On the Effect of Relaxation in the Convergence and Quality of Statistical Image Reconstruction for Emission Tomography Using Block-iterative Algorithms

Elias Salomão Helou Neto
Universidade Estadual de Campinas
Inverse Problems Group
IMECC
e950514@ime.unicamp.br

Álvaro Rodolfo De Pierro
Universidade Estadual de Campinas
Inverse Problems Group
IMECC
alvaro@ime.unicamp.br

Abstract

Relaxation is widely recognized as a useful tool for providing convergence in block-iterative algorithms [1], [2], [6]. In the present article we give new results on the convergence of RAMLA (Row Action Maximum Likelihood Algorithm) [2], filling some important theoretical gaps. Furthermore, because RAMLA and OS-EM (Ordered Subsets - Expectation Maximization) [4] are the algorithms for statistical reconstruction currently being used in commercial emission tomography scanners, we present a comparison between them from the viewpoint of a specific imaging task. Our experiments show the importance of relaxation to improve image quality.

1. Introduction

1.1. Statistical reconstruction in positron emission tomography

Emission tomography refers to the reconstruction of cross-sectional images of radioactive intensity from radiation detected outside the object being studied. In medical applications the emission is produced by the use of biochemical metabolites labelled by a radioactive element. This allows the study of metabolic activity, which may help to achieve diagnostic goals that cannot be reached with the aid of other tomographic techniques. The emission is proportional to the concentration of the labeled metabolite and there are obvious limitations on this concentration inside the patient's body. This implies that the emission rates may be quite low even in the regions with comparatively large concentration of the metabolite. Because radioactivity has a statistical nature, low emission causes the measured data to be highly distorted by noise in a way that one cannot neglect its probabilistic characteristics when reconstructing the desired image. Even though the results below can be easily adapted to SPECT (Single Photon Emission Computed Tomography), we will be focusing on PET (Positron Emission Tomography).

1.2. Positron emission tomography

In PET, the radioactive compound used decays emitting a positron, which is annihilated after colliding with a nearby electron. This results in two photons with 511keV energy traveling in almost opposite directions. See Figure 1 for a schematic representation of a PET scan. In there, $\times$ represents the place where a positron was emitted; the dotted line shows the trajectory of the resulting photons and the ring around the image is the detector grid. Several other features related to photon behavior (distance until annihilation, scatter, attenuation, directions of the photons not exactly opposites, etc) will not be considered. For more information about this, see [7]. It is worth mention that most of these effects can be included in a natural way in the statistical model to be presented below.

The data resulting from the scan is the number of almost simultaneous detections for each pair of detectors. These photons detected nearly at the same time must be result of a positron-electron collision which, in turn, is a consequence of a positron emission. From this we know, for every coincidence, a region in space (with an associated probability density), where the positron that generated it could have been emitted. Some scanners record the time between near detections, increasing our information about the location of the emission. This can also be incorporated into the statistical model [7].

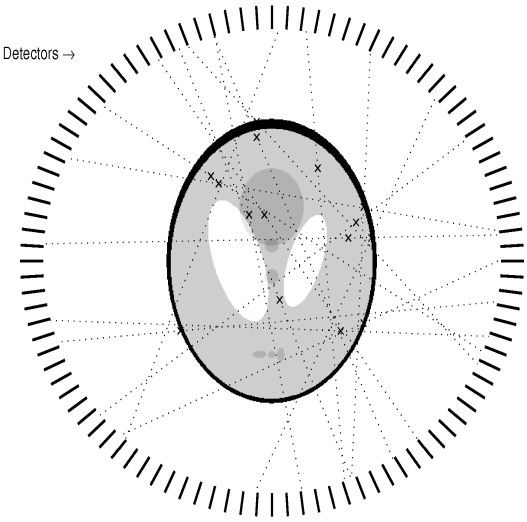

Figure 1. Schematic representation of a short PET scan.

1.3. Block-iterative algorithms

The first algorithm proposed to solve the optimization problem arising from the statistical approach in emission tomography was the EM (Expectation-Maximization) algorithm [7] which is quite slow, often taking hundreds of iterations to reach convergence. The most successful efforts in speeding up the convergence, although with very simple implementation, were those based on an incremental update of the solution. In all of these block-iterative methods, one iteration is composed of several subiterations, each of which considers only one subset of the tomograph data at a time, reducing the computational cost required proportionally to the size of the considered subset. At the end of an iteration all data must have been considered and if, furthermore, the subsets do not overlap, the total computational effort is equal to a gradient evaluation plus the updates between consecutive subiterations, whose cost is negligible. One then hopes to improve the solution at each subiteration as much as an EM iteration with much less computational overhead.

