Iterative reconstruction for helical CT: a simulation study

John Nuyts†, Bruno De Man‡, Patrick Dupont†, Michel Defrise§,
Paul Suetens‡ and Luc Mortelmans†

† Department of Nuclear Medicine, Katholieke Universiteit, Leuven, Belgium
‡ ESAT-MI2, Katholieke Universiteit, Leuven, Belgium
§ Vrije Universiteit, Brussels, Belgium

Abstract. Iterative reconstruction algorithms for helical CT are presented. The algorithms are derived from two-dimensional reconstruction algorithms, by adapting the projector/backprojector to the helical orbit of the source, and by constraining the axial frequencies with a Gaussian sieve. Simulations have been carried out and the performance of the iterative algorithms is compared to that of filtered backprojection of synthetic (interpolated) two-dimensional sinograms. The iterative algorithms produce superior bias–noise curves. Axial resolution is superior, but disturbing edge-artefacts are introduced.

1. Introduction

Helical computed tomography (spiral CT) combines x-ray source rotation with continuous patient translation (along the rotation axis), which means that consecutive projections (at consecutive angles) are measured at increasing axial positions. The acquired data can be represented in a sinogram, where the column indicates the position on the detector, and the row is proportional to both the projection angle and the table position. In current clinical systems, the three-dimensional (3D) sinogram is converted using linear interpolation to a set of approximate 2D sinograms, which are reconstructed with filtered backprojection (FBP). However, the interpolation reduces the axial resolution. Moreover, this method is only approximate, and artefacts may be produced in regions with high axial gradients. Because of their higher computation times, iterative algorithms have not yet received much attention for this application. However, in contrast with the classical approach, they can handle a very accurate model of the acquisition process. In addition, iterative algorithms have been shown to outperform filtered backprojection in some PET and CT applications (Mumcuoğlu *et al* 1996, Wang *et al* 1996). Ever-increasing computer power and effective acceleration techniques (Hudson and Larkin 1994, Fessler *et al* 1997) are bringing iterative reconstruction for helical CT within reach. The aim of this simulation study is to investigate the possible benefit of iterative methods with respect to non-iterative Fourier methods, regardless of any speed considerations. All simulations and reconstructions have been carried out assuming a parallel-beam geometry because of the availability of parallel-beam projection/backprojection software. Because the difference between a fan-beam and parallel-beam geometry vanishes with decreasing object size, the findings from a parallel-beam simulation are relevant for fan-beam helical CT. Adaptation of the algorithms to a fan-beam geometry is straightforward.

2. Methods

2.1. Reconstruction algorithms

The following algorithms have been included in this study:

- NN180: 180° nearest-neighbour interpolation FBP
- LIN180: 180° linear interpolation FBP
- ILIN180: iterated 180° linear interpolation FBP
- ML–EM: maximum likelihood–expectation maximization
- ML–TRANS: maximum likelihood–gradient ascent for transmission tomography.

The first two methods are included for comparison. LIN180 is the current standard in clinical practice. NN180 produces better axial resolution, but at the cost of more severe reconstruction artefacts near axial gradients. ILIN180 is a heuristic algorithm based on iterative filtered backprojection (Pan *et al* 1993, Riddell *et al* 1995), which is expected to converge faster than the maximum-likelihood methods. ML–TRANS is a maximum-likelihood algorithm, which treats the transmission data as realizations from a Poisson distribution. Poisson noise is indeed the dominating noise contributor (Guan and Gordon 1996). ML–EM on the other hand, treats the log-converted sinograms as Poisson data. This assumption is wrong. However, ML–EM is known to be a robust algorithm (at least at low iteration numbers): it is less sensitive to inconsistencies in the sinograms than e.g. FBP. Comparison between ML–TRANS and ML–EM is of interest, since it helps to find out whether the power of the ML method is in its robustness or in its accurate noise model. Below, a more exact description of the algorithms is given, using the following definitions:

μ_j is the linear attenuation coefficient at position j, represented by voxel j

y_i is the measured value at i, where i indicates both the angle θ of the projection line and its position d within the detector array ($d = 0$ is the center of the array).

b_i is the blank scan value at i

$m_i = \ln(b_i/y_i)$, if $y_i > 1$ and $m_i = \ln b_i$ otherwise. This is the log-converted sinogram, protected against zero division.

