A New Parameters Choice Method for Ill-Posed Problems with Poisson Data and its Application to Emission Tomographic Imaging

Reginaldo J. Santos[1] **and Álvaro R. De Pierro**[2]

[1]Departamento de Matemática-ICEx,
Universidade Federal de Minas Gerais,
Belo Horizonte-MG-Brazil,
http://www.mat.ufmg.br/~regi
[2]Applied Mathematics Department-IMECC,
University of Campinas,
Campinas-SP-Brazil,
http://www.ime.unicamp.br/~alvaro

ABSTRACT

We present a new approach to estimate regularization parameters arising in very large scale inverse problems with Poisson noise. We apply the derived formulae as a stopping rule for the EM and ART methods in Positron Emission Tomography, comparing the results to those obtained by the χ^2 method.

Keywords: regularization, Poisson noise, stopping criterion, iterative methods, EM algorithm, Positron Emission Tomography.

2000 Mathematics Subject Classification: 47A52, 65F22.

1 Introduction

When dealing with discrete linear ill-posed inverse problems, like those arising in image restoration and reconstruction, one aims at finding a solution x of

$$Kx + \epsilon = b\,, \tag{1.1}$$

where K is a real $m \times n$ matrix, b is an m-vector of observations and ϵ an error vector. If ϵ is white gaussian noise (uncorrelated components, zero mean and constant unknown variance σ^2), and K is full rank, a least squares solution of (1.1) is the best (minimum variance) unbiased linear estimator. However, ill-conditioning of K makes this minimum still too large and some kind of regularization is needed. One way to do this is by considering solutions of regularized problems of the form

$$\min_x \quad ||Kx - b||^2 + \lambda x^t B x\,, \tag{1.2}$$

for a positive real number λ, where B is an $n \times n$ matrix, that introduces *a priori* information on the problem. For example, if we take $B = L^t L$, where L is the discretization of first order

derivatives, increasing λ restricts first order variations of x. Another regularization approach consists of applying an iterative method with a 'smooth' starting point, and stopping before deterioration caused by ill-posedeness shows up. In this case, the regularization parameter is the iteration number and the choice is equivalent to a stopping criterion. For some special cases, it can be proven that both approaches are mathematically equivalent (Santos, 1996).

A crucial question is how to choose the regularization parameter, in such a way that the information loss is minimized. Many methodologies have been proposed in the scientific literature to accomplish this task. Some of them (see for example (Zhou, Leahy and Qi, 1997; Higdon, Bowsher, Johnson, T. G. Turkington and Jaszczak, 1997; Hebert and Leahy, 1992)) are based on Maximum Likelihood estimation, but they are expensive; others, sometimes called frequentists methods (Archer and Titterington, 1995), are more suitable for estimating the iteration number, like the chi-squared method (Hebert, 1990), the equivalent degrees of freedom method (Thomson, Brown, Kay and Titterington, 1991), and generalized cross-validation (GCV) (Wahba, 1987). Cross-validation procedures have been used also in tomography as stopping rules (Coakley, 1991).
One possibility is to approximate a value of λ in (1.2) such that the average of the mean square error is minimum, i.e., λ solves the problem

$$\min \; E\{T(\lambda)\}\,, \tag{1.3}$$

where E stands for the expectation,

$$T(\lambda) = \frac{1}{m}||Kx - Kx_\lambda||^2\,, \tag{1.4}$$

and, x_λ and x are solutions of (1.2) and (1.1) respectively. In spite of the fact that we do not know the values of x and σ^2, methods like GCV provide good approximations for the solution of (1.3) when the error is gaussian.
Defining the influence operator (a matrix in this case) as

$$A(\lambda)b = Kx_\lambda\,. \tag{1.5}$$

for the above example, if $A(\lambda)$ in (1.5) is symmetric positive definite, the GCV functional to be minimized, suggested by Wahba (Wahba, 1987), (Craven and Wahba, 1979) for the approximation of (1.3), is given by

$$V(\lambda) = \frac{\frac{1}{m}||b - Kx_\lambda||^2}{[\frac{1}{m}\mathrm{Tr}(I - A(\lambda))]^2}\,. \tag{1.6}$$

where Tr stands for the trace of the matrix. The minima of (1.6) are supposed to approximate those of (1.3).
Essentially, a GCV functional relates the residual, measuring the consistency of the regularized solution (numerator of $V(\lambda)$) with the data, and a denominator containing the trace of the influence operator. Usually, gaussian white noise and simplicity induce the use of the square norm for the consistency measure and the linearity of the influence operator determines a simple trace calculation in the denominator. Sometimes, the previous assumptions (gaussian white noise) are not true and the use of square norms is not appropriate. This is the case in Emission Tomography, where the data is essentially Poisson (Rockmore and Macovski, 1976).

On the other hand, for very large systems, a regularization using (1.2) involves the solution of a large system of equations, just for one evaluation of $V(\lambda)$ at one point, an extremely expensive task in tomography. So, instead of regularizing using a penalization term, it is faster to do it by using an appropriate stopping criterion for an iterative method. Now, the parameter to be chosen is the iteration number and the influence operator becomes nonlinear because of the method and because of the different way of measuring consistency.
If the error is Poisson a very well known stopping criterion is Hebert's chi-squared (Hebert, 1990) that is an enhancement of that from Veklerov-Llacer (Llacer and Veklerov, 1989) (See also (Vogel, 2002)). Our experience indicates that, in Emission Tomography problems, this criterion tends to choose an earlier, therefore oversmoothed image. Apparently, the reason is the fact that the chi-squared sequence always chooses a parameter that approximates the minimum of the squared distance to the 'true image' (see (Hebert, 1990) and the experiments in Section 4). (Perry and Reeves, 1994; Reeves, 1995) applied a randomized GCV stopping criterion derived for gaussian error to problems with Poisson noise, without taking into account the nature of the noise. The square norm quickly increases for images with high frequencies content. Because of this, we were led to search for a method that directly takes into account the Poisson noise characteristics of the problem.
Inspired by the GCV approach, in this paper we introduce a general functional, like (1.6), which uses as a consistency measure the Kullback-Leibler distance instead of the square norm and that is appropriate for nonlinear influence operators. This allows us to apply the new formula as a stopping criterion for iterative algorithms for problems with Poisson data as it the case of Positron Emission Tomography (PET).
The main goal of this paper is to present the new method as an option for choosing parameters for ill-posed problems with Poisson data and to test its applicability in the particular case of Emission Tomography. In the next section, we briefly describe the PET problem and the mathematical background, in section 3 we present the main theoretical result that gives the foundation for the formulae that are used in the application to tomography presented in section 4, where it is compared to chi-squared statistic criteria from (Hebert, 1990). Section 5 contains some concluding remarks.

