A Focus-of-Attention EM-ML Algorithm for PET Reconstruction

Jens Gregor and Dean A. Huff
Department of Computer Science
University of Tennessee
107 Ayres Hall
Knoxville, TN 37996–1301 USA

May 28, 1996

Abstract

The EM–ML algorithm belongs to a family of algorithms that compute PET (positron emission tomography) reconstructions by iteratively solving a large linear system of equations. We describe a preprocessing scheme for focusing the attention, and thus the computational resources, on a subset of the equations and unknowns in order to reduce both the time and space requirements of such algorithms. The approach is completely data-driven and uses no prior anatomic knowledge. Experimental results are given for a CM–5 parallel computer implementation of the EM-ML algorithm using a simulated phantom as well as real data obtained from an ECAT 921 PET scanner.

Keywords: positron emission tomography, image reconstruction, expectation-maximization, parallel computing.

I. Introduction

Positron emission tomography (PET) is a method for non-invasively studying the physiology of the human body. A patient is injected with a radio-isotope which collects in the part of the body to be studied, then emits positrons as it decays. Each positron annihilates with an electron causing two photons to be emitted in opposite directions. From measurements of these photon pairs obtained externally by a ring of detectors surrounding the patient, a discretized emission map, i.e., an image, that reflects the internal isotope concentration must be reconstructed by computer.

The *de facto* standard reconstruction algorithm is Fourier based filtered backprojection. Many alternatives have been studied, e.g., the ART [1], EM-ML [2, 3, 4], Landweber-LS [5, 6], and CG-LS [7, 8] algorithms which all iteratively solve the linear system of equations

$$P'\boldsymbol{\lambda} \;=\; \boldsymbol{n}^* \tag{1}$$

where $\boldsymbol{n}^*$ is a D-dimensional vector representing the sinogram, i.e., the number of coincident photons recorded by the detectors, $\boldsymbol{\lambda}$ is a B-dimensional vector representing the unknown image, i.e., the unknown number of photon pairs actually emitted, and P is a $B \times D$ matrix that describes the connection between the internal emission activity and the observed data. We henceforth concentrate on the EM-ML algorithm which finds the $\boldsymbol{\lambda}$ that maximizes the likelihood $P(\boldsymbol{n}^*|\boldsymbol{\lambda})$ based on a Poisson model of the emission process.

Two main concerns with the EM-ML algorithm are: (i) a visually dissatisfying checkerboarding tends to appear in the reconstructions after a number of iterations, and (ii) the algorithm is computationally expensive in that it requires many time consuming iterations as well as significant amounts of memory. The checkerboarding has been addressed in numerous ways, e.g., regularization by the method of kernel of sieves [9, 10, 11], smoothing [12], and MAP estimation which finds the $\boldsymbol{\lambda}$ that maximizes the posterior probability $P(\boldsymbol{\lambda}|\boldsymbol{n}^*) \propto P(\boldsymbol{n}^*|\boldsymbol{\lambda})\, P(\boldsymbol{\lambda})$ using Poisson and Gaussian priors [13, 14, 15, 16, 17], penalized likelihoods [18], Good's measure of roughness [19], and Gibbs priors that induce Markov random fields [20, 21, 22, 23, 24]. With respect to the computational cost, various attempts have been made to reduce the number of iterations by accelerating convergence, e.g., overrelaxation [25], line search [7, 8], multigrids [26, 5], incremental partitioning [27], and preconditioning [28]. In order to achieve acceptable reconstruction times, the use of parallel computing has been studied and many implementations have been reported, e.g., for the iPSC/2 hypercube and the BBN Butterfly GP–1000 shared memory computer [29], the Cray–1 vector computer [30], the Alliant FX/8 shared memory multiprocessor [31], the MasPar 4096 and AMT-DAP 4096 SIMD computers [19, 32, 33], the Thinking

Machines CM–5 MIMD computer and a network of loosely coupled SUN workstations [34].

We present a focus-of-attention preprocessing scheme for reducing the time and space requirements of the EM-ML algorithm. Simply stated, our approach is to first determine which of the equations and the unknowns are relevant to the reconstruction and then concentrate the computational resources thereon. We note that the approach does not prohibit regularization by means of priors and is generally applicable to the family of iterative reconstruction algorithms mentioned above.

The paper is organized as follows. In Section II we introduce our notation, review the EM-ML algorithm, and outline how we compute the P-matrix. We describe the focus-of-attention preprocessing scheme in Section III with emphasis on issues pertinent to real data processing. Finally, in Section IV, we give experimental results obtained from runs on a CM–5 parallel computer using a simulated phantom and data obtained from an ECAT 921 PET scanner.

II. Background

A. Notation

We consider a ring with M detectors where M is even. Any pair of opposing detectors is said to form a tube. Using D to denote the number of tubes, we have that $D = M(M-1)/2$. All tubes having the same slope constitute a projection. There is a total of M projections. With reference to Fig. 1, we shall distinguish between full-angle projections, for which all tubes consist of detector pairs whose indices are odd-even or even-odd, and half-angle projections, for which the detector pair indices are either both odd or both even. Regarding the image, we assume that it is square and has $B \leq D$ pixels, but we make no assumptions about its physical dimension or its precise location as long as it remains fully contained within the ring of detectors whose radius we refer to as R.

We denote the number of photon pairs recorded by tube d by n_d^*, the expected number of photon pairs emitted from pixel b by λ_b, and the expected number of photon pairs recorded by tube d by $\lambda_d^* = \sum_b \lambda_b p_{bd}$ where p_{bd} is the conditional probability that a photon pair is recorded by tube d given that it originated from pixel b; we assume normalization such that $\sum_d p_{bd} = 1$. When referring to these entities collectively, we use vectors $\boldsymbol{n}^* = [n_d^*]$, $\boldsymbol{\lambda} = [\lambda_b]$, $\boldsymbol{\lambda}^* = [\lambda_d^*]$, and matrix $P = [p_{bd}]$.

B. EM-ML Algorithm

Based on the observation that the photon emissions can be modeled as independent Poisson processes, the reconstruction problem can be formulated as finding the $\boldsymbol{\lambda}$ that maximizes the log-likelihood

$$\log P(\boldsymbol{n}^*|\boldsymbol{\lambda}) \quad = \quad f(\boldsymbol{n}^*, \boldsymbol{\lambda}^*) - D(\boldsymbol{n}^*, \boldsymbol{\lambda}^*) \tag{2}$$

where function $f(\boldsymbol{n}^*, \boldsymbol{\lambda}^*)$ represents terms that are constant for a given $\boldsymbol{n}^*$ and nonnegative Kullback function

$$D(\boldsymbol{n}^*, \boldsymbol{\lambda}^*) \quad = \quad \sum_d n_d^* \log \frac{n_d^*}{\lambda_d^*} \tag{3}$$

represents terms that are not. The EM-ML (expectation-maximization, maximum-likelihood) algorithm finds the $\boldsymbol{\lambda}$ that solves $D(\boldsymbol{n}^*, \boldsymbol{\lambda}^*) = 0$, and thus also (1), using the multiplicative iteration scheme [2, 3, 4]

$$\forall b: \quad \lambda_b^{k+1} \quad = \quad \lambda_b^k \sum_d \frac{n_d^*}{\lambda_d^{*k}} p_{bd}. \tag{4}$$

Typically, $\boldsymbol{\lambda}^0$ is chosen to be a uniform, positive vector. Important features of the algorithm are that $\boldsymbol{\lambda}^k$ remains nonnegative for all k and $\sum_b \lambda_b^k = \sum_d n_d^*$ for $k > 0$. The algorithm ensures that $D(\boldsymbol{n}^*, \boldsymbol{\lambda}^{*k}) \to 0$ [2, 3, 4, 35]; also, $D(\boldsymbol{n}^*, \boldsymbol{\lambda}^{*k})$ is $o(k^{-1})$ which means that the convergence to 0 takes place faster than $k \to \infty$ [35].

C. Computing P

Several different methods have been suggested for computing the elements of the P-matrix, e.g., angle-of-view [2, 29], line intersection [14, 17], and area intersection based on circular pixel approximations [36, 34]. We compute p_{bd} as the exact area intersection of tube d with pixel b relative to the sum-total intersection of all tubes with that pixel; thus, by design, $\sum_d p_{bd} = 1$.

Due to the circular arrangement of the detectors, the width of a tube depends on its location; a tube located near the center of the ring of detectors is wider than one located near the periphery (cf. Fig. 1). To take this non-linear nature of the sampling into consideration, we model each tube by the trapezoid obtained by connecting the corners of the two detectors in question. This makes geometric arc correction by means of count redistribution within each projection obsolete.