$i(d, \theta)$ is the index i as a function of d and θ.

$\theta(z)$ is the projection angle at axial position z. Similarly, $z(\theta)$ is the axial position corresponding to projection angle θ (assuming θ increases from 0 to ∞).

c_{ij} is the contribution of voxel j to the projection line i, normalized to the case in which the projection line intersects the centre of the voxel, so $0 \leqslant c_{ij} \leqslant 1$, $\forall i, j$.

N is the number of samples (voxels) along the x- or the y-axis.

$L_y(\mu)$ is the likelihood of the reconstruction μ, given the measured sinogram y.

In all cases, FBP is implemented using convolution in the spatial domain, to avoid artefacts due to FFT.

(i) *NN180.* A 2D 180° sinogram for position z_0 is constructed as

$$\{m_{i(d,\phi)} \mid \theta(z_0) - \pi/2 < \phi < \theta(z_0) + \pi/2\}. \tag{1}$$

This sinogram is reconstructed with FBP.

(ii) *LIN180.* A 2D 180° sinogram for position z_0 is constructed as

$$\{q_{i(d,\phi)} \mid \theta(z_0) - \pi/2 < \phi < \theta(z_0) + \pi/2\} \tag{2}$$

$$q_{i(d,\phi)} = (1 - w)m_{i(d,\phi)} + wm_{i(-d,\phi+\pi)} \qquad \text{if } z(\phi) < z_0$$

$$= (1 - w)m_{i(d,\phi)} + wm_{i(-d,\phi-\pi)} \qquad \text{if } z(\phi) > z_0$$

$$w = \frac{|\phi - \theta(z_0) \bmod 2\pi|}{\pi}.$$

This sinogram is reconstructed with FBP.

(iii) *ILIN180*. In each iteration new values for μ_j are computed as

$$\mu_j^{\text{new}} = \mu_j + \text{LIN180}(\Delta m) \qquad \text{with } \Delta m_i = m_i - \sum_j c_{ij}\mu_j. \qquad (3)$$

(iv) *ML–EM*. This is the algorithm proposed by Shepp and Vardi (1982) and Lange and Carson (1984) for emission tomography, but applied to the log-converted sinogram.

$$\mu_j^{\text{new}} = \frac{\mu_j}{\sum_i c_{ij}} \sum_i c_{ij} \frac{m_i}{\sum_\xi c_{i\xi}\mu_\xi}. \qquad (4)$$

(v) *ML–TRANS*. Lange and Carson (1984) have derived an ML–EM algorithm for transmission tomography. However, the maximization step leads to a transcendental equation, which can be solved by approximation as a truncated series expansion (Lange and Carson 1984) or by numerical optimization (Ollinger 1994) at the cost of increased complexity per iteration. In addition, the complete variables (the number of photons entering each pixel) do not disappear from the equations as in emission tomography. As a result, modified projector/backprojectors have to be used. In the appendix, we derive a simple gradient ascent algorithm, which directly maximizes the exact likelihood function and uses only regular projection/backprojection, making implementation relatively easy. A potential weakness is that it accepts negative values in the solution. We set negative values to zero in every iteration. The algorithm can be written as:

$$\mu_j^{\text{new}} = \mu_j + \frac{\alpha}{N}\left(1 - \frac{\sum_i c_{ij} y_i}{\sum_i c_{ij} b_i \exp(-\sum_\xi c_{i\xi}\mu_\xi)}\right) \qquad (5)$$

where α is a relaxation parameter. Because of its similarity to the ML–EM expression, the ordered subsets acceleration approach (Hudson and Larkin 1994) can be applied directly to this algorithm. The resulting behaviour is similar to that of accelerated ML–EM: large acceleration factors are achieved, but when few projections per subset are used in noisy images, convergence is affected. For this work, both the ML–EM and the ML–TRANS algorithm were applied without acceleration.