2 The PET Problem and Mathematical Preliminaries

In Positron Emission Tomography (PET) as well as in several other very large scale inverse problems arising in image processing and reconstruction, iterative methods have been successfully applied to compute approximate solutions. Emission Computed Tomography (ECT) (Budinger, Gullberg and Huesman, 1979) is the quantitative determination of the moment-to-moment changes in the chemistry and flow physiology of radioactive labeled compounds inside the body. Mathematically, the problem consists of reconstructing a function representing the distribution of radioactivity in a body cross-section from measured data that are the total activity along lines of known location. One of the main differences between this problem and that arising in x-ray Computed Tomography (CT) (Herman, 1980) is that measurements are much more noisy. As such, in ECT it is desirable to have a reconstruction approach incorpo-

rating an estimation procedure that takes into account the statistical nature of the noise. For example, in positron Emission Tomography (PET) (Ter-Pogossian, Raichle and Sobel, 1980), the isotope used emits positrons which annihilate with nearby electrons, thereby generating two photons traveling away from each other in (approximately) opposite directions. The collected data are the number of photon pairs detected in time coincidence for each line determined by a pair of detectors. Since the emission and detection of photons are Poisson processes (Herman, 1980), it was suggested by Rockmore and Macovski (Rockmore and Macovski, 1976) that improvements in image reconstruction from ECT data may be achieved by estimating the radioactive distribution in the body cross-section using the maximum likelihood (ML) approach. Mathematically, this approach consists of estimating the emission densities x_j (for each pixel j) by solving the optimization problem

$$\max_{x \geq 0} \; L(x) = \sum_{i=1}^{m} b_i \log\langle a_i, x\rangle - \langle a_i, x\rangle, \tag{2.1}$$

where $b_i \, (i = 1, \ldots, m)$ represents the total emission counts for the i-th pair of detectors (data), $a_i = \{a_{ij}\} \, (j = 1, \ldots, n)$ is the vector derived from the discretization of the i-th line integral, $x = (x_1, \ldots, x_n)^t$ is the image vector of densities and $\langle \, , \, \rangle$ denotes the standard inner product.

Shepp and Vardi (Shepp and Vardi, 1982), and independently Lange and Carson (Lange and Carson, 1987), proposed the use of the expectation maximization (EM) (Dempster, Laird and Rubin, 1977) algorithm to maximize the Poisson likelihood in PET. Since then, the ML-EM approach has become very popular among researchers in ECT, since it has been shown to provide in many cases better results compared to reconstructions produced with the conventional filtered backprojection (FBP) algorithm (Shepp, Vardi, Ra, Hilal and Cho, 1984). Recently, Herman and Meyer (Herman and Meyer, 1993) proposed to use ART (a very popular method in image processing and X-ray Tomography (Gordon, Bender and Herman, 1970)) also in Emission Tomography, specifically in PET. In theory, ART does not increase the Poisson likelihood, but it produces reasonable images with a high likelihood faster than the EM (in less iterations). Emission Tomography is an ill-conditioned problem; so, if we apply an iteration process in order to obtain high likelihood images, after some number of iterations, the images start deteriorating (Levitan and Herman, 1987). Therefore some kind of regularization is needed. As pointed out in the introduction, this regularization can be obtained by adding a penalization term to (2.1), like in (Levitan and Herman, 1987) and (Herman, Pierro and Gai, 1992). Another, less expensive approach, consists of applying an iterative method that theoretically converges to a solution and to use some stopping procedure, that chooses an iterate before the instabilities show up. We will address the second approach in this paper for the EM and the ART algorithms.

Let us consider the following general problem of finding a solution x of the system of equations

$$f(x) + \epsilon = b, \tag{2.2}$$

where f is a mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$ with components f_i, $(i = 1, \ldots, m)$, b is the m-vector of data $b_i (i = 1, \ldots, m)$ and ϵ an error vector. Although in our case of interest, $f_i(x)$ equals $\langle a_i, x\rangle$, we will use (2.2), because the notation is simpler and our proofs are still valid in the

more general nonlinear case. We assume that ϵ is a random vector satisfying

$$E\{\epsilon\} = 0\,, \quad E\{\epsilon\epsilon^t\} = \mathrm{diag}(E(b_i))\,, \tag{2.3}$$

Conditions (2.3) are satisfied when b is a Poisson distributed random variable; in this case the undetermined system (2.2) can be substituted by

$$\arg\min_x -\frac{1}{m}L(b, f(x))\,, \tag{2.4}$$

where $L(b, f(x)) = \sum_{i=1}^{m}(b_i \log f_i(x) - f_i(x))$ is the log-likelihood. That is, (2.4) is the general form of (2.1).

The crucial question is how to choose the iteration index k, in such a way that the information loss is minimized. In other words, if x_λ is the regularized solution and x the "true solution" of (2.2) (the solution without noise), we need to introduce some measure of the inconsistency, in order to minimize it. The relative entropy, or Kullback-Leibler distance (or better, divergence measure) (Cover and Thomas, 1993) appears as the natural choice because it is intimately associated with Poisson noise, in the same way that the square norm is associated with a Gaussian likelihood. As a matter of fact, maximizing the Poisson likelihood (2.1) is the same as maximizing the Kullback-Leibler distance between the data vector and the matrix range, as it can be easily deduced from the definitions. The relative entropy between two probability mass functions p_i and q_i is defined by

$$\mathrm{KL}(p, q) = \sum_{i=1}^{m}\left(p_i \log \frac{p_i}{q_i}\right). \tag{2.5}$$

(2.5) measures the inefficiency of assuming that the distribution is q when the true distribution is p. When dealing with quantities that are not necessarily probabilities (sum not necessarily equal to one), (2.5) should be substituted by

$$\mathrm{KL}(p, q) = \sum_{i=1}^{m}\left(p_i \log \frac{p_i}{q_i} - p_i + q_i\right). \tag{2.6}$$

Following the previous rationale, a reasonable choice for λ (or k) is the one that solves the problem

$$\arg\min_\lambda \frac{1}{m}E\{\mathrm{KL}(f(x), f(x_\lambda))\} \tag{2.7}$$

where

$$\mathrm{KL}(f(x), f(x_\lambda)) = \sum_{i=1}^{m}\left(f_i(x) \log \frac{f_i(x)}{f_i(x_\lambda)} - f_i(x) + f_i(x_\lambda)\right)$$

is the Kullback-Leibler distance defined by (2.6). The inconvenience of this approach is that we do not know the value of x. So, using the hypotheses (2.3) on the error, we will introduce some approximation to $\mathrm{KL}(f(x), f(x_\lambda))$. We suggest to minimize the following stochastically defined function

$$\mathrm{REKL}(\lambda) = -\frac{1}{m}L(b, f(x_\lambda)) + \frac{w^t \mathrm{diag}(b_i)(A(\lambda)(b+w) - A(\lambda)(b-w))}{2\delta w^t w}) \tag{2.8}$$

where w is a random vector normally distributed, zero mean and variance matrix $I_{m\times m}$, i.e., $w \sim \mathcal{N}(0, I_{m\times m})$, and $A(\lambda)$ is the influence operator defined by

$$A(\lambda)(y) = (\log f_1(x_\lambda), \ldots, \log f_m(x_\lambda))^t, \tag{2.9}$$

for a given data y, and x_λ is an estimator of (2.2) associated with the data b and the regularizing parameter λ.
In the next section we present a theoretical result that justifies the choice of (2.8) as an approximation for $\mathrm{KL}(f(x), f(x_\lambda))$.

3 Main Theoretical Result

The next theorem shows that the function

$$\mathrm{REKL}(\lambda) + \frac{1}{m} L(f(x), f(x))$$

is a nearly unbiased estimator of the Kullback-Leibler distance. That is, the difference between the expected Kullback-Leibler distance and the function defined by (2.8) is approximately a constant (second order) . The Monte-Carlo approximation was inspired by the approximation suggested in (Girard, 1979) to compute the trace.

