# A multi-scale stopping criterion for MLEM reconstructions in PET

Nicolai Bissantz*‡, Bernard A. Mair† and Axel Munk‡

*Department of Mathematics
Ruhr-University of Bochum, Germany
Email: nicolai.bissantz@rub.de

†Department of Mathematics
University of Florida, Gainesville, USA.

‡Institute of Mathematical Stochastics
University of Göttingen, Germany

*Abstract*— **In this paper we propose and test a new method for terminating the maximum likelihood expectation maximization algorithm for reconstructing positron emission tomography images. The method is based on a stochastic multiresolution analysis which involves all partial sums (scales) of normalized differences between the projected images and the detector data for each row of the sinogram. Previous methods involved only the single total sum of these differences for all detectors. Our method is theoretically founded on recent results from probability theory on the almost sure behaviour of the maximum of the partial sum process for Poisson data. Preliminary tests indicate that this method produces predictions for the optimal stopping iterations which are robust relative to modeling errors in the system matrix and has a signal-to-noise ratio which is $80\%$ of the maximal SNR available from the MLEM iterates.**

## I. INTRODUCTION

It is well–known that if the MLEM (maximum likelihood expectation maximization) algorithm is initialized with a uniform image, the iterates initially improve, then after a certain point gradually deteriorate (become more noisy and lose resolution) in appearance and accuracy [1]. As a result, there was an effort in the last two decades to develop suitable stopping criteria for the MLEM algorithm [2], [3], [4], [5], [6]. On the whole, these methods worked well for the case in which the projection matrix used in the MLEM algorithm exactly modeled the scanner system response function, but had serious problems in the more realistic case of an approximate model. In [5] a method is described that seeks to obtain a set of "feasible images" from the MLEM algorithm. However, this method depends on a suitable choice for a parameter that measures the model inaccuracies and does not provide a finite set of feasible images. It was also found that early termination of the MLEM algorithm produced images with low contrast, which represented the coarse scale image features, but did not do as well with fine scale structures. This situation led to the rapid development of a large body of research on penalized ML, or, maximum a posteriori (MAP) reconstruction algorithms. Although these algorithms are also iterative, there

is no need to terminate early, and one seeks to obtain the actual limiting image. However, these methods rely on choosing appropriate penalty functions, and regularization parameters. There is significant interest in developing data–based methods for determining the regularization parameter.

Despite the extensive development of MAP methods, the only iterative algorithm that is currently in commercial use is the OSEM (ordered subsets EM) algorithm [7], [8], which is a fast version of the basic MLEM algorithm obtained by iterative applications to subsets of the detector data. Thus, in our estimation, the determination of suitable stopping criteria for the MLEM algorithm remains an area of interest.

In this paper, we adopt the usual probabilistic model and consider PET data which are observations of the Poisson random variables

$$Y_i \sim \text{Poiss}([A\mathbf{x}]_i), \quad i = 1, \ldots, m, \qquad (1)$$

where $A$ is the projection matrix representing the scanner system response function, $\mathbf{x}$ is the $n-$ dimensional vector of emission intensities, and $[A\mathbf{x}]_i$ is the $i^{\text{th}}$ entry of the vector $A\mathbf{x}$. That is, $[A\mathbf{x}]_i$ is the mean number of detections in the $i^{\text{th}}$ detector tube. Of the methods developed in [2], [3], [4], [5], [6], due to space limitations, we will only discuss those in [5], [6] as they are the ones most relevant to this paper.

For each iterate $\hat{\mathbf{x}}$, the methods in [5], [6] are based on testing the null hypothesis that the data $\{Y_i, \ i = 1, 2, \ldots, m\}$ are Poisson random variables with means $\{[A\hat{\mathbf{x}}]_i, \ i = 1, 2, \ldots, m\}$. By using a certain transformation, the null hypothesis results in random variables $\{Z_i, \ i = 1, 2, \ldots, M\}$ which are uniformly distributed on $[0, 1]$. Then, if the total number of detections is large, Pearson's test can be applied to test the uniformity assumption. The test statistic, $V$ is a scaled, normalized sum of all the transformed data which is asymptotically $\chi^2_{M-1}$. Thus, the null hypothesis was accepted at the $\alpha\%$ level of significance if the test statistic fell below the $1 - \alpha$ quantile of the $\chi^2_{M-1}$ distribution. Results in [6] showed that the method worked well when the matrix $A$ was