In the FBP-based methods (NN180, LIN180, ILIN180), the high-frequency noise is suppressed by convolving the reconstructed slices with a 2D Gaussian convolution mask. In the iterative methods (ILIN180, ML–EM, ML–TRANS), noise is suppressed by stopping the iterations before convergence is obtained.

Because of the helical orbit, data are not sufficient to obtain high axial resolution. Consequently, axial constraining is required. This is accomplished by including an axial Gaussian sieve in the iterative algorithms (Snyder and Miller 1985): prior to each projection $(\sum_j c_{ij}\mu_j)$ and after each backprojection $(\sum_i c_{ij} y_i)$, a convolution with the sieve is applied. The resulting reconstruction must be convolved with the sieve to produce the final image. In this study, the full width at half maximum of the sieve was set equal to the width of the rectangular axial point spread function of the simulated scanner.

In the iterative reconstruction algorithms, the axial point spread function (PSF) is taken into account in an approximate way. In each iteration, the reconstructed images are convolved with the axial PSF (before projection and after backprojection). This is an approximation: because in reality, axial blurring occurs during detection of the photons, the convolution should actually be carried out on the sinograms, prior to log-conversion. The latter approach, however, requires computation of multiple parallel projections, resulting in an unacceptable increase of computation time.

2.2. Simulated objects

Three objects have been simulated. The first is a single clinical CT image of the thorax, in which the grey values were replaced by typical attenuation coefficients. Using this 2D simulation is equivalent to testing the algorithms in the case of an object which is constant in the z-direction. The image contained 300×300 pixels, there were 300 detectors per angle and 400 projection angles were simulated. For this object, NN180 and LIN180 reduce to FBP and ILIN180 to iterative FBP. The second object consists of an elliptical attenuating cylinder. This cylinder contains a smaller circular cylinder with 10% higher attenuation coefficient. The smaller cylinder occupies only half of the axial field of view, thus producing a sharp axial gradient. This object is used to study performance near axial gradients, since axial gradients are the dominant source of artefacts in helical CT. The third object consists of an elliptical attenuating cylinder, containing multiple ellipsoids with different attenuation coefficient. The purpose is to verify the findings obtained on the previous two phantoms. Six slices were reconstructed for the second object and 15 for the third object, using a slice separation of 0.5 times the axial PSF width, 150 angles per $180°$ and 120 detectors. For all objects, Poisson noise was simulated assuming 10^6 photons per detector prior to attenuation.

2.3. Simulation algorithm

For the first object, projections were computed at twice the resolution as compared to the projection/backprojection in the reconstruction programs (i.e. two projection lines were computed per detector pixel, using twice the number of points per projection line). Finite spatial resolution was simulated by convolving the computed counts with the mask [0.25, 0.5, 0.25]. For the two other objects, projections were computed analytically as line integrals and finite resolution was simulated with the same convolution mask. Here, the axial PSF was simulated by computing for every angle and every detector ten projections at different axial position (uniformly sampling the rectangular PSF) and adding the 'detected' counts.

Poisson noise was approximated as a Gaussian pseudo-random realization, with variance equal to the mean. In the reported results, the axial increment per $180°$ was 0.9 times the axial (rectangular) PSF width.

2.4. Evaluation

Evaluation is done qualitatively by visual inspection, and quantitatively by computing bias–noise curves for each algorithm. Bias and noise in a region are computed as:

$$\text{bias} : \sqrt{\frac{\sum_j (q_j - r_j)^2}{M}} \qquad \text{noise} : \sqrt{\frac{\sum_j (x_j - q_j)^2}{M}} \tag{6}$$

where M is the number of pixels in the region, q is the reconstruction of a noise-free simulation, r is the reference image and x is a reconstruction of noisy data.

