



M Ú E G Y E T E M 1 7 8 2

Master's Thesis

Improving convergence properties of ML-EM reconstruction technique for Poisson statistics

Vencel Somai

Supervisor: Dr. Dávid Légrády

BME Institute of Nuclear Techniques

2018.

Alulírott Somai Vencel a Budapesti Műszaki és Gazdaságtudományi Egyetem fizika Msc szakos hallgatója kijelentem, hogy ezt a szakdolgozatot meg nem engedett segédeszközök nélkül, önállóan, a témavezető irányításával készítettem, és csak a megadott forrásokat használtam fel. minden olyan részt, melyet szó szerint, vagy azonos értelemben, de átfogalmazva más forrásból vettetem, a forrás megadásával jelöltetem.

Budapest, 2018.

.....
Aláírás

Köszönetnyilvánítás:

Köszönöm Légrády Dávidnak és Tolnai Gábornak a munkám során számos alkalommal nyújtott szakmai segítséget, a konzultációkra szánt idejüket és energiájukat. Külön köszönöm Légrády Dávidnak a dolgozat javítását, tartalmi, formai tanácsait. Köszönötettel tartozom családomnak minden otthonról nyújtott támogatásért, és különösképpen Zitának, amiért segített elérni az álmomat.



POISSON STATISZTIKÁJÚ ML-EM REKONSTRUKCIÓ KONVERGENCIATALAJDONSÁGAINAK JAVÍTÁSA

✉ Nyomatotáborát változat (<http://www.ttk.bme.hu/print/2713>)

Típus:

MSc diplomamunka téma - orvosi fizika

Félév: 2016/17/2.

Témavezető:

Név:	Dr. Légrády Dávid
Email cím:	legrady@reak.bme.hu
Intézet/Tanszék/Cégnév:	Nukleáris Technikai Intézet
Beosztás:	egy. docens

Hallgató:

Név:	Somai Vencel
Képzés:	Fizikus MSc - orvosi fizika

Elvárások:

Orvosi Képalkotás, funkcionálanalízis, valószínűségszámítás és statisztika magas szintű ismerete

Leírás:

A Maximum Likelihood Expectation Maximization (ML-EM) iteratív paraméterbecslő séma az orvosi képalkotás emissziós tomográfiás ágban egyre elfogadottabbá, lassan bevett eljárássá válik. A Nukleáris Technikai Intézetben évek óta fejlesztés alatt álló Monte-Carlo (MC) alapú, videokártyákon történő (GPU-n történő) számítást felhasználó ML-EM képrekonstruktív szoftver fejlesztésébe kapcsolódhat be a hallgató. A diplomamunka során meg kell vizsgálnia a hagyományos ML-EM séma konvergenciabességének felgyorsítási lehetőségeit, annak érdekében, hogy MC alapú modelllezési motor pontosságában rejlő előnyöket ki lehessen aknázní. A felgyorsítási lehetőségekkel a mérési zajok (kevés adat, random beütések, hamis beütések) tekintetében kell elvégzni. A gyorsítási eljárások közül a GPU-hoz és a MC motorhoz illeszkedő eljárásokat választhat a hallgató, mint például szingulárisérték-alapú felbontás, vagy nem strukturált rácsok használata. A konvergenciagyorsítási keretben meg kell vizsgálnia a MC motor felgyorsításának lehetőségeit is a kódban implementált szóráscsökkenési eljárások optimalizációjával. A hallgatónak eredményeit valós, mért adatokon tesztelnie és értkeznie kell.

Contents

1 Abstract and motivation	5
2 Introduction	5
2.1 Basics of PET	7
2.2 The PANNI code	8
2.3 Notations	9
2.4 ML-EM reconstruction	10
2.5 Anomalous behaviour of PANNI code	11
2.6 SVD analysis and convergence properties	13
2.7 Advantage of faithful modelling	17
2.8 Proposed improvement	18
3 Practical implementation of the SVD filter	20
3.1 Deconvolution	22
3.2 Modelling of the PSF	25
3.3 Fitting the PSF	25
3.4 Square root of the PSF	29
4 Reconstruction results	31
4.1 Homogeneous cylinder ring phantom	32
5 Position dependence of the PSF	34
5.1 Position dependence in due to scanner geometry	34
5.2 Karhunen-Loëve PSF decomposition	36
5.3 Incorporating material dependence	36
5.4 Homogeneous cylinder ring phantom v2	39
5.5 Simulated Derenzo phantom	41
5.6 Measured Derenzo phantom	44
6 Remaining problems	45
7 Conclusion	47

1 Abstract and motivation

The main motivation of this thesis is to bridge the theoretical results of the SVD filter proposed in my bachelor's thesis and the possibility of practical implementation in order to accelerate the convergence of ML-EM PET image reconstruction without modifying the fix point of the algorithm and deteriorating its stability to noise. To the best knowledge of the author and the editorial team of Radiology and Oncology journal the SVD filter method is a novelty in the field of PET image reconstruction.

The SVD filter has proven to be very efficient with respect to convergence acceleration but due to the need of a numerical singular value decomposition it brings serious computational challenge to the role making the implementation nearly impossible and highly impractical. The implied storage requirement is also problematic. In order to solve this problem, a (de)convolution model is presented in this thesis to approximate the effect of the SVD filter without introducing its calculation and storage related difficulties. The key point of a proper deconvolution approach is the ability of accounting for the position dependence of the kernel to be deconvolved which is an inherent property of the SVD-filter. If this requirement is met the (de)convolution approach is a faithful approximation of the SVD-filter but excluding the computational difficulties. As a powerful solution, a subkernel method is presented which is a novel contribution altogether with the overall SVD-filtering and its deconvolution model.

The resulted deconvolution-like filtering successfully abolishes the material dependent inhomogeneity artefact which previously appeared when faithful physical modelling was used in the back projection. Also, overall convergence speed is increased significantly above the level achieved by the algorithm when simplified modelling is used in the back projection giving the best reconstruction result so far.

The presented method was tested on simulations with different mathematical phantoms and an FDG micro Derenzo phantom measurement as well, and the pitfalls and difficulties are discussed together with possible future development directions.

2 Introduction

In emission tomography maximum likelihood expectation maximization (ML-EM) image reconstruction technique [1, 2] has replaced the analytical approaches (e.g. the widely used filtered back projection) in several applications, since ML-EM offers improvements in spatial resolution and stability due to the more accurate modelling of the system and to the ability of accounting for noise structure [3].

In exchange ML-EM has only a linear rate of convergence [2] and its computational cost is still tedious even with the rapidly increasing processing capacity of current computers. Thus a significant part of recent research activities aims at accelerating the algorithm [3–5]. Another important property partly connected to the low convergence rate is the maximal resolution achievable for a given reconstruction method given a certain noise level towards which most of the developments are directed [6–9]. In order to achieve improvement in both convergence rate and spatial resolution an ML-EM positron emission tomography (PET) reconstruction code has been developed with graphical processing unit (GPU) based Monte Carlo (MC) engine [10]. GPUs parallel threads allow for running the inherently parallel neutral particle MC transport simulations approximately hundred times faster than on a comparably priced CPU thus significantly reducing the time required for the reconstruction. As increased computational capacity allows for better physics modelling the main novelty of this code is the ability of full particle transport modelling as accurate as it is worthwhile in hope of improving image quality [11].

Contrary to expectations such faithful physics modelling in the back-projecting (BP) step of the algorithm causes strong artefacts: modelling positron range leads to tissue dependent inhomogeneity artefacts in the reconstructed image. Furthermore, these inhomogeneities disappear when simplified MC simulations are used without. All the differences between the two cases occur in the system matrix (derived from the MC simulations) of the BP step. These differences were analysed with respect to the convergence properties and stability to noise in a smaller test system by means of singular value decomposition (SVD) which is a powerful tool when analyzing rectangular matrices. According to the SVD analysis there happens to be a significant advantage of the matrix belonging to the simplified simulations in terms of both singular values and vectors that characterized the convergence properties and stability of the algorithm. In other words more accurate physical modelling is less efficient in terms of convergence and these differences explained the perceived artefacts.

Taking advantage of these results an a posteriori filtering matrix was created applied in each iteration after the BP step with which these differences could be further amplified for speeding up the convergence, but without spoiling the stability to noise. However, the proposed filtering method requires the calculation and storage of a matrix of a size $\text{Number of voxels} \times \text{Number of voxels}$. For a real scanner it can take up to several terabytes.

The main contribution of this thesis is the bridging of the theoretical results and the practical implementation of the proposed filtering method which was impossible before. A convolution model is formed in order to replace the filter matrix multiplication in the back projection with deconvolution which obviates the need of the computationally intensive and usually infeasible numerical SVD.

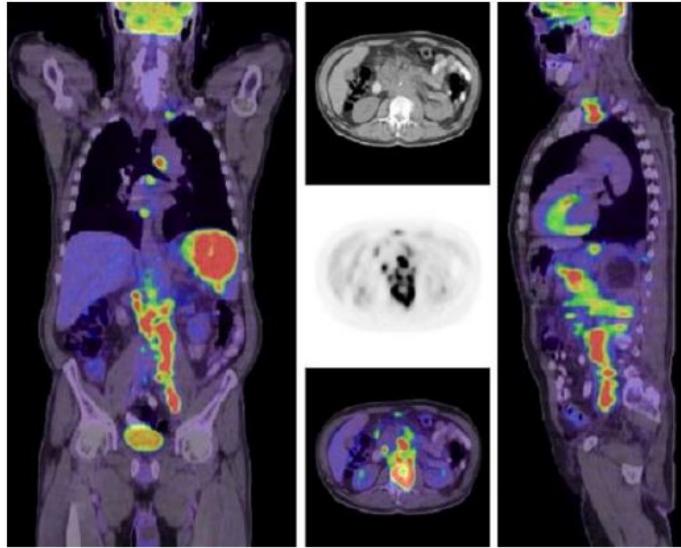


Figure 1: Cooregistration of functional (in colour) and anatomical (grayscale) data

The proposed deconvolution is easy to calculate by means of convolution theorem and fast Fourier-transform (FFT) and only requires the storage of the point spread function (PSF) of the system. Despite all the calculations were implemented in MATLAB the proposed method increases the duration of one ML-EM iteration only by 10% (MATLAB calling takes most of the extra time), but accelerates the reconstruction to perform significantly better than the best setting so far.

The text of this section is partly published in [19]

2.1 Basics of PET

Positron Emission Tomography (PET) is a functional imaging modality. It observes physiological functions *in vivo* by looking at for example blood flow, metabolism, neurotransmitters and radiolabelled drugs and capturing their spatial distribution and kinetics. PET is frequently used in clinics to monitor the increased fluorodeoxyglucose(FDG) uptake of cancerous or inflamed cells in order to help diagnosis and guide therapy.

PET has been proven a powerful tool in preclinical imaging as well, among others in the field of drug tests and cognitive disease related medical research.

In order to give useful information anatomical data is also needed which is used for the cooregistration of the functional image. Certain algorithms also use the anatomical data in the reconstruction itself to generate the functional image. Further reading can be found in [12].

2.2 The PANNI code

A 3D MC based ML-EM image reconstruction code named PANNI (PET Aimed Novel Nuclear Imager) has been developed in the framework of the TeraTomo project in the Intitute of Nuclear Techniques of BME. [13, 14]. PANNI is a MC based image reconstruction software written for GPUs using C and CUDA environments surmising roughly 40 000 lines [11]. The key feature of our software is the possibility of faithful MC modelling which accounts for positron range, gamma photon-matter interaction and detector response supported by advanced variance reduction methods. Detectors around the object are positioned on a quasi-cylindrical surface with dodecagon cross section. Detector response is either simulated or a pre-generated tabulated response function may be used. Positron range modelling simplifies to the following probability density function [15]

$$\rho(r) = aA^2re^{-Ar} + bB^2re^{-Br} \quad (2.2.1)$$

with r being the positron range distance, a , A , b and B material dependent constants. As sampling each of the terms is equivalent to sampling the sum of two exponentially distributed random variables x can be obtained by using double exponential sampling [16]. Advanced variance reduction methods are implemented for source angular sampling outgoing direction and energy biasing and for free flight sampling. The MC engine has been validated against MCNP5 [17]. The code is capable of simulating 10^8 photon pairs per second on a commercially available GPU (NVidia GeForce 690). Both the forward projection and the BP steps are carried out via the MC method. In the BP step some of the physics modelling may be turned off. The code has been tested with two geometries, a sophisticated scanner geometry (“full system”) and a simplified smaller system (“1D model”). Acquisition geometry for the full system can be set as wished, in our current setup it consists of a dodecagon with inscribed circle radius of 8.7cm , packed on each side with an array of 39×81 LYSO detector pixels of 1.17mm sided squares comparable to a small animal PET scanner similar to the Mediso nanoScan PET/CT scanner (figure 2). Coincidence counting is accepted between detector pixels on opposite and next to opposite dodecagon sides (1:1 and 1:3 coincidence). If not stated otherwise the voxel space of the full system is divided into $128 \times 128 \times 128$ voxels ($0,3\text{mm}$ sided) and contains a water-cylinder (light grey area in figure 3 and 4) except for a smaller cylindrical area containing bone material (dark grey area in figure 3 and 4). Activity phantom for the evaluation is a cylindrical ring of ^{15}O partially located in bone material (the more commonly used ^{18}F gives less conspicuous results). From now on simplified modelling means the neglect of the positron range effect in the back-projecting MC simulations in contrast with faithful modelling which accounts for positron range.

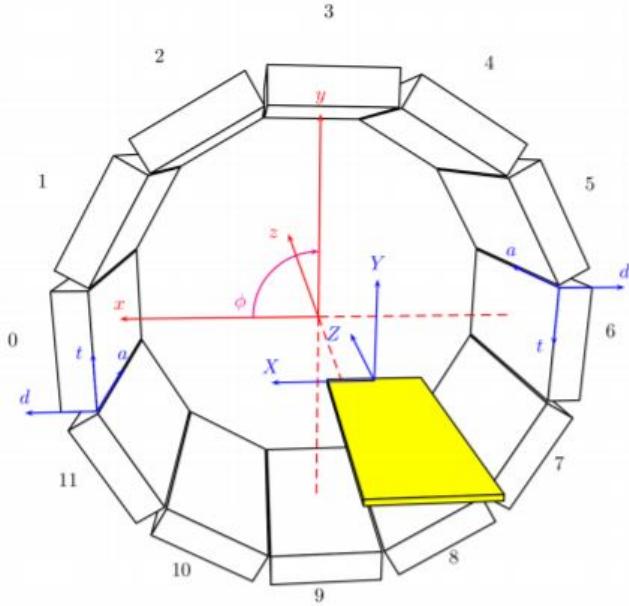


Figure 2: The sophisticated scanner geometry of PANNI¹⁷

The voxel space of the 1D model is a 38,4mm long interval containing 256 voxels half of which is located in bone material the other half in water. Detector pixels are assigned in two parallel sections each containing 81 crystals of a size 1,17 × 1,17mm (see figure 5). Every pixel is in coincidence with every pixel on the opposite side. Roughly speaking the 1D model is a cross-section of the full system geometry of PANNI.

2.3 Notations

If not stated otherwise the following notations are used:

* stands for convolution

x : vector of activity estimate in the voxels, $x \in R_{\geq 0}^{N_{voxel}}$

A : system (response) matrix

Ax : forward projection

y_m : vector of measured data

$y = Ax$: vector of forward-projected data

y_r : pointwise (i.e. Hadamard) ratio vector of measured and forward-projected data $y_r = \frac{y_m}{y}$

v_i : i^{th} singular vector corresponding to voxel space

σ_i : i^{th} singular value

Lor: Line of response

u_i : i^{th} singular vector corresponding to Line of response - space

T : in superscript means transpose

$A^T y_r$: back projection

Posrange: Positron range

BP: back projection

BP posrange OFF: simulation neglects positron range in the BP

BP posrange ON: simulation accounts for positron range in the BP

MC: Monte Carlo

FWHM: Full Width at Half Maximum

FWTM: Full Width at Tenth Maximum

SVD: singular value decomposition

RL: Richardson-Lucy deconvolution

$L_2 - norm$: Vectorial $L_2 - norm$ divided by the $L_2 - norm$ of the activity distribution and multiplied by 100: $L_2 - norm = 100 * \sqrt{\frac{\sum_{i=1}^{N_{voxel}} (x_{i,phantom} - x_{i,recon})^2}{\sum_{i=1}^{N_{voxel}} x_{i,recon}^2}}$

$CC - norm$: Vectorial cross-correlation norm, normalized to the 0 – 100 interval: $CC - norm = 100 * \|1 - \frac{x_{i,phantom}^T x_{i,recon}}{\|x_{i,phantom}\| \|x_{i,recon}\|}\|$

2.4 ML-EM reconstruction

Maximum Likelihood Expectation Maximization is an iterative method to find maximum likelihood or maximum a posteriori (MAP) estimates of parameters in statistical models (x), where the model depends on unobserved latent variables (z). The EM iteration alternates between performing an expectation (E) step, which creates a function for the expectation of the log-likelihood evaluated using the current estimate for the parameters ($x^{(n)}$) and the given realisation (y), and a maximization (M) step, which computes parameters maximizing the expected log-likelihood found in the E step:

- E - step:

$$Q(x, x^{(n)}) = \mathbb{E}_{z|y, x^{(n)}} [\log \mathbb{L}(x^{(n)}, y, z)] \quad (2.4.1)$$

- M - step:

$$x^{(n+1)} = \underset{x}{argmax} [Q(x, x^{(n)})] \quad (2.4.2)$$

In emission tomography the latent variable z_{ij} denotes the (unobserved, latent) number of events detected in i^{th} LoR (y_i) which come from the j^{th} voxel. Therefore $\sum_j z_{ij} = y_i$ and z_{ij} follows Poisson

distribution. Its expected value conditioned on given y_i : $\mathbb{E}[z_{ij}|y_i] = \frac{A_{ij}x_j}{\sum_j A_{ij}x_j}y_i$ as its conditional distribution is multinomial.