an exact model of the true system response function. We will refer to this as the "exact model" case. In that case, as the number of iterations increased, the corresponding test statistic decreased past the critical value, then increased to values exceeding the critical value. This resulted in a well–defined, finite set of "feasible iterates". However, it is impossible to accurately model all the factors affecting the performance of a real scanner. In this more realistic "inexact case" the proposed method in [6] failed, as observed by the authors. Sometimes the test statistic always remained above the critical value, and on other occasions when it did fall below the critical value, it remained below for all successive iterations. Thus there were either no feasible iterates or infinitely many [5]. In order to obtain a more robust stopping rule, a modification was proposed in [5] which required choosing an interval of uncertainty around each $Y_i$. Then similar (but more complex) tranformations were performed on the data to obtain uniformly distributed r.v.'s as in [6]. As a result, the same Pearson test involving a $\chi^2$ test statistic could be applied. In this case, the simulations indicated that all iterations past a certain point were feasible. The method relies on an appropriate choice for the interval of uncertainty, which depends on the total number of detections. The authors acknowledge the unsuitability of this feature as they were unable to develop a data–based method for determining its value.

In this paper we propose and test a new method for terminating the MLEM algorithm for the reconstruction of PET images. The method is based on a stochastic multiresolution analysis which involves all partial sums (scales) of normalized differences between the projected images and the detector data for each row of the sinogram. Previous stochastic methods involved only the single total sum of these differences for all detectors. Our method is theoretically founded in recent results from probability theory on the almost sure behaviour of the maximum of the partial sum process for Poisson data.

## II. THEORY

For each estimator $\hat{\mathbf{x}}$ of the true emission intensity vector $\mathbf{x}$, consider the normalized residuals

$$R_i(\hat{\mathbf{x}}) = \frac{Y_i - [A\hat{\mathbf{x}}]_i}{\sqrt{[A\hat{\mathbf{x}}]_i}}, \quad i = 1, \ldots, m. \tag{2}$$

If $\hat{\mathbf{x}}$ is a "good" approximation to $\mathbf{x}$ the residuals in (2) approximately have mean zero and variance 1. In particular, they should, in distribution, neither be "too large" (which would indicate substantial remaining signal in the residuals) nor "too small" (which would indicate overfitting).

Our proposed method is based on a modification of the test statistic

$$\mathcal{D}_m = \max_{0 \leq j < m} \max_{1 \leq k \leq m-j} \frac{|\sum_{i=j+1}^{j+k} R_i|}{k\,\alpha(k/\log(m))}, \tag{3}$$

with $R_i = R_i(\hat{\mathbf{x}})$. That is, $\mathcal{D}_m$ is the maximum magnitude of scaled partial sums of the residuals $R_i$, where the scaling function $\alpha(\cdot)$ will be determined later. Note that this multiresolution test statistic (3) involves all possible re-binnings of

the data into increasing bin sizes in the detector space, and at all locations. It therefore tests simultaneously whether the residuals are consistent with the distribution of the noise – or whether there are systematic deviations anywhere, at all bin sizes.

Fundamental to our method is the following result adapted to our setting from Steinebach [9] on the asymptotic behaviour of $\mathcal{D}_m$ as $m \to \infty$.

*Theorem 1:* If $\{R_i : i = 1, 2, \ldots, m\}$ are independent, identically distributed Poisson random variables with mean $\lambda$ (i.e. $R_i \sim \text{Poiss}(\lambda)$), and $\alpha(\cdot)$ is the "inverse Chernoff function" of the Poiss($\lambda$)) distribution, then $\lim_{m\to\infty} \mathcal{D}_m = 1$, with probability 1.

Note that $\alpha(\cdot)$ depends on the Poisson parameter, and the residuals will only be identically distributed if $\hat{\mathbf{x}}$ is a constant vector. Therefore we propose a modification $\mathcal{B}_m(\hat{\mathbf{x}})$ of the test statistic $\mathcal{D}_m$ for the case $R_i = R_i(\hat{\mathbf{x}})$.

Given the PET data $\{Y_i : i = 1, 2, \ldots, m\}$, let $\mu = \sum_{i=1}^{m} Y_i/m$, and $\alpha_{\text{mean}}$ be the inverse Chernoff function for the normalized residual $R = \frac{Y-\mu}{\sqrt{\mu}}$ where $Y \sim \text{Poiss}(\mu)$. Then, the Chernoff function for $R$ is given by $\rho(y) = \inf_{t\geq 0}\exp(-ty)M(t)$, where $M(t)$ is the moment generating function for the random variable $R$. It can be shown that $\rho$ is decreasing and positive, so its inverse function, $\rho^{-1}$ exists. As a result, the inverse Chernoff function, $\alpha_{\text{mean}}(c) = \rho^{-1}(\exp(-1/c))$, is easily computed. We replace $\mathcal{D}_m$ with the following test statistic:

$$\mathcal{B}_m(\hat{\mathbf{x}}) = \max_{0 \leq j < m} \max_{1 \leq k \leq m-j} \frac{|\sum_{i=j+1}^{j+k} R_i(\hat{\mathbf{x}})|}{k\,\alpha_{\text{mean}}(k/\log(m))}. \tag{4}$$

