Accelerated convergence with image-block iterative reconstruction

Katrien Van Slambrouck and Johan Nuyts

***Abstract*—It is known that the convergence of iterative algorithms is better when updating pixels sequentially rather than simultaneously, but sequential updating is more computationally demanding and less practical for implementation issues. We propose to divide the image in a matrix of equally sized cuboids and update them separately as groups of pixels in a grouped coordinate ascent algorithm. We apply this to transmission and emission tomography. For both applications we see a substantial increase in convergence per iteration.**

I. Introduction

Statistical iterative reconstruction algorithms have several advantages over filtered backprojection (FBP) based algorithms. They explicitly take into account the Poisson nature of the measurement, which results in better noise characteristics. They usually give better results in case of incomplete or truncated data and they offer more flexibility for defining a reconstruction model. The main disadvantage is the increased computation time.

It is known that the convergence of iterative algorithms is better when pixels are updated sequentially rather than simultaneously. In [1] and [2], Fessler et al. proposed a grouped coordinate ascent algorithm for transmission tomography where groups of pixels are updated sequentially. The optimal way of choosing these groups is by selecting pixels that have a low coupling, which means that the update in one pixel is nearly independent of the values of the other pixels in the group. To achieve this low coupling, each group should consist of a grid of pixels with maximum inter-pixel distance. Updating these groups in less practical than applying a simultaneous update where one can use a regular projector and backprojector.

In our work on local models for artefact reduction in computed tomography [3], we created little subareas around metals to change the reconstruction model. The different subareas were updated separately and sequentially, as a group of pixels in [1]. We observed a clearly improved convergence for these smaller areas. We called these areas patches. Our work presented here, is a more elaborated study on the influence of patches or image-blocks on the convergence of iterative algorithms.

We applied a grouped coordinate, *patched*, algorithm on transmission (CT - computed tomography) and emission data (PET - positron emission tomography).

Dept. of Nuclear Medicine, K.U.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).

II. Methods

All reconstruction schemes used in this study are maximum likelihood (ML) based. In the first part of this section the algorithms for transmission and emission data are explained. The second part explains how the use of the so-called *patches* can improve the convergence.

A. Maximum likelihood reconstruction

The study on the reconstruction with patches is based on statistical iterative reconstruction with Poisson likelihood:

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

with i the index of the projection lines, y_i the measurement and $\hat{y}_i$ the estimation and $\vec{\tau}$ the attenuation coefficients in case of transmission tomography and the activity for emission tomography.

To obtain a gradient ascent algorithm, we calculate a truncated series expansion around the current reconstruction ([4], [5]):

$$\begin{aligned} &L(\vec{\tau} + D(\alpha)\Delta x) \simeq T_1(\vec{\tau}, \Delta x) \\ &= L(\vec{\tau}) + \sum_j \frac{\partial L}{\partial \tau_j}\alpha_j \Delta x_j + \frac{1}{2}\sum_{j,k} \frac{\partial L^2}{\partial \tau_j \partial \tau_k}\alpha_j \alpha_k \Delta x_j \Delta x_k \\ &\geq L(\vec{\tau}) + \sum_j \frac{\partial L}{\partial \tau_j}\alpha_j \Delta x_j + \frac{1}{2}\sum_{j,k} \frac{\partial L^2}{\partial \tau_j \partial \tau_k}\alpha_j \alpha_k (\Delta x_j)^2, \end{aligned} \tag{2}$$

where D is a diagonal matrix with $D_{jj} = \alpha_j$. The last expression is a surrogate function for T_1. The update for $\vec{\tau}$ is obtained by maximizing the surrogate function, with $\Delta\tau_j = \alpha_j \Delta x_j$ this becomes:

$$\tau_j^{\text{new}} = \tau_j^{\text{old}} - \frac{\left.\frac{\partial L}{\partial \tau_j}\right|_{\vec{\tau}^{\text{old}}}}{\sum_h \left.\frac{\partial^2 L}{\partial \tau_j \partial \tau_h}\right|_{\vec{\tau}^{\text{old}}}}. \tag{3}$$

The final update for transmission and emission data is explained in the following.

