{'}\mathcal{C}^{'}}}{2\mathcal{A}^{'}}$ in step 3, where $\mathcal{A}^{'} = 2\alpha$, $\mathcal{B}^{'} = 1 - 2\alpha\hat{\lambda}_{MEAN}(jb)$, $\mathcal{C}^{'}$ is the same as (20). The computational costs of ACRMLE2 and ACRMLE3 are of the same order.

4 On The Penalty Parameter

Because the prior boundary information may be incomplete or incorrect, it is important to choose the penalty parameter to balance the trade-off between the log-likelihood term and the penalty term. As $\alpha \to 0$, the log-likelihood of ACRMLE1 approaches to that of the MLE. But if $\alpha \to \infty$, the log-likelihood of the ACRMLE1 approaches to that of local average $\bar{\lambda}_j$ and n^*_d. A proper α is sought such that the log-likelihood of ACRMLE1 is similar to that of the true phantom. In practice, the true phantom is unknown. Therefore, one can consider a dynamic graph function, like a slider, to choose a proper α with user interactions. Besides, there is an automatically data-driven approach to select a suitable penalty parameter. Now, a fast way to select the penalty parameter from data is investigated based on the generalized approximate cross validation (GACV) method [17].

Similar to the equation (1.1) and I_λ in [17], the cross-reference log-likelihood is rewritten as

$$I_\alpha = -l(\hat{\lambda}, \hat{\lambda}_d^*) + \frac{\alpha}{N_1 + N_2}\eta^T \Sigma \eta, \tag{26}$$

where $l(\hat{\lambda}, \hat{\lambda}_d^*)$ is the log-likelihood of ACRMLE1 for a given α. For a fixed α and $d = 1,\ 2,\ \ldots,\ D$, it is supposed that $\hat{\lambda}^*(d)$ is $\hat{\lambda}^*_{ACRMLE1}(d)$, $\hat{\lambda}^*_d(d)$ is $\hat{\lambda}^*_{d,ACRMLE1}(d)$, $\hat{\lambda}(jb)$ is the ACRMLE1 at the bth box in the jth region, $\hat{\hat{\lambda}}_j$ is $\frac{1}{n_j}\sum_{k=1}^{n_j}\hat{\lambda}(jk)$, N_1 is the count of $\{d : \hat{\lambda}^*(d) + \hat{\lambda}^*_d(d) \neq 0 \ or\ 1\}$, N_2 is the count of $\{d : \hat{\lambda}^*_d(d) \neq 0 \ or\ 1\}$, $R = \sum_{j=1}^m \sum_{b=1}^{n_j} [\hat{\lambda}(jb) - \hat{\hat{\lambda}}_j]^2 + \left\|\hat{\lambda}^*_d - n^*_d\right\|^2$, $S(d) = \left[\log\left(\hat{\lambda}^*(d) + \hat{\lambda}^*_d(d)\right)\right]^2$, and $T(d) = \left[\log\left(\hat{\lambda}^*_d(d)\right)\right]^2$. Then, η is a $(N_1 + N_2) \times 1$ vector, $\eta = [\log\left(\hat{\lambda}^*(1) + \hat{\lambda}^*_d(1)\right), \ldots, \log\left(\hat{\lambda}^*(N_1) + \hat{\lambda}^*_d(N_1)\right), \log\left(\hat{\lambda}^*_d(1)\right), \ldots, \log\left(\hat{\lambda}^*_d(N_2)\right)]^T$, and Σ is a $(N_1 + N_2) \times (N_1 + N_2)$ diagonal matrix with $\Sigma = diag(\frac{R}{S(1)}, \ldots, \frac{R}{S(N_1)}, \frac{R}{T(1)}, \ldots, \frac{R}{T(N_2)})$.

By the equation (2.8) in [17], the GACV function of ACRMLE1 for PET turns out to be

$$\begin{aligned} & GACV(\alpha) \\ = \ & \frac{1}{N_1 + N_2}\left[\sum_{d=1}^{N_1}\mathcal{A}_3(d) + \sum_{d=1}^{N_2}\mathcal{B}_3(d)\right] \\ + \ & \frac{1}{N_1 + N_2}\left[\sum_{d=1}^{N_1}\mathcal{C}_3(d) + \sum_{d=1}^{N_2}\mathcal{D}_3(d)\right] \\ \times \ & \left[\frac{\sum_{d=1}^{N_1} n^*_p(d)(n^*_p(d) - \mathcal{E}_3(d)) + \sum_{d=1}^{N_2} n^*_d(d)(n^*_d(d) - \mathcal{F}_3(d))}{N_1 + N_2 - \left[\sum_{d=1}^{N_1}\mathcal{C}_3(d) \times \mathcal{E}_3(d) + \sum_{d=1}^{N_2}\mathcal{D}_3(d) \times \mathcal{F}_3(d)\right]}\right], \end{aligned} \tag{27}$$

where

$\mathcal{A}_3(d) = -n^*_p(d)\log\left(\hat{\lambda}^*(d) + \hat{\lambda}^*_d(d)\right) + \hat{\lambda}^*(d) + \hat{\lambda}^*_d(d),$

