Photogrammetrie • Fernerkundung • Geoinformation 3/2006, S. 229–238, 6 figs.

ICM Algorithm for the Bayesian Reconstruction of Tomographic Images

KARL-RUDOLF KOCH, Bonn

Keywords: image processing, tomographic images, Bayesian reconstruction, image smoothing

Summary: The ICM (iterated conditional modes) algorithm of BESAG (1986) is derived for the Bayesian reconstruction of three-dimensional images of the positron emission tomography. As prior distribution a modified density function of HUBER (1964) for a robust parameter estimation is used. This gives a fast algorithm for the reconstruction which smoothes the image while preserving the edges. The results of the ICM algorithm are compared with the ones of the Gibbs sampler, a frequently applied Markov Chain Monte Carlo method. For an example with simulated data it turns out that the ICM algorithm gives a MAP (maximum a posteriori) estimate.

Zusammenfassung: *ICM Algorithmus für die Rekonstruktion tomographischer Bilder mit dem Bayes Verfahren*. Der ICM (iterated conditional modes) Algorithmus von BESAG (1986) wird für die Rekonstruktion dreidimensionaler Bilder der Positronen-Emissions-Tomographie nach dem Bayes-Verfahren abgeleitet. Als Priori-Verteilung wird eine modifizierte Dichte von HUBER (1964) für die robuste Parameterschätzung eingeführt. Man erhält für die Rekonstruktion einen schnellen Algorithmus, der das Bild glättet und die Kanten im Bild erhält. Die Ergebnisse des ICM Algorithmus werden mit denen des Gibbs-Verfahrens verglichen, einer häufig angewandten Methode der Monte-Carlo-Verfahren mit Markoff-Ketten. Für ein Beispiel mit simulierten Daten wird festgestellt, dass der ICM Algorithmus auf eine MAP (maximum a posteriori) Schätzung führt.

1 Introduction

Positron emission tomography is applied to study metabolic activities, like the distribution of a pharmaceutical in a part of a body. The pharmaceutical is combined with a radioactive isotope which produces a positron. The positron finds a nearby electron and annihilates with it to form a pair of photons. The two photons move in opposite directions along a straight line. They collide with a pair of detectors which are placed around the body on several rings forming a tube so that a coincidence line between the two detectors is determined. From the photon counts of pairs of detectors the three-dimensional image of the positions of the photon emitters is reconstructed, cf. LEAHY & QI (2000). The image is represented by a three-dimensional array of voxels with varying intensities proportional to the number of photon emissions.

For a statistical analysis of the photon counts it is generally assumed that the photons are generated by a Poisson process. The maximum likelihood estimation is then solved by the EM (expectation maximation) algorithm independently proposed by SHEPP & VARDI (1982) AND LANGE & CARSON (1984). This algorithm has two disadvantages, it is slow to converge and the reconstruction has high variance so that it needs smoothing to eliminate the noise. The defects can be overcome by a Bayesian reconstruction. For a faster convergence a gamma distributed prior may be introduced, cf.

1432-8364/06/2006/0229 $ 2.50

Lange et al. (1987), Wang & Gindi (1997). For the smoothing one should keep in mind that the intensities of the voxels of the image represent a random field for which the Markov property can be assumed because the intensity of a voxel is mainly influenced by the ones of the voxels of the neighbourhood, cf. Koch & Schmidt (1994, p. 299). The prior information can then be expressed by the Gibbs distribution which is defined such that large density values of the posterior distribution follow for smooth images and small ones for rough images so that a smooth image is obtained from the prior information, see Geman & McClure (1987). However, the smoothing has to stop at the edges, where sudden changes of the intensities of the voxels occur.

Random fields of line elements have therefore been introduced to represent the edges in two-dimensional applications, cf. Busch & Koch (1990), but it is difficult to define priors for line elements. A promising way of handling the edges has been obtained by modeling the Gibbs distribution by the density function of Huber (1964) for the robust parameter estimation, cf. Fessler et al. (2000), Qi et al. (1998). Voxels beyond edges are considered outliers and are accordingly downweighted. A similar effect results from the use of the median root prior (Alenius & Ruotsalainen (1997)) which gives good spatial details as shown by Bettinardi et al. (2002). For a better edge preserving property Koch (2005) modified Huber's density function such that pixels beyond edges of two-dimensional images do not contribute to the smoothing. The method was tested for photographic images and showed an excellent edge preserving quality.

The same modification of the density function of Huber (1964) for a robust parameter estimation is used here for the reconstruction of three-dimensional images of the positron emission tomography by deriving the ICM (iterated conditional modes) algorithm, first sugggested for digital image analysis by Besag (1986). The prior information of this algorithm is controlled by two parameters, a weight and a parameter for intensity differences. Depending on the weight the algorithm converges faster or slower than the EM algorithm. It reduces the noise of the imge by smoothing it while preserving the edges. The technique is applied for a numerical example from data of the small animal PET scanner ClearPET^{TM} Neuro of the Forschungszentrum Jülich (Ziemons et al. 2005). The results are compared with the ones of the Gibbs sampler which is a special Markov Chain Monte Carlo method first suggesed by Geman & Geman (1984) for the Bayesian restoration of digital images. It turns out that the ICM algorithm gives a MAP (maximum a posteriori) estimate, therefore not an approximate but an optimal solution.