Theorem 3.1. Let $w = (w_1, \ldots, w_m)^t$ be a vector of random components with normal distribution, that is, $w \sim \mathcal{N}(0, I_{m\times m})$. Suppose that $A(\lambda)(\cdot)$ is a continuously differentiable operator in some open set Ω containing $f(x)$ For each λ we have that

$$\begin{aligned} E(\mathrm{REKL}(\lambda)) + \frac{1}{m} L(f(x), f(x)) \quad &= \quad \frac{1}{m} E(\mathrm{KL}(f(x), f(x_\lambda)) + \\ &+ \frac{1}{m} E(\epsilon^t O_\lambda(\epsilon^t \epsilon)) + \\ &E\left(\frac{w^t \mathrm{diag}(b_i) O_\lambda(\epsilon^t \epsilon + 2\delta|\epsilon^t w| + \delta^2 w^t w)}{w^t w} \right) \end{aligned} \tag{3.1}$$

where $||O_\lambda(y)|| \leq M|y|$, for some $M > 0$.

Proof. Taking into account that

$$\mathrm{KL}(y^1, y^2) = \sum_{i=1}^{m} \left(y_i^1 \log \frac{y_i^1}{y_i^2} - y_i^1 + y_i^2 \right)$$

and

$$L(y^1, y^2) = \sum_{i=1}^{m} \left(y_i^1 \log y_i^2 - y_i^2 \right),$$

we get that

$$\mathrm{KL}(y^1, y^2) = -L(y^1, y^2) + L(y^1, y^1). \tag{3.2}$$

Thus, applying the expectation and considering (3.2), we get that

$$E(L(b, f(x_\lambda)) + \mathrm{KL}(f(x), f(x_\lambda))) = L(f(x), f(x)) + E(L(b, f(x_\lambda)) - L(f(x), f(x_\lambda)) \tag{3.3}$$

and from (2.2), that

$$L(b, f(x_\lambda)) - L(f(x), f(x_\lambda)) = \sum_{i=1}^{m} (\epsilon_i \log f_i(x_\lambda)) \ . \tag{3.4}$$

Let $A(\lambda)(b) = \log f(x) = (\log f_1(x_\lambda), \ldots, \log f_m(x_\lambda))^t$. Then

$$L(b, f(x_\lambda)) - L(f(x), f(x_\lambda)) = \epsilon^t A(\lambda)(f(x) + \epsilon) \tag{3.5}$$

Expanding the operator $A(\lambda)(\cdot)$ at $f(x) + \epsilon$, $f(x) + \epsilon + \delta w$ and $f(x) + \epsilon + \delta w$ at the point $f(x)$ using Taylor's formula gives:

$$A(\lambda)(f(x) + \epsilon) = A(\lambda)(f(x)) + DA(\lambda)(f(x))(\epsilon) + O_\lambda(\epsilon^t \epsilon)\,, \tag{3.6}$$

$$A(\lambda)(f(x) + \epsilon + \delta w) = A(\lambda)(f(x)) + DA(\lambda)(f(x))(\epsilon + \delta w) + O_\lambda[(\epsilon + \delta w)^t(\epsilon + \delta w)] \tag{3.7}$$

and

$$A(\lambda)(f(x) + \epsilon - \delta w) = A(\lambda)(f(x)) + DA(\lambda)(f(x))(\epsilon - \delta w) + O_\lambda[(\epsilon - \delta w)^t(\epsilon - \delta w)], \tag{3.8}$$

where $DA(\lambda)(z)$ denotes the Jacobian of the operator at z. Thus, from (3.7) and (3.8) we obtain the equality

$$A(\lambda)(b + w) - A(\lambda)(b - w) = 2\delta DA(\lambda)(f(x))w + O_\lambda(\epsilon^t\epsilon + 2\delta|\epsilon^t w| + \delta^2 w^t w) \tag{3.9}$$

and using (3.6), (3.9) and applying the expectation we get that

$$\begin{aligned} E\left(\frac{1}{m}\epsilon^t A(\lambda)(f(x) + \epsilon) - \frac{w^t \mathrm{diag}(b_i)(A(\lambda)(b + w) - A(\lambda)(b - w))}{2\delta w^t w}\right) \ = \\ \frac{1}{m}E(\epsilon^t DA(\lambda)(f(x))\epsilon) - E(\frac{w^t \mathrm{diag}(b_i) DA(\lambda)(f(x))w}{w^t w}) + \\ + \frac{1}{m}E(\epsilon^t O_\lambda(\epsilon^t\epsilon)) + E(\frac{w^t \mathrm{diag}(b_i) O_\lambda(\epsilon^t\epsilon + 2\delta|\epsilon^t w| + \delta^2 w^t w)}{w^t w}) \end{aligned} \tag{3.10}$$

But from (2.3), an easy calculation gives that

$$E(\epsilon^t DA(\lambda)(f(x))\epsilon) = \mathrm{Tr}(\mathrm{diag}(E(b_i))DA(\lambda)(f(x))) \tag{3.11}$$

and taking into account that $w \sim \mathcal{N}(0, I_{m\times m})$ we have that

$$E(\frac{w^t \mathrm{diag}(b_i) DA(\lambda)(f(x))w}{w^t w}) = \frac{1}{m}\mathrm{Tr}(\mathrm{diag}(E(b_i))DA(\lambda)(f(x)))\,, \tag{3.12}$$

annihilating the 1st and 2nd terms in (3.10).
The result follows from (3.3), (3.5), (3.10), (3.11) and (3.12). □

The theorem above provides a theoretical foundation to the fact that minimizing $\mathrm{REKL}(\lambda)$ gives a good approximation, in the average, to the minimum of $\mathrm{KL}(f(x), f(x_\lambda))$, also in the average, taking into account that $\frac{1}{m}L(f(x), f(x))$ is constant in λ. The approximation is second order with respect to the error contained in the second term of (3.1) and first order times the small parameter δ in the third term. In the next section, this fact will be used to develop a stopping criterion for ART and the EM algorithms.

4 Application to Positron Emission Tomography

We consider in this section two very well known algorithms, that have been applied to Emission Computed Tomography (ECT). The first is the Expectation Maximization (EM) algorithm, whose application in ECT was first suggested by Shepp and Vardi (Shepp and Vardi, 1982) and, since then, became an important tool in the area. For a given a positive starting image $x^{(0)}$, the EM sequence is defined by

$$x_j^{(k+1)} = \frac{x_j^k}{W_{jj}} \sum_{i=1}^{m} \frac{a_{ij} b_i}{\sum_{l=1}^{n} a_{il} x_l^k} \tag{4.1}$$

where $W_{jj} = \sum_{i=1}^{m} a_{ij}$, being $K = [a_1, \ldots, a_m]^t = \{a_{ij}\}$ the projection matrix (Usually, a_{ij} is taken as the intersection of the i-th ray with the j-th pixel).

The second algorithm to be tested is the Algebraic Reconstruction Technique (ART), originally proposed for X-ray Computed Tomography (CT) (Herman, 1980), (Gordon et al., 1970). ART was applied in Positron Emission Tomography by Herman and Meyer (Herman and Meyer, 1993), achieving very high likelihood values in few iterations. So, we expect, also in this case, the chosen number of iterations, to be lower than the EM. It is defined by

$$\begin{cases} x^{(k+1,1)} = x^k \\ x^{(k+1,i+1)} = x^{(k+1,i)} + \omega \frac{(b_i - a_i^t x^{(k+1,i)})}{||a_i||^2} a_i \\ x^{k+1} = x^{(k+1,m+1)}, \end{cases} \tag{4.2}$$

for $i = 1, \ldots, m$; where ω is a relaxation parameter in the interval $(0, 2)$.

