Simultaneous reconstruction of attenuation and activity in TOF-PET: analysis of the convergence of the MLACF algorithm.

Michel Defrise[1], Ahmadreza Rezaei[2], Johan Nuyts[2]

I. Introduction

Various algorithms have been recently proposed to estimate both the attenuation and the activity from time-of-flight (TOF) PET emission data [1-6], without requiring additional CT or MR information. Initial tests with simulated, phantom, and clinical data demonstrate that this simultaneous estimation problem can be solved with a surprisingly good accuracy. These results corroborate previous works demonstrating that the TOF emission data contain significant information on the attenuation [7,8], and they offer promising perspectives for various applications including for instance studies with partially known or mismatched CT data [5,6].

This work presents a mathematical and numerical analysis of the convergence of MLACF [4], a maximum likelihood algorithm that jointly estimates the activity distribution and the attenuation factors. This algorithm does not reconstruct the attenuation image, in contrast with the MLAA algorithm [3]. We consider the simplest case where there is no background due to scatter or random events, and demonstrate that i/ the MLACF algorithm is monotonic and asymptotically regular, ii/ if the likelihood function has a unique global maximum and if there are no local maxima, MLACF converges to the global maximizer. Although the mathematical analysis of the continuous version of the problem showed that noise free emission data determine the attenuation correction factors up to a scale factor [1,2,6], it is still unclear whether uniqueness also applies to the Poisson likelihood, except if the TOF data are consistent (uniqueness in that case was proven in [4]). Therefore, we present a numerical analysis of convergence on a simple 2D problem, the results of which suggests that local maxima of the likelihood, if any, would only occur for extremely low count data.

II. The MLACF algorithm

Let $\lambda_j \geq 0, j = 1, \ldots, M$ be the unknown activity image discretized on a grid of M voxels, and let $y_{i,t} \in I\!N_+, i = 1, \ldots, N, t = 1, \ldots, T$ be the measured emission data for line of response i and TOF bin t. We wish to find an activity estimate λ^* and an attenuation sinogram estimate a^* which maximize the log-likelihood

$$L(y, \lambda, a) = \sum_{i=1}^{N} \sum_{t=1}^{T} \{-a_i \, p_{i,t} + y_{i,t} \log(a_i \, p_{i,t})\} \qquad (1)$$

where $c_{i,j,t} \geq 0$ is the known system matrix, $0 < a_i \leq 1$ are unknown attenuation factors, and $p_{i,t} = \sum c_{i,j,t} \lambda_j$ is the expectation of the unattenuated data. Note the scale invariance $L(y, \lambda, a) = L(y, \alpha\lambda, a/\alpha)$ for any $\alpha > 0$, and the fact that L is concave in both λ and a, but not jointly concave in λ, a. Owing to the scale invariance, the constraint $a_i \leq 1$ can be ignored provided the unconstrained optimizer belongs to the open set

$$H = \left\{ x \in I\!R^M \,|\, x_j \geq 0, \text{ and } p_i = \sum_j c_{i,j} x_j > 0 \text{ if } y_i > 0 \right\} \qquad (2)$$

Thus ignoring the constraint, the likelihood is easily maximized w.r.t. a at fixed activity λ, yielding the ML estimate of the attenuation factor,

$$a_i^* = \frac{y_i}{p_i} \quad i = 1, \ldots, N \qquad (3)$$

where $y_i = \sum_t y_{i,t}$ and similarly $p_i = \sum_t p_{i,t}$. The same convention is used everywhere: quantities indexed by i denote quantities summed over the TOF bins. Substituting a^* in L reduces the problem to the maximization of the *reduced log-likelihood* (see [4] for details),

$$\tilde{L}(y, \lambda) = \sum_{i=1}^{N} \sum_{t=1}^{T} y_{i,t} \, \log \frac{p_{i,t}}{p_i}. \qquad (4)$$

