ANALYSIS OF THE ML-EM ALGORITHM FOR NONLINEAR RECONSTRUCTION OF POSITRON EMISSION TOMOGRAPHY IMAGES

F. Boschen, A. Kummert

Laboratory for Communication Theory
Department of Electrical Engineering
University of Wuppertal
Fuhlrottstr. 10
D - 42097 Wuppertal
Germany
Phone: +49-202-439-2961
Fax: +49-202-439-2959

H. Herzog

Institute of Medicine
Research Center Juelich KFA
D - 52425 Juelich
Germany
Phone: +49-2461-615913
Fax: +49-2461-612820

ABSTRACT

Positron emission tomography (PET) is a technique that has opened new facilities to study the metabolic activity of human body. In the last years many algorithms have been developed for reconstructing tomography images. The oftenly used maximum likelihood expectation maximization algorithm (ML-EM) seems to be a stable method and was developed by Shepp and Vardi [1] in 1982. However, the ML-EM algorithm causes some serious problems in the context of the application considered. It is an iterative procedure and converges to a stationary point, however, the reconstructed image seems to be distored by superposed high frequency noise. In this paper it is shown, that the ML-EM-Algorithm is not based on significant statistical properties in our problem, which has been verified by respective investigations. It is shown, that the algorithm converges to an 'optimal' solution in a mathematical sense. The convergence characteristics of the algorithm are discussed by means of examples.

1. INTRODUCTION

Positron emission tomography (PET) is a special form of Computer Tomography (CT) with the following properties. The most important feature of PET is the use of the emission of radioactive nuclides in contrast to illuminating the patient by X-rays in classical tomographs. The mathematical description of X-ray based CT and PET is the same. The measurements (projections) $p(l,\theta)$ and the underlying image $g(x,y)$ are

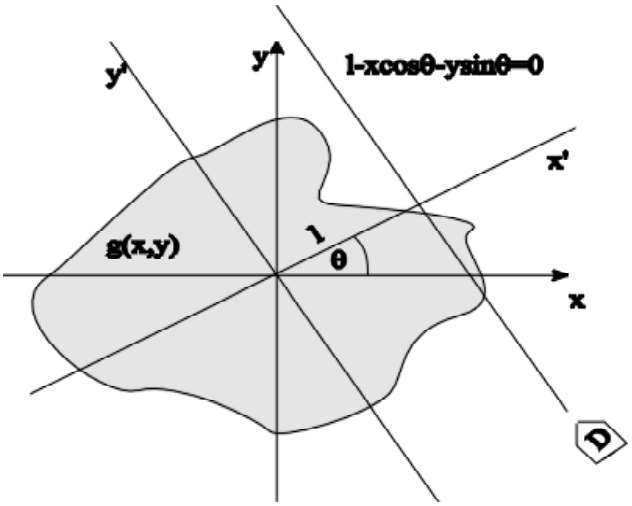

Figure 1: Principles of Tomography

related via the Radon-Transform

$$p(l,\theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\delta(l - x\cos\theta - \sin\theta)dxdy, \quad (1)$$

where l is the line coordinate for the projection under the angle θ and x, y are the cartesian coordinates of the image (Fig. 1). The measured projection is the integral of the emitted energy (PET) or the attenuation (CT), respectively, along a line. This is illustrated by Fig. 1.

Measurements are made under different angels θ. With this data it is possible to reconstruct the image of one slice (plane) of the inspected object. The analytical

solution of (1) is given by

$$g(x,y) = \frac{1}{2\pi^2} \int_0^{\pi} \int_{-\infty}^{\infty} p(l',\theta) \frac{1}{(l-l')^2} dl' d\theta, \qquad (2)$$

with

$$l = x\cos\theta + y\sin\theta. \qquad (3)$$

However, the numerical solution of the above integral is very delicate due to the pole of the kernel at $l' = l$. Hence, alternative methods for reconstruction of images had to be developed. The most familiar method is the filtered backprojection, which can be characterized as follows

$$g(x,y) = \int_0^{\pi} p'(x\cos\theta + y\sin\theta, \theta)\, d\theta, \qquad (4)$$

where $p'(l,\theta)$ is the filtered (with respect to l) version of $p(l,\theta)$, and the transfer function of the filter is described by $G(j\omega) = |\omega|$. In other words, the Fourier transforms $P'(j\omega,\theta)$ and $P(j\omega,\theta)$ of the signals $p'(l,\theta)$ and $p(l,\theta)$ are related by

$$\begin{aligned} P'(j\omega,\theta) &= P(j\omega,\theta)\, G(j\omega) \qquad (5) \\ &= P(j\omega,\theta)\, |\omega|. \end{aligned}$$