The P-matrix is quite sparse which makes it advantageous to store only its nonzero elements. When the image is centered in the ring of detectors, then certain symmetry considerations can be applied to further reduce the number of elements which have to be stored [7, 29]. But since we allow the image to be located anywhere within the ring of detectors these symmetry considerations are not exploited here.

III. Focus-of-Attention

The computational requirements of the EM-ML algorithm, which are quite extensive, can be reduced by realizing that certain parts of the linear system of equations described by (1) do not contribute to the reconstruction. Consider, for example, the equation

$$\sum_b \lambda_b p_{bd} \quad = \quad n_d^*. \tag{5}$$

All numbers being nonnegative, $n_d^* = 0$ implies that $\lambda_b = 0$ for all b for which $p_{bd} > 0$. By identifying and discarding such equations and unknowns we can obtain a smaller system of equations and thus reduce the amount of computation needed. Knowing whether pixel b is intersected by tube d is enough to say whether $p_{bd} > 0$. Consequently, we can reduce the amount of storage needed by computing and storing only the actual values of the p_{bd} coefficients of the smaller system of equations.

These observations have lead to the following focus-of-attention preprocessing based EM-ML algorithm:

1. Compute an $\boldsymbol{n}^*$-mask that indicates which equations, i.e., tubes, to focus on.
2. Compute a $\boldsymbol{\lambda}$-mask that indicates which unknowns, i.e., pixels, to focus on.
3. Compute the P-matrix for the tubes in $\boldsymbol{n}^*$-mask and the pixels in $\boldsymbol{\lambda}$-mask.
4. When necessary, compensate for edge packing.
5. Apply the EM-ML algorithm to solve the reduced system of equations.

The premise of the computations described below is that all photon emissions originate from a single, connected subset of pixels [1]. Under ideal circumstances, this implies that each projection contains a single, connected sequence of tubes whose counts are non-zero. This allows us to compute $\boldsymbol{n}^*$-mask (step 1) by locating the corresponding boundary tubes and, in turn, set $\boldsymbol{\lambda}$-mask (step 2) to be the convex hull obtained by intersecting

[1] In applications where this premise does not hold, our algorithm will eliminate only a subset of the non-contributing equations and unknowns and the system of equations produced will therefore not be a small as it could be.

the halfplanes defined thereby. However, when processing real data, which typically is very noisy, the detection of the boundary tubes becomes a difficult task and, unless an optimization scheme is applied to the $\boldsymbol{n}^*$-mask to ensure that the boundary tubes are, or become, consistent with one another, it is very likely that a non-convex hull will be obtained for $\boldsymbol{\lambda}$-mask. Figure 2 illustrates the difference between consistent and inconsistent boundary tubes in terms of the shape of the hull.

All the emission counts attributed to tubes located within $\boldsymbol{n}^*$-mask will be distributed across pixels located within $\boldsymbol{\lambda}$-mask. For ideal data this is as it should be, but for real data it means that the counts which reflect only noise will be distributed over fewer pixels. Since the elements of the P-matrix are computed only for the subset of tubes and pixels respectively identified by $\boldsymbol{n}^*$-mask and $\boldsymbol{\lambda}$-mask (step 3), then the pixels located on the edge of the convex hull may be estimated to have high emission rates although in reality they should be low. When necessary, we compensate for this edge packing by reducing some of the tube counts (step 4) before applying the EM-ML algorithm to compute the reconstruction (step 5).

Since steps 3 and 5 were addressed in the previous section, we now concentrate on steps 1, 2, and 4.

A. Computing n^*-mask

Suppose we want to find the leftmost (resp. rightmost) boundary tube for a given projection. Scanning from left-to-right (resp. right-to-left) until $n_d^* > 0$ is sufficient for ideal data. But when the data is noisy, this approach may cause the search to stop prematurely due to spurious occurrences of tubes whose non-zero count only reflects noise. We therefore use moving average based thresholding in combination with the following optimization procedure.

Define $\boldsymbol{h} = [h_m]$ to be the $2M$-dimensional vector whose elements, also called support values, denote the distance from the center of each projection to the center of its boundary tubes. In particular, let h_m (resp. h_{M+m}) denote the distance from the center of the mth projection to the center of its leftmost (resp. rightmost) boundary tube; see Fig. 3. Then the boundary tubes are consistent with one another, i.e., $\boldsymbol{h}$ is a support vector, if and only if [37]

$$C\boldsymbol{h} \;\le\; 0 \tag{6}$$

where C is a $2M \times 2M$ Toeplitz matrix of the form

$$C = \begin{bmatrix} 1 & -k & 0 & \cdots & 0 & -k \\ -k & 1 & -k & 0 & \cdots & 0 \\ 0 & -k & 1 & & & \vdots \\ \vdots & 0 & & \ddots & & 0 \\ 0 & \vdots & & & 1 & -k \\ -k & 0 & \cdots & 0 & -k & 1 \end{bmatrix} \tag{7}$$

and $k = 1/(2\cos\pi/M)$.

Let $\boldsymbol{h}^\circ$ denote the initial estimate produced by the moving average based thresholding, and let $\boldsymbol{h}^*$ denote a support vector obtained on basis thereof. Then $\boldsymbol{h}^*$ is the support vector closest to $\boldsymbol{h}^\circ$ (in the Euclidean sense) if it solves the quadratic programming problem [37]

$$\min_{\boldsymbol{h}} \frac{1}{2}\boldsymbol{h}'\boldsymbol{h} - \boldsymbol{h}'\boldsymbol{h}^0 \quad s.t. \ \ C\boldsymbol{h} \leq 0. \tag{8}$$

Because of its simpler constraint set, it is more convenient to consider the dual problem [38]

$$\min_{\boldsymbol{u}} \frac{1}{2}\boldsymbol{u}'\Omega\boldsymbol{u} + \boldsymbol{r}'\boldsymbol{u} \quad s.t. \ \ \boldsymbol{u} \geq 0 \tag{9}$$

where $\Omega = CC'$ and $\boldsymbol{r} = -C\boldsymbol{h}^0$. The connection between the primary (original) and dual problems is that the former is solved by

$$\boldsymbol{h}^* = \boldsymbol{h}^0 - C'\boldsymbol{u}^* \tag{10}$$

if $\boldsymbol{u}^*$ solves the latter.

We note that the dual problem may be solved using a Gauss-Seidel iteration which, when the mth element is updated, has the form [38]

$$u_m = \max\left\{ 0, \ \ u_m - \frac{1}{\omega_{mm}} \left(r_m + \sum_{j=1}^{2M} \omega_{mj} u_j \right) \right\}. \tag{11}$$

By furthermore exploiting the block structure of matrix Ω and, in particular, the Toeplitz non-zero structure of matrix C, we obtain the simpler iteration scheme

$$u_m = \max\left\{ 0, \ \ -\frac{1}{1+2k^2} \left(r_m - 2k(u_{m-1} + u_{m+1}) + k^2(u_{m-2} + u_{m+2}) \right) \right\} \tag{12}$$

where the indices are computed modulo $2M$.

In practice, the discrete nature of the optimization problem, i.e., the facts that only a finite number of support values correspond to actual tube locations and that each full-angle projection is followed by a half-angle projection and vice versa, means that a computed $\boldsymbol{h}^*$ most likely has no physical interpretation. Rather than proceeding with the above Gauss-Seidel iteration until we find the $\boldsymbol{h}^*$ closest to $\boldsymbol{h}^\circ$, we therefore stop when

$$\max \left| C\, f_Q(\boldsymbol{h}^*) \right| \quad \leq \quad \varepsilon \tag{13}$$

where function f_Q quantizes $\boldsymbol{h}^*$ into its physical equivalent and

$$\varepsilon \quad = \quad R \sin(2\pi/M)/2 \tag{14}$$

is half the width of the widest tube which is the largest discrepancy that has to be compensated for. In other words, we use quadratic programming to obtain a reasonably close support vector that has a physical interpretation but do not carry the optimization through in order to find the closest one.

Figure 4 provides an example of the process. The initial estimate $\boldsymbol{h}^\circ$ shown in Fig. 4(a) contains a number of inconsistencies as indicated by the peaks in Fig. 4(b). After a number of iterations, typically very few, we obtain the $\boldsymbol{h}^*$ shown in Fig. 4(c) for which the inconsistencies have been resolved. The remaining oscillation between positive and negative values seen in Fig. 4(d) is due to the interlacing of half-angle and full-angle projections.

B. λ-mask Computation

As mentioned above, we set $\boldsymbol{\lambda}$-mask to be the convex hull obtained by intersecting the halfplanes defined by the boundary tubes of $\boldsymbol{n}^*$-mask. Thus, $\boldsymbol{\lambda}$-mask consists of the set of all pixels which have any area in common with the hull including those pixels for which the intersection is only partial. To help prevent cutting off part of the emitter region, we move all the halfplane cross-over points out by a distance of ε (cf. (14)).