In the following section the ICM algorithm is derived for the image reconstruction of emission tomography. Section 3 deals with the Gibbs sampler while Section 4 presents numerical examples. The paper finishes with conclusions.

2 Image Reconstruction

Let Ω be the set of voxels forming a three-dimensional array with

$$\Omega = \{j = (l,m,o),\ 0 \le l \le L,\ 0 \le m \le M,\ 0 \le o \le O\},\ u = (L+1)(M+1)(O+1) \quad (2.1)$$

and let β_j with $j \in \{1,\ldots,u\}$ be the unknown intensity of voxel j which is proportional to the number of photon emissions of voxel j. The vector $\boldsymbol{\beta}$ with $\boldsymbol{\beta} = (\beta_j)$ is therefore the vector of unknown parameters of the reconstruction. Let y_i with $i \in \{1,\ldots,n\}$ be the number of photon emissions or the number of coincidence lines measured by the pair i of detectors so that $\boldsymbol{y} = (y_i)$ gives the vector of observations.

Let y_{ij} be the number of photons emitted by voxel j and counted at detector pair i. This number is nonobservable because of the superposition of the photon streams. The expected value $\bar{y}_{ij}$ of y_{ij} is connected to the unknown intensity β_j of voxel j by

$$\bar{y}_{ij} = E(y_{ij}) = p_{ij}\beta_j \quad (2.2)$$

where p_{ij} gives the probability of detecting an emission from voxel j at detector pair i.

It is determined by the geometry of the scanner and is therefore known. By summing over the voxels j which are cut by the coincidence line i the observation y_i and its expectation $\bar{y}_i$ follow from

$$y_i = \sum_j y_{ij} \quad \text{and} \quad \bar{y}_i = E(y_i) = \sum_j p_{ij}\beta_j \tag{2.3}$$

and the expectation $\bar{\boldsymbol{y}}$ of the vector $\boldsymbol{y}$ of observations from

$$\bar{\boldsymbol{y}} = E(\boldsymbol{y}) = \boldsymbol{P}\beta \tag{2.4}$$

where $\boldsymbol{P} = (p_{ij})$ denotes the so-called system matrix with elements p_{ij}. The system matrix $\boldsymbol{P}$ does not only contain the probabilities p_{ij} but also a number of corrections, cf. Leahy & Qi (2000), which need not be considered here, since in the application we will use simulated data.

The random number y_{ij} results from counting photons so that it is assumed as Poisson distributed, cf. Koch (1999, p. 87). Since the y_{ij} are independent, the measurement y_i with expectation $\bar{y}_i$ is also Poisson distributed with density function

$$p(y_i|\boldsymbol{\beta}) = \frac{\bar{y}_i^{y_i}\exp(-\bar{y}_i)}{y_i!} = \frac{(\Sigma_j p_{ij}\beta_j)^{y_i}\exp(-\Sigma_j p_{ij}\beta_j)}{y_i!}. \tag{2.5}$$

The measurements y_i are independent, too, the joint density function for $\boldsymbol{y}$ therefore follows from

$$p(\boldsymbol{y}|\boldsymbol{\beta}) = \prod_i \frac{(\Sigma_j p_{ij}\beta_j)^{y_i}\exp(-\Sigma_j p_{ij}\beta_j)}{y_i!}. \tag{2.6}$$

This is the likelihood function for the Bayesian reconstruction.

The intensity β_j of voxel j with $j \in \Omega$ represents a Markoff random field as already mentioned in the introduction. A special Gibbs distribution defined for cliques with two sites of the three-dimensional neighbourhood N_p of order p is therefore chosen as prior distribution (Koch 2005)

$$p(\boldsymbol{\beta}) \propto \exp\left\{-\frac{c_\beta}{2}\sum_{r\in\Omega}\sum_{s\in N_p}(\beta_j - \beta_{j+s})^2\right\}. \tag{2.7}$$

This is a normal distribution where $\propto$ means proportionality. The constant c_β acts as a weight and determines the contribution of the prior information. The index s defines the indix of a voxel in half of the neighbourhood of voxel j, because one has to sum in (2.7) over all cliques with two sites in the set Ω. This is accomplished by summing over the cliques of half the neighbourhood N_p, cf. Koch & Schmidt (1994, p. 277). For the application in Section 4 a three-dimensional neighbourhood of 32 voxels surrounding voxel j has been chosen. The larger the intensity difference in (2.7) between voxel j and voxel $j+s$ the smaller is the density value. The reconstruction of a rough image is therefore less likely than the reconstruction of a smooth one.