The MLACF algorithm is based on optimization transfer [9], it maximizes at iteration n the following concave and separable surrogate of $\tilde{L}$:

$$\begin{aligned} \tilde{L}^{sur}(y, \lambda, \lambda^n) &= \sum_{i=1}^{N} \Big(- y_i \, \log p_i^n - \frac{y_i \, (p_i - p_i^n)}{p_i^n} \\ &+ \sum_{t=1}^{T} \sum_{j=1}^{M} \frac{y_{i,t} \, c_{i,j,t} \lambda_j^n}{p_{i,t}^n} \log(\frac{\lambda_j p_{i,t}^n}{\lambda_j^n}) \Big) \end{aligned} \qquad (5)$$

with $p_{i,t}^n = \sum_j c_{i,j,t} \lambda_j^n$ and $p_i^n = \sum_t p_{i,t}^n$, yielding the iteration [4]

$$\lambda_j^{n+1} = T(\lambda^n)_j = \frac{\lambda_j^n}{\sum_{i=1}^{N} \frac{y_i \, c_{i,j}}{p_i^n}} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{y_{i,t} c_{i,j,t}}{p_{i,t}^n}. \qquad (6)$$

[1]Dept. of Nuclear Medicine, Vrije Universiteit Brussel, B-1090, Brussels, Belgium, mail: mdefrise@vub.ac.be, [2]Dept. of Nuclear Medicine, Katholieke Universiteit Leuven, B-3000, Leuven, Belgium, mail: ahmadreza.rezaei@uzleuven.be, johan.nuyts@uzleuven.be.

Like ML-EM, this algorithm maintains strict positivity if the initial activity estimate is strictly positive, this property also guarantees the absence of singularity because $p_{i,t}^n > 0$. However, convergence to zero cannot be excluded, i.e. one may have $\lambda_j^n \to 0$ for some voxels. Since $\tilde{L}$ is scale invariant, we also define the normalized MLACF iteration $T_{\mathcal{N}}(\lambda) = T(\lambda)/||T(\lambda)||$, with $||x||$ the euclidean norm in $\mathbb{R}^M$. See [4,5] for the extension of MLACF to the case with background, which will not be considered here.

Some notations will be needed. We define the active set of data bins for a given voxel as:

$$\tau_j = \{(i,t) \,|\, c_{i,j,t} > 0 \text{ and } y_{i,t} > 0\} \quad j = 1, \ldots, M \tag{7}$$

and adopt an equivalent definition for the TOF-summed data,

$$\iota_j = \{i \,|\, c_{i,j} > 0 \text{ and } y_i > 0\} \quad j = 1, \ldots, M. \tag{8}$$

Note that if $\tau_j = \emptyset$,

$$\frac{\partial \tilde{L}(y,\lambda)}{\partial \lambda_j} = \sum_{i \in \iota_j} \left\{ -\frac{y_i c_{i,j}}{p_i} \right\} \leq 0 \tag{9}$$

and therefore any maximizer of the reduced log-likelihood satisfies $\lambda_j = 0$ (because of the non-negativity constraint). If in addition $\iota_j = \emptyset$ the reduced log-likelihood does not depend on the value of voxel j, the maximizer is undefined and as a logical (but arbitrary) estimate we take $\lambda_j = 0$. These voxels without active data can be set to zero and must not be further considered when maximizing the reduced log-likelihood. Therefore we assume below that $\tau_j \neq \emptyset$ and $\iota_j \neq \emptyset$ for $j = 1, \ldots, M$. We also assume that $y_i \geq 1$ and $y_{i,t} \geq 1$ for all non-zero data bins. Finally we define the constants

$$\begin{aligned} \xi &= \min_j \min_{i \in \iota_j} c_{i,j} > 0 \quad , \eta = \min_j \min_{(i,t) \in \tau_j} c_{i,j,t} > 0 \\ \omega &= \max_i \sum_j c_{i,j} \quad , \quad \sigma = \max_{i,t} \sum_j c_{i,j,t} > 0. \end{aligned} \tag{10}$$

III. Properties of MLACF

Though previous tests with MLACF showed good convergence and encouraging practical results, two questions remain unanswered: i/ is the maximizer of the reduced likelihood unique, are local maxima or saddle points possible?, and ii/ does MLACF converge? This section investigates the latter question, while the former is adressed through numerical tests in section IV. Here we summarize our main results, including only brief comments on the proofs.

Lemma 1. The sequence of *normalized* iterates $\lambda^{n+1} = T_{\mathcal{N}}(\lambda^n)$ with $\lambda_j^0 > 0$ is such that the sequence $\tilde{L}(y, \lambda^n), n = 0, 1, 2, \ldots$ is non-decreasing and converges.

• This is the standard monotonicity property of surrogate-based algorithms, and convergence uses the upper bound $\tilde{L}(\lambda) \leq 0$, which follows from $p_{i,t} \leq p_i$.

