Reconstruction scheme for accelerated maximum likelihood reconstruction: the patchwork structure

Katrien Van Slambrouck and Johan Nuyts

***Abstract*—Convergence of iterative algorithms can be improved by updating groups of voxels sequentially rather than updating the whole image simultaneously. The optimal way is to choose groups of uncoupled voxels, i.e. voxels spread over the reconstruction volume. While this is most efficient for convergence reasons, updating groups of spread voxels is less efficient regarding memory access and computational burden. In this work, an image-block update scheme is presented that updates relatively large groups of voxels simultaneously while keeping a considerable gain in convergence. The sequential image-block update can also be combined with ordered subsets. This image-block or patchwork scheme is applied both to transmission and emission maximum likelihood algorithms.**

I. Introduction

Iterative reconstruction has several advantages over analytical, filtered backprojection (FBP) type, reconstruction. A forward model, as used in iterative reconstruction, is much more suitable for model adaptations, compared to the inverse model used in analytical reconstruction methods. Consequently, with iterative reconstruction one can implement more accurate acquisition models, resulting in more accurate reconstructions. Moreover, iterative reconstruction can cope better with incomplete data and truncation. Several groups have shown that iterative reconstruction can lead to dose reduction with preserved image quality, for example in [1], [2], [3]. This is partly due to the use of non-linear constraining techniques [4], [5] and partly to the use of an accurate noise model [6]. The most important disadvantage of iterative reconstruction is the increased computation time. For CT (computed tomography), this is still an important obstacle for introduction into daily clinical practice.

Many iterative algorithms update all image voxels simultaneously, based on a computation involving all projections. Examples are: MLEM [7], ISRA [8], [9], generalized MLEM/ISRA [10], NEGML [11], ABML [12] MLTR [13], convex algorithm [14], and SIRT [15]. The convergence of iterative algorithms is often improved by using blocks of projection data sequentially instead of using all projections simultaneously. This idea is used in the ordered subsets algorithms (OSEM[16], OSTR[17] and Ordered subsets convex algorithm [18]) or in ART [19], [20] and SART [21]. Convergence can be improved as well by updating voxels sequentially rather than simultaneously as was done by Sauer and Bouman in [22]. The improved convergence of this coordinate ascent approach is (partly) offset by the increased computation time per iteration. The sequential update requires to access the system matrix column by column (retrieving all elements for a single image voxel), while conventional projector/backprojector software typically enables row by row access only (retrieving elements for a single projection line). In CT reconstruction sequential updates also cause an increased number of exponentiation operations compared to one simultaneous update. A compromise between convergence and computational burden is made by Fessler et *al.* in the space-alternating generalized EM algorithms (SAGE [23]) and the grouped coordinate ascent algorithm [24], [25] where groups of voxels are updated. The optimal way of choosing these groups is by selecting voxels that have low coupling. This means that the update for one voxel is nearly independent of the updates of other voxels in the group. The lower the coupling between voxels in a group, the faster the convergence. For most system geometries this results in groups of voxels spread over the image. The main drawback of updating groups spread over the image is inefficient memory access.

In our work on metal artifact reduction in CT we developed a reconstruction algorithm based on a patchwork structure [26]. The patchwork structure divides the reconstruction volume in several subareas or patches. In case of metal artifact reduction the patches were chosen such that every metal object and its close surroundings were considered as one patch, the remaining volume was another nonmetal patch. The original idea of the method was to apply different reconstruction models in different patches. By updating the regions separately, a grouped coordinate algorithm was applied to the patches (one patch is one group). A substantial increase in convergence was observed for the smaller (metal) patches, although the voxels in one group were adjacent and quite strongly coupled.

Benson et *al.* present in [27] another technique for updating coupled blocks of voxels in iterative CT reconstruction. The technique is an extension of the ICD technique described in [22]. The reconstruction volume is divided in several small blocks (maximum 8×8 voxels). In the block raster strategy, the different blocks are sequentially updated while the voxels in a block are simultaneously updated. The different axial planes are updated one by one. A linear system of equations for all voxels in the block is solved by inversion of the block matrix. This block matrix includes off-diagonal elements that describe the correlation between the voxels in a group. By including this correlation, larger (but still rather small groups) of voxels can be updated with improved convergence with respect to voxel by voxel updating but at the cost of increased

Corresponding author: katrien.vanslambrouck@uzleuven.be. Department of Imaging and Pathology: Division of Nuclear Medicine and Medical Imaging Research Center, KU Leuven, B-3000 Leuven, Belgium. This work is supported by the SBO-project QUANTIVIAM (060819) of the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen).

computational burden caused by the inversion of a matrix with (non-zero) off-diagonal elements.

The idea of updating relatively large groups of connected voxels as in [26] will be further explored in this work. Instead of local improved convergence, the patchwork structure presented here aims to improve the overall convergence level of the reconstruction. The framework is a rather simple and intuitive gradient ascent maximum likelihood optimization algorithm. In the next section the maximum likelihood framework is introduced, followed by a description of the patch structure. The convergence of patched updates with respect to fully simultaneous updates with and without ordered subsets will be analyzed for different CT geometries and for 3D PET (positron emission tomography).

II. Methods

A. Maximum likelihood gradient ascent optimization

The study presented in this work is based on a gradient ascent maximization of the Poisson (log)likelihood, which is given by:

$$L(\vec{\tau}) = \sum_i \left(y_i \ln \hat{y}_i(\vec{\tau}) - \hat{y}_i(\vec{\tau})\right), \tag{1}$$

with i the index of the projection lines, y_i the measurement, $\hat{y}_i$ the estimate of y_i calculated from the current reconstructed image $\vec{\tau}$, containing the attenuation coefficients in case of transmission tomography or the activity for emission tomography.