After performing E and M steps the following iteration formula is obtained [2]

$$x_i^{(n+1)} = x_i^{(n)} \left\{ \frac{\sum_i A_{ij} \frac{y_i}{[\sum_j A_{ij}x_j]}}{\sum_i A_{ij}} \right\} \quad (2.4.3)$$

Square brackets contain the result of the so-called forward projection and the braces contain the result of the so-called back projection.

2.5 Anomalous behaviour of PANNI code

The key feature of the PANNI code is the use of full MC transport modelling by following particles through every possible collision in the object of observation as faithfully to the laws of nature as practically worthwhile thanks to GPU implementation and the applied advanced variance reduction methods. However, faithful physical modelling in the backprojection seemed to be deteriorating with respect to the quality of the reconstructed image.

Two reconstruction results were compared for the sophisticated scanner geometry: one with full physical modelling in the BP and one omitting positron range in the BP. After 80 iterations faithful modelling produced the reconstruction in figure 3. The cross section of the cylindrical ring phantom in radial direction is originally a box function which is blurred due to gridding and averaging in a given voxel (similarly to partial volume effect). Therefore, it can be well approximated by a Gaussian with a Full Width at Half Maximum of 3 voxels which is indicated by the red line on the figure. The Full Width at Half Maximum/Full Width at Tenth Maximum is calculated by fitting a Gaussian in radial direction along the ring separately for each angular position with resolution of 1 degree:

Accounting for positron range in the BP caused systematic inhomogeneity in the reconstructed image. The activity estimation in the bone material is appropriate ($FWHM = 3,5$ voxel) in contrast with the activity of the voxels located in water which is underestimated ($FWHM = 5$ voxels). The inhomogeneity reduces with simplified BP, when positron range is neglected in the MC simulation. Also the FWHM is reduced in the water area see figure 4.

The effect of modelling any physical phenomenon appears in the system matrix, thus any possible analysis aims at finding the differences in the BP system matrix caused by positron range. Such an analysis was performed in [18] and [19] based on SVD and focusing on the convergence properties of the resulted reconstruction algorithm. (The SVD calculations are similar to those in [20])

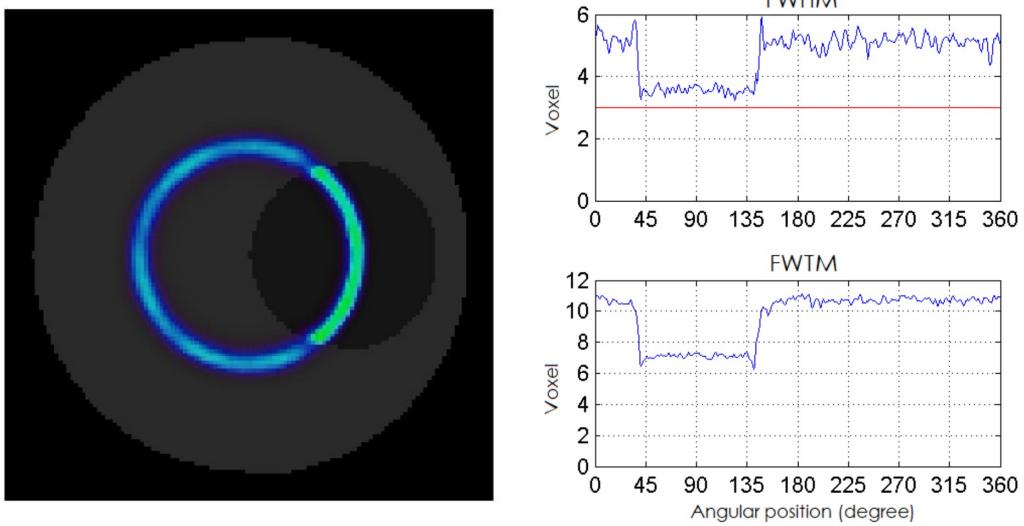


Figure 3: Top view of the reconstruction of the cylinder-ring mathematical phantom of the full system with faithful modelling in the BP. Light gray area represents water, dark gray area represents bone material. Underestimated activity and increased FWHM/FWTM can be seen for voxels located in water. FWHM and FWTM are calculated along the ring. Red line indicates the phantom ideal FWHM.

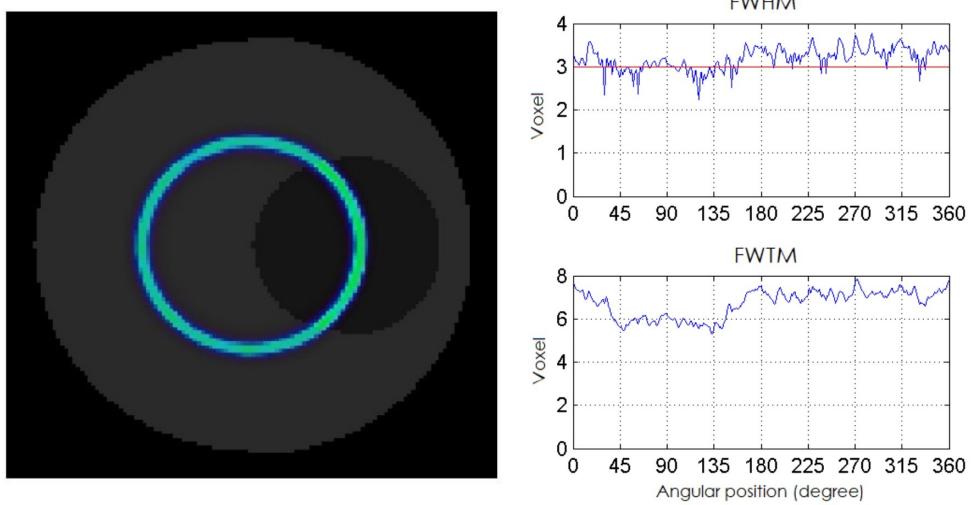


Figure 4: Top view of the reconstruction of the cylinder-ring mathematical phantom of the full system with simplified modelling in the BP. Homogeneous activity estimate and FWHM can be seen along the ring. Neglect of the positron range in the BP abolished the artefact of figure 3. and phantom ideal FWHM is reached.

2.6 SVD analysis and convergence properties

The effect of positron range modelling and the average positron range is accounted for by the system matrix. As an obvious tool for the analysis of rectangular matrices (as usually there is more LoR than voxel), the algorithm and the BP step are examined by means of singular value decomposition. SVD is a factorisation of any $m \times n$ real or complex matrix A of the form $A = UDV^T$. The following notation is used: U is an $m \times n$ matrix, D and V are $n \times n$ square matrices. In general, matrix V has k orthogonal columns where k is the rank of the system matrix A . V can be completed to a $n \times n$ dimensional matrix by adding $n - k$ orthogonal vectors from the null space of A^T to form a basis in the voxel space. The first k columns of U are also orthogonal and can also be completed to a basis in the sinogram space by adding $m - k$ orthogonal vectors from the null space of A . In this last case, D is zero filled to a $m \times n$ dimensional matrix. In the point of view of our analysis the completion of U is not needed and we chose the nomenclature where U has only n (as A has full rank, thus $k = n$) columns. D and V are $n \times n$ in this case. SVD was used for the analysis of convergence speed of the reconstruction algorithm with respect to the applied BP. According to reference [18, 20] the speed of convergence of PET ML-EM algorithm particularly depends on the singular values of the BP system matrix. Singular values represent relative weights for the voxel space basis vectors (i.e. corresponding voxel space singular vectors) in the update process of the previous activity estimate in a given iteration. Sinogram space singular vectors can measure the information content of a given measurement-forward projection Hadamard ratio in the corresponding BP step by means of Picard condition formalism which states that for the existence of a square integrable solution to the problem $y = \mathcal{A}x$ the following has to be true (\mathcal{A} is the integral operator the discretization of which is the system matrix A) [21]:

$$\sum_{i=1}^{\infty} \left(\frac{u_i^T y}{\sigma_i} \right)^2 < \infty \quad (2.6.1)$$

In case of matrices instead of integral operators, the discrete Picard condition requires the spectral coefficients $|u_i^T y|$ to decay faster in average than the singular values [21]. Despite BP is not a direct inversion from this point of view the faster is the decay of the spectral coefficients $|u_i^T y_r|$ of the ratio as the index increases the heavier the blurring of the BP. Higher frequency components level off at a plateau which is dominated by noise and can be regarded as an error-level estimate [21] because these components do not contain information for the corresponding BP. Even accounting for voxel space effects only (e.g. positron range) sinogram space singular vectors are not the same for the simplified and faithful modelling. The BP step of the algorithm back projects the Hadamard ratio of the measured and the currently forward-projected data

$$A^T y_r = V D^T U^T y_r \quad (2.6.2)$$

The aforementioned difference in the sinogram space basis affects the $U^T y_r$ product, i.e. the spectral coefficients of the Hadamard ratio.

In consequence of the analysis above, it can be concluded that modelling the positron range i.e. accounting for a physical effect in the voxel space influences the convergence properties of the reconstruction algorithm in three different manners, which are explained in details in the following by means of the 1D model:

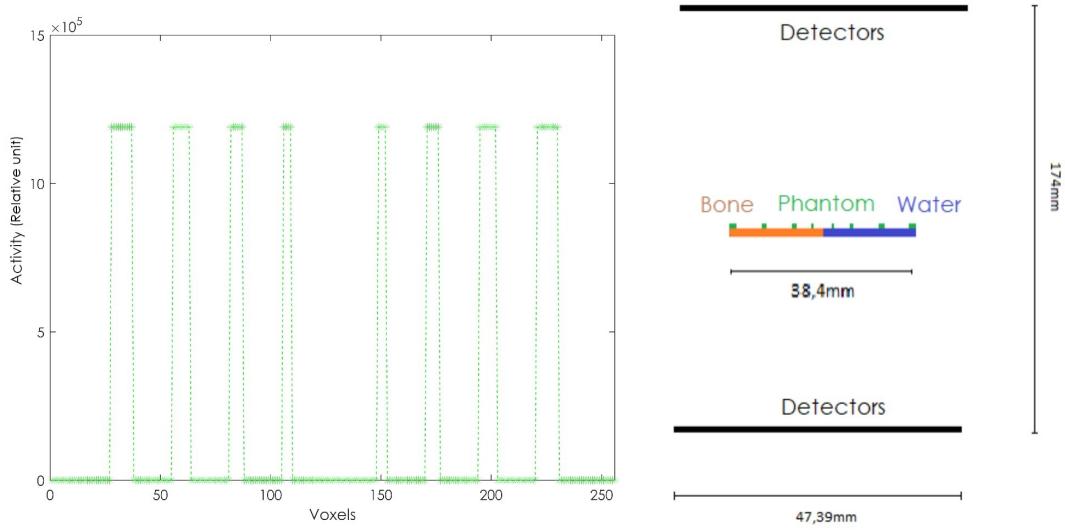


Figure 5: Mathematical phantom and system geometry for 1D model. 1-128 voxels are located in bone material, 129-256 voxels are located in water

The 1D model contains only positron range modelling, neither gamma photon-matter interaction nor detector response modelling is included. Detection is based on the angle of view of the detector from a given voxel. The two analysed settings are: positron range neglected (BP posrange OFF) and positron range modelled (BP posrange ON) in the BP. Forward projection always accounts for positron range.

- Differences in singular values:

Comparing the BP posrange OFF and BP posrange ON case in terms of singular values of the system matrix, figure 6 shows the positive difference for the first 133 index belonging to the former setting.

- Differences in voxel space singular vectors:

In the light of the convergence analysis of reference [18,20] smaller singular values mean that the

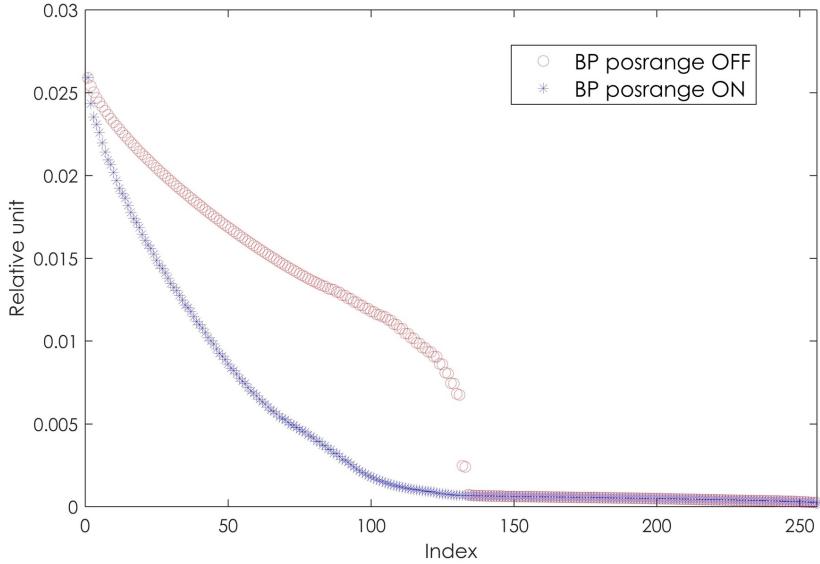


Figure 6: Singular values of the system matrix for positron range neglecting and modelling case. Increased values of the former imply the faster convergence of the corresponding (first 133) basis component of the activity estimate

corresponding frequency components of the solution are later reconstructed with positron range modelling compared to the positron range neglecting BP. This space frequency characterises the singular vectors of the voxel space, which is a second but not less significant difference between the two types of system matrices (see Figure 7).

Figure 7 on the left accounts only for the symmetries of the system while the right figure reflects also the tissue map of the volume. As the average positron free path is much longer in water than in the bone material space frequency of every basis vector is smaller in the water area. Thus the reconstruction of the activity of these voxels is significantly slower and this property is the reason of the obtained artefact resulted from faithful modelling in the BP and partly the solution to the perceived anomalous behaviour.

- Differences in the sinogram space:

The third and final difference occurs in the sinogram space basis vectors with which the measurement-forward projection Hadamard ratio can be unfolded in a given iteration. The absolute value of the obtained spectral coefficients can be seen in Figure 8 after 15 iterations for the positron range neglecting and modelling case respectively (the spectrum varies slowly through iterations).

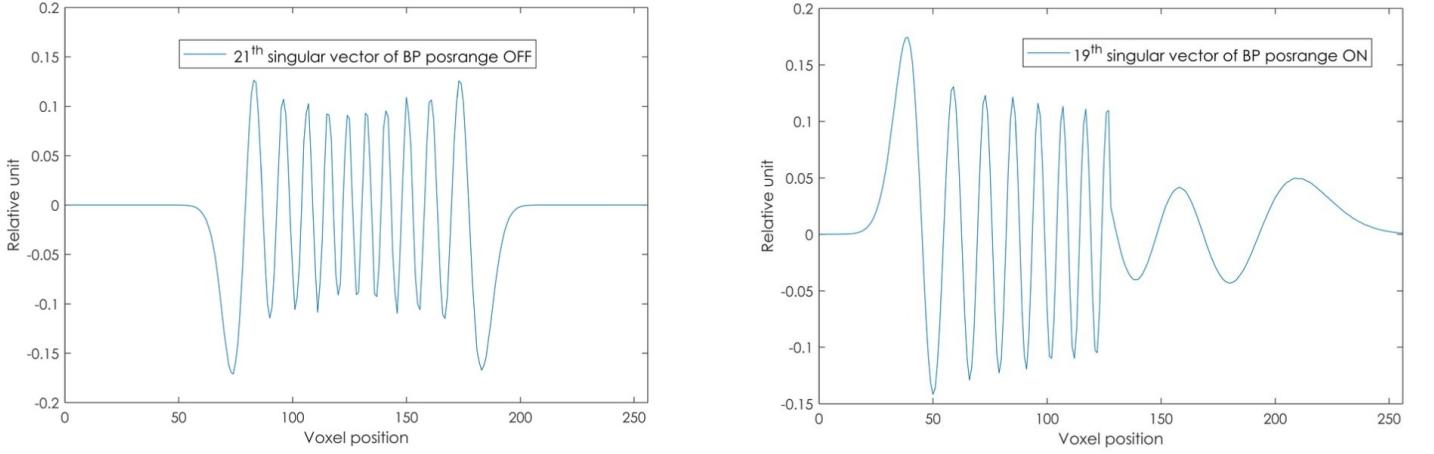


Figure 7: One of the voxel space singular vectors of the system matrices corresponding to positron range neglect (left - BP posrange OFF) and positron range modelling (right - BP posrange ON). BP posrange OFF reflects only the symmetries of the geometry. BP posrange ON accounts for the material map as well, increased position uncertainty due to longer average positron free path implies lower space-frequency in water area

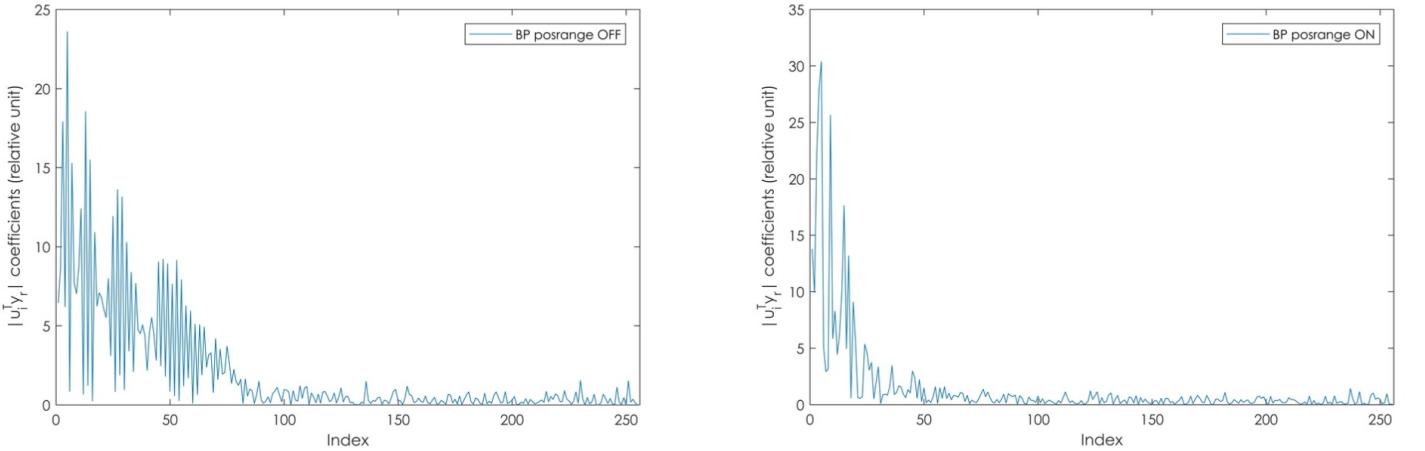


Figure 8: Absolute value of the spectral coefficients of the measurement-forward projection Hadamard ratio in the LoR basis corresponding to the positron range neglect (left - BP posrange OFF) and positron range modelling (right - BP posrange ON). Faster decay means less information gathered as the coefficients of the horizontal plateau are corrupted by noise thus it represents an error level estimate. Due to one to one correspondance property of SVD between LoR and voxel space singular (basis) vectors these basis coefficients of the activity are not hoped to be correctly estimated

Faster decay in the spectral coefficients equals to heavier blurring in the BP [21]. This means that the positron range modelling gathers less information from the Hadamard ratio in a given iteration than the positron range neglecting BP. This also implies the faster convergence.