Lemma 2. Let $\tilde{\lambda} \in \mathbb{R}^M$ be any positive vector with $||\tilde{\lambda}|| = 1$.

$$\tilde{L}(y, T(\tilde{\lambda})) - \tilde{L}(y, \tilde{\lambda}) \geq C \frac{||T(\tilde{\lambda}) - \tilde{\lambda}||^2}{\max(1, ||T(\tilde{\lambda})||)} \tag{11}$$

with T the mapping (6) and $C = (1/2)\min(\xi/\omega, \eta/\sigma) > 0$.

• The proof follows the same line as in [10]. It is based on a 2-term Taylor development of the surrogate $\tilde{L}^{sur}(y, x, \tilde{\lambda})$ around its unique maximizer $T(\tilde{\lambda})$, and on a lower bound on the curvature of the surrogate (5).

Lemma 3. Asymptotic regularity. The sequence of normalized iterates $\lambda^{n+1} = T_{\mathcal{N}}(\lambda^n)$ with $\lambda_j^0 > 0$ is such that

$$\begin{aligned} &\lim_{n \to \infty} ||T(\lambda^n) - \lambda^n|| = \lim_{n \to \infty} ||T_{\mathcal{N}}(\lambda^n) - \lambda^n|| = 0 \\ &\lim_{n \to \infty} ||T(\lambda^n)|| = 1. \end{aligned} \tag{12}$$

• The proof follows easily from the previous Lemma. Note that asymptotic regularity is not sufficient to prove convergence of the algorithm.

Proposition 4. The sequence of *normalized* iterates $\lambda^{n+1} = T_{\mathcal{N}}(\lambda^n)$ with $\lambda_j^0 > 0$ has an accumulation point λ^*. If $\lambda^* \in H$ (see eq. (2)), then $\nabla_j \tilde{L}(y, \lambda^*) = 0$ for any voxel satisfying $\lambda_j^* > 0$. All accumulation points of the sequence $\lambda^{n+1} = T_{\mathcal{N}}(\lambda^n)$ have the same value of the reduced log-likelihood.

• It seems reasonable to assume that a limit point belongs to the set H because otherwise there are lines of response i for which events have been detected ($y_i > 0$) but which only receive contributions from voxels with zero activity ($p_i = 0$). This theorem does not guarantee convergence, and even if there is convergence, it may be to a local maximum or to a saddle point. Further investigation on the existence of saddle points and local maxima is therefore warranted. In favorable cases, one has the

Corollary 5. If the only stationary point of the reduced likelihood is a unique global maximum $\lambda^\dagger \in H$, then the sequence of *normalized* iterates $\lambda^{n+1} = T_{\mathcal{N}}(\lambda^n)$ with $\lambda_j^0 > 0$ converges to $\lambda^\dagger$.

IV. Numerical results

A. Simulation parameters

We digitized a 2D thorax phantom on a $M = 64 \times 64$ image with pixel size 8.027 mm. Simulated TOF-PET data were generated by forward projecting this phantom with radial pixel size 8.027 mm, 64 angular samples on $[0, \pi)$, and $T = 8$ TOF bins with sampling $\Delta\tau = 64.0$ mm. The TOF profile was a gaussian with FWHM 80 mm. The aim of this study is to get insight into convergence and uniqueness, hence this coarse discretization was chosen to allow performing a very large number of iterations with various initial estimates λ^0. All calculations were done in double precision, using a matched backprojector and without data subset. The emission and attenuation phantoms are shown in Figure 1. The phantom support is an ellipse with axes of 300 mm and 470 mm and the minimum attenuation factor a_i was 0.015. A vial with activity 0.5, diameter 40 mm, and water attenuation was added outside the phantom and used to scale the reconstructed activity image after reconstruction. Poisson noise was added to the data to generate three noisy data sets S1, S2 and S3 with respectively a total of 479705, 15990 and 3198 events, corresponding to

respectively 300, 10, and 2 average events in the maximum data bin $y_{i,t}$. A large number of data bins are equal to zero in data set S3 (see Figure 2), allowing to challenge the algorithm's behaviour at the edge of the admissible domain.

The MLACF iteration was run up to 10^5 iterations, starting with a uniform image estimate $\lambda_j^0 = 1$ and with a set of 20 random initial images generated as $\lambda_j^0 = 1 + 0.8R$ where R is a pseudo-random number with uniform distribution in $(0, 1)$.