To determine the analog of the limiting value 1 in Theorem 1, we generate a frequency histogram of values of

$$\tilde{\mathcal{B}}_m = \max_{0 \leq j < m} \max_{1 \leq k \leq m-j} \frac{|\sum_{i=j+1}^{j+k} \tilde{R}_i|}{k\,\alpha_{\text{mean}}(k/\log(m))} \tag{5}$$

where

$$\tilde{R}_i = \frac{\tilde{Y}_i - \mu}{\sqrt{\mu}}, \quad \text{and } \tilde{Y}_i \sim \text{Poiss}(\mu) \tag{6}$$

and determine its median $\nu$.

We propose stopping the MLEM algorithm at the first iterate $\hat{\mathbf{x}}$ for which $\mathcal{B}_m(\hat{\mathbf{x}}) \leq \nu$. The idea is that if the test statistic, $\mathcal{B}_m(\hat{\mathbf{x}})$, computed from the residuals (2), is larger than $\nu$ the projected means are not consistent with the random data, and if it is smaller than this critical value the projected means overfit the data (i.e. undersmooth the image).

## III. EXPERIMENTAL METHODS

In this section we apply the proposed multiresolution method to stopping the MLEM reconstruction of PET (positron emission tomography) images. We compare it with a previous algorithm developed by Llacer and Veklerov [5], [6], and with the oracle–type method which chooses the iterate with maximum signal-to-noise ratio (SNR). Of course, this latter method cannot be applied in practice.

As usual we consider the detector data in sinogram format, which in the Siemens ECAT Exact scanner, is a $192 \times 160$ matrix, containing data in 192 angles (or "views") and 160 bins in each view. First, we only consider partial sums within each view. That is, the partial sums do not contain residuals from more than one view. Then, to avoid very small means, we consider "pooled residuals". In our simulations, it is sufficient to pool the original residuals in 8 adjacent bins for each view, before forming the partial sums. Of course, this requires a further normalization (dividing by $\sqrt{8}$) for the pooled residuals (2).

In [5], [6], an empirical histogram $\mathcal{H}(\hat{\mathbf{x}})$ giving the number of detectors $j$ for which the observed count rate $Y_j$ is in a certain bin of a slightly modified Poisson distribution function with rate $[A\hat{\mathbf{x}}]_i$ is first computed. If the null hypothesis holds, i.e. the observations are Poisson distributed with the respective rate predicted by the projected model for every detector, then the distribution underlying the histogram is uniform. In the second step, a Pearson $\chi^2$-test of $\mathcal{H}(\hat{\mathbf{x}})$ for uniformity is applied. This method, proposed in [6] produced reasonable results in the exact case, when there was no modeling error in the system matrix $A$. However, in the inexact case, where there was modeling error, the method sometimes did not produce any "feasible" images [5]. Therefore, they proposed an improved version of their selection criterion, which we will call the "LV" method. The improved method tests the relaxed hypothesis that there exists some model among the set of models with Poisson parameters in the intervals $[[A\hat{\mathbf{x}}]_i \cdot (1 - \epsilon), [A\hat{\mathbf{x}}]_i \cdot (1 + \epsilon)]$, for which the hypothesis that the observations are Poisson distributed with rates predicted by the projected model cannot be discarded. All such iterates are called "feasible".

We performed a study with 20 simulations of noisy data generated by a $128 \times 128$ slice of the Hoffmann phantom with 1 million total counts. To apply our method, we estimated the median $\nu$ of the distribution of $\tilde{\mathcal{B}}_m$ in (5) from 100 simulations (under the hypothesis that the model holds). Application of the LV method requires the user to determine the number of bins $N_{\text{classes}}$ in the histogram $\mathcal{H}(\hat{\mathbf{x}})$, and the uncertainty parameter $\epsilon$. We used the values $N_{\text{classes}} = 50, 500$ and $\epsilon = 0.001, 0.01$ and $\epsilon = 0.1$. The feasible iterates for the LV method were those iterates for which the hypothesis that the observations follow a Poisson distribution with rates predicted by the projected means cannot be rejected at the 95% level.