Let us now describe how to compute (2.8) for $\lambda = k$. For both algorithms, following (2.9), the influence operator at y is

$$A(k)(y) = (\log(a_1^t x^{(k)}), \ldots, \log(a_m^t x^{(k)})). \tag{4.3}$$

That is, the operator at the point y and iteration k, is computed by applying the matrix A to the current k-th iterate starting with data y and taking logarithms of the resulting row scalar products. For the computation of the second term in (2.8), let x^{k+} be the estimator that corresponds to the data vector $b + \delta w$, where δ is a small number, x^{k-} the one that corresponds to the vector $b - \delta w$, and w randomly generated. After a little trial and error, we found that δ of the order of 10^{-4} worked well. In order to simplify notation, we define

$$\log(A(x^{k+})/A(x^{k-})) = (\log(a_1^t x^{(k+)}/a_1^t x^{(k-)}), \ldots, \log(a_m^t x^{(k+)}/a_m^t x^{(k-)}))$$

and (2.8) becomes

$$\mathrm{REKL}(k) = -\frac{1}{m} L(b, Ax^k) + \frac{w^t \mathrm{diag}(b_i) \log(A(x^{k+})/A(x^{k-}))}{2\delta w^t w}). \tag{4.4}$$

Therefore, the computational task for $\mathrm{REKL}(k)$ requires the application of the algorithm for three different data sets: b for the first term and the perturbed data $b + \delta w$ and $b - \delta w$ for the second term (we observe that this can be done in parallel).

Remark: We have used second order differences for the approximation of the derivative because we have preferred better accuracy, in spite of the fact that first order approximations reduce the complexity by $1/3$. In order to prevent possible problems with the reconstruction algorithms, another w was to be selected in case of $b - \delta w < 0$, although, this is quite unlikely to happen.

We performed a large number of experiments, simulating typical PET reconstructions of emission activity inside the brain. Phantoms were digitized using a 95x95 pixels grid and projection data generated assuming that the scanner consists of a ring of 300 detectors with each detector in coincidence with 101 detectors opposite to it, a total number of m = 30292 equations (8 rays were not considered because of their lack of intersection with the image region). Poisson noise was introduced into measurements and two extreme situations considered: a relatively low noise corresponding to a high photon count number of 2,022,085 and a higher noise level corresponding to a total count of 495,609 photons. For the experiments we used SNARK93, a reconstruction software package developed by the Medical Image Processing Group of the University of Pennsylvania (Browne, Herman and Odhner, 1993). The relaxation parameter for ART, was $\omega = 0.025$ (at this point, we observe that this parameter can be also considered as a variable and chosen by using the same functional, but with a two dimensional minimization procedure).

The results of a typical run are shown in figures 1 and 2 for the EM algorithm and ART respectively. We observe that, for ART, the dashed line, corresponding to the proposed functional approximates very accurately the shape of the function whose maximum is the criterion for choosing the parameter (thick line). For the EM, this approximation is less accurate, probably because of the accumulation of errors, due to the larger number of iterations. Although not so clear visually, the detection of the maximum is numerically very easy. For the EM, the number of iterations for high and low photon counts was respectively around 35 and 23. For ART, 7 and 4. Also, as expected, for higher noise levels (low photon counts), the maxima of the curves are better visualized and the chosen number of iterations is of course smaller, due to the earlier appearance of the high frequencies generated by noise.

For the same data, we performed the experiments using the χ^2 stopping criterion (Hebert, 1990). The results are shown in figures 3 and 4.

The tables below show the iteration values (the location of the minimum) for three different PET phantoms, for the standard distance, our functional and the chi-squared. The chi-squared method consistently chooses values closer (and in general lower) to the distance, producing excessively smooth images.

TABLE 1- ART ($5 \cdot 10^5$ counts)

Phantom	$\min \|\|x^k - x^{\text{true}}\|\|$	REKL	χ_p^2
1a	2	4	2
2a	4	4	2
3a	4	4	2

TABLE 2- EM ($5 \cdot 10^5$ counts)

Phantom	$\min \|\|x^k - x^{\text{true}}\|\|$	REKL	χ_p^2
1a	12-16	23	10
2a	11-14	20	10
3a	9-16	20	9

TABLE 3- ART ($2 \cdot 10^6$ counts)

Phantom	$\min \|\|x^k - x^{\text{true}}\|\|$	REKL	χ_p^2
1a	4	7	5
2a	3	7	5
3a	3	7	5

TABLE 4- EM ($2 \cdot 10^6$ counts)

Phantom	$\min \|\|x^k - x^{\text{true}}\|\|$	REKL	χ_p^2
1a	20-24	35	19
2a	18-20	35	17
3a	18	33	19

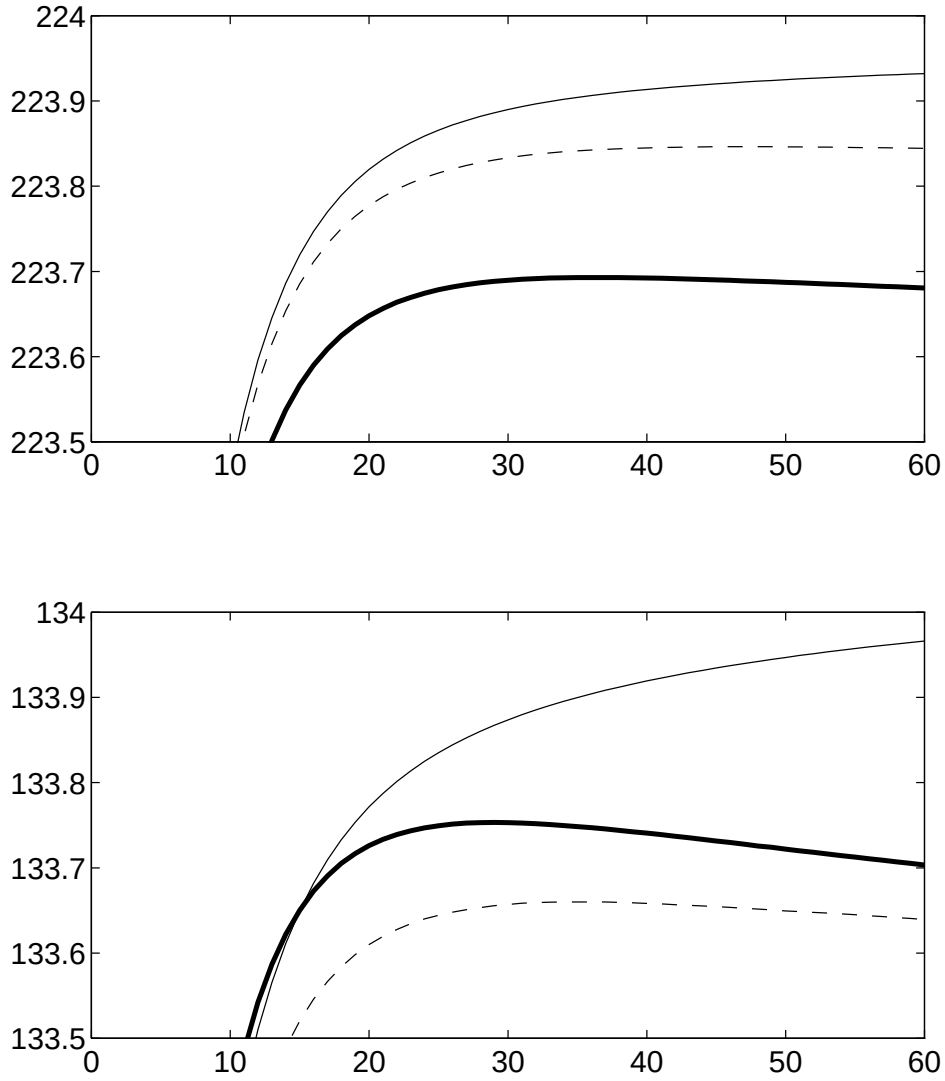

Figure 1: The EM Algorithm. Total photon count number is 2,022,085 (top) and 495,609 (botton). The solid thin line corresponds to $\frac{1}{30292}L(b, Ax^k)$, the thick one $\frac{1}{30292}(-\mathrm{KL}(Ax, Ax^k) + L(Ax, Ax))$; the dashed line to $-\mathrm{REKL}(k)$.