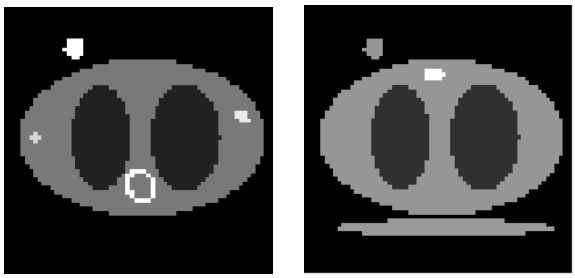

Fig. 1. The simulated phantom. Emission (left): activity is 0.2 (background tissues), 1.7 ("heart"), 0.05 ("lungs"), 0.40 and 0.45 ("tumors"), and 0.5 (vial). Attenuation (right): 0.00966/mm ((background tissues and vial), 0.00266/mm ("lungs"), 0.0187/mm ("spine"), and 0.01/mm ("bed").

Fig. 2. One central TOF-bin sinogram of the high noise data set S3, showing the large number of data bins for which no event have been detected.

B. Results

Figure 3 shows the convergence of the reduced log-likelihood cost function, with the three data sets and three of the random initial images. Note the irregular convergence, especially with data set S2. This phenomenon is tentatively attributed to the existence of regions where $\tilde{L}$ is almost flat for this specific data set. This behaviour is not observed when the same data sets are reconstructed with ML-EM assuming exact knowledge of the attenuation factors (Figure 4). Figure 5 shows numerical evidence for the asymptotic regularity of MLACF.

The three data sets have been reconstructed with MLACF, using a uniform initial activity and 20 random images. The relative difference between the maximum and minimum of these 21 values of the reduced likelihood after 10^5 iterations was $3.3\,10^{-14}$, $1.5\,10^{-9}$ and $1.77\,10^{-4}$ for S1, S2 and S3 respectively. The maximum relative RMSE difference between all pairs of images among the 21 reconstructions was $1.73\,10^{-8}$, $1.29\,10^{-7}$ and 0.25 for S1, S2 and S3 respectively. The difference between the two images corresponding to these maximum RMSE is negligible visually for S1 and S2. Even for the highest noise data set S3, the difference is small (Figure 6) and unlikely to be relevant in practice, where regularization would be applied. Nevertheless the plot of the RMSE between all pairs of images (Figure 7) reveals clusters; it is unclear at this point whether this observation reflects a lack of convergence, the digitization of the Poisson data (the number of events in the data bins of S3 take only 6 different values, between 0 and 5), or the convergence to different local maxima of the cost function. For the two other data sets, the corresponding RMSE values (plot not shown) are not only much smaller (see above) but do not show any structure or clustering.

Finally the loss of image quality caused by the absence of prior knowledge of the attenuation is illustrated for our example in Figure 8, which compares the MLACF and ML-EM reconstructions "at convergence" (meaning here 10^5 iterations).

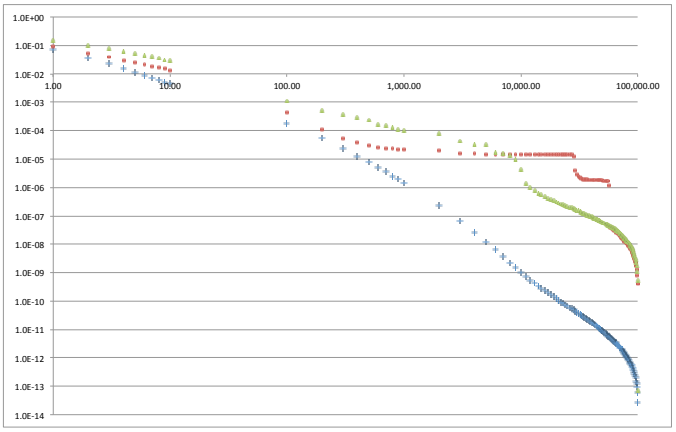

Fig. 3. The reduced log-likelihood $|\tilde{L}(y, \lambda^n) - \tilde{L}(y, \lambda^{100000})|/|\tilde{L}(y, \lambda^0)|$ for the data sets S1 (low noise, blue $+$), S2 (red circles), and S3 (high noise, green triangles). The horizontal scale is the number of MLACF iterations n.