The derivation of the patchwork update scheme is based on a gradient ascent approach [11][13]. A surrogate for the likelihood is optimized based on a quadratic approximation:

$$L(\vec{\tau} + D(\alpha)\Delta x) \simeq T_1(\vec{\tau}, \Delta x) \tag{2}$$

$$= L(\vec{\tau}) + \sum_j \left.\frac{\partial L}{\partial \tau_j}\right|_{\vec{\tau}} \alpha_j \Delta x_j + \frac{1}{2}\sum_{j,k} \left.\frac{\partial L^2}{\partial \tau_j \partial \tau_k}\right|_{\vec{\tau}} \alpha_j \alpha_k \Delta x_j \Delta x_k. \tag{3}$$

D is a diagonal matrix with $D_{jj} = \alpha_j \geq 0$. The second derivatives of L are all negative and

$$2\Delta x_j \Delta x_h \leq \Delta x_j^2 + \Delta x_h^2. \tag{4}$$

Using (3) and (4) a surrogate function T_2 is constructed which is equal to T_1 in the current solution and smaller than T_1 elsewhere.

$$\begin{aligned} T_1(\vec{\tau}, \Delta x) &\geq T_2(\vec{\tau}, \Delta x) && (5)\\ &= L(\vec{\tau}) + \sum_j \left.\frac{\partial L}{\partial \tau_j}\right|_{\vec{\tau}} \alpha_j \Delta x_j \\ &\quad + \frac{1}{2}\sum_{j,k} \left.\frac{\partial^2 L}{\partial \tau_j \partial \tau_k}\right|_{\vec{\tau}} \alpha_j \alpha_k (\Delta x_j)^2. && (6)\end{aligned}$$

The update for $\vec{\tau}, \Delta\tau_j = \alpha_j \Delta x$, is obtained by maximizing the surrogate function T_2:

$$\Delta\tau_j = -\frac{\alpha_j \left.\frac{\partial L}{\partial \tau_j}\right|_{\vec{\tau}}}{\sum_h \alpha_h \left.\frac{\partial^2 L}{\partial \tau_j \partial \tau_h}\right|_{\vec{\tau}}} \tag{7}$$

This update maximizes T_2, which guarantees an increase of T_1. This in turn is expected to increase the likelihood as well, provided that the quadratic approximation in (4) is 'sufficiently' accurate. The new reconstruction is $\vec{\tau}_{\text{new}} = \vec{\tau} + \Delta\vec{\tau}$. When choosing an appropriate projection model this update scheme can be applied on both transmission and emission data. The models and their corresponding update will be described in the following sections.

1) Transmission model: In transmission tomography, attenuation values μ_j, for all voxels j are reconstructed, thus $\tau_j = \mu_j$. The simplest projection model is given by:

$$\hat{y}_i(\vec{\mu}) = b_i \exp(-\sum_j l_{ij}\mu_j) \tag{8}$$

with i the projection ray index, j the pixel index and b_i the blank value for projection ray i. A more complex model includes a scatter term and resolution modeling.

$$\begin{aligned} \hat{y}_i(\vec{\mu}) &= \sum_n w_{in}\hat{\psi}_n + r_i && (9)\\ \hat{\psi}_i(\vec{\mu}) &= b_i \exp(-\sum_j l_{ij}\mu_j) && (10)\end{aligned}$$

$$\text{with } \sum_n w_{in} = 1 \text{ and } w_{in} = w_{ni},$$

modeling blurring with convolution kernel w_{in} and scatter with a known sinogram r_i. The first and second derivative of the likelihood become:

$$\begin{aligned} \frac{\partial L}{\partial \mu_j} &= \sum_i l_{ij}\hat{\psi}_i \sum_n \frac{\hat{y}_n - y_n}{\hat{y}_n} && (11)\\ \frac{\partial^2 L}{\partial \mu_j \partial \mu_h} &= -\sum_i l_{ij}\hat{\psi}_i \sum_n w_{in} && (12)\\ &\quad \left((1 - \frac{y_n}{\hat{y}_n})l_{ih} + \frac{y_n}{\hat{y}_n^2}\sum_m l_{mh} w_{nm}\hat{\psi}_m\right).\end{aligned}$$

We now apply the update given in equation (7):

$$\Delta\mu_j = \frac{\alpha_j \sum_i l_{ij}\hat{\psi}_i \sum_n \frac{\hat{y}_n - y_n}{\hat{y}_n}}{\sum_i l_{ij}\hat{\psi}_i \sum_n w_{in} M_n} \tag{13}$$

$$M_n = \left((1 - \frac{y_n}{\hat{y}_n})(\sum_h \alpha_h l_{ih}) + \frac{y_n}{\hat{y}_n^2}\sum_m (\sum_h \alpha_h l_{mh}) w_{nm}\hat{\psi}_m\right). \tag{14}$$

For particular choices of α_j, e.g. $\alpha_j = 1$, we can assume that $\sum_h \alpha_h l_{ih}$ is smooth and hardly modified by the blurring kernel w_{in}. This results in:

$$\Delta\mu_j = \frac{\alpha_j \sum_i l_{ij}\hat{\psi}_i \sum_n \frac{\hat{y}_n - y_n}{\hat{y}_n}}{\sum_i l_{ij}\hat{\psi}_i (\sum_h \alpha_h l_{ih}) \sum_n w_{in} \frac{\hat{y}_n^2 - y_n r_n}{\hat{y}_n^2}}. \tag{15}$$

For model (8) $w_{in} = \delta_{in}$ and $r_i = 0$ the update given in (15) becomes:

$$\Delta\mu_j = \frac{\alpha_j \sum_i l_{ij}(\hat{y}_i - y_i)}{\sum_i l_{ij}(\sum_k \alpha_k l_{ik})\hat{y}_i}. \tag{16}$$

The same algorithm is obtained by introducing an exact separable surrogate function for the likelihood, followed by

a quadratic approximation, as was proposed by Fessler et *al.* in [24].

The voxel weight α_j is a design parameter that influences the convergence. If the weight α_j for voxel j is increased relative to weight of the other voxels, the update (16) for voxel j is increased as well, because (16) has weight α_j in the numerator and a sum over all weights in the denominator. Choosing α_j equal to 1 for all voxels results in the MLTR algorithm described in [13]. Equal update weights lead to uniform convergence for all regions in the object. When the update weights are proportional to the attenuation, Eq.(16) becomes the so-called convex algorithm [14], which has faster convergence for highly attenuating regions at the expense of reduced convergence for regions with low attenuation.