1) Transmission data: In transmission tomography $\hat{y}_i$ is given by $\hat{y}_i = b_i \exp\left(-\sum_j l_{ij}\mu_j\right)$ with b_i the blank value for projection line i and μ_j the linear attenuation in pixel j. Maximizing (2) with this model for $\hat{y}_i$ gives:

$$\mu_j^{\text{new}} = \mu_j^{\text{old}} + \frac{\alpha_j \sum_i l_{ij}(y_i - \hat{y}_i)}{\sum_i l_{ij}\hat{y}_i \sum_h l_{ih}\alpha_h}. \tag{4}$$

978-1-4673-0120-6/11/$26.00 ©2011 IEEE

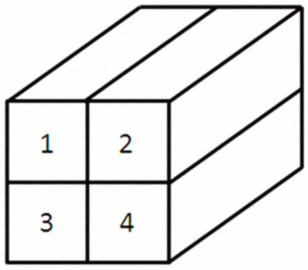


Fig. 1. Three dimensional reconstruction volume divided into four patches

With $\alpha_j = 1$ for every j this is MLTR (maximum likelihood for transmission [4]), with $\alpha_j = \mu_j^{old}$ this is the convex algorithm [1], [6]. Note that in [1], Fessler et al. introduced a parameter α_{ij} with similar function as α_j described here.

2) Emission data: For emission tomography $\hat{y}_i$ becomes $\hat{y}_i = \sum_j a_{ij}\lambda_j$ with a_{ij} the sensitivity of detector i for activity in voxel j and λ_j the activity in voxel j.

Maximizing (2) by assuming that $y_i \simeq \hat{y}_i$ for the second derivative yields the update step:

$$\lambda_j^{\text{new}} = \lambda_j^{\text{old}} + \frac{\alpha_j \sum_i a_{ij} \frac{(y_i - \hat{y}_i)}{\hat{y}_i}}{\sum_i l_{ij} \frac{1}{\hat{y}_i} \sum_h a_{ih}\alpha_h}. \quad (5)$$

With $\alpha_j = 1$ for every j this is a NEGML-update as described in [5], with $\alpha_j = \lambda_j^{old}$ we obtain the MLEM-update (maximum likelihood expectation maximization - [7]).

B. Patched reconstruction

A patched reconstruction is a reconstruction where the reconstruction volume is divided into several equally sized cuboids or patches. Figure 1 shows an example of a three-dimensional volume divided into 4 patches.

During the reconstruction the patches will be updated separately and sequentially as a group of pixels in a grouped-coordinate algorithm [1]. A patch-update can be considered as an update with $\alpha_j \geq 0$ for voxels belonging to the patch and $\alpha_j = 0$ for other voxels. The improved convergence is partly due to the sequential update but the main reason is the increased step size of the update step. In equations (4) and (5) the sum $\sum_h l_{ih}\alpha_h$ or $\sum_h a_{ih}\alpha_h$ in the denominator will be smaller and the step size for the update will be bigger for smaller patches. Because voxels have only a weak (or no) coupling in axial direction, we only split up the reconstruction volume transaxially as in figure 1. The patched reconstruction is based on a distance driven (back)projector [8].

III. Convergence analysis

In this work we investigate the influence of the use of patches on the convergence of different algorithms. The patches are combined with different numbers of subsets. The degree of convergence of the different reconstruction schemes will be calculated. The quadratic difference with a reference image is chosen to be a measure for convergence. This reference image is a long iterated standard reconstruction. We calculate the convergence level for an one-patch-one-subset reconstruction that iterated for 100 or 200 iterations. Next, we calculate for which iteration number a reconstruction scheme with patches and/or subsets achieves the same convergence level. This is done by linear interpolation of the convergence curves, and therefore a non-integer number of iterations may be obtained.

Sequentially updating the patches gives some convergence difference between the patches for the first iterations (or subsets). This can result in visible edges between the patches, which persist in subsequent iterations because they are very thin and have zero mean in most projections. To minimize this effect we apply an underrelaxation to the first five subsets (or to the first iteration if no subsets are used). The underrelaxation was chosen such that it is the same as a non-patched update.

IV. Experiments

A. CT

The reconstruction for the transmission data is done by MLTR which means that α_j equals 1 or 0.