C. Edge Packing

The P-matrix is computed for the subset of tubes and pixels given by $\boldsymbol{n}^*$-mask and $\boldsymbol{\lambda}$-mask, respectively. Consequently, a tube will have its emission count distributed across a limited number of pixels. For ideal data this is as it should be, but for real data it may result in the activity of some of the edge pixels becoming artificially high; we refer to this phenomenon as edge packing.

To alleviate the problem of edge packing, we reduce the emission activity of the boundary tubes as follows. Let the area covered by boundary tube d of $\boldsymbol{\lambda}$-mask and the original image respectively be denoted by A_d^{λ} and A_d^{Σ}; clearly, $A_d^{\lambda} \leq A_d^{\Sigma}$. Then, under the assumption that the noise recorded by a tube should be distributed evenly across all the pixels it intersects, we simply multiply n_d^* by $A_d^{\lambda}/A_d^{\Sigma}$.

The emission activity of the interior tubes also reflect noise and should have their counts reduced accordingly. But it is unclear how to do so in a straightforward manner, and we leave it as an open problem for further study.

IV. Experimental Results

We have implemented both the original and the focus-of-attention EM-ML algorithm in C on a 32-node Thinking Machines CM–5 parallel computer. The nodes, which communicate via high-speed internal networks, are all equipped with a 32 MHz RISC processor and 32 Mbytes of RAM. The algorithm is parallelized by letting each node maintain its own copy of $\boldsymbol{\lambda}^k$ while distributing $\boldsymbol{n}^*$ and the corresponding columns of the P-matrix across the nodes. After every iteration, the nodes communicate their subresults, i.e., partial multiplier sums, to each other using built-in combiner operators [39].

The experimental work is based on the CTI/Siemens ECAT 921 PET scanner which generates 384×80 sinograms and 128×128 images for arbitrary offsets and zoom factors; the P-matrix is designed accordingly. We use both simulated phantom and real scan data. The simulated phantom data consists of a single, noise-free sinogram obtained by forward projecting a version of the Shepp-Vardi phantom [2] through the P-matrix. The real scan data is obtained from an ECAT 921 PET scanner located at the University of Tennessee Medical Center, Knoxville (UTMCK). Three real scans are used: an FDG head scan, a 13-N ammonia chest scan, and a (low count) C-11 ACBC abdomen scan. The patient scans follow routine clinical protocols and data processing approved by the UTMCK Institutional Review Board. The real sinograms are normalized for detector variability, corrected for attenuation using transmission scans obtained prior to radiopharmaceutical injection, and corrected for estimated randoms. Scatter correction is not applied.

Below, we first provide a quantitative assessment of the original and the focus-of-attention EM-ML algorithms using the simulated phantom. Then we carry out a qualitative assessment for the real scanner data followed by an analysis of the computational savings. Traditionally, sinograms and images alike are displayed white-on-black, but in order to point out certain details we use reverse video, i.e., black-on-white.

A. Quantitative Assessment

There are many ways to quantify the quality of a reconstructed image. The Kullback function $D(\boldsymbol{n}^*, \boldsymbol{\lambda}^*)$, for example, can be used as a global error measure since the closer it is to 0, the more likely is the reconstruction. Local characteristics are more difficult to quantify, but region-of-interest (ROI) statistics (such as mean pixel value, standard deviation, and regional bias which is the relative difference between the observed and the true ROI mean pixel values) may be appropriate [36]. Figure 5 shows the simulated phantom used for this quantitative assessment together with an outline of seven uniform ROIs that cover 6287, 521, 425, 37, 24, 24, and 140 pixels, respectively.

Figure 6 plots $D(\boldsymbol{n}^*, \boldsymbol{\lambda}^{*k})$ as a function of iteration index k for reconstruction of the phantom using both the original and the focus-of-attention EM-ML algorithms; since the sinogram is noise-free by design, we create the $\boldsymbol{n}^*$-mask for the latter using simple thresholding instead of moving average and apply neither the quadratic programming optimization nor the edge packing compensation scheme. We see that both algorithms converge nicely and that the focus-of-attention version produces a more likely image at every iteration. When comparing the two Kullback functions, one must be aware of the logarithmic ordinate axis. For the first 25-30 iterations the relative difference is about an order of magnitude, but it settles to a factor of 2-3 for the later iterations. Many fewer iterations are thus needed for the focus-of-attention EM-ML algorithm to attain the same likelihood as the original EM-ML algorithm. The reason is, of course, that the latter requires a number of iterations to overcome the initial assignment of the same activity to background and foreground pixels.

Figure 7 plots the observed mean pixel value plus-minus one standard deviation for each iteration between 8 and 1024 which is a power of 2 for the two EM-ML algorithms. The true mean pixel value of each ROI is indicated by a dotted line. The plot for ROI 5 is almost identical to the plot for ROI 4 and is therefore not shown. Given enough iterations, both the original EM-ML algorithm and the focus-of-attention version produce observed mean values that are very close to the true values for all ROIs. During the early iterations, however, the regional emission activities are somewhat off the mark, especially for the smaller ROIs; note that the mean values for ROIs 1–5 computed by the focus-of-attention EM-ML algorithm are as close to the true values as those computed by the original algorithm, but significantly closer for ROIs 0 and 6. The standard deviations are practically indistinguishable for the two algorithms for all ROIs and most iterations, and the regional biases behave similarly to the observed mean values.

B. Qualitative Assessment

Figure 8 shows the three ECAT 921 sinograms with the $\boldsymbol{n}^*$-masks overlayed. The moving average based segmentation uses a window width of 5 and a conservative threshold corresponding to 1% of the peak value in $\boldsymbol{n}^*$. The number of Gauss-Seidel iterations required to make the $\boldsymbol{n}^*$-mask consistent is 6, 4, and 4, respectively, for the head, chest, and abdomen sinogram. Narrowing the moving average window would produce more irregular initial $\boldsymbol{n}^*$-masks and therefore call for more Gauss-Seidel iterations. Similarly, increasing the threshold would produce less irregular initial $\boldsymbol{n}^*$-masks and consequently require fewer Gauss-Seidel iterations. The window width and the threshold used here are a compromise we have found to work well.

Figures 9–12 show the reconstructed images after 32 iterations of the two EM-ML algorithms; the z=1 and z=2 annotations refer to an image being unzoomed and zoomed a factor of two, respectively. The results of applying the original algorithm and the focus-of-attention version are virtually indistinguishable for the head reconstructions (Figs. 9, 10) and the chest reconstruction (Fig. 11). For these images, edge packing is not a serious problem although the transition from foreground to background is somewhat smoother when compensating for it; this is especially true for the zoomed head reconstruction (Fig. 10). With respect to the abdomen reconstruction (Fig. 12), the focus-of-attention preprocessing produces a few relatively dark pixels in the left arm region. Rather than being a result of edge packing which refers to erroneous concentration of noise activity, these pixels are a result of imperfect thresholding. The abdomen sinogram (cf. Fig. 8) shows very little activity for the tubes that specifically intersect the left arm region which, in turn, results in $\boldsymbol{\lambda}$-mask cutting off a portion thereof. We see that the edge packing compensation scheme lessens the impact somewhat but, as should be expected, fails to completely eliminate the problem. The solution would be to improve on the thresholding and/or incorporate more knowledge into the quadratic programming optimization.

C. Computational Aspects

The computational benefits of applying focus-of-attention can be seen in a comparison of array sizes relating to memory consumption. Table 1 therefore lists the number of tubes, pixels, and non-zero elements in the P-matrix for the reconstruction of the real scanner data. The original listings for $\boldsymbol{n}^*$ and $\boldsymbol{\lambda}$-mask refer to the $384 \times 80 = 30,720$ tube recordings in each sinogram and the $128 \times 128 = 16,328$ pixels in each image, respectively. When applying focus-of-attention, these numbers drop significantly. In the best case (head, z=1), for example, about 60% of the tubes and 90% of the pixels are not needed for the reconstruction. Even in the worst case (abdomen), 20% of the tubes and

70% of the pixels can be ignored. When it comes to the number of non-zero elements in the P-matrix, which is the governing factor with respect to memory consumption, we see reductions on the order of 90% in the best case (head, z=1) and 65% in the worst case (head, z=2; abdomen, z=1).

Table 2 lists the computational benefits translated into elapsed CPU time (in seconds). Notice that the time it takes to compute the $\boldsymbol{n}^*$ and $\boldsymbol{\lambda}$-masks is negligible. The fact that the P-matrix computes in time proportional to the size of $\boldsymbol{n}^*$-mask and is less affected by the size of the $\boldsymbol{\lambda}$-mask is due to the implementation which is tube rather than pixel driven. The compensation for edge packing is not listed but takes only about 0.15 seconds. The overall time spent on initialization is thus reduced in all cases. But perhaps more importantly, the time required for a single iteration of the EM-ML algorithm is reduced by 80% in the best case (head, z=1) and 60-70% otherwise (head, z=2; chest and abdomen; z=1).