Because of inconsistency introduced by noise, algorithms considering subsets of the data usually do not converge unless either the level of incrementality is gradually reduced or a diminishing stepsize rule is adopted. The former seems to be a reasonable procedure, however one will invariably end up with a non-incremental method so that we may consider the block-iterative method just as an initialization procedure for a usual algorithm. Diminishing stepsize rules (also known as diminishing relaxation parameters because of historical reasons) seem to be intrinsically more pleasant, since the incrementality of the method is maintained while convergence can still be assured.

1.4. Organization of the paper

The paper is organized in the following manner: in section 2 the optimization problem posed by the maximum likelihood approach is presented and the EM algorithm is described; in section 3 we present the algorithms that will be used to asses the importance of relaxation in the image reconstruction, namely RAMLA and OS-EM; section 4 is devoted to explain how the free parameters of both algorithms should be properly chosen and, simultaneously, a comparison is shown of both algorithms behavior; we present in section 5 the demonstration of convergence for RAMLA under conditions similar to those in [1], but with weaker hypotheses; finally, in section 6 we give our concluding remarks.

2. The optimization problem

In statistical image reconstruction, one seeks for the maximum likelihood estimator of the image. What is meant by this is that the image is chosen among all possible ones in order to maximize the probability of obtaining the given data. Discretizing the image in a regular grid of non overlapping pixels (picture elements, or voxels in the tridimensional case) where the emission intensity is considered to be constant, let us consider the corresponding Poisson likelihood, that is, the probability of obtaining the given data as:

$$P(\mathbf{b}|\mathbf{x}) = \prod_{i=1}^{m} \mathrm{e}^{-(A\mathbf{x})_i} \frac{(A\mathbf{x})_i{}^{b_i}}{b_i!}, \tag{1}$$

where $\mathbf{b} = (b_i) \in \mathbb{R}_+^m$ is a vector containing the number of photons counted in the i^{th} bin, $\mathbf{x} = (x_j) \in \mathbb{R}_+^n$ is the discretized image and $A = (a_{ij}) \in \mathbb{R}_+^{m \times n}$ is a system matrix that could incorporate the various effects that act in an emission scan such as attenuation, scattering, detector efficiency, etc. We simplify the problem by taking logarithms and discarding constant terms, resulting in the nonlinear optimization program:

$$\begin{aligned} &\max L(\mathbf{x}) := \sum_{i=1}^{m} \{b_i \log(A\mathbf{x})_i - (A\mathbf{x})_i\} \\ &\text{s.t. :} \\ &\quad x_j \geq 0 \quad j = 1, \ldots, n. \end{aligned} \tag{2}$$

The function $L(\mathbf{x})$ defined above is easily shown to be concave. Conscious manipulation of the K.K.T. conditions for this problem lead to the following necessary

condition for $\mathbf{x}^*$ to be an optimal point:

$$x_j^* = \frac{x_j^*}{\sum_{i=1}^m a_{ij}} \sum_{i=1}^m \frac{a_{ij} b_i}{(A\mathbf{x}^*)_i} \quad j = 1, \ldots, n. \tag{3}$$

This motivates the well known EM algorithm:

$$x_j^{(k+1)} = \frac{x_j^{(k)}}{\sum_{i=1}^m a_{ij}} \sum_{i=1}^m \frac{a_{ij} b_i}{(A\mathbf{x}^{(k)})_i} \quad j = 1, \ldots, n. \tag{4}$$

3. The algorithms

3.1. OS-EM

As proposed by Hudson and Larkin [4], OS-EM (Ordered Subsets - Expectation Maximization) is an algorithm where EM iterations are taken successively for subsets of the tomograph data. When far from the optimal solution each subiteration imposes to the present estimate a direction very similar to that given by a complete EM iteration. The final result of a whole cycle through the subsets is roughly the same as N EM iterations, where N is the number of subsets, being the computational cost essentially the same of a single EM iteration. However, the presence of noise makes the system $A\mathbf{x} = \mathbf{b}$ inconsistent and the algorithm reaches a limit cycle within the subiterations and does not converge to the optimal solution. The definition of the algorithm is the following:

$$\begin{aligned} \mathbf{x}^{(k,0)} &= \mathbf{x}^{(k)} \\ x_j^{(k,l)} &= \frac{x_j^{(k,l-1)}}{\sum_{i \in \mathcal{N}_l} a_{ij}} \sum_{i \in \mathcal{N}_l} \frac{b_i a_{ij}}{(A\mathbf{x}^{(k,l-1)})_i} \\ \mathbf{x}^{(k+1)} &= \mathbf{x}^{(k,N)}, \end{aligned} \tag{5}$$

where $\mathcal{N}_l$, $l = 1, \ldots, N$ are the sets of indices of the data for each subset. From now on these subsets are always supposed to be such that the following conditions hold:

$$\begin{aligned} &\bigcup_{l=1}^N \mathcal{N}_l = \{1, \ldots, m\}; \\ &\mathcal{N}_i \cap \mathcal{N}_j = \emptyset, \quad i \neq j. \end{aligned} \tag{6}$$

The conditions above assure that all data is considered in a complete cycle and that the subsets are non overlapping so that the computational burden is essentially the same as an EM iteration.