In the experiments below the simplified update with $w_{in} = \delta_{in}$ and $r_i = 0$ will be used. The voxel weights α_j will be set to 1 to obtain MLTR.

2) Emission model: In emission tomography $\tau_j = \lambda_j$, the activity in voxel j. The model is given by:

$$\hat{y}_i(\vec{\lambda}) = \sum_j a_{ij}\lambda_j + r_i \tag{17}$$

where a_{ij} represents the probability that a photon (pair) emitted in pixel j will be detected in detector (pair) i, including attenuation and sensitivity for line of response i. Unlike for transmission model, resolution effects can directly be included as well in a_{ij}. For transmission tomography this is not possible due to the exponentiation operation which is nonlinear. r_i represents the contamination by randoms and/or scatter for pixel i.

The derivatives of the likelihood are:

$$\frac{\partial L}{\partial \lambda_j} = \sum_i a_{ij}\frac{y_i - \hat{y}_i}{\hat{y}_i} \tag{18}$$

$$\frac{\partial^2 L}{\partial \lambda_j \partial \lambda_h} = -\sum_i a_{ij}\frac{y_i}{\hat{y}_i^2}a_{ih}. \tag{19}$$

Assuming $y_i \simeq \hat{y}_i$ for the second derivative, the update for activity λ_j becomes:

$$\Delta\lambda_j = \frac{\alpha_j \sum_i a_{ij}\frac{y_i - \hat{y}_i}{\hat{y}_i}}{\sum_i a_{ij}\frac{1}{\hat{y}_i}\sum_h a_{ih}\alpha_h}. \tag{20}$$

The separable surrogate technique in [24] could also be applied to the model described in Eq. (17). This would result in the same update step (also under the assumption that $y_i \simeq \hat{y}_i$ for the second derivative).

The update weight α_j has the same role as for the transmission model. Setting $\alpha_j = 1$ for all j produces the NEGML algorithm of [11][28]. When $\alpha_j = \lambda_j$, Eq. (20) resembles the MLEM (maximum likelihood expectation maximization [7]) algorithm and becomes identical to it when $r_i = 0$.

Note that both for transmission and emission tomography, the sinogram $\sum_j l_{ij}\alpha_j$ or $\sum_j a_{ij}\alpha_j$ could be calculated beforehand when α_j is constant over all iterations. For $\alpha_j = 1$ for all j, this sinogram can be approximated by a single value, the diameter of the (transaxial) field of view. For PET only the geometrical part can be approximated by this value. Other effects included in a_{ij} like attenuation, should still be considered. This approximation is conservative since this is the maximal value found after projection of the field of view. The advantage of this approximation is that no extra sinogram needs to be stored.

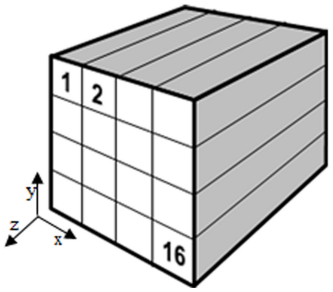

Fig. 1. Division of three dimensional reconstruction volume into sixteen patches: x and y are the transaxial directions, z is the axial direction: CT rotational axis or PET symmetry axis.

B. Patchwork reconstruction

The update weight α_j described in the previous section is the basis of the patchwork reconstruction. By choosing update weights equal to 1 for a certain subset of voxels and setting the weights to 0 for the rest the object, the convergence for that group of voxels is improved but there is no update for the voxels outside that region. One main iteration can then be defined as a number of subiterations where each time another group of voxels receives nonzero update weights. All voxels should be part of one single group. This principle was applied in the grouped coordinate algorithm [24] where each group consists of voxels spread over the reconstruction volume. This way the voxels have the lowest coupling, minimizing the denominator in Eq. (16) or (20). In the patchwork reconstruction, the group of voxels, called patches, will be groups of neighboring voxels. The different patches are determined by dividing the reconstruction volume in P equally sized image-blocks. The image volume is only subdivided in transaxial (x- and y-) direction and not for the axial or z-direction. The reason is that updating a voxel in one plane has only little or no influence on voxel values in other planes [25]. Therefore dividing the volume in axial direction will not further improve convergence. The division of a reconstruction volume into sixteen patches is shown in figure 1. Only the total projection estimate $\hat{y}_i$ for the whole image is stored. Each time a patch is updated, that update is forward projected and added to $\hat{y}_i$.

The order of the patch updates is chosen as indicated in figure 1. Other ordering schemes were investigated but had negligible influence on the result for the patch schemes presented in this study. However, when the patches become smaller, the algorithm becomes more similar to a (grouped) coordinate ascent algorithm and the order of updating the patches might become more important.

The gain in convergence of using patches can be estimated when considering the sum in denominator of Eq. (16) and (20). The difference in the update step will be determined by the change in $\sum_h l_{ih}\alpha_h$ or $\sum_h a_{ih}\alpha_h$. In the described

configuration, the convergence is expected to be proportional to the square root of the number of patches in the patch structure.

The patched update will be combined with ordered (projection) subsets. A reconstruction scheme using patches and projections subsets means that for one subset all patches will be updated and then the algorithm goes on to the next subset.

III. Materials

A. Transmission tomography

Three different sets of transmission data were used: circular CT, cone beam CT and helical CT. For the three transmission studies, resolution effects and scatter were ignored during reconstruction.

1) Circular CT: A circular scan was performed on a Siemens Sensation 16 system (part of the Biograph 16 PET/CT). A circular PMMA phantom with four aluminum inserts (two of 1 cm diameter and two of 3 cm diameter) was scanned with tube voltage 120 kV and tube current 300 mA. The collimation used for this scan was 2 x 1.00 mm. The reconstruction is two-dimensional by using only one of the slices. It was a 360° scan, with 1160 angles. The field of view was 50 cm and represented by a reconstruction grid of 512 x 512 pixels.

