Analysis and Control of the Accuracy and Convergence of the ML-EM Iteration

Milán Magdics[1,2], László Szirmay-Kalos[2], Balázs Tóth[2], Anton Penzov[3]

[1] University of Girona, Spain
[2] Budapest University of Technology, Hungary
[3] Bulgarian Academy of Sciences, Bulgaria

Abstract. In inverse problems like tomography reconstruction we need to solve an over-determined linear system corrupted with noise. The ML-EM algorithm finds the solution for Poisson noise as the fixed point of iterating a forward projection and a non-linear back projection. In tomography we have several hundred million equations and unknowns. The elements of the huge matrix are high-dimensional integrals, which cannot be stored, but must be re-computed with Monte Carlo (MC) quadrature when needed. In this paper we address the problems of how the quadrature error affects the accuracy of the reconstruction, whether it is possible to modify the back projection to speed up convergence without compromising the accuracy, and whether we should always take the same MC estimate or modify it in every projection.

1 Introduction

In Positron Emission Tomography (PET) we need to find the spatial density of radioactive tracer materials [4]. The tracer density is computed from the statistics of detected hits, which is the *inverse problem* of particle transport in scattering and absorbing media. Inverse problems are usually solved iteratively, by alternating the simulation of the forward problem and a correction step.

The output of the reconstruction is the activity density that is defined on a 3D voxel grid $\mathbf{x} = (x_1, x_2, \ldots, x_{N_{voxel}})$. The inputs of the reconstruction algorithm are the measured coincident photon hits in detector pairs, called LORs: $\mathbf{y} = (y_1, y_2, \ldots, y_{N_{LOR}})$. Using *maximum likelihood estimation* (ML-EM), vector $\mathbf{x}$ is found by maximizing the probability of the actually measured data $\mathbf{y}$ [5], alternating forward projection and back projection that together update estimate $\mathbf{x}^{(n)}$:

$$\text{Forward: } \tilde{\mathbf{y}} = \mathbf{A} \cdot \mathbf{x}^{(n)}, \qquad \text{Back: } \frac{\mathbf{x}^{(n+1)}}{\mathbf{x}^{(n)}} = \frac{\mathbf{A}^T \cdot \frac{\mathbf{y}}{\tilde{\mathbf{y}}}}{\mathbf{A}^T \cdot \mathbf{1}}$$

where vector division is defined in an element-wise manner, $\mathbf{A}_{LV}$ or $\mathbf{A}$ in short is the *System Matrix* (*SM*), which is the probability that a photon pair born in voxel V is detected by LOR L of expected value $\tilde{y}_L$.

The true solution $\mathbf{x}^*$ of the reconstruction is the fixed point of this scheme, which satisfies:

$$\mathbf{A}^T \cdot \frac{\mathbf{y}}{\mathbf{A} \cdot \mathbf{x}^*} = \mathbf{A}^T \cdot \mathbf{1}. \tag{1}$$

In tomography we have several hundred million LORs and voxels, thus an SM may have more than 10^{16} elements. To handle the huge SM, it can be *factored* [3], and simpler physical phenomena may be obtained by on-the-fly analytic approximations. However, as these approximations are part of an iteration process, even a small error can accumulate unacceptably. Accurate and consistent estimations can be obtained with MC quadrature, but its computational burden is high [1]. There is an important difference between applying MC for estimating a quadrature and using MC as a part of an iteration process [6]. While the goal is an integral quadrature, the convergence rate is known and the error can be minimized by variance reduction techniques and increasing the number of samples. When MC is applied in an iteration, the accuracy of a single estimate is not so relevant since later iteration steps may correct the error of an earlier estimate. However, decreasing the samples in a single step means that we can make more iterations under the given budget of samples or computation time.

This paper examines the process of iteration with random MC estimates. Furthermore, we also investigate the potential of using simplified back projection matrices to speed up the projection.

