

Vector-Extrapolated Fast Maximum Likelihood Estimation Algorithms for Emission Tomography

N. Rajeevan, *Member, IEEE*, K. Rajgopal, *Member, IEEE*, and G. Krishna

Abstract—A new class of fast maximum likelihood estimation (MLE) algorithms for emission computed tomography (ECT) is developed. In these cyclic iterative algorithms, vector extrapolation techniques are integrated with the iterations in gradient-based MLE algorithms, with the objective of accelerating the convergence of the base iterations. This results in a substantial reduction in the effective number of base iterations required for obtaining an emission density estimate of specified quality. The mathematical theory behind the minimal polynomial and reduced rank vector extrapolation techniques, in the context of emission tomography, is presented. With the EM and EM search algorithms in the base iterations, these extrapolation techniques are implemented in a positron emission tomography system. Using computer experiments, with measurements taken from simulated phantoms, the new algorithms are evaluated. It is shown that, with minimal additional computations, the proposed approach results in substantial improvement in reconstruction, in terms of both qualitative visual performance and quantitative measures of likelihood and residual error, of the image.

I. INTRODUCTION

EMISSION computed tomography (ECT) is a powerful medical imaging modality suitable for the study of human physiology and organ functions. In ECT, the image reconstruction problem is basically the estimation of radionuclide concentrations inside the body from measurements of photon counts taken from outside. Since the processes involved in ECT are intrinsically stochastic in nature, a statistical approach for estimating the emission densities has been felt necessary. A maximum likelihood estimation (MLE) approach to ECT image reconstruction was first suggested by Rockmore and Macovski [21]. An efficient, and simple to implement, iterative MLE algorithm for emission tomography was introduced by Shepp and Vardi [22] and by Lange and Carson [11] independently. This algorithm was based on the expectation maximization (EM) formalism of maximum likelihood estimation from incomplete data (see Dempster et al. [4]). It was shown that the MLE, implemented using the EM algorithm, optimally compensates the statistical noise in measurement data and that with low-count measurement data it produces better reconstructed images than those produced by deterministic

reconstruction algorithms [3]. Moreover, it can easily incorporate many of the physical processes involved in ECT, such as photon attenuation, scatter, positron range and angulation, and detector efficiency. This algorithm has been successfully applied in positron emission tomography (PET) [9], [22], [24], time of flight positron emission tomography (TOFPET) [30], and single photon emission computed tomography (SPECT) [17].

The EM algorithm for ECT image reconstruction produces a sequence of images which converge to the maximum likelihood estimate (MLE) of the emission densities. Though the quality of the MLE image produced by this algorithm has been very encouraging, it has not been found suitable for routine practical applications owing to the excessive computational requirement of each EM iteration step and slow convergence. This drawback has prompted many researchers to suggest various ways of efficiently implementing and accelerating the convergence of the EM algorithm. For example, exploitation of the symmetries of the physical system [9], use of a multigrid approach [20], and the high frequency enhanced filtered iterative reconstruction (FIR) method [31] have been proposed for efficient implementation. A number of traditional numerical analysis techniques have been suggested for accelerating the convergence of the EM algorithm. Once such approach is the expectation maximization search (EMS) algorithm [10]. In this modified EM algorithm each EM iteration step is used to define a direction along which a one-dimensional search for maximizing the likelihood is made to obtain the new estimate. Lewitt and Muehllehner [13] give a simpler version of the EMS algorithm, in which the image update, calculated using the standard EM algorithm, is multiplied at each iteration step by an overrelaxation parameter. A class of rescaled gradient algorithms has also been suggested in [10]. All these algorithms have a common feature of finding a direction at each iteration step and obtaining a new estimate in that direction.

In this paper we introduce a class of fast cyclic iterative algorithms suitable for the maximum likelihood estimation of emission densities in ECT. The approach is based on the acceleration of convergence by vector extrapolation [2], [5], [16], [26], [28]. Each cycle of these algorithms involves a few iterations of a standard gradient-based algorithm. The sequence of image estimates thus generated is vector extrapolated to yield a new image estimate. Various strategies for vector extrapolation are presented. In particular, two promising methods, the minimal polynomial and reduced rank extrapolation methods, have been adapted to the proposed algorithm and

Manuscript received July 23, 1990; revised August 14, 1991.

N. Rajeevan was with the Department of Electrical Engineering, Indian Institute of Science, Bangalore, India 560012. He is now with the Department of Nuclear Medicine, University of Massachusetts Medical Center, Worcester, MA 01655.

K. Rajgopal is with the Department of Electrical Engineering, Indian Institute of Science, Bangalore, India 560012.

G. Krishna is with the Department of Computer Science and Automation, Indian Institute of Science, Bangalore, India 560012.

IEEE Log Number 9104848.

implemented. A preliminary result on this work appeared in [19].

The organization of the paper is as follows. In Section II, we discuss the standard EM algorithm and some of its variants. The EMS algorithm and the rescaled gradient approach are presented. In Section III, the mathematical theory behind the vector extrapolation approach, in the context of emission tomography, leading to the development of the new class of cyclic iterative algorithms is presented. The issues related to the practical implementation of these algorithms are also discussed. Section IV is on computer-simulated experiments. The proposed new algorithms are implemented in a positron emission tomography system and their convergence properties are compared with standard algorithms. In the concluding section we explore the potential of the proposed approach and address some of the issues remaining unsolved.

II. THE EM ALGORITHM AND ITS VARIANTS

The EM algorithm is an iterative procedure for finding the maximum likelihood estimate of a probability distribution function from an incomplete set of measurement data, where the incomplete data may be viewed as a many-to-one function of some unobserved complete data. This algorithm is attractive in situations where it is impractical to maximize the incomplete data likelihood function, whereas the complete data likelihood has a functional form which is relatively easy to maximize.

By assuming a spatially independent Poisson model for the emission process, the measurements of photon counts in ECT can also be modeled as independent Poisson processes. With these models for the emission and the measurement processes, the loglikelihood function is given by

$$L(\boldsymbol{\lambda}) = \sum_{j=1}^M \left\{ - \sum_{i=1}^N \lambda_i p_{ij} + y_j \log \left(\sum_{i=1}^N \lambda_i p_{ij} \right) - \log(y_j!) \right\} \quad (2.1)$$

where N is the number of pixels in the image and M is the number of measurements. The vector $\boldsymbol{\lambda} = \{\lambda_i, i = 1, \dots, N\}$ is the parameter vector of the emission process, which is to be estimated, and $\mathbf{y} = \{y_j, j = 1, \dots, M\}$ is the measurement vector. Each p_{ij} is the probability that a photon emitted in the i th pixel is detected by the j th detector. In the case of PET, p_{ij} is the probability of detecting a pair of annihilation photons, created in the i th pixel, by the j th detector pair. This loglikelihood function has been shown [22] to be concave; hence, a vector $\boldsymbol{\lambda}^{\text{ML}}$, the maximum likelihood estimate at which $L(\boldsymbol{\lambda})$ attains its maximum, exists. However, a closed-form formula for finding $\boldsymbol{\lambda}^{\text{ML}}$ as a maximizer of the loglikelihood function $L(\boldsymbol{\lambda})$ does not exist.

The EM algorithm is an iterative procedure which is appropriate in the above situation for finding the maximum likelihood estimate $\hat{\boldsymbol{\lambda}}^{\text{ML}}$ of the emission densities. Starting from a strictly positive initial estimate $\hat{\boldsymbol{\lambda}}^0$, this algorithm computes a sequence of estimates which converge to the MLE, $\hat{\boldsymbol{\lambda}}^{\text{ML}}$. At the k th iteration step, the new estimate $\hat{\boldsymbol{\lambda}}^{k+1}$ is

computed from the current estimate $\hat{\boldsymbol{\lambda}}^k$ as

$$\hat{\lambda}_i^{k+1} = \frac{\hat{\lambda}_i^k}{\sum_j p_{ij}} \sum_j \frac{y_j p_{ij}}{\sum_{i'} \hat{\lambda}_{i'}^k p_{i'j}}. \quad (2.2)$$

This algorithm has the property that $L(\hat{\boldsymbol{\lambda}}^{k+1}) > L(\hat{\boldsymbol{\lambda}}^k)$, unless it has already converged. It has been shown that the estimates $\hat{\boldsymbol{\lambda}}^k$ converge to the maximum likelihood estimate $\boldsymbol{\lambda}^{\text{ML}}$ [23]. Images of emission densities produced by the EM algorithm are superior to those produced by deterministic reconstruction algorithms, such as the convolution backprojection (CBP) method, especially for low photon counts [3]. However, the major drawbacks of the EM algorithm that limit its use in routine practical applications are its high computational requirement and slow convergence. Various faster algorithms, based on traditional numerical analysis techniques, have been suggested for the maximum likelihood estimation of emission densities. These include the EM search (EMS), rescaled gradient, and Newton-type algorithms.

