Activity and attenuation recovery from activity data only in emission computed tomography

ALVARO R. DE PIERRO* and FABIANA CREPALDI**

State University of Campinas, Department of Applied Mathematics

CP. 6065, 13081-970 Campinas, SP, Brazil

E-mail: alvaro@ime.unicamp.br

Abstract. We describe the continuous and discrete mathematical models for Emission Computed Tomography (ECT) and the need for attenuation correction. Then we analyse the problem of retrieving the attenuation directly from the emission data, nonuniqueness and its consequences. We present the existing and new approaches for solving the problem. Methods are compared and illustrated by numerical simulations. The presentation focuses on Positron Emission Tomography, but we also discuss the extensions to Single Photon Emission Computed Tomography.

Mathematical subject classification: 44A12, 65R30, 65R32.

Key words: Attenuated Radon Transform, maximum likelihood, emission tomography.

1 Introduction

Emission Computed Tomography (ECT), in its two modalities, SPECT and PET (Single Photon and Positron Emission), is the main tool to understand physiological processes inside the body at a molecular level. The pure mathematical model (no attenuation, no scattering, high statistics, etc) for ECT is the Radon transform, the same as for X-ray Computed Tomography (CT). In this article we describe the main issues concerning the introduction of attenuation in the ECT model and the different approaches to the direct computation of the activity and

the attenuation maps from emission data only. Our emphasis will be on iterative methods and Maximum Likelihood models for PET, although, it is worth mentioning that each method for PET generates a similar approach for SPECT and viceversa. We discuss the nonuniqueness issue for PET and present some new methods for solving the problem as well as numerical simulations. Open problems and research directions are discussed in Section 5.

1.1   *What is Emission Computed Tomography (ECT)*

In ECT we are trying to understand a physiological process, like glucose consumption, blood flow, oxidative metabolism, neurotransmission, and many others (see [17] for a long list for PET). How this is done? There is a compound that is metabolized during that physiological process, and that could be tagged using a radioative isotope. In the case of PET, the isotope is artificial and decays emitting positrons. After a very short path, the positron meets an electron, generating two photons in almost opposite directions, that are detected by pairs of detectors. Each pair of detectors determines a line and the data is the total number of coincidences counted along all possible lines (Figure 1 shows a PET scanner with its ring of detectors).

We aim at reconstructing the emission density that is correlated with the process being studied. The most common example in PET is glucose, that is tagged with F_{18}, giving rise to FDG (fluorodeoxyglucose). This measures glucose consumption, an indirect description of many processes for clinical diagnosis (malignancy, heart failure, etc) as well as medical reserach (brain function, blood flow, etc). In the case of SPECT [5], the isotope is natural and decays by generating a single photon (see Figure 2), so, the line is determined by using a collimator.

Many photons are absorbed by the collimator before detection, thus obtaining a very low statistics, poorer than PET.

1.2   *A continuous 'pure' mathematical model*

In a continuous setting, we aim at reconstructing a function $f(x)$ representing the mean emission density at the point x given the integrals along lines L, that is,

$$\int_L f \, ds \, = \, d_L \qquad (1.1)$$
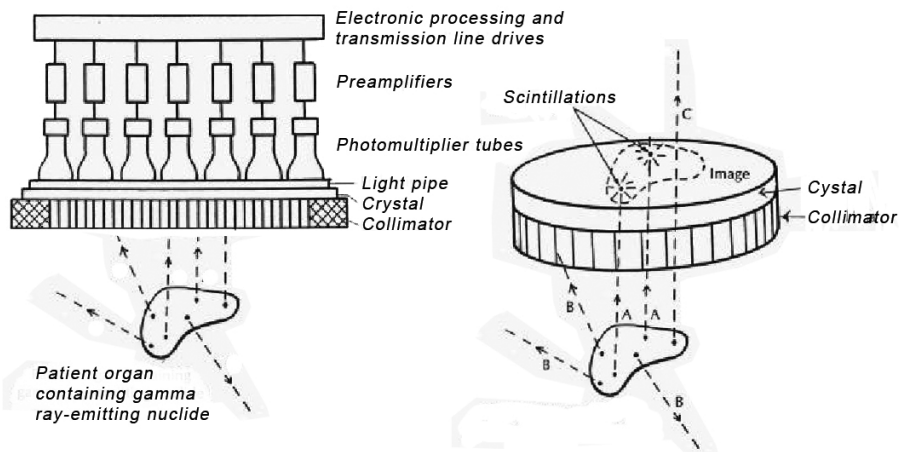
Figure 1 – PET scanner.

Figure 2 – SPECT data acquisition.

where d_L is the mean number of coincidences detected by the pair of detectors (or by a single detector and its collimator for SPECT) that determine L. Equation (1.1) means that the 'pure' problem reduces to invert the well known Radon Transform [13]

$$f \xrightarrow{\mathcal{R}} \int_L f \, ds = p(l, \theta) \tag{1.2}$$

1.2.1   The difference between ECT and Computed Tomography (CT)

We observe that the inversion above is the basic model also for Transmission Computed Tomography (CT), where d_L, instead of number of coincidences stands for $log \; \frac{p_L^e}{p_L^d}$, being in this case p_L^e and p_L^d the mean number of photons emitted by a source and the detected ones respectively.

Inversion formulas for (1.2) can be easily obtained using the Fourier Projection Theorem (see [13]). However, the higher levels of noise in ECT, as compared to CT, made models that take into account noise characteristics and the use of iterative methods more reliable, producing better images [25]. This leads to to the necessity of discretization, in order to develop Maximum Likelihood (ML) models.

1.3   A discrete model

We consider now the discretized two-dimensional PET model. In what follows, x_j denotes the expected number of coincidences per unit area in the jth pixel $(j = 1, \ldots, n)$ (we are assuming here that the "tubes" are of constant cross-section and that dividing the radioactivity concentration per unit volume by the area of this constant cross section gives us the "number of coincidences per unit length"); x represents the n-dimensional vector whose jth component is x_j and will be referred to as the image vector (see Figure 3).