$\mathcal{B}_3(d) = -n^*_d(d)\log\hat{\lambda}^*_d(d) + \hat{\lambda}^*_d(d),$

$$\mathcal{C}_3(d) = \left\{\hat{\lambda}^*(d) + \hat{\lambda}^*_d(d) + \frac{2\alpha}{N_1 + N_2}\frac{\sum_{j=1}^m \sum_{b=1}^{n_j}[\hat{\lambda}(jb) - \hat{\hat{\lambda}}_j]^2 + \left\|\hat{\lambda}^*_d - n^*_d\right\|^2}{\left[\log\left(\hat{\lambda}^*(d) + \hat{\lambda}^*_d(d)\right)\right]^2}\right\}^{-1},$$

$$\mathcal{D}_3(d) = \left\{\hat{\lambda}^*_d(d) + \frac{2\alpha}{N_1 + N_2}\frac{\sum_{j=1}^m \sum_{b=1}^{n_j}[\hat{\lambda}(jb) - \hat{\hat{\lambda}}_j]^2 + \left\|\hat{\lambda}^*_d - n^*_d\right\|^2}{\left[\log\hat{\lambda}^*_d(d)\right]^2}\right\}^{-1},$$

$\mathcal{E}_3(d) = \hat{\lambda}^*(d) + \hat{\lambda}^*_d(d)$, and

$\mathcal{F}_3(d) = \hat{\lambda}_d^*(d)$.

If the ACRMLE3 is applied, the corresponding GACV function is similar to (27) with the replacement of $\mathcal{C}_3(d)$ and $\mathcal{D}_3(d)$ by

$$\mathcal{C}_3(d) = \left\{ \hat{\lambda}^*(d) + \hat{\lambda}_d^*(d) + \frac{2\alpha}{N_1 + N_2} \frac{\left\| \hat{\lambda} - \hat{\lambda}_{MEAN} \right\|^2 + \left\| \hat{\lambda}_d^* - n_d^* \right\|^2}{\left[\log \left(\hat{\lambda}^*(d) + \hat{\lambda}_d^*(d) \right) \right]^2} \right\}^{-1}, \text{ and}$$

$$\mathcal{D}_3(d) = \left\{ \hat{\lambda}_d^*(d) + \frac{2\alpha}{N_1 + N_2} \frac{\left\| \hat{\lambda} - \hat{\lambda}_{MEAN} \right\|^2 + \left\| \hat{\lambda}_d^* - n_d^* \right\|^2}{\left[\log \hat{\lambda}_d^*(d) \right]^2} \right\}^{-1}, \text{ respectively}$$

where $\hat{\lambda}$ is the ACRMLE3, $\hat{\lambda}^*(d)$ is $\hat{\lambda}^*_{ACRMLE3}(d)$, and $\hat{\lambda}_d^*(d)$ is $\hat{\lambda}^*_{d,ACRMLE3}(d)$.

5 Simulation Results

In order to see the performance of ACRMLE and GACV in finite samples, Monte Carlo studies are simulated for different phantoms. The number of pixels, B, is equal to 128×128. The number of detectors is 128 and the number of detector tubes, D, is smaller than $\binom{128}{2}$ [1]. The ratio of $\lambda_d^*(d)$ and $\lambda^*(d)$ is set to be 0.3 to simulate the situation that there may be 30% AC events in practice [18].

Firstly, the test phantom 1 shown in Figure 1 is under investigation. Suppose the boundary information is as shown in Figure 2. This boundary information is incomplete or incorrect in the sense that the anatomic boundary in the upper part of Figure 2 is only a half of the functional boundary in the same area of Figure 1. If $\alpha = 0.001$, the reconstruction of ACRMLE1 is displayed in Figure 3. The square errors between reconstructions and the test phantom 1 are displayed in Figure 4 with respect to the penalty parameters. When the penalty parameter is $\alpha = 0.001$, the ACRMLE1 has the minimum square error as shown in Figure 4. When $\alpha \to 0$, the error of ACRMLE1 approaches to that of MLE. For a wide range of α, the error of ACRMLE1 is small than that of the MLE as shown in Figure 4. Furthermore, the ACRMLE1 makes up the missing part of the functional boundary as shown in Figure 3. The user can also find out that the prior boundary information is incomplete or incorrect through Figure 3. One is able to try to correct this prior boundary information and look for the ACRMLE based on the corrected boundary information to improve the reconstruction afterwards. The data-driven selection of α can be found by the GACV curve as shown in Figure 5. The minimum point of GACV in Figure 5, $\alpha = 0.0001$, is different from that of square errors in Figure 4, $\alpha = 0.001$. However, the square error of the ACRMLE1 with respect to this penalty parameter, $\alpha = 0.0001$, is still smaller than that of the MLE as shown in Figure 4. The reconstructions of ACRMLE2 and ACRMLE3 are of the same quality as that of ACRMLE1.