2 Error and convergence analysis

SM estimations may be different in forward projection and back projection, and due to the numerical errors both differ from the exact matrix. Let us denote the forward projection SM by $\mathbf{F} = \mathbf{A} + \mathbf{\Delta F}$ and the back projection estimation by $\mathbf{B} = \mathbf{A} + \mathbf{\Delta B}$. We use the following notations for the normalized back projectors

$$\bar{\mathbf{A}}_{LV} = \frac{\mathbf{A}_{LV}}{\sum_{L'} \mathbf{B}_{L'V}},\ \bar{\mathbf{B}}_{LV} = \frac{\mathbf{B}_{LV}}{\sum_{L'} \mathbf{B}_{L'V}} \implies \bar{\mathbf{B}} = \bar{\mathbf{A}} + \mathbf{\Delta}\bar{\mathbf{B}} \text{ and } \mathbf{\Delta}\bar{\mathbf{B}} \cdot \mathbf{1} = \mathbf{0}.$$

Note that $\mathbf{\Delta}\bar{\mathbf{B}} \cdot \mathbf{1} = \mathbf{0}$ is the consequence of the normalization of matrix $\mathbf{\Delta}\bar{\mathbf{B}}$, i.e. each element is divided by the row sum.

The question is how these approximations modify the convergence and the fixed point of the iteration scheme. Let us express the activity estimate in step n as $\mathbf{x}^{(n)} = \mathbf{x}^* + \mathbf{\Delta x}^{(n)}$. Substituting this into the iteration formula and replacing the terms by first order Taylor's approximations we obtain:

$$\mathbf{\Delta x}^{(n+1)} \approx \left(\mathbf{1} - \langle x_V^* \rangle \cdot \bar{\mathbf{B}}^T \cdot \langle \frac{y_L}{\tilde{y}_L^2} \rangle \cdot \mathbf{F}\right) \cdot \mathbf{\Delta x}^{(n)} + \langle x_V^* \rangle \cdot \bar{\mathbf{B}}^T \cdot \langle \frac{y_L}{\tilde{y}_L} \rangle \cdot \frac{\mathbf{\Delta}\tilde{\mathbf{y}}}{\tilde{\mathbf{y}}} - \mathbf{\Delta}\bar{\mathbf{B}}^T \cdot \frac{\mathbf{y}}{\tilde{\mathbf{y}}}.$$

where $\langle x_V^* \rangle$ is an N_{voxel}^2 element diagonal matrix of true voxel values, $\langle \frac{y_L}{\tilde{y}_L^\alpha} \rangle$ is an N_{LOR}^2 element diagonal matrix of ratios $\frac{y_L}{\tilde{y}_L^\alpha}$, and $\mathbf{\Delta}\tilde{\mathbf{y}} = \mathbf{\Delta F} \cdot \mathbf{x}$ is the error of the expected LOR hits made in the forward projection. Note that Taylor's approximation is acceptable only if function $1/y$ can be well approximated by a line in $\tilde{y}_L \pm \Delta\tilde{y}_L$. The iteration is convergent if

$$\mathbf{T} = \mathbf{1} - \langle x_V^* \rangle \cdot \bar{\mathbf{B}}^T \cdot \langle \frac{y_L}{\tilde{y}_L^2} \rangle \cdot \mathbf{F}$$

is a *contraction* after certain number of iterations (note that $\mathbf{T}$ is not constant but depends on $\mathbf{x}^{(n)}$ via $\tilde{y}_L$). Even for convergent iteration, the limiting value will be different from $\mathbf{x}^*$ due to the errors of the forward and back projections:

$$\mathbf{\Delta x}^{(\infty)} = \mathbf{S} \cdot \left(\mathbf{\Delta \bar{B}}^T \cdot \frac{\mathbf{y}}{\tilde{\mathbf{y}}} - \mathbf{A}^T \cdot \langle \frac{y_L}{\tilde{y}_L} \rangle \cdot \frac{\mathbf{d\tilde{y}}}{\tilde{\mathbf{y}}} \right) \quad \text{where} \quad \mathbf{S} = \left(\mathbf{A}^T \cdot \langle \frac{y_L}{\tilde{y}_L^2} \rangle \cdot \mathbf{A} \right)^{-1}. \tag{2}$$