We performed tests for both the exact and inexact case. In the first test, we used the same matrix $A$ to generate the data and to reconstruct the images by the MLEM algorithm. For the inexact case we generated the data using an approximate matrix $\tilde{A}$, obtained from uniformly distributed detector gains between $\pm 5\%$ in $A$.

## IV. RESULTS AND CONCLUSIONS

Our simulations of the values of the test statistic $\mathcal{B}$ resulted in a small variance of $0.006$ and the median value $\nu = 0.95$. Additional testing with different numbers of total counts produced only a relatively small variation in the value of the
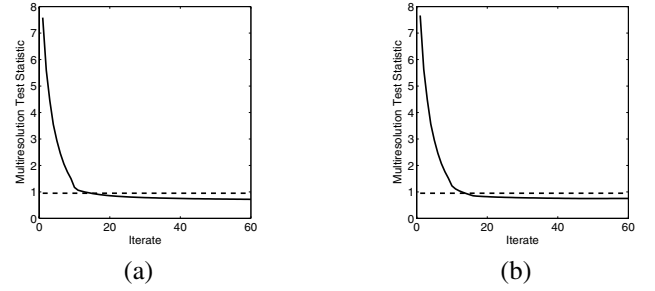


Fig. 1. Multiresolution test statistic for one realization in (a) exact case and (b) inexact case. The iterate corresponding to the dashed line is the multiresolution stopping iterate.
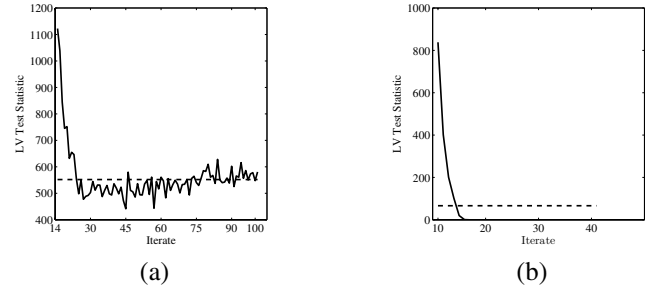


Fig. 2. LV test statistic for one realization in (a) exact case with , $N_{classes} = 50$, and (b) inexact case, with $N_{classes} = 50$, $\epsilon = 0.01$. The feasible iterates for the 5% level of significance correspond to values below the dashed line.

median. So, the method is stable relative to variations in the total counts.

Fig. 1 contains plots of the values of the multiresolution test statistic $\mathcal{B}_m(\mathbf{x}^{(k)})$ for the $k^{\text{th}}$ iteration $\mathbf{x}^{(k)}$ in the (a) exact case, and (b) inexact case. The broken line represents the crtical value, $\nu$. So, the first iterate that falls below the line is the predicted multiresolution stopping iterate. Although this plot is for a single realization, randomly chosen among from those that had the same stopping iterate as the mean, it is typical of all realizations. Fig. 2 contains plots of the LV test statistic $\mathcal{H}(\mathbf{x}^{(k)})$ (as $k$ varies) for the $k^{\text{th}}$ iteration $\mathbf{x}^{(k)}$ in the (a) exact case, with $N_{classes} = 500$, and (b) inexact case, with $N_{classes} = 50$, $\epsilon = 0.01$. The broken line represents the crtical value for the $\chi^2$ distribution. So, all iterates that fall below this value are feasible. This plot is for a single, randomly chosen realization. Observe, in Fig. 2, the LV mehod produced infinitely many feasible iterates in both cases. This behavior occurred in all our simulations. It is important to note that in every simulation, the LV method failed to reject very late iterates which were clearly too speckled to be considered reasonable reconstructions. In contrast, the proposed multiresolution stopping criterion always produced a definite stopping iterate, with the property that images obtained from iterations preceeding this value were not consistent with the random data, and images obtained from succeeding iterations overfitted the data, resulting in undersmoothed estimates.