V. Conclusion

We have a presented a preprocessing scheme for reducing the computational requirements of the EM-ML algorithm for PET reconstruction. The approach is simple to implement, very general, and could be applied to other iterative reconstruction algorithms. The experimental work indicates that quite significant savings can be obtained with respect to both time and space without compromising the quality of the reconstructed images.

Acknowledgement

We thank Dr. Gary Smith, Director of Nuclear Medicine at the University of Tennessee Medical Center, Knoxville, for providing the ECAT 921 PET scanner data.

References

[1] G. T. Herman, A. Lent, and S. W. Rowland, "ART: Mathematics and application," *J Theor. Biol.*, vol. 42, pp. 1–32, 1973.

[2] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. Med. Imaging*, vol. 1, pp. 113–122, 1982.

[3] Y. Vardi, L. A. Shepp, and L. Kaufman, "A statistical model of positron emission tomography," *J. Am. Stat. Assoc.*, vol. 80, pp. 8–20, 1985.

[4] K. Lange and R. Carson, "EM reconstruction algorithms for emission and transmission tomography," *J. Comp. Asst. Tomog.*, vol. 8, pp. 306–316, 1984.

[5] T.-S. Pan and A. E. Yagle, "Numerical study of multigrid implementations of some iterative image reconstruction algorithms," *IEEE Trans. Med. Imaging*, vol. 10, pp. 572–588, 1991.

[6] T.-S. Pan and A. E. Yagle, "Acceleration of Landweber-type algorithms by suppression of projection on the maximum singular vector," *IEEE Trans. Med. Imaging*, vol. 11, pp. 479–487, 1992.

[7] L. Kaufman, "Implementing and accelerating the EM algorithm for positron emission tomography," *IEEE Trans. Med. Imaging*, vol. 6, pp. 37–50, 1987.

[8] L. Kaufman, "Maximum likelihood, least squares, and penalized least squares for PET," *IEEE Trans. Med. Imaging*, vol. 12, pp. 200–214, 1993.

[9] D. L. Snyder and M. I. Miller, "The use of sieves to stabilize images produced with the EM algorithm for emission tomography," *IEEE Trans. Nucl. Sci.*, vol. 32, pp. 3864–3870, 1985.

[10] M. I. Miller, D. L. Snyder, and S. M. Moore, "An evaluation of the use of sieves for producing estimates of radioactivity distributions with the EM algorithm for PET," *IEEE Trans. Nucl. Sci.*, vol. 33, pp. 492–495, 1986.

[11] T. R. Miller and J. W. Wallis, "Clinically important characteristics of maximum-likelihood reconstruction," *J. Nucl. Med.*, vol. 33, pp. 1678–1684, 1992.

[12] B. W. Silverman, M. C. Jones, J. D. Wilson, and D. W. Nychka, "A smoothed EM approach to indirect estimation problems with particular reference to stereology and emission tomography," *J. Roy. Stat. Soc.*, vol. 52, pp. 271–324, 1990.

[13] H. Hart and Z. Liang, "Bayesian image processing in two dimensions," *IEEE Trans. Med. Imaging*, vol. 6, pp. 201–208, 1987.

[14] E. Levitan and G. T. Herman, "A maximum a posteriori probability expectation maximization algorithm for image reconstruction in emission tomography," *IEEE Trans. Med. Imaging*, vol. 6, pp. 185–192, 1987.

[15] Z. Liang and H. Hart, "Bayesian reconstruction in emission computerized tomography," *IEEE Trans. Nucl. Sci.*, vol. 35, pp. 788–792, 1988.

[16] Z. Liang, R. Jaszczak, C. Floyd, and K. Greer, "A spatial interaction model for statistical image processing," in *Information Processing in Medical Imaging, IPMI89*, pp. 29–43, Wiley-Liss, New York, 1991.

[17] G. T. Herman, A. R. De Pierro, and N. Gai, "On methods for maximum a posteriori image reconstruction with a normal prior," *J. Visual Comm. and Image Rep.*, vol. 3, pp. 316–324, 1992.

[18] P. J. Green, "On the use of the EM algorithm for penalized likelihood estimation," *J. Roy. Stat. Soc.*, vol. 52, pp. 443–452, 1990.

[19] M. I. Miller and B. Roysam, "Bayesian image reconstruction for emission tomography incorporating Good's roughness prior on massively parallel processors," *Proc. Natl. Acad. Sci.*, vol. 88, pp. 3223–3227, 1991.

[20] T. Hebert and R. Leahy, "A generalized EM algorithm for 3–D Bayesian reconstruction from Poisson data using Gibbs priors," *IEEE Trans. Med. Imaging*, vol. 8, pp. 194–202, 1989.

[21] P. J. Green, "Bayesian reconstructions from emission tomography data using a modified EM algorithm," *IEEE Trans. Med. Imaging*, vol. 9, pp. 84–93, 1990.

[22] V. E. Johnson, W. H. Wong, X. Hu, and C.-T. Chen, "Bayesian restoration of PET images using Gibbs priors," in *Information Processing in Medical Imaging, IPMI89*, pp. 15–28, Wiley-Liss, New York, 1991.

[23] D. S. Lalush and B. M. W. Tsui, "A generalized Gibbs prior for maximum a posteriori reconstruction in SPECT," *Phys. Med. Biol.*, vol. 38, pp. 729–741, 1993.

[24] S. Geman, K. M. Manbeck, and D. E. McClure, "A comprehensive statistical model for single-photon emission tomography," in *Markov Random Fields: Theory and Application* (R. Chellappa and A. Jain, eds.), pp. 93–130, San Diego, CA: Academic Press, 1993.

[25] R. M. Lewitt and G. Muehllehner, "Accelerated iterative reconstruction for positron emission tomography based on the EM algorithm for maximum likelihood estimation," *IEEE Trans. Med. Imaging*, vol. 5, pp. 16–22, 1986.

[26] M. V. Ranganath, A. P. Dhawan, and N. Mullani, "A multigrid expectation maximization reconstruction algorithm for positron emisson tomography," *IEEE Trans. Med. Imaging*, vol. 7, pp. 273–278, 1988.

[27] T. Hebert, R. Leahy, and M. Singh, "Three-dimensional maximum likelihood reconstruction for an electronically collimated single-photon-emission imaging system," *J. Opt. Soc. Am.*, vol. 7, pp. 1305–1313, 1990.

[28] N. H. Clinthorne, T.-S. Pan, P.-C. Chiao, W. L. Rogers, and J. A. Stamos, "Preconditioning methods for improved convergence rates in iterative reconstructions," *IEEE Trans. Med. Imaging*, vol. 12, pp. 78–83, 1993.

[29] C. M. Chen, C. Y. Lee, and Z. H. Cho, "Parallelization of the EM algorithm for 3–D PET image reconstruction," *IEEE Trans. Med. Imaging*, vol. 10, pp. 513–522, 1991.

[30] L. Kaufman, "Solving emission tomography problems on vector machines," *Ann. Op. Res.*, vol. 22, pp. 325–353, 1990.

[31] G. T. Herman, D. Odhner, K. D. Toennies, and S. A. Zenios, "A paralleized algorithm for image reconstruction from noisy projections," in *Workshop on Large Scale Optimization* (T. Coleman and Y. Li, eds.), Philadelphia: SIAM, 1990.

[32] A. W. McCarthy and M. I. Miller, "Maximum likelihood SPECT in clinical computation times using mesh-connected parallel computers," *IEEE Trans. Med. Imaging*, vol. 10, pp. 426–436, 1991.

[33] C. S. Butler and M. I. Miller, "Maximum a posteriori estimation for SPECT using regularization techniques on massively parallel computers," *IEEE Trans. Med. Imaging*, vol. 12, pp. 84–89, 1993.

[34] S. P. Olesen, J. Gregor, M. G. Thomason, and G. Smith, "EM-ML PET reconstruction on multiple processors with reduced communications," to appear in *International Journal of Imaging Systems & Technology.*

[35] J. Gregor, S. P. Olesen, and M. G. Thomason, "On convergence of the EM-ML algorithm for PET reconstruction," submitted.

[36] J. Llacer, E. Veklerov, K. J. Coakley, E. J. Hoffman, and J. Nunez, "Statistical analysis of maximum likelihood estimator images of human brain FDG PET studies," *IEEE Trans. Med. Imaging*, vol. 12, pp. 215–213, 1993.

[37] J. L. Prince and A. S. Willsky, "Reconstructing convex sets from support line measurement," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 12, pp. 377–389, 1990.