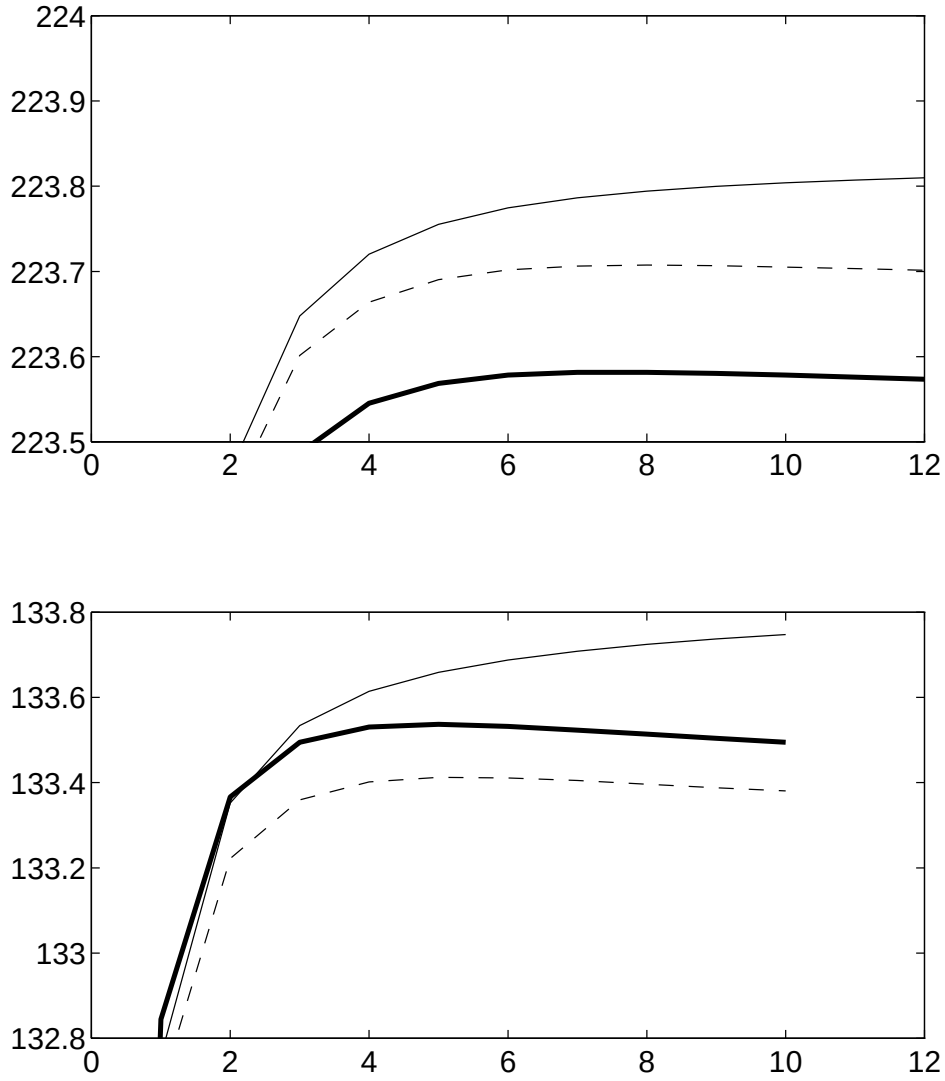

Figure 2: The ART Algorithm. Total photon count number is 2,022,085 (top) and 495,609 (bottom). The solid thin line corresponds to $\frac{1}{30292}L(b, Ax^k)$, the thick one $\frac{1}{30292}(-\mathrm{KL}(Ax, Ax^k)+ L(Ax, Ax))$; the dashed line to $-\mathrm{REKL}(k)$.

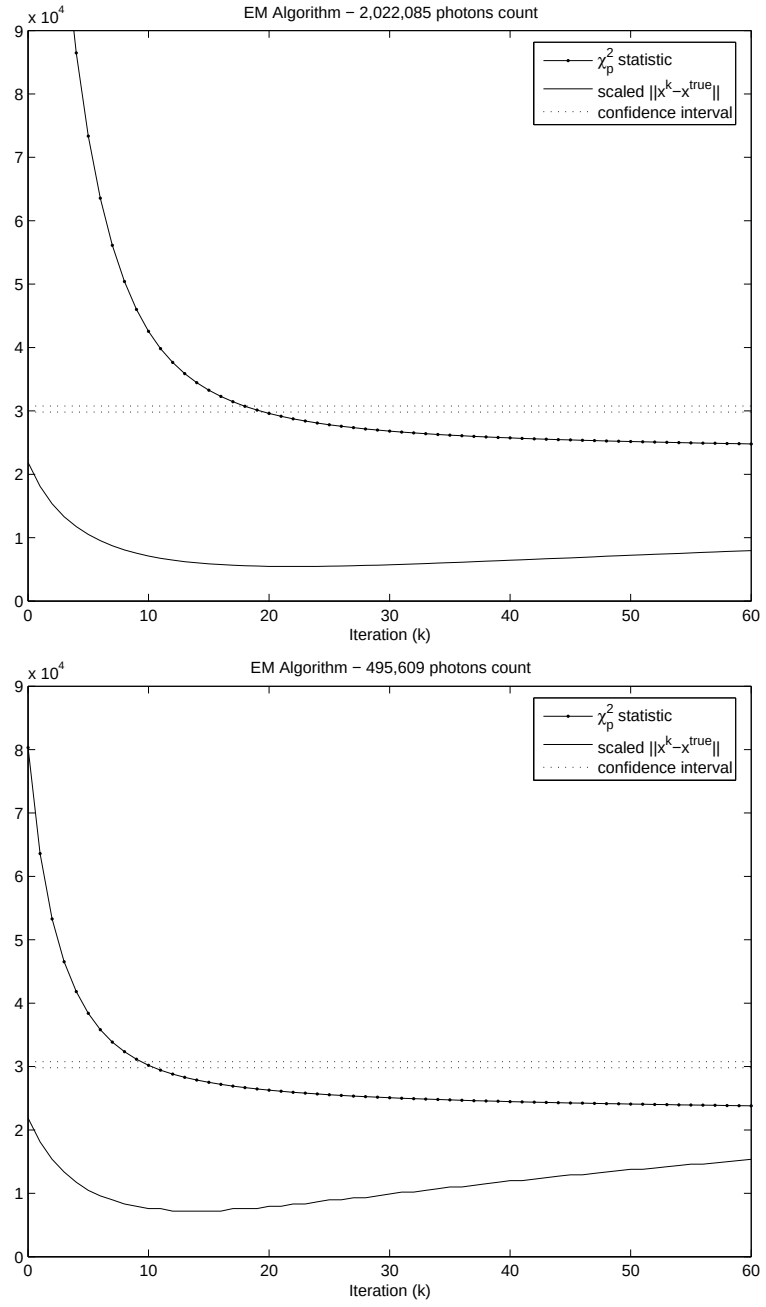

Figure 3: The EM Algorithm. Total photon count number is 2,022,085 (top) and 495,609 (botton). The solid thin line corresponds to scaled $||x^k - x^{\text{true}}||$ and the dashed line to χ^2_p statistic.