If voxel j and voxel $j+s$ are seperated by an edge, a sudden change in the intensity, the voxel $j+s$ should not contribute to the smoothing of voxel j. The density function of Huber (1964) for a robust parameter estimation is therefore modified such that we use in (2.7) for a given index s

$$\begin{aligned} p(\beta_j) &\propto \exp(\beta_j - \beta_{j+s})^2/2 \quad \text{for} \quad |\beta_j - \beta_{j+s}| \le c \\ p(\beta_j) &= 0 \qquad\qquad\qquad\quad \text{for} \quad |\beta_j - \beta_{j+s}| > c \end{aligned} \tag{2.8}$$

where the constant c is set according to the jumps in the intensities of the edges which one wants to preserve. Substituting (2.8) in (2.7) means that we set in (2.7) (Koch 2005)

$$\begin{aligned} c_\beta &\ne 0 \quad \text{for} \quad |\beta_j - \beta_{j+s}| \le c \\ c_\beta &= 0 \quad \text{for} \quad |\beta_j - \beta_{j+s}| > c\,. \end{aligned} \tag{2.9}$$

Since the number of voxels $j+s$ changes which contribute to the smoothing, we form the mean of their contributions and obtain with q being the number of voxels $j+s$, for which $c_\beta \ne 0$ holds, instead of (2.7) together with (2.9)

$$p(\boldsymbol{\beta}) \propto \exp\left\{-\frac{c_\beta}{2}\sum_{r\in\Omega}\sum_{s\in N_p}(\beta_j - \beta_{j+s})^2 q\right\}. \tag{2.10}$$

The prior (2.10) together with the likelihood function (2.6) gives with Bayes' the-

orem, cf. Koch (2020, p. 32), the posterior density function for $\boldsymbol{\beta}$

$$p(\boldsymbol{\beta}|\boldsymbol{y}) \propto \exp\left\{-\frac{c_\beta}{2}\sum_{r\in\Omega}\sum_{s\in N_p}(\beta_j-\beta_{j+s})^2/q\right\} \prod_i (\sum_j p_{ij}\beta_j)^{y_i}\exp(-\sum_j p_{ij}\beta_j). \quad (2.11)$$

The conditional density function for β_j given the unknown intensities $\partial\beta_j$ in the neighbourhood N_p of voxel j follows from (2.11) with t being now the summation index, cf. Koch & Schmidt (1994, p. 262),

$$p(\beta_j|\partial\beta_j,\boldsymbol{y}) = \frac{1}{C}\exp\left\{-\frac{c_\beta}{2}\sum_{\pm s\in N_p}(\beta_j-\beta_{j+s})^2/q\right\} \prod_i (\sum_t p_{it}\beta_t)^{y_i}\exp(-\sum_t p_{it}\beta_t). \quad (2.12)$$

where C denotes the normalization constant and where the sum has to be extended over the whole neighbourhood N_p of voxel j so that the index s becomes positive and negative.

If we derive the MAP estimate for β_j and apply it iteratively for $j\in\Omega$, the ICM algorithm of Besag (1986) results. Taking the logarithm of (2.12)

$$\ln p(\beta_j|\partial\beta_j,\boldsymbol{y}) = -\frac{c_\beta}{2}\sum_{\pm s\in N_p}(\beta_j-\beta_{j+s})^2/ q + \sum_i (y_i \ln\sum_t p_{it}\beta_t - \sum_t p_{it}\beta_t) - \ln C \quad (2.13)$$

and the derivative with respect to β_j

$$\frac{d\ln p(\beta_j|\partial\beta_j,\boldsymbol{y})}{d\beta_j} = -c_\beta\sum_{\pm s\in N_p}(\beta_j-\beta_{j+s})/ q+\sum_i\left(\frac{p_{ij}y_i}{\Sigma_t p_{it}\beta_t}-p_{ij}\right) \quad (2.14)$$

and setting the result equal to zero gives the condition the MAP estimate for β_j has to fulfill given in a form explained below

$$1 = \frac{1}{\Sigma_i p_{ij} + c_\beta\Sigma_{\pm s\in N_p}(\beta_j-\beta_{j+s})/q}\sum_i\frac{p_{ij}y_i}{\Sigma_t p_{it}\beta_t}. \quad (2.15)$$

It leads to the ICM algorithm for the kth step of the iteration

$$\beta_j^{(k+1)} = \frac{\beta_j^{(k)}}{\Sigma_i p_{ij} + c_\beta\Sigma_{\pm s\in N_p}(\beta_j^{(k)}-\beta_{j+s}^{(k)})/q}\sum_i \frac{p_{ij}y_i}{\Sigma_t p_{it}\beta_t^{(k)}}. \quad (2.16)$$

The prior information of this algorithm is weighted by c_β with respect to the contribution of the observations y_i. If c_β is too large, the iterations will not converge anymore because the intensities will continuously increase unconstrained by the observations. If c_β is very small or equal to zero, the EM (expectation maximation) algorithm of Shepp & Vardi (1982) and Lange & Carson (1984) is obtained, see also Vardi et al. (1985). The second parameter which controlls the prior information in the ICM algorithm is according to (2.9) the constant c which determines the intensity difference of the edges one wants to preserve.