[38] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods.* Englewood Cliffs, NJ: Prentice-Hall, 1989.

[39] Thinking Machines Corporation, Cambridge, MA, *CMMD Reference Manual*, May 1993.

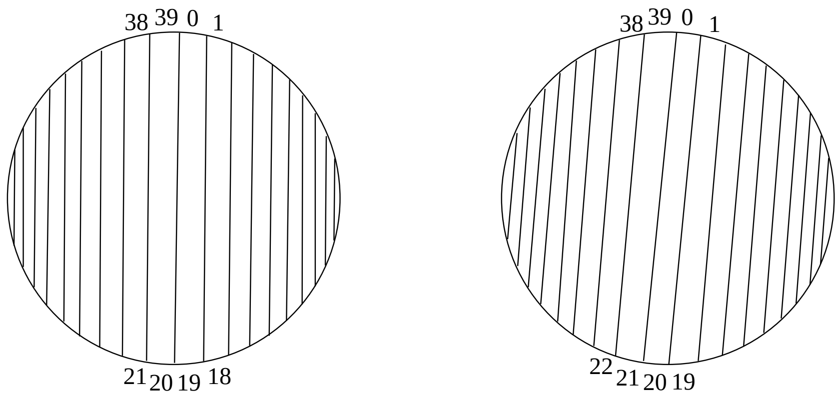

Figure 1: A projection is either (left) full-angle or (right) half-angle.

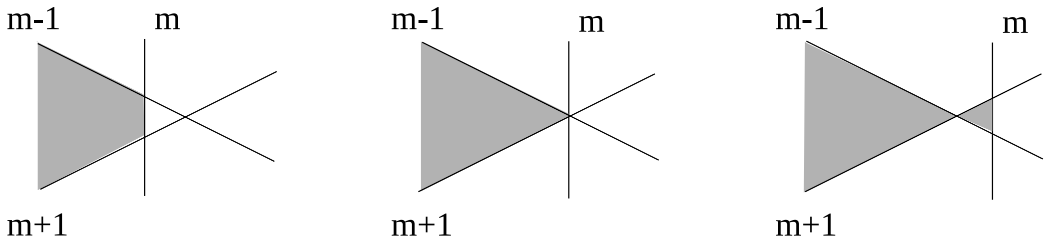

Figure 2: (left),(middle) Consistent boundary tubes result in a convex-hull whereas (right) inconsistent ones do not. (Adapted from [37].)

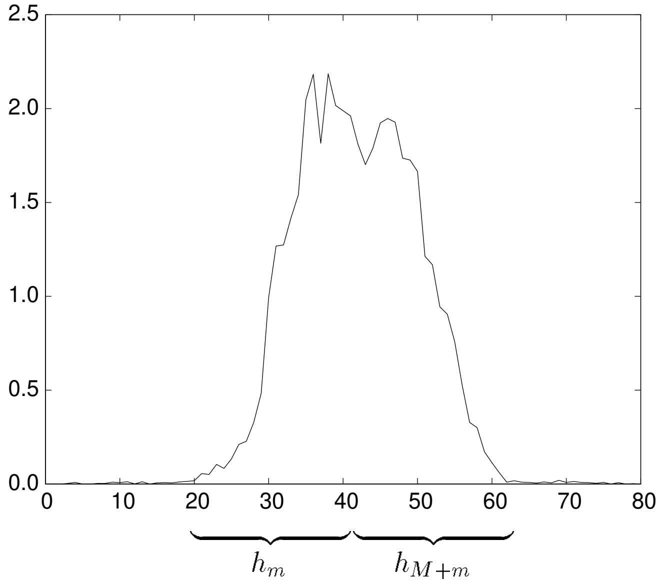

Figure 3: Estimated support values for the mth projection.

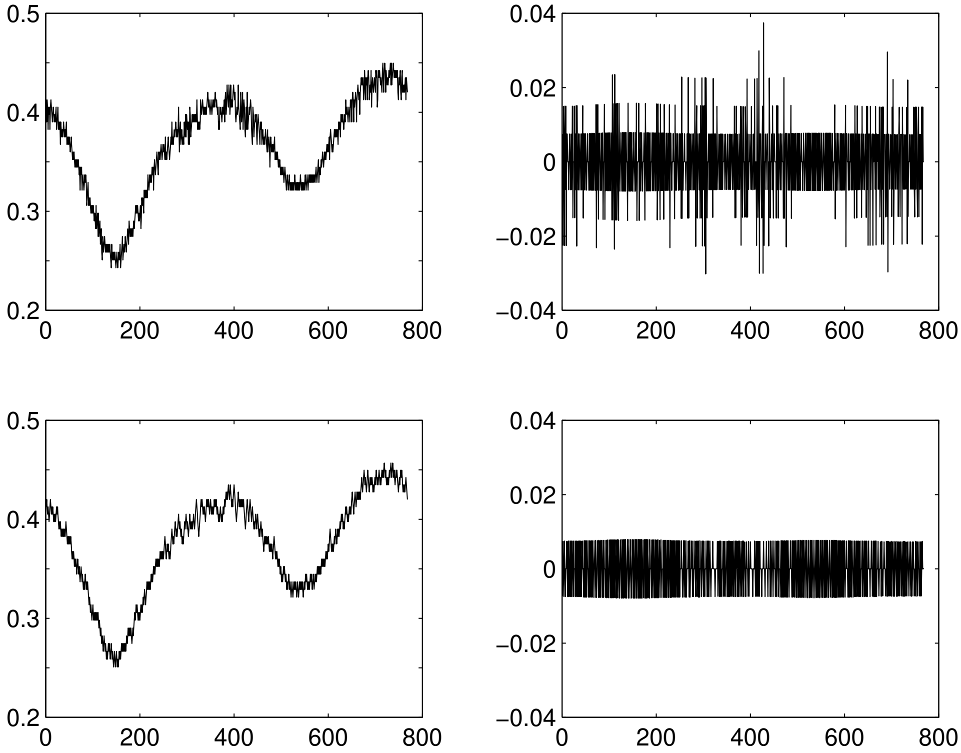

Figure 4: Illustration of (upper-left) $\boldsymbol{h}^{\circ}$, (upper-right) $C\boldsymbol{h}^{\circ}$, (lower-left) $\boldsymbol{h}^{*}$, and (lower-right) $C\boldsymbol{h}^{*}$.

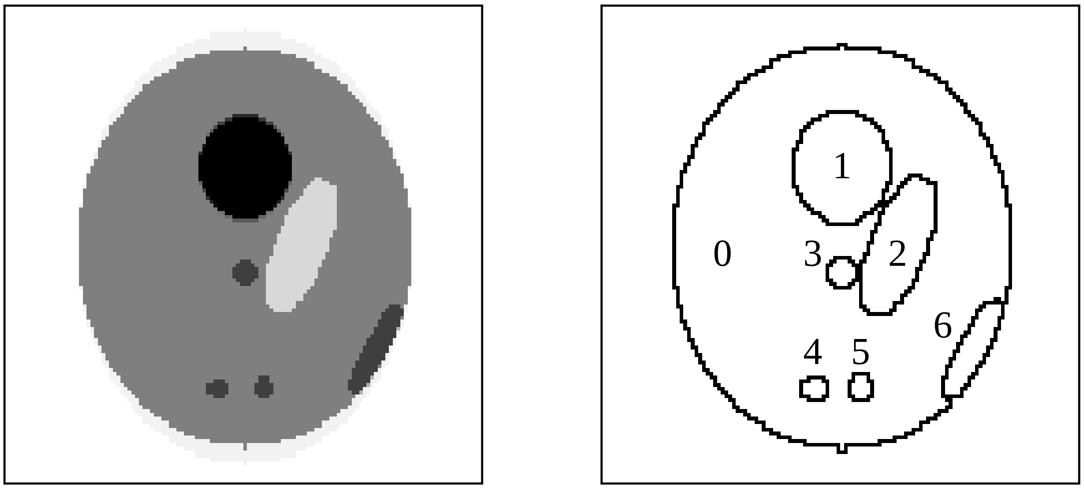

Figure 5: (left) Simulated phantom and (right) seven regions-of-interest (ROIs).