Another test phantom 2 in Figure 6 is also investigated. There are two different organs. One organ has homogeneous intensity inside and the other one does not. With the anatomic boundary information in Figure 7 and $\alpha = 0.00005$, the reconstruction of ACRMLE1 is shown in Figure 8. The calculation of the square error shows that the ACRMLE1 outperforms the MLE for a range of the penalty parameter and the square error is minimum when $\alpha = 0.00005$. The value of α is relatively to the ratio of the log-likelihood and the penalty term. One can standardize the penalty term to have a large value of the penalty parameter instead of a small value, like $\alpha = 0.00005$ here. The resulting estimators are equivalent. The GCV function can be used to suggest a penalty parameter from the observed data.

Figure 9 displays the penalized log-likelihood $\Phi(\lambda^{(i)}, \lambda_d^{*(i)})$ versus computational complexity for the modified ECM, SAGE-2, and SAGE-3 algorithms when they are applied to a simulation of PET data. It is clear to see that the convergence speed of the modified SAGE-3 algorithm is the fastest, the modified SAGE-2 is the next, and the modified ECM is the slowest. It is also observed that if the initial value in Algorithm 3.1 or 3.2 is the method-of-moments estimate (MME), then the penalized log-likelihood increases faster from the beginning than an arbitrary initial value, like 1.0. It is plotted in Figure 10. Another interesting phenomena is observed that if one use more iterations in updating λ^{new} than those in updating λ_d^{*new}, the speed of convergence is even faster.

6 Conclusion and Discussion

Whether a reconstruction algorithm can be realized in a clinical environment strongly depends on two factors. One is the correctness of the reconstructed images and the other is the reconstruction time. The accomplishment of this paper has led to a fast reconstruction with a better accuracy. Specifically, this paper promises the following achievements.

1. The one-, two-, or three- step approaches of ACRMLE for PET with AC events and attenuation have been developed. When the boundary information is known *in prior*, the one-step approach can be applied. If a primary reconstruction is needed in order to find out the boundaries, the two- or three-step approaches are utilized. Thus, there are a variety of ACRMLEs that are available in different situations.

2. The iterative methods based on row operations, such as the modified ECM algorithms, has been developed. Hence, it is efficient in computation and storage. The monotonic increasing of the object function and the convergence are proved. Combining with the modified SAGE-3 or SAGE-2 method, the convergence speeds of iterative methods are accelerated. Thus, the iterative methods are both practical and theoretical appealing.

3. The GACV method provides a fast way to select the penalty parameter automatically from observed data. The user may use this selected value as a

reference and perform the fine tune to get the most out of the correlated boundary information.

The success of this paper not only represents an academic progress in image reconstruction area, but also make it more practical to employ iterative algorithms to obtain a better image faster. This is useful to improve the reconstruction images of PET and establish a bridge to an expert system for various tomograph systems.

Finally, the convergence speed of the modified SAGE-3 is faster than the modified ECM with respect to the same computational complexity. But the minimization in step 4 of Algorithm 3.2 for every iteration consumes extra computation time. In order to overcome this problem, one way is to combine the modified ECM algorithm and SAGE-3 algorithms. For instance, the modified SAGE-3 algorithm can be applied in the beginning and then followed by the modified ECM algorithm. Thus, the modified SAGE-3 algorithm accelerates the convergence in the beginning iterations and provides a better initial estimate for the modified ECM algorithm. The convergence of this mixed algorithm is assured by the monotonic convergence of these algorithms.

Acknowledgments

The authors would like to thank Professors I.-S. Chang, C.-M. Chen, C. A. Hsiung, B. M. W. Tsui and W. H. Wong for their discussions. This research was supported by the National Science Council of Taiwan.

References

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

[2] K. Lange and R. Carson, "EM reconstruction algorithms for emission and transmission tomography," *J. Comput. Assist. Tomog.*, vol. 8, pp. 306–316, 1984.

[3] D. G. Politte and D. L. Snyder, "Corrections for accidental coincidences and attenuation in maximum-likelihood image reconstruction for positron-emission tomography," *IEEE Transactions on Medical Imaging*, vol. 10, no. 1, pp. 82–89, 1991.

[4] J. A. Fessler, N. H. Clinthorne, and W. L. Rogers, "On complete-data spaces for PET reconstruction algorithms," *IEEE Transactions on Nuclear Science*, vol. 40, no. 4, pp. 1055–1061, 1993.