Green (1990) proposed the OSL (one step late) approximation in order to solve the Bayesian reconstruction also by the EM algorithm. The name was chosen because the derivative of the prior density function with respect to the unknown intensity of voxel j is evaluated at the current estimate of the unknown parameter during the iterations. For the prior density function (2.7) used here the OSL algorithm gives

$$\beta_j^{(k+1)} = \frac{\beta_j^{(k)}}{\Sigma_t p_{ij} + c_\beta\Sigma_{s\in N_p}(\beta_j^{(k)}-\beta_{j+s}^{(k)})/q}\sum_i \frac{p_{ij}y_i}{\Sigma_t p_{it}\beta_t^{(k)}}. \quad (2.17)$$

By comparing it with the ICM algorithm (2.16) one recognizes that the ICM algorithm is presented in the shape of the OSL algorithm. This was the reason that for the condition of the MAP estimate the form (2.15) had been chosen.

The ICM algorithm (2.16) differs from the OSL algorithm (2.17) only by the fact that for computing the prior information the sum in (2.16) is extended over the whole neighbourhood of voxel j, while in (2.17) the

sum is taken over half of the neighbourhood. It is therefore obvious and test computations have confirmed it that the ICM algorithm converges faster than the OSL algorithm so that the ICM algorithm has to be preferred.

3 Gibbs Sampler

The Gibbs sampler, a special Markov Chain Monte Carlo method, is also applied to compute the reconstruction by a MAP estimate and in addition by a Bayes estimate. Because it is much slower than the ICM algorithm it is only used as a check to numerically find out whether the solution of the ICM algorithm agrees with the MAP estimate.

The Gibbs sampler generates random samples $\boldsymbol{\beta}^{(w)}$ from a posterior density function $p(\boldsymbol{\beta}|\boldsymbol{y})$ by using the conditional density functions

$$p(\beta_j|\beta_1,\ldots,\beta_{j-1},\beta_{j+1},\ldots,\beta_u,\boldsymbol{y}) \quad \text{for} \quad j\in\{1,\ldots,u\} \tag{3.1}$$

The Gibbs sampler begins with arbitrary starting values

$$\beta_1^{(0)},\ldots,\beta_u^{(0)}$$

Then, random values

$$\begin{aligned}
&\beta_1^{(1)} \quad \text{from} \quad p(\beta_1|\beta_2^{(0)},\ldots,\beta_u^{(0)},\boldsymbol{y})\\
&\beta_2^{(1)} \quad \text{from} \quad p(\beta_2|\beta_1^{(1)},\beta_3^{(0)},\ldots,\beta_u^{(0)},\boldsymbol{y})\\
&\beta_3^{(1)} \quad \text{from} \quad p(\beta_3|\beta_1^{(1)},\beta_2^{(1)},\beta_4^{(0)},\ldots,\beta_u^{(0)},\boldsymbol{y})\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
&\beta_u^{(1)} \quad \text{from} \quad p(\beta_u|\beta_1^{(1)},\ldots,\beta_{u-1}^{(1)},\boldsymbol{y})
\end{aligned} \tag{3.2}$$

are sequentially generated to complete the first step of an iteration. After a burn-in phase of for instance o iterations the Gibbs sampler can be shown to converge to the posterior density function $p(\boldsymbol{\beta}|\boldsymbol{y})$ so that the generated samples $\boldsymbol{\beta}^{(w)}$ with $w > o$ are distributed like $p(\boldsymbol{\beta}|\boldsymbol{y})$, cf. Gelman et al. (2004, p. 287). The Bayes estimate $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ then follows as a mean, cf. Koch (2000, p. 206),

$$\hat{\boldsymbol{\beta}} = \frac{1}{W}\sum_w \boldsymbol{\beta}(w) \tag{3.3}$$

with W being the number of generated samples beyond the burn-in phase. The MAP estimate $\hat{\boldsymbol{\beta}}_M$ of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}}_M = \arg\max_{\boldsymbol{\beta}^{(w)}} p(\boldsymbol{\beta}^{(w)}|\boldsymbol{y})\,. \tag{3.4}$$

The conditional density functions to sample from by the Gibbs sampler (3.2) are taken because of simpler computing not from (2.12) but from (2.13)

$$p(\beta_j|\partial\beta_j,\boldsymbol{y}) = \exp\left\{-\frac{c_\beta}{2}\sum_{\pm s\in N_p}(\beta_j-\beta_{j+s})^2/\right.$$
$$\left. q+\sum_i (y_i \ln\sum_t p_{it}\beta_t - \sum_t p_{it}\beta_t) - \ln C\right\}. \tag{3.5}$$

To find the MAP estimate from (3.4) the density values for the generated samples $\beta^{(w)}$ have been obtained for simpler computations from taking the logarithm of (2.11) where the normalization constant need not be computed.

A method to sample from (3.5) could not be found. It is therefore assumed that the unknown intensity β_j can be approximated within an interval I_j by discrete values $\beta_{jk}^{(w)}$ for $k\in\{1,\ldots,d\}$ with the current value $\beta_j^{(w)}$ of the Gibbs sampler lying in the middle of the interval I_j. The random variable β_j is therefore considered a discrete random variable with values β_{jk} for which it is simple to generate random values, cf. Koch (2000, p. 185). However, this sampling method is very slow since the number of discrete values may not be small. Thus, in such an application the Gibbs sampler is not a practical algorithm.