To summarise the findings above, all of the three SVD matrices from the factorisation of the system matrix were analysed. The results explained the perceived artefact as the convergence speed of the scheme with positron range modelling BP was material dependent and the further advantage of the simplified BP in terms of overall convergence speed. These properties led to the systematic inhomogeneities in the reconstruction.

Despite the convergence speed analysis explained the perceived artefact the result is still anomalous as full physical modelling means additional information in the reconstruction (compared to simplified modelling) which should reflect in better image quality. The next subsection discusses this anomalous behaviour and presents the advantage of the faithful modelling.

2.7 Advantage of faithful modelling

Despite the lower and material dependent convergence speed, it was showed in [18, 19] that after high number of iterations the reconstruction algorithm with full physical modelling in the BP which therefore uses matched projector pairs can reach better activity estimate, than the one with unmatched projector pair resulted by simplified modelling which therefore only approximates the solution even in the noiseless case [1, 2, 5]. Figure 9 shows convergence of the solution through iterations in the 1D model. $L_2 - norms$ of the distance of the solution from phantom distribution are plotted against the number of iterations. Figure 9 on the left represents the noiseless test case while on the right captures very similar characteristics but in a simulated (noise containing) reconstruction. It can be seen that the algorithm with simplified back projection approximates the optimum faster but converges to another fix point afterwards due to unmatched projector pair (probably this is the reason of the accumulating activity in the corner of the reconstructed images when simplified modelling is used in the back projection, discussed later). Contrarily, the initial convergence speed with faithful modelling is lower but after numerous iteration it can achieve at least the same $L_2 - norm$, but does converge to the desired fix point. This means that the optimal stopping criterion is much more relaxed with faithful modelling in the BP because the reconstruction gives the optimal result in a wide interval of iteration number. This result resolves the contradiction as additional information indeed leads to better image reconstruction (as finding an optimal stopping criterion is even nowadays an active research area) thus the perceived anomaly was only apparent.

These advantageous properties imply that the faithful modelling in the back projection is worth the effort only the convergence of the iteration scheme has to be accelerated. If it is successfully done, a robust reconstruction algorithm is resulted which is more stable to noise with a much more relaxed stopping criterion compared to the algorithm with simplified modelling in the back projection. Also,

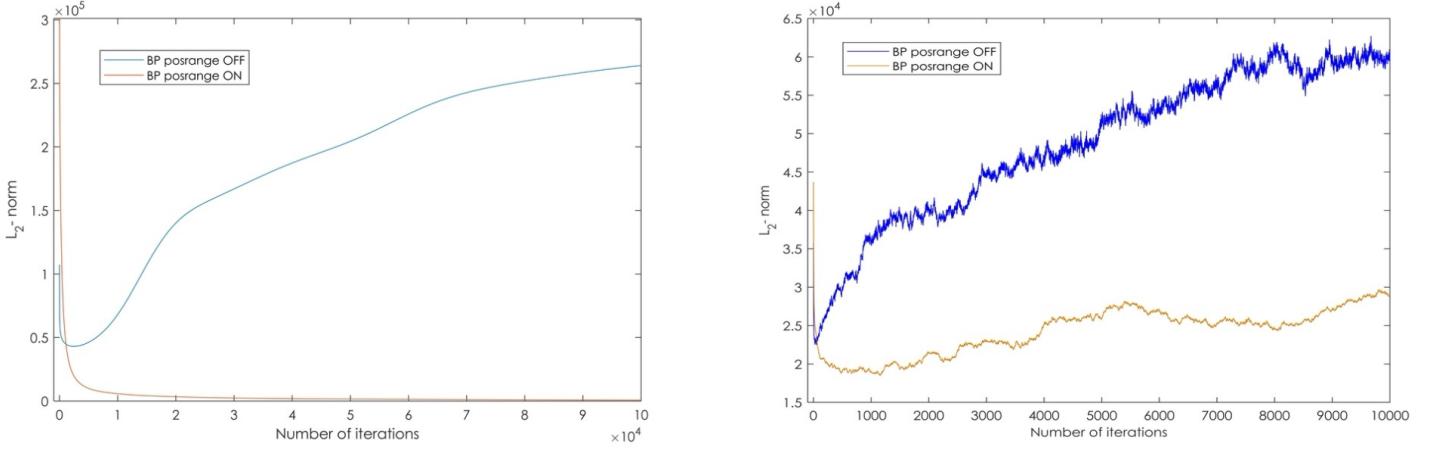


Figure 9: L_2 – norm curves for noiseless (on the left) and MC simulated (on the right) reconstructions comparing the convergence path of the settings using faithful and simplified modelling in the back projection

the activity accumulation in the image corners is avoided. In the next subsection such a possible solution for convergence acceleration is presented.

2.8 Proposed improvement

According to the discussion of the previous subsections, simplification of the MC modelling in the BP just luckily affects the behaviour of the algorithm from a mathematical point of view. However this form is not the ideal BP operator but the one that is easy to implement without much modification to the original algorithm. To obtain a better BP operator the previously listed advantages can be amplified with a posteriori manipulation and a better form can be reached.

In order to combine the positive features of both simplified and faithfully modelling BP settings, namely the accelerated convergence and proper fix points respectively the only possible degree of freedom for the modification is to further increase the singular values of the BP operator using full physical modelling, similarly to the accompanying effect of the simplified modelling. In this case U and V matrices of the SVD are unchanged. The possible fix points of the algorithm can be obtained from the next equations (ratio is in a Hadamard sense) as the update process multiplies (also in Hadamard sense) the current estimate by 1 when the following is true:

$$\frac{A^T y_r}{|A|} = \frac{A^T y_r}{A^T 1_{LoR}} = 1_{vox} \quad (2.8.1)$$

Rearranging:

$$A^T(y_r - 1_{LoR}) = A^T \tilde{y} = 0_{vox} \quad (2.8.2)$$

Using the dyadic definition of SVD:

$$A^T \tilde{y} = \left(\sum_{r=1}^{\text{rank}(A)} \sigma_r v_r u_r^T \right) \tilde{y} \quad (2.8.3)$$

v_r and u_r are the columns of matrix V and U respectively. As V is an orthogonal matrix the linear combination above equals to zero precisely when every $\sigma_r u_r^T \tilde{y}$ coefficient equals to zero. The singular values are all nonzero thus the $u_r^T \tilde{y}$ dot product equals to zero for $\forall r$. This implies that $U^T \tilde{y} = 0_{\text{vox}}$. Matrix U remains the same with singular value modification so the possible fix points are unchanged.

The easiest way to modify the singular value spectrum of the BP system matrix is the application of a matrix of the form:

$$B = VD^*V^T \quad (2.8.4)$$

D^* is the matrix with which D^*D has the desired form, i.e. the desired singular value spectrum in the diagonal. Applying the SVD filter to the ML-EM algorithm (in matrix form) results in the following formula ($B^T = B$ as being symmetric and $V^T V = I$ as V is orthogonal. I is the identity matrix. The ratio and the multiplication in the update process of $x^{(n)}$ is in Hadamard sense):

$$x^{(n+1)} = x^{(n)} \frac{BA^T y_r}{A^T B 1_{LoR}} = x^{(n)} \frac{VD^* V^T V D U^T y_r}{V D^* V^T V D^T U^T 1_{LoR}} = x^{(n)} \frac{V D^* D U^T y_r}{V D^{*T} D^T U^T 1_{LoR}} \quad (2.8.5)$$

In noiseless (test) case, i.e. when there is no noise added to simulations, this form is (the scalar multiple of) the identity matrix, in agreement with the convergence analysis [5, 20, 22], as the singular values are clustered together as far as possible. The SVD filter fastens the convergence with two orders of magnitude [18]. Illustratively, Figure 10 shows the $L_2 - \text{norm}$ curves of a noiseless reconstruction of the mathematical phantom using the SVD filter in the 1D model. The $L_2 - \text{norm}$ curve of the unmodified ML-EM reconstruction is also plotted for comparison.

However, the SVD filter cannot be applied straightforward for the real, noisy case. The measurement process equals to $Ax = UDV^T x$ where x stands for the activity distribution. Thus, the measurement attenuates its frequency components according to the singular value spectrum (multiplication with matrix D) and adds some noise to the result. As so, only those components which fit to the discrete Picard condition can be amplified.

The resulted method was tested in the 1D model and gave promising results [19]. Figure 11 shows a representative reconstruction result comparing different back projection settings with respect to $L_2 - \text{norm}$. The mathematical phantom of the 1D model (see figure 5) is reconstructed with faithful modelling, simplified modelling and SVD filtered faithful modelling in the back projection, $L_2 - \text{norms}$ after a given number of iteration are plotted.

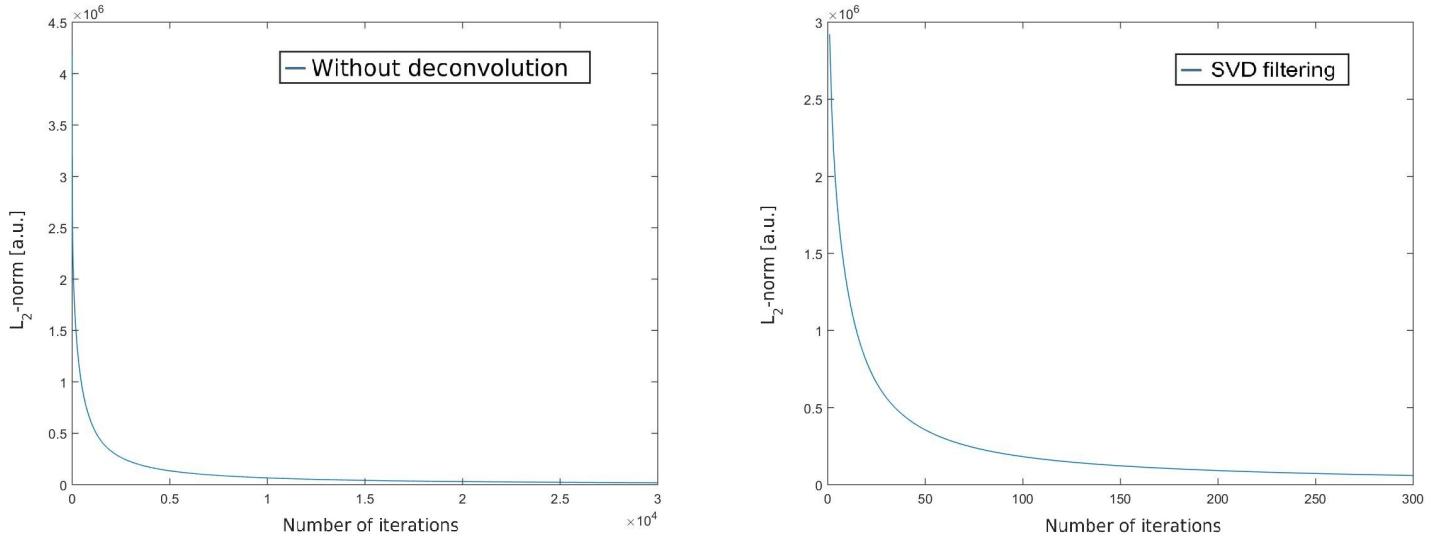


Figure 10: In both reconstructions full physical modelling was used. Comparison of the unmodified ML-EM algorithm (on the left) and its SVD filtered version (on the right). Important to highlight the different range of the $x - \text{axes}$. Noise is neglected at this point, the reconstructions aim at verifying the result of the convergence analysis

Despite the promising reconstruction results in the 1D model, there were some problems with the full system implementation, therefore the applicability of the SVD filter in the proposed form was very limited. B matrix is of a size $N_{\text{voxel}} \times N_{\text{voxel}}$ which is in the order of terabyte with respect to variable size. Generally, it is impossible to efficiently store such a matrix directly for the full system on one hand, and even more difficult to calculate on the other hand, because the method requires the SVD of the system matrix, which is even several folds bigger than matrix B .

In the following sections a possible remedy for this problem is presented which opens the way towards practical implementation and everyday use of the proposed modification. The performance of the resulted method is evaluated on the full system by means of several simulated phantoms and a measurement as well.

3 Practical implementation of the SVD filter

This section describes the main theoretical contribution of thesis. The SVD filter has proven to be very efficient but its capacities could not be exploited for a real size system due to computational issues. This problem is solved with the theoretical considerations detailed in the following.

The effect of the multiplication of matrix $B = VD^*V^T$ with any voxel space vector equals to

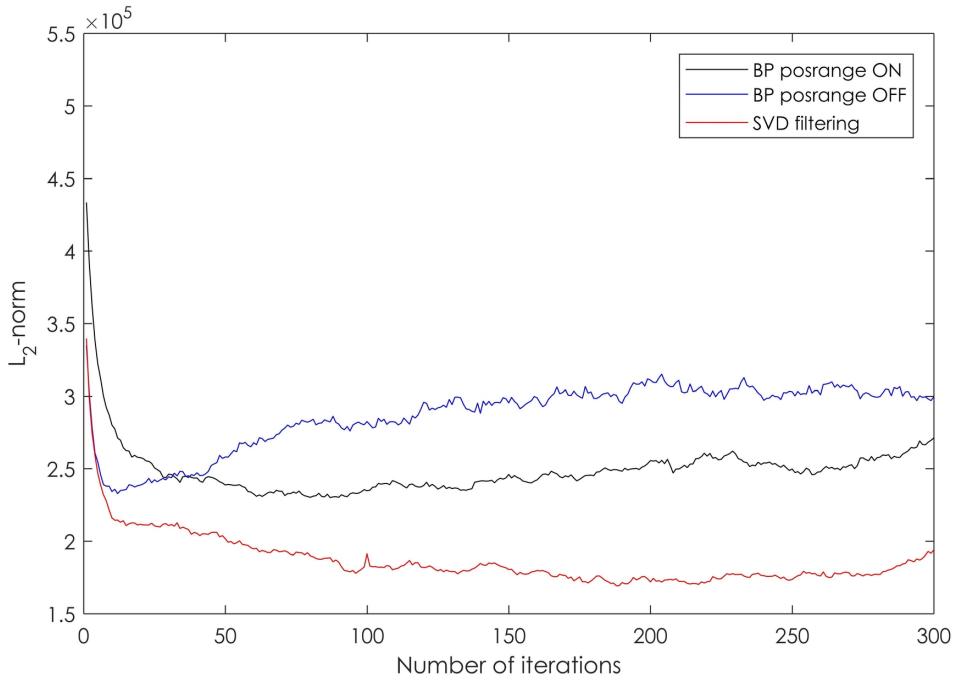


Figure 11: The L2-norm of the difference between the activity distribution and the current estimate after a given number of iterations. Smaller value means better agreement. Reconstruction with SVD filter outperforms the best setting so far in terms of faster initial convergence and the farther starting point of increasing discrepancy due to semi-convergence and unmatched fix point resulting in more relaxed stopping criterion. Also the faster initial convergence of positron range neglecting back projection can be seen compared to positron range modelling back projection (without filtering) in agreement with the convergence analysis

the following: unfolding the vector in the eigenbasis of the voxel space, multiplying the resulted basis coefficients pointwise with the eigenvalues of B (which equal to the diagonal entries of D^*) and transforming back the result into the voxel space. The multiplication with the eigenvalues in the spectral domain counteract the decay of the singular values of the system matrix i.e. the blurring of the corresponding compact operator.

The process is analogous to conventional deconvolution except for the used eigenbase. Conventional deconvolution implies Fourier basis which is the eigenbase precisely when the operator composition of a consecutive forward projection - back projection can be discretised as a circular matrix i.e. $A^T A = C$ where C is a $N_{voxel} \times N_{voxel}$ circular matrix. Despite this is rarely the case for a general setup a reasonable approximation can be made. Observing the structure of matrix B (see figure 12) it happens to be approximately circulant which may have been expected from its deconvolution-like acting. The discrepancy (compared to circulant form) comes from the position dependence of the

latent (de)convolution. Position dependence arises due to material map, scanner geometry and the proximity of the edge of the voxel space where the neighbourhood of a given voxel is different from that of a more central voxel. The eigenbase of a given system inherently accounts for these effects, hence the structure of matrix B .