For NN180 and LIN180, these curves are obtained by varying the standard deviation of the Gaussian smoothing from 0 to 3 pixels. For ML–EM and ML–TRANS, the number of iterations is varied. For ILIN180, both the number of iterations and the width of the Gaussian mask are varied. Lower bias for the same noise (and lower noise for the same bias) is suggestive for more accurate reconstruction. As pointed out in the discussion, mean squared error is a simple but poor figure of merit, and for a final analysis, task-dependent assessment of image quality will have to be carried out.
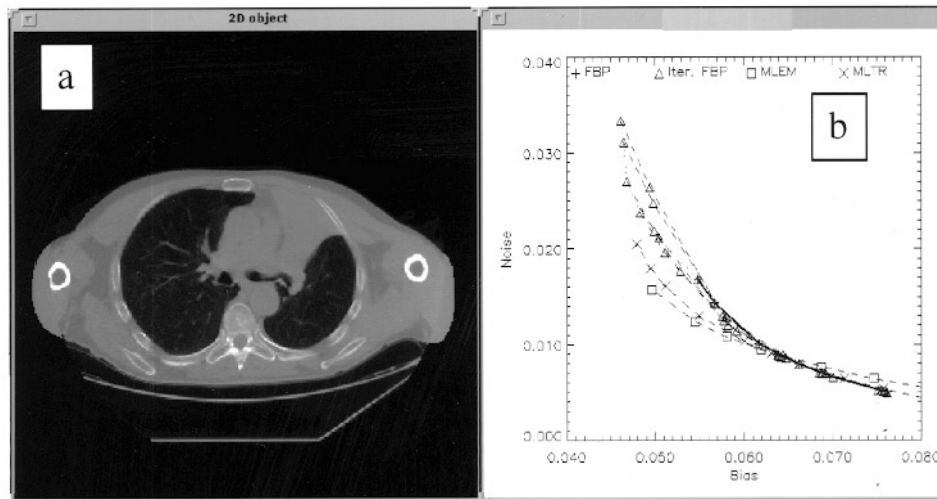
Figure 1. (a) The CT phantom. (b) Bias–noise curves using Gaussian smooths with $\sigma = 0$, 0.5, 1, 1.5, 2 and 3 pixels for FBP (solid line) and iterative FBP (triangles, dotted line). For iterative FBP, iterations 1, 2, 3, 5, 10 and 15 are shown as independent dashed lines. The points for 15 iterations are connected with a dotted line. Iteration 1 is identical to FBP, shown as a solid line. For ML–EM, iterations 1, 3, 6, 10, 20, 30, 50, 70, 100, 200 are shown, and for ML–TRANS also iterations 400, 600, 1000.

3. Results

Figure 1(a) shows the 2D object, and figure 1(b) shows the corresponding bias–noise curves. Note that in the 2D case NN180 and LIN180 both turn into FBP, and ILIN180 turns into iterative FBP. For iterative FBP, one curve per iteration is shown, where the curve is produced using different smoothing masks. Initially, iteration increases noise and decreases bias. At ten iterations, however, increased bias is observed. For the other algorithms, bias decreases monotonically with iteration number. Figure 2(a) shows five central reconstructed slices of the axial gradient object at similar noise level for each of the algorithms. Figure 2(b) presents the bias–noise curves for a central region enclosing the central cylinder. When the region encloses the entire elliptical cylinder, the curves of the ML algorithms shift towards lower bias and lower noise, relative to the other curves (data not shown). Figure 2(c) shows axial profiles, computed as the mean of a region in the centre of the small cylinder as a function of plane position. Figure 3 shows five transaxial slices for the object with multiple ellipsoids (a), and a profile intersecting two ellipsoids (b). For this simulation, the bias–noise curves of ILIN180 overlapped those of NN180 and LIN180, while those of the two ML methods had about 20% lower bias for the same noise level.