Suppose we count coincidences along m lines. We denote by g the m-dimensional vector whose ith element, g_i $(g_i \geq 0)$, is the number of coincidences which are counted for the ith line $(i = 1, \ldots, m)$ during the data collection period; we shall refer to g as the measurement vector. If a_{ij} $(a_{ij} \geq 0)$ denotes the probability that a positron emitted from pixel j results in a coincidence at the i-th line of response (in this work, as usual, this probability is
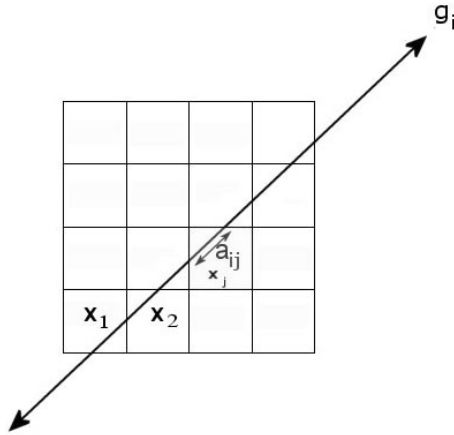
Figure 3 – Discretization.

approximated by the weighted length of the intersection), then g_i is a sample of a Poisson distribution whose expected value is

$$\langle a^i, x \rangle = \sum_{j=1}^{n} a_{ij} x_j \qquad (1.3)$$

where $\langle , \rangle$ is the standard inner product and a^i is the ith column of the transpose A^T of the $m \times n$ projection matrix $A = (a_{ij})$ [28].

The probability of obtaining the measurement vector g if the image vector is x (i.e., the likelihood function) is

$$P_L(\mathbf{g}|x) = \prod_{i=1}^{m} \left[\frac{\langle a^i, x \rangle^{g_i}}{g_i!} \exp(-\langle a^i, x \rangle) \right]. \qquad (1.4)$$

The reconstruction problem is to estimate the image vector x given the data measurements $\mathbf{g}$.

As mentioned before one approach to this problem, first suggested in [24], is the ML method which estimates the x that maximizes $P_L(\mathbf{g}|x)$ subject to nonnegative constraints on x, or, equivalently, finds the $x \geq 0$ which maximizes

$$L(x) = \sum_{i=1}^{m} g_i \log\langle a^i, x \rangle - \langle a^i, x \rangle. \qquad (1.5)$$

In [26] and [16], the Expectation Maximization (EM) algorithm was proposed for the maximization of (1.5). This results in an iterative algorithm, which, starting with a strictly positive vector $x^{(0)}$, successively estimates the image vector by

$$x_j^{(k+1)} = \frac{x_j^{(k)}}{\sum_{i=1}^{m} a_{ij}} \sum_{i=1}^{m} \frac{a_{ij} g_i}{\langle a^i, x \rangle}, \quad j = 1, \ldots, n \qquad (1.6)$$

The EM algorithm's convergence rate is extremely slow and a significant amount of computation is often needed to achieve an acceptable image, (essentially, applying filtered backprojection costs one iteration of the EM, and, in general, at least fifty iterations are required). In the 90's faster methods, like ordered subsets [14] and RAMLA [4], made possible the adoption of iterative methods by commercial scanners.

2 The need for attenuation correction: attenuated transforms, problems

In practice, a more accurate model for ECT should be considered because of the attenuation effect, that is, the fact that a substantial number of photons is absorbed by the tissue, and the formulas in (1.1) should be substituted by (see [21])

$$e^{-\int_L \mu(\eta)\, d\eta} \int_L f(x)\, dx \ = d_L \qquad (2.7)$$

for PET and

$$\int_L f(x)\, e^{-\int_L \mu(x+t\theta^\perp)dt}\, d\eta dx \ = d_L \qquad (2.8)$$

for SPECT, where μ is now the attenuation density and d_L as before. Considering that the previous formulas are now strongly nonlinear, the usual way to compensate for atenuation is to estimate it by a previous (or simultaneous) CT scan for estimating μ. An important improvement could be achieved if it is possible to retrieve the attenuation map directly from the activity data (meaning less time, less patient motion, automatic registration, no error amplification, etc). Recently, PET scanners coupled with CT scanners, called PET-CT, have been developed, overcoming the problem (at a very high cost). However, still many of the existing scanners are not PET-CT, and there is not such a capability in most of the gamma-camera based systems [22] with coincidence detection

(SPECTscanners adapted for coincidence detection), and in small high resolution tomographs for research purposes [19].

Methods trying to perform this task, without CT corrections, tend to retrieve solutions that generate an undesired cross talk between the activity and attenuation images, unless a considerable amount of information about attenuation values is assumed as in [20]. That is, visually, the main drawback is that the shadow of the attenuation appears in the emission image and vice versa. This is a consequence of nonlinearity and nonuniqueness of the inversion of the operators just described. In the following we analyse the problem and we describe the existing and new approaches for its solution, now restricted to PET ((2.7)).

3   Recovering the attenuation from activity data only: nonuniqueness and different approaches

It is very easy to show that, for data generated by circularly symmetric activity and attenuation images, the inversion in (2.7) has an infinite number of solutions. Also, examples of the influence of 'approximate' nonuniqueness could be found in [2]. In any case, methods based on consistency of the data, first proposed by Natterer [20] have been the most effective ones, producing the best images. In the following μ and λ will stand for the images corresponding to the attenuation and the activity values respectively. The results of the numerical experiments corresponding to the methods described in **3.2** and **3.3** are described in Section 5.