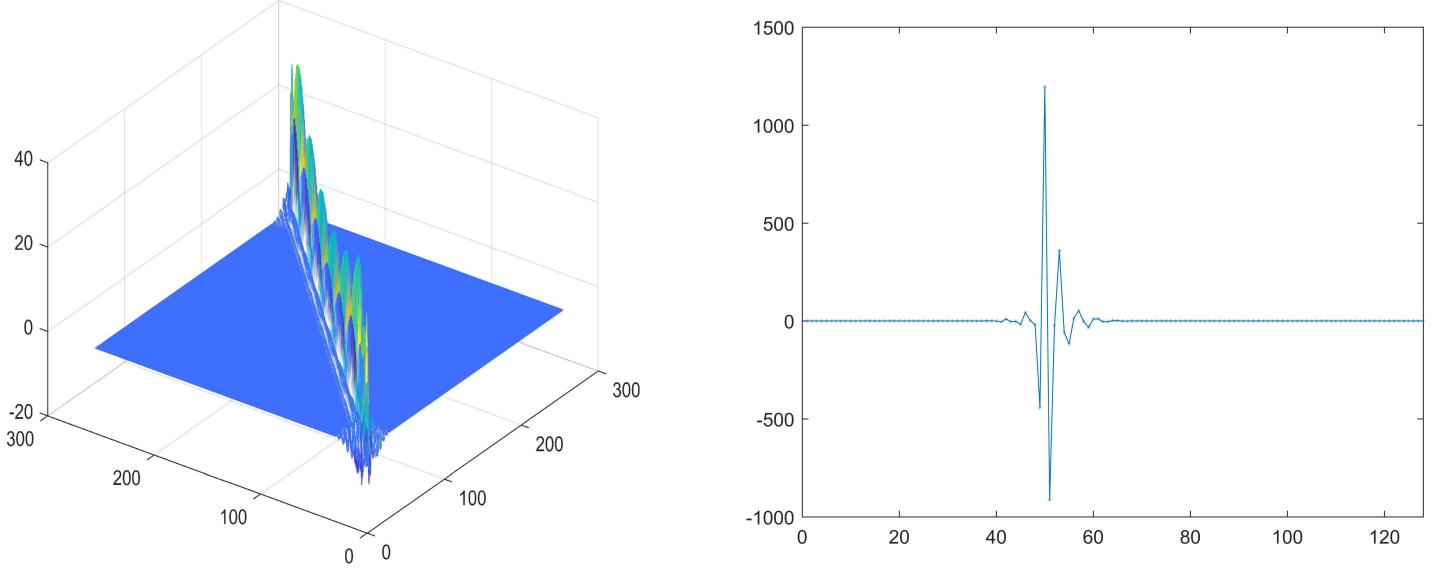


Figure 12: Mesh plot of matrix B and a characteristic (50^{th}) row of the matrix for the 1D model

In conclusion, a convolution approximation which can account for position dependence is the voxel space is a strong candidate for the replacement of the multiplication with matrix B in order to avoid computational difficulties and therefore can worth a consideration. The details of such an approximation are presented in the following subsections, with the corresponding concepts and notations.

3.1 Deconvolution

Unwanted convolution is an inherent problem in transferring analogue information. Deconvolution is an algorithm-based process used to reverse this effect on the recorded data. In general, the objective of deconvolution is to find the solution of a convolution equation of the form:

$$f * g = h \quad (3.1.1)$$

h is the recorded signal, f is the data to be recovered and g is the convolution kernel of the corrupting process. Its inversion is simple in the Fourier domain using the convolution theorem:

$$f = \mathbb{F}^{-1} \left\{ \frac{\mathbb{F}\{h\}}{\mathbb{F}\{g\}} \right\} \quad (3.1.2)$$

However, in physical measurements, the situation is usually closer to:

$$(f * g) + \epsilon = h \quad (3.1.3)$$

ϵ represents the unavoidable noise. Consequently, the inverse problem is ill conditioned and regularisation is required, but the exact data is still not hoped to be recovered. The quality of the recovered approximate solution highly depends on the noise level of the measurement and the applied regulariser. In the point of view of the applied algorithm deconvolution processes can be classified into two groups, direct and iterative methods. Roughly speaking, direct methods use a regularisation term in the Fourier inversion, iterative methods are controlled through the number of iterations as a regularisation parameter. Being a complete analogue of ML-EM reconstruction (for Poisson noise), Richardson-Lucy (abbreviated as RL from now on) algorithm can be an obvious candidate for the deconvolution. With normalised convolution kernel the formula looks like as follows (the ratio and the multiplication is in Hadamard sense, \bar{g} means the complex conjugate of g):

$$f^{(n+1)} = f^{(n)} \left(g^T * \frac{h}{x^{(n)} * g} \right) \quad (3.1.4)$$

Or, in the computationally more efficient Fourier form:

$$f^{(n+1)} = f^{(n)} \left(\mathbb{F}^{-1} \left\{ \overline{\mathbb{F}\{g\}} \mathbb{F} \left\{ \frac{h}{\mathbb{F}^{-1}\{\mathbb{F}\{f^{(n)}\} \mathbb{F}\{g\}} \} \right\} \right\} \right) \quad (3.1.5)$$

The choice of the RL method is implied by the following considerations: efficiently deconvolves Gaussian blurring in presence of Poisson noise which is the case in the current setup and location dependence can be more easily incorporated being iterative method which contains only forward steps i.e. convolutions.

As the efficient algorithm is given the remaining task for the deconvolution is to find the convolution kernel. In this current problem the undesired convolution is the blurring of the back-projecting operator which is unavoidable and location dependent as well. Suppose, that there is an ideal back-projecting operator which back projects a given sinogram space signal to the voxel space ideally i.e. without any blurring. Then the latent convolution kernel blurs this resulted voxel space distribution so that it coincides with the result of the Monte Carlo simulated back projection.

According to the SVD of the system matrix and the one to one correspondence of the left and right i.e. sinogram and voxel space singular vectors the blurring of a consecutive forward projection - back projection the contribution to the resulted blurring is equal for both projections:

$$A = UDV^T \quad (3.1.6)$$

$$A^T = V^T D^T U = V^T D U \quad (3.1.7)$$

$$u_i = \frac{Av_i}{\sigma_i}; v_i = \frac{A^T u_i}{\sigma_i} \quad (3.1.8)$$

If $\sigma_i = 1 \forall i$ a voxel space distribution is forward projected without any blurring to the sinogram space and also the sinogram distribution is back projected into the initial distribution. In this imaginary situation the image of a point source is the point source itself. However, this is never the case. The forward projection weights the spectral coefficients of any voxel space distribution with the diagonal entries of matrix D before mapping it to the sinogram space, and similarly the back projection also weights the spectral coefficients of any voxel space distribution but after it is mapped to the voxel space from the sinogram space. This weighting is responsible for the blurring of the operator and is modelled as a convolution in the voxel space with a given kernel. From now on this kernel is denoted by g .

According to the convergence analysis only the blurring caused by the back projection has to be deconvolved. Therefore, an advantage of the symmetric blurring contribution model is that only the result of the back projection has to be modified which can be performed solely in the voxel space where the physical effect are well known and there is no need for deconvolution in the sinogram space.

This symmetry property of the system matrix (and the corresponding operator) also significantly facilitates the calculation of g . The image of a point source (after forward projection-back projection) is easy to simulate (measured data cannot be used for the point spread function (PSF) determination because the PSF of the back projection is in demand. Back projection is always MC simulated, therefore simulation is required) and with the notations and considerations above can be modelled as follows (noise is neglected at this point):

$$h = A^T Ax = VD^2V^T x := g * g * x \quad (3.1.9)$$

From now on $g * g$ is referred to as the PSF of the system. Once the PSF is known from either simulation or measurement g is hoped to be found as well. The deconvolution is then performed with kernel g and this process well approximates the effect of the multiplication with matrix B , therefore the SVD-filter becomes possible to apply on a real size system which is a big step towards practical implementation and usability.

The remaining task after the model formation is the collection of the physical effects of PET imaging which cause blurring giving contribution to the PSF of the system. Then, a suitable smooth function is needed to fit on the simulated PSF in order to avoid noise amplification in any further calculation. Once the fitting is successful and a proper smooth representation of the PSF is given, kernel g can be computed for the deconvolution.

3.2 Modelling of the PSF

The PSF of the system can be obtained from point source simulation. As it was highlighted in the previous subsection, simulation is used in every PSF calculation of this thesis and is always recommended if the proposed method is used. Despite its simplicity the resulted PSF is always contaminated by statistical fluctuations therefore the deconvolution becomes unstable. To overcome this difficulty the PSF must be fitted with analytical functions. In the following the relevant physical effects and their analytical fit functions to model the resulted blurring (in the voxel space) are itemized.

- Positron range: the density function of the positron free path can be fitted with biexponential curve, the direction distribution is isotropic [23]
- Acollinearity: the voxel space blurring caused by acollinearity can be well approximated by a Gaussian [23]
- Scatter in the detector and the variance of penetration depth: Gaussian fit is satisfactory [23]
- Scatter in the imaged volume: the fit of the sum of multiple Gaussians is required [24]
- Geometric coverage of the voxel space by the LoRs: Gaussian [23]

The consecutive convolution with the individual effects results in the convolution with the PSF of the system. As the convolution of Gaussian is also Gaussian the PSF itself can be well approximated with the sum of multiple Gaussians. Multiple different Gaussians are required to modell scattering in the voxel space and also to approximate the biexponential characteristics of the positron range. In the reconstructions of this thesis the sum has three terms (regardless of the offset) as the sum of three Gaussians has proven to fit the PSF efficiently.

The fitting process is one of the key points of the method because the agreement of the resulted PSF (and kernel g) with the corresponding real function has the highest impact on the performance of the proposed algorithm. Therefore, the fit function has to be as close as possible to the real latent function coming from the physics of the system and also the fitting procedure itself has several constraints in order to let the deconvolution perform properly.

3.3 Fitting the PSF

In order to fit the PSF one iteration step with PANNI code was performed. The voxel space was homogeneously filled by the given material (simulation consisted of water, bone and plexi glass) and an ^{15}O point source (δ_i) was placed in the middle of the voxel space i.e. in the middle of the simulated

nanoPET geometry (identified as the i^{th} voxel with 1D list ordering). 2×10^{12} positron was used to simulate the "measurement" and in the back projection as well. As the first iteration only back projects the "measured" dataset which is in this case the (simulation) the forward projection of the point source the result is no other than the PSF:

$$A^T A \delta_i = g * g * \delta_i = PSF \quad (3.3.1)$$

Next, the sum of three Gaussians was fitted to the PSF. To facilitate the fitting procedure, the separability of the Gaussian kernel has been exploited. A multidimensional Gaussian is separable i.e. it can be constructed as the direct product of 1D Gaussians [24]. Therefore, instead of fitting a sum of three 3D Gaussians onto the PSF directly which would be hardly feasible and unsatisfactory with respect to precision (according to the efforts and attempts of the author), the sum of three 1D Gaussians was fitted onto the line profiles of the PSF along all the three main axes: x - transverse, y - sagittal and z - coronal (taking into account the circular symmetry of the PET scanner the transverse and sagittal profiles are the same). The result of the fit in water filled voxel space is presented for illustration:

- Gaussian: $G_i(x) = h_0 + A_x \exp^{-\left(\frac{x-x_0}{width}\right)^2}$

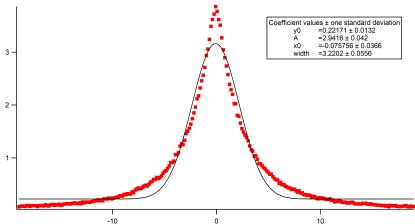


Figure 13: Transverse profile

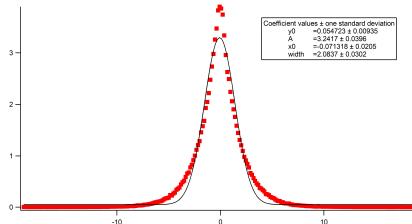


Figure 14: Coronal profile

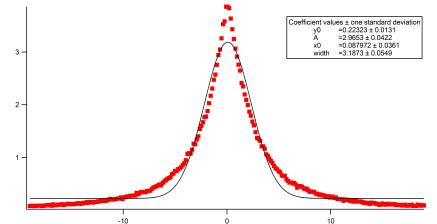


Figure 15: Sagittal profile

- Sum of three Gaussians: $G_i(x) = h_0 + A_{1,x} \exp^{-\left(\frac{x-x_{01}}{width_1}\right)^2} + A_{2,x} \exp^{-\left(\frac{x-x_{02}}{width_2}\right)^2} + A_{3,x} \exp^{-\left(\frac{x-x_{03}}{width_3}\right)^2}$

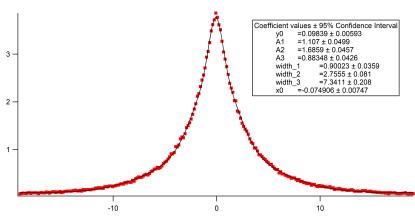


Figure 16: Transverse profile

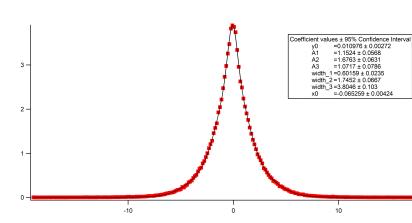


Figure 17: Coronal profile

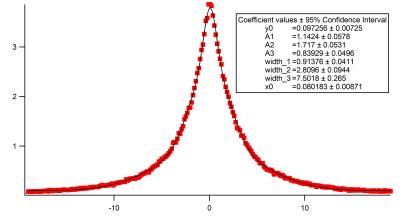


Figure 18: Sagittal profile

Once the 1D fit was performed three 3D Gaussians were constructed from the three terms of the fitted 1D Gaussians. Finally, the three 3D Gaussians were added, hence the PSF (Figure 20,23). In mathematical form the procedure looks as follows:

$$PSF = h_0 + \sum_{i=1}^3 G_{i,transverse} \otimes G_{i,sagittal} \otimes G_{i,coronal} = h_0 + \sum_{i=1}^3 A_{i,x} A_{i,y} A_{i,z} e^{-(\frac{x-x_{0,i}}{width_{i,x}})^2 - (\frac{y-y_{0,i}}{width_{i,y}})^2 - (\frac{z-z_{0,i}}{width_{i,z}})^2} \quad (3.3.2)$$

1D fitting is relatively easy, the only constraint is the equality of the relative weights of the individual Gaussians in the three term sum for all the three directions, i.e $A_{i,x} = A_{i,y} = A_{i,z} = A_i \forall i$, because A_i weights the i^{th} 3D Gaussians in the three term sum, hence there is no degree of freedom for $A_{i,j}$'s to be different for given i and different j (this latent degree of freedom comes from the 1D fitting approach).

This requirement was usually very hard to met. The fit of the coronal and transverse (/sagittal) profiles resulted in quite different A_i weights, so the $A_{i,x} = A_{i,y} = A_{i,z} = A_i \forall i$ constraint was successfully met only after several iteration of the individual fits when A_i was manually constrained and fine tuned in the range of the resulted $A_{i,j}$'s trying to obtain the best fit for the individual profiles using the same A_i for every profile (and leaving only the FWHM of the directions in the 3D Gaussians as a free parameter for the fit). However, the result was still not satisfactorily precise in some cases. The proper fit of the PSF is a remaining challenge and it could significantly improve the performance of the proposed algorithm.

From the line profile fit, it would be possible to simply compose the dot product of the three 1D profile, i.e the fitted sum of 1D Gaussians.

$$PSF = h_0 + \sum_{i=1}^3 G_{i,transverse} \otimes \sum_{i=1}^3 G_{i,sagittal} \otimes \sum_{i=1}^3 G_{i,coronal} \quad (3.3.3)$$

Despite this approach is much more easier as the coefficient constraint is avoided, it gives an erroneous result with respect to the PSF (Figure 21,24). However, it still has proven quite efficient in most of the reconstructions.

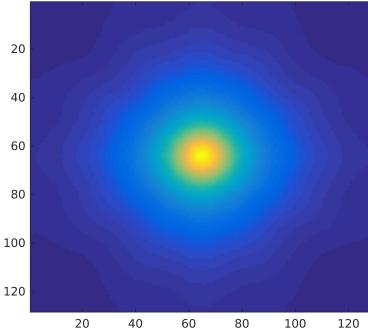


Figure 19: Measured PSF in water filled voxel space, $z = 0$ cross section. C_{12} symmetry due to dodecagon detector panel can be seen

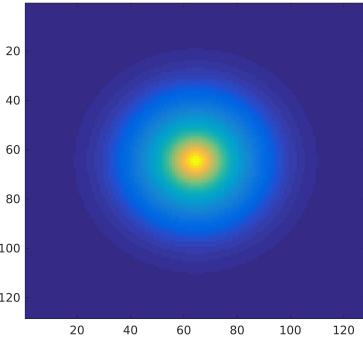


Figure 20: Sum of three 3D Gaussians composed each from the direct product of the corresponding terms in the fitted 1D profiles, $z = 0$ cross section

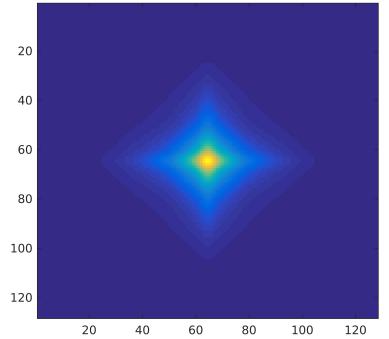


Figure 21: Direct product from the fitted sum of 1D Gaussians, $z = 0$ cross section

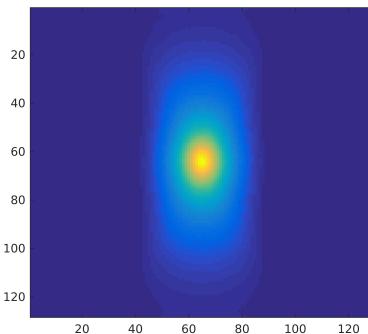


Figure 22: Measured PSF in water filled voxel space, $y = 0$ cross section.

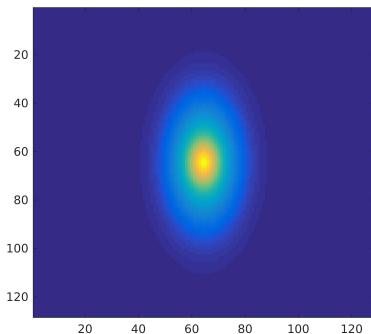


Figure 23: Sum of three 3D Gaussians composed each from the direct product of the corresponding terms in the fitted 1D profiles, $y = 0$ cross section

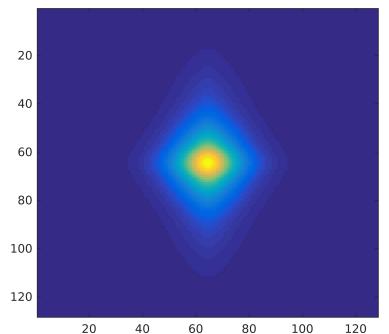


Figure 24: Direct product from the fitted sum of 1D Gaussians, $y = 0$ cross section

Once the PSF is fitted properly the remaining task for the deconvolution is the calculation of kernel g from the relation $PSF = g * g$. This implies a square root approach when g is calculated as the square root of the PSF.