4. Discussion

For the 2D object, the bias–noise curves of iterative FBP and FBP overlap, but iterative FBP allows us to reach lower bias values at the cost of increased noise. The maximum-likelihood algorithms both produce very similar bias–noise curves, which are slightly better than those of (iterative) FBP. At higher iteration numbers, all images are of high quality and approximately equivalent. These findings suggest that the ML methods perform at least as well as the FBP methods for nearly ideal data (in the absence of axial gradients).
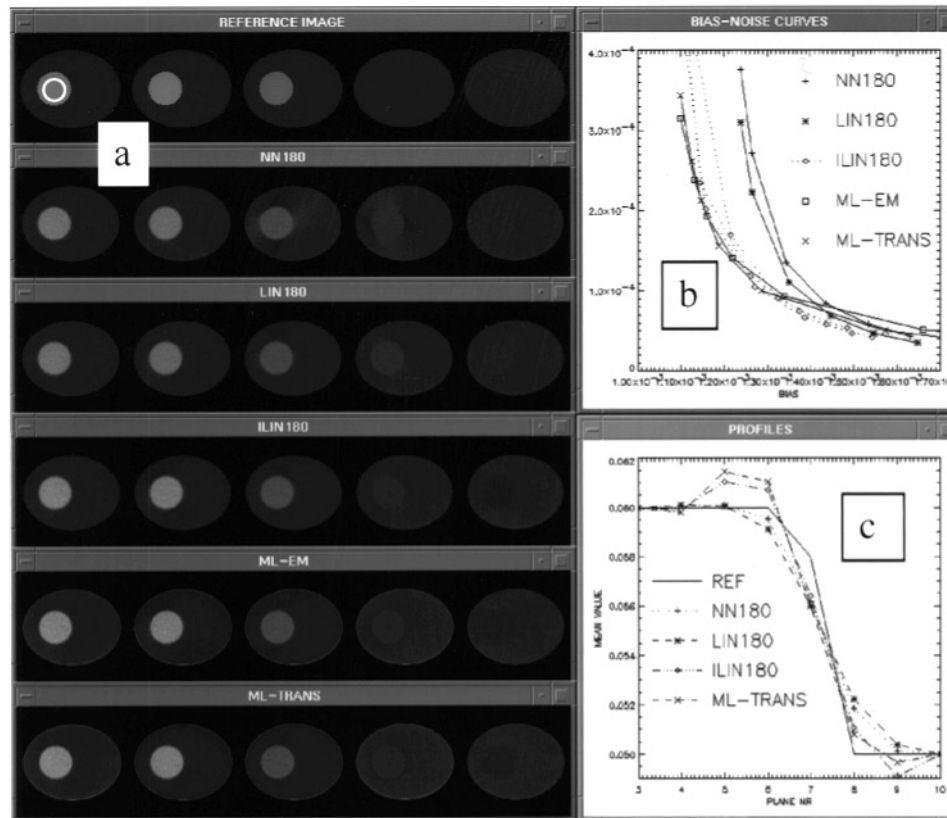
Figure 2. (a) Reconstructions for the axial gradient object, from top to bottom: reference image, NN180, LIN180, ILIN180, ML–EM, ML–TRANS. For ML–EM and ML–TRANS; images at highest iteration number are shown; for ILIN180, it is the image at 20 iterations, smoothed with $\sigma = 1$. (b) Bias–noise curves for a region around the central cylinder. Gaussian smooths with $\sigma = 0$, 0.5, 1, 1.5, 2 and 2.5 pixels were applied. For iterative FBP, iterations 5, 10 and 20 are shown, as independent dotted lines. For ML, iterations 5, 10, 15, 30, 50, 75, 100 and 150 are shown. (c) Mean value of the circular region (shown on top left slice) as a function of axial position for the reference image (solid line) and for NN180, LIN180, ILIN180 and ML–TRANS.