2) Cone beam micro CT: The second data set is a microCT scan of a mouse bone in a plastic tube filled with water. The scan is performed on a Skyscan 1172 microCT (cone beam), at 49 kV and 200μA with an 0.5 mm Al-filter. This was an 180 degree scan with 264 views. The detector pixel size was 11.57 μm and this was rebinned to a pixel size of 46.28 μm during reconstruction. The reconstruction grid is 552 $\times$ 552 $\times$ 158 with isotropic voxel size of 12.0 μm.

3) Helical CT: The last set is a helical whole body patient scan on a Siemens Sensation 16 system (CT part of Biograph 16). The mean effective tube current was 99 mA and the tube voltage 120 kV (DLP 476.8 mGy cm, scan length 975 mm). The table feed was 18 mm per rotation, the rotation time 0.5s per rotation and the collimation 16 x 1.50 mm. The reconstruction grid is 512 $\times$ 512 $\times$ 48, representing 1 mm $\times$ 1 mm $\times$ 3 mm voxels.

B. Emission tomography

One emission data set was used. It was a 3D 30 min. ^{18}F-FDG PET brain study acquired on a Siemens HR+ system. The system resolution was modeled by a 2D Gaussian with FWHM of 1.78 mm in the transaxial direction and 2.25 mm in the axial direction. The reconstruction grid was 144 $\times$ 144 $\times$ 63. The randoms contribution was corrected during acquisition using the delayed coincidence technique, scatter was ignored during reconstruction.

C. Implementation and system information

The projector used in this study is the distance driven projector [29]. The distance driven projector was chosen because it implements the backprojector as the exact transpose of the projector, which is important for most iterative reconstruction algorithms and because it is considered as more efficient compared to pixel and ray driven methods regarding memory access. The reconstruction software is written in IDL and C and is not optimized for parallel computing on CPUs or GPUs. The system used in this study is a Solaris 64bit workstation with an Intel E5440 CPU with 12 MB of L2 cache and 32 GB RAM.

IV. Convergence analysis

This study investigates the gain in convergence by using a patched reconstruction update. The convergence for different reconstructions schemes will be compared: reconstructions with and without projection subsets (classical ordered subsets reconstruction [16]), with and without patches and combinations of projections subsets and patches. The number of patches will be 1 (regular reconstruction), 4, 16 or 64, the number of projections subsets will be 1, 20, or 40 for CT and 1, 10, and or 20 for PET.

To determine the degree of convergence of a particular reconstruction, a measure based on the quadratic difference with a reference image is calculated:

$$\text{Quadratic difference} = \sum_j \left(\tau_j^{\text{ref}} - \tau_j^{\text{recon}}\right)^2 \qquad (21)$$

with τ_j^{recon} the image value in the reconstruction under investigation and τ_j^{ref} the corresponding image value in the reference image. This reference image is obtained by a long iterated reconstruction with decreasing number of subsets to prevent limit cycle solutions [16]. The particular iteration scheme can be found in table I as well as the equivalent number of iterations without subsets. The equivalent number of iterations is calculated based on the assumption that N iterations using M projections subsets per iteration is equivalent to N $\times$ M iterations without subsets. The degree of convergence of a certain reconstruction will be measured by the quadratic difference of that certain reconstruction with the reference image. The reference images are shown in figure 2.

For comparing different reconstruction schemes with and without patches and/or subsets, we calculate the convergence level obtained at 200 standard iterations. Next, we calculate when the same convergence level is reached for schemes with subsets and/or patches. This value is calculated based on linear interpolation of the convergence curves, and therefore a non-integer number of iterations may be obtained.

Sequentially updating results in faster convergence of higher frequencies [22], [24]. The more patches are used, the more pronounced this effect will be. Moreover, a convergence difference between neighboring patches can arise, depending on the order of patch updates. Both effects can give rise to artifacts in an early stage of the reconstruction. Some of these artifacts, especially those at patch borders can persist during the rest of the reconstruction. The artifacts appear as very thin lines (one or two voxels) at the patch borders and have zero means in most projections, through which it can take very long to reduce them. This effect can be avoided by underrelaxation of the first updates. A relaxation is applied to the first five updates, i.e. for the first five subsets or, if no subsets are used,

TABLE I
RECONSTRUCTION SCHEME TO OBTAIN THE REFERENCE RECONSTRUCTION

Data sets	Reconstruction scheme	Equivalent iterations
Circular CT	20 × 232 + 20 × 116 + 20 × 58 + 20 × 29 + 20 × 1	8720
Cone beam CT	50 × 66 + 50 × 44 + 50 × 22 + 50 × 11 + 50 × 1	7200
Helical CT	20 × 232 + 20 × 116 + 20 × 58 + 20 × 29 + 20 × 1	8720
PET	50 × 36 + 50 × 27 + 50 × 18 + 50 × 9 + 50 × 1	4550

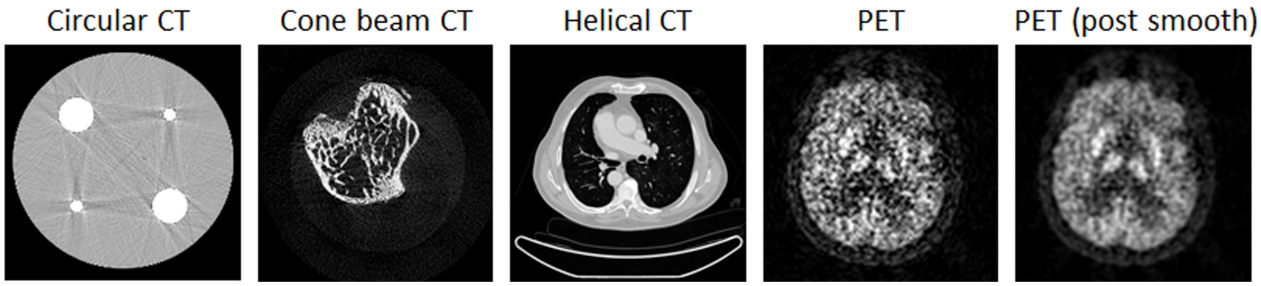

Fig. 2. The reference images after reconstruction as indicated in table I. The PET image is shown with and without post smoothing (Gaussian with full width half maximum 1.5 voxels).

the first five iterations. The underrelaxation was chosen such that the update step is the same as in a non-patched update. In practice iterative reconstruction is often started from a filtered backprojection (FBP) reconstruction. In that case no relaxation is needed since the convergence level is already relatively high. Since convergence is reached in fewer iterations when starting with FBP images, that approach was not used in this study to facilitate the observation of convergence differences. However, we verified that very similar results are obtained when starting from FDK images in the microCT study (results not shown). For the results shown below, the reconstructions were started from a uniformly filled (rough) body contour (filled with water attenuation for CT, and with the mean of the sinogram divided by the image size for PET).