The standard EM iteration step, (2.2), can be rewritten as

$$\hat{\lambda}_i^{k+1} = \hat{\lambda}_i^k + \frac{\hat{\lambda}_i^k}{\sum_j p_{ij}} \sum_j \frac{y_j - \sum_{i'} \hat{\lambda}_{i'}^k p_{i'j}}{\sum_{i'} \hat{\lambda}_{i'}^k p_{i'j}} p_{ij}. \quad (2.3)$$

In vector matrix notation this is expressed as

$$\begin{aligned} \hat{\boldsymbol{\lambda}}^{k+1} &= \hat{\boldsymbol{\lambda}}^k + \mathbf{r}(\hat{\boldsymbol{\lambda}}^k) \mathbf{g}(\hat{\boldsymbol{\lambda}}^k) \\ &= \hat{\boldsymbol{\lambda}}^k + \Delta(\hat{\boldsymbol{\lambda}}^k), \end{aligned} \quad (2.4)$$

where the diagonal matrix $\mathbf{r}(\hat{\boldsymbol{\lambda}}^k)$ has elements $\left(\frac{\hat{\lambda}_i^k}{\sum_j p_{ij}}, i = 1, \dots, N \right)$, and $\mathbf{g}(\hat{\boldsymbol{\lambda}}^k)$ is the loglikelihood gradient vector evaluated at $\boldsymbol{\lambda} = \hat{\boldsymbol{\lambda}}^k$. The i th component of $\mathbf{g}(\hat{\boldsymbol{\lambda}}^k)$ is given by

$$\left[\mathbf{g}(\hat{\boldsymbol{\lambda}}^k) \right]_i = \sum_j \frac{y_j - \sum_{i'} \hat{\lambda}_{i'}^k p_{i'j}}{\sum_{i'} \hat{\lambda}_{i'}^k p_{i'j}} p_{ij}. \quad (2.5)$$

The quantity $\mathbf{A}(\hat{\boldsymbol{\lambda}}^k)$ is the correction vector, which, when added to the current estimate, $\hat{\boldsymbol{\lambda}}^k$, gives the new EM estimate, $\hat{\boldsymbol{\lambda}}^{k+1}$.

Instead of considering $\mathbf{A}(\hat{\boldsymbol{\lambda}}^k)$ as the correction vector at the k th iteration step, it can be viewed that $\Delta(\hat{\boldsymbol{\lambda}}^k)$ defines the direction along which $\hat{\boldsymbol{\lambda}}^k$ moves to the new estimate. In the EMS algorithm a one-dimensional search in this direction is made to obtain the new estimate, $\hat{\boldsymbol{\lambda}}^{k+1}$, as the point at which the likelihood function attains its maximum in that direction. The EMS iteration step is given by

$$\hat{\boldsymbol{\lambda}}^{k+1} = \hat{\boldsymbol{\lambda}}^k + \tau \Delta(\hat{\boldsymbol{\lambda}}^k), \quad (2.6)$$

where $\tau > 0$ is a constant such that $\{\hat{\boldsymbol{\lambda}}^k + \tau \mathbf{A}(\hat{\boldsymbol{\lambda}}^k), \tau > 0\}$ defines the line in R^n , originating from $\hat{\boldsymbol{\lambda}}^k$ and having the direction of the vector $\mathbf{A}(\hat{\boldsymbol{\lambda}}^k)$. A one-dimensional Newton-Raphson search algorithm for finding the τ for which

the loglikelihood function $L(\hat{\lambda}^k + \tau \Delta(\hat{\lambda}^k))$ is maximum is given in [10].

The EM iteration has an important geometrical interpretation. At each iteration step the gradient vector of the loglikelihood function $g(\hat{\lambda}^k)$ is calculated and rescaled using the diagonal matrix $\Gamma(\hat{\lambda}^k)$ to obtain the correction vector $\Delta(\hat{\lambda}^k)$. This rescaling enhances the movement of the components with higher values and damps the movement of those with smaller component values. This has the effect of moving in a direction different from that of the gradient. The EMS algorithm can be seen as maximizing the likelihood function along this modified direction. This is in contrast to the steepest ascent method, where the new estimate is found in the gradient direction. In a more general class of rescaled gradient algorithms, the gradient vector is rescaled by a power of the diagonal matrix $\Gamma(\hat{\lambda}^k)$ to obtain the correction vector. This algorithm has the form

$$\hat{\lambda}^{k+1} = \hat{\lambda}^k + \left(\Gamma(\hat{\lambda}^k)\right)^n g(\hat{\lambda}^k) \quad (2.7)$$

The EM algorithm can be seen as a special case of this algorithm with n equal to unity.

All the algorithms discussed so far in the present section have a general form of finding a direction at each iteration step, and moving in this direction to obtain the new estimate. The acceleration of convergence of the iterates is then achieved by making a search for maximum likelihood in this direction. On the other hand, the Newton-type algorithms use a kind of curvature information in the form of a Hessian matrix. Though these algorithms have much better convergence properties than the gradient algorithms, the computation of the Hessian and its inversion is highly impractical in ECT image reconstruction problems. Attempts to unify the EM methodology and Newtons methods [15] were also not successful in this context.

In the next section we develop a new class of fast ML algorithms for emission density estimation in ECT. These algorithms are based on vector extrapolation for accelerating the convergence of the iterates. In the gradient and Newton-type algorithms the new estimate is computed using the measurement data and the present estimate. The algorithms we develop here make use of a finite number of previous estimates also, along with the present estimate and the measurement data, for computing the new estimate.

III. VECTOR-EXTRAPOLATED FAST ML ALGORITHMS

The gradient-based algorithms and their modifications, discussed in the previous section, compute the new estimate from the current estimate and the measurement data. The sequence of iterates produced by these algorithms converges asymptotically to the maximum likelihood solution. The class of fast cyclic iterative algorithms for emission density estimation that we develop in this paper is based on the acceleration of convergence by vector extrapolation. In this approach, each cycle of the algorithm consists in generating a finite number of consecutive estimates by a gradient-based algorithm which are used along with the measurement data to obtain a new

extrapolated estimate of the emission densities. This density estimate is used in the next cycle as the starting estimate for the gradient-based algorithm. This process of cyclic iteration is continued until acceptable convergence is obtained. In this section we present the mathematical theory behind this approach.

Accelerating the convergence of a vector sequence by extrapolation has been an active area of research since the early 1960's. The extrapolation methods we develop for maximum likelihood estimation in this paper are nonlinear, in the sense that the coefficients of the extrapolating polynomials are functions of the terms of the sequence itself. The basic ideas of nonlinear vector extrapolation methods were first developed for accelerating the convergence of a sequence of vectors generated by a linear operator. In this context, Cabay and Jackson [2] introduced the minimal polynomial extrapolation (MPE), and Eddy [5] and Mesina [16] developed the reduced rank extrapolation (RRE) method. The extension of these methods to the acceleration of convergence of a vector sequence generated by a nonlinear operator has been given by Skelboe [27] and Smith *et al.* [28]). For a theoretical study on the convergence and stability properties of MPE and RRE algorithms, see Sidi [25]. In the following we adapt these results and develop a class of fast algorithms for maximum likelihood estimation of emission densities.