3.4 Square root of the PSF

In the previous subsection the calculation of the PSF was presented. However, in order to reproduce the effect of the multiplication with matrix B the convolution kernel g is required. In the following the 1D case is used for sake of simplicity for the derivation but the separability of the kernel implies the generalisability. Using the matrix notation of SVD and the convolution model for the forward projection - back projection composition the following is true:

$$C_{PSF} = C_{g*g} = C_g^2 \approx VD^2V^T \quad (3.4.1)$$

C_{PSF} and C_g is the matrix belonging to the convolution with the PSF and g respectively. To get matrix C_g the square root of C_{PSF} is needed. In general, an $n \times n$ matrix with n distinct nonzero eigenvalues has 2^n square roots. However, as the VD^2V^T is a positive definite matrix it has precisely one positive definite square root called the principal square root [25] which equals to VDV^T as all the singular values (diagonal entries of matrix D) of the system matrix are positive. The blurring interpretation of the matrix composition also implies the positive definiteness. Similarly, it can be shown that the eigenvalues of C_g are real, C_g being the circulant matrix corresponding to the vector which contains convolution kernel g : $[g_0, \dots, g_n]$. As g is supposed to be symmetric its Fourier-transform $\mathbb{F}\{g\}$ is real symmetric [25] which is no other than the eigenvalue spectrum of C_g :

$$C_g = F_n^{-1} \text{diag}(\mathbb{F}\{g\}) F_n \quad (3.4.2)$$

F_n is the n -dimensional discrete Fourier-transform. The same holds for matrix C_{PSF} . This implies that the calculation of the suitable square root is as follows:

- Simulation of the PSF
- Fitting the sum of three Gaussians to the simulated PSF
- Fourier-transform the result
- Take the absolute value of the Fourier spectrum (in order to avoid imaginary square root) and take its square root

- Use the square root in the Fourier formulation of the forward-problem type RL deconvolution to deconvolve kernel g of both the nominator and the denominator in the back projection of ML-EM formula (multiplications and the ratio is in Hadamard sense):

$$(de)nominator^{(k+1)} = (de)nominator^{(k)} \left(\mathbb{F}^{-1} \left\{ \mathbb{F}\{\bar{g}\} \mathbb{F} \left\{ \frac{h}{\mathbb{F}^{-1}\{\mathbb{F}\{(de)nominator^{(k)}\} \mathbb{F}\{g\}} \right\} \right\} \right) \quad (3.4.3)$$

$k \in \{1, \dots, K\}$ indexes the number of iterations, K is the maximum number of iterations which serves as a regularisation parameter.

Hence the modified ML-EM formula:

$$x^{(n+1)} = x^{(n)} \frac{BA^T y_r}{A^T B 1_{LoR}} = x^{(n)} \frac{\mathbb{D}_{\bar{g}}\{A^T y_r\}}{\mathbb{D}_{\bar{g}}\{A^T 1_{LoR}\}} \quad (3.4.4)$$

$\mathbb{D}_{\bar{g}}$ denotes deconvolution with respect to kernel g .

At one point of the derivation special care shall be taken. The eigenvalue spectrum of matrix C_{PSF} i.e. the Fourier-transform of the PSF is not guaranteed to be positive (but it guaranteed to be real) [25], so any sign change caused by taking the absolute value would spoil the $PSF = g * g$ square root model of the PSF, i.e. that the convolution with the PSF can be written as a double convolution with the kernel g . However, the derivation above is not restrictive thanks to the special form of the PSF ($PSF = G_1 + G_2 + G_3$ where G_i is Gaussian $\forall i$). Due to the linearity of the Fourier-transform and the special invariance of the Gaussian function, the Fourier-transform of the sum of Gaussians is real and positive, therefore taking the absolute value makes no change:

$$\mathbb{F}\{PSF\} = \mathbb{F}\{G_1 + G_2 + G_3\} = \mathbb{F}\{G_1\} + \mathbb{F}\{G_2\} + \mathbb{F}\{G_3\} \quad (3.4.5)$$

The Fourier-transform of a Gaussian is again a Gaussian:

$$G_i(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \quad (3.4.6)$$

Differentiating:

$$\frac{dG_i(x)}{dx} = -\frac{x}{\sigma^2} G_i(x) \quad (3.4.7)$$

Applying Fourier-Transform on both sides (\tilde{G}_i is the Fourier-transform of G_i):

$$i\omega \tilde{G}_i(\omega) = \frac{1}{i\sigma^2} \frac{d\tilde{G}_i(\omega)}{d\omega} \quad (3.4.8)$$

Rearranging:

$$\frac{d\tilde{G}_i(\omega)}{\tilde{G}_i(\omega)} = -\omega\sigma^2 d\omega \quad (3.4.9)$$

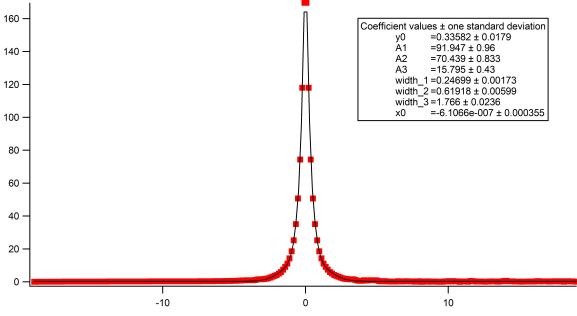


Figure 25: Sum of 3 Gaussians fitted to the Fourier-transform of the transversal line profile of the PSF (water)

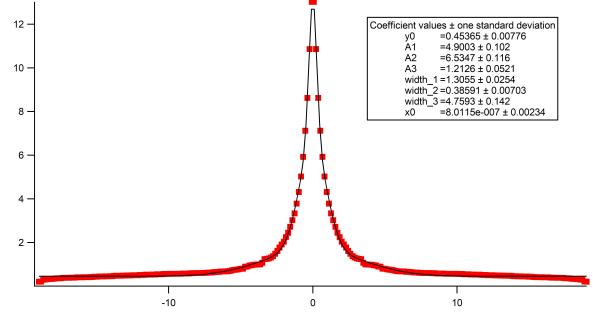


Figure 26: Sum of 3 Gaussians fitted to the square root of the Fourier-transform of the transversal line profile of the PSF (water)

Integrating the differential equation from 0 to ω :

$$\ln \tilde{G}_i(\omega) - \ln \tilde{G}_i(0) = \frac{\sigma^2 \omega^2}{2} \quad (3.4.10)$$

$\tilde{G}_i(0) = 1$ as $G_i(x)$ is normalised to 1, therefore $\ln \tilde{G}_i(0) = 0$. After exponentiating:

$$\tilde{G}_i(\omega) = e^{-\frac{\sigma^2 \omega^2}{2}} \quad (3.4.11)$$

This implies that the Fourier-transform of the PSF and hence the eigenvalue spectrum of C_{PSF} is positive valued, so there is no contradiction in the derivation of the convolution kernel g . g is the principal square root of the PSF, consequently $\mathbb{F}\{g\}$ is also positive valued, and using the Taylor expansion of the square root and the multiplication invariance of Gaussians [26], g is also well approximated with the linear combination of Gaussians in good agreement with the physical intuition:

For all these derivations to be true the special invariance property of the 3D PSF fit function has to be highlighted.

4 Reconstruction results

The applicability of the convolution model to matrix B opens the way for the SVD filter towards practical implementation. The PSF can be stored in an array which has the same size as the voxel space, so means no difficulty at all. RL deconvolution is easy to implement, the convolution steps can be replaced by FFTs and run on GPU for faster computation. The modified ML-EM algorithm was tested with PANNI on several simulations. The deconvolved ML-EM algorithm uses full physical modelling in the back projection in all cases (If deconvolution is not used, this setting is denoted as "without deconvolution"). The reconstructions used the cylinder ring phantom for comparison with respect to inhomogeneity artefact, a simulated Derenzo phantom and a measured Derenzo phantom.

4.1 Homogeneous cylinder ring phantom

The reconstruction of the homogeneous cylinder ring phantom observes the presence of the inhomogeneity artefact with the deconvolved back projection. $128 \times 128 \times 128$ voxel space, 0.298mm sided voxels and 2×10^9 positron used in each of the 80 iterations. 120 RL iteration is performed in every ML-EM iteration. The input data is the simulation of an 1000s long measurement of ^{15}O isotope, the result can be seen in Figure 27. Figure 28 shows the line profile of the voxel space summed along the $z - axis$.

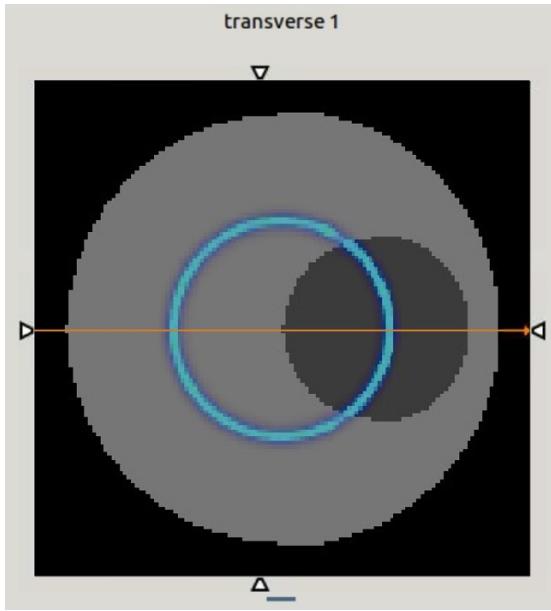


Figure 27: The result of the reconstruction after 80 iterations, voxel space is summed along the z -axis

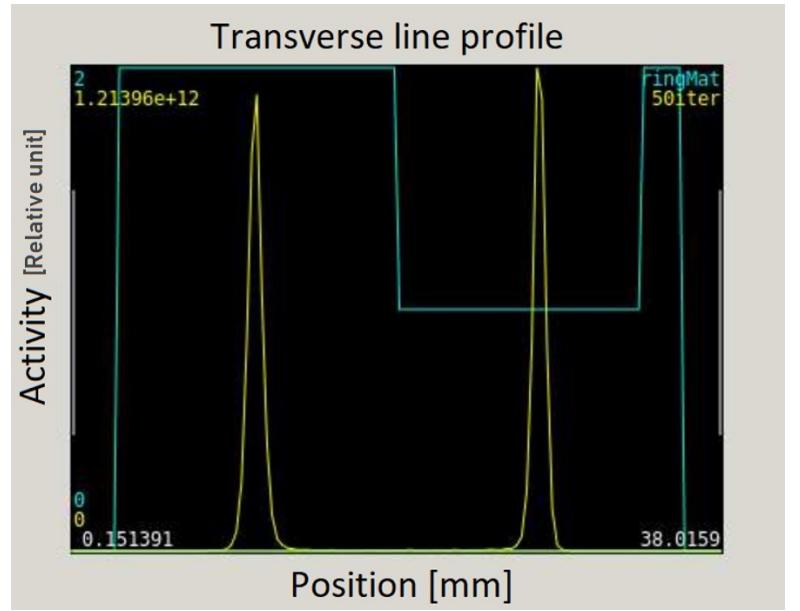


Figure 28: Line profile of the reconstructed activity along the orange line in figure 27. Identical FWHM and activity estimate can be seen regardless of the material map

Deconvolved back projection successfully makes the artefact disappear and the peak activity estimates are identical (up to the noise level) in bone and water area. The only remaining artefact is the "double step" in the estimated activity on the bone-water material border along the ring. This artefact also appears in Figure 3 when full physical modelling is used in the back projection and is not present in Figure 4 when simplified modelling is used. This suggests, that the deconvolution is not perfect with respect material map induced position dependence incorporation, and therefore is unable to make this "double step" disappear. The limitations of the homogeneous PSF approximation and possible solutions are discussed later.

However, quantitative results are more promising. Figure 29 shows the $L_2 - norm$ of the difference of the reconstructed estimate and the phantom distribution with respect to the number of iterations. The deconvolved ML-EM algorithm with full physical modelling achieves faster initial convergence and

much better activity estimate in every iteration. Also, the initial faster convergence of the simplified back projection can be seen compared to full physical modelling (without deconvolution) similarly to the result obtained from the analysis with the 1D model. No additional filtering was used in the deconvolved reconstruction, the reconstruction with simplified modelling in the back projection used trimmed median filtering to be able to give any result. This suggest that the original noise stability of full physical modelling is preserved to high extent.

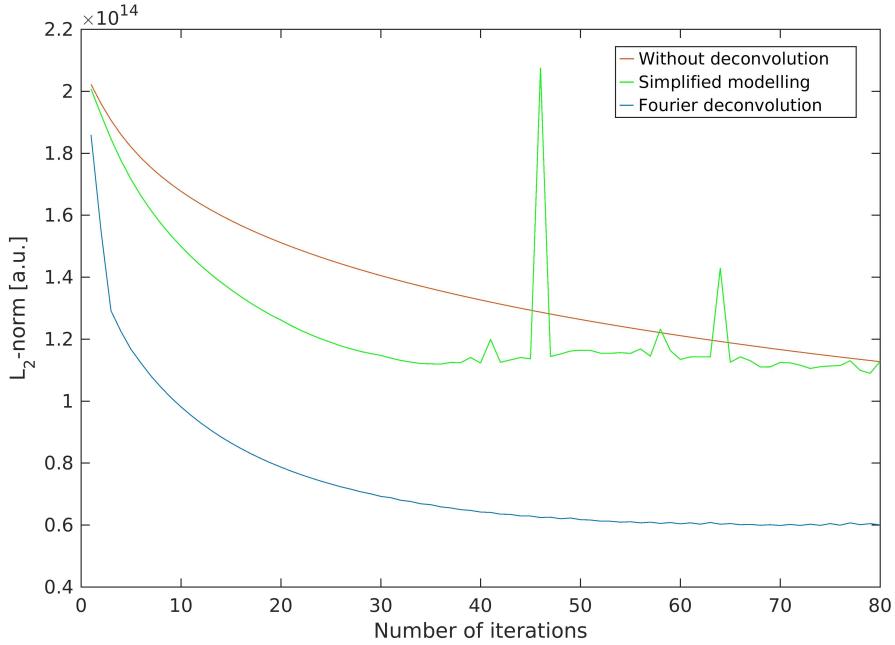


Figure 29: The $L_2 - \text{norm}$ of the deviation of the activity estimate from the ideal through iterations. Lower value means better agreement and usually better reconstruction. However, this metric gives no information about the contrast. After 80 iterations the simplified modelling and the full physical modelling without deconvolution achieves the same $L_2 - \text{norm}$, but the former still shows the inhomogeneity artefact and hence the estimate is far from the ideal, while the latter estimate is far from ideal due to noise build up.

Based on the test simulation with the cylinder ring phantom it can be concluded that deconvolution approximation of the SVD filtering method successfully reduces the perceived inhomogeneity artefact and at the same time increases the overall speed of convergence of the full physical modelling back projection significantly above that of the simplified one. However, incorporating the position dependence of the PSF still there remained to solve. Even in the case of a relatively simple phantom and material map the algorithm was unable to remove a significant artefact arose from the presence of a material border. In a more complex situation like a Derenzo phantom or a real (pre)clinical mea-

surement more significant deteriorations are expected. So, it is important to itemize the main factors which cause position dependence and discuss the possibility of any incorporation to the deconvolution model. These factors are analysed in the following with a proposed solution as well.

5 Position dependence of the PSF

So far the PSF used for the deconvolution process was simulated by means of a point source placed in the middle of the scanner geometry. However, it is a well known result of PET imaging that the PSF of the system is not homogeneous but position dependent within the scanner. This section analyses the effect of the central PSF approximation in the deconvolution and discusses the possible remedies and performance improvements.

5.1 Position dependence in due to scanner geometry

Dodecagon scanner geometry can be well approximated by a circle so rotational symmetry in the $x - y$ plane is a reasonable assumption. Therefore, the position dependence of the PSF is significant in radial direction and along the $z - axis$. For testing, the voxel space is still $128 \times 128 \times 128$ but the linear voxel size is doubled: $0.596mm$ so that the outer voxels are located outside of the detector ring in the $z - direction$ and closer to the detector ring in radial direction. The input data is the simulation of a ^{15}O measurement and the iteration uses 2×10^{12} positron in agreement with previous PSF simulations. First, the point source is placed in the $(48, 0, 0)$ position, i.e. located on the $x - axis$ in the 48^{th} voxel $24,885mm$ apart from the center. The PSF can be seen in Figures 30,31. The distortion comes from rather edge effects i.e. the fact that there is no voxel after the 64^{th} coordinate thus the neighbourhood of the voxels located at the edges of the voxel space is different from that of the more centered ones. The similarity with the central PSF (PSF_0) is measured by means of cross-correlation norm (CC): $CC(PSF, PSF_0) = 100 \left(1 - \left| \frac{PSF^T PSF_0}{\sqrt{\|PSF\|_2 \|PSF_0\|_2}} \right| \right)$. PSF and PSF_0 denote the vector of 1D list ordering of the 3D PSFs. In this case, the similarity is 66,37%.

Second, the point source is placed in the $(0, 0, 48)$ position, i.e. located on the $z - axis$ in the 48^{th} voxel $24,885mm$ apart from the center which is already outside of the scanner as the edge of the detector panel is at $22,815mm$ in the $z - direction$. The PSF can be seen in Figures 32,33. No distortion can be seen and the similarity with the central PSF is 93,84%.