V. RESULTS

Table II, III, and IV show the convergence results for the circular CT, cone beam CT and helical CT data sets respectively. Table V shows the convergence results for the PET data set. The PET study was repeated, evaluating the difference with the reference image after post-smoothing. The results were very similar (not shown). The numbers in the table represent the number of iterations necessary to achieve the same level of convergence as obtained by 200 iterations without subsets and patches. For instance, the convergence level of the microCT data at 200 iterations without subsets and patches, can be obtained in 2.8 iterations when 20 subsets and 16 patches are used. The reconstruction with 20 subsets but without patches reaches this level in 9.8 iterations. This is a convergence gain of almost 4.

The data for the helical CT scan with subsets and 4 or 16 patches is plotted in figure 3. The left curve shows convergence per iteration, the right curve convergence per time. In addition the calculation time per iteration for the helical CT data set is given in table VI. The convergence level as a function of iteration demonstrates the same results as in the table. Using more patches and/or subsets increases the convergence.

TABLE II
CONVERGENCE RESULTS FOR CIRCULAR CT

Patches	1 subset	20 subsets	40 subsets
1	200.0	10.0	5.0
4	104.3	5.1	2.7
16	54.5	2.8	1.5
64	29.6	1.7	0.1

TABLE III
CONVERGENCE RESULTS FOR CONE BEAM - MICRO CT

Patches	1 subset	20 subsets	40 subsets
1	200.0	9.8	5.0
4	103.2	5.0	2.7
16	53.1	2.8	1.6
64	29.0	1.8	1.2

TABLE IV
CONVERGENCE RESULTS FOR HELICAL CT

Patches	1 subset	20 subsets	40 subsets
1	200.0	9.6	5.4
4	101.3	5.0	2.8
16	52.2	2.8	1.6
64	27.7	1.6	0.8

TABLE V
CONVERGENCE RESULTS FOR PET

Patches	1 subset	10 subsets	20 subsets
1	200.0	20.0	10.0
4	102.5	10.3	5.1
16	53.7	5.4	2.7
64	29.4	3.0	1.6

TABLE VI
AVERAGE CALCULATION TIME PER ITERATION FOR THE HELICAL CT DATA SET.

Patches	1 subset	20 subsets	40 subsets
1	1.14 h	1.18 h	1.15 h
4	0.67 h	0.64 h	0.64 h
16	0.64 h	0.66 h	0.64 h
64	0.67 h	0.67 h	0.68 h

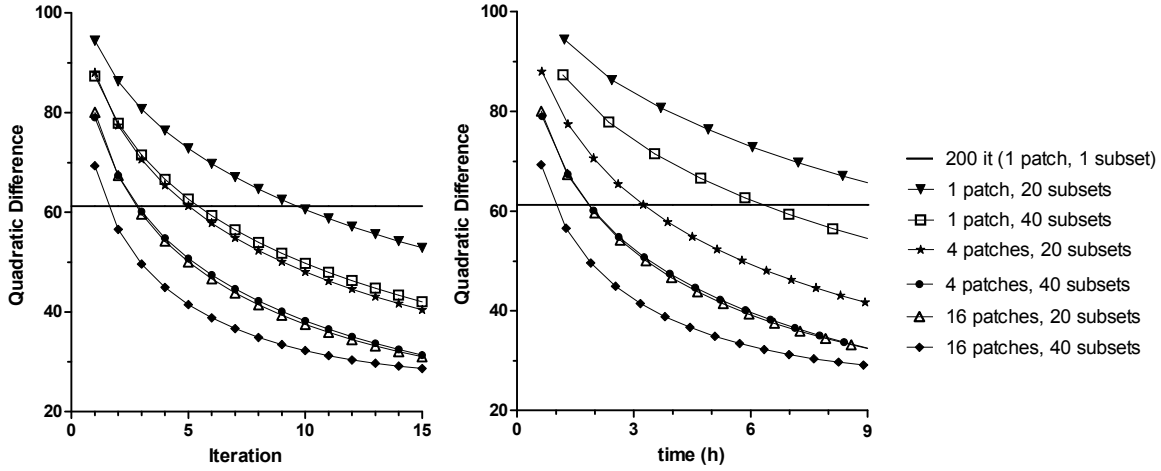

Fig. 3. Helical CT data set: convergence per iteration (left) and per time (right). The horizontal line is a reference line representing the convergence level obtained after 200 iterations without patches or subsets.

The curves *1 patch, 40 subsets* and *4 patches, 20 subsets* coincide, also *4 patches, 40 subsets* coincides with *16 patches, 20 subsets*. Only one of these pairs also coincides in the convergence versus time curve. While *4 patches, 20 subsets* has the same convergence per iteration as *1 patch, 40 subsets*, less time per iteration is required. No improvement in time per iteration is observed when increasing the number of patches from 4 to 16. A visual example of the different convergence can be found in figure 4 and figure 5 where reconstructions with the different patch schemes are depicted.