It is straightforward to show that OS-EM iterations retain positivity and assure boundedness of the sequence $\mathbf{x}^{(k)}$ from well known results on the EM algorithm. It can be shown [4] that, in the consistent case (i.e., the system $A\mathbf{x} = \mathbf{b}$ has a solution), the algorithm converges to a maximum likelihood solution.

Notice that we can write (5) as:

$$\begin{aligned} \mathbf{x}^{(k,0)} &= \mathbf{x}^{(k)} \\ \mathbf{x}^{(k,l)} &= \mathbf{x}^{(k,l-1)} + D_l\big(\mathbf{x}^{(k,l-1)}\big) \nabla L_l\big(\mathbf{x}^{(k,l-1)}\big) \\ \mathbf{x}^{(k+1)} &= \mathbf{x}^{(k,N)}. \end{aligned} \tag{7}$$

Or, equivalently:

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \sum_{l=1}^N D_l\big(\mathbf{x}^{(k,l-1)}\big) \nabla L_l\big(\mathbf{x}^{(k,l-1)}\big), \tag{8}$$

where $D_l\big(\mathbf{x}\big)$ is a diagonal definite positive matrix defined as $D_l(\mathbf{x}) := \operatorname{diag}\left\{\frac{x_j}{\sum_{i \in \mathcal{N}_l} a_{ij}}\right\}$ and the functions $L_l(\mathbf{x})$ are $L_l(\mathbf{x}) := \sum_{i \in \mathcal{N}_l} \{b_i \log(A\mathbf{x})_i - (A\mathbf{x})_i\}$.

Because $\sum_{l=1}^N L_l(\mathbf{x}) = L(\mathbf{x})$, (8) can be regarded as an approximate diagonally scaled gradient iteration[1]. However, if we are close enough to the solution, the error in the approximation of the gradient becomes larger than the error in the present image estimate and this causes the limit cycle to appear.

3.2 RAMLA

RAMLA (Row Action Maximum Likelihood Algorithm), from Browne e De Pierro [2], although proposed in an independent framework, is closely related to OS-EM. The algorithm was first formulated with subsets consisting of a single component of the vector $\mathbf{b}$ but is readily generalized, as well as the convergence results, to subsets that satisfy the conditions (6). We also present the algorithm with the use of a scaling function suggested by Ahn and Fessler [1] which makes stepsize selection more natural and also preserves convergence properties:

$$\begin{aligned} \mathbf{x}^{(k,0)} &= \mathbf{x}^{(k)} \\ x_j^{(k,l)} &= x_j^{(k,l-1)} + \\ &\quad \frac{\lambda_k N x_j^{(k,l-1)}}{\sum_{i=1}^m a_{ij}} \sum_{i \in \mathcal{N}_l} a_{ij} \left(\frac{b_i}{(A\mathbf{x}^{(k,l-1)})_i} - 1 \right) \\ \mathbf{x}^{(k+1)} &= \mathbf{x}^{(k,N)}. \end{aligned} \tag{9}$$

We recast the algorithm in a way similar to what we have done to OS-EM in order to make the relation be-

[1] In fact, because the scaling matrices $D_l(\mathbf{x})$ are different on each subiteration, this does not approximate a diagonally scaled gradient iteration to the function $L(\mathbf{x})$ but to a weighted log-likelihood instead.

tween the algorithms clearer:

$$\begin{aligned}\mathbf{x}^{(k,0)} &= \mathbf{x}^{(k)} \\ \mathbf{x}^{(k,l)} &= \mathbf{x}^{(k,l-1)} + \lambda_k D\big(\mathbf{x}^{(k,l-1)}\big)\nabla L_l\big(\mathbf{x}^{(k,l-1)}\big) \\ \mathbf{x}^{(k+1)} &= \mathbf{x}^{(k,N)}.\end{aligned} \tag{10}$$

Which is the same as:

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \lambda_k \sum_{l=1}^{N} D\big(\mathbf{x}^{(k,l-1)}\big)\nabla L_l\big(\mathbf{x}^{(k,l-1)}\big), \tag{11}$$

where:

$$D(\mathbf{x}) := \operatorname{diag}\left\{\frac{Nx_j}{\sum_{i=1}^{m} a_{ij}}\right\}. \tag{12}$$

Now, the difficulties presented in (8) are circumvented by the fact that the scaling functions are the same in each iteration and that imposing $\lambda_k \to 0$ makes the error in the approximation to the gradient gradually diminish to zero.