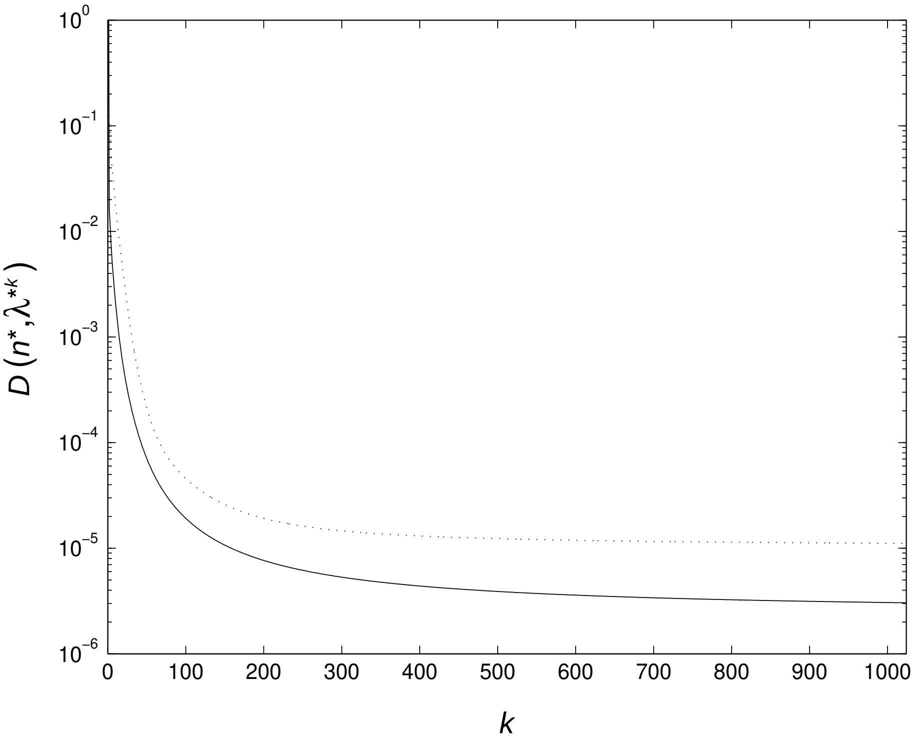

Figure 6: Convergence of the phantom reconstructions for (dotted line) the original EM-ML algorithm and (solid line) the focus-of-attention version.

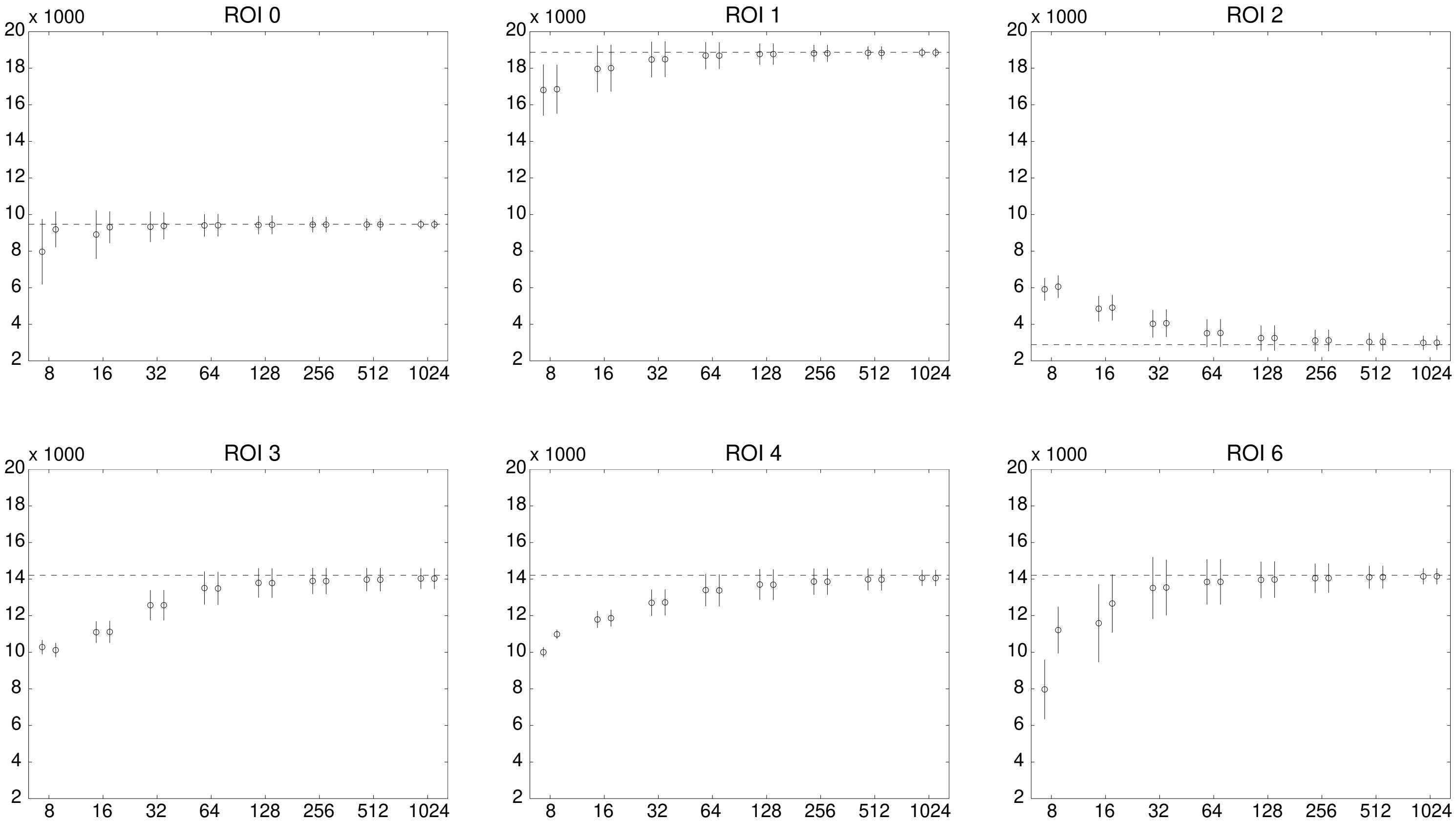

Figure 7: Per ROI mean pixel value plus-minus one standard deviation.

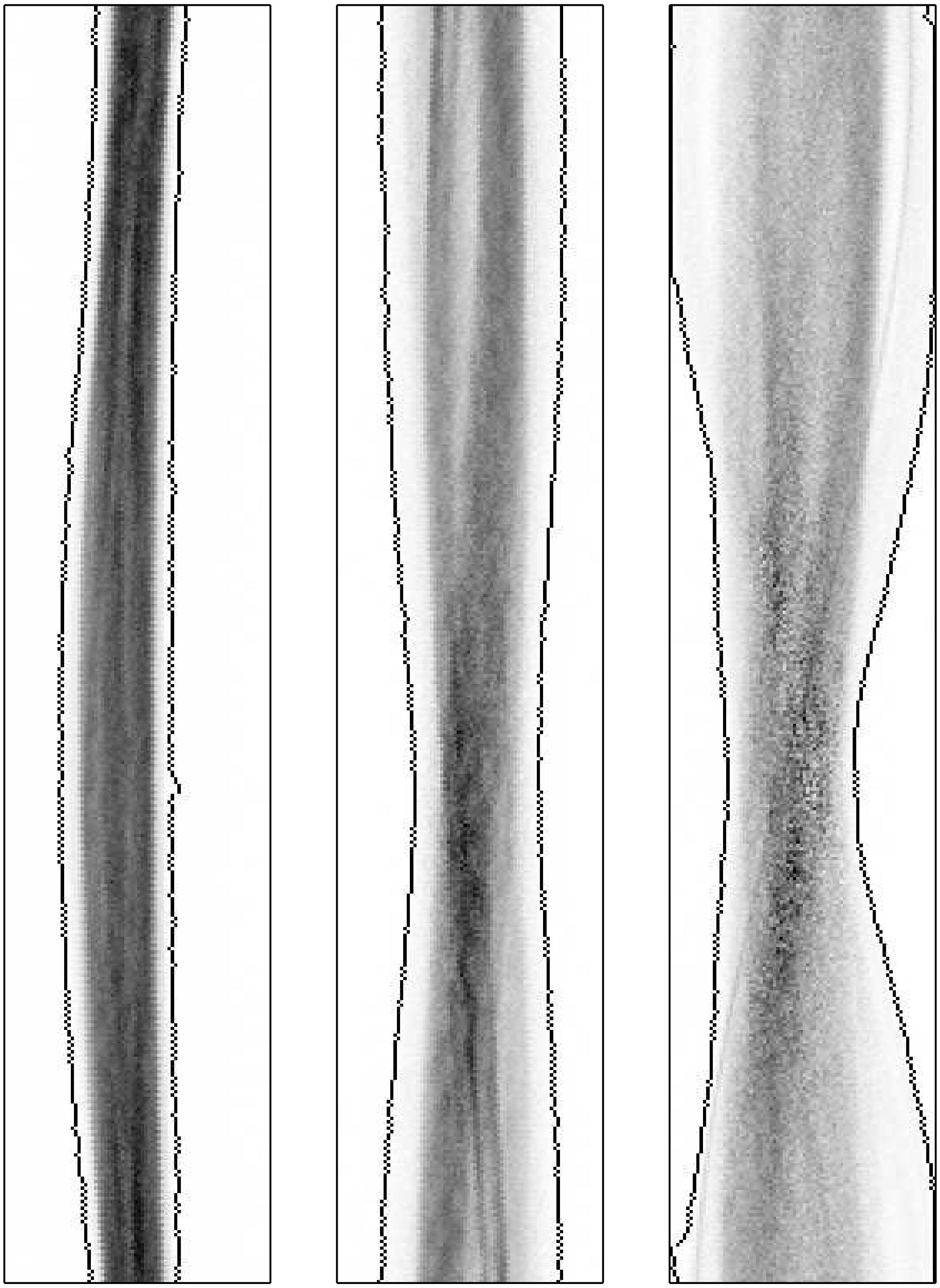

Figure 8: Sinograms and $\boldsymbol{n}^*$-masks for (left) the head scan, (middle) the chest scan, and (right) the abdomen scan.