3.1   Natterer

For the sake of completeness, and because it was the first to introduce in the problem the necessity of data consistency, we briefly describe in this section Natterer's approach [20]. As we said before, this approach relies heavily on the knowledge of additional information on the attenuation, but, on the other hand, it is the point of departure of the other successful approaches based on consistency. From the Helgason-Ludwig conditions for PET data [11], [18], if we take the Fourier Transform of the m-th moment, all coefficients greater than m should be

zero, that is

$$\int_0^{2\pi} \int_{-\infty}^{+\infty} s^m e^{-ik\theta} e^{T(s,\theta)} g(s,\theta) \, ds d\theta = 0, \quad \forall k > m \in \mathcal{Z}_{\geq 0} \qquad (3.9)$$

and $T(s,\theta)$ is the Radon Transform of μ. g is the projection data as before. (3.9) determines a set of equations to be satisfied by the attenuation. However, these equations are not enough to completely determine μ. So, the idea is to use additional information in order to reduce the indetermination. To do this, Natterer et al. [29] consider that the sought attenuation is a shifted version of a given one. That is, they consider $\mu(x) = \mu_0(Ax + a)$ where A is a 2x2 matrix and a a 2-vector. If the attenuation is not too different from μ_0 sometimes it works fine.

3.2   Bronnikov

In this Section we present a method for the attenuation recovery from PET emission data, that follows closely Bronnikov's approach for SPECT. Our implementation uses known properties of pseudoinverses, allowing a relatively low cost reconstruction, as compared to Bronnikov's implementation for SPECT. We show experimentally that, even without regularization, the obtained images are stable. Following [7, 6], the discretization of the attenuated Radon transform for PET (2.7) can be described by the equation:

$$A_\mu \lambda = \mathbf{g} \,, \qquad (3.10)$$

where $A_\mu \in \mathbb{R}^{m \times n}$ represents the projection matrix, $\mu \in \mathbb{R}^n$, the attenuation vector, $\lambda \in \mathbb{R}^n$, the activity, and $\mathbf{g} \in \mathbb{R}^m$ the emission data vector as before. As in [7], μ appears as a matrix parameter. The elements of the matrix $A_\mu = (a_{ij})$ are now defined by:

$$a_{ij} = \sigma_{ij} e^{-\sum_k \mu_k s_{ik}} \qquad (3.11)$$

where σ_{ij} and s_{ik} stand for the intersection length of the ith ray with the jth and kth pixels of activity and attenuation, respectively. In our approach, for the sake of simplicity, we consider the same discretization for both, activity and attenuation.

The number of equations m in (3.10) should be greater than or equal to the number of variables, that is, the number of attenuation voxels (l) plus the number of activity voxels (n): $m \geq n + l$. In the particular case where the images have the same resolution, ($l = n$), Bronnikov observed experimentally that the choice $m = 3n$ is the more efficient one. For a general Hilbert space H, let $\overline{H}$ be a closed subspace of H and $\overline{H}^{\perp}$ its orthogonal complement. If we assume that $g \in H$, it can be written as

$$g = \bar{g} + g_{\perp} \tag{3.12}$$

where $\bar{g} \in \overline{H}$ and $g_{\perp} \in \overline{H}^{\perp}$. In our case $\overline{H}$ is $\mathbb{R}^m$ and $\overline{H}$ is the range of the operator defined by (3.10). So, let $R(A_{\mu})$ be the range of the operator A_{μ} and $R(A_{\mu})^{\perp}$. Its orthogonal complement. then $\mathbb{R}^m = R(A_{\mu}) \oplus R(A_{\mu})^{\perp}$ and we can write $\mathbf{g} = \bar{\mathbf{g}} + \mathbf{g}_{\perp}$, where $\bar{\mathbf{g}} \in R(A_{\mu})$ and $\mathbf{g}_{\perp} \in R(A_{\mu})^{\perp}$ and $\mathbf{g} \in \mathbb{R}^m$. Then the consistency condition for (3.10) is that the data should not have components other than those in the operator's range, that is,

$$\mathbf{g}_{\perp} = 0. \tag{3.13}$$

If $A_{\mu}^{+} \in \mathbb{R}^{n \times m}$ stands for the Moore-Penrose pseudoinverse of the matrix A_{μ} [3], then:

$$\begin{aligned} \bar{\mathbf{g}} &= P_{\mu}\mathbf{g} \\ \mathbf{g}_{\perp} &= P_{\mu}^{\perp}\mathbf{g} \end{aligned} \tag{3.14}$$

where

$$\begin{aligned} P_{\mu} &= A_{\mu}A_{\mu}^{+} \\ P_{\mu}^{\perp} &= I - A_{\mu}A_{\mu}^{+}. \end{aligned} \tag{3.15}$$

are the orthogonal projections onto $R(A_{\mu})$ and $R(A_{\mu})^{\perp}$, respectively.

With the previous notation, the consistency condition assumes the form:

$$P_{\mu}^{\perp}\mathbf{g} = 0. \tag{3.16}$$

In the case of PET, A_{μ} is the product of two matrices: a diagonal matrix, that depends on the attenuation, and a projection matrix, S, that does not, that is:

$$A_{\mu} = D_{\mu-}S \tag{3.17}$$

with

$$(D_{\mu-})_{ii} = e^{-\sum_k \mu_k s_{ik}} (S)_{ij} = \sigma_{ij}.$$

Considering that, in this particular case ($D_{\mu-}$ diagonal and positive), $(D_{\mu-}S)^+ = S^+(D_\mu)^{-1}$, and replacing (3.17) in (3.16), we obtain

$$I - D_{\mu-}S(D_{\mu-}S)^+ = I - D_{\mu-}SS^+D_\mu = I - D_{\mu-}(I - SS^+)D_\mu\mathbf{g} = 0 \quad (3.18)$$

where

$$D_\mu = (D_{\mu-})^{-1}. \quad (3.19)$$

Therefore the consistency condition is equivalent to:

$$D_{\mu-}(I - SS^+)D_\mu\mathbf{g} = 0 \quad (3.20)$$

In system (3.20), the matrix $D_{\mu-}$ can be eliminated, by multiplying by the inverse. We have, then, a 'linear system' for $x = D_\mu\mathbf{g}$, nonlinear for μ, that can be written in a more convenient form as:

$$Mx = 0 \quad (3.21)$$

with

$$M = (I - SS^+)x_i = u_i g_i, \quad x_i \geq g_i u_i = e^{\sum_k \mu_k s_{ik}}.$$

In the next section we describe two different approaches for solving (3.21)

