An EM-Algorithm for TOF-PET to Reconstruct the Activity and Attenuation Simultaneously

Koen Salvo¹, Michel Defrise¹, Johan Nuyts² and Ahmadrreza Rezaei²

Abstract—The ‘Simultaneous Maximum-Likelihood Attenuation Correction Factors’ (SMLACF) algorithm presented here, is an iterative algorithm to calculate the maximum-likelihood (ML) estimate of the activity $\lambda$ and the attenuation correction factors $\mu$ in time-of-flight (TOF) positron emission tomography (PET), and this from emission data only. SMLACF is derived using the expectation-maximization (EM) principle by introducing an appropriate set of complete data. The resulting iteration step yields a simultaneous update of $\lambda$ and $\mu$ which, in addition, enforces in a natural way the constraint $\mu \leq 1$. Hence, providing an alternative to the MLACF algorithm.

I. INTRODUCTION

In time-of-flight (TOF) positron emission tomography (PET) attenuation correction is required to reconstruct the activity image $\lambda$ accurately.

Currently the standard way to apply attenuation correction makes use of a computed tomography (CT) scan. However, when no CT is available, such as in stand-alone PET or PET combined with magnetic resonance imaging (MRI), one possible strategy consists in reconstructing the attenuation together with the activity from emission data only.

Current methods based on a maximum-likelihood (ML) estimation alternate (i) a usual ML expectation-maximization (EM) update of the activity $\lambda$, and (ii) an update of the linear attenuation coefficients $\mu$ for MLAA [1] or an update of the attenuation correction factors $\mu$ for MLACF [2].

Alternatively the EM principle [3] can be used to obtain a simultaneous algorithm, as in [4] for single photon emission computed tomography (SPECT) and in [5] for (TOF-)PET. Those algorithms reconstruct $\lambda$ and $\mu$. We present a simultaneous version of MLACF, which reconstructs $\lambda$ and $\mu$, and uses a simpler set of complete data than in [4], [5].

II. THE SMLACF ALGORITHM

A. Definitions

Consider TOF PET data $y = \{y_{it}\}_{it}$ binned in Lines-Of-Response (LORs) $i = 1, \ldots, J$ and TOF-bins $t = 1, \ldots, T$. The expectation value of the data bin $it$ given an activity map $\lambda = \{\lambda_j\}$, in the voxels $j = 1, \ldots, J$ and an attenuation correction factors map $\mu = \{\mu_j\}$, is modeled as

$$\langle y_{it} | \lambda, \mu \rangle = n_{it}\lambda_{it}\mu_{it} + \langle b_{it} \rangle \quad (1a)$$

where $n_{it}$ is the sensitivity of detector pair $i$, $\langle b_{it} \rangle$ the known expectation value of the background and $p_{it}$ the projected activity

$$p_{it} = \sum_{j=1}^{J} c_{ijt} \lambda_j \quad (1b)$$

with system matrix elements $c_{ijt}$. When the index $t$ is omitted, it means that the time-bins are summed, e.g. $p_i = \sum_t p_{it}$.

B. The likelihood

The goal is to estimate $\lambda$ and $\mu$ by maximizing the Poisson log-likelihood

$$L(\lambda, \mu; y) = \sum_{it} \left( -\langle y_{it} | \lambda, \mu \rangle + y_{it} \ln \langle y_{it} | \lambda, \mu \rangle \right) \quad (2)$$

This likelihood is scale invariant, i.e. $L(\beta \lambda, \beta \mu; y) = L(\lambda, \mu; y)$, $\beta > 0$. So the solution is determined up to a global scale factor [6]. Moreover this likelihood is not concave in both $\lambda$ and $\mu$ (although concave in $\lambda$ and $\mu$ separately) [7]. Hence maximizing (2) w.r.t. $\lambda$ and $\mu$ involves the risk of converging to a saddle point or a local maximum.

C. Expectation-Maximization

Direct maximization of the likelihood for the incomplete (measured) data $y$ in general is impractical because setting the derivatives to zero yields a huge set of coupled equations.

However, if appropriate (unmeasured) complete data $x$ would have been available, setting the derivatives of the resulting complete data likelihood $L_x$ to zero, yields a set of equations that is uncoupled in the unknowns, and hence, easy to solve.

Because the complete data are not known, the expectation values of the complete data, given the measured data and the current estimate of the unknowns, are calculated in a first expectation step. This complete data are then substituted in $L_x$ and in a second maximization step, $L_x$ is maximized. Repeating those two steps concludes the Expectation-Maximization (EM) algorithm [3] and maximizes the original likelihood (2).

D. Choice of complete data

The ML-EM algorithm for (TOF-)PET with known attenuation was derived to reconstruct $\lambda$ by choosing as complete data the number of measured photon-pairs $x_{ijt1}$ that originated from voxel $j$ and were emitted along LOR $i$ (and TOF-bin $t$) [8].

To reconstruct $\mu$ as well, we propose to add a second set of complete data: the number of photon-pairs $x_{ijt0}$ that originated
from voxel $j$, were emitted along LOR $i$ and TOF-bin $t$, but were **attenuated**.

$$
\langle x_{ijt} \mid \lambda_j, a_i \rangle = \begin{cases} 
  n_i a_i c_{ijt} \lambda_j : \alpha = 1 \\
  n_i (1 - a_i) c_{ijt} \lambda_j : \alpha = 0
\end{cases} \quad (3)
$$

Assuming that the activities in every voxel are Poisson distributed, the complete data will be Poisson distributed too. Moreover all the complete data are independent of each other. Hence the complete data likelihood is given by

$$
L_x (\lambda, \alpha; x) = \sum_{ijt} \left( \langle x_{ijt} \mid \lambda_j, a_i \rangle + x_{ijt} \ln \langle x_{ijt} \mid \lambda_j, a_i \rangle \right),
$$

which is simpler than the likelihood obtained in [5].