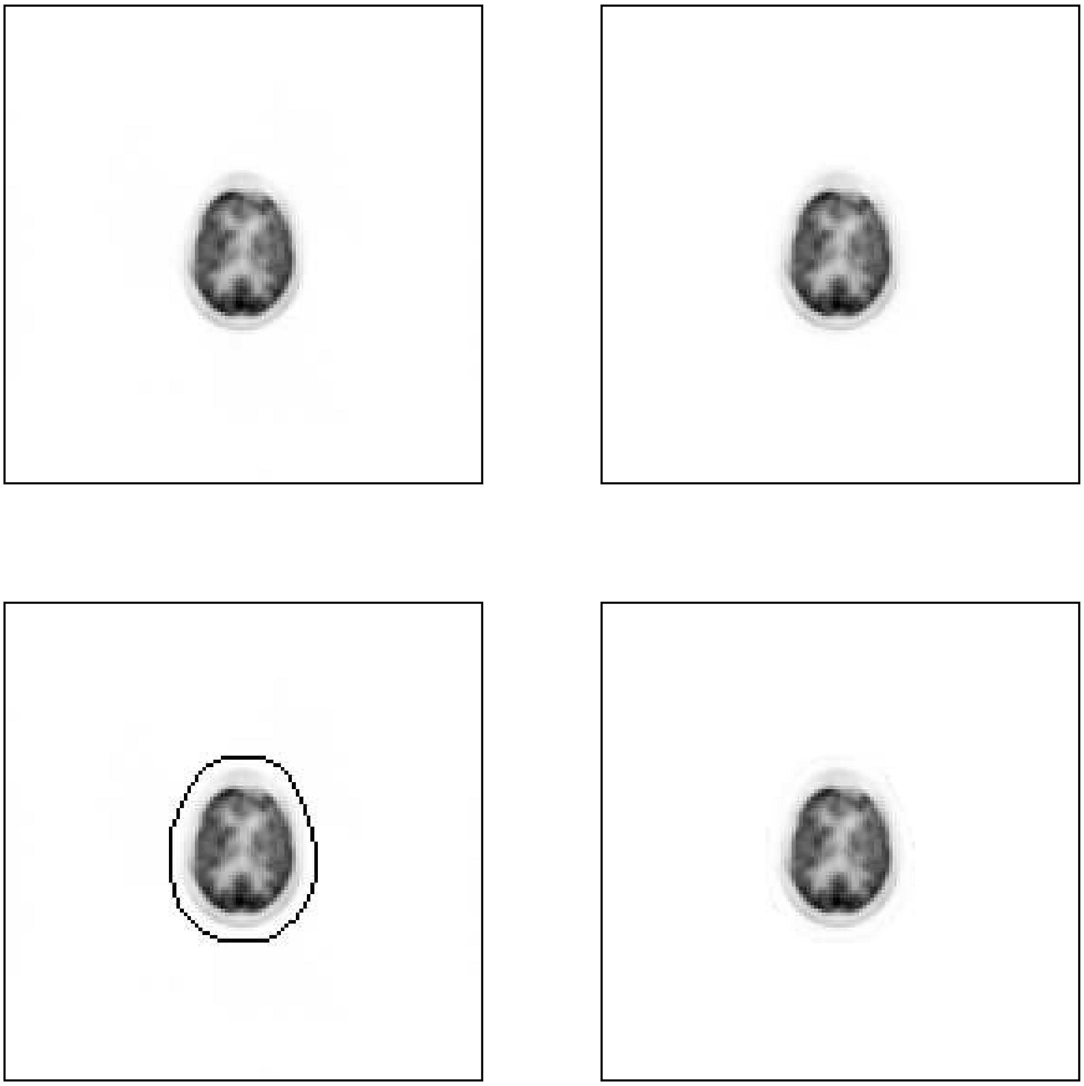

Figure 9: Reconstructed images for the head scan (z=1): (upper-left) original EM-ML result, (lower-left) $\boldsymbol{\lambda}$-mask overlay, (upper-right) focus-of-attention EM-ML including edge packing compensation, and (lower-right) focus-of-attention EM-ML without edge packing compensation.

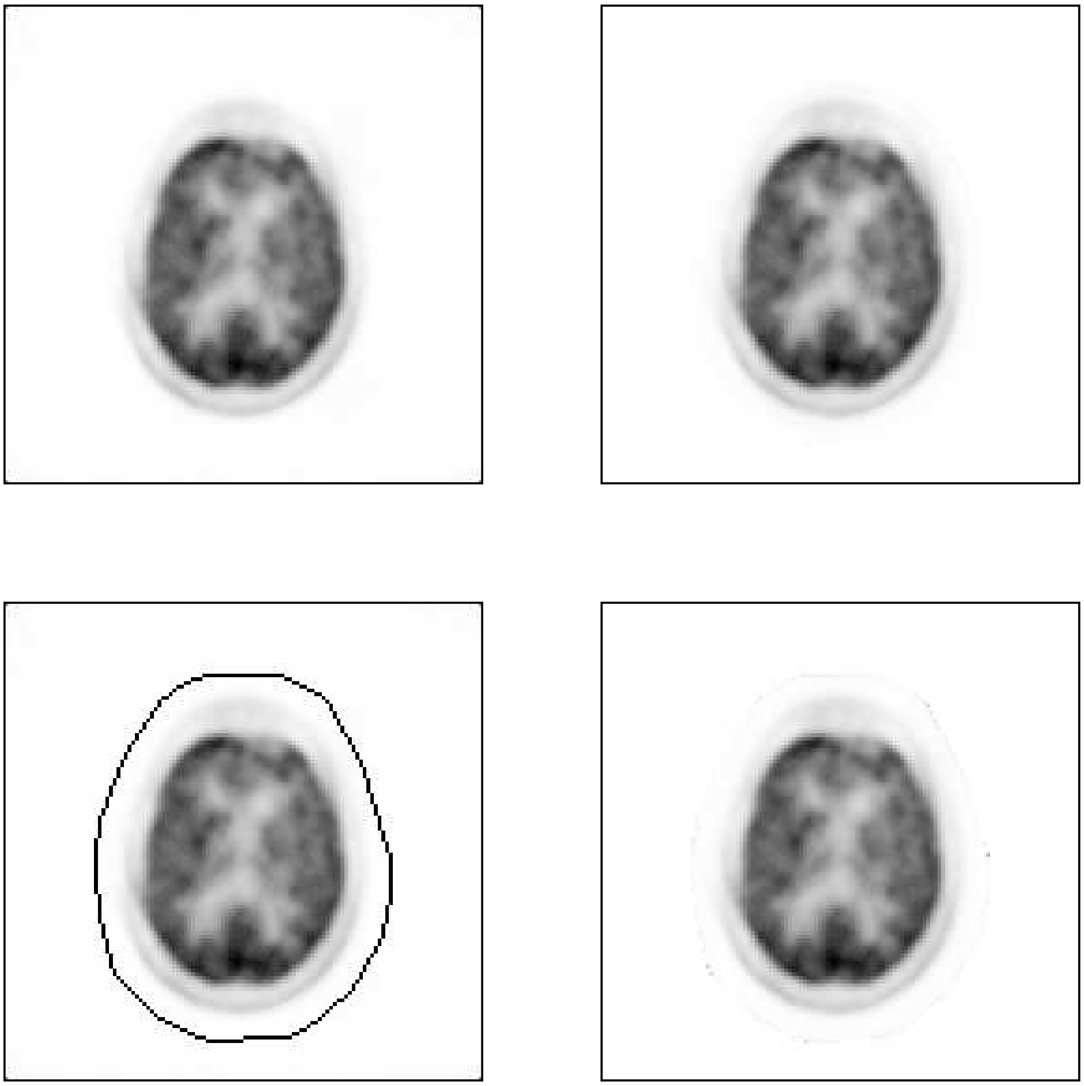

Figure 10: Reconstructed images for the head scan (z=2). See Fig. 9 for layout.