4 Numerical Example

To apply the ICM algorithm (2.16) and to compare the results with the Gibbs sampler (3.2), a small example with generated data is chosen because of the large amount of computing time needed for the Gibbs sampler. The starting point is a test image obtained by the scanner Clear PET^{TM} Neuro of the Forschungszentrum Jülich. The image consists of the intensities of a three-dimensional array of voxels with 96 slices, 81

rows and 81 columns. The image shows a cylinder with holes open at the top parallel to the axis . The mantle of the cylinder extends a bit beyond the bottom to which two small cylinders are attached. The image has been reconstructed by the EM algorithm from about 8 000 000 coincidence lines. The image shows excellent contrasts for the cylinder and its holes exept for the bottom of the cylinder where low intensities with low contrasts prevail so that smoothing can be applied. In this area a three-dimensional array of voxels with 9 slices, 45 rows and 33 columns has been selected for the example. To avoid boundary effects for the smoothing only the intensities of an array of voxels with 5 slices, 41 rows and 29 columns within the larger array are reconstructed. Thus, the vector $\boldsymbol{\beta}$ contains 5945 unknown intensity values. Fig. 1 shows the intensities expressed by grey values of the test image for the slice in the middle of the array of voxels for the example. The circular ring of light grey values indicates the mantle of the cylinder and the light grey values to the right close at the mantle show one of the small cylinders attached to the bottom. The area of light grey values to the left does not have a special meaning.

Instead of searching the 8 000 000 measured coincidence lines for the ones going through the array of voxels for the example the observations were generated. It was assumed that the emissions of the photons of the example result in 8000 coincidence lines symmetrically distributed over the array of voxels. The element p_{ij} of the system matrix $\boldsymbol{P}$ in (2.4) was computed by the ratio of the length of the coincidence line i within the voxel j to the length of the coincidence line within the array of voxels. The method of SIDDON (1985) was applied. The voxel volume was set to 1 mm^3 and the radius of the scanner was reduced to 34 mm according to the ratio of the size of the image of the example to the size of the test image. The system matrix $\boldsymbol{P}$ thus obtained multiplied by the intensities of the voxels of the test image restricted to the example gives the generated data.

To reconstruct the intensities of the voxels from the simulated data, the intensities of the test image after introducing normally distributed noise with a standard deviation of 0.00005 are used as starting values for the iterations. The intensities of the voxels lie within the interval of [0.00004, 0.00825]. The EM algorithm without the term c_β in (2.16) was first applied to reconstruct the image. The result is given in Fig. 2 where the intensities are expressed by grey values and the same slice as in Fig. 1 is shown. The number of iterations was determined such that 1000 additional iterations do not change the image which was checked by a visual inspection. The number of iterations thus obtained was 2000 for the EM algorithm. By comparing the results with Fig. 1, where as mentioned above the intensities of the test image are shown, one recognizes especially from the dark voxels that the EM algorithm introduces variances.

The reconstructed intensities of Fig. 2 vary within the interval [0.0001, 0.0043] with only a few voxels in the lower left corner of Fig. 2 surpassing the intensity 0.0015. To obtain sufficient contrast for the image, the intensities of these voxels are reduced to 0.0015. The intensities are then transformed to grey values and shown in Fig. 2. This procedure has been applied to all Figures 1 to 6. The consequence is that the images appear rougher than the images where the whole range of intensities are depicted. On the other hand it is clearly visible how the reconstruction algorithm builds the image.

By applying the ICM algorithm (2.16) it turns out that the iterations do not converge anymore for $c_\beta \geq 200$. The prior information is not constrained for these weights by the observations so that the intensities continuously grow. The reconstruction by (2.16) with a low weight of $c_\beta = 20$ and an intensity difference of $c = 0.00017$ for the edges to be preserved is shown in Fig. 3 again for the same slice as in Fig. 1 and Fig. 2. The comparison with Fig. 2 reveals that the image is smoother and the edges forming the mantle of the cylinder are well kept. Prior information in addition to the observations need to be processed so that 2500 iterations are necessary. Inspecting the results of 1000 more iterations reveals that

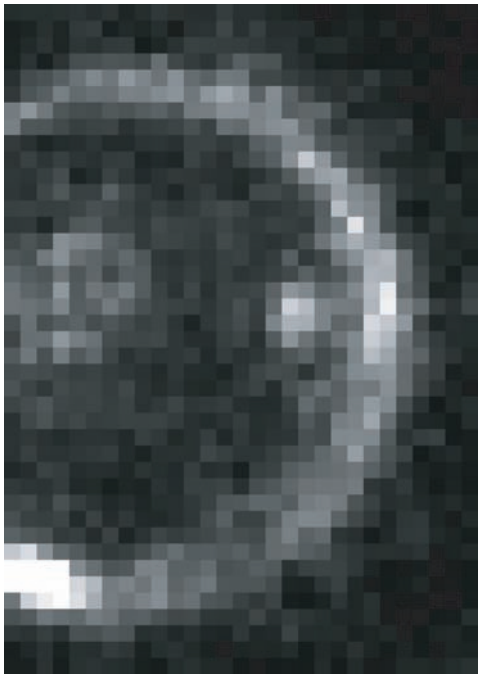

Fig. 1: Test Image of the Slice in the Middle of the Array of Voxels for the Example.