We now compare our method with the oracle of choosing the maximal SNR. In the exact case the multiresolution and maximal SNR stopping iterates had means of 15, 47, with
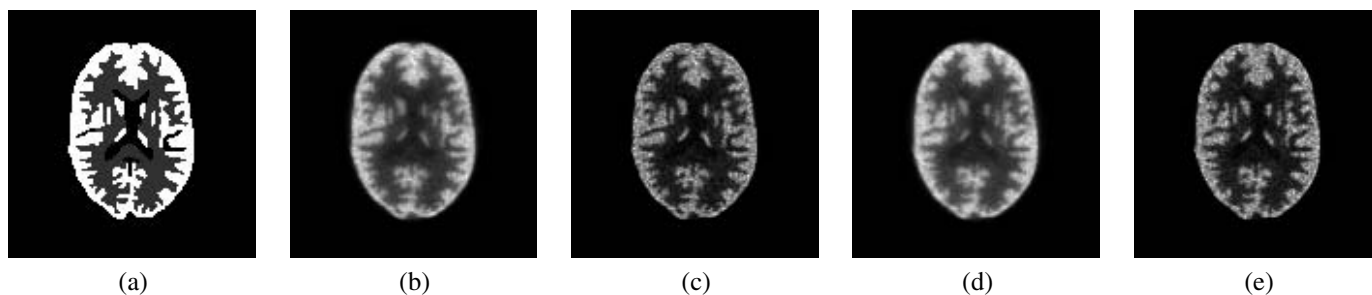
Fig. 3. (a) True phantom; (b) Multiresolution MLEM iterate: exact case; (c) Multiresolution MLEM iterate: inexact case; (d) MLEM iterate with maximum SNR: exact case; (e) MLEM iterate with maximum SNR: inexact case

standard deviations of 2.7, 1.3 respectively. In the inexact case the multiresolution and maximal SNR stopping iterates had means of 14, 47, with standard deviations of 2.7, 1.1 respectively. The small variation between the exact and inexact results were most likely due to the relatively small model error in the system matrix. We expect more significant variation when the method is applied to real data. Although the multiresolution iterate was obtained in one-third the time taken to achieve the maximal SNR, the SNR of the multiresolution iterate was always approx $80\%$ of the maximal SNR.

Fig. 3 shows the true phantom, and reconstructions obtained by stopping the MLEM algorithm using the proposed method and at the optimal SNR, in the exact and inexact cases, for a randomly chosen realization. Visually, the main features of the image are recovered reasonably well (compared to the maximal SNR images) by the iterates predicted from the multiresolution criterion.

In conclusion, we have proposed a completely data–based method for stopping the MLEM algorithm. Our stopping criterion depends on the observed data, the forward projections of the reconstructed image at each iteration and on a pre-processing step to determine a critical model parameter. This parameter is the median of a simulated test statistic, which only depends on the total number of observations. This dependence is very stable, thus a few simulated values will serve a wide range of applications. The method has been demonstrated to be very effective in determining suitable stopping iterates for reconstructing PET images with the MLEM algorithm applied to a uniform initialization. The method is also very stable in the presence of modeling errors in the system

matrix and adds only a modest computational burden to the basic MLEM algorithm. In our preliminary tests, this method produced images that have a signal-to-noise ratio (SNR) which is $80\%$ of the maximal SNR available from the MLEM iterates, and in one–third the time taken to achieve maximal SNR.

### REFERENCES

[1] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. Med. Imag.*, vol. 1, pp. 113–122, 1982.
[2] K. J. Coakley, "A cross-validation procedure for stopping the em algorithm and deconvolution of neutron depth profiling spectra," *IEEE Trans. Nucl. Sci.*, vol. 38, pp. 9–15, 1991.
[3] T. Hebert, R. Leahy, and M. Singh, "Fast MLE for SPECT using an intermediate polar representation and a stopping criterion," *IEEE Trans. Nucl. Sci.*, vol. 35, no. 1, pp. 615–619, 1988.
[4] G. N. Kontaxakis and G. S. Tzanakos, "Practical application of a new stopping criterion for the EM-MLE image reconstruction for PET," in *IEEE Conf. on Eng. in Med. and Biol. Soc.* IEEE, 1996, pp. 848–849.
[5] J. Llacer and E. Veklerov, "Feasible images and practical stopping rules for iterative algorithms in emission tomography," *IEEE Trans. Med. Imag.*, vol. 8, pp. 186–193, 1989.
[6] E. Veklerov and J. Llacer, "Stopping rule for the MLE algorithm based on statistical hypothesis testing," *IEEE Trans. Med. Imag.*, vol. 6, no. 4, pp. 313–319, 1987.
[7] S. Ahn and J. A. Fessler, "Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms," *IEEE Trans. Med. Imag.*, vol. 22, no. 5, pp. 613–626, 2003.
[8] H. Hudson and R. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Trans. Med. Imag.*, vol. 13, no. 4, pp. 601–609, 1994.
[9] J. Steinebach, "On a conjecture of revesz and its analogue for renewal processes," in *Asymptotic Methods in Probability and Statistics (ICAMPS 1997)*, 1997, pp. 311–322.