As shown in figures 2(a) and (c), NN180 has somewhat better axial resolution, but shows severe artefacts near the gradient. LIN180 reduces these artefacts at the cost of poorer axial resolution. The iterative methods produce an axial resolution superior to that of NN180 combined with strong artefact reduction. However, a new artefact is introduced: noticeable axial Gibbs undershoots and overshoots show up in slices adjacent to the gradient. The ML methods also produce in-plane Gibbs overshoots (ringing) at the edges of the outer cylinder. The Gibbs over- and undershoots are due to reconstruction of strong gradients without incorporating the highest frequencies. In the transaxial direction, the highest frequencies are not yet fully converged, even at high iteration numbers. Snyder *et al* (1987) argue that the Gibbs artefact is due to the ill-posedness of the reconstruction problem (high spatial frequencies are lost during detection) and they show that post-smoothing reduces the artefact. Without smoothing, the Gibbs artefact also appears in ILIN180 reconstructions, but it is suppressed by the 2D Gaussian post-filtering. In the axial direction, the high frequencies are suppressed by the sieve. Omitting the sieve, however, results in still stronger artefacts (data
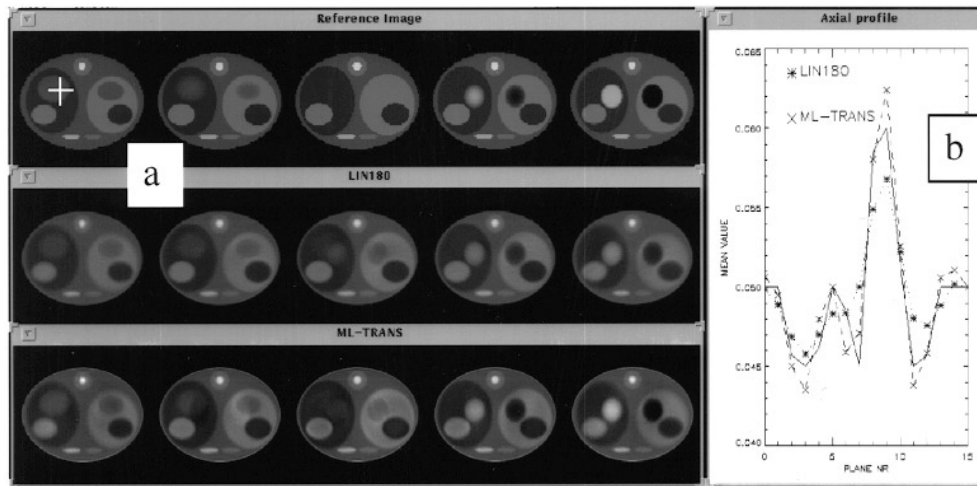
Figure 3. (a) Five transaxial slices of reconstructions for the third simulation. From top to bottom: reference image, LIN180 and ML–TRANS. (b) An axial profile (position shown on top left slice), for the reference image (solid line), LIN180 (*, dotted) and ML–TRANS (×, dashed).

not shown), because only few projection angles are actually acquired for every reconstructed slice.

Figure 2(b) shows that in the gradient region, the bias–noise curves of the iterative methods are superior to those of the non-iterative methods, which is attributed to the artefact reduction and recovery of axial resolution. All three iterative methods perform equivalently near the gradient. When the entire region is considered, we found that the ML methods produce better bias noise curves. Visual inspection reveals that in homogeneous regions, ILIN180 tends to produce higher noise than the ML methods.

Figure 3 confirms that the iterative algorithm yields better axial resolution than LIN180, resulting in a better contrast. The axial Gibbs effect produces shadows of opposite contrast in slices adjacent to objects with strong axial gradients.

In this exploratory study, we hoped to find evidence that iterative reconstruction based on a superior mathematical model of the acquisition, would produce a clear improvement in image quality, as compared to the classical approach. If sufficient evidence were found, two (overlapping) tasks would have to be carried out: optimization of the acquisition and reconstruction parameters (axial PSF width, sieve width, acceleration algorithms etc), and more elaborate quantitative evaluation. Quantitative assessment of the performance of reconstruction algorithms is a complicated matter, because some of the algorithms are non-linear, and because image quality is task dependent (Barrett *et al* 1993, Furuie *et al* 1994). However, we feel that the current results indicate that axial constraining should be improved before the iterative methods can be considered for clinical applications.