The similarity results above suggest that edge effect can cause significant distortion in the PSF compared to the central value (PSF_0) even when the edge of the voxel space is far from the detector panel, thus there is no real geometry dependence introduced. Also, the assumed symmetry of the

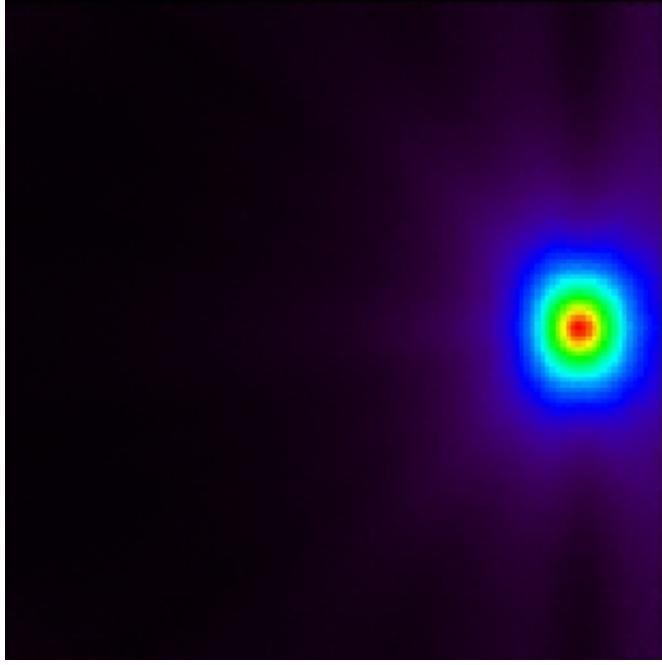


Figure 30: PSF in $(48, 0, 0)$ position, transverse view

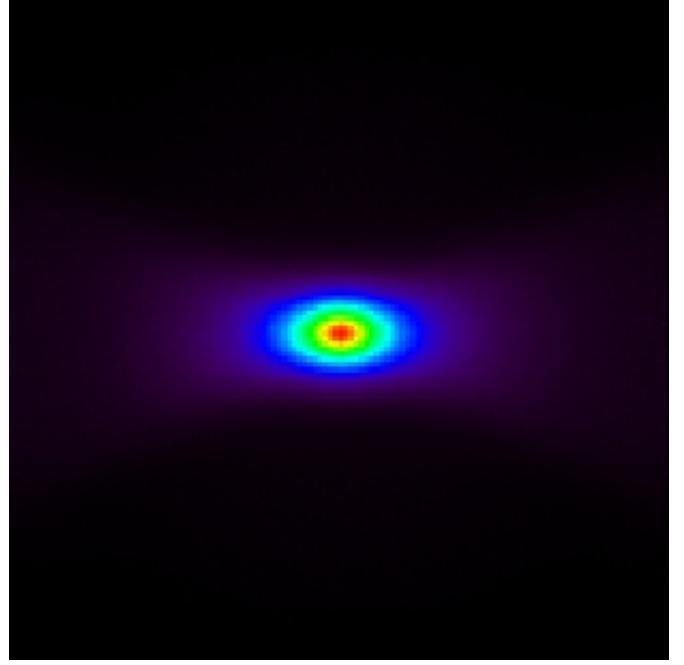


Figure 31: PSF in $(48, 0, 0)$ position, coronal view

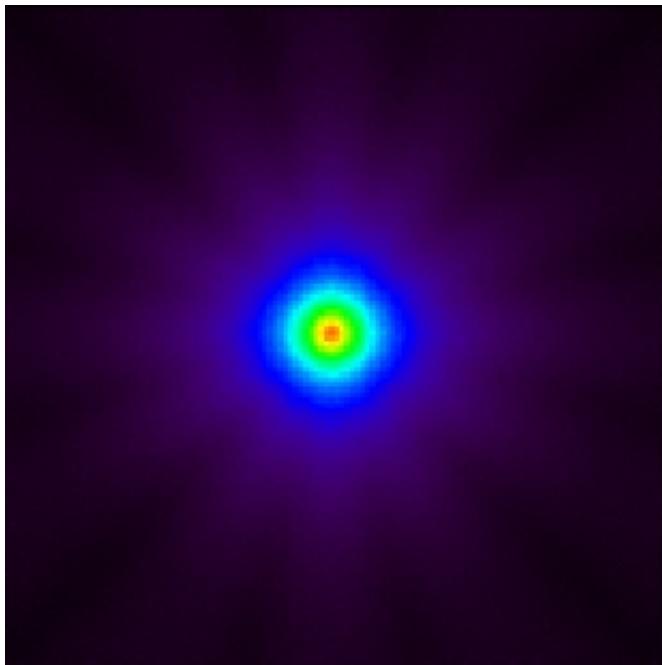


Figure 32: PSF in $(0, 0, 48)$ position, transverse view

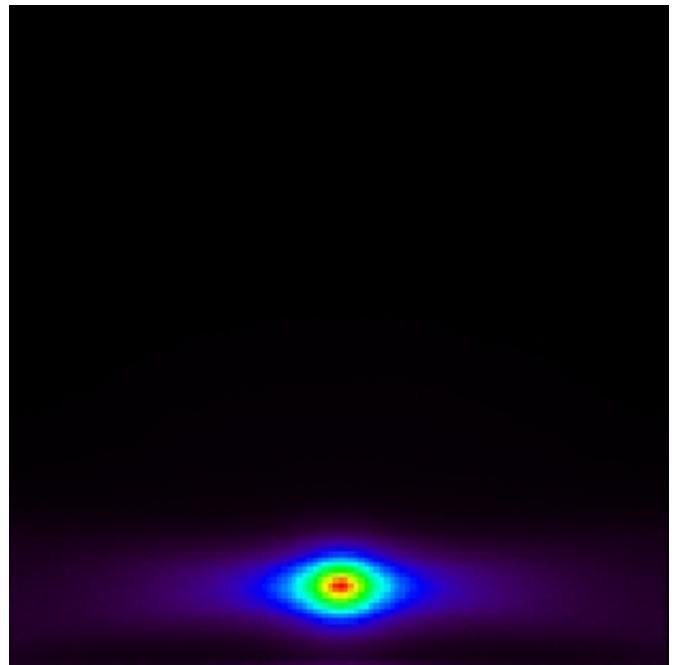


Figure 33: PSF in $(0, 0, 48)$ position, coronal view

PSF is broken. Another significant effect occurs on material borders, i.e. when the center of the PSF is close to another material/tissue type which is almost always the case due to the spread of the PSF. Consequently, using only the central Gaussian as the overall PSF is a very rough approximation. A real anatomical structure can introduce various voxel configurations considering the neighbouring

materials and their relative positions. This means that even the Gaussian fit is infeasible for a precise description of every voxel and the corresponding PSF.

First, the position dependence due to scanner geometry is discussed. As this gives the less significant contribution to the PSF distortions the following principal component - like method was not incorporated to the deconvolution process (as increased the computation time linearly with the number of principal component kept) but it can be important for further improvement and fine tuning.

5.2 Karhunen-Loëve PSF decomposition

A more complex approximation of modelling the position dependence of the PSF is the Karhunen-Loëve decomposition which is related to principal component analysis. PSFs $\{P_i^*\}_{i=1}^N$ have to be simulated at various points of the voxel space to capture the dominant structures in the differences, then their covariance matrix is composed: $C_{ij} = \langle P_i^*, P_j^* \rangle$. The $\{p_i\}_{i=1}^K$ basis PSFs of the voxel space are obtained from the eigen decomposition of matrix C as follows:

$$p_i = \sum_{j=1}^K x_{ij} P_j^* \quad (5.2.1)$$

x_{ij} denotes the j^{th} coordinate of the i^{th} eigenvector. Then the position dependence of the PSF is modelled by weight fields $\{a_i(u, v)\}_{i=1}^K$. The $PSF(x, y, z)$ at a given (x_0, y_0, z_0) point:

$$PSF_{(x_0, y_0, z_0)}(x, y, z) = \sum_{i=1}^K a_i(x_0, y_0, z_0) p_i(x, y, z) \quad (5.2.2)$$

The remaining difficulty is that the behaviour of $\{a_i(u, v)\}_{i=1}^K$ weight fields has to be modelled over the image domain so that they can be expressed as continuous coefficient fields (usually interpolation is used). This type of decomposition can well approximate the changes in the PSF due to geometry and voxel space edge effects, but increases dramatically the computation time and the corresponding deterioration is has not proven to be significant. Contrarily, the position dependence due to material map is a key factor and the following subsections aim at its incorporation to the deconvolution process.

5.3 Incorporating material dependence

Karhunen-Loëve decomposition gives the possibility of accounting for the geometry related distortions of the PSF. However, according to the test simulations performed and analysed in section 5.1 the distortion is not significant in z – *direction*, i.e. along the symmetry axis of the PET scanner where the PSF has smaller spread, and the radial distortion has smaller importance as usually the

imaged object is located further from the edges of the voxel space in the $x - y$ plane, i.e. in those direction where the PSF has large spread.

Furthermore, simulations showed that the most significant distortion in the PSF is introduced near the material borders where the PSF spreads to another type of material. Figures 34,35 illustrates this phenomenon in the voxel space used in section 2.5 and section 4.1 for the homogeneous cylinder ring phantom. At the border of the bone cylinder and the water significant distortion can be seen.

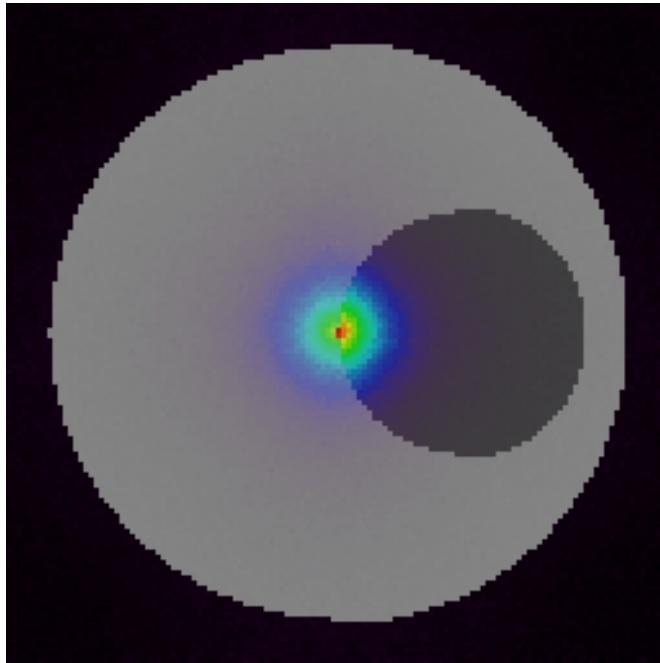


Figure 34: PSF distortion on bone-water border, transverse view

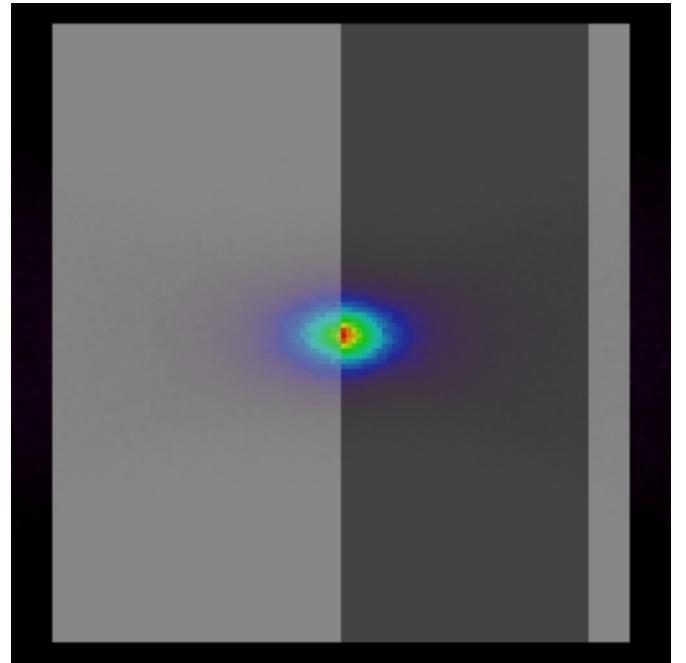


Figure 35: PSF distortion on bone-water border, coronal view

Having a closer look on figures 34,35 it is suggested that using one single PSF in the deconvolution process is a very rough and insufficient approximation if material borders are present which is almost always the case. Nor the Gaussian fit is appropriate neither the symmetry is true, therefore a solution has to be found which allows for these assumption on which the deconvolution model has been built so far.

In order to model the material dependent characteristics of the system the following method is proposed. Once the activity of a voxel located in e.g. water reaches the bone-water border pending one of the convolution steps of RL, in the remaining part of the convolution step it has to propagate according to the bone specific PSF. Converting this into mathematical form the convolution kernel g has to be divided into subkernels generalizing the square root taking step of the derivation is section 3.4. Performing convolution N times with the N^{th} root kernel results in the convolution with kernel g . Then every convolution step of RL is divided into N subconvolutions. After every subconvolution

the activity which crossed material border is assigned to the voxel group corresponding to the new material. In the voxel space of section 2.5 and section 4.1 this means that the activity coming from the water area crossing the bone-water border is deleted from the 3D array stacking the activity in water and added to the 3D array corresponding to activity in bone area.

At this point the semigroup property of the Gaussian kernel can be exploited [28]. It was concluded that the convolution kernel g can be well approximated as a sum of Gaussians. Therefore, the same N^{th} root calculations are valid than in section 3.4 for the square root of the PSF. This means that in the deconvolution process in every forward step the kernel g can be divided into N subkernels denoted by $g_{1/n}$ (which equals to the N^{th} principal root of kernel g) and the convolution can be performed in N consecutive steps:

$$g = g_{1/n} * \dots * g_{1/n} \quad (5.3.1)$$

Then in each subconvolution step the activity contributions are revised with respect to material map and the material specific PSF can be used for a given area containing a given material:

$$g_{material} = g_{1/n}^{mat_1} * \dots * g_{1/n}^{mat_n} \quad (5.3.2)$$

mat_1, \dots, mat_n denotes the material in which the activity is located in the $1^{st}, \dots, n^{th}$ subconvolution step. To validate the algorithm a PSFs located on bone-water material border was simulated and compared to the PSF $g * g$ built from subconvolutions according to the derivation above:

$$PSF_{material} = g_{material} * g_{material} = g_{1/n}^{mat_1} * \dots * g_{1/n}^{mat_n} * g_{1/n}^{mat_1} * \dots * g_{1/n}^{mat_n} \quad (5.3.3)$$

If not stated otherwise the reconstructions in the following used 10 subkernels in the RL iterations. This number has proven a proper compromise between computation time and precise approximation of the real PSF and also offers a tuning possibility with respect to image quality versus invested time.

Figure ?? shows the simulation of a point source placed in the center of the voxel space, therefore close to bone-material border. The simulation used 2×10^{12} positron. Figure ?? shows the corresponding PSF obtained from the (subkernel) convolution of a point source (placed into the same central voxel) with $PSF_{material}$ using 10 subkernels. The similarity is 92,18%, so it can be concluded that the proposed subkernel approach well approximates the real behaviour of the PSF, hence the deconvolution can account for material dependence to a high extent.

Once the proposed algorithm can account for the material map in the voxel space the biggest requirement with respect to performance is met, so significant improvement in the reconstruction quality is expected. The reconstruction results of the proposed method tested on simulation and measurement are presented in the following subsections.

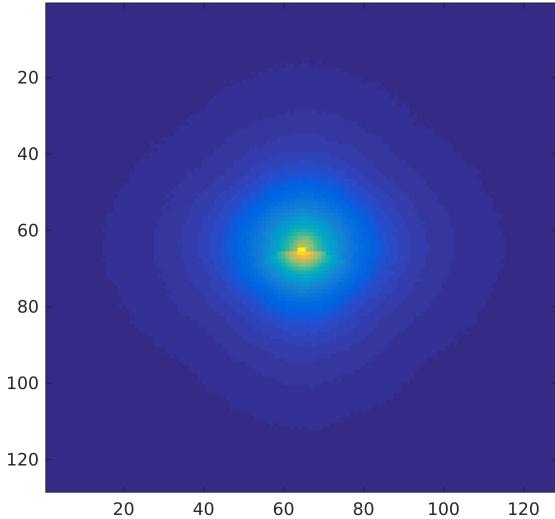


Figure 36: Distortion on bone-water border, simulated PSF, summed transverse view

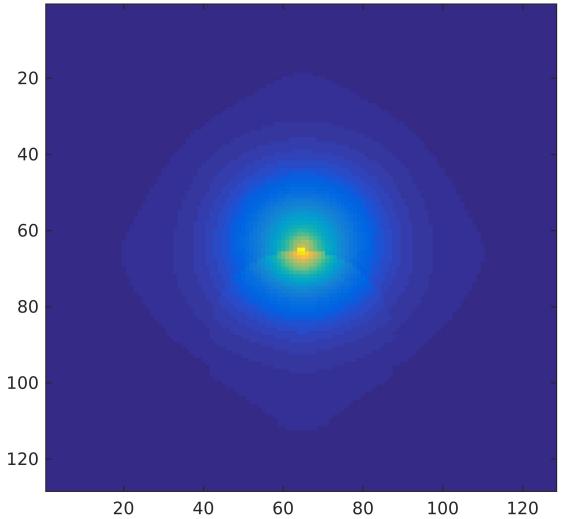


Figure 37: Distortion on bone-water border, subkernel convolved PSF, summed transverse view

5.4 Homogeneous cylinder ring phantom v2

In subsection 4.1 the deconvolution method was tested on the cylinder ring phantom and proved its efficiency as successfully abolished the inhomogeneity artefact and accelerated the overall convergence speed of the reconstruction. However, the method was still not able to handle the "double-step" artefact on the bone-water material border along the activity ring. As the PSF in water is more spread than the PSF in bone ML-EM algorithm with full physical modelling in the back projection tends to assign the activity on the bone-water border rather to the voxels containing bone than to water filled ones. Therefore this artefact is also an attribute of faithful physics modelling together with the inhomogeneity artefact. In agreement, reconstruction scheme with simplified modelling introduces none of the these artefacts.

Despite the impressive results, the proposed deconvolution method with homogeneous PSF could not handle the "double-step" artefact because it arose from the latent material map and the presence of material border. As a first test for the subkernel approach a reconstruction with the same parameters as in subsection 4.1 was performed. Figure ?? shows the resulted image summed along the $z - axis$. Significant improvement with respect to activity homogeneity on the material border can be seen, the use subkernels in RL iterations reduced the "double-step" artefact to a high extent. For comparison zoomed images of the corresponding parts of the ring with the following settings are presented in figures ??, ??, ?? respectively: subkernel approach, homogeneous PSF and without deconvolution.