Although this method is numerical stable, the reconstruction, however, is marked with radial distortions caused by interpolation errors. Another method is based on the attempt to solve the problem by dicretisation and is known as algebraic method. More precisely, the image $g(x,y)$ is interpreted as a two-dimensional discrete signal of finite size. Its samples (pixels) can be arranged in vector form $\mathbf{g} = [g_1, ..., g_B]^T$. The projection signal is discretised in a natural way since measurements are performed by an array of equispaced sensors. Each sensor signal is interpreted as a sample of the discretised projection. These samples are summarized to a vector $\mathbf{p} = [p_1, ..., p_D]^T$ which is related to $\mathbf{g}$ via

$$\begin{bmatrix} p_1 \\ p_2 \\ \cdot \\ \cdot \\ p_D \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdot & \cdot & a_{1B} \\ a_{21} & a_{22} & \cdot & \cdot & a_{2B} \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ a_{D1} & a_{D2} & \cdot & \cdot & a_{DB} \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ \cdot \\ \cdot \\ g_B \end{bmatrix}, \qquad (6)$$

with

$$\sum_{d=1}^{D} a_{db} = 1,\ b = 1, ..., B, \qquad (7)$$

where $\mathbf{a} = (a_{db})$ is a constant coefficient matrix that describes the system. The coefficient matrix is sparse and badly conditioned. The number of equations is oftenly different from the number of unknowns in the system above. Consequently, a direct inversion of the system matrix $\mathbf{a}$ is numerically critical and thus, iterative procedures for solving this system of equations appear to be more promising. One of the most severe problems associated with PET is the statistical fluctuation of the radioactive emission processes which causes noisy projection data. In order to reduce the influence of these effects on the reconstructed image, the so called ML-EM algorithm has been introduced by Vardi and Shepp [1], where the samples of the image are modeled as independent Poisson distributed random variables.

2. ML-EM ALGORITHM

In this section we have to distinguish explicitly between random variables and their realizations (outcomes of experiments, measured data). Since this aspect is of fundamental importance for the correct interpretation of our results, we represent random variables by capital letters and their respective realizations by the corresponding small letters. As just mentioned, the generation of photons in the body of the patient can be modeled as a Poisson process. Consequently, the image samples will be considered as Poisson distributed random variables G_b, $b = 1, ..., B$, with mean value

$$\lambda_b = E\{G_b\}, \qquad (8)$$

where E is the expectation operator. The random variable G_b itself can be considered [4] as a sum of independent Poisson distributed random variables W_{db}, where W_{db} models the number of photons detected in sensor number d and emitted from image point b, i.e.

$$G_b = \sum_{d=1}^{D} W_{db} \qquad (9)$$

and

$$g_b = \sum_{d=1}^{D} w_{db.} \qquad (10)$$

Furthermore, it can be shown [4] that the mean values $\lambda_{db} = E\{W_{db}\}$ are associated with λ_b via

$$\lambda_{db} = a_{db}\lambda_b, \qquad (11)$$

where the a_{db} are defined by (6) and can be determined from the geometry of the tomograph. Due to the independence of the random variables W_{db} we can define the likelihood function

$$l(\boldsymbol{\lambda}) = \prod_{b=1}^{B} \prod_{d=1}^{D} e^{-a_{db}\lambda_b} \frac{(a_{db}\lambda_b)^{w_{db}}}{w_{db}!}. \qquad (12)$$

However, in order to obtain parameter estimators with small variances, large sets of measured data are required, which, however are usually not available in the application considered.