The maximum likelihood estimate $\hat{\lambda}^{\text{ML}}$ to which the iterates generated by a gradient-based algorithm converge can be written as

$$\begin{aligned} \hat{\lambda}^{\text{ML}} &= \hat{\lambda}^0 + \sum_{i=0}^{\infty} \Delta(\hat{\lambda}^i) \\ &= \hat{\lambda}^k + \sum_{i=0}^{\infty} \Delta(\hat{\lambda}^{k+i}), \end{aligned} \quad (3.1)$$

where $\Delta(\hat{\lambda}^k)$ is the correction vector computed at the k th iteration step. The estimates $\hat{\lambda}^k$ converge asymptotically to $\hat{\lambda}^{\text{ML}}$, and

$$\|\Delta(\hat{\lambda}^k)\| \rightarrow 0 \quad \text{as } k \rightarrow \infty, \quad (3.2)$$

where $\|\mathbf{u}\|$ denotes the Euclidean length of the vector \mathbf{u} . At each iteration step k , the new estimate $\hat{\lambda}^{k+1}$ can be considered as generated by a nonlinear operator A_k , as

$$\hat{\lambda}^{k+1} = A_k \hat{\lambda}^k. \quad (3.3)$$

The new class of cyclic iterative algorithms we develop here uses a local linearization of this operator. Each cycle of these algorithms consists in generating a small number, m , of consecutive estimates by a gradient-based algorithm. Using the local linearization approximation, we assume that these m estimates are generated by a fixed linear operator. Hence, for the n th cycle, the estimate $\hat{\lambda}_n^k$, $k = 1, \dots, m$, produced by the gradient-based algorithm with $\hat{\lambda}_n^0$ as the initial estimate, is considered to be generated as

$$\hat{\lambda}_n^k = B_n \hat{\lambda}_n^{k-1} + b_n, \quad k = 1, \dots, m \quad (3.4)$$

where B_n is a linear operator and b_n is a constant vector, both fixed for the n th cycle. With this local linearization assumption

we can use vector extrapolation techniques for obtaining an estimate $\hat{\lambda}_{n,m}$ as the vector to which the estimates $\hat{\lambda}_n^k$ would converge under the iterations in (3.4). This estimate $\hat{\lambda}_{n,m}$ is used as the starting estimate in the next cycle. This process of cyclic iteration, which is aimed at alleviating the errors introduced by the local linearization of the nonlinear operator \mathbf{A}_k , is continued until acceptable convergence is achieved.

Various approaches can be adopted for obtaining $\hat{\lambda}_{n,m}$. In one such approach we compute \mathbf{A}_m as a weighted sum of the estimates $\hat{\lambda}_n^k, k = 0, \dots, m$. An algorithm where these weights are computed from the coefficients of the minimal polynomial of the linear operator \mathbf{B}_n is given below. Alternatively, $\hat{\lambda}_{n,m}$ can be computed by adding a weighted sum of the correction vectors $\mathbf{A}(\hat{\lambda}_n^k), k = 0, \dots, m-1$ to the initial estimate of the cycle $\hat{\lambda}_n^0$, where $\mathbf{A}(\hat{\lambda}_n^k)$ is defined as

$$\Delta(\hat{\lambda}_n^k) = \hat{\lambda}_n^{k+1} - \hat{\lambda}_n^k. \quad (3.5)$$

The reduced rank extrapolation algorithm discussed later in this section is based on this approach.

A. Minimal Polynomial Extrapolated ML (MPEML) Algorithms

In the MPEML algorithms each cycle consists in finding an extrapolated estimate as a weighted sum of a finite number of intermediate estimates generated by a gradient-based algorithm. For the n th cycle this extrapolated estimate, \mathbf{A}_m is given by

$$\hat{\lambda}_{n,m} = \sum_{k=0}^m w_k \hat{\lambda}_n^k, \quad (3.6)$$

where the weights $w_k, k = 0, \dots, m$, are related to the coefficients of the minimal polynomial of \mathbf{B}_n , the linear operator generating the intermediate estimates $\hat{\lambda}_n^k, k = 1, \dots, m$, with respect to the correction vector $\mathbf{A}(\hat{\lambda}_n^0)$. The minimal polynomial of an operator \mathbf{B} with respect to a vector \mathbf{v} is defined as the unique monic polynomial $P(\mathbf{B})$ of minimal degree which annihilates the vector \mathbf{v} . For the linear operator \mathbf{B}_n and the correction vector $\mathbf{A}(\hat{\lambda}_n^0)$, this polynomial, $P(\mathbf{B}_n)$, is such that

$$P(\mathbf{B}_n)\Delta(\hat{\lambda}_n^0) = \mathbf{0}. \quad (3.7)$$

In the mathematical development of the MPEML algorithm we consider the number of intermediate estimates used for extrapolation to be equal to the degree of the minimal polynomial $P(\mathbf{B}_n)$. However, the linear operator \mathbf{B}_n is not known explicitly; hence the degree of $P(\mathbf{B}_n)$ cannot be computed. In the practical implementation of these algorithms we arbitrarily choose this number to be a small positive integer. As shown by computer-simulated experiments in the next section, a value for m as small as 2 has produced substantial improvement in the convergence of the iterates.

The development of the minimal polynomial algorithm for the extrapolated estimate $\lambda_{n,m}$ is as follows: we first show how the coefficients $c_k, k = 0, \dots, m$, of the minimal polynomial $P(\mathbf{B}_n)$ with respect to $\mathbf{A}(\hat{\lambda}_n^0)$ can be computed from the

vectors $\Delta(\hat{\lambda}_n^k), k = 0, \dots, m$ (eqn. (3.17)). We then prove that, if $P(\mathbf{B}_n)$ is the minimal polynomial with respect to $\Delta(\hat{\lambda}_n^0)$, then it is also the minimal polynomial with respect to $(\hat{\lambda}_n^0 - \hat{\lambda}_{n,m})$, the difference vector between the initial estimate and the extrapolated estimate (eqn. (3.23)), i.e.,

$$P(\mathbf{B}_n)\Delta(\hat{\lambda}_n^0) = \mathbf{0} \implies P(\mathbf{B}_n)(\hat{\lambda}_n^0 - \hat{\lambda}_{n,m}) = \mathbf{0}. \quad (3.8)$$

Using this result, we obtain an expression for $\lambda_{n,m}$ as a weighted sum of $\hat{\lambda}_n^k, k = 0, \dots, m$, where the weights are related to the coefficients of the minimal polynomial (eqns. (3.26) and (3.27)).

Let the minimal polynomial $P(\mathbf{B}_n)$ of \mathbf{B}_n with respect to $\Delta(\hat{\lambda}_n^0)$ be given by

$$P(\mathbf{B}_n) = \sum_{k=0}^m c_k \mathbf{B}_n^k, \quad (3.9)$$

where m is the degree of the minimal polynomial, and $c_k, k = 0, \dots, m$, are the coefficients with $c_0 = 1$. Then from (3.9) we have

$$\sum_{k=0}^m c_k \mathbf{B}_n^k \Delta(\hat{\lambda}_n^0) = \mathbf{0}. \quad (3.10)$$

From the definition of $\mathbf{A}(\hat{\lambda}_n^k)$ we have

$$\mathbf{B}_n \Delta(\hat{\lambda}_n^k) = \Delta(\hat{\lambda}_n^{k+1}). \quad (3.11)$$

Using this in (3.10) we get

$$\sum_{k=0}^m c_k \Delta(\hat{\lambda}_n^k) = \mathbf{0}. \quad (3.12)$$

Since $P(\mathbf{B}_n)$ is a monic polynomial, $c_m = 1$ and (3.12) can be rewritten as

$$\sum_{k=0}^{m-1} c_k \Delta(\hat{\lambda}_n^k) = -\Delta(\hat{\lambda}_n^m). \quad (3.13)$$

In vector matrix notations this is expressed as

$$\mathbf{D}\mathbf{c} = -\Delta(\hat{\lambda}_n^m), \quad (3.14)$$

where the matrix \mathbf{D} , made up of m column vectors, is defined as

$$\mathbf{D} \doteq \left[\Delta(\hat{\lambda}_n^0) \Delta(\hat{\lambda}_n^1) \dots \Delta(\hat{\lambda}_n^{m-1}) \right], \quad (3.15)$$

and \mathbf{c} is the vector of coefficients given by

$$\mathbf{c} \doteq [c_0 c_1 \dots c_{m-1}]^t. \quad (3.16)$$

A least-squares solution to the overdetermined system of equation (3.14) is obtained as [1]

$$\mathbf{c} = -\mathbf{D}^\dagger \Delta(\hat{\lambda}_n^m), \quad (3.17)$$

where \mathbf{D}^\dagger is the Moore Penrose generalized inverse of \mathbf{D} , which is computed as

$$\mathbf{D}^\dagger = [\mathbf{D}^t \mathbf{D}]^{-1} \mathbf{D}^t. \quad (3.18)$$

Hence, the coefficients c_k , $k = 0, \dots, m$, of the minimal polynomial $P(B_n)$ can be computed from the estimates $\hat{\lambda}_n^k$, $k = 0, \dots, m$, using (3.5), (3.18), and (3.17). We will now see how these coefficients can be used for the computation of the weights w_k in (3.6) for finding $\hat{\lambda}_{n,m}$.