3.2.1   Using the Newton-Raphson method

System (3.21) can be solved using the Newton-Raphson method for non-linear systems with nonnegative constraints, defined by the iteration:

$$\mathbf{J}_\mu \Delta\mu = \mathbf{F}(\mu^k) \quad (3.22)$$

where

$$\Delta\mu \equiv \mu^k - \mu^{k+1}, \quad (3.23)$$

$$\mathbf{F}(\mu) = M\mathbf{x}(\mu) \quad (3.24)$$

($\mathbf{x}$ is given by (3.21)) and J_μ is the Jacobian matrix with components:

$$j_{ij} = \frac{\partial f_i(\mu)}{\partial \mu_j}. \quad (3.25)$$

J_μ or:

$$j_{ij} = \sum_l m_{il} \frac{\partial}{\partial \mu_j} x_l = \sum_l m_{il} x_l \sigma_{lj}. \quad (3.26)$$

In matrix form,

$$J_\mu = MB_\mu = Q_2 Q_2^t B_\mu \,, \tag{3.27}$$

where B_μ is a matrix with components defined by

$$b_{ij} = x_i \sigma_{ij}. \tag{3.28}$$

Using the QR decomposition and (3.21), the matrices S and M can be written as:

$$S = (Q_1 \ \ Q_2) \begin{pmatrix} R \\ 0 \end{pmatrix} \quad M = Q_2 Q_2^t$$

with $Q_1 \in \mathbb{R}^{m \times n}$ and $Q_2 \in \mathbb{R}^{m \times (m-n)}$. As in the original Bronnikov's approach, the Jacobian matrix is singular with rank $m - n$. An alternative, equivalent, but more expensive way to obtain M is the SVD decomposition. In [9], details could be found on the different ways to deal with the matrix M.

3.3 The ML Approach. Crepaldi and De Pierro

According to our analysis in (1.3), the ML approach, that takes into account the statistical properties of the noise, tends to retrieve better solutions than nonstatistical models. So, back to the discretized model, given the activity and the attenuation maps for PET, where the attenuation is independent of the position along the projection line, the expected number of counts is given by

$$r_i = a_i b_i \tag{3.29}$$

with

$$a_i \equiv \exp\left\{-\sum_{j=1}^n s_{ij}\mu_j\right\} \tag{3.30}$$

and

$$b_i \equiv \sum_{j=1}^n \sigma_{ij}\lambda_j \tag{3.31}$$

with σ_{ij} being the probability that an annihilation occurred at j is detected in tube i and s_{ij} the intersection of tube i with box k:

Now, if g_i is the number of counts measured at the i-th tube of detectors, as in Section 1.3 the g_i's, are independent Poisson variables with mean r_i. So, after

excluding terms not containing λ or μ, the new log-likelihood function that we want to maximize is

$$L(\lambda, \mu) = \sum_{i=1}^{m} (-r_i + g_i \ln r_i). \tag{3.32}$$

or equivalently,

$$L\left(\lambda, \mu\right) = \sum_i \left\{ - \exp\left(-\sum_j s_{ij}\mu_j \right)\left(\sum_{j'} \sigma_{ij'}\lambda_{j'} \right) \right.$$
$$\left. - g_i \sum_j s_{ij}\mu_j + g_i \ln \sum_j \sigma_{ij}\lambda_j \right\} \tag{3.33}$$

It can be proven that, for fixed μ, L is a concave function of λ and vice versa [23]. However, L is not jointly concave for (λ, μ). Following [23], we develop an alternate maximization method, where each maximization step is based on the minorizing functions approach proposed in [10]. The difference between the algorithm in [23] and ours is essentially the guaranteed monotonicity for the latter. If we assume first that the attenuation map is constant, using the logarithm's concavity, and, taking into account that

$$\sum_j \frac{\sigma_{ij}\lambda_j^k}{\langle \sigma_i, \lambda^k \rangle} = \frac{\sum_j \sigma_{ij}\lambda_j^k}{\langle \sigma_i, \lambda^k \rangle} = \frac{\langle \sigma_i, \lambda^k \rangle}{\langle \sigma_i, \lambda^k \rangle} = 1, \tag{3.34}$$

where λ^k is the current emission at iteration k and, for any given vectors $u = (u)_j$ and $v = (v)_j$ $\langle u, v \rangle = \sum_j u_j v_j$ stands for the scalar product, we have that:

$$\ln \sum_j \sigma_{ij}\lambda_j = \ln \sum_j \frac{\lambda_j^k \langle \sigma_i, \lambda^k \rangle}{\lambda_j^k \langle \sigma_i, \lambda^k \rangle} \sigma_{ij}\lambda_j \tag{3.35}$$

is greater equal than

$$\ln \sum_j \sigma_{ij}\lambda_j \geq \sum_j \frac{\sigma_{ij}\lambda_j^k}{\langle \sigma_i, \lambda^k \rangle} \ln \frac{\langle \sigma_i, \lambda^k \rangle}{\lambda_j^k} \lambda_j. \tag{3.36}$$

On the other hand, using the same procedure to separate the μ variables for the (negative) exponential term. That exponential part is greater or equal than

$$- \exp\left(-\sum_j s_{ij}\mu_j \right) \geq -\sum_j \frac{s_{ij}\mu_j^k}{\langle s_i, \mu^k \rangle} \exp\left(-\frac{\langle s_i, \mu^k \rangle}{\mu_j^k} \mu_j \right). \tag{3.37}$$

Using (3.36) and (3.37), we obtain, for each k:

$$L(\lambda, \mu) \geq \Phi(\lambda, \mu, \lambda^k, \mu^k) \tag{3.38}$$

with

$$\Phi(\lambda, \mu, \lambda^k, \mu^k) = \sum_i \left\{ - \sum_j \frac{s_{ij} \mu_j^k}{\langle \mathbf{s}_i, \mu^k \rangle} \exp\left(- \frac{\langle \mathbf{s}_i, \mu^k \rangle}{\mu_j^k} \mu_j \right) \langle \sigma_i, \lambda \rangle \right. \tag{3.39}$$
$$\left. - g_i \langle \mathbf{s}_i, \mu \rangle + g_i \sum_j \frac{\sigma_{ij} \lambda_j^k}{\langle \sigma_i, \lambda^k \rangle} \ln \lambda_j \right\}$$