5. Conclusion

We conclude that iterative ML reconstruction allows improvement of the axial resolution beyond that of nearest-neighbour interpolation (NN180) without introducing the in-plane distortions seen in NN180 and without excessive noise amplification. Iterative FBP

(ILIN180) can produce lower bias than LIN180 or NN180, but for the same bias, noise was equivalent or higher than for the ML methods, depending on the region studied. Performance of ML–EM and ML–TRANS was rather similar in this study (although we have some preliminary evidence indicating that when small objects with very high attenuation are present, bias–noise curves by ML–TRANS are significantly better). This suggests that, at least in these simulations, the exactness of the noise model was less important than the robustness of the numerical procedures.

However, constraining axial resolution is necessary, and the current sieve-based constraint introduces unacceptable axial Gibbs artefacts, which completely offset the advantage of improved axial resolution, in-plane artefact reduction and noise suppression. A potential solution is the use of a nonlinear axial constraint, such as a Gibbs prior (Geman and McClure 1987).

Acknowledgments

Appendix A

The log-likelihood for transmission tomography is (see e.g. Lange and Carson 1984):

$$L = \sum_i (y_i \ln t_i - t_i - \ln(y_i!)) \qquad \text{with } t_i = b_i \, e^{-\sum_j c_{ij}\mu_j}. \tag{A1}$$

It can be shown that this function has a unique maximum (Lange and Carson 1984). Therefore, an iterative algorithm of the form $\mu_j^{\text{new}} = \mu_j + \Delta\mu_j$, with $\Delta\mu_j$ the same sign as $\partial L/\partial\mu_j$, $|\Delta\mu_j|$ sufficiently small and $|\Delta\mu_j| > 0$ if $|\partial L/\partial\mu_j| > 0$, increases the likelihood at every iteration.

We will estimate now the step $\Delta\mu_j$ in the neighbourhood of the maximum $\boldsymbol{\mu}_0$. Expanding the first derivative up to first order, and assuming that $\Delta\boldsymbol{\mu}$ is such that $\boldsymbol{\mu} + \Delta\boldsymbol{\mu} = \boldsymbol{\mu}_0$, we obtain:

$$0 = \frac{\partial L}{\partial\mu_k}(\boldsymbol{\mu} + \Delta\boldsymbol{\mu}) = \frac{\partial L}{\partial\mu_k}(\boldsymbol{\mu}) + \sum_l \frac{\partial^2 L}{\partial\mu_k\partial\mu_l}\Delta\mu_l \qquad \forall k. \tag{A2}$$

This implies that $\Delta\boldsymbol{\mu}$ should satisfy:

$$\sum_k \left(\frac{\partial L}{\partial\mu_k}(\boldsymbol{\mu}) + \sum_l \frac{\partial^2 L}{\partial\mu_k\partial\mu_l}\Delta\mu_k \right) = 0. \tag{A3}$$

Instead of solving (A2), which is much more complicated, we estimate $\Delta\boldsymbol{\mu}$ as:

$$\Delta\mu_k = -\frac{(\partial L/\partial\mu_k)(\boldsymbol{\mu})}{\sum_l (\partial^2 L/\partial\mu_k\partial\mu_l)(\boldsymbol{\mu})}. \tag{A4}$$

This expression satisfies (A3) and has the same sign as $\partial L/\partial\mu_k$ since $(\partial^2 L/\partial\mu_k\partial\mu_l)(\boldsymbol{\mu}) < 0, \forall k, l$. Introducing a relaxation factor $\alpha > 0$ we obtain the following algorithm:

$$\mu_k^{\text{new}} = \mu_k + \frac{\alpha}{N}\left(1 - \frac{\sum_i c_{ik}y