We can make several observations examining these formulae:

1. As measured hits y_L are Poisson distributed with expectations $\tilde{y}_L$, ratios $y_L/\tilde{y}_L$ have expected value 1 and variance $1/\tilde{y}_L$, thus $E[\mathbf{\Delta \bar{B}}^T \cdot \mathbf{y}/\tilde{\mathbf{y}}] = \mathbf{0}$ and even the variance caused by the back projector error diminishes when the measurement is high dose and thus the result is statistically well defined. Thus, for high dose measurement, the error made in forward projection is mainly responsible for the accuracy of the reconstruction, which adds the following error in each iteration step:

$$\langle x_V^* \rangle \cdot \bar{\mathbf{B}}^T \cdot \langle \frac{y_L}{\tilde{y}_L} \rangle \cdot \frac{\mathbf{\Delta \tilde{y}}}{\tilde{\mathbf{y}}} = \langle x_V^* \rangle \cdot \bar{\mathbf{B}}^T \cdot \langle \frac{y_L}{\tilde{y}_L} \rangle \cdot \frac{\mathbf{\Delta F} \cdot \mathbf{x}}{\tilde{\mathbf{y}}} \tag{3}$$

2. If the back projection accuracy is not so important, it is worth using a modified normalized SM $\bar{\mathbf{B}}$ to increase the contraction of $\mathbf{T}$ and thus speeding up the iteration.

3 ML-EM iteration using MC quadrature

In tomography the size of the SM is enormous, thus matrix elements cannot be pre-computed and stored, but must be re-computed each time with MC quadrature when a matrix element is needed. It means that forward projector $\mathbf{F}$ and back projector $\bar{\mathbf{B}}$ are random variables. We use unbiased MC estimates, i.e.

$$E[\mathbf{F}] = \mathbf{A}, \quad E[\bar{\mathbf{B}}] = \bar{\mathbf{A}}.$$

As these estimates are re-made in every iteration, we can choose whether the same random estimate is used in all iterations, the estimate is modified in each iteration, or even between the forward projection and back projection. Note that as we have to re-compute the matrix elements anyway, the computation costs of different options are the same, the algorithms differ only in whether or not the seed of the random number generator is reset.

The contribution to the error of a single iteration is defined by Eq. 3. Errors of different iteration steps accumulate. However, the accuracy can be improved if we use an MC estimation where the expectation value of this contribution is zero since it means that the error contributions of different iteration steps compensate each other and we may get a precise reconstruction even with inaccurate SM estimates. So, our goal is to guarantee that

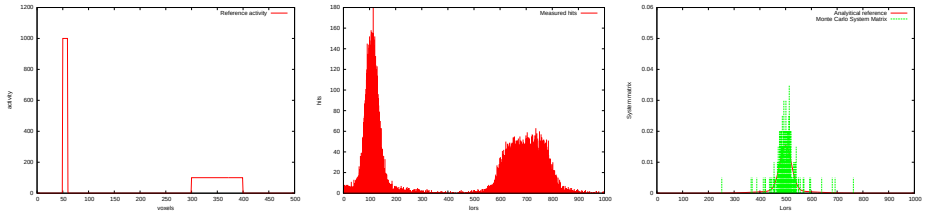
$$E\left[\langle x_V^* \rangle \cdot \bar{\mathbf{B}}^T \cdot \langle \frac{y_L}{\tilde{y}_L} \rangle \cdot \frac{\mathbf{\Delta F} \cdot \mathbf{x}}{\tilde{\mathbf{y}}} \right] = E\left[\langle x_V^* \rangle \cdot (\bar{\mathbf{A}}^T + \mathbf{\Delta \bar{B}}^T) \cdot \langle \frac{y_L}{\tilde{y}_L} \rangle \cdot \frac{\mathbf{\Delta F} \cdot \mathbf{x}}{\tilde{\mathbf{y}}} \right] = 0$$

which, taking into account that both the forward projector and the back projector are unbiased estimators, is held if

$$E\left[\mathbf{\Delta}\bar{\mathbf{B}}^T \cdot \langle \frac{y_L}{\tilde{y}_L} \rangle \cdot \mathbf{\Delta}\mathbf{F}\right] = 0.$$

Note that this is true if the forward projector is statistically independent from the back projector, but is false when they are correlated. This means that it is worth using independent random samples in each iteration and re-sampling even between forward projection and back projections.