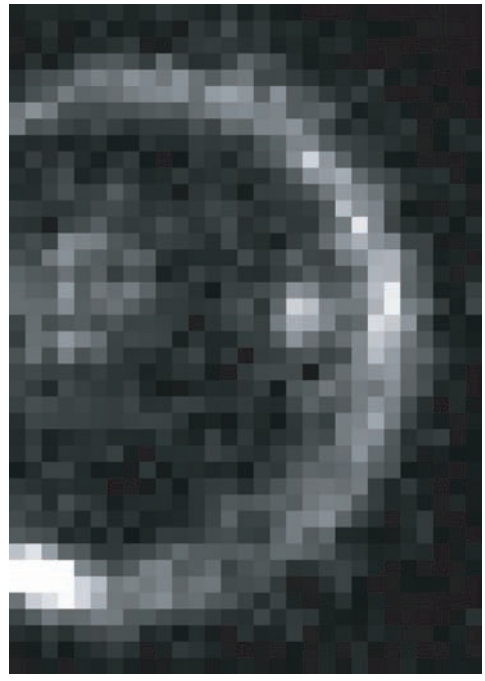

Fig. 2: Image Reconstructed by the EM Algorithm.

they do not change the image. Only 1500 iterations are needed if the weight is increased to $c_\beta = 100$ with the indentical value $c = 0.00017$, see Fig. 4. This results in a stronger smoothing. One has to keep in mind that the ICM algorithm (2.16) applies a three-dimensional smoothing with a neighbourhood of 32 voxels as explained in Section 2. This is the reason that patterns appear in Fig. 3 and Fig. 4, which would not show up in a two-dimensional smoothing, like the break in the ring of voxels with light grey values near the upper left corner of Fig. 4.

Decreasing the parameter for the intensity difference to $c = 0.00008$ together with the low weight $c_\beta = 20$ gives after 2500 iterations, as shown in Fig. 5, a moderate smoothing if one compares it with Fig. 3. Fig. 5 is less rough than Fig. 2 and it is almost identical with the original test image of Fig. 1. This is an important result because the original image can with good approximation be recovered by the ICM algorithm (2.16) which the EM algorithm cannot. Increasing the weight to $c_\beta = 100$ with $c = 0.00008$ gives after 1500 iterations stronger smoothing, see Fig. 6, but less pronounced than in Fig. 4.

The solution shown in Fig. 3 of the ICM algorithm (2.16) with $c_\beta = 20$ and $c = 0.00017$ is investigated by the Gibbs sampler (3.2). As starting values for the Gibbs sampler the intensities of the solution of the ICM algorithm are taken. Since it is expected that these intensities are already distributed according to (2.11), no burn-in phase is assumed. The Gibbs sampler is applied twice, the first time with the lengths of the intervals I_j equal to 0.000 001 and the second time equal to 0.000 002. In both cases $d = 21$ discrete values equally spaced in the intervals I_j are taken and 1500 iterations are computed. For sampling from (3.5) the normalization constant C has to be known. Because of the discretization it cannot be ac-

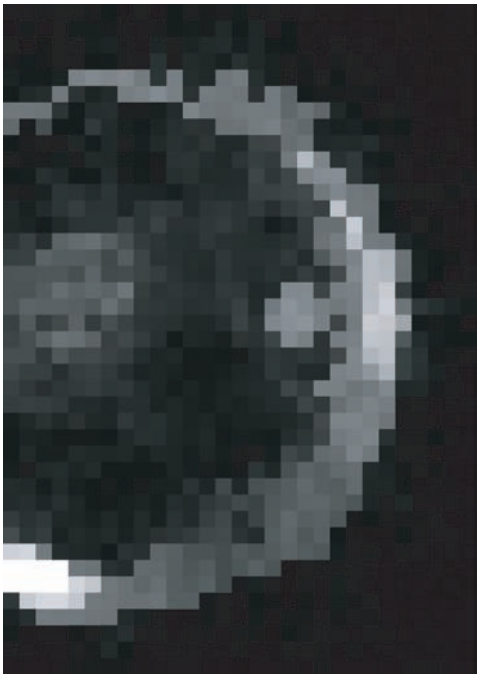

Fig. 3: Image Smoothed with $c_\beta = 20$ and $c = 0.00017$.

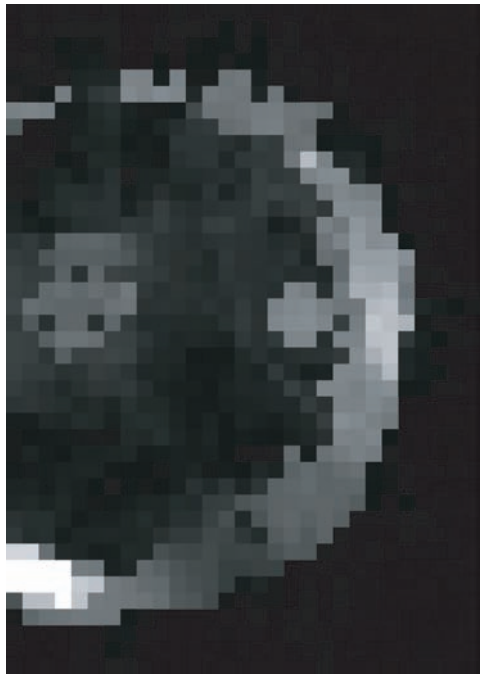

Fig. 4: Image Smoothed with $c_\beta = 100$ and $c = 0.00017$.