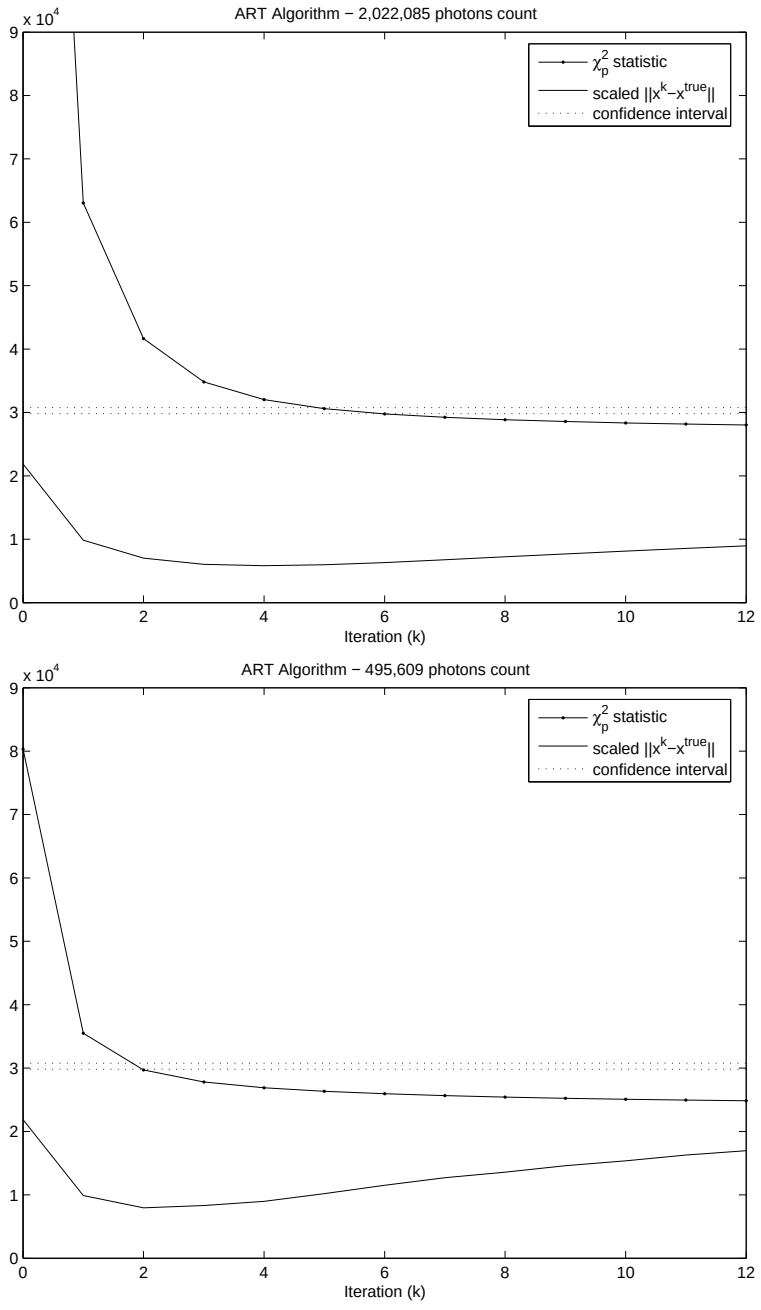

Figure 4: The ART Algorithm. Total photon count number is 2,022,085 (top) and 495,609 (bottom). The solid thin line corresponds to scaled $||x^k - x^{\text{true}}||$ and the dashed line to χ^2_p statistic.

Our experiments have demonstrated that the proposed method is robust with respect to the random vector w. Figures 5 and 6 illustrate how consistently the second term of (2.8) approximates $\frac{1}{m}\mathrm{Tr}(\mathrm{diag}(b_i)DA(k)(b))$. The terms containing the random approximation of the trace (equality (3.12)) are almost the same along the iterations for every choice of w (we show only five).

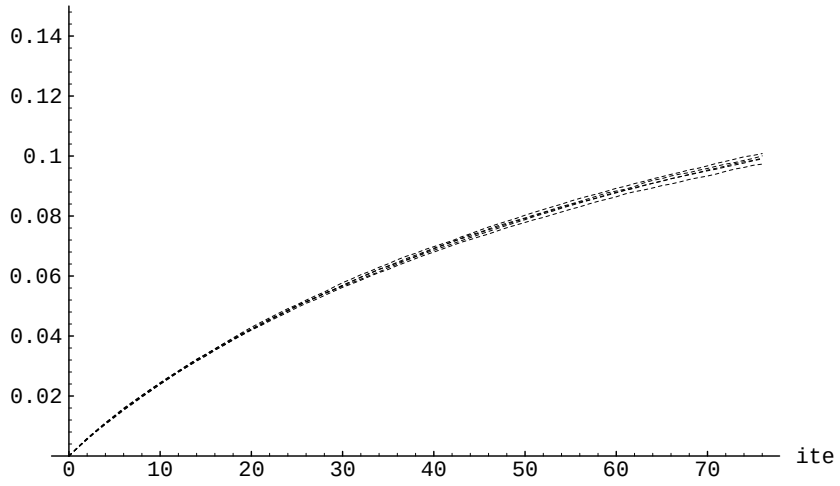

Figure 5: Five estimates of $\frac{1}{30292}\mathrm{Tr}(\mathrm{diag}(b_i)DA(k)(b))$ for the EM algorithm

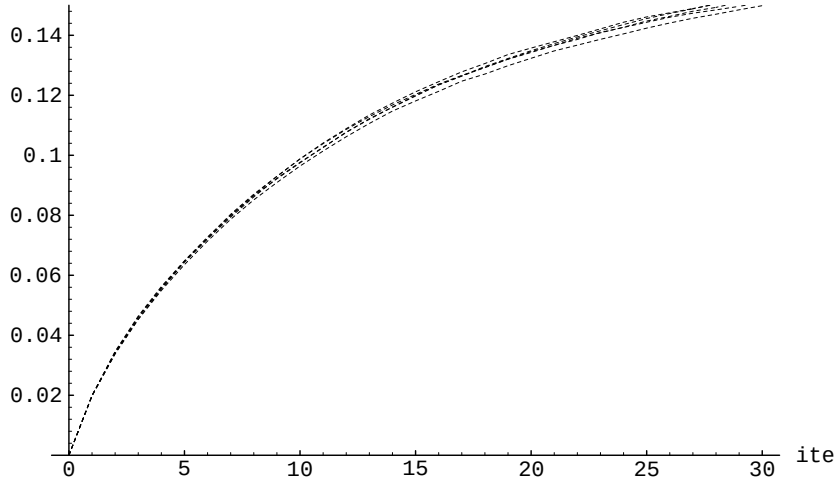

Figure 6: Five estimates of $\frac{1}{30292}\mathrm{Tr}(\mathrm{diag}(b_i)DA(k)(b))$ for the ART algorithm with $\omega = 0.025$

Finally, figures 7 and 8 show reconstructions obtained by both algorithms at different iterations.