The estimate $\hat{\lambda}_{n,m}$ has been taken as the vector to which $\hat{\lambda}_n^k$ would converge under the iterations in 13.4). If we assume that 1 is not an eigenvalue of B_n , then $\lambda_{n,m}$ is unique and satisfies

$$\hat{\lambda}_{n,m} = B_n \hat{\lambda}_{n,m} + \mathbf{b}_n. \quad (3.19)$$

From the definition of $A(\hat{\lambda}_n^k)$ and using (3.4) we have

$$\Delta(\hat{\lambda}_n^0) = (B_n - I)\hat{\lambda}_n^0 + \mathbf{b}_n, \quad (3.20)$$

where I is the unity matrix. Substituting for \mathbf{b}_n from (3.19), we get

$$\Delta(\hat{\lambda}_n^0) = (B, -I)(\hat{\lambda}_n^0 - \hat{\lambda}_{n,m}). \quad (3.21)$$

Premultiplying both sides of (3.21) by $P(B_n)$ and using (3.7) we have

$$P(B_n)(B_n - I)(\hat{\lambda}_n^0 - \hat{\lambda}_{n,m}) = 0. \quad (3.22)$$

By observing that $(B, -I)$ commutes with any polynomial in $(B,)$ and if we assume that $(B, -I)$ is invertible, then (3.20) gives

$$P(B_n)(\hat{\lambda}_n^0 - \hat{\lambda}_{n,m}) = 0 \quad (3.23)$$

i.e.,

$$\sum_{k=0}^m c_k B_n^k (\hat{\lambda}_n^0 - \hat{\lambda}_{n,m}) = 0. \quad (3.24)$$

Using (3.4) and (3.19), (3.24) gives

$$\sum_{k=0}^m c_k \hat{\lambda}_{n,m} = \sum_{k=0}^m c_k \hat{\lambda}_n^k. \quad (3.25)$$

From this $\hat{\lambda}_{n,m}$ is obtained as

$$\hat{\lambda}_{n,m} = \sum_{k=0}^m w_k \hat{\lambda}_n^k, \quad (3.26)$$

where the weights w_k are given by

$$w_k = \frac{c_k}{\sum_{j=0}^m c_j}. \quad (3.27)$$

The estimate $\hat{\lambda}_{n,m}$ is used as the starting estimate in the next cycle by setting $\hat{\lambda}_{n+1}^0 = \hat{\lambda}_{n,m}$ and the cyclic iteration is continued until convergence is obtained.

If the gradient-based algorithm which generates the intermediate estimates has the self-normalization property of preserving the total activity in the estimate, i.e.,

$$\sum_i [\hat{\lambda}_n^k]_i = \sum_j y_j, \quad k = 0 \dots m, \quad (3.28)$$

then the MPEML estimate $\hat{\lambda}_{n,m}$ also preserves the total activity. In the case of EM and EMS algorithms this property is automatically satisfied. With such self-normalizing estimation algorithms in the base iterations, the self-normalization property of the MPEML algorithm can be given as follows:

$$\begin{aligned} \sum_i [\hat{\lambda}_{n,m}]_i &= \sum_i \sum_{k=0}^m w_k [\hat{\lambda}_n^k]_i \\ &= \sum_{k=0}^m w_k \sum_i [\hat{\lambda}_n^k]_i \\ &= \sum_{k=0}^m w_k \sum_j y_j \\ &= \sum_j y_j, \end{aligned} \quad (3.29)$$

since $\sum_{k=0}^m w_k = 1.0$.

Since the coefficients (c_k 's) determining the weights (w_k 's) are obtained as the minimum norm solution by (3.17), some of the weights can be negative. In such a case the positivity of the extrapolated image cannot be guaranteed. In particular, where image values are small, the extrapolated image might become negative. These negative pixels are offset to a "tiny" positive value before starting the next cycle. If the base iteration has the self-normalizing property, the total counts will automatically be normalized in subsequent iterations.

As mentioned earlier, the error introduced in the estimate by the local linearization assumption is reduced by the process of cyclic iteration. It has been shown [25] that with cyclic iteration, nonlinear vector extrapolation methods in general converge quadratically. This algorithm, the minimal polynomial extrapolated maximum likelihood (MPEML) algorithm, is given below.

Algorithm: MPEML

Initialization:

$\hat{\lambda}_n^0$ = strictly positive initial estimate normalized to have total activity equal to that of the measurements.

m = degree of the minimal polynomial (taken as a small positive integer).

$n = 0$, the cycle index.

MPEML:

do

{

generate $\hat{\lambda}_n^k$, $k = 1, \dots, m+1$, using a gradient-based algorithm.

find $\Delta(\hat{\lambda}_n^k) = \hat{\lambda}_n^{k+1} - \hat{\lambda}_n^k$, $k = 0, \dots, m$.

find the coefficient vector, $\mathbf{c} = [c_0 c_1 \dots c_{m-1}]^t$, of the minimal polynomial using

$$\mathbf{c} = -D^t \Delta(\hat{\lambda}_n^0), \quad c_m = 1.0,$$

where

$$\mathbf{D} \doteq \left[\Delta(\hat{\lambda}_n^0) \Delta(\hat{\lambda}_n^1) \cdots \Delta(\hat{\lambda}_n^{m-1}) \right]$$

and \mathbf{D}^\dagger , the Moore–Penrose generalized inverse, is computed as

$$\mathbf{D}^\dagger = [\mathbf{D}^t \mathbf{D}]^{-1} \mathbf{D}^t.$$

find weights w_k , $k = 0, \dots, m$, from

$$w_k = \frac{c_k}{\sum_{i=0}^m c_i}, \quad k = 0, \dots, m.$$

compute $\lambda_{n,m}$, the minimal polynomial extrapolated estimate as the weighted sum of $\hat{\lambda}_n^k$, $k = 0, \dots, m$, as

$$\hat{\lambda}_{n,m} = \sum_{k=0}^m w_k \hat{\lambda}_n^k$$

Initialize for next cycle

$$\begin{aligned} \hat{\lambda}_{n+1}^0 &= \lambda_{n,m} \\ n &\leftarrow n + 1 \end{aligned}$$

} until convergence.

$\hat{\lambda}_n^0$ is the MPEML estimate of the emission density.

B. Reduced Rank Extrapolated ML (RREML) Algorithms

In RREML algorithms, for each cycle n , the extrapolated estimate $\lambda_{n,m}$ is obtained by adding a weighted sum of the correction vectors $\mathbf{A}(\hat{\lambda}_n^k)$, $k = 0, \dots, m-1$, to the estimate $\hat{\lambda}_n^0$, i.e.,

$$\hat{\lambda}_{n,m} = \hat{\lambda}_n^0 + \sum w_k \Delta(\hat{\lambda}_n^k), \quad (3.30)$$

where the w_k 's are the weights. As in the case of the MPEML algorithms we assume that, for a given cycle n , the intermediate estimates $\hat{\lambda}_n^k$, $k = 1, \dots, m$, are generated by a fixed linear operator B_n as given by (3.4). Further, we also assume that 1 is not an eigenvalue of B_n so that (3.19) holds. Multiplying both sides of (3.30) by \mathbf{B}_n , we have

$$\mathbf{B}_n \hat{\lambda}_{n,m} = \mathbf{B}_n \hat{\lambda}_n^0 + \sum_{k=0}^{m-1} w_k \mathbf{B}_n \Delta(\hat{\lambda}_n^k). \quad (3.31)$$

Using (3.19) and (3.4) this is equivalent to

$$\hat{\lambda}_{n,m} = \hat{\lambda}_n^1 + \sum_{k=0}^{m-1} w_k \Delta(\hat{\lambda}_n^{k+1}). \quad (3.32)$$

Subtracting (3.30) from (3.32) we have

$$0 = \Delta(\hat{\lambda}_n^0) + \sum_{k=0}^{m-1} w_k (\Delta(\hat{\lambda}_n^{k+1}) - \Delta(\hat{\lambda}_n^k)). \quad (3.33)$$

This can be rewritten as

$$\sum_{k=0}^{m-1} w_k \Delta^2(\hat{\lambda}_n^k) = -\Delta(\hat{\lambda}_n^0), \quad (3.34)$$

where $\Delta^2(\hat{\lambda}_n^k)$ is the difference of the correction vectors, given by

$$\Delta^2(\hat{\lambda}_n^k) \doteq \Delta(\hat{\lambda}_n^{k+1}) - \Delta(\hat{\lambda}_n^k). \quad (3.35)$$

In matrix vector notation (3.34) is expressed as

$$\mathbf{D}^2 \mathbf{w} = -\Delta(\hat{\lambda}_n^0), \quad (3.36)$$

where the matrix \mathbf{D}^2 is defined as

$$\mathbf{D}^2 \doteq \left[\Delta^2(\hat{\lambda}_n^0) \Delta^2(\hat{\lambda}_n^1) \cdots \Delta^2(\hat{\lambda}_n^{m-1}) \right], \quad (3.37)$$

and the vector of weights, \mathbf{w} , is given by

$$\mathbf{w} \doteq [w_0 \ w_1 \ \cdots \ w_{m-1}]^t. \quad (3.38)$$

A least-squares solution of (3.36) is obtained as

$$\mathbf{w} = -[\mathbf{D}^2]^\dagger \Delta(\hat{\lambda}_n^0), \quad (3.39)$$

where $[\mathbf{D}^2]^\dagger$, the Moore Penrose generalized inverse of the matrix \mathbf{D}^2 , is computed using

$$[\mathbf{D}^2]^\dagger = \left[[\mathbf{D}^2]^t \mathbf{D}^2 \right]^{-1} [\mathbf{D}^2]^t. \quad (3.40)$$