In the next step, the log-likelihood function is considered where products in (12) are transformed into sums

$$L(\boldsymbol{\lambda}) = \ln l(\boldsymbol{\lambda}) = \sum_{b=1}^{B}\sum_{d=1}^{D}(-a_{db}\lambda_b + w_{db}\ln\lambda_b + w_{db}\ln a_{db} - \ln(w_{db}!)), \quad (13)$$

where w_{db} is a realization of W_{db}. In this equation the quantities w_{db} are unknown and have to be estimated in the so called expectation step of the algorithm. This is done in an iterative procedure where the conditional expectation

$$\widehat{w}_{db}^{[k+1]} = E\left\{W_{db}\left|p_d, \widehat{\boldsymbol{\lambda}}^{[k]}\right.\right\}, \quad (14)$$

is used and $\widehat{w}_{db}^{[k]}$ represents the estimate of w_{db} in the k-th iteration step, p_d is the measured projection value and $\widehat{\boldsymbol{\lambda}}^{[k]}$ in the estimate of $\boldsymbol{\lambda}$ in the k-th iteration step. As outlined in [4], this procedure leads to

$$\widehat{w}_{db}^{[k+1]} = \frac{p_d a_{db}\widehat{\lambda}_b^{[k]}}{\sum_{b'=1}^{B} a_{db'}\widehat{\lambda}_{b'}^{[k]}}, \quad (15)$$

Next, we consider the so called maximization step of the ML-EM algorithm. In order to maximize the log-likelihood function one has to set its partial derivatives with respect to the unknown parameters λ_b to zero

$$\frac{\partial}{\partial\lambda_b}\sum_{b=1}^{B}\sum_{d=1}^{D}(-a_{db}\lambda_b + \widehat{w}_{db}^{[k+1]}\ln\lambda_b + \widehat{w}_{db}^{[k+1]}\ln a_{db} - \ln(\widehat{w}_{db}^{[k+1]}!)) = 0 \quad (16)$$

which leads to

$$\widehat{\lambda}_b^{[k+1]} = \sum_{d=1}^{D}\widehat{w}_{db}^{[k+1]}. \quad (17)$$

On the other hand, equation (10) implies that the sum of realizations of w_{db} (belonging to the same experiment) over $d = 1, ..., D$, represents the realization (of the same experiment) g_b of the random variable G_b. Hence, a corresponding sum over estimations $\widehat{w}_{db}^{[k+1]}$ of the realizations w_{db} can be interpreted as an estimate $\widehat{g}_b^{[k+1]}$ of g_b, i.e.

$$\widehat{\lambda}_b^{[k+1]} = \widehat{g}_b^{[k+1]}. \quad (18)$$

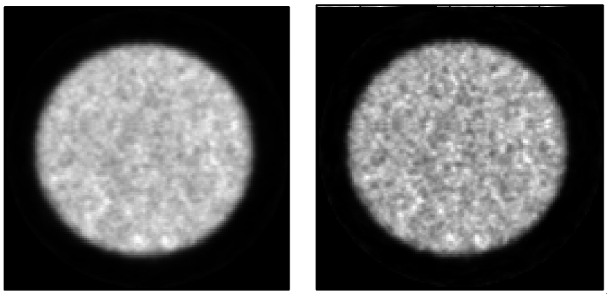

Figure 2: Reconstructed image of a cylinder phantom after 20 iterations (left) and 50 iterations (right)

Obviously, the 'best' estimate $\widehat{\lambda}_b$ of the expected value λ_b of the random variable G_b is the current realization g_b of G_b, if we apply the ML-EM algorithm. This astonishing result gives deep insight into the mechanism of this algorithm. Obviously, the algorithm is not able to deliver a reliable estimate of λ_b, but simply tries to reconstruct the current radioactive activities, which, due to the stochastic nature of these processes, can be interpreted as the λ_b's superposed by 'noise'. In other words, the algorithm is not able to reduce statistical fluctuations by means of averaging over independent measurements which would be the only possible mechanism to improve the SNR (signal to noise ratio) of the reconstructed image. Experiments with real data confirm this result, that the reconstructed image is distorted by superposed high frequency noise (cf. Fig. 2, where reconstructed images (after 20 and 50 iterations) of a cylinder phantom are depicted). However, the stochastic nature of the radioactive processes is not the only reason for noisy effects in reconstructed images.

One has to interpret simulation results carefully, since, additionally, the convergence behaviour of the algorithm itself amplifies noisy artefacts during a long time of the iteration cycle. This aspect will be discussed in the next section.