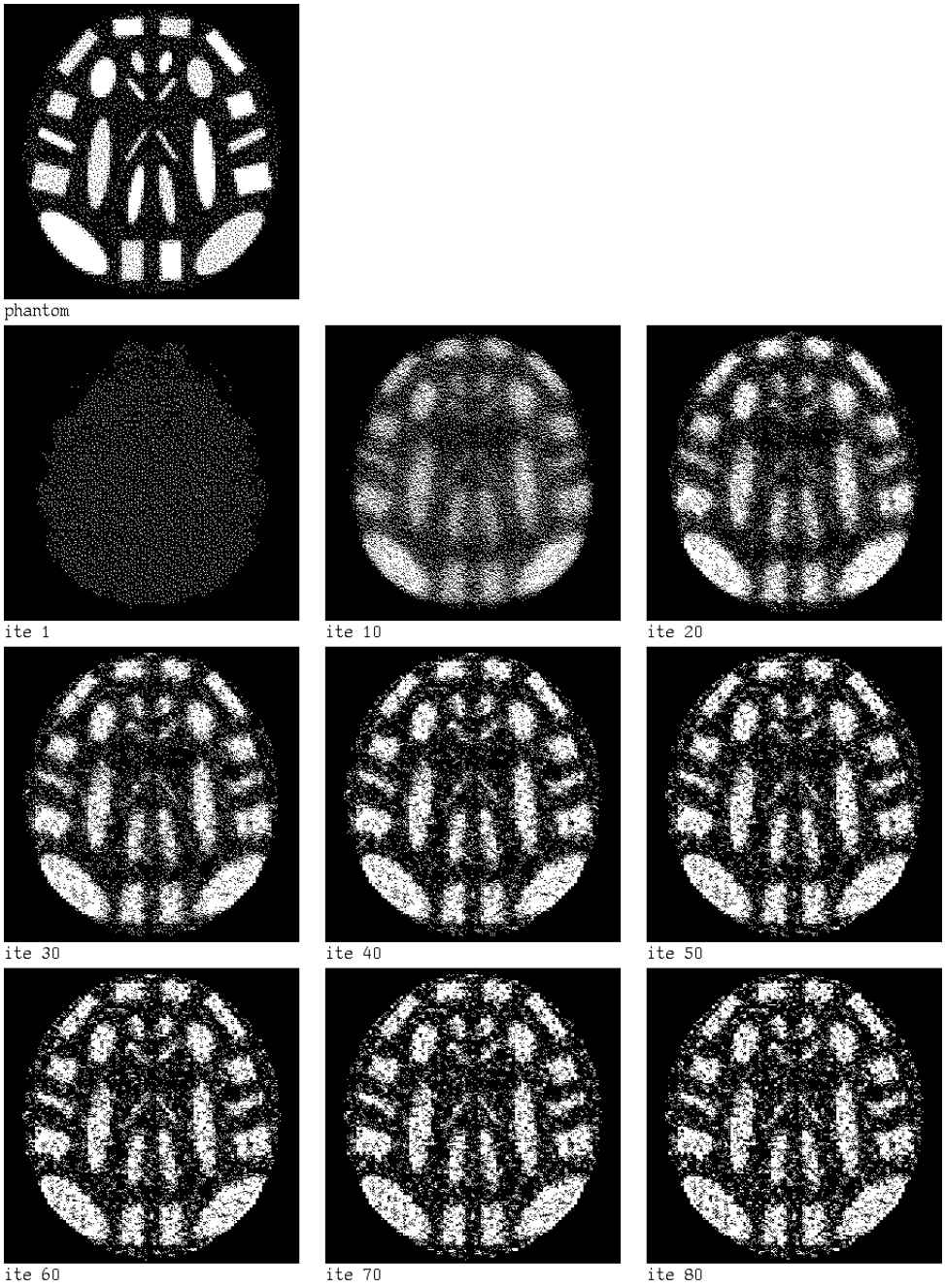

Figure 7: A 95×95 phantom and its reconstructions using the EM algorithm

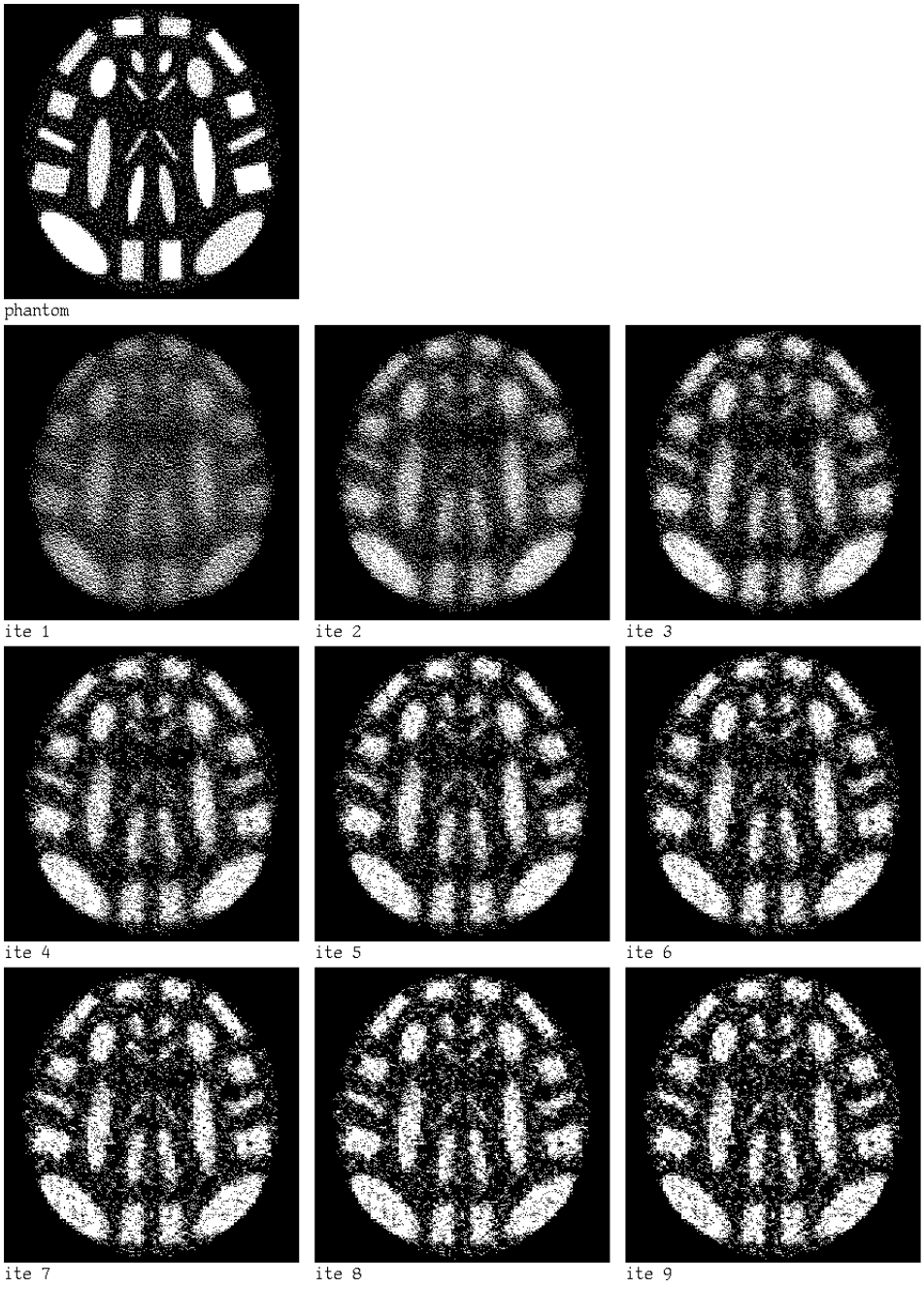

Figure 8: A 95×95 phantom and its reconstructions using ART with $\omega = 0.025$