Once the weights w_k , $k = 0, \dots, m-1$, are obtained from (3.39), the extrapolated estimate $\hat{\lambda}_{n,m}$ is computed using (3.30). In vector matrix notation, (3.30) can be rewritten as

$$\hat{\lambda}_{n,m} = \hat{\lambda}_n^0 + \mathbf{D} \mathbf{w}. \quad (3.41)$$

Substituting for \mathbf{w} from (3.39)

$$\hat{\lambda}_{n,m} = \hat{\lambda}_n^0 - \mathbf{D} [\mathbf{D}^2]^\dagger \Delta(\hat{\lambda}_n^0). \quad (3.42)$$

In the next cycle $\hat{\lambda}_{n,m}$ is used as the starting estimate by taking $\hat{\lambda}_{n+1}^0 = \hat{\lambda}_{n,m}$. The cyclic iteration is continued until acceptable convergence is obtained.

Equation (3.42) gives the reduced rank extrapolated estimate $\hat{\lambda}_{n,m}$. In this case the matrices \mathbf{D} and \mathbf{D}^2 , each consisting of m column vectors, have ranks much smaller than the rank of the operator \mathbf{B}_n . If the number of intermediate estimates is taken equal to the dimension of the solution space, then \mathbf{D} and \mathbf{D}^2 are square matrices and an expression for the full rank extrapolation can be found as follows. If $(\mathbf{B}_n, -\mathbf{I})$ is invertible, then from (3.19)

$$\hat{\lambda}_{n,m} = -(\mathbf{B}_n - \mathbf{I})^{-1} \mathbf{b}_n. \quad (3.43)$$

From the definitions of $\mathbf{A}(\hat{\lambda}_n^k)$ and $\Delta^2(\hat{\lambda}_n^k)$ we have

$$\mathbf{A}(\hat{\lambda}_n^0) = (\mathbf{B}_n, -\mathbf{I}) \hat{\lambda}_n^0 + \mathbf{b}_n, \quad (3.44)$$

and

$$\Delta^2(\hat{\lambda}_n^k) = (\mathbf{B}_n, -\mathbf{I}) \Delta(\hat{\lambda}_n^k). \quad (3.45)$$

Substituting for \mathbf{b}_n from (3.44) into (3.43)

$$\hat{\lambda}_{n,m} = \hat{\lambda}_n^0 - (\mathbf{B}_n, -\mathbf{I})^{-1} \Delta(\hat{\lambda}_n^0). \quad (3.46)$$

From (3.45) the matrix D^2 is given by

$$D^2 = (B - I)D. \quad (3.47)$$

From (3.46) and (3.47), $\hat{\lambda}_{n,m}$ for full rank extrapolation is obtained as

$$\hat{\lambda}_{n,m} = \hat{\lambda}_n^0 - D[D^2]^{-1} \Delta(\hat{\lambda}_n^0). \quad (3.48)$$

This equation for full rank extrapolation compares with the reduced rank extrapolation equation, (3.42) in which the inverse of D^2 is replaced by its generalized inverse.

If the gradient-based algorithm generating the intermediate estimates has the property of self-normalization, then

$$\sum [\Delta(\hat{\lambda}_n^k)]_i = 0, \quad k = 0, \dots, m. \quad (3.49)$$

In this case the RREML algorithm also has the desirable property of self-normalization. This can be shown as follows:

$$\begin{aligned} \sum_i [\hat{\lambda}_{n,m}]_i &= \sum_i [\hat{\lambda}_n^0]_i + \sum_i \sum_{k=0}^{m-1} w_k [\Delta(\hat{\lambda}_n^k)]_i \\ &= \sum_i [\hat{\lambda}_n^0]_i + \sum_{k=0}^{m-1} w_k \sum_i [\Delta(\hat{\lambda}_n^k)]_i \\ &= \sum_i [\hat{\lambda}_n^0]_i \\ &= \sum_j y_j. \end{aligned} \quad (3.50)$$

As in the case of MPEML algorithms, the positivity of the extrapolated image cannot be guaranteed. Where pixel values become negative, they are offset to a small positive value and the iteration is continued. The error introduced by the local linearization assumption is alleviated by cyclic iteration. The RREML algorithm is given as follows.

Algorithm: RREML

Initialization:

- $\hat{\lambda}_n^0$ = strictly positive initial estimate normalized to have total activity equal to that of the measurements.
- m = a small positive integer, the number of intermediate estimates used in each cycle for extrapolation.
- $n = 0$, the cycle index.

RREML:

- do
- {
- generate $\hat{\lambda}_n^k, k = 1, \dots, m + 1$, using a gradient-based algorithm.
- find $\mathbf{A}(\hat{\lambda}_n^k) = \hat{\lambda}_n^{k+1} - \hat{\lambda}_n^k, k = 0, \dots, m$.
- find $\Delta^2(\hat{\lambda}_n^k) = \Delta(\hat{\lambda}_n^{k+1}) - \Delta(\hat{\lambda}_n^k), k = 0, \dots, m - 1$.
- find weights $\mathbf{w} = [w_0, w_1, \dots, w_{m-1}]$ from

$$\mathbf{w} = -[D^2]^\dagger \Delta(\hat{\lambda}_n^0)$$

where

$$D^2 \doteq [\Delta^2(\hat{\lambda}_n^0) \Delta^2(\hat{\lambda}_n^1) \dots \Delta^2(\hat{\lambda}_n^{m-1})]$$

and $[D^2]^\dagger$, the Moore–Penrose generalized inverse of D^2 , is computed as

$$D^{2\dagger} = [[D^2]^t D^2]^{-1} [D^2]^t$$

compute \mathbf{A} the reduced rank extrapolated estimate, as

$$\hat{\lambda}_{n,m} = \hat{\lambda}_n^0 + \sum_{k=0}^{m-1} w_k \Delta(\hat{\lambda}_n^k).$$

Initialize for next cycle

$$\begin{aligned} \hat{\lambda}_{n+1}^0 &= \hat{\lambda}_{n,m} \\ n &\leftarrow n + 1 \end{aligned}$$

} until convergence.

$\hat{\lambda}_n^0$ is the RREML estimate of the emission density.

C. Implementation Issues

The additional computations required by the proposed MPEML and RREML algorithms are negligible compared with that required by a single iteration of the base algorithm. This is seen by noting that, for both the MPEML and the RREML algorithm, the excess computations required are in calculating the weights w_k and in finding the weighted average of m image or correction vectors. Finding the weights w_k involves the generalized inversion of a matrix D for MPEML and D^2 for RREML, made up of m column vectors. Since, in practice, m is a very small positive integer, the generalized inversion using (3.18) or (3.40) requires only the inversion of a small matrix of dimension $m \times m$. Further the two matrix multiplications involved are of sizes $N \times m$ each, whereas a single iteration of the EM algorithm involves multiplications of pixel detector probability matrices of huge dimensions.

Though the computational overheads of the proposed algorithms are minimal, they require extra memory space for storing $(m + 1)$ intermediate image estimates. It should be noticed that, in the MPEML algorithm, the array of correction vectors, $\Delta(\hat{\lambda}_n^k), k = 0, \dots, m - 1$, need not be stored separately, and can be computed easily from the estimates $(\hat{\lambda}_n^k), k = 0, \dots, m$. In the case of the RREML algorithm, the estimates $(\hat{\lambda}_n^k), k = 1, \dots, m + 1$, are required only for computing the vectors $\Delta(\hat{\lambda}_n^k), k = 0, \dots, m$. Hence once the $\hat{\lambda}_n^k$'s are obtained, $\Delta(\hat{\lambda}_n^k), k = 0, \dots, m$, are computed and stored in place of $(\hat{\lambda}_n^k), k = 1, \dots, m + 1$. Again in this case, the vectors $\Delta^2(\hat{\lambda}_n^k), k = 0, \dots, m - 1$, need not be stored, since they can be computed easily from the $\Delta(\hat{\lambda}_n^k)$'s. Thus for both the MPEML and the RREML algorithm, the excess memory required is for storing m vectors of sizes $N \times 1$ each. Even in this case, this extra memory required is negligible compared with that required for storing the pixel-detector probability matrix of the base MLE algorithm.