1) **Expectation step**: As $y_{ijt} = \sum_j x_{ijt}$, the measured data $y$ give additional information about the unattenuated complete data $x_{ijt}$. They do not however provide information about the **attenuated** complete data $x_{ijt}$.

So, with $\hat{\lambda}^k$ and $\hat{a}^k$ the current estimates at iteration $k$, the expectation values of the complete data are

$$
\langle x_{ijt} \mid \hat{\lambda}^k, \hat{a}^k; y_{ijt} \rangle = \begin{cases} 
  n_i \hat{a}_i c_{ijt} \hat{\lambda}_j : \alpha = 1 \\
  n_i (1 - \hat{a}_i) c_{ijt} \hat{\lambda}_j : \alpha = 0
\end{cases} \quad (5)
$$

2) **Maximization step**: Plugging the expectations (5) into $L_x$ (4) and setting the derivatives w.r.t. $\lambda$ and $\alpha$ to zero, gives the simultaneous update of the sMLACF algorithm

$$
\hat{\lambda}_j^{k+1} = \hat{\lambda}_j^k \frac{\sum_{it} c_{ijt} n_i \left( 1 - \hat{a}_i^k \right) + \gamma_i^k y_{ijt} / \langle y_{ijt} \rangle}{\sum_{it} c_{ijt} n_i} \quad (6a)
$$

$$
\hat{a}_i^{k+1} = \hat{a}_i^k \frac{\gamma_i^k}{1 + \hat{a}_i^k (\gamma_i^k - 1)} \quad (6b)
$$

with

$$
\gamma_i^k = \frac{1}{\sum_{it} \hat{p}_{ijt}} \sum_{it} \hat{p}_{ijt} \frac{y_{ijt}}{\langle y_{ijt} \rangle} \quad (7)
$$

E. Basic properties

1) **Monotonicity**: Because (6) is an EM-algorithm, it is ensured that the likelihood increases during every iteration [3].

2) **Naturally constrained solution**: With all $\hat{\lambda}_j^k \geq 0$ and $0 \leq \hat{a}_i^k \leq 1$, also $\hat{\lambda}_j^{k+1} \geq 0$ and $0 \leq \hat{a}_i^{k+1} \leq 1$. Note that both the MLAA and MLACF updates enforce the positivity of $\hat{\lambda}_j^{k+1}$ and $\hat{a}_i^{k+1}$, but —when no additional truncation is applied— do not enforce that $\hat{a}_i^{k+1} \leq 1$.

3) **Scaled gradient**: The updates (6) can be rewritten as

$$
\hat{\lambda}_j^{k+1} = \hat{\lambda}_j^k + \frac{\lambda_j^k}{\sum_{it} c_{ijt} n_i} \cdot \frac{\partial L}{\partial \lambda_j} \left( \hat{\lambda}_j^k, \hat{a}^k \right) \quad (8a)
$$

$$
\hat{a}_i^{k+1} = \hat{a}_i^k + \frac{1 - \hat{a}_i^k}{\left( 1 - \hat{a}_i^k \right) + \hat{a}_i^k \lambda_i^k} \frac{\partial L}{\partial a_i} \left( \hat{\lambda}_j^k, \hat{a}^k \right) \quad (8b)
$$

Compared to ML-EM, the activity update step is scaled with a factor $\sum_{it} c_{ijt} n_i \hat{a}_i^k / \sum_{it} c_{ijt} n_i$. Because $0 \leq \hat{a}_i^k \leq 1$, the sMLACF activity-update is smaller and one expects a slower convergence of $\lambda$. Compared to MLACF, the $a$ update step is scaled by $\xi_i^k = \left( 1 - \hat{a}_i^k \right) / \left( 1 - \hat{a}_i^k + \hat{a}_i^k \gamma_i^k \right) \leq 1$. So also the convergence of $\alpha$ will be slower. Note that $\xi_i^k$ tends to zero when $\hat{a}_i^k$ tends to one.

4) **Preservation of summed attenuation-corrected data**: From (6) it follows that

$$
\sum_{i} \langle \hat{a}_{i}^{k+1} \rangle \approx \sum_{i} \hat{y}_{i} / \hat{a}_{i}^{k+1} \quad (9)
$$

where the similarity $\simeq$ becomes an equality when there is no background.

5) An interior fixed point is also a stationary point: For a fixed point \( \left( \hat{\lambda}_j^{k+1}, \hat{a}^{k+1} \right) = (\hat{\lambda}_j^k, \hat{a}^k) \) such that $\hat{\lambda}_j^k > 0$, $j = 1, \ldots, J$ and $0 < \hat{a}_i^k < 0$, $i = 1, \ldots, I$, it follows from (6) that the gradient of the likelihood (2) is zero.

III. RESULTS

Figure 1 shows reconstructions of FDG Siemens mCT brain data after 10 iterations with 12 subsets. Using a uniform attenuation map $\mu_0$ (the support of the object filled with water), ML-EM overestimates the activity in regions with a small attenuation like air cavities, see fig. 1 (d). Both MLACF and sMLACF alleviate this, see fig. 1 (e-f). As expected the sMLACF converges more slowly. Strategies to accelerate the convergence are under study.

![Figure 1](image-url)

**Figure 1.** Sagittal section of the attenuation (a-b) and activity (c-f) of the brain. Note the air cavity in the dashed circle.

REFERENCES