It is very simple to show that one can find $\lambda > 0$ such that $0 < \lambda_k \leq \lambda$, $k \geq 0$ ensures positivity and boundedness of the iterates. Also important to notice is that, if one wishes convergence to a global maximizer, the condition $\sum_{k=0}^{\infty} \lambda_k = \infty$ is necessary, otherwise, if $\mathbf{x}^{(0)}$ is far from the solution, we may not reach it.

4. Numerical comparison

Both algorithms presented in the previous section leave the number, choice and order of processing of the subsets open to the user and in the case of RAMLA the sequence of the relaxation parameters needs also to be chosen. In this section we discuss how to properly choose these parameters and compare the performance of RAMLA versus OS-EM.

4.1. The imaging task

In order to have a precise way to evaluate the algorithms, we introduce the pointwise accuracy, defined as:

$$-\sqrt{\frac{\sum_{j=1}^{n}(p_j - x_j)^2}{\sum_{j=1}^{n}(p_j - \overline{p})^2}}, \tag{13}$$

where $\mathbf{p} := (p_j)$ is the digital phantom used in the simulations, $\mathbf{x}$ is the reconstructed image and $\overline{p} := \frac{1}{n}\sum_{j=1}^{n} p_j$ is the average pixel intensity of the phantom. Higher values of this quantity indicate better reconstructed images.

The image being reconstructed is a discrete version of the well known Shepp-Logan Head Phantom [5] which can be seen in Figure 2 below. An ideal PET scan was performed and then, Poisson noise was simulated. When we needed to vary the number of coincidences, it was enough to scale properly the test image prior to the scan simulation.

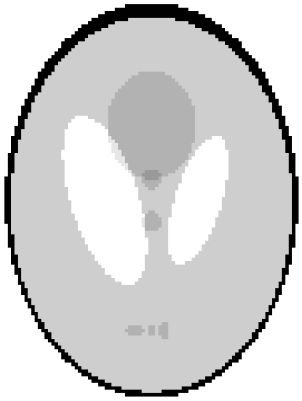

Figure 2. Digital Shepp-Logan head phantom, 128×128 pixels.

4.2. Subsets

How many subsets to use is a rather difficult question to answer. Using only a few would mean slow convergence. However, using lots of subsets may not be desirable. In the case of OS-EM this would pose the problem of reaching a limit cycle fast and far from the optimal solution. We shall not try to answer the question of how many subsets one should use, but, instead, to explain how to choose the parameters once the number of subsets is decided. Another point to be considered is the composition and ordering of these subsets. Helou [3] shows that, for a given task, subsets composed by equally spaced views (each view is the collection of all parallel bins) are less sensitive to ordering. Our experiments use this kind of subset and sequential ordering, even though similar relative behavior for the algorithms was found whatever combination of subsets and ordering was being used.

In PET, the views' angles are equally spaced, so we can say that the i^{th} view is composed by those bins which make an angle of $(i-1)\frac{180^o}{m}$, $i = 1, \ldots, m$ with some properly chosen reference axis. With this convention, we define the l^{th} subset as composed by the views $\big\{l, l+N, l+2N, \ldots, l+N(\frac{m}{N}-1)\big\}$. These

were the subsets used in the experiments to come. The ordering was the natural one, the 1^{st} subset, then the 2^{nd}, etc.

4.3. Relaxation parameters

As there are no such parameters in OS-EM, the choice of the relaxation parameters in RAMLA cannot be time consuming, which excludes the possibility of using training sets. Doing so would bias the comparison towards RAMLA and the importance of relaxation may become overestimated. Relying on the proper scaling of the algorithm, we choose the initial parameter to be $\lambda_0 = 1$ and adopt a simple rule for the rest of the sequence. For RAMLA's convergence, we will show in section 5, it is sufficient that (with $0 < \lambda_k \leq \lambda$ in order to maintain positivity and boundedness) the stepsizes satisfies[2]:

$$\sum_{k=0}^{\infty} \lambda_k = \infty, \quad \sum_{k=0}^{\infty} \lambda_k^2 < \infty. \tag{14}$$

There are many ways of setting up rules of diminishing relaxation parameters that accomplish to these simple conditions. The most obvious one is $\lambda_k = \frac{\lambda_0}{k+1}$. However, intuition says that the relaxation parameters should decrease less for a smaller number of subsets. We illustrate this by means of a simple example: Tests were made with 48, 24, 12 and 6 subsets always with the same diminishing rule. Results are shown in figure 3, where we can see that the effects of relaxation are less beneficial when there are few subsets because it slows down excessively the algorithm. This suggests diminishing stepsize rules of the form $\lambda_k = \frac{\lambda_0}{\alpha(N)k+1}$ where $\alpha(N)$ is a function that will dictate how much the stepsize diminishes between successive iterations for each value of N.