As mentioned earlier in this section, we do not actually compute the exact number of intermediate estimates which is

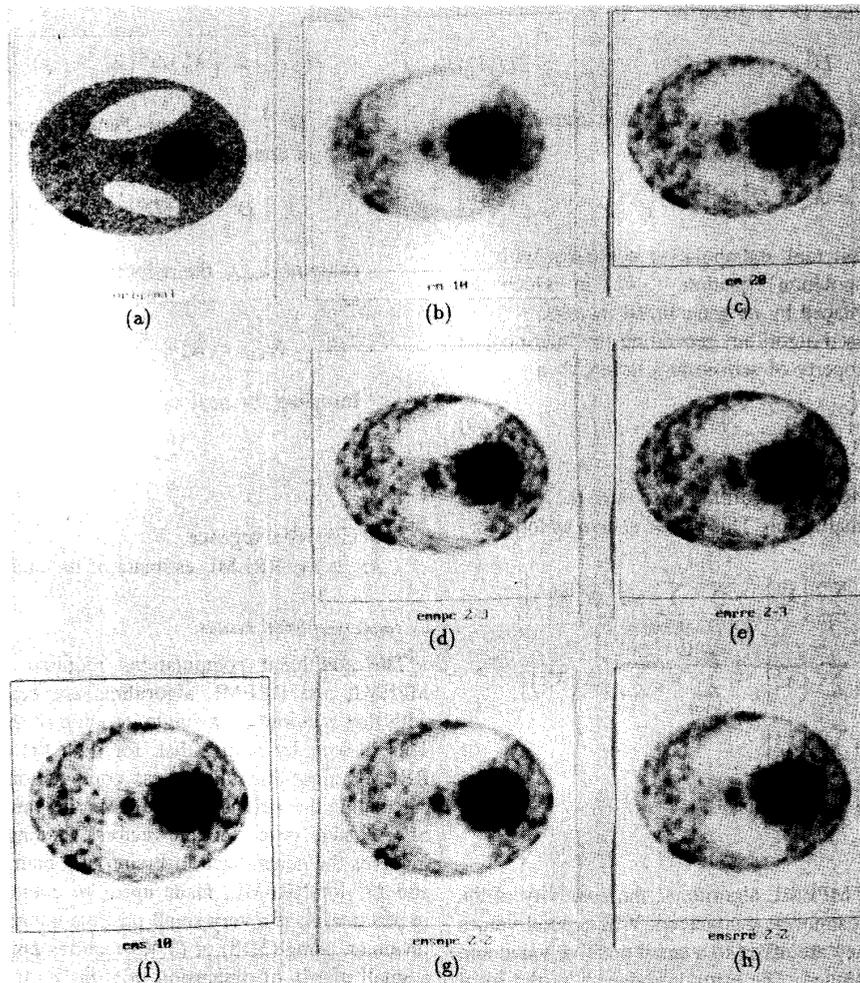


Fig. 1. Reconstructed images of the mathematical phantom. (a) Histogram of emissions in the original image. (b) **em10**: ten iterations of the EM algorithm. (c) **em20**: 20 iterations of the EM algorithm. (d) **emmpe 2-3**: three cycles of EMMPE with two EM iterations per cycle. (e) **emrrre 2-3**: three cycles of EMRRE with two EM iterations per cycle. (f) **ems10**: ten iterations of the EMS algorithm. (g) **emsmpe 2-2**: two cycles of EMSMPE with two EMS iterations per cycle. (h) **emsrre 2-2**: two cycles of EMSRRE with two EMS iterations per cycle.

optimum for the vector extrapolation in each cycle. Instead, we take the value of m to be a small positive integer. If this number is too large, it is observed that the extrapolation may result in instability, mainly because of the unconditioning of the matrices D and D^2 , in the cases of MPEML and RREML, respectively. A practical strategy for implementing these algorithms is to start with a very small positive integer value for m in the first cycle, and step up this value in the subsequent cycles until the change in residual does not show signs of instability.

IV. COMPUTER SIMULATION EXPERIMENTS

The proposed MPEML and RREML algorithms were implemented in a simulated positron emission tomography (PET) system and their performance was evaluated. The PET system used consisted of 128 detectors in a ring geometry and a 128×128 square pixel decomposition of the object space. The object was assumed to be confined within the circle inscribed

in the square region. The pixel-detector probabilities, p_{ij} , were calculated as the angle of view in the j th detector pair from the center of the pixel i . Exploiting the eightfold symmetries in the system, these probabilities were precomputed and stored as arrays. Since we were interested only in investigating the efficacy of vector extrapolation techniques for accelerating the convergence of gradient-based MLE algorithms for emission density estimation, in our simulation studies we did not consider the effects of such factors as photon attenuation and scatter.

For the simulation study two sets of measurement data were collected using two different phantoms. In the first case a mathematical phantom made up of eight elliptical objects was used and the second one used was a Hoffman brain phantom with gray matter, white matter, and cerebro spinal fluid (CSF) as distinct regions, with activities 8, 2, and 1, respectively. These values approximately represent the metabolic rate of glucose for these three regions [14]. In both cases one million

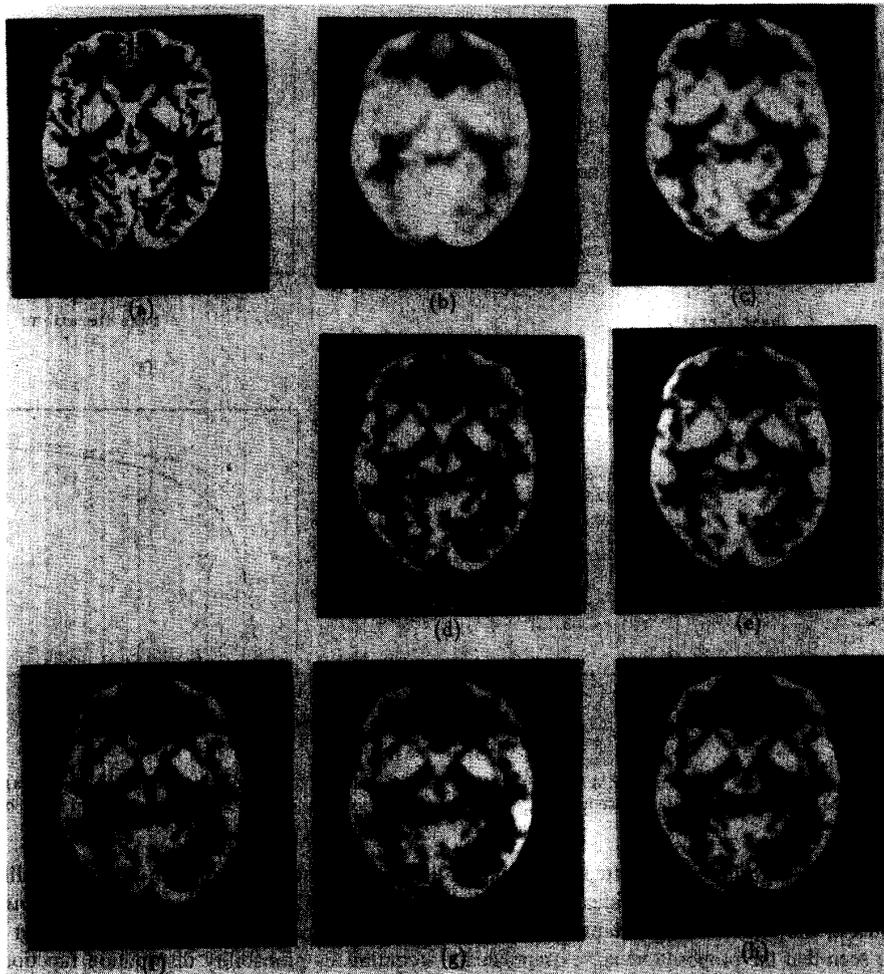


Fig. 2. Reconstructed images of the Hoffman brain phantom. (a) Histogram of emissions in the original image. (b) **em10**: ten iterations of the EM algorithm. (c) **em20**: 20 iterations of the EM algorithm. (d) **emmpe 2-3**: three cycles of EMMPE with two EM iterations per cycle. (e) **emrre 2-3**: three cycles of EMRRE with two EM iterations per cycle. (f) **ems10**: ten iterations of the EMS algorithm. (g) **emsmpe 2-2**: two cycles of EMSMPE with two EMS iterations per cycle. (h) **emsrre 2-2**: two cycles of EMSRRE with two EMS iterations per cycle.

coincidence counts were detected in 4160 detector pairs, and these were used as the measurement data for the reconstruction algorithms. The corresponding histograms of the emissions in the object are shown in Fig. 1(a) for the mathematical phantom, and in Fig. 2(a) for the Hoffman brain phantom.