To demonstrate this, we analyze a simple analytical problem, when an SM of dimensions $N_{LOR} = 1000$ and $N_{voxel} = 500$ is defined as the sum of two Gaussian density functions.



ref. activity measured values 1 column of the SM

Fig. 1. The measured function (left), the distribution of the hits in different LORs for the high dose case (middle), and the MC estimate of the SM (1 column is shown) obtained with 10^5 random samples.

The ground truth activity is another simple function of the left of Fig. 1. The measured values are obtained by sampling Poisson distributed random variables setting their means to the product of the SM and the reference activity. We examined a high dose and a low dose case, which differ in a factor of 10 of their activities. The middle of Fig. 1 shows the measurement of the high activity case. The error of the reconstruction is tested with random SM approximations, which are obtained by replacing the $5 \cdot 10^5$ analytical SM elements by unbiased MC estimates calculated with 10^4, 10^5, and 10^6 discrete samples in total, respectively.

In the first set of experiments we examine the L2 error of the reconstruction process of the *fixed case*, i.e. when the same SM approximation is used in all iteration steps (Fig. 2). These results indicate that working with the same MC estimate during an EM iteration is generally a bad idea. Reconstructing with a modified SM means that we altered the physical model, so the EM iteration converges to a different solution. *Matched sampling* takes the same samples in the forward and back projections of a single iteration but regenerates samples for each iteration. Matched sampling does not help, the error curves are quite similar to those of generated with fixed SM.

Independent sampling, where samples of forward projection are independent of the samples in back projection, has advantages and disadvantages as well. If the sample number is small, then the error curves are strongly fluctuating. The explanation is that matrix $\mathbf{T}$ is just probably a contraction, so the iteration have convergent and divergent stages. If the number of samples is higher, then the iteration gets similar to iterating with the analytic SM.

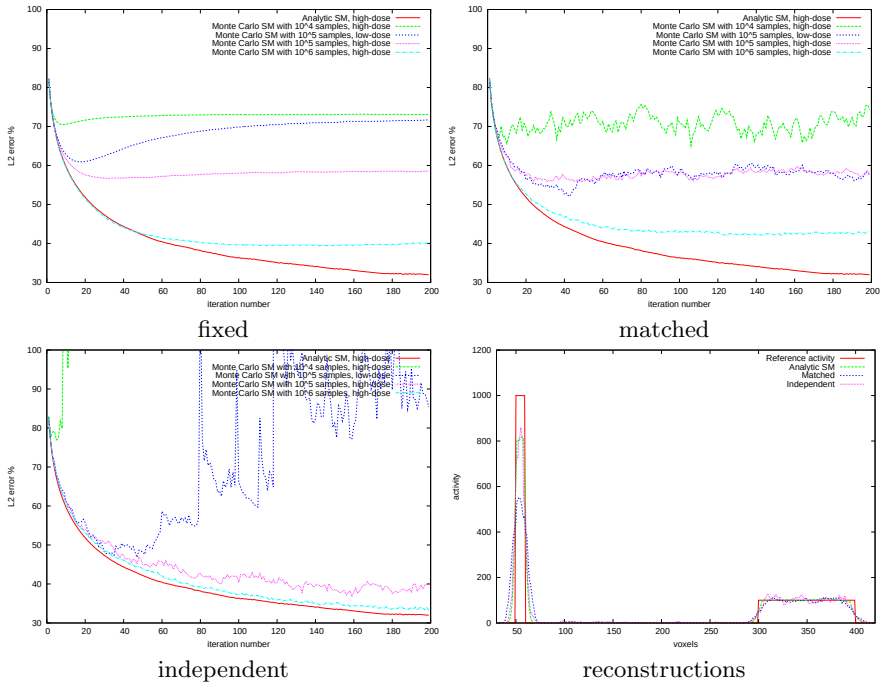


Fig. 2. L2 error curves of different sampling strategies and the reconstructed results. Fixed sampling takes the same samples in every projection. Matched sampling keeps the samples of forward even for back projection of the same iteration step. Independent sampling uses different random numbers in all projections.