[5] D. L. Snyder, M. I. Miller, L. J. J. Thomas, and D. G. Politte, "Noise and edge artifacts in maximum-likelihood reconstructions for emission tomography," *IEEE Transactions on Medical Imaging*, vol. 6, pp. 228–238, 1987.

[6] E. Veklerov and J. Llacer, "Stopping rule for the MLE algorithm based on statistical hypothesis testing," *IEEE Trans. on Medical Imaging*, vol. 6, no. 4, pp. 313–319, 1987.

[7] D. L. Snyder and M. I. Miller, "The use of sieves to stabilize images produced with the EM algorithm for emission tomography," *IEEE Transactions on Nuclear Science*, vol. NS-32, no. 5, pp. 3864–3872, 1985.

[8] I.-S. Chang and C. A. Hsiung, "Asymptotic consistency of the maximum likelihood estimate in positron emission tomography and applications," *Ann. Statist.*, vol. 22, no. 4, pp. 1871–1883, 1994.

[9] T. Hebert and R. Leahy, "A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors," *IEEE Trans. on Medical Imaging*, vol. 8, no. 2, pp. 194–202, 1989.

[10] P. J. Green, "Bayesian reconstructions from emission tomography data using a modified EM algorithm," *IEEE Trans. Med. Imaging*, vol. 9, pp. 84–93, Mar. 1990.

[11] G. T. Herman, A. R. D. Pierro, and N. Gai, "On methods for maximum a posteriori image reconstruction with a normal prior," *Journal of Visual Communication and Image Representation*, vol. 3, no. 4, pp. 316–324, 1992.

[12] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 6, pp. 721–740, 1984.

[13] X. Ouyang, W. H. Wong, V. E. Johnson, X. Hu, and C.-T. Chen, "Incorporation of correlated structural images in PET image reconstruction," *IEEE Transactions on Medical Imaging*, vol. 13, no. 4, pp. 627–640, 1994.

[14] X. L. Meng and D. B. Rubin, "Maximum likelihood estimation via the ECM algorithm: A general framework," *Biometrika*, vol. 80, no. 2, pp. 267–278, 1993.

[15] J. A. Fessler and A. O. Hero, "Space-alternating generalized expectation-maximization algorithm," *IEEE Trans. on Signal Processing*, vol. 42, pp. 2664–2677, 1994.

[16] J. A. Fessler and A. O. Hero, "Penalized maximum-likelihood image reconstruction using space-alternating generalized expectation-maximization algorithms," *IEEE Trans. on Image Processing*, vol. 4, no. 10, pp. 1417–1429, 1995.

[17] D. Xiang and G. Wahba, "A generalized approximate cross validation for smoothing splines with non-Gaussian data," *Statistica Sinica*, vol. 6, pp. 675–692, 1996.

[18] D. G. Politte and D. L. Snyder, "Corrections for accidental coincidences and attenuation in maximum-likelihood image reconstruction for positron-emission tomography," *IEEE Trans. Med. Imaging*, vol. 10, no. 1, pp. 82–89, 1991.

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

[20] P. J. Green, "On use of the EM algorithm for penalized likelihood estimation," *J. Roy. Statist. Soc.*, vol. B-52, no. 3, pp. 443–452, 1990a.

[21] C. F. J. Wu, "On the convergence properties of the EM algorithm," *The Annals of Statistics*, vol. 11, pp. 95–103, 1983.

[22] H. H.-S. Lu, C.-M. Chen, and I.-H. Yang, "Cross-reference weighted least square estimates for positron emission tomography." Technical Report, 1996.

[23] C.-M. Chen, H. H.-S. Lu, and Y.-P. Hsu, "Cross-reference maximum likelihood estimates for positron emission tomography." Technical Report, 1996.

[24] G. H. Golub, M. Heath, and G. Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," *Tchnometrics*, vol. 21, no. 2, pp. 215–223, 1979.

Figure Captions

Figure 1: The test phantom 1.

Figure 2: The prior boundary information for the test phantom 1.

Figure 3: The ACRMLE1 reconstruction with a penalty parameter of 0.001.

Figure 4: The 2-norm errors of the ACRMLE1 and the MLE with respect to the penalty parameters. The base of logarithm function is 10.

Figure 5: The GACV curve of ACRMLE1 with respect to the penalty parameters. The base of logarithm function is 10.

Figure 6: The test phantom 2.

Figure 7: The prior boundary information for the test phantom 2.

Figure 8: The ACRMLE1 reconstruction with a penalty parameter of 0.00005.

Figure 9: The convergence speeds of the modified ECM, SAGE-2, and SAGE-3 algorithms with respect to the computational complexity.

Figure 10: The convergence speeds for different initial values with respect to the computational complexity.