Once in the case $N = 1$ we do not need to diminish the stepsize because $\lambda_k \equiv 1$ is the EM algorithm and for $N = 48$ the choice $\alpha(N) = 1$ has done well, we take $\alpha(N)$ as the linear function such that $\alpha(1) = 0$ and $\alpha(48) = 1$:

$$\lambda_k = \frac{\lambda_0}{\frac{N-1}{47}k + 1}. \tag{15}$$

As wished, if $N = 1$ we have constant parameters and for $N = 48$, we retain the rule used in the above tests. The results in figure 4 show that even this simple formula is able to provide a better balance between the diminishing rule and the degree of incrementality. We see also that the quality of the image degrades faster in the tests that made use of more subsets. Maybe one could use the rule $\lambda_k = \frac{\lambda_0}{\frac{N-1}{23}k+1}$ instead of (15) if we are interested in using a higher level of incrementality or convergence speed is not crucial.

[2]Also, the condition (16) must hold, as will for the diminishing rules to be presented.

4.4. Further testing

One may wish to asses the robustness of the method when varying the noise level and the number of views in data acquisition. We have varied the image being reconstructed, the noise level and the scanner geometry, always finding the same relative behavior between the algorithms. It is worth noting that the image size was, however, kept constant. As an example, we show a test where only 120 angles between 0^o and 180^o were used in the data acquisition. Results are shown in figure 5, where one can see that they are quite the same as the ones previously obtained. A discussion about these results can be found in section 6.

5. Theoretical results

In this section we prove the convergence of RAMLA under the assumptions of **(i)** strict concavity[3] and **(ii)** the conditions on the stepsize (14). We further assume that **(iii)** $\mathbf{x}^{(0)}$ is a strictly positive image, **(iv)** the subsets satisfy (6), **(v)** $0 < \lambda_k \leq \lambda$ where λ is to be chosen according to proposition 1 and that **(vi)**:

$$\frac{\lambda_k}{\lambda_{k+1}} < M \tag{16}$$

for some $M > 0$.

From now on $\mathbf{x}^{(k)}$ denotes a sequence generated by RAMLA with some positive $\mathbf{x}^{(0)}$.

Proposition 1 *There is $\lambda > 0$ such that $0 < \lambda_k \leq \lambda$ implies $x_j^{(k,l)} > 0, \forall j, k, l$.*

PROOF:

$$\begin{aligned} x_j^{(k,l)} &= x_j^{(k,l-1)} \left(1 + \right. \\ &\quad \left. \lambda_k \frac{N}{\sum_{i=1}^m a_{ij}} \sum_{i \in \mathcal{N}_l} a_{ij} \left(\frac{b_i}{(A\mathbf{x}^{(k,l-1)})_i} - 1 \right) \right) \\ &\geq x_j^{(k,l-1)} \left(1 - \lambda_k \frac{N}{\sum_{i=1}^m a_{ij}} \sum_{i \in \mathcal{N}_l} a_{ij} \right). \end{aligned} \tag{17}$$

[3]One can easily show that $L(\mathbf{x})$ is strictly concave for $A \in \mathbb{R}^{m \times n}$ with $m > n$, if and only if $\operatorname{rank}(W(\mathbf{x})A) = n$, where $W(\mathbf{x}) := \operatorname{diag}\{\sqrt{b_i}/(A\mathbf{x})_i\}$ [7].

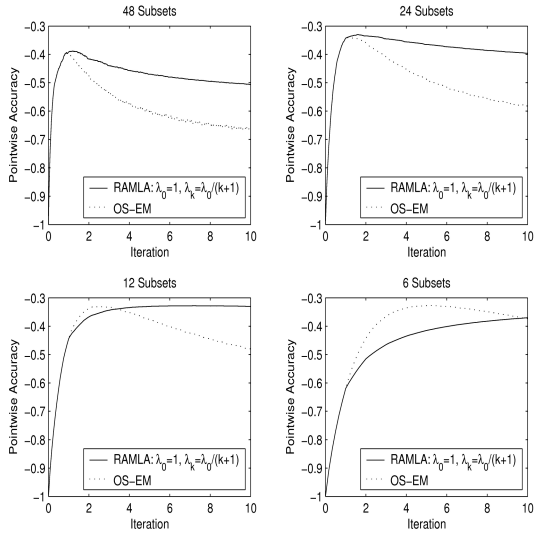

Figure 3. OS-EM×RAMLA. $\sum_{i=1}^{m} b_i = 764,713$**; 384 views.**

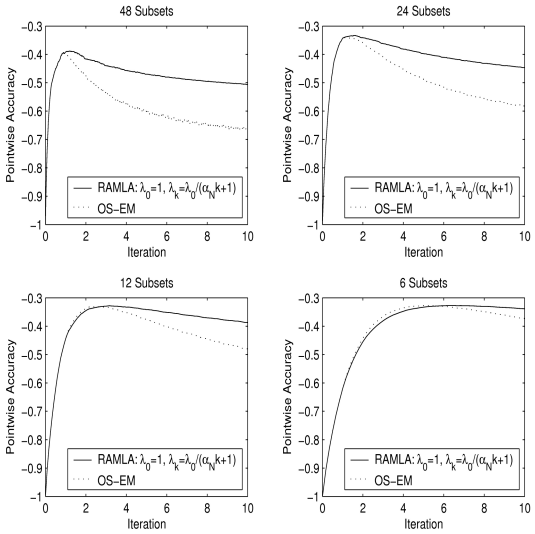

Figure 4. OS-EM×RAMLA. $\alpha_N = \frac{N-1}{47}$**;** $\sum_{i=1}^{m} b_i = 764,713$**; 384 views.**