Three data sets of transmission data were used. The first is a sequential CT scan of a PMMA (polymethyl methacrylate) circular phantom with four aluminium inserts. The scan was performed on a Siemens Sensation 16 (part of Biograph 16 PET/CT) with collimation 2 x 1.00, 120 kV and 300 mA. A full 360 degree scan with 1160 views was used. A two-dimensional reconstruction is performed for one of the slices. The reconstruction grid was 512 x 512 with isotropic pixel size 0.97 mm. The iteration scheme for obtaining the reference image was: MLTR, 50x116 + 50x58 + 50x29 + 50x1 (iterations x subsets).

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 reconstruction grid is 200x200x22 with isotropic voxel size of 33.4 μm. The reference image for the microCT was: MLTR, 100x50 + 100x25 + 100x13 + 100x1 (iterations x subsets). For this data set we used a field-of-view mask, all voxels outside the field of view (reconstruction domain covered by all projections) are set to zero after each iteration.

The last CT scan is a patient whole body scan, minimal dose (11 mAs, DPL 57.8 mGy cm, scan length 975 mm) on a Siemens Sensation 16 (part of Biograph 16 PET/CT, helical scan, 0.42 s/rotation, table feed 24 mm per rotation and collimation 16 x 0.75 mm). The projection data were rebinned with a factor of two for the detector elements and the views. The final reconstruction grid consisted of 2 mm isotropic voxels. The iteration scheme for the reference image was: MLTR, 50x116 + 50x58 + 50x29 + 50x1 (iterations x subsets).

B. PET

Here we apply the NEGML scheme (α_j equals 1 or 0).

One emission data set was used. It was a 3D 30 min. FDG PET brain study on a Siemens HR+ system. The system resolution was modelled by a 2D Gaussian with FWHM of 1.78 mm in the transaxial direction and 2.25 mm in the axial direction. The reference reconstruction is an 100x36 +

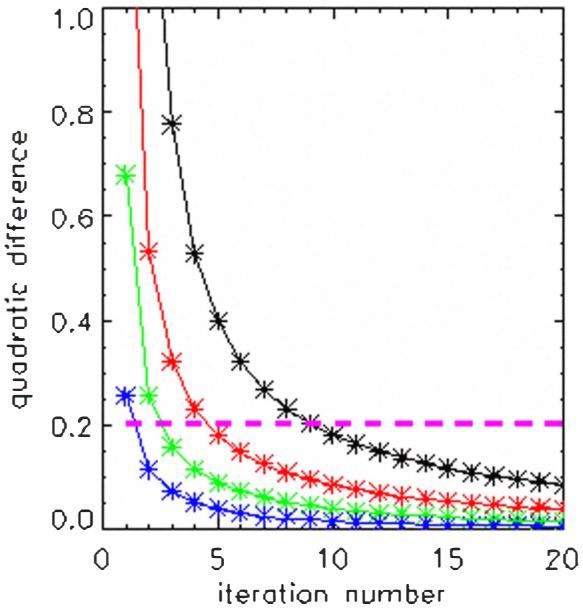


Fig. 3. Convergence curves for sequential CT. Reconstruction with regular MLTR for 20 (black) and 40 (red) subsets and reconstruction with 16 patches for again 20 (green) and 40 (blue) subsets. The pink horizontal line represents the convergence degree achieved after 200 regular one-subset MLTR iterations.

100x24 + 100x12 + 100x1 NEGML reconstruction (iterations x subsets). Also here a field-of-view mask is used.

V. Results

Figure 2 shows the convergence analysis for the different transmission data sets. On the left a table is shown with the iteration number at which the same convergence level is achieved as in the one-subset-one-patch case. On the right an example of the obtained images at 3 iterations with 40 subsets is shown without patches and with 16 patches. In all cases an improved convergence can be seen when using the patches.

An example of a convergence curve is given in figure 3 for the sequential CT data set. The figure shows the convergence curve for regular MLTR with 20 and 40 subsets and a reconstruction with 16 patches with again 20 and 40 subsets. The horizontal dashed line is drawn at the convergence degree achieved at 200 iterations of regular one-subset MLTR. The other curves achieve the same convergence level at respectively 9.0, 4.7, 2.6 and 1.4 iteration as can also be seen in the table. The curves are obtained with simple linear interpolation between the points for each iteration.