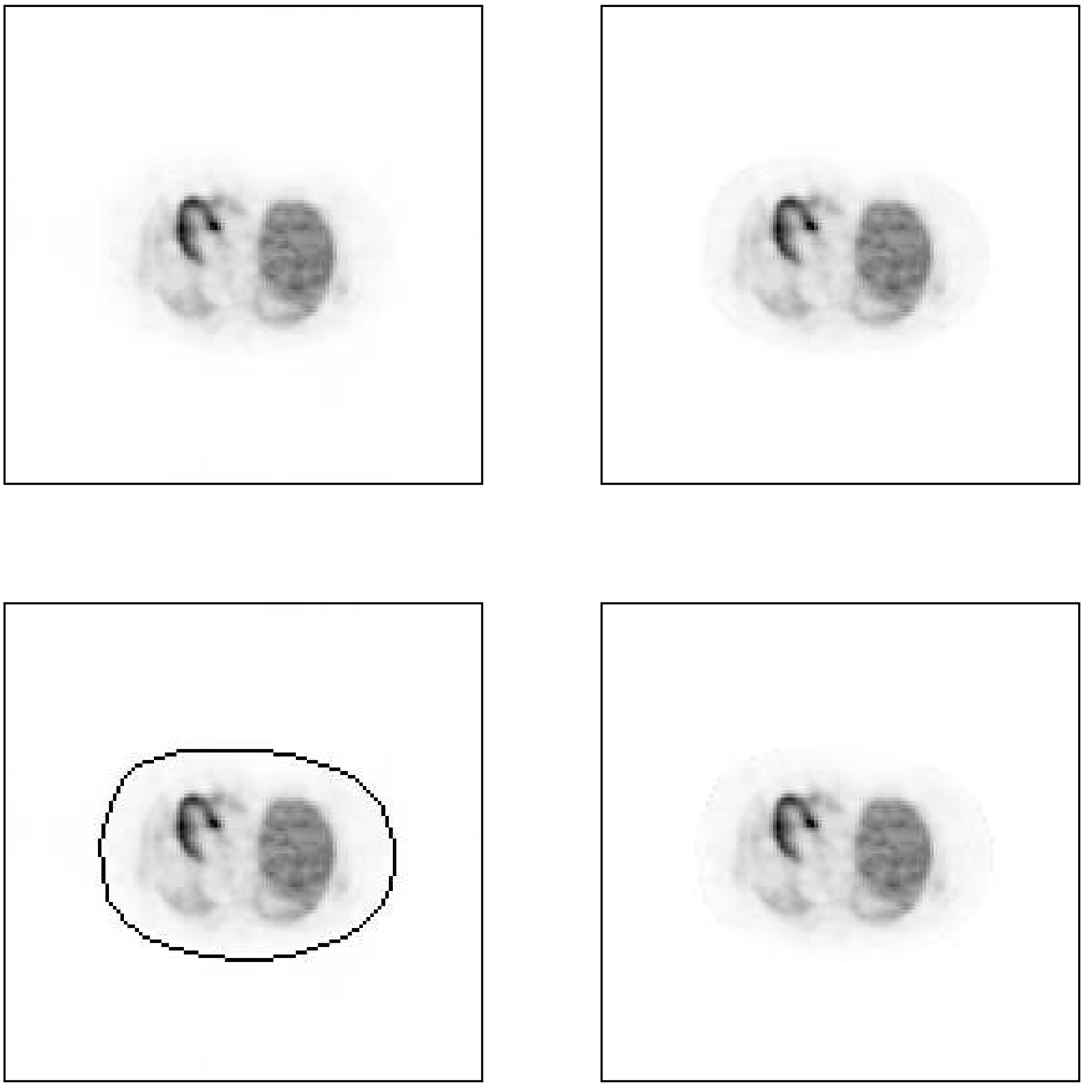

Figure 11: Reconstructed images for the chest scan (z=1). See Fig. 9 for layout.

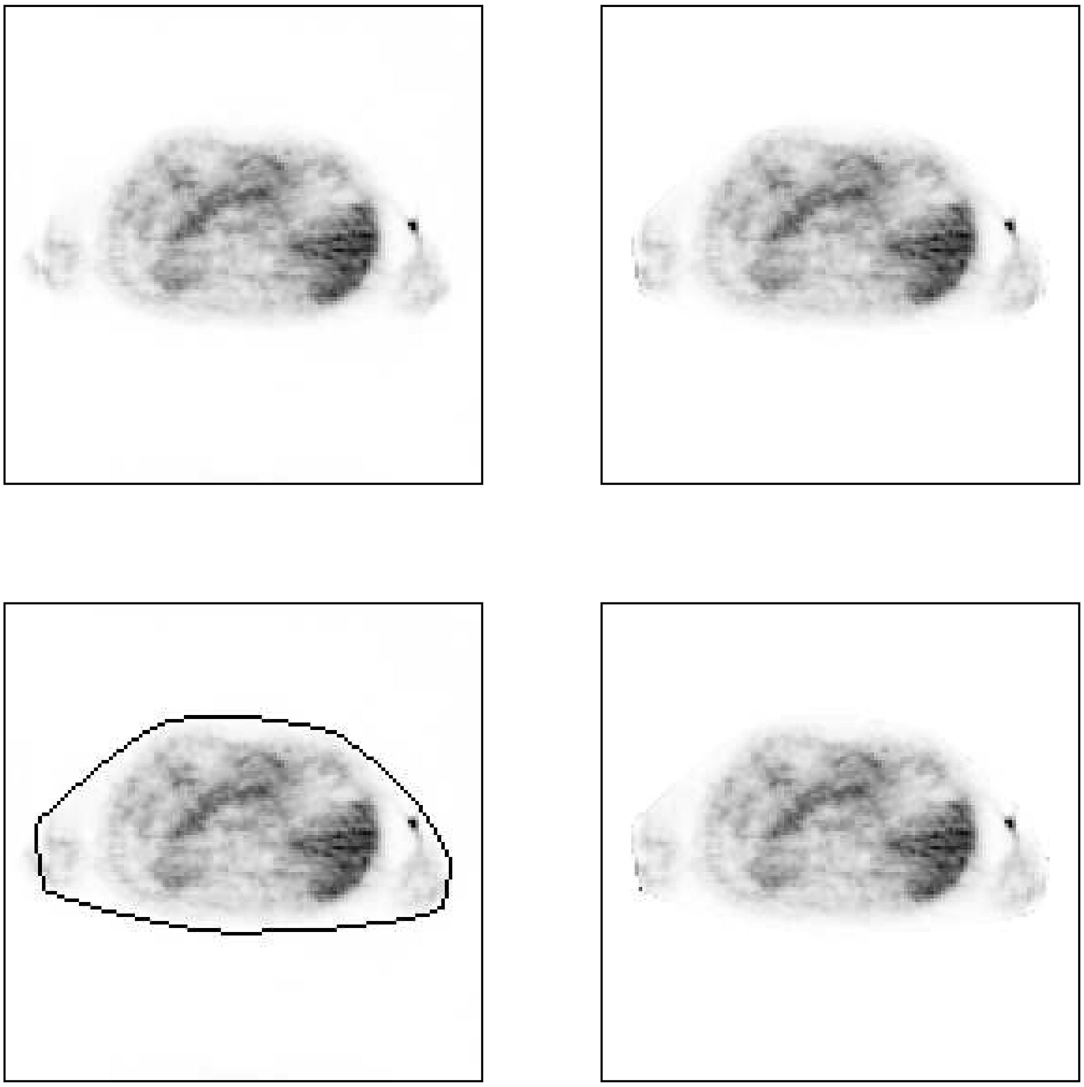

Figure 12: Reconstructed images for the abdomen scan (z=1). See Fig. 9 for layout.

Table 1: Array dimensions (no. elements).

	$\boldsymbol{n}^*$-mask	$\boldsymbol{\lambda}$-mask	P-matrix
Original	30.7K	16.4K	10.2M
Head, z=1	11.9K	1.6K	1.2M
Head, z=2	11.9K	6.3K	3.5M
Chest	18.6K	3.7K	2.7M
Abdomen	24.0K	5.1K	3.7M

Table 2: Timing results (CPU seconds).

	$\boldsymbol{n}^*$-mask	$\boldsymbol{\lambda}$-mask	P-matrix	Initialization	Per Iteration
Original	—	—	4.69s	4.69s	0.96s
Head, z=1	0.18s	0.17s	1.09s	1.44s	0.19s
Head, z=2	0.18s	0.15s	1.60s	1.93s	0.36s
Chest	0.15s	0.16s	2.03s	2.34s	0.31s
Abdomen	0.15s	0.16s	2.85s	3.16s	0.39s