VI. Discussion

The popularity of iterative reconstruction in clinical practice is increasing. The main drawback is the computational burden of most iterative techniques. The patchwork reconstruction presented in this work aims to reduce this problem. By using a sequential update of image parts the convergence of the reconstruction is accelerated. The likelihood is an obvious measure of convergence. However, we found that often the images still change significantly with iteration, while the corresponding changes to the likelihood are lost due to the finite numerical precision. For that reason, we have used the difference with a reference reconstruction as a measure of convergence. Using the proposed patch structure, an acceleration of the square root of the number of patches was expected. The patchwork scheme was also combined with an ordered subset scheme where we expect an acceleration by the number of subsets. The experimental convergence results for all four modalities show a convergence gain that agrees well with these expectations. Some values are slightly higher than expected, which may be due to the linear interpolation which gives a slight overestimation for a concave monotonic decreasing function. In cases without subsets the underrelaxation of the first five updates also influences the convergence result. In practice one would often use subsets while starting from FBP which makes relaxation unnecessary. A visual perception of the improved convergence is given in figures 4 and 5. The same iteration scheme (iterations $\times$ subsets) with patches results in sharper edges, more details and fewer metal artifacts.

It might seem beneficial to further subdivide the image into more patches to increase the convergence even more. In the patch structure as described above this is most probably not advisable. By using relatively large image blocks with connected neighboring voxels, the patchwork structure enables the use of regular projector and backprojector software, which is usually optimized for dealing with large series of neighboring image voxels and projections pixels. When patches become smaller the advantage of using these projectors becomes smaller. Moreover, since the different regions are sequentially updated, the level of convergence is different for patches being updated first compared to patches that are updated later on. This difference becomes more prominent when there are more patches which also implies that it would take longer (i.e. more iterations) to obtain a relatively uniform convergence over the image. Besides, when more image parts are sequentially updated, the computational overhead becomes more important, which increases the required computation time to update the whole volume. Finally, when using relatively large patches, the patchwork structure can be combined with ordered projections subsets, giving a two-sided acceleration. The smaller the patches become, the higher the change that the subset balance is strongly violated. More patches would eventually lead to a similar algorithm as ABCD in [25] where all axial rows are sequentially updated and which requires dedicated projection/backprojection software.

For large data sets, the patch reconstruction showed another advantage. By applying the patch update, i.e. by sequentially using subparts of the reconstruction volume, a decreased computation time per iteration was observed (figure 3 and table VI). The time required for one iteration with patches is less then 60% of the time required for one regular iteration. We believe this is because a patch update reduces the direct memory requirements of the system since less data need to be handled at the same time. The full reconstruction volume of the helical CT data set is 48 MB. This size has to be divided by the number of patches to obtain the size of an individual patch. Probably when using patches, the reconstruction volume fits better in the cache memory (the L2 cache of our system is

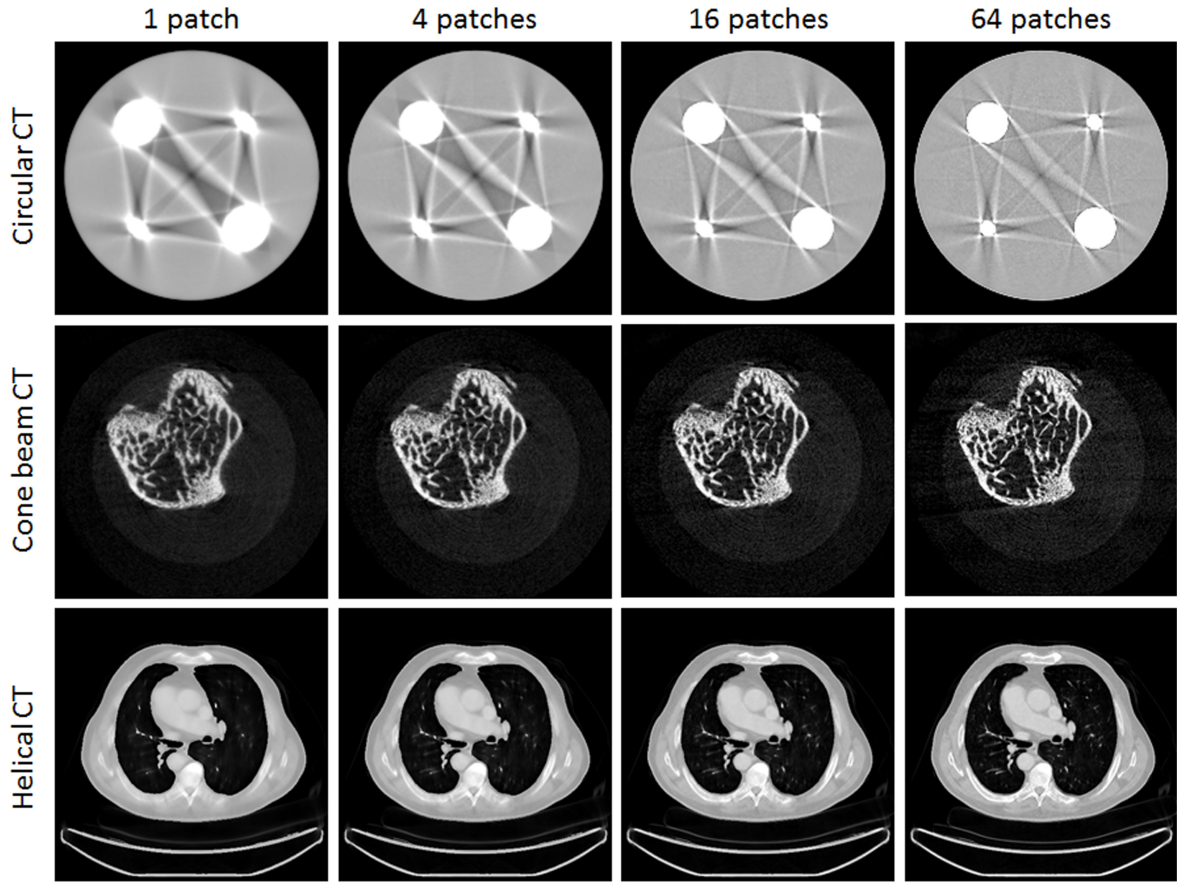

Fig. 4. MLTR reconstructions for circular, cone beam, and helical CT for 2 iterations and 40 subsets. MLTR without patches and with 4, 16, or 64 patches.