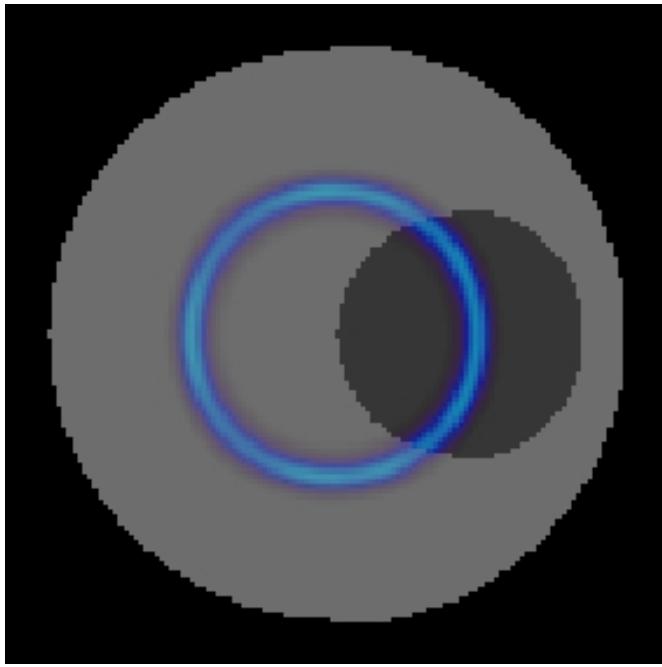


Figure 38: Reconstruction after 20 iterations, summed transverse view. The "double-step" artefact is significantly reduced even after 20 iterations

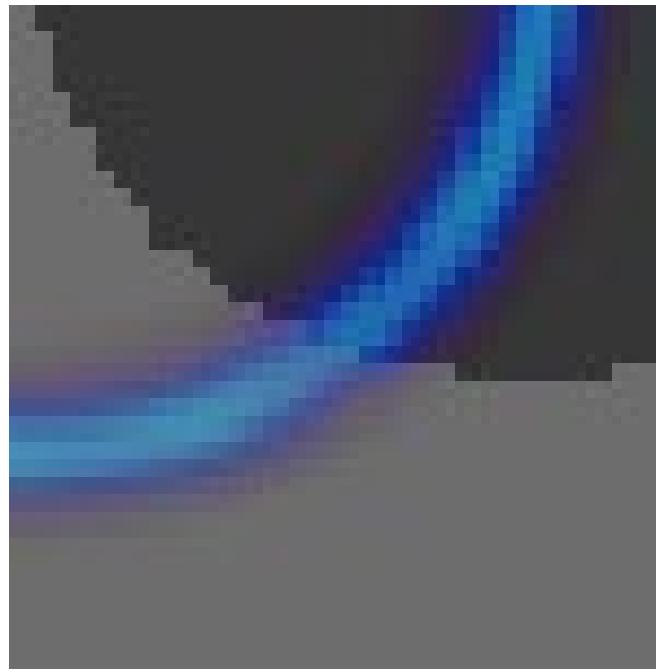


Figure 39: Zoomed part of figure ??

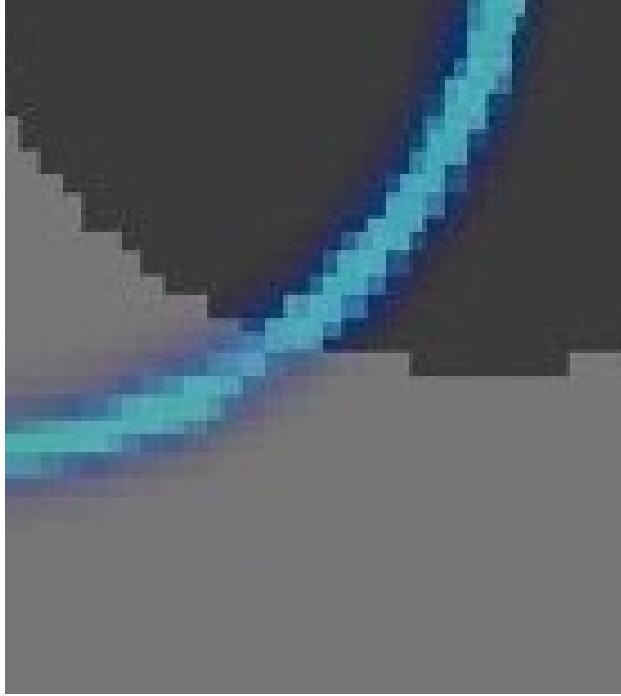


Figure 40: Corresponding zoomed part of figure 27 where homogeneous PSFs were used in the deconvolution

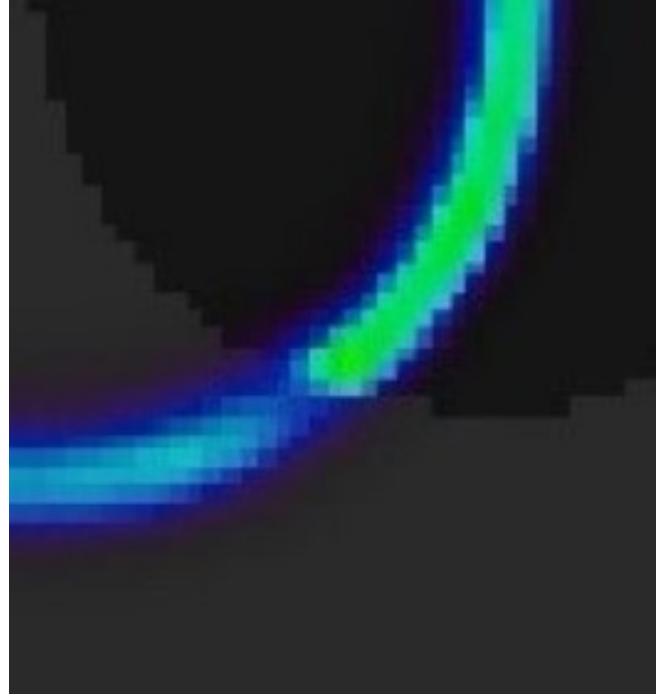


Figure 41: Corresponding zoomed part of figure 3 where no deconvolution was used with the fully physical modelling BP

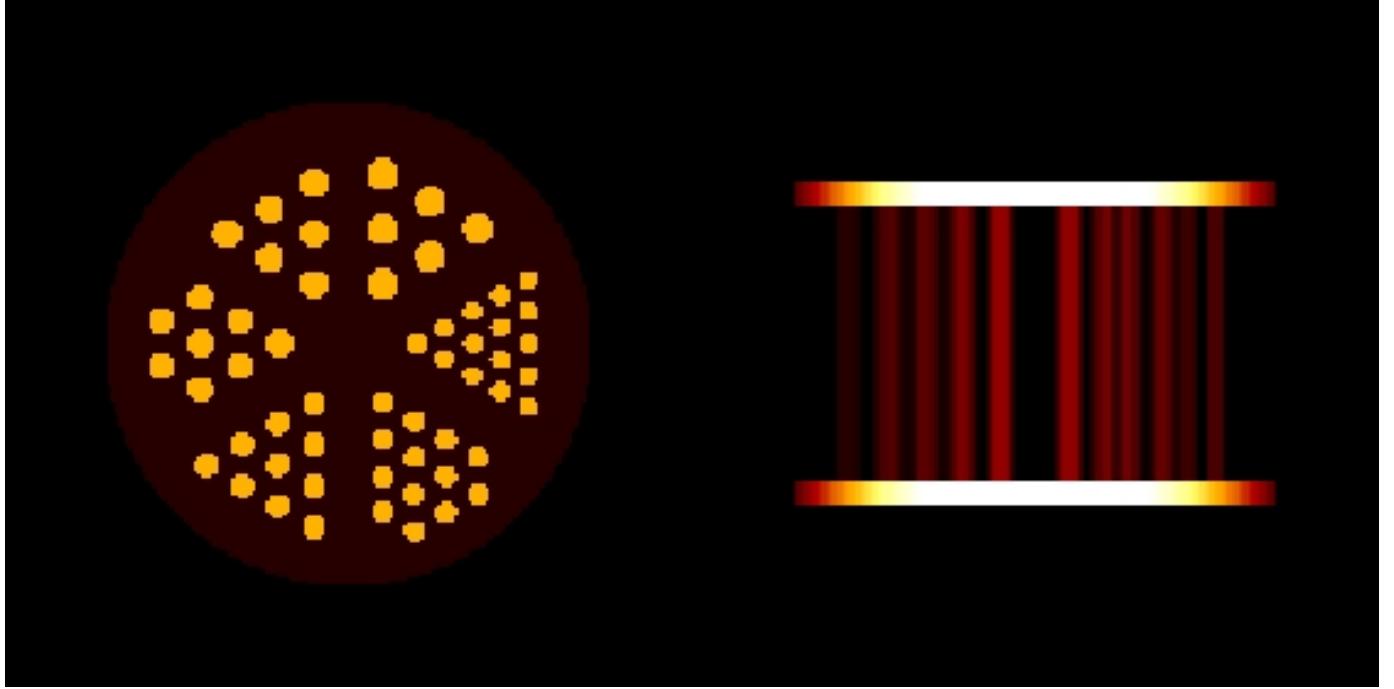


Figure 42: Derenzo phantom, transverse view

Figure 43: Derenzo phantom, coronal view

5.5 Simulated Derenzo phantom

The deconvolution method was tested on the simulation of a Derenzo standard PET phantom which is full of material borders, so is a strong candidate for testing the effectiveness of the subkernel approach. Standard PET phantoms allow for inter-scanner or inter-reconstruction comparison. One of the most common quality control phantoms for nuclear medicine imaging is the Derenzo phantom. The phantom pattern, shown in figures 42,43 is a series of positron emitting rods separated by twice their diameter in a triangular close-packed configuration. Several rod diameters are typically employed as a way to determine the diameter at which resolution breakdown occurs [27]. In this current setup the diameter of the rods varies between 0.7mm and 1.2mm .

The voxel space is $200 \times 200 \times 200$, voxels are 0.1907mm sided, the iterations use 5×10^9 positrons. The body of the phantom is composed of plexi glass, the phantom volume (rods and closing plates) is filled with water (containing the activity). A measurement of ^{15}O is simulated as input data. The reconstructions used simplified modelling (reference setting) and full physical modelling with subkernel deconvolution in the back projection for performance comparison. As it was highlighted in section 2.7 simplified modelling results in activity accumulation in the image corners. For quantitative comparison two metrics are used, $L_2 - \text{norm}$ and $CC - \text{norm}$. In the former case this corner activity is untouched, while in the latter case a band of 10 voxel is neglected on all the six sides of the cubic voxel space ($CC - \text{norm}$ is always normalised into the $0 - 100$ interval, so the truncation of the voxel

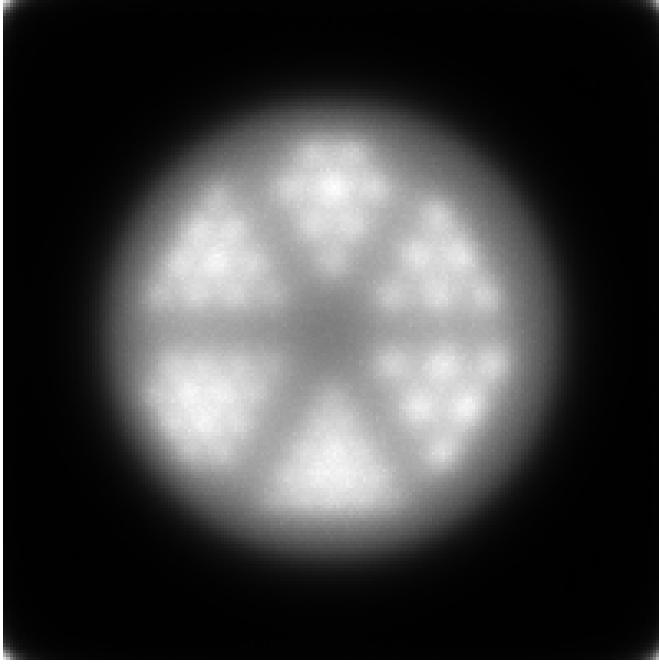


Figure 44: Derenzo phantom reconstruction, summed transverse view after 10 iteration with simplified modelling in the BP. Activity accumulation in the corners can be seen

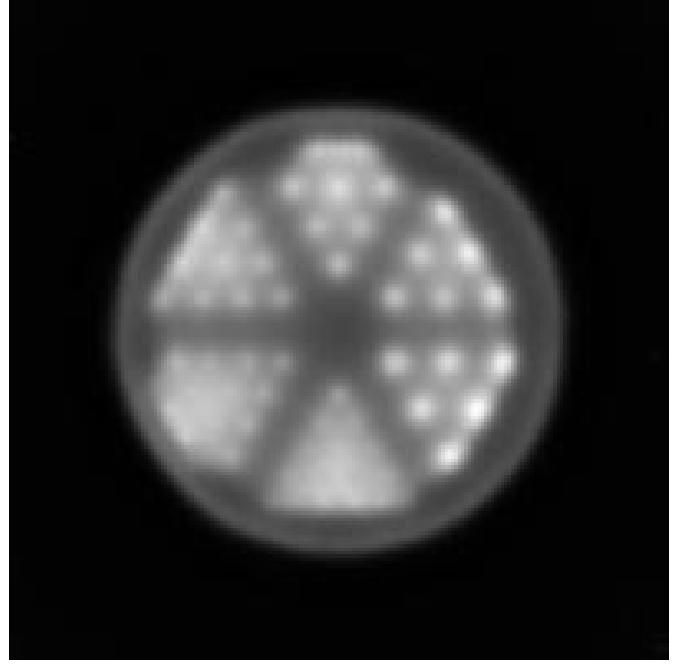


Figure 45: Derenzo phantom reconstruction, summed transverse view after 10 iteration with subkernel deconvolution in the BP

space is possible in contrast with $L_2 - \text{norm}$ which gives smaller number if smaller number of voxel is used). This means that the disadvantage of the simplified modelling is manually removed and the deconvolved back projection has to prove more efficiency in this handicapped situation. Figures 44 and 45 shows the reconstructions after 10 iterations. The results corresponding to simplified modelling are located on the left, those corresponding to the deconvolved full physical modelling are located on the right. For qualitative comparison $CC - \text{norms}$ are plotted in figure 46 and $L_2 - \text{norms}$ are plotted in figure 47. It can be concluded that the deconvolved scheme represents introduces much faster initial convergence but after a characteristic number of iterations the decay of the norms flattens and the reconstruction with simplified modelling gives better qualitative result. At this point it has to be highlighted that the activity accumulation in the corners was neglected. Otherwise, the $L_2 - \text{norm}$ shows significant advantage on the side of the deconvolved full physical modelling.

The flattening of the $CC-norm$ after a characteristic number of iterations is surmised to originate from the quality of the fit of the PSF. Once the reconstruction resolution reaches the precision of the PSF fit the deconvolution loses its impact and the further (CCL_2)norm path becomes very similar to the one specific to the full physical modelling without deconvolution. This happens almost regardless to the number of RL iterations in the deconvolution process i.e. the amount of regularisation. Below a certain amount (i.e. above a certain number of RL iterations), deconvolution rather causes noise accumulation than convergence acceleration in correspondence with theory.

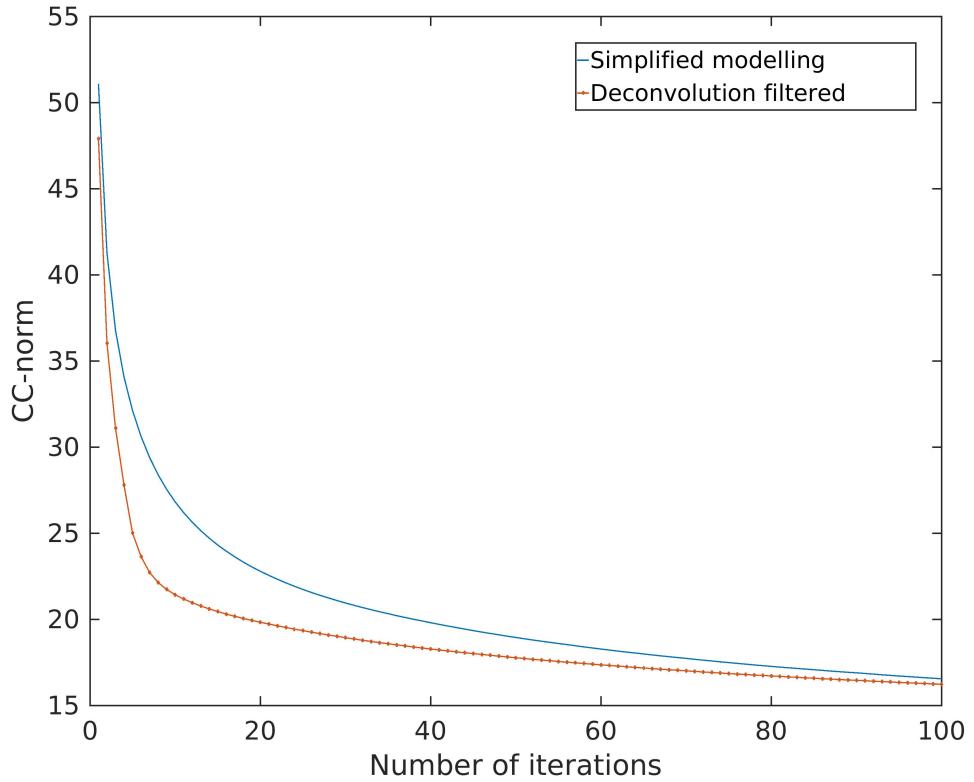


Figure 46: $CC - norm$ comparison of the algorithm with simplified modelling in the BP and the proposed algorithm with subkernel approach in the deconvolution with respect to the number of ML-EM iterations. The faster initial convergence implied by the proposed method is clearly visible. However, its advantage reduces as the resolution of the reconstruction reaches the precision of the PSF fit

To overcome this limitation of PSF fitting several forms have been tested. In subsection 3.4 the PSF is assumed to be the sum of three Gaussian and therefore its Fourier-transform is real valued. However, this is true only approximately and there is small but non-zero phase component of the result. Accounting for this phase component in the deconvolution can stabilize the process to a certain