The MPEML and RREML algorithms were implemented using both EM and EMS algorithms in the base iterations. With EM in the base iterations, the MPEML and RREML algorithms are designated as EMMPE and EMRRE, respectively. The corresponding names with the EMS algorithm in the base iterations are EMSMPE and EMSRRE, respectively. A number of computer experiments were carried out using these algorithms with different numbers of base iterations per cycle. Some of the estimated images for the case of two base iterations per cycle are shown in Fig. 1 for mathematical phantom and in Fig. 2 for Hoffman brain phantom. The reconstructed images using both EM and EMS algorithms are also given in these figures for comparison. Visually, it can be seen that the

EMMPE and EMRRE algorithms with three cycles of two EM iterations per cycle produced qualitatively much better images than those produced by ten iterations of the EM algorithm and are comparable to those generated by 20 EM iterations. Similarly, the EMSMPE and EMSRRE algorithms with two cycles of two EMS iterations produced qualitatively better images than those produced by ten EMS iterations.

The performance of the proposed new algorithms were evaluated using two different image-based quantitative criteria. The first one was the residual error, which measures the deviation of the pseudomeasurements, generated from the reconstructed image, from the actual measurement data. The residual error $e(\lambda)$, for the image λ is calculated as

$$e(\lambda) = \sum_j \left(y_j - \sum_i \lambda_i p_{ij} \right)^2. \quad (4.1)$$

In Fig. 3(a) the residual errors of images estimated using the

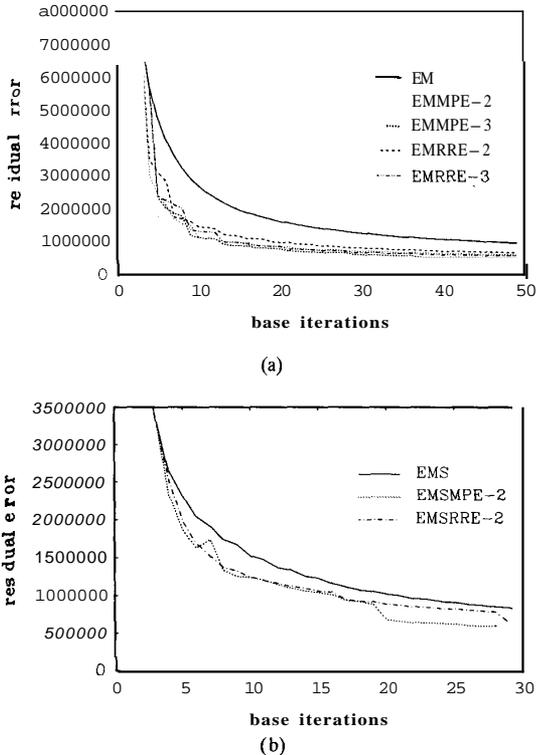


Fig. 3. Plots of residual error versus base iterations. (a) EM algorithm in the base iteration. (b) EM search algorithm in the base iteration.

EM, EMMPE, and EMRRE algorithms are plotted against base iterations. Similarly, Fig. 3(b) shows the residual error plots for the EMS, EMSMPE, and EMSRRE algorithms. From these two plots it can be seen that the proposed vector-extrapolated algorithms require only fewer base iterations for obtaining an image with a desired low residual error.

Since all these algorithms iteratively compute the maximum likelihood estimate of the emission densities, an appropriate measure for quantitative evaluation of these algorithms is the likelihood function. For an estimated image we calculate the loglikelihood function $L(\hat{\lambda})$ as

$$L(\hat{\lambda}) = \sum_{j=1}^M \left\{ - \sum_{i=1}^N \hat{\lambda}_i p_{ij} + y_j \cdot \log \left(\sum_{i=1}^N \hat{\lambda}_i p_{ij} \right) - \log(y_j!) \right\}. \quad (4.2)$$

In Fig. 4(a) the loglikelihood values of images obtained using the EM, EMMPE, and EMRRE algorithms are plotted against base iterations. Fig. 4(b) shows similar plots for the EMS, EMSMPE, and EMSRRE algorithms. The potential of the proposed vector extrapolation approach in conjunction with gradient-based MLE algorithms is clearly evident from these plots.

As also observed by other researchers [29], [7], [12], the image of the ML estimate turns increasingly noisy as the

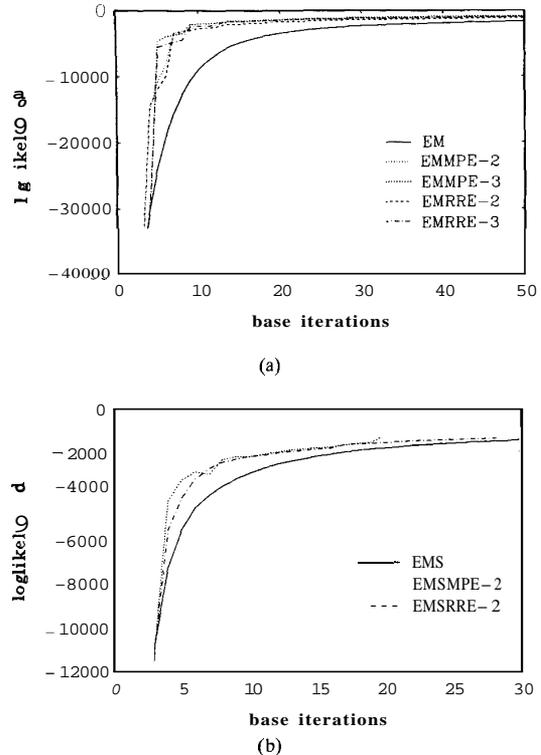


Fig. 4. Plots of loglikelihood versus base iterations. (a) EM algorithm in the base iteration. (b) EM search algorithm in the base iteration.

iterations continue to move up the “likelihood hill.” This has been attributed to the *dimensional instability* problem common to maximum likelihood estimation of parameters of a continuous probability distribution function on the basis of a finite set of measurement data [29]. Two approaches have been suggested for dealing with this problem. In the first, the iterations are stopped prematurely before the image begins to turn noisy [32], [8]. The second approach is to incorporate smoothness criteria in the reconstruction process [7], [6], [12]. We have taken the former approach and the iterations are stopped using a statistical criterion suggested by Veklerov and Llacer [32]. In Fig. 5(a) the hypothesis test value (H value in [32]) is plotted against base EM iterations. Fig. 5 (b) shows the corresponding plots with the EMS algorithm in the base iteration. The EM algorithm took about 35 iterations to produce the optimal image based on this stopping criterion. Images of similar quality were obtained with three cycles of EMMPE-2 (EMMPE with two iterations per cycle) and EMRRE-2 and with two cycles of EMMPE-3 and EMRRE-3. With the EMS algorithm in the base iteration, both EMSMPE-2 and EMSRRE-2 produced similar images with two cycles each.

V. CONCLUSION AND DISCUSSION

In this paper we have presented a new approach for fast maximum likelihood estimation of emission densities in ECT. This was based on integrating vector extrapolation techniques

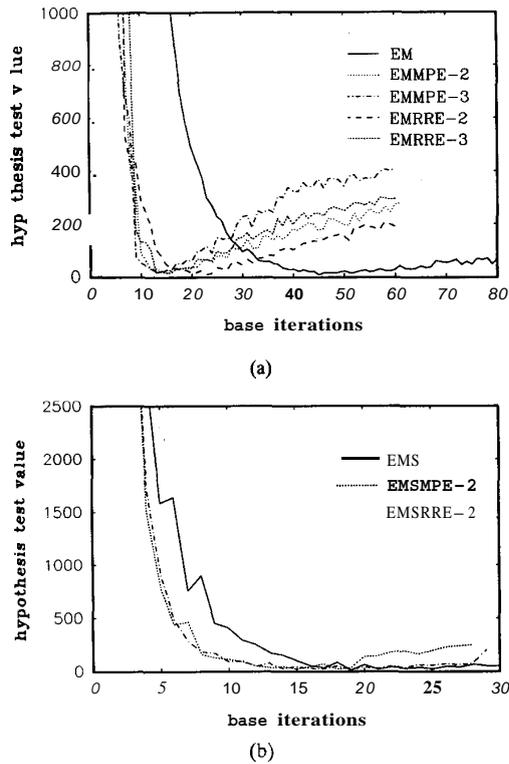


Fig. 5. Plots of hypothesis test value versus base iterations. (a) EM algorithm in the base iteration. (b) EM search algorithm in the base iteration.

with gradient-based MLE algorithms. Based on this approach, a class of fast cyclic iterative estimation algorithms for emission tomography was developed.