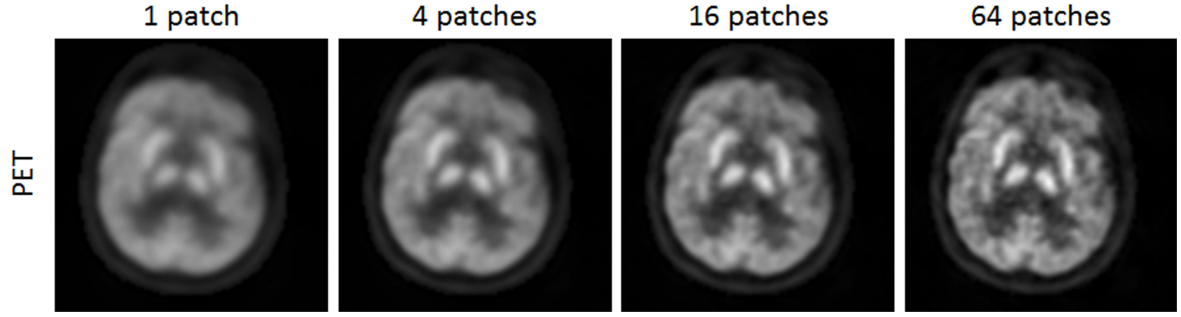

Fig. 5. NEGML reconstructions for PET for 2 iterations and 20 subsets. NEGML without patches and with 4, 16, or 64 patches. Post-smoothed with a Gaussian kernel with a full width half maximum of 1.5 voxels.

12 MB, see section III-C). Of course this effect is dependent on the system. This advantage disappears when data sets are small like for the circular CT.

Nonuniform convergence of certain image parts can also be used on purpose. In our previous work on metal artifact reduction [26], the image was divided in several metal patches and one patch without metals. Since the metal patches were smaller, their update was larger, giving an increased convergence compared to the non-metal patch that receives an update with equal step size as when updating the whole image simultaneously. This resulted in faster and more efficient artifact reduction. In this work the projection model for each patch could be chosen separately. The voxel size and energy model could be different for the patches. Using different voxels size over the image also allows for region-of-interest reconstruction [30], [31], [32], [33], [34], [35]. The region of interest would be one patch, reconstructed with small voxels, the remaining region is another patch, reconstructed with larger pixels. This gives two advantages, the background is reconstructed with larger voxels which lowers the computation time and the patch structure is applied which will increase the convergence of the region-of-interest patch (especially when it is considerably smaller than the total volume). Besides, the patchwork reconstruction combined with non-uniform resolution modeling was used for digital breast tomosynthesis in Michielsen et *al.* [36].

VII. Conclusion

We have presented a simple maximum likelihood gradient ascent algorithm to obtain update formulas for both emission and transmission tomography reconstruction. The algorithm includes a design parameter α_j that can alter relative convergence of particular image regions or image voxels. A very similar approach was previously used by Fessler et *al.* [24] to derive a grouped coordinate ascent algorithm for transmission tomography. By sequentially setting this parameter to 1 for a certain region and 0 for the rest of the object, a sequential 'patchwork' reconstruction can be obtained. Unlike other sequential algorithms, the different regions or patches that are sequentially updated are rather large groups of neighboring voxels. We have shown that this still gives a considerable gain in convergence while projection and backprojection can be done efficiently with conventional (back)projectors. Moreover, the larger groups allow for combining sequential voxel updates with ordered (projection) subsets. For very large data sets, the scheme also reduces the system memory requirements, which may further reduce the computation time.

References

[1] M. J. Willemink, T. Leiner, P. A. de Jong, L. M. de Heer, R. A. J. Nievelstein, A. M. R. Schilham and R. P. J. Budde, "Iterative reconstruction techniques for computed tomography part 2: initial results in dose reduction and image quality," *Eur. Radiol.* - published online on 16 January 2013.

[2] A. K. Hara, R. G. Paden, A. C. Silva, J. L. Kujak, H. J. Lawder and W. Pavlicek, "Iterative Reconstruction Technique for Reducing Body Radiation Dose at CT: Feasibility Study," *Gastrointestinal Imaging*, vol. 193, pp. 764-771, 2009.

[3] M. Katsura, I. Matsuda, M. Akahane, J. Sato, H. Akai, K. Yasaka, A. Kunimatsu and K. Ohtomo, "Model-based iterative reconstruction technique for radiation dose reduction in chest CT: comparison with the adaptive statistical iterative reconstruction technique," *Eur. Radiol.*, vol. 22, pp 16131623, 2012.

[4] E. Y. Sidky, Y. Duchin, X. Pan and C. Ullberg, "A constrained, total-variation minimization algorithm for low-intensity x-ray CT," *Med. Phys.*, vol. 38, pp. S117-S125, 2011.

[5] J. Wang, T. Li and L. Xing, "Iterative image reconstruction for CBCT using edge-preserving prior," *Med. Phys.*, vol. 36, no. 1, pp. 252-260, 2009.

[6] P. J. La Rivière, "Penalized-likelihood sinogram smoothing for low-dose CT," *Med. Phys.*, vol 32, no. 6, pp. 1676-1683, 2005.

[7] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *Trans. Med. Im.*, vol. 1, no.2, pp. 113-122, 1982.

[8] M. E. Daube-Witherspoon and G. Muehllehner, " An iterative image space reconstruction algorithm suitable for volume ECT," *IEEE Trans. Med. Imag.*, vol. 5, no. 2, pp. 61-66, 1986.

[9] G. E. B. Archer and D. M. Titterington, "The iterative image space reconstruction algorithm (ISRA) as an alternative to the EM algorithm for solving positive linear inverse problems. *Statistica Statica*, vol. 5, pp. 77-96, 1995.

[10] A. R. De Pierro. "In the relation between ISRA and the EM algorithm for positron emission tomography, *IEEE Trans. Med. Imag.*, vol. 12, no. 2, pp. 328-333, 1993.