In order to obtain $x_j^{(k,l)} > 0$ it suffices:

$$\lambda_k \frac{N}{\sum_{i=1}^{m} a_{ij}} \sum_{i \in \mathcal{N}_l} a_{ij} < 1, \quad \forall j,k,l. \tag{18}$$

Thus, if $\lambda_k \leq \lambda$, the inequality above is satisfied whenever:

$$\lambda < \min_{j,l} \frac{\sum_{i=1}^{m} a_{ij}}{N \sum_{i \in \mathcal{N}_l} a_{ij}}. \tag{19}$$

Assuming that $\sum_{i=1}^{m} a_{ij} > 0$ [4]for all j, we can always find a positive λ satisfying the inequality above.□

Now we start demonstrations about of the regularity of the derivatives $\nabla L(\mathbf{x}^{(k,l)})$ that can be used as tools for proving the convergence of the sequence $L(\mathbf{x}^{(k,l)})$, result needed to show the convergence of $\mathbf{x}^{(k,l)}$ to the desired maximizer.

Proposition 2 *If, for given C and I, $\eta > 0$ is a lower bound to the sequence $(A\mathbf{x}^{(k,C)})_I$, generated with $0 < \lambda_k \leq \lambda$ chosen as in proposition 1 then $(A\mathbf{x}^{(k,l)})_I$, $l = 1,\ldots,N$, $k \geq 0$ also has a positive lower bound.*

PROOF: According to the hypothesis, if we make:

$$\gamma_j^{(l)} := 1 - \lambda \frac{N}{\sum_{i=1}^{m} a_{ij}} \sum_{i \in \mathcal{N}_l} a_{ij}; \quad \gamma := \min_{j,l} \gamma_j^{(l)}.$$

Then, because of the way λ was chosen we have that $1 > \gamma > 0$. Furthermore, applying (17) we have $\mathbf{x}^{(k,l+1)} \geq \gamma \mathbf{x}^{(k,l)}$. This leads to the set of inequalities:

$$\mathbf{x}^{(k,L+i)} \geq \gamma^i \mathbf{x}^{(k,L)}, \quad i = 1,\ldots,N-1 \tag{20}$$

Let $\mu = \gamma^{N-1}$, then:

$$\mathbf{x}^{(k,L+i)} \geq \mu \mathbf{x}^{(k,L)}, \quad i = 1,\ldots,N-1, \tag{21}$$

which implies:

$$(A\mathbf{x}^{(k,L+i)})_I \geq \mu (A\mathbf{x}^{(k,L)})_I \geq \mu\eta > 0, \quad i = 1,\ldots,N-1. \tag{22}$$

This ends the demonstration.□

Proposition 3 *With λ as in proposition 1, $0 < \lambda_k \leq \lambda$ and* (16) *being respected, then the sequence generated by RAMLA is such that $(A\mathbf{x}^{(k,l)})_I$ has a positive lower bound whenever $b_I > 0$ and $\mathbf{x}^{(0)} > \mathbf{0}$.*

PROOF: Suppose that $b_I > 0$ and let $\mathcal{N}_C$ be the subset which contains I. We can describe RAMLA's iteration as:

$$\begin{aligned} x_j^{(k,C)} = x_j^{(k-1,C)} + \frac{\lambda_k N}{\sum_{i=1}^{n} a_{ij}} \Bigg\{ & \\ \frac{\lambda_{k-1}}{\lambda_k} \sum_{l=C+1}^{N} x_j^{(k-1,l-1)} \sum_{i \in \mathcal{N}_l} a_{ij} \left(\frac{b_i}{(A\mathbf{x}^{(k-1,l-1)})_i} - 1 \right) & \\ + \sum_{l=1}^{C} x_j^{(k,l-1)} \sum_{\substack{i \in \mathcal{N}_l \\ i \neq I}} a_{ij} \left(\frac{b_i}{(A\mathbf{x}^{(k,l-1)})_i} - 1 \right) & \\ + x_j^{(k,C-1)} a_{Ij} \left(\frac{b_I}{(A\mathbf{x}^{(k,C-1)})_I} - 1 \right) \Bigg\}. & \end{aligned} \tag{23}$$

[4]If that's not the case, there is little sense in trying to estimate the j^{th} pixel.