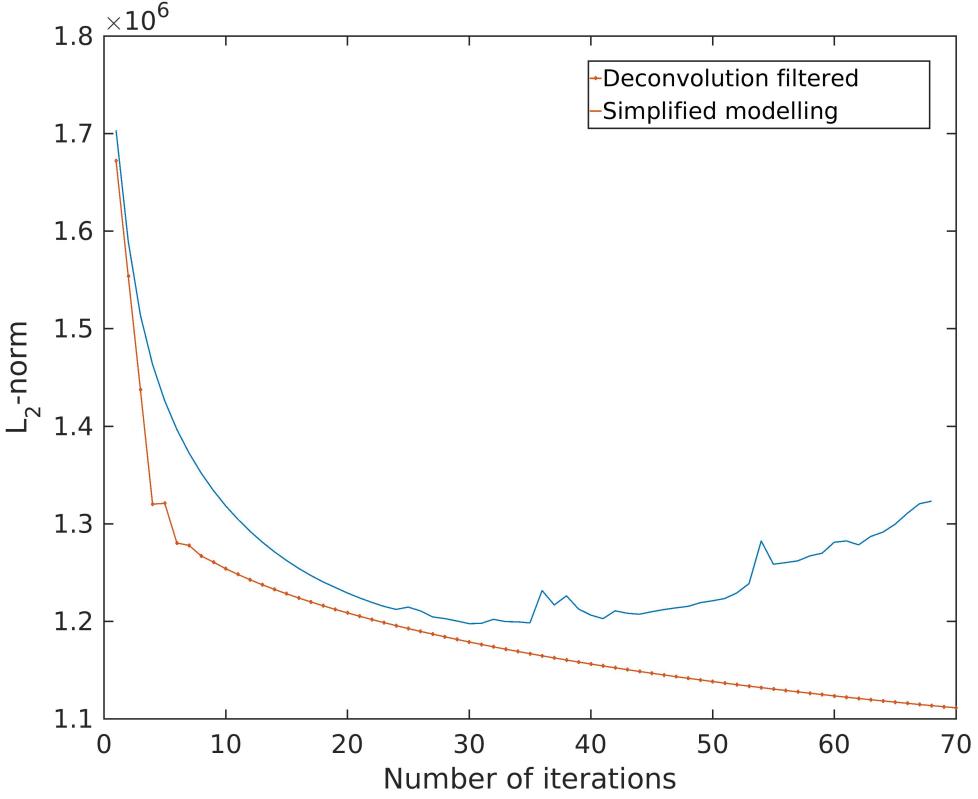


Figure 47: $L_2 - \text{norm}$ comparison of the algorithm with simplified modelling in the BP and the proposed algorithm with subkernel approach in the deconvolution with respect to the number of ML-EM iterations. Simplified modelling iterates away from the solution due to activity accumulation in the corners of the voxel space

extent provided that the phase map remains smooth enough after taking the N^{th} root of the PSF in the Fourier domain. This modification has indeed proven to be useful with respect to stability. This difference was much more significant when measured data was used as the input of the reconstruction. In this case the MC modelling is not same process as the measurement generation only an approximation. Therefore, any imperfection in the MC modelling is amplified and the deconvolution task also becomes more challenging as the PSF is based on simulation and the question of noise stability is more emphasized in this case. Results are presented in the next subsection.

5.6 Measured Derenzo phantom

Performance of the deconvolved ML-EM algorithm was evaluated against measured data as well. The reconstruction uses a measurement of a Derenzo phantom with identical geometry and reconstruction parameters to that used in the simulated Derenzo subsection, the total number of LoR counts is

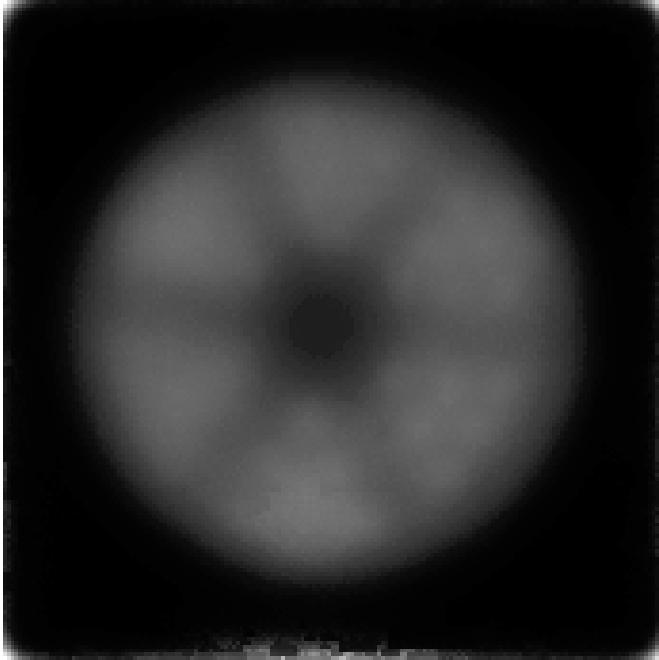


Figure 48: Derenzo phantom reconstruction, summed transverse view after 10 iteration with simplified modelling in the BP. Brightness is tuned so that there is anything visible beside activity in the corners

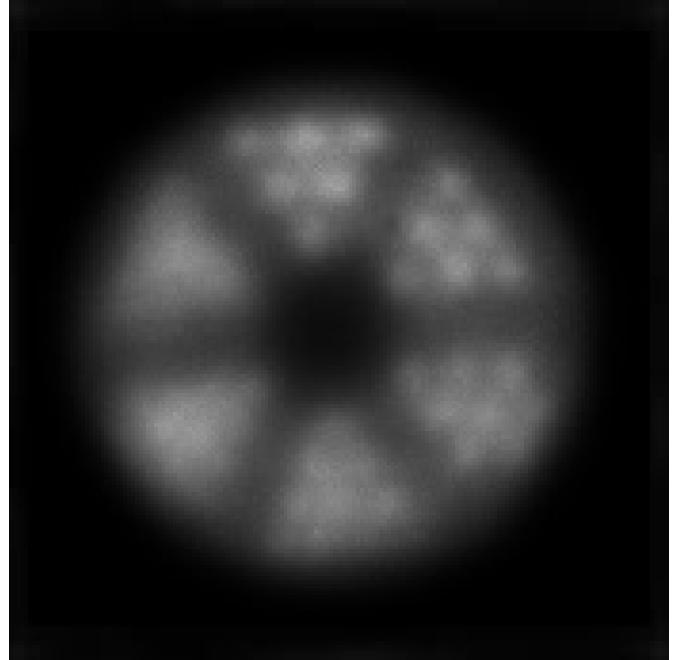


Figure 49: Derenzo phantom reconstruction, summed transverse view after 10 iteration with deconvolution in the BP

1.2423×10^9 . Unlike in the previous tests, in this case the isotope is ^{18}F . Consequently, positron range effects are less conspicuous but the other blurring factors are still strongly present, so the proposed deconvolution filtering can still bring something to the role. Due to lack of phantom reference the results are qualitative only a line profile through the rods gives some qualitative information with respect to resolution. Results after 10 ML-EM iteration (activity in the corners accumulates much faster than in case of simulation) can be seen in figures 48,49.

After qualitative comparison it can be concluded that the proposed method might be a useful image quality improving modification for standard ML-EM reconstruction scheme and also full physical modelling is worth the effort implementing on real scanners. However, further fine tuning is required.

6 Remaining problems

Throughout this thesis it was highlighted several times that the biggest hindrance for the proposed algorithm is the imperfection of the PSF fit. Many efforts were made to improve the performance of the fitting process or replace it with extra long simulations. However, the precision required by

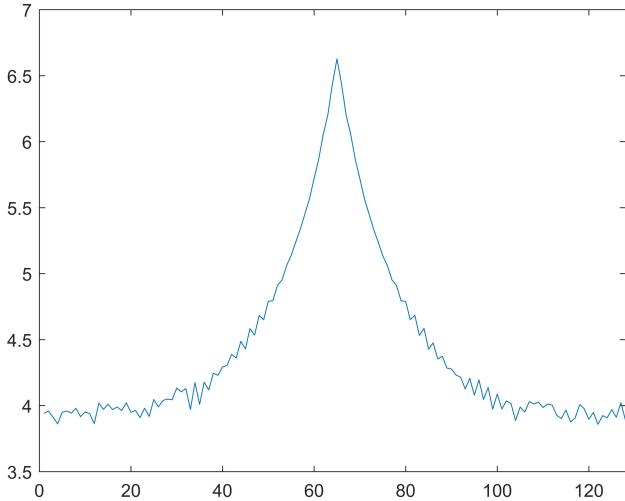


Figure 50: Fourier-transform of the 10^{th} root of the transverse line profile of kernel g . the simulation took approximately 50 hours, but the noise is still significantly present.



Figure 51: Resulted salt and pepper - like noise accumulation in the reconstruction. The subsequent ML-EM iteration step gave only a pattern similar to a Dirac-delta and the reconstruction is broken

the subkernel approach is still not met even in the case of a 2 day long PSF simulation for each material. Representative result can be seen in figure 50. The subkernel is still noisy and therefore cannot be used directly for deconvolution due to noise amplification which leads to the deterioration of the reconstruction algorithm very quickly, see figure 51.

Improving the performance of the fit procedure is a remaining task for future development and is hoped to increase dramatically the efficiency of the proposed deconvolution approach of the SVD-filter. Reformulation of the model function for the PSF (instead of the sum of three Gaussians plus offset) can be a possible direction. Also, precise enough implementation of the direct 3D fit would also improve reconstruction quality as the coefficient equality constraint for individual 3D Gaussians with respect to direction (see subsection ...) was very hard to met and only approximately succeeded in some cases.

Every reconstruction which used the presented deconvolution SVD-filtering lacked any other type of filtering method due to trustworthy performance evaluation not being influenced by other effects caused by additional manipulation. Further research could analyse the behaviour of the proposed deconvolution approach when incorporating other filtering to the algorithm. As it is well known that deconvolution amplifies noise, performance improvement is expected with well tailored supplementary

modifications.

7 Conclusion

In this work a possible convergence accelerating method for ML-EM PET image reconstruction was presented as a continuation of the SVD based convergence analysis and the corresponding filtering method from my bachelor's thesis. Despite the promising results of the proposed SVD filter in the 1D model the practical implementation for a real scanner seemed infeasible due to computational and storage related difficulties.

This thesis presents a possible remedy for these problems. A convolution model was formed to approximate as faithfully as possible the effect of the SVD filtering without introducing any computational burden and the only requirement is the storage of some voxel space sized array. The proposed method uses the PSF of the imaging system for the deconvolution of the nominator and denominator of the back projecting step of ML-EM PET reconstruction. The calculation of the convolution kernel g from the PSF incorporating material dependence is a novel contribution, to the best knowledge of the author, and makes the approximation of the SVD-filter possible which was also firstly used in [19]. The modification hardly lengthened the reconstruction time but gave significantly better results in term of image quality and overall convergence speed and also abolished the inhomogeneity artefact of back projection using full physical modelling which artefact had been present without deconvolution.

The algorithm uses the PSF of the system which is known to be position dependent. The quality of the homogeneous approximation when every voxel is deconvolved with the central PSF was tested and the approximation has proven to be sufficient. The most significant distortion of the PSF is introduced by the material map and material borders. A possible remedy for this problem was proposed which divides the convolution kernel into subkernels. The algorithm was tested on both simulated and measured Derenzo phantom with several configurations. The reconstruction results outperformed the results of the simplified back projection using algorithm and the method is still time-saving (even in MATLAB environment) thanks to FFT and GPU implementation.

To summarize, the proposed method of this thesis successfully reproduced the advantageous effects of the SVD filter (accelerated convergence, preserved fix point and noise stability) and hence opened the way for the full physical modelling in the back projection towards clinical and pre-clinical use resulting in a better reconstruction than the best setting so far.

References

- [1] Y. Vardi, L.A. Shepp and L. Kaufman, *A statistical model for positron emission tomography (with discussion)*, Journal of the American Statistical Association 1985. Volume 80, pages 8-20, issue 389
- [2] Geoffrey McLachlan, Thriyambakam Krishnan, *The EM Algorithm and Extensions*, 1997. ISBN: 0-471-12358-7
- [3] Gengsheng Lawrence Zeng, *Medical Image Reconstruction*, Springer Heidelberg Dordrecht London New York 2010. ISBN: 978-3-642-05367-2
- [4] Laszlo Szirmay-Kalos, Milan Magdics, Balazs Toth and Tamas Bukki, *Averaging and Metropolis Iterations for Positron Emission Tomograph*, LECTURE NOTES IN COMPUTER SCIENCE 8353: pp. 147-154. (2014)
- [5] Gengsheng L. Zeng and Grant T. Gullberg, *Unmatched Projector/Backprojector Pairs in an Iterative Reconstruction Algorithm*, IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 19, NO. 5, MAY 2000
- [6] Fan Xin, Wang Hai-Peng, Yun Ming-Kai, Sun Xiao-Li, Cao Xue-Xiang, Liu Shuang-Quan et al., *PET image reconstruction with a system matrix containing point spread function derived from single photon incidence response*, Chinese Physics B, Volume 24, Number 1
- [7] Kuang Gong, Simon R Cherry, and Jinyi Qi, *On the assessment of spatial resolution of PET systems with iterative image reconstruction*, Phys Med Biol. 2016 Mar 7; 61(5): N193–N202, doi: 10.1088/0031-9155/61/5/N193
- [8] Lucia B. Chávez-Rivera , Leticia Ortega-Máynez, José Mejía, Boris Mederos, *ML-EM reconstruction model including total variation for low dose PET high resolution data*, Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2015 IEEE, doi: 10.1109/NSS-MIC.2015.7582221
- [9] Xiao-Yue Gu, Wei Zhou, Lin Li, Long Wei, Peng-Fei Yin, Lei-Min Shang et al., *High resolution image reconstruction method for a double-plane PET system with changeable spacing*, Chinese Physics C, Volume 40, Number 5
- [10] Á. Cserkaszy and Dr. D. Légrády, A. Wirth, T. Bükkii and G. Patay, *GPU Based Monte Carlo for PET image reconstruction: Parameter optimization*, International Conference on Mathematics

and Computational Methods Applied to Nuclear Science and Engineering (MC 2011), Rio de Janeiro, RJ, Brazil, May 8-12, 2011.

- [11] PANNI – PET Aimed Novel Nuclear Imager; Forward Monte Carlo based image reconstruction software for Positron Emission Tomography corresponding to code Revision 664; General description; Coding: Á. Cserkaszky, Á. Szlávecz, D. Légrády, G. Tolnai; contact: legrady@reak.bme.hu
- [12] Magdy M. Khalil, *Basic science of PET imaging*, Springer 2017, ISBN 978-3-319-40068-6
- [13] A. Wirth, Á. Cserkaszky, B. Kéri, D. Légrády, S. Fehér, S. Czifrus, and B. Domonkos, *Implementation of 3D Monte Carlo PET Reconstruction Algorithm on GPU*, in IEEE Medical Imaging Conference (2009.)
- [14] Mediso Ltd. Product page of nanoScan
- [15] Stephen E. Derenzo, *Mathematical removal of positron range blurring in high resolution tomography*, IEEE Transactions on Nuclear Science, Vol. 33, No. 1, February 1986
- [16] Wencke Lehnert, Marie-Claude Gregoire, Anthonin Reilhac and Steven R. Meikle: *Analytical positron range modelling in heterogeneous media for PET Monte Carlo simulation*, 2011 Phys. Med. Biol. 56 3313
- [17] Briesmeister, *MCNP – a General Monte Carlo N-Particle Transport Code*, Los Alamos National Laboratory report LA-13709-M, April 2007
- [18] V. Somai, Bachelor's Thesis, *The effect of positron range modelling on spatial resolution of ML-EM PET iterative image reconstruction*, Institute of Nuclear Techniques BME 2016.
- [19] Somai, V., Legrady, D., Tolnai, G., *Singular value decomposition analysis of back projection operator of maximum likelihood expectation maximization PET image reconstruction*, Radiology and Oncology 2018., doi:10.2478/raon-2018-0013
- [20] Zheng Liu, Takashi Obi, Masahiro Yamaguchi, and Nagaaki Ohyama, An investigation of convergence rates in expectation maximization (EM) iterative reconstruction, Nuclear Science Symposium, 1999. Conference Record. 1999 IEEE, doi: 10.1109/NSSMIC.1999.842824
- [21] Per Christian Hansen, *Numerical tools for analysis and solution of Fredholm integral equations of the first kind*, Inverse Problems 8 (1992) 849-872, doi: <https://doi.org/10.1088/0266-5611/8/6/005>

- [22] Anna Pyzara, Beata Bylina, Jarosław Bylina, *The influence of a matrix condition number on iterative methods' convergence*, Institute of Mathematics, Marie Curie-Skłodowska University, ISBN: 978-83-60810-22-4
- [23] A. Rahmim, M. A. Lodge, J. Tang, S. Lashkari, M. R. Ay: *Analytic System Matrix Resolution Modeling in PET: An Application to Rb-82 Cardiac Imaging*, October 2008, Physics in Medicine & Biology, Volume 53, Number 21, DOI: 10.1088/0031-9155/53/21/004
- [24] W Ulmer: *Inverse problem of linear combinations of Gaussian convolution kernels (deconvolution) and some applications to proton/photon dosimetry and image processing*, 2010 IOP Publishing Ltd, Inverse Problems, Volume 26, Number 8, DOI: 10.1088/0266-5611/26/8/085002
- [25] Roger A. Horn, Charles R. Johnson, *Matrix Analysis, second edition*, Cambridge University Press 2013. ISBN 978-0-521-83940-2
- [26] Smith J.O., *Spectral Audio Signal Processing*, Stanfor University webpage, online book, 2011 edition, accessed 2018
- [27] Benjamin L Cox, Stephen A Graves, Mohammed Farhoud, Todd E Barnhart, Justin J Jeffery, Kevin W Eliceiri, Robert J Nickles, *Development of a novel linearly-filled Derenzo microPET phantom*, Am J Nucl Med Mol Imaging 2016. 6(3) 199–204.
- [28] Hong Yan, *Signal Processing for Magnetic Resonance Imaging and Spectroscopy*, Marcel Dekker 2002, ISBN: 0-8247-0653-6