[11] J. Nuyts, S. Stroobants, P. Dupont, S. Vleugels, P. Flamen and L. Mortelmans, "Reducing loss of image quality due to the attenuation artifact in uncorrected PET whole body images," *J. Nucl. Med.*, vol. 43, pp. 1054-1062.

[12] C. Byrne, "Iterative algorithms for deblurring and deconvolution with constraints,, *Inverse Problems* vol. 14, pp. 1455-146, 1998.

[13] J. Nuyts, B. De Man, P. Dupont, M. Defrise, P. Suetens and L. Mortelmans, "Iterative reconstruction for helical CT: a simulation study," *Phys. Med. Biol.*, vol. 43, pp. 729-737, 1998.

[14] K. Lange and F. A. Fessler, "Globally convergent algorithms for maximum a posteriori transmission tomography," *Tran. Im. Proc.*, vol. 4, no. 10, pp. 1430-1438, 1995.

[15] P. Gilbert, "Iterative methods for the three-dimensional reconstruction of an object from projections," *J. Theor. Biol.*, vol. 36, pp. 105-177, 1971.

[16] H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *Trans. Med. Im.*, vol. 13, no. 4, pp. 601-609, 1994.

[17] H. Erdoğan and J. A. Fessler, "Ordered subsets algorithms for transmission tomography," *Phys. Med. Biol*, vol. 44, pp. 2835-2851, 1999.

[18] C. Kamphuis and F. J. Beekman, "Accelerated iterative transmission CT reconstruction using an ordered subsets convex algorithm," *IEEE Trans. Med. Imag*, vol. 17, no. 6, pp. 1101-1105, 1998.

[19] R. Gordon, R. Bender and G. T. Herman, "Algebraic Reconstruction Techniques (ART) for three-dimensional electron microscopy and x-ray photography, *J. Theor. Biol.*, vol. 29, pp. 471-481, 1970.

[20] G. N. Houndsfield, "A method and apparatus for examination of a body by radiation such as X or Gamma radiation," Patent Specification 1283915, London, England, 1972.

[21] A. H. Andersen and A. C. Kak, "Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm," *Ultrasonic Imaging*, vol. 6, pp. 81-94, 1984.

[22] K. Sauer and C. Bouman, "A local update strategy for iterative reconstruction from projections," *IEEE Trans. Sign. Proc.*, vol. 41, no. 2, pp. 534-548, 1993.

[23] J. A. Fessler and A. O. Hero, "Space-alternating generalized expectation maximization algorithm," *IEEE Trans. Sign. Proc.*, vol. 42, no. 10, pp. 2664-2677, 1994.

[24] J. A. Fessler, E. P. Ficaro, N. H. Clinthorne and K. Lange , "Grouped-coordinate ascent algorithm for penalized-likelihood transmission image reconstruction," *IEEE Trans. Med. Imag.*, vol. 16, no. 2, pp. 166-175, 1997.

[25] J. A. Fessler and K. Donghwan, "Axial block coordinate descent (ABCD) algorithm for X-ray CT image reconstruction," *11th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, Potsdam, Germany, 2011, pp. 262-265.

[26] K. Van Slambrouck and J. Nuyts. "Metal artifact reduction in computed tomography using local models in an image block-iterative scheme," *Med. Phys.*, vol. 39, no. 11, pp. 7080-7093, 2012.

[27] T. Benson, B. De Man, L. Fu, J.-B. Thibault, "Block-based Iterative coordinate descent," *IEEE Nucl Sci Symp Conf Record*, Knoxville, TN, USA, 2010, pp. 2856-2859.

[28] J. Nuyts, S. Stute, K. Van Slambrouck, F. van Velden, R. Boellaard, and C. Comtat, "Maximum-likelihood reconstruction based on a modified Poisson distribution to reduce bias in PET," *IEEE Nucl Sci Symp Conf Record*, Valencia, Spain, 2011, pp. 4337-4341.

[29] B. De Man and S. Basu, "Distance-driven projection and backprojection in three dimensions." *Phys. Med. Biol.*, vol. 49, pp. 2463-2475, 2004.

[30] B. Hamelin, Y. Goussard, J. Dussault, G. Cloutier, G. Beaudoin and G. Soulez, "Design of iterative ROI transmission tomography reconstruction procedures and image quality analysis," *Med. Phys.*, vol. 37, no. 9, pp. 4577-4589, 2010.

[31] B. Hamelin, Y. Goussard, D. Gendron, J. Dussault, G. Cloutier, G. Beaudoin and G. Soulez, "Iterative reconstruction of real data with metal artifact reduction," *IEEE Int. Symp. Bio. Im.*, Paris, France, 2008, pp. 1453-1456.

[32] Z. Yu, J. Thibault, C. A. Bouman, K. D. Sauer and J. Hsieh, "Edge-localized iterative reconstruction for computed tomography," *Proc. Intl. Mtg. on Fully 3D image Recon. in Rad. and Nuc. Med.*, Beijing, China, 2009, pp. 255-258.

[33] L.J. Meng and N. Li, "Non-uniform object-space pixelation (NUOP) for penalized maximum-likelihood image reconstruction for a single photon emission microscope system," *IEEE Trans. Nucl. Sci.*, vol. 56, no. 5, pp. 2777-2788, 2009.

[34] J. S. Maltz, "Optimal time-activity basis selection for exponential spectral analysis: application to the solution of large dynamic emission tomographic reconstruction problems," *IEEE Trans. Nucl. Sc.*, vol. 49, no. 4, pp. 1452-1464, 2001.

[35] J. G. Brankov and M. N. Wernick, "Tomographic image reconstruction using content-adaptive mesh modeling," *Proc. IEEE Intl. Conf. on Image Processing*, Thessaloniki, Greece, 2001, pp. 690-693.

[36] K. Michielsen, K. Van Slambrouck, A. Jerebko and J. Nuyts, "Patchwork Reconstruction with Resolution Modeling for Digital Breast Tomosynthesis, *Med. Phys.*, vol. 40, no. 3, pp. 031105 (1-10), 2012.