3. CONVERGENCE BEHAVIOUR

The convergence behaviour of the algorithm is analyzed by means of artificially constructed projection data. This is done by projecting an artificially created image $\mathbf{g}$ (which is deterministic) of an ideal disc via matrix $\mathbf{a}$, i.e. the projections

$$\mathbf{p} = \mathbf{ag} \quad (19)$$

are not distorted by noisy (statistical) effects. In other words, a good reconstruction algorithm should be able to determine $\mathbf{g}$ on basis of the given projections $\mathbf{p}$. Fig-

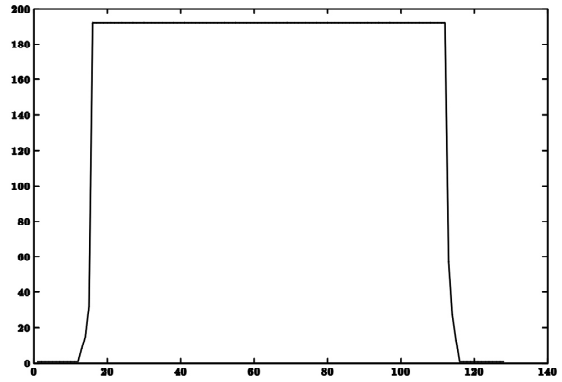

Figure 3: Profile of an artificial disc

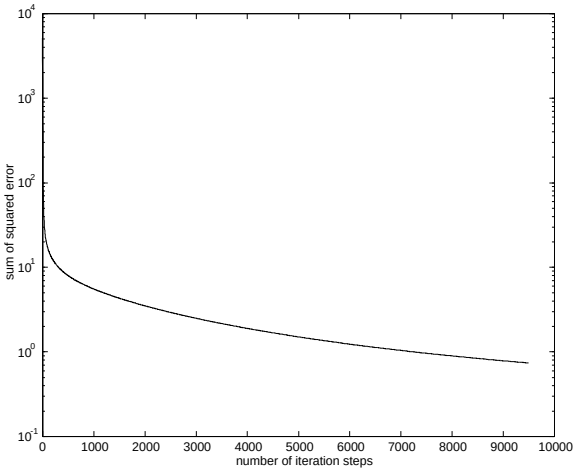

Figure 4: Sum of squared error as a function of the number of iteration steps

ure 3 shows a profile of this disc.

The euclidean distance f between reconstruction $\widehat{\mathbf{g}}$ and original image $\mathbf{g}$ represents the reconstruction error and served as convergence criterion

$$f = \sum_{b=1}^{B} (g_b - \widehat{g}_b)^2. \tag{20}$$

In a computer simulation with 19000 iteration steps, this deviation has been measured. The algorithm converges slowly, but monoton. Figure 4 shows the results of the first 9500 iteration steps.

In spite of the ideal data used in our experiments, the well known artefacts occurring during the iteration procedure had been also observed. These artefacts appear as superposed noisy texture in the reconstructed

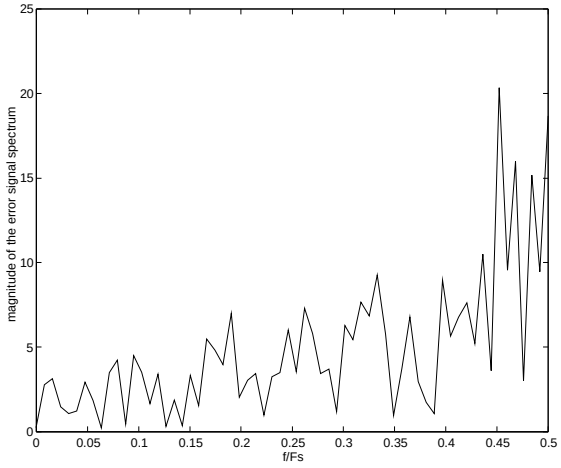

Figure 5: Magnitude of the difference of spectra between reconstructed and original profile after 19000 iteration steps. Fs=sampling frequency

images, whose granularity seems to become finer during the iteration process. Since fine structured noise is more disturbing for humans, many people associated with these effects a declining performance of the ML-EM algorithm after approximately 20 iterations. This subjective interpretation is definitely incorrect if we apply the mathematically based quality measure f. The latter indeed was monotonically decreasing during all of our simulations. In order to explain this discrepancy between subjective and objective (mathematically based) quality judgement one has to investigate the spectral properties of the reconstruction error. Respective investigations have been performed on basis of the already described ideal disc image. The magnitude of the difference between the spectra of reconstructed and original profile after 19000 iterations is shown in Figure 5.