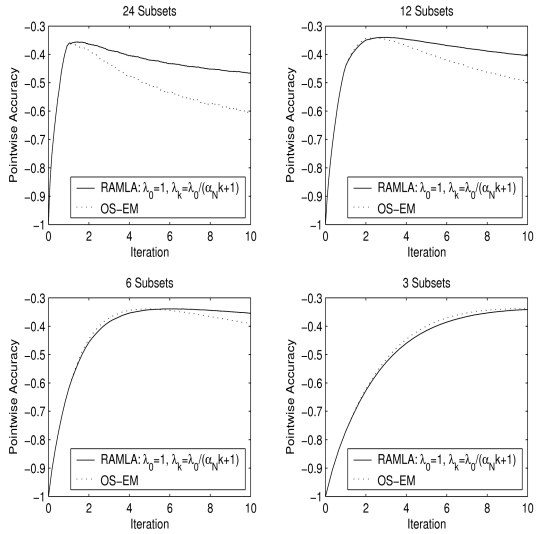

Figure 5. OS-EM×RAMLA. $\alpha_N = \frac{N-1}{47}$**;** $\sum_{i=1}^{m} b_i = 715,863$**;** 120 **views.**

This implies:

$$x_j^{(k,C)} \geq x_j^{(k-1,C)} + \frac{\lambda_k N}{\sum_{i=1}^{n} a_{ij}} \Bigg\{ -\frac{\lambda_{k-1}}{\lambda_k} \sum_{l=C+1}^{N} x_j^{(k-1,l-1)} \sum_{i \in \mathcal{N}_l} a_{ij} - \sum_{l=1}^{C} x_j^{(k,l-1)} \sum_{\substack{i \in \mathcal{N}_l \\ i \neq I}} a_{ij} + x_j^{(k,C-1)} a_{Ij} \left(\frac{b_I}{(A\mathbf{x}^{(k,C-1)})_I} - 1 \right) \Bigg\}. \quad (24)$$

Now, notice that $\mathbf{x}^{(k,l)} \geq \gamma \mathbf{x}^{(k,l-1)} \Rightarrow \mathbf{x}^{(k,l-1)} \leq \frac{\mathbf{x}^{(k,l)}}{\gamma}$. This with (16) implies that there is $K > 0$ such that:

$$x_j^{(k,C)} \geq x_j^{(k-1,C)} + \frac{\lambda_k N}{\sum_{i=1}^{n} a_{ij}} \Bigg\{ -K x_j^{(k,C-1)} + x_j^{(k,C-1)} a_{Ij} \left(\frac{b_I}{(A\mathbf{x}^{(k,C-1)})_I} - 1 \right) \Bigg\}. \quad (25)$$

At this point, it suffices to draw our attention to the set of indices $J_I := \{j | a_{Ij} > 0\}$ because they are the only that have influence in $(A\mathbf{x})_I$. Let $\epsilon_I > 0$ be such that $a_{Ij} \left(\frac{b_I}{\epsilon_I} - 1 \right) > K, \forall j \in J_I$. Now we proceed by splitting the possibilities in two cases:

1. If $(A\mathbf{x}^{(k,C-1)})_I \geq \epsilon_I$ then $(A\mathbf{x}^{(k,C)})_I \geq \gamma \epsilon_I$;
2. The remaining possibility is $(A\mathbf{x}^{(k,C-1)})_I < \epsilon_I$. However, this implies $x_j^{(k,C)} > x_j^{(k-1,C)}, \forall j \in J_I$ due to (25) and the construction of ϵ_I. This clearly means $(A\mathbf{x}^{(k,C)})_I > (A\mathbf{x}^{(k-1,C)})_I$. This lower limit can be used to a recursive application of this analysis. If no lower bound is met going to case 1 we have $(A\mathbf{x}^{(k,C)})_I > (A\mathbf{x}^{(0,C)})_I \geq \gamma^C (A\mathbf{x}^{(0)})_I$.

The above arguments imply that for all $\mu < \min\{\gamma \epsilon_I, \gamma^C (A\mathbf{x}^{(0)})_I\}$ then $(Ax^{(k,C)})_I > \mu$, $\forall k, l$. Once $\epsilon_I > 0$ and $(A\mathbf{x}^{(0)})_I > 0$ then μ can be set higher than zero and thus by proposition 2 the result is proved.□

The importance of this proposition can be expressed in the following corollary:

Corollary 1 *Let* $\mathbf{C} = \overline{\text{conv}} \left\{ \mathbf{x}^{(k)} \right\}$*, where* $\mathbf{x}^{(k)}$ *is the sequence generated by RAMLA with assumptions* **(ii)-(vi)** *satisfied. Then* $L(\mathbf{x})$*,* $L_l(\mathbf{x})$*,* $\nabla L(\mathbf{x})$ *and* $\nabla L_l(\mathbf{x})$ *are bounded and Lipschitz continuous in an open set containing* $\mathbf{C}$*.*