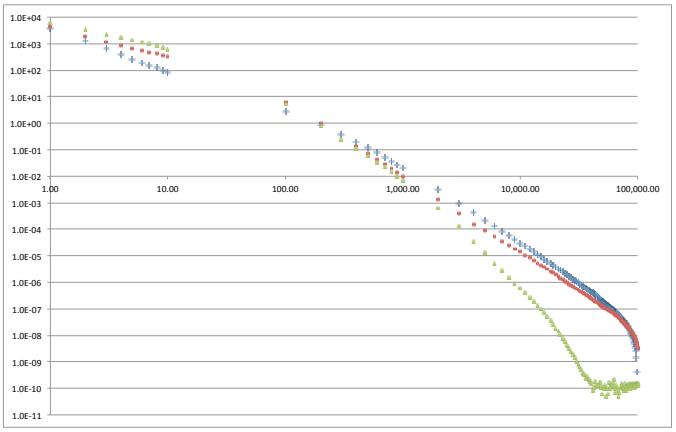

Fig. 4. The log-likelihood $|L(y, \lambda^n, a) - L(y, \lambda^{100000}, a)|$ for the data sets S1 (low noise, blue $+$), S2 (red circles), and S3 (high noise, green triangles) reconstructed using ML-EM with the exact attenuation factors a_i. The horizontal scale is the number of ML-EM iterations n.

V. Conclusion

In practice the MLACF algorithm would be stopped before convergence and/or would be regularized by adding a smoothing penalty. Other modifications include an accelerated ordered-subset implementation and the generalization to account for a scatter or random background [4,5]. Understanding the convergence properties of the basic MLACF algorithm

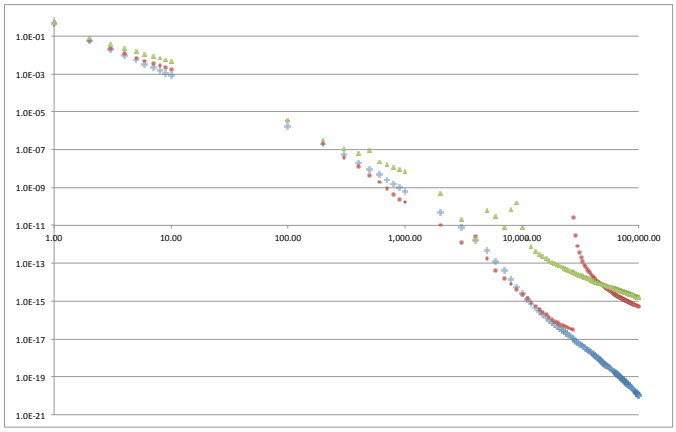

Fig. 5. The value of $||\lambda^{n+1} - \lambda^n||^2/||\lambda^n||^2$ for the data sets S1 (low noise, blue $+$), S2 (red circles), and S3 (high noise, green triangles). The horizontal scale is the number of MLACF iterations n.

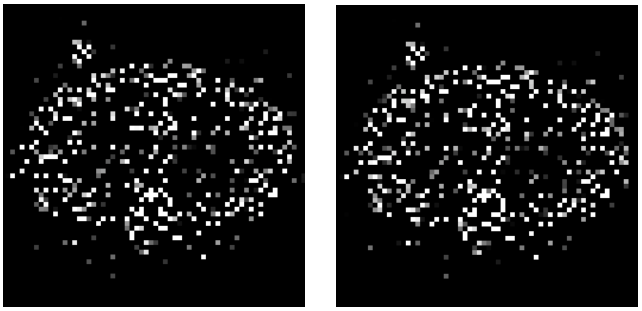

Fig. 6. The activity image reconstructed from data set S3 using the two initial images that yielded the largest difference. Grey scale $(0, 0.5)$.

(6) is nevertheless important. In this work we showed that MLACF is a monotonic and asymptotically regular algorithm, and that it converges to the maximum likelihood estimate of the activity image if the only stationary point of the reduced likelihood is a unique global maximizer, provided this maximizer satisfies condition (2). The results of the numerical study agree with these mathematical properties.

Uniqueness, up to the global scale factor, of the ML estimate is an open problem. The discrete Poisson model (1) investigated here does not include any specification on the range of the system matrix $c_{i,j,t} : \mathbb{R}^M \to \mathbb{R}^{N\times T}$. Clearly, uniqueness cannot hold in such a general setting, and counter-examples can easily be found, e.g. by building two disconnected voxel subsets that are not connected by any "active" ($y_{i,t} > 0$) line of response. In [2], the proof of uniqueness for the continuous model uses the range (consistency) conditions for the TOF Radon transform, and similar conditions will need to be added to the discrete model to give any hope of proving uniqueness. The numerical results seem to indicate that uniqueness does hold for the simple 2D problem considered, except possibly for very low statistics data such as S3.