4 Speeding up the convergence with simplified back projectors

We concluded that the reconstruction accuracy of high dose measurements is just slightly affected by the accuracy of the back projector. In a special case when $\mathbf{B} = \mathbf{A} \cdot \mathbf{P}$ where $\mathbf{P}$ is an invertible square matrix of N_{voxel}^2 elements, the fixed point is preserved, which can be seen if both sides of Eq. 1 are multiplied

with matrix $\mathbf{P}$. The convergence speed depends on the contraction of matrix $\mathbf{T}$, which is strong if

$$\langle x_V^* \rangle \cdot \bar{\mathbf{B}}^T \cdot \langle \frac{y_L}{\tilde{y}_L^2} \rangle \cdot \mathbf{A}$$

is close to the identity matrix. We need to find matrix $\mathbf{P}$ so that for every voxel V just the most significant $\mathbf{A}_{LV}$ elements are kept while others are replaced by zero during the multiplication with $\mathbf{P}$. As the SM represents a sequence of physical phenomena, this means ignoring voxel space blurring effects.

Using the example of the previous section, we examined the convergence of the reconstruction for different activity levels (recall that back projection accuracy becomes important only for low dose measurements).

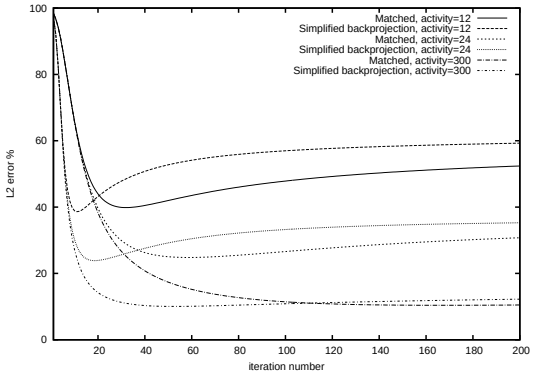


Fig. 3. Convergence in L2 for matched and simplified back projectors for different activities.

The results are shown by Fig. 3. Note that simplified and original back projectors converge to the same result, the approximation is more accurate when the measurement is of high dose. The initial convergence of the simplified back projector is much faster and it becomes poorer only when the iteration overfits the result and therefore the iteration is worth stopping anyway (such overfitting may be avoided with regularization).

5 Application in 3D Positron Emission Tomography

To test the presented method with a realistic 3D PET reconstruction, we took a LOR-centric, i.e. ray-based forward projection and a voxel-based back projection (Fig. 4). The forward projection samples are multiple rays or line segments connecting two uniformly distributed points on the two crystals' surfaces of the LOR. The line integral is evaluated between the two endpoints by sampling points being equal size but having a random starting offset. In back projection,

a discrete point is sampled in each voxel and the solid angles subtended from this point by the two crystals' surfaces of each LOR are randomly sampled by a line. The SM element can then be computed from the solid angle, and the total attenuation of the line between the two detectors.

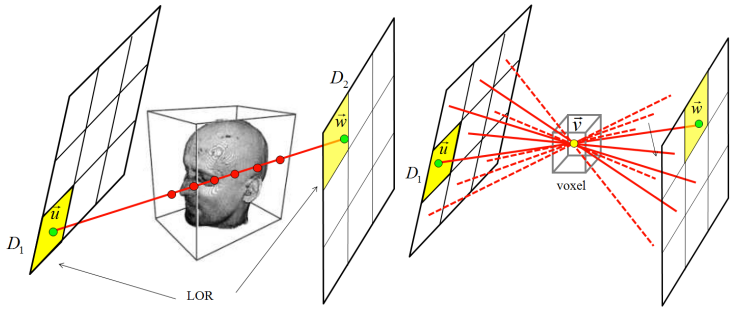


Fig. 4. Forward projection samples are line segments connecting uniformly distributed sample points on the crystals. Back projection samples are points in voxels and then points of detectors.