5 Concluding Remarks

We have presented a functional that gives rise to a new method for choosing parameters based on the Kullback-Leibler divergence as a measure of consistency. As stated before, the main goal was to introduce this method as a general tool when dealing with ill-posed inverse problems with Poisson data arising in image processing. We tested the method as an automatic way to stop iterations for two very well known algorithms in Emission Tomography. Our experimental results are consistent with what expected, suggesting that this approach can be an alternative for problems with similar noise properties. The method is stable, the minima numerically easy to be determined, and, when compared to the chi-squared method, it produces images with less smoothing. For future research it would be important to check the validity of the proposed criterion by comparing the results for several algorithms, using different figures of merit and a more statistical methodology, as the one proposed in (Herman and Meyer, 1993). It is also possible to apply this new automatic stopping criterion to EM-like block projection methods, (Hudson and Larkin, 1994), (Browne and Pierro, 1996), that have become very popular in research and in commercial applications in recent years.

Acknowledgements. We are grateful to J. Browne and G. T. Herman for their support on the use of SNARK93.

References

Archer, G. and Titterington, D. M. 1995. On some bayesian/regularization methods for image restoration, *IEEE Trans. Imag. Proc.* **4**(7): 989–995.

Browne, J. A., Herman, G. T. and Odhner, D. 1993. SNARK93 - a programming system for image reconstruction from projections, *Technical Report MIPG198*, Department of Radiology, University of Pennsylvania.

Browne, J. A. and Pierro, A. R. D. 1996. A row-action alternative to the em algorithm for maximizing likelihoods in emission tomography, *IEEE Transactions on Medical Imaging* **15**(4): 687–699.

Budinger, T. F., Gullberg, G. T. and Huesman, R. H. 1979. Emission computed tomography, *in* G. T. Herman (ed.), *Image Reconstruction from Projections: Implementation and Applications*, Springer Verlag, New York Berlin Heidelberg Tokyo.

Coakley, K. J. 1991. A cross-validation procedure for stopping the em algorithm and deconvolution of neutron depth profiling spectra, *IEEE Trans. on Nuclear Science* **38**(1): 9–15.

Cover, T. M. and Thomas, J. A. 1993. *Elements of Information Theory*, John Wiley, New York.

Craven, P. and Wahba, G. 1979. Smoothing noisy data with spline functions, *Numer. Math.* **31**: 377–403.

Dempster, A. P., Laird, N. M. and Rubin, D. D. 1977. Maximum likelihood from incomplete data via the em algorithm, *J. Roy. Stat. Soc., Series B* **39**: 1–38.

Girard, D. 1979. A fast 'monte-carlo cross-validation' procedure for large least squares problems with noisy data, *Numer. Math* **56**: 1–23.

Gordon, R., Bender, R. and Herman, G. T. 1970. Algebraic reconstruction techniques (art) for three-dimensional electron microscopy and x-ray photography, *J. Theo. Biol.* **29**: 471–481.

Hebert, T. J. 1990. Statistical stopping criteria for iterative maximum likelihood reconstruction of emission images, *Phys. Med. Biol.* **35**(9): 1221–1232.

Hebert, T. J. and Leahy, R. 1992. Statistic-based map image reconstruction from poisson data using gibbs priors, *IEEE Trans. Sig. Proc.* **40**(9): 2290–2303.

Herman, G. T. 1980. *Image Reconstruction from Projections: The Fundamentals of Computerized Tomography*, Academic Press, New York.

Herman, G. T. and Meyer, L. B. 1993. Algebraic reconstruction techniques can be made computationally efficient, *IEEE Trans. Med. Imaging* **12**: 600–609.

Herman, G. T., Pierro, A. R. D. and Gai, N. 1992. On methods for maximum a posteriori image reconstruction with a normal prior, *Journal of Visual Comm. Image Rep.* **3**(4): 316–324.

Higdon, D. M., Bowsher, J. E., Johnson, V. E., T. G. Turkington, D. R. G. and Jaszczak, R. J. 1997. Fully bayesian estimation of gibbs hyperparameters for emission for emission computed tomography data, *IEEE Trans. Med. Imag.* **16**(5): 516–526.

Hudson, H. M. and Larkin, R. S. 1994. Accelerated image reconstruction using ordered subsets of projection data, *IEEE Trans. Med. Imaging* **13**(4): 601–609.

Lange, K. and Carson, R. 1987. Em reconstruction algorithms for emission and transmission tomography, *J. Comput. Assisted Tomog.* **8**: 306–316.

Levitan, E. and Herman, G. T. 1987. A maximum a posteriori probability expectation maximization algorithm for image reconstruction in emission tomography, *IEEE Trans. Med. Imaging* **6**(3): 185–192.

Llacer, J. and Veklerov, E. 1989. Feasible images and practical stopping rules for iterative algorithms in emission tomography, *IEEE Trans. Med. Imaging* **8**(2): 186–193.

Perry, K. M. and Reeves, S. J. 1994. Generalized cross-validation as a stopping rule for the richardson-lucy algorithm, *in* R. J. Hanisch and R. L. White (eds), *The Restoration of HST Images and Spectra II*, Space Telescope Science Institute.

Reeves, S. J. 1995. Generalized cross-validation as a stopping rule for the richardson-lucy algorithm, *International J. Imaging Syst. and Tech.* **6**: 387–391.

Rockmore, A. and Macovski, A. 1976. A maximum likelihood approach to emission image reconstruction from projections, *IEEE Trans. Nucl. Sci.* **23**: 1428–1432.

Santos, R. J. 1996. Equivalence of regularization and truncated iteration for general ill-posed problems, *Linear Algebra Appl.* **236**: 25–33.

Shepp, L. A., Vardi, Y., Ra, J. B., Hilal, S. K. and Cho, Z. 1984. Maximum likelihood pet with real data, *IEEE Trans. Nucl. Sci.* **31**: 910–913.

Shepp, L. and Vardi, Y. 1982. Maximum likelihood reconstruction for emission tomography, *IEEE Trans. Med. Imaging* **48**: 1135–1149.

Ter-Pogossian, M. M., Raichle, M. and Sobel, B. E. 1980. Positron emission tomography, *Scientific Amer.* **243**(4): 170–181.

Thomson, A. M., Brown, J. C., Kay, J. W. and Titterington, D. M. 1991. A study of methods of choosing the smoothing parameter in image restoration by regularization, *IEEE Trans. Pattern Anal. Machine Intell.* **13**: 326–339.

Vogel, C. 2002. *Computational methods for inverse problems*, SIAM, Philadelphia.

Wahba, G. 1987. Three topics in ill-posed problems, *in* H. Engl and C. Groetsch (eds), *Inverse and Ill-Posed Problems*, Academic Press, New York.

Zhou, Z., Leahy, R. M. and Qi, J. 1997. Approximate maximum likelihood hyperparameter estimation for gibbs priors, *IEEE Trans. Imag. Proc.* **6**(6): 844–861.