Similar results are obtained for the emission PET data set in figure 4. The table on the left compares the degree of convergence. On the right a reconstruction with and without patches is shown. As a comparison, a regular MLEM (here OSEM) reconstruction for the same iteration scheme is also given.

VI. Discussion

In this study we investigated the influence of a sequential updating algorithm using image-blocks. The convergence is compared with regular simultaneous reconstructions. Also, the combination with ordered projection subsets was examined.

The first rows of the tables show the acceleration that is achieved when using projection subsets. The improvement is, as expected, proportional to the number of subsets.

Theoretically the gain in convergence of using patches is proportional to the square root of the number of patches, as the field of view for each patch $\sum_h l_{ih}\alpha_h$ or $\sum_h a_{ih}\alpha_h$ in (4) and (5) becomes smaller with the same factor. The results in figure 2, 3 and 4 follow this predicted behaviour. The reconstructions on the right also show the improved convergence when patches are used, resulting in sharper image, better visual contrast and less artefacts.

Note that linear interpolation slightly underestimates the iteration number where a particular degree of convergence is obtained.

An important advantage of simultaneously updating all pixels rather than sequentially, is that one can use a conventional efficient projector and backprojector. The sequential update of the image blocks as presented here, keeps the possibility of using a regular projector and backprojector. In an efficient implementation the extra time for accessing the memory should be limited. However, when using subsets an extra projection per iteration is necessary.

A disadvantage of using the sequential update is the difference in convergence of the different patches, which can be seen in early iterations and results in visible patch edges in later iteration that are hard to remove (because they are very thin and have zero mean in most projections). We applied an underrelaxed update step for the first 5 subsets as explained in section III to reduce this effect. Still, for a low number of iterations (subsets) some edges can be seen. The importance of the edges depends on the final application. When CT is used for attenuation correction, this is definitely not a problem. In radiology examinations this could possibly be more important, although if there are edges, one knows exactly where they can be located and they are unlikely to be confused with body structures. For PET this causes less problems because of the high noise level and the often used post-smoothing.

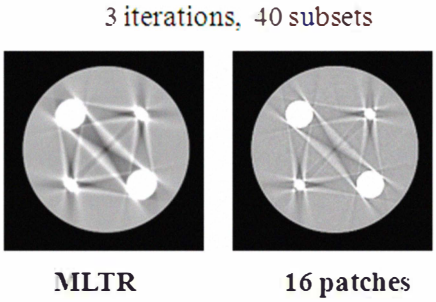
The convergence gain seems to be limited when combining many patches with many subsets, as can be seen in the second table in figure 2 for micro CT. The cause of this problem needs some further investigation.

In the reconstruction for PET and micro CT data, a field-of-view mask was used. This was done because the combination of subsets and patches is not stable outside the field-of-view. This is possibly because the subset balance is violated for the update of individual patches [9]. When using only subsets or only patches, the problem disappears.

When comparing the NEGML reconstruction with MLEM we see that the NEGML algorithm converges slower. This is because MLEM makes use of an 'intrinsic patch system'. By putting $\alpha_j = \lambda_j$, α_j will become very small or zero in the background. The same holds for the convex algorithm for transmission data. MLEM and the convex algorithm result in an improved convergence for regions with higher activity/attenuation values but the convergence of low value pixels is getting infinitely slow when their value is approaching zero.

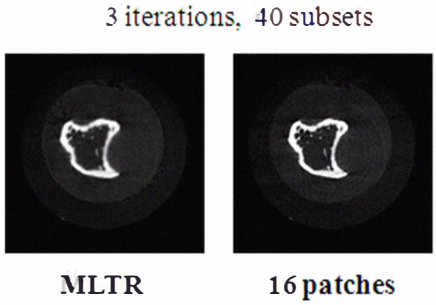
Sequential CT
PMMA phantom with aluminum inserts

Subsets / *Patches*	1	20	40
1	200it	9.0it	4.7it
4	100.4it	4.8it	2.5it
16	50.7it	2.6it	1.4it
64	26.1it	1.6it	<1it



Micro CT (cone beam)
Mouse bone in water

Subsets / *Patches*	1	20	40
1	200it	10.1it	4.9it
4	100.3it	5.2it	2.7it
16	50.6it	2.8it	1.7it
64	25.9it	1.8it	*1.7it*