The results of our computer simulations can be interpreted as follows. At the beginning of the iteration process, the reconstruction error signal has a relatively high energy, which is nearly uniformly distributed over the whole spatial frequency range (Fig. 6). With progressing iteration, the total error signal energy declines, however, it is more and more concentrated in the high spatial frequency band. In other words, after some iterations, the low frequency components of the original image are nearly perfectly reconstructed whereas the reconstruction of higher frequency components needs a large number of iteration steps. In particular, the latter statement is of special interest. Most people working with the ML-EM algorithm have observed a 'conver-

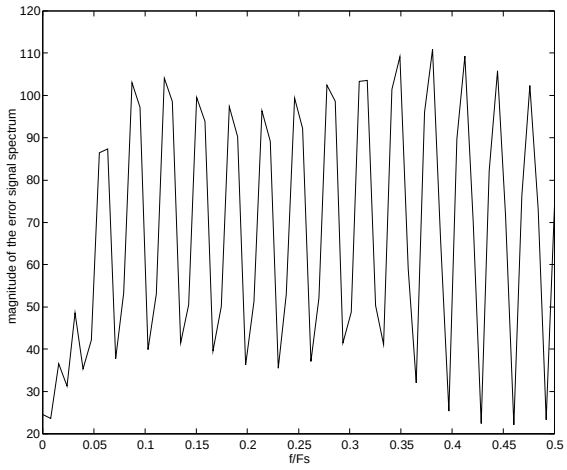

Figure 6: Magnitude of the difference of spectra between reconstructed and original profile after 20 iteration steps. Fs=sampling frequency

gence' of the algorithm after approximately 20 iterations [3] resulting in a poor reconstruction. However, in light of the above discussion, the algorithm is still far away from its equilibrium point. Thus, improvements are still achievable. However, due to the extremely slow convergence of the algorithm with respect to the reconstruction of high frequency components, many thousands of iteration steps would be needed for significantly improving the results, what would be impossible in practice. In order to verify the last statement we tested the ML-EM algorithm with an ideal disc image of lower resolution. Indeed, after a sufficiently long iteration time, even the higher frequency components of the reconstruction error signal had been so small that they lay below the visual perception level.

In summary, differences between the low frequency components of reconstructed and original image in the k-th iteration step lead to a large correction signal. In contrast to this, the algorithm is very insensitive with respect to high spatial frequency components of the error signal.

4. CONCLUSIONS AND REMARKS

The results of the above investigations could be used as a basis for improvements of the ML-EM algorithm or the construction of alternative reconstruction methods. The ML-EM algorithm is an iterative procedure, however, the statistical properties of the measured data are not sufficiently taken into account. The convergence behaviour of the algorithm with respect to high spatial frequency components of the image is rather poor.

As a consequence, the expectation step of the algorithm should be replaced by an algorithm with better convergence behaviour. Promising results have been achieved by using conjugate gradient methods. In general, parameter estimation methods always need a large set of measurements in order to reduce the variance of the estimator. In our application, however, the number of unknown independent parameters λ_b is not substantially smaller than the number of statistically independent measured parameters (projection data). Only in special cases, where a priori knowledge concerning the spatial smoothness of the images under reconstruction is available [2], statistically significant results are achievable. In other words, there are two possibilities for constructing better estimators. First, many independent measurements of projection data can be used for achieving a small variance of the estimator. Second, the number of independent parameters can be reduced by using a priori knowledge, e.g. with respect to smoothness of the image or prior anatomical information from other tomography images [5].

5. REFERENCES

[1] Y. Vardi, L. A. Shepp. *Maximum Likelihood Reconstruction for Emission Tomography.* IEEE Transactions on Medical Imaging, VOL. MI-1, NO 2, October 1982.

[2] E. Levitan and G. T. Herman. *A Maximum a posteriori Probability Expectation Maximization Algorithm for Image Reconstruction in Emission Tomography.* IEEE Transactions on Medical Imaging, VOL. MI-6, NO. 3, September 1987.

[3] J. Llacer, E. Veklerov. *Feasible Images and Practical Stopping Rules in Emission Tomography.* IEEE Transactions on Medical Imaging, VOL. 8, NO. 2, June 1989.

[4] T. K. Moon. *The Expectation Maximization Algorithm.* IEEE Signal Processing Magazine, November 1996.

[5] B. Lipinski, H. Herzog, E. Rota Kops, W. Oberschelp, H.W. Müller-Gärtner. *Expectation Maximization Reconstruction of Positron Emission Tomography Images using Anatomical Magnetic Resonance Information.* IEEE Transactions on Medical Imaging, VOL MI-16, NO 2, April 1997.