Now, the algorithm's global iteration is defined in two steps: for a given activity-attenuation vector pair (λ^k, μ^k), the first step consists of solving the problem

$$\lambda^{k+1} = \arg\max_{\lambda \geq 0} \Phi(\lambda, \mu^k, \lambda^k, \mu^k) \tag{3.40}$$

and the second step iterates in μ solving the problem

$$\mu^{k+1} = \arg\max_{\mu \geq 0} \Phi(\lambda^{k+1}, \mu, \lambda^k, \mu^k). \tag{3.41}$$

The solution of the first problem (3.40) is nothing but the EM algorithm applied to L with fixed μ and this gives the updating

$$\lambda_l^{k+1} = \frac{\lambda_l^k}{a_l} \sum_i \frac{g_i \sigma_{il}}{\langle \sigma_i, \lambda^k \rangle} \tag{3.42}$$

with

$$a_l = \sum_i \left\{ \sigma_{il} \exp\left(- \langle \mathbf{s}_i, \mu^k \rangle \right) \right\}. \tag{3.43}$$

The solution of the problem defined by the second step (3.41) requires solving a set of nonlinear optimization problems in one variable that can be very easily solved using the Newton-Raphson method with a positivity constraint. In our implementation, enough iterations of Newton-Raphson's method were performed in such a way that the maximum in (4.41) is guaranteed (in practice one iteration of Newton Raphson is enough to increase the function value) as well as the monotonicity of the algorithm (4.40–41). In [15], λ and μ are updated sequentially by applying Newton's method to the first order necessary conditions

for the maximum. Trial and error is needed to choose the ascent parameters, in order to preserve monotonicity, as in [23].

Unfortunately, the application of the algorithm above in its pure form, as well as Nuyts', still generates images with the undesired 'cross-talk', so, following Natterer's original idea, consistency should be imposed. The way we found to do that is by using Iterative Data Refinement.

3.3.1 *Iterative Data Refinement (IDR)*

In this Section we briefly describe the general methodology known as Iterative Data Refinement (IDR). A more detailed description together with some historical background and applications in tomography could be found in [8].

Let x be a mathematical representation of a physical object. It might represent, in our problem, the activity density. The *idealized measurer* is defined as the operator $\mathcal{A}$ that provides the *ideal measurement vector*, $\hat{y}$, when applied to the vector x:

$$\mathcal{A}x = \hat{y}.$$

The matematical operator that recovers the input from the output is called the *recovery operator* and denoted by $\mathcal{R}$:

$$\mathcal{R}\hat{y} = \hat{x}.$$

In general $\hat{x}$ is a good approximation for x.

However, in practice we do not have an exact physical implementation of the idealized measurer $\mathcal{A}$. We have an actual measurer $\mathcal{B}$ which, given x, produces the *actual measurement vector* $\tilde{y}$:

$$\mathcal{B}x = \tilde{y}.$$

The problem arises because, in general, $\tilde{x} = \mathcal{R}\tilde{y}$ is **not** a good approximation for x.

This difficulty is the motivation for *Iterative Data Refinement* (IDR). The main idea is: if y^k is a reasonable approximation for $\hat{y}$, then

$$\mathcal{A}\mathcal{R}y^k - \eta(k)\mathcal{B}\mathcal{R}y^k \approx \hat{y} - \eta(k)\tilde{y}$$

where $\eta(k)$ is a relaxation parameter. From this equation, it is natural to propose the iterative process:

$$y^{k+1} = \mathcal{A}\mathcal{R}y^k + \eta(k)(\tilde{y} - \mathcal{B}\mathcal{R}y^k). \tag{3.44}$$

Suppose that y^k is a good approximation for $\hat{y}$. Then

$$y^k \approx \mathcal{A}\mathcal{R}y^k.$$

and (3.44) becomes

$$y^{k+1} = y^k + \eta(k)(\tilde{y} - \mathcal{B}\mathcal{R}y^k). \tag{3.45}$$

In the next Section we describe how we apply IDR to the activity attenuation problem.

3.3.2   IDR and the optimization approach

The main numerical problem that we have to solve for the optimization model is the fact that iterative methods like (3.40)-(3.41) tend to choose the 'wrong' solution (fixed point). Solutions showing the undesired 'crosstalk' look more equilibrated and highly likely than the ones we are looking for. This rationale plus the observation of the results of many experiments led us to the introduction of a multiplicative factor in (3.42), that, essentially "breaks" the equilibrium, but still converging to an ML solution. That is, we substituted (3.42) by

$$\lambda_l^{k+1} = \phi_\xi(k)\frac{\lambda_l^k}{a_l} \sum_i \frac{g_i\sigma_{il}}{\langle \sigma_i, \lambda^k \rangle} \tag{3.46}$$

where a_l is as before and

$$\phi_\xi(k) = \frac{k + \xi}{k + 1} \tag{3.47}$$

is the multiplicative factor that retains the asymptotic convergence properties of the EM part of the algorithm, ξ is a constant parameter to be chosen. $\phi_\xi(k)$ tends to one, so that the monotonicity property of the likelihood remains essentially valid when updating the activity variables. An important observation is that the multiplicative factor enhances the data, preparing a better starting point for the next step, IDR.

In order to apply IDR to our problem, now x represents the vector whose elements are the activity and attenuation densities, that is, $\mathbf{x} = (\lambda, \mu)$, $\mathcal{B}$ is the operator that we want to 'invert', that is, the discrete Radon transform with the corresponding attenuation, given by the left hand side of (2.7), and $\mathcal{R}$ the recovery method of simultaneous reconstruction, previously described (the alternated maximization using the multiplicative factor for the EM step). In the appropriate notation, we can write (3.45) as

$$\mathbf{g}^{k+1} = \mathbf{g}^k + \eta(k)(\tilde{\mathbf{g}} - \mathcal{B}\mathcal{R}\mathbf{g}^k). \tag{3.48}$$

Observe that, in (3.48), if there is convergence to a fixed point, say $\tilde{\mathbf{g}}*$, the ideal measurement (and that is the case), that means that the data $\tilde{\mathbf{g}}$ has been modified in such a way that our activity-attenuation solution is one that the forward operator (the attenuated Radon Transform) applied to it is consistent with the original data (actual measurement).