PROOF: Demonstration follows easily from the expression to the functions and gradients and proposition 3.□

Proposition 4 *The sequence* $\mathbf{x}^{(k)}$*, generated under the assumptions of proposition 3 is bounded.*

PROOF: It is sufficient to prove boundedness for each of the sequences $x_J^{(k,l)}$. First we note that as a consequence of the first equality in (17) and the proposition 3 above, $\exists \Gamma > 0$ such that $x_J^{(k,l)} \leq \Gamma x_J^{(k,l-1)}$ $\forall k, l$. Let now $M > 0$ be such that $\frac{b_i}{a_{iJ} M} < 1$ $\forall i | a_{iJ} > 0$. Observing again the first equality in (17) we see that if $x_J^{(k,l-1)} \geq M$ then $x_J^{(k,l)} < x_J^{(k,l-1)}$. Joining both rationales we can conclude that an upper bound for the sequence $x_J^{(k,l)}$ is $\max \left\{ \Gamma M, x_J^{(0)} \right\}$.□

Now we state the final result:

Theorem 1 *If conditions* **(i)-(vi)** *are satisfied then the sequence* $\mathbf{x}^{(k)}$ *generated by RAMLA converges to the maximizer of* $L(\mathbf{x})$ *subject to the nonnegativity constraints.*

PROOF: The demonstration makes use of proposition 1 above to ensure nonnegativity, proposition 4 to ensure boundedness of the sequence $\mathbf{x}^{(k)}$ and corollary 1 above so that that lemma 3 in [1], which uses $\sum_{k=0}^{\infty} {\lambda_k}^2 < \infty$ and regularity conditions on $\nabla L_l(\mathbf{x}^{(k,l)})$ can be applied to our case demonstrating that $L(\mathbf{x}^{(k,l)})$ converges and that there is an accumulation point $\mathbf{x}^*$ such that $D(\mathbf{x}^*) \nabla L(\mathbf{x}^*) = \mathbf{0}$. These results, when put in proposition 3 in [2] show that $\mathbf{x}^*$ is the maximizer of L, which in turn implies, by continuity of L in $\mathbf{C}$, convergence of $L(\mathbf{x}^{(k,l)})$ and boundedness of $\mathbf{x}^{(k,l)}$, in convergence of the whole sequence.□

6 Conclusion

Since its appearance in [2], a detailed convergence proof was missing for RAMLA. In [2], convergence results follow from a strong hypothesis on the convergence of $L(\mathbf{x}^{(k,l)})$. In a more recent paper [1] the condition $\sum_{k=0}^{\infty} {\lambda_k}^2 < \infty$ was used to show convergence of $L(\mathbf{x}^{(k,l)})$ whenever $\nabla L_l(\mathbf{x})$ is Lipschitz continuous. This is not guaranteed to be the case in RAMLA and that is the reason why we introduced the condition $\frac{\lambda_{k+1}}{\lambda_k} < M$, which is likely to be respected in most practical situations. One contribution of the present paper was to show that, in most realistic situations, the needed regularity on the gradients is naturally achieved, allowing us to give a satisfactory proof of convergence without any *a priori* assumption.

We have also assessed the task of choosing a proper sequence of the relaxation parameter. The results we exhibit show that it is easy to find a sequence that meets the convergence conditions and favours the quality of the reconstruction. Our results show that even if we are not looking for full convergence of the algorithm, the use of RAMLA would be preferable over OS-EM because the quality of the former's reconstruction is often superior to that obtained by the latter for the same number of iterations.

References

[1] S. Ahn and J. A. Fessler. Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms. *IEEE Trans. Med. Imag.*, 22:613–626, 2003.

[2] J. A. Browne and A. R. D. Pierro. A row-action alternative to the em algorithm for maximizing likelihoods in emission tomography. *IEEE Trans. Med. Imag.*, 15:687–699, 1996.

[3] E. S. Helou Nt. Teoria e experimentação com métodos incrementais relaxados em tomografia por emissão. M.Sc. Dissertation (in portuguese), 2005.

[4] H. M. Hudson and R. S. Larkin. Accelerated image reconstruction using ordered subsets of projection data. *IEEE Trans. Med. Imag.*, 20:601–609, 1994.

[5] A. C. Kak and M. Slaney. *Principles of Computerized Tomographic Imaging*. IEEE Press, N. York, NY, 1999 (Electronic Copy).

[6] A. Nedić and D. P. Bertsekas. Incremental subgradient methods for nondifferentiable optimization. *SIAM J. Optim.*, 12:109–138, 2001.

[7] Y. Vardi, L. A. Shepp, and L. Kaufman. A statistical model for positron emission tomography. *Journal of the American Statistical Association*, 80(389):8–20, March 1985.