Helical CT
Patient data, thorax

Subsets / *Patches*	1	20	40
1	100it	4.4it	2.4it
4	52.0it	2.3it	1.3it
16	27.1it	1.4 it	<1it

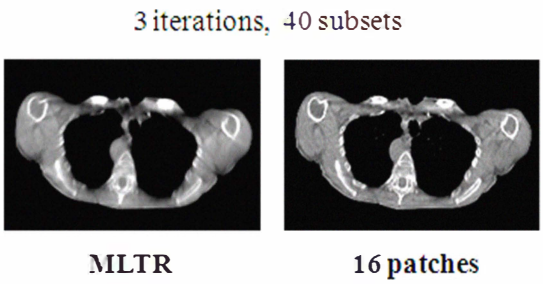


Fig. 2. Convergence analysis for MLTR with and without subsets and patches. Top: sequential CT, middle: micro CT (cone beam), bottom: helical CT. On the left the degree of convergence compared to an one-subset-one-patch iteration is shown in the tables. At the right the reconstructions after 3 iterations with 40 subsets in shown.

PET
FDG Brain scan

Subsets / *Patches*	1	10	20
1	200it	20.0it	9.9it
4	100.7it	10.3it	5.1it
16	51.2it	5.4it	2.7it
64	27.3it	3.0it	1.5it

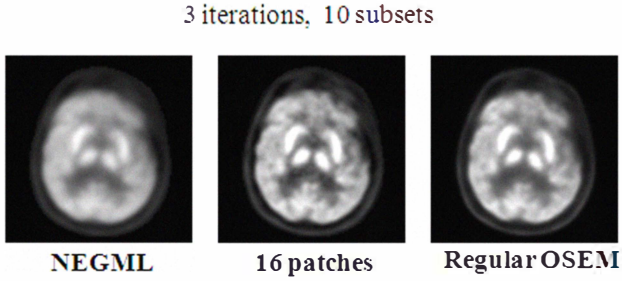


Fig. 4. Convergence analysis for NEGML with and without subsets and patches for the FDG PET scan. On the left the degree of convergence compared to an one-subset-one-patch iteration is shown in the tables. At the right the reconstructions after 3 iterations with 10 subsets in shown and compared to a regular OSEM reconstruction.

Other choices of α_j can be made to influence the degree of convergence over the image, this can also be combined with patches.

Note that the NEGML algorithm has another difference with MLEM. It has no positivity constraint which allows negative voxels values and can reduce bias [5].

VII. Conclusion

Dividing the reconstruction volume into equally sized cuboids, *patches*, and updating them sequentially leads to an improved convergence degree proportional to the square root of the number of patches. This structures still allows the use of a conventional (back)projection, which is not the case for pixel by pixel updates.

References

[1] Fessler J.A. et al, "Grouped-coordinate ascent algorithm for penalized-likelihood transmission image reconstruction." *Trans. Med. Im.*, 1997; 16 (2): 166-175.

[2] Fessler J.A. and Donghwan K., "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*, 2011, Potsdam, Germany; 262-265.

[3] Van Slambrouck K. and Nuyts J., "A patchwork (back)projector to accelerate artifact reduction in CT reconstruction". *IEEE Nucl Sci Symp Conf Record*, Knoxville, TN, USA, oct-nov 2010.

[4] Nuyts J. et al, "Iterative reconstruction for helical CT: a simulation study." *Phys. Med. Biol.*, 1998; 43: 729-737.

[5] Nuyts J. et al, "Reducing loss of image quality due to the attenuation artifact in uncorrected PET whole body images." *J. Nucl. Med.*, 2002; 43: 1054-1062.

[6] Lange K. and Fessler F.A., "Globally convergent algorithms for maximum a posteriori transmission tomography." *Tran. Im. Proc.*, 1995; 4 (10): 1430-1438.

[7] Shepp L.A. and Vardi Y., "Maximum likelihood reconstruction for emission tomography. " *Trans. Med. Im.*, 1982; 1 (2): 113-122.

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

[9] Hudson H.M. and Larkin R.S. , "Accelerated image reconstruction using ordered subsets of projection data" *Trans. Med. Im.*, 1994; 13 (4):601 - 609.