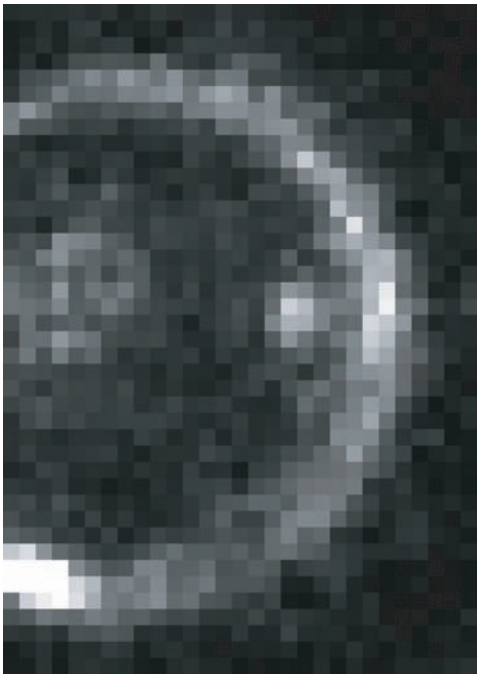

Fig. 5: Image Smoothed with $c_\beta = 20$ and $c = 0.00008$.

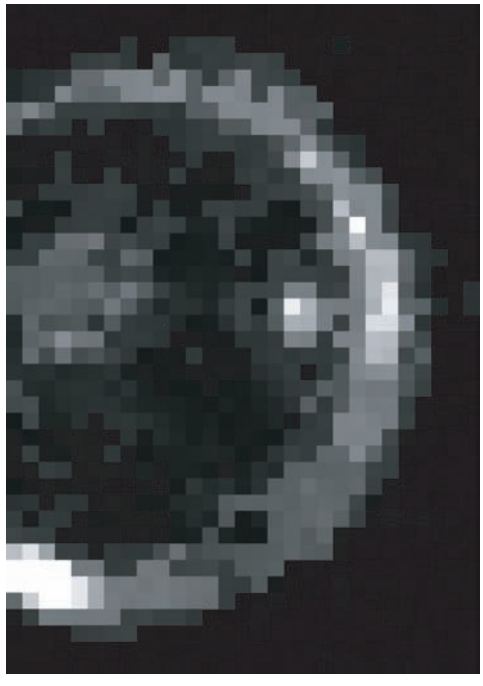

Fig. 6: Image Smoothed with $c_\beta = 100$ and $c = 0.00008$.

curately determined. When applying the ICM algorithm (2.16) with $c_\beta = 20$ the fifth significant digit in the sum of the denominator is affected. Thus, instead of subtracting $\ln C$ in (3.5) the first term of the right hand side of (3.5) is multiplied by a factor which is chosen such that the 5th significant digit of the sum is affected. This is, of course, only an approximation with the consequence that the weight c_β in (3.5) is slightly changed as compared to (2.16). But it alters the solution only a little.

In the first application of the Gibbs sampler the maximum of $p(\boldsymbol{\beta}^{(w)}|\boldsymbol{y})$ in (3.4) is found after 14 iterations, in the second case after 15 iterations. For the first case the maximum difference between an intensity value from the ICM algorithm and the MAP estimate of the Gibbs sampler is equal to 0.000 0043 and the square root of the mean squared differences equals 0.000 0011. For the second case the corresponding values are 0.000 0099 and 0.000 0023. Thus, the results of the ICM algorithm and the MAP estimate of the Gibbs sampler practically agree. The ICM algorihm therefore does not give an approximate solution but a MAP estimate.

To confirm this result, the Bayes estimate (3.3) is also computed by the Gibbs sampler. For the first application the maximum difference between an intensity value obtained by the ICM algorithm and the Gibbs sampler is equal to 0.000 024 and the square root of the mean squared differences equals 0.000 007. For the second application the corresponding values are 0.000 052 and 0.000 014. Theoretically the maximum difference could be half of the lengths of the intervals I_j times the number of iterations which gives 0.00075 for the first case and 0.00150 for the second one. By comparing these values with the maximum differences given above one recognizes that the Bayes estimate stays close to the MAP estimate which confirms the MAP estimate.

5 Conclusions

Numerically it is shown that the ICM algorithm derived here for the Bayesian reconstruction of tomograhic images does not give an approximate solution but an optimal solution in form of a MAP estimate. Due to the modified density of Huber (1964) for the robust parameter estimation the ICM algorithm smoothes the image while preserving the edges. The smoothing is controlled by two parametrs which allows flexibility as demonstrated by the examples. For instance, using simulated observations the original image can be recovered with an excellent approximation which the EM algorithm cannot.

Acknowledgement

The author is indebted to B. Gundlich and S. Weber, Central Institute for Electronics, Forschungszentrum Jülich, for their advice and for providing the test data of the scanner Clear PET^{TM} Neuro.