We modeled the Mediso's small animal *nanoScan PET/CT* [2], which consists of twelve detector modules of 81×39 crystals detectors of surface size 1.12×1.12 mm^2, thus the total number of LORs is 180 million when crystals of a module are connected by LORs to crystals of three opposite modules. We examined the *Micro Derenzo phantom* with rod diameters $1.0, 1.1, \ldots, 1.5$ mm in different segments. The Derenzo is virtually filled with 1.6 MBq activity and we simulated a 1000 sec measurement.

The error curves and slices of reconstructions when the random number generator is reset in each iteration and when independent samples are generated are shown by Fig. 5. We considered the cases when integrals are estimated with many and with fewer samples. Note that for low sampling density fixed iteration diverges, while independent sampling oscillates. For higher sampling density, both of them are stable and independent sampling has better results. Examining the reconstruction results we can observe that fixed sampling distorts the uniform activity distribution in rods, while independent sampling better preserves the ground truth.

6 Conclusion

This paper proposed the application of independent sampling and simplified back projector in inverse problems when elements are re-computed in each iteration step. The independent re-sampling has the advantage that it can gather more

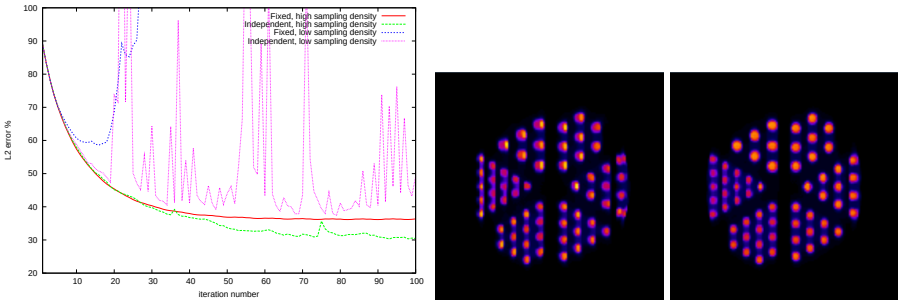


Fig. 5. L2 error curves and reconstructions of the Derenzo phantom.

information about the system, probably not in a single step but as iteration proceeds. This additional information helps increase the accuracy. Independent sampling in forward and back projectors has a drawback that solution oscillates if the sample density is low, so sample numbers should be carefully selected. We also shown that if back projector is properly simplified, then not only its computation can be speeded up, but also the iteration can be made faster.

Acknowledgement

This work has been supported by the OTKA K-104476 and by TÁMOP - 4.2.2.B-10/1–2010-0009. The GATE simulation of the Derenzo phantom has been executed by Gergely Patay.

References

1. I. Buvat and I. Castiglioni. Monte Carlo simulations in SPET and PET. *J Nucl Med*, 46(1):48–61, 2002.
2. nanoscan PET/CT. http://www.mediso.com/products.php?fid=2,11&pid=86.
3. J. Qi, R. M. Leahy, S. R. Cherry, A. Chatziioannou, and T. H. Farquhar. High-resolution 3D Bayesian image reconstruction using the microPET small-animal scanner. *Physics in Medicine and Biology*, 43(4):1001, 1998.
4. A. J. Reader and H. Zaidi. Advances in PET image reconstruction. *PET Clinics*, 2(2):173–190, 2007.
5. L. Shepp and Y. Vardi. Maximum likelihood reconstruction for emission tomography. *IEEE Trans. Med. Imaging*, 1:113–122, 1982.
6. L. Szirmay-Kalos, M. Magdics, B. Tóth, and Bükki T. Averaging and Metropolis iterations for positron emission tomography. *IEEE Trans Med Imaging*, 32(3):589–600, 2013.