Summarizing, our method consists of two iterative processes: the IDR iterations and the alternated iterations from section 2 that recover both images simultaneously. This can be described by the following steps; for a given starting point, and the data $\tilde{\mathbf{g}}$, and, in general, for a point (λ^k, μ^k) and the modified data $\mathbf{g}^k$:

Step 1: Apply (4.41-42) until convergence using $\mathbf{g}^k$ as data, that is, compute $\mathcal{R}\mathbf{g}^k$.

Step 2: Update the data in (4.45) using the forward operator $\mathcal{B}$.

Step 3: Go back to **Step 1**. Stop if a suitable criterion is satisfied.

In the next sections we present our simulations and describe how the various parameters have been chosen, including the value of the constant ξ in the multiplicative factor and the number of iterations used for each method, the alternated maximization and IDR.

4   Simulations

In this Section we present a visual comparison of the methods described in Sections 4.2 and 4.3, and the results for pure ML approaches without IDR. That is, we consider only methods not adding any information about the solution,

through the use of approximate. For the simulations we used a simple model of a 2D thorax cross section phantom. The activity source was modeled as a ring, representing the heart, inside an ellipse, which represents the background activity in the human body, approximately 10 times less than the ring's activity. The attenuation map was an ellipse representing the body, two ellipses representing the lungs and a small circle for the spine bone (see Figure 4). For all the 50×50 reconstructions it is also shown the profile of the the 35th line and 16th line for 32×32 reconstructions.
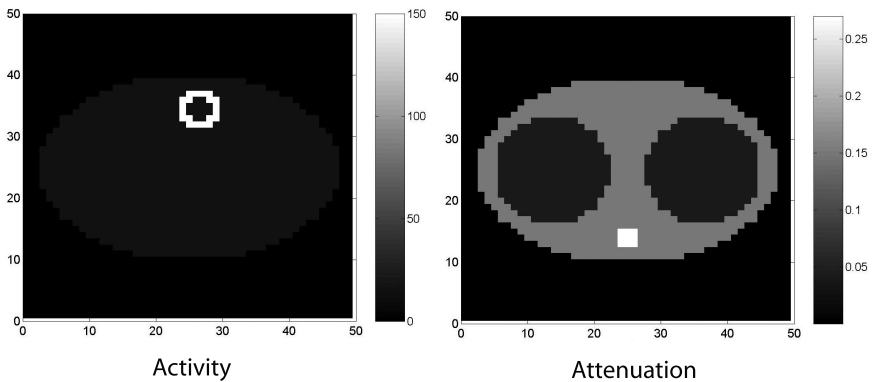
Figure 4 – Phantom.

For all the experiments, the starting images were uniform using an average value; there were no modifications in the results when varying these values in an acceptable range. Images were reconstructed using a 50×50 grid, except for Bronnikov's, that was 32×32, because of the huge amount of computations.

Figures 5 and 6 show the reconstructions for Nuyts et al's method and pure Crepaldi & De Pierro's (without IDR) respectively. The cross-talk between images is clearly seen.

Figure 7 shows the result of applying Bronnikov's approach (that computes only the attenuation) and Figure 8 the result for Crepaldi & De Pierro's method with IDR. In both cases the cross-talk almost dissapeared.

Finally Figure 9 shows the result obtained by Bronnikov's approach with noise. Noise was generated by using a Poisson generator with mean given by the data for each projection. Our results with noise for IDR did not show substantial changes.
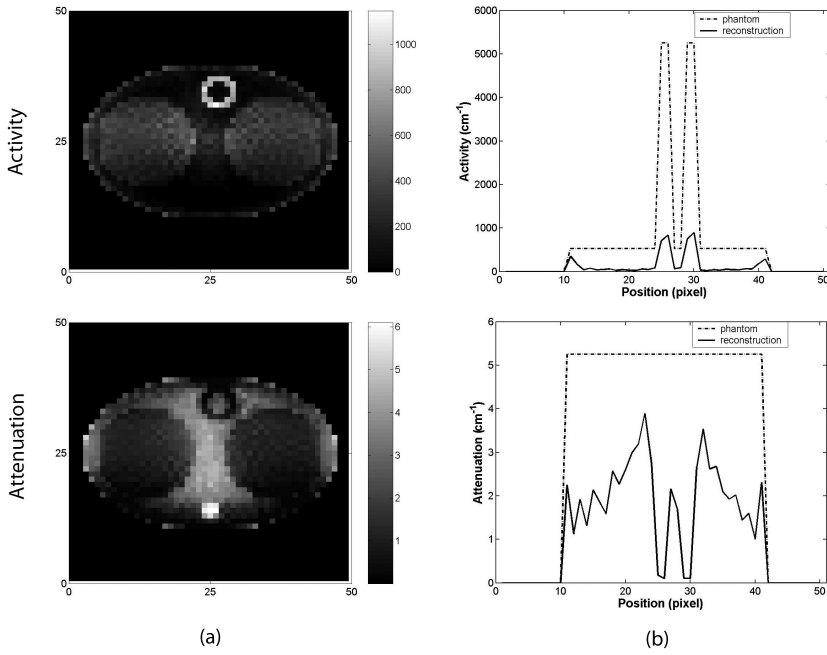
(a)   (b)