REFERENCES

[1] A. Rezaei et al. Simultaneous reconstruction of activity and attenuation in time-of-flight PET, IEEE Nucl Sci Symp Conf Record, Valencia, Oct 2011, MIC8-7.

[2] M Defrise et al. Time-of-flight PET data determine the attenuation sinogram up to a constant, Phys. Med. Biol., 2012, 57 (4), 885-899.

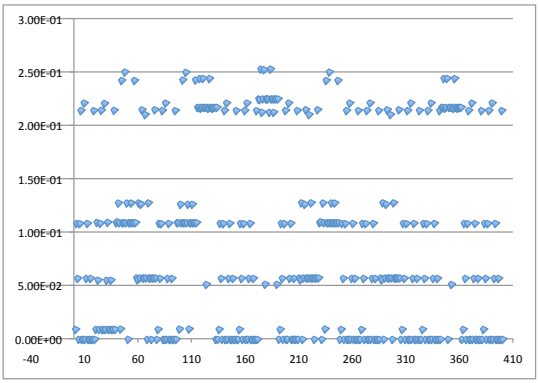

Fig. 7. Unordered plot of the relative RMSE differences between all pairs of images reconstructed from data set S3 with 100000 iterations of MLACF and various initial images. All points appear twice.

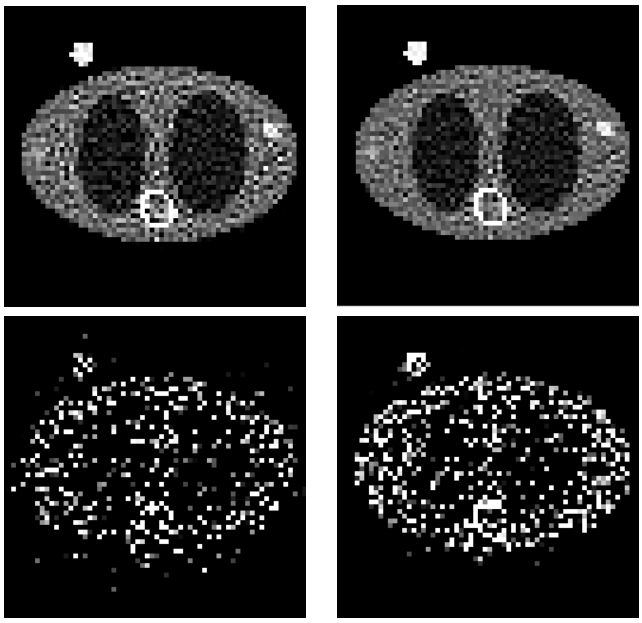

Fig. 8. The emission image reconstructed from data set S1 (top row) and S3 (bottom row) with 10^5 iterations. Left: MLACF. Right: ML-EM with known attenuation. Uniform initial images. Grey scale $(0, 0.5)$.

[3] A. Rezaei, et al, "Simultaneous reconstruction of activity and attenuation in time-of-flight PET", IEEE Trans. Med. Imag., 2012, 31 (12), 2224 - 2233.

[4] J. Nuyts et al., "ML-reconstruction for TOF-PET with simultaneous estimation of the attenuation factors", IEEE Nucl Sci Symp Conf Record, Anaheim, Oct 2012, M04-7.

[5] V. Panin et al."Reconstruction of Uniform Sensitivity Emission Image with Partially Known Axial Attenuation Information in PET-CT Scanners", IEEE Nucl Sci Symp Conf Record, Anaheim, Oct 2012, M04-1

[6] S. Ahn et al., "Convergent Iterative Algorithms for Joint Reconstruction of Activity and Attenuation from Time-of-Flight PET Data", IEEE Nucl Sci Symp Conf Record, Anaheim, Oct 2012, M19.5.

[7] M Conti. Why is TOF PET reconstruction a more robust method in the presence of inconsistent data, Phys. Med. Biol., 2011, 56, 155-168.

[8] A. Salomon et al. Simultaneous reconstruction of activity and attenuation for PET/MR, IEEE Trans Med Imaging, 2011, 30, 804-813.

[9] K. Lange et al., "Optimization transfer algorithms using surrogate objective functions", J. Comput. Graph. Stat., 2000, 9, 1-59.

[10] A. De Pierro, "A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography", IEEE Trans. Med. Imag., 1995, 14, 132-137.