After discussing various gradient MLE algorithms and the need for accelerating their convergence, we proposed the vector extrapolation technique as a practical convergence accelerator. The mathematical theory behind the minimal polynomial and reduced rank extrapolation techniques in the context of emission tomography was presented in detail. To further substantiate the potential of the new approach in the fast maximum likelihood estimation of emission densities, we implemented the above vector extrapolation methods, in conjunction with the EM and EM search algorithms, on a PET system. The extensive computer-simulated experiments carried out proved our proposition. It was also shown that the proposed approach results in a substantial reduction in the number of base iterations required for obtaining an image of emission density estimate satisfying a given specification. It is to be noted that the proposed vector extrapolation techniques require only minimal additional computation and the resulting estimate has the desirable property of preserving total activity in the image equal to that in the measurements.

One of the issues which require further study has to do with the stability properties of the proposed cyclic iterative algorithms. The effect of local linearization of the nonlinear operator A_n , discussed in Section III, on the stability of the algorithms needs further analytical study. A discussion

on these issues is given in [18]. It should be noted that the MPEML algorithm uses first-order difference vectors for computing the weights for extrapolation, whereas the RREML algorithm uses second-order difference vectors. In this sense the MPEML algorithm might be more stable than the RREML algorithm. The stability is also influenced by the number of intermediate estimates used for vector extrapolation in each cycle of the algorithm. We have not given a method for computing the optimum value for this number. However, even with the small number of intermediate estimates used for extrapolation in each cycle, our simulation experiments have given very encouraging results.

ACKNOWLEDGMENT

The first author wishes to express his thanks to Dr. S.K. Sinha and Dr. N. Sreenivasa of the Department of Electrical Engineering, Indian Institute of Science, Bangalore, India, for many useful discussions.

REFERENCES

- [1] A. Albert, *Regression and the Moore-Penrose Pseudoinverse*. New York: Academic, 1972.
- [2] S. Cabay and L.W. Jackson, "A polynomial extrapolation method for finding limits and antilimits of vector sequences," *SIAM J. Numer. Anal.*, vol. 13, pp. 734-752, 1976.
- [3] E. S. Chernoboy, C.J. Chen, M. I. Miller, T. R. Miller, and D. L. Snyder, "An evaluation of maximum likelihood reconstruction for SPECT," *IEEE Trans. Med. Imaging*, vol. 9, pp. 99-110, 1990.
- [4] A. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc.: Series B*, vol. 39, pp. 1-38, 1977.
- [5] R.P. Eddy, "Extrapolating to the limit of a vector sequence," in *Information Linkage Between Applied Mathematics and Industry*, P. C. C. Wang, Ed. New York: Academic, 1979.
- [6] P.J. Green, "Bayesian reconstruction from emission tomography data using a modified EM algorithm," *IEEE Trans. Med. Imaging*, vol. 9, pp. 84-93, Mar. 1990.
- [7] T. Hebert and R. Leahy, "A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors," *IEEE Trans. Med. Imaging*, vol. MI-8, pp. 194-202, June 1989.
- [8] T.J. Hebert, R. Leahy, and M. Singh, "Fast MLE for SPECT using an intermediate polar representation and a stopping criterion," *IEEE Trans. Nucl. Sci.*, vol. 34, pp. 615-619, 1988.
- [9] L. Kaufman, "Implementing and accelerating the EM algorithm for positron emission tomography," *IEEE Trans. Med. Imaging*, vol. MI-6, pp. 37-51, Mar. 1987.
- [10] K. Lange, M. Bahn, and R. Little, "A theoretical study of some maximum likelihood algorithm for emission and transmission tomography," *IEEE Trans. Med. Imaging*, vol. MI-6, pp. 106-114, June 1987.
- [11] K. Lange and R. Carson, "EM reconstruction algorithm for emission and transmission tomography," *J. Comput. Assist. Tomogr.*, vol. 8, pp. 306-316, 1984.
- [12] E. Levitan and G.T. Herman, "A maximum a posteriori probability expectation maximization algorithm for image reconstruction in emission tomography," *IEEE Trans. Med. Imaging*, vol. MI-6, pp. 184-192, Sept. 1987.
- [13] R. M. Lewitt and G. Muehllehner, "Accelerated iterative reconstruction for positron emission tomography based on the EM algorithm for maximum likelihood estimation," *IEEE Trans. Med. Imaging*, vol. MI-5, pp. 16-22, Mar. 1986.
- [14] D. K. Mahony *et al.*, "A realistic computer simulated brain phantom for evaluation of PET characteristics," *IEEE Trans. Med. Imaging*, vol. MI-6, pp. 250-257, Sept. 1987.
- [15] I. Meilijson, "A fast improvement to the EM algorithm on its own terms," *J. Roy. Statist. Soc.: Series B*, vol. 51, no. 1, pp. 127-138, 1989.
- [16] M. Mesina, "Convergence acceleration for the iterative solution of the equation $X = AX + f$," *Comp. Math. Appl. Mech. Eng.*, vol. 10, pp. 165-173, 1977.

- [17] M.I. Miller, D.L. Snyder, and T.R. Miller, "Maximum likelihood reconstruction for single photon emission computed tomography," *IEEE Trans. Nucl. Sci.*, vol. NS-23, pp. 769–778, 1985.
- [18] N. Rajeevan, "A stochastic estimation approach to emission tomography," Ph.D. thesis, Dept. of Electrical Engineering, Indian Institute of Science, Bangalore, India, 1991.
- [19] N. Rajeevan, "Vector extrapolated fast ML algorithms for emission tomography," in *Proc. 12th Ann. Int. Conf. IEEE EMBS Soc.* (Philadelphia), 1990, pp. 370–371.
- [20] M.V. Ranganath, A.P. Dhawan, and N. Mullani, "A multigrid expectation maximization reconstruction algorithm for positron emission tomography," *IEEE Trans. Med. Imaging*, vol. 7, pp. 273–278, 1988.
- [21] A.J. Rockmore and A. Macovski, "A maximum likelihood approach to emission image reconstruction from projection," *IEEE Trans. Nucl. Sci.*, vol. NS-23, pp. 1428–1432, 1976.
- [22] L.A. Shepp and Y. Vardi, "Maximum likelihood estimation for emission tomography," *IEEE Trans. Med. Imaging*, vol. MI-1, pp. 113–121, 1982.
- [23] L.A. Shepp, Y. Vardi, and L. Kaufman, "A statistical model for positron emission tomography," *J. Amer. Statist. Ass.*, vol. 80, pp. 8–37, 1985.
- [24] L.A. Shepp, Y. Vardi, J.B. Ra, S.K. Hilal, and Z.H. Cho, "Maximum likelihood PET with real data," *IEEE Trans. Nucl. Sci.*, vol. NS-31, pp. 910–913, 1984.
- [25] A. Sidi, "Convergence and stability properties of minimal polynomial and reduced rank extrapolation algorithms," *SIAM J. Numer. Anal.*, vol. 23, no. 1, pp. 197–209, 1986.
- [26] A. Sidi, W.F. Ford, and D.A. Smith, "Acceleration of convergence of vector sequences," *SIAM J. Numer. Anal.*, vol. 23, no. 1, pp. 178–196, 1986.
- [27] S. Skelboe, "Computation of the periodic steady state response of nonlinear network by extrapolation methods," *IEEE Trans. Circuits Syst.*, vol. CAS-27, pp. 161–175, 1980.
- [28] D.A. Smith, W.F. Ford, and A. Sidi, "Extrapolation methods for vector sequences," *SIAM Rev.*, vol. 29, no. 2, pp. 199–233, 1987.
- [29] D.L. Snyder and M.I. Miller, "The use of sieves to stabilize images produced with the EM algorithm for emission tomography," *IEEE Trans. Nucl. Sci.*, vol. NS-32, no. 5, pp. 3864–3872, 1985.
- [30] D.L. Snyder and D.G. Polite, "Image reconstruction from list mode data in an emission tomography system having time-of-flight measurements," *IEEE Trans. Nucl. Sci.*, vol. NS-30, pp. 1843–1949, 1983.
- [31] E. Tanaka, "A fast reconstruction algorithm for stationary positron emission tomography based on a modified EM algorithm," *IEEE Trans. Med. Imaging*, vol. MI-31, pp. 98–105, 1987.
- [32] E. Veklerov and J. Llacer, "Stopping rule for the MLE algorithm based on statistical hypothesis testing," *IEEE Trans. Med. Imaging*, vol. MI-6, pp. 313–319, 1987.