Figure 5 – Nuyts et al.
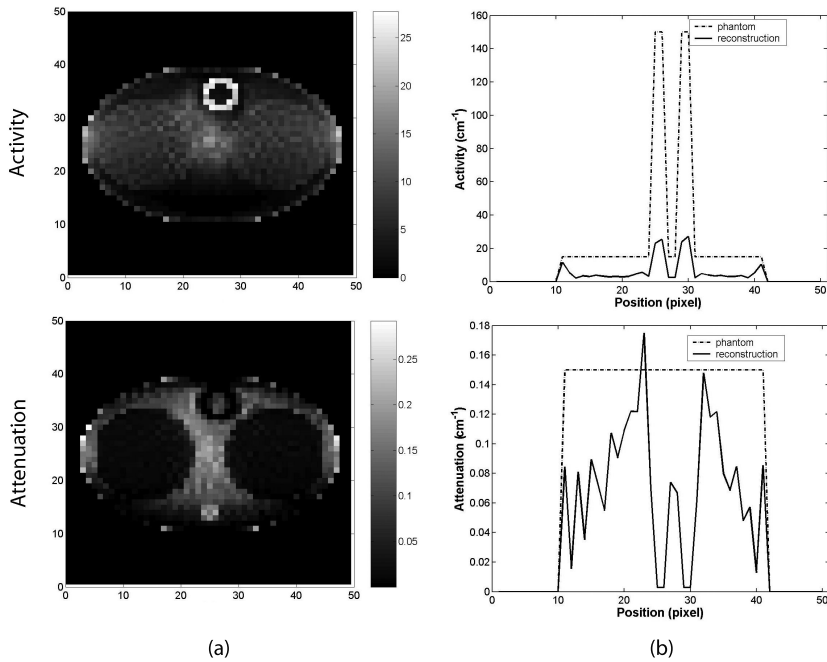

(a)   (b)

Figure 6 – Crepaldi-De Pierro.

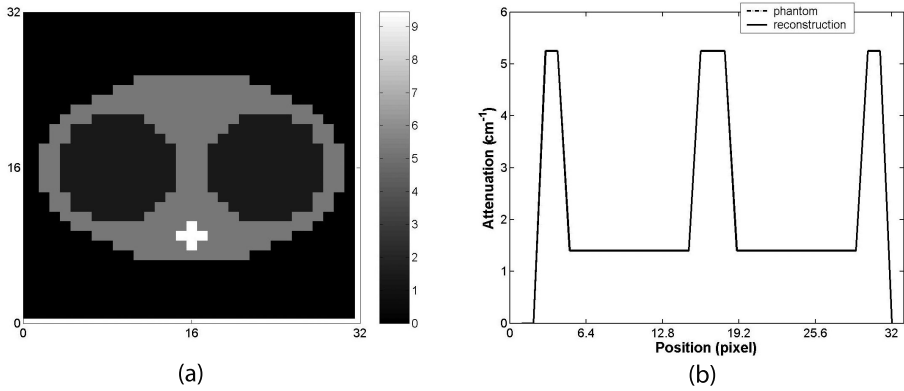Figure 7 – Bronnikov's.

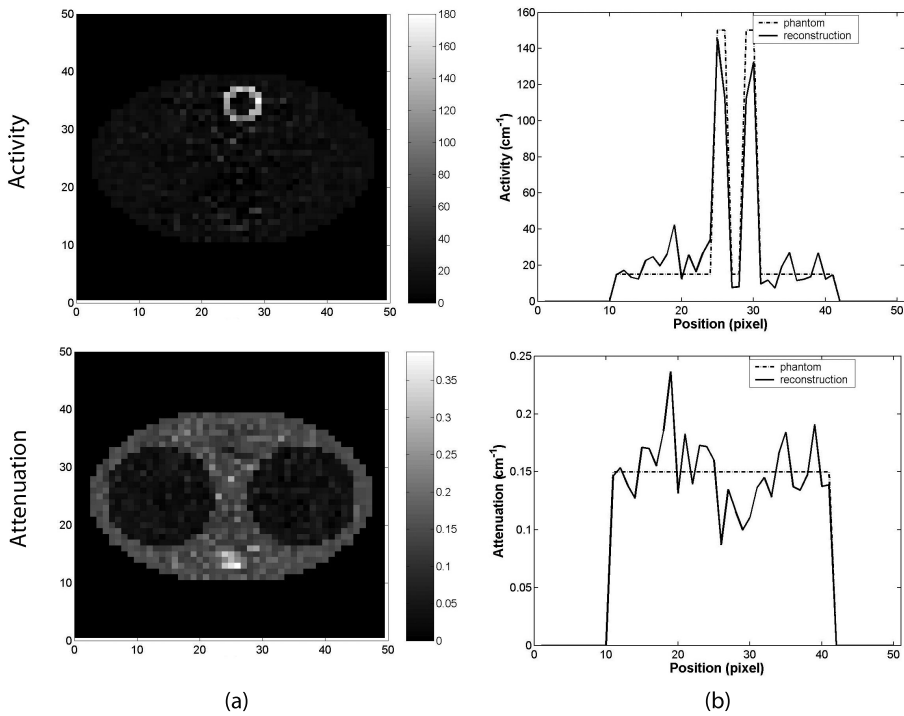

Figure 8 – Crepaldi-De Pierro with IDR.

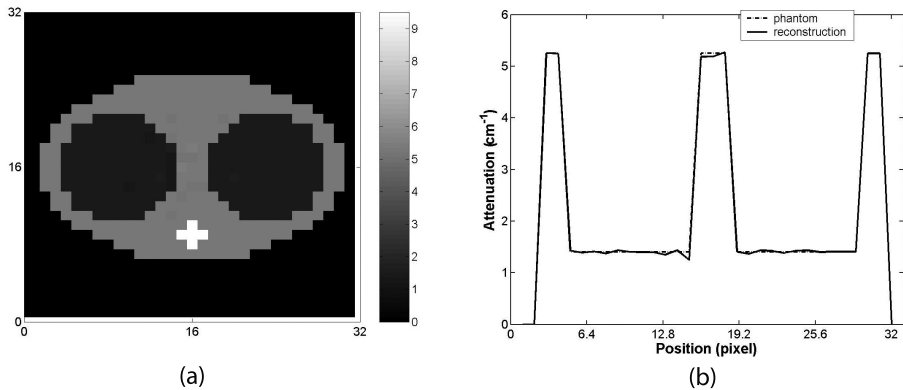(a)   (b)

Figure 9 – Bronnikov's with noise.

5   Some conclusions and questions still to be answered

The methods presented in Sections 4.2 and 4.3 are a first step to obtain quality images in ECT from emission data only, without any additional information. However, there is a relatively long list of open problems still to be addressed. It follows.

1. One advantage of iterative methods with IDR, as compared to Bronnikov's approach is the possibility of considering ML models for the noise, but the other one is the computational burden. Iterative methods 'could be' computationally much less time consuming if faster methods like RAMLA [4] are used for the inner part in (3.40)-(3.41). The overall number of iterations (including IDR) should be also reduced.

2. Reasonable stopping criteria are needed for the outer (IDR) and inner iterations.

3. Stability: together with the number of iterations, several other parameters are involved. Also, there is some dependence of these parameters with respect to the discretization.

4. Some kind of regularization should be considered. That could help to improve the images in the presence of noise, but could also help for the choice of the 'right' solution.

5. Local convergence results could be obtained along the lines of [12].

6. We are currently working on simulations for more complex objects, for example, nonconvex. Preliminary results show that our approach with IDR does not produce a convexification as in [23].

7. We are now implementing the methods for SPECT simulations, and with real data, in a joint research with the Heart Institute of the University of São Paulo.

It is also worth mentioning that these methods for problems with two 'separated' unknowns and many possible solutions depending on the data, could be a reference in other important inverse poblems like optical tomography and phase retrieval.

REFERENCES

[1] M. Avriel, Nonlinear Programming: Analysis and Methods, Prentice Hall, Englewood Cliffs, NJ, 1976.

[2] C. Bai, P. Kinahan, D. Brasse, C. Comtat, D.W. Townsend, C.C. Metzer, V. Villemagne, M. Charron and M. Defrise, An analytic study of the effects of attenuation on tumor detection in whole-body PET oncology imaging, J. Nucl. Med., **44** (2003), 1855–1861.

[3] A. Ben Israel and T. Greville, Generalized Inverses: Theory and Applications, Springer, 2 edition, 2003.

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

[5] T.F. Budinger, G.T. Gullberg and R.H. Huesman, Emission computed tomography, in Image Reconstruction from Projections: Implementation and Applications, G.T. Herman (Ed), Springer Verlag, Berlin, Heidelberg. **5** (1979), 147–246.

[6] A.V. Bronnikov, Numerical solution of the identification problem for the attenuated Radon transform, Inverse Problems, **15** (1999), 1315–1324.

[7] A.V. Bronnikov, Reconstruction of attenuation map using discrete consistency conditions, IEEE Trans. Medical Imaging, **19**(5) (2000), 451–462.

[8] Y. Censor, T. Elfving and G.T. Herman, A method of iterative data refinement and its applications, Math. Meth. Appl. Sc., **7** (1985), 108–123.

[9] F. Crepaldi, On the computation of the attenuation and the activity in emission tomography from activity data, PHd Dissertation, Institute of Physics, State University of Campinas, 2004.

[10] A.R. De Pierro, A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography, IEEE Trans. Med. Imaging, **14**(1) (1995), 132–137.

[11] S. Helgasson, The Radon Transform, Boston, Birkhauser, 1980.

[12] E.S. Helou and A.R. De Pierro, Convergence results for scaled gradient algorithms in positron emission tomography, Inverse Problems, **21**(6) (2005), 1905–1914, October.

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

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

[15] M. Landmann, S.N. Reske and G. Glatting, Simultaneous iterative reconstruction of emission and attenuation images in positron emission tomography from emission data only, Med. Phys., **29** (2002), 1962–1967.

[16] K. Lange and R. Carson, EM reconstruction algorithms for emission and transmission tomography, J. Comput. Assisted Tomog., **8** (1987), 306–316.

[17] http://www.crump.ucla.edu/software/lpp/lpphome.html

[18] D. Ludwig, The Radon Transform on euclidean space, Comm. Pure Appl. Math. **XIX** (1966), 49–81.

[19] R.H. Huesman and T.F. Budinger, Design of a high resolution, high sensitivity PET camera for human brains and small animals, IEEE Trans. Nucl. Sci. **44** (1997), 1487–1491.

[20] F. Natterer, Determination of tissue attenuation in Emission Tomography of optically dense media, Inverse Problems, **9** (1993), 731–736.

[21] F. Natterer, The Mathematics of Computerized Tomography, SIAM, 2001.

[22] P. Nelleman, H. Hines, W. Braymer, G. Muehllehner and M. Geagan, Performance characteristics of a dual head SPECT scanner with PET capability, IEEE Nuclear Science Symposium and Medical Imaging Conference, Conference record, (1995), 1751–1755.

[23] J. Nuyts, P. Dupont, S. Stroobants, R. Benninck, L. Mortelmans and P. Suetens, Simultaneous Maximum A Posteriori reconstruction of attenuation and activity distributions from emission sinograms, IEEE Trans. Med. Imaging, **18**(5) (1999), 393–403.

[24] A.J. Rockmore and A. Macovski, A maximum likelihood approach to emission image reconstruction from projections, IEEE Trans. Nucl. Sci., **23** (1976), 1428–1432.

[25] M.S. Rosenthal, J. Cullom, W. Hawkins, S.C. Moore, B.M.W. Tsui and M. Yester, Quantitative SPECT imaging: a review and recommendations by the focus committee of the Society of Nuclear Medicine Computer and Instrumentation Council, J. Nucl. Med., **36** (1995), 1489–1513.

[26] L.A. Shepp and Y. Vardi, Maximum likelihood reconstruction for emission tomography, IEEE Trans. Med. Imaging, **1** (1982), 113–121.

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

[28] Y. Vardi, L.A. Shepp and L. Kaufman, A statistical model for positron emission tomography, J. Amer. Statist. Assoc., **80** (1985), 8–20.

[29] A. Welch, C. Campbell, R. Clackdoyle, F. Natterer, M. Hudson, A. Bromiley, P. Mikecz, F. Chillcot, M. Dodd, P. Hopwood, S. Craib, G.T. Gullberg and P. Sharp, Attenuation correction in PET using consistency information, IEEE Trans. Nuclear Science, **45**(6) (1988).