References

Alenius, S. & Ruotsalainen, U., 1997: Bayesian image reconstruction for emission tomography based on median root prior. – Eur. J. Nucl. Med., **24:** 258–265.

Besag, J.E., 1986: On the statistical analysis of dirty pictures. – J. Royal Statist. Society, B **48:** 259–302.

Bettinardi, V., Pagani, E., Gilardi, M.C., Alenius, S., Thielemans, K., Teras, M. & Fazio, F., 2002: Implementation and evaluation of a 3D one-step late reconstruction algorithm for 3D positron emission tomography brain studies using median root prior. – Eur. J. Nucl. Med., **29:** 7–18.

Busch, A. & Koch, K.R., 1990: Reconstruction of digital images using Bayesian estimates. – Z. Photogrammetrie und Fernerkundung, **58:** 173–181.

Fessler, J.A., Erdogan, H, & Wu, W.B., 2000: Exact distribution of edge-preserving MAP estimators for linear signal models with Gaussian measurement noise. – IEEE Trans Im Proc, **9** (6): 1049–1056.

Gelman, A., Carlin, J.B., Stern, H.S. & Rubin, D.B., 2004: Bayesian Data Analysis. – 2nd ed., Chapman and Hall, Boca Raton.

Geman, S. & Geman, D., 1984: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. – IEEE Trans. Pattern. Anal. Machine Intell., PAMI, **6:** 721–741.

Geman, S. & McClure, D.E., 1987: Statistical methods for tomographic image reconstruction. – Bull. Int. Statist. Inst., 52–21.1: 5–21.

Green, P.J., 1990: Bayesian reconstruction from emission tomography data using a modified EM algorithm. – IEEE Trans. Med. Imaging, **9:** 84–93.

Huber, P.J., 1964: Robust estimation of a location parameter. – Annals Mathematical Statistics, **35:** 73–101.

Koch, K.R., 1999: Parameter Estimation and Hypothesis Testing in Linear Models. – 2nd ed., Springer, Berlin.

Koch, K.R., 2000: Einführung in die Bayes-Statistik. – Springer, Berlin.

Koch, K.R., 2005: Bayesian image restoration by Markov Chain Monte Carlo methods. – ZfV (Z. Geodäsie, Geoinformation und Landmanagement), **130:** 318–324.

Koch, K.R. & Schmidt, M., 1994: Deterministische und stochastische Signale. – Dümmler, Bonn.

Lange, K. & Carson, R., 1984: EM reconstruction algorithms for emission and transmission tomography. – J. Comput. Assist. Tomogr., **8:** 306–316.

Lange, K., Bahn, M. & Little, R., 1987: A theoretical study of some maximum likelihood algorithms for emission and transmission tomography. – IEEE Trans. Med. Imaging, MI-**6:** 106–114.

Leahy, R.M. & Qi, J., 2000: Statistical approaches in quantitative positron emission tomography. – Statistics and Computing, **10:** 147–165.

Qi, J., Leahy, R.M., Cherry, S.R., Chatziioannou, A. & Farquhar, T.H., 1998: High-resolution 3D Bayesian image reconstruction using the microPET small-animal scanner. – Phys. Med. Biol., **43:** 1001–1013.

Shepp, L.A. & Vardi, Y., 1982: Maximum likelihood reconstruction for emission tomography. – IEEE Trans. Med. Imaging, MI-1: 113–122.

Siddon, R.L., 1985: Fast calculation of the exact radiological path for a three-dimensional CT array. – Medical Physics, **12:** 252–255.

Vardi, Y., Shepp, L.A. & Kaufman, L., 1985: A statistical model for positron emission tomography. – J. American Statist. Ass., **80:** 8–37.

Wang, W. & Gindi, G., 1997: Noise analysis of MAP-EM algorithms for emission tomography. – Phys. Med. Biol., **42:** 2215–2232.

Ziemons, K., Auffray, E., Barbier, R., Brandenburg, G., Bruyndonckx, P., Choi, Y., Christ, D., Costes, N., Declais, Y., Devroede, O., Dujardin, C., Federovd, A., Heinrichs, U., Korjik, M., Krieguer, M., Kuntner, C., Largeron, G., Lartizien, C., Larue, H., Lecoq, P., Leonard, S., Marteau, J., Morel, Ch., Mosset, J.B., Parl, Ch., Pedrini, Ch., Petrosyan, A.G., Pietrzyk, U., Rey, M., Saladino, S., Sappey-Marinier, D., Simon, L., Streun, M., Tavernier, S. & Vieira, J.M. (2005): The ClearPET^{TM} project: development of a 2nd generation high performance small animal PET scanner. – Nuclear Instruments and Methods in Physics Research, **A 537:** 307–311.

Author's address:

Prof. em. Dr.-Ing. Dr.-Ing. h.c. mult.
Karl-Rudolf Koch
Institute of Theoretical Geodesy
University of Bonn
Nussallee 17, D-53115 Bonn
e-mail: koch@theor.geod.uni-bonn.de

Manuskript eingereicht: Dezember 2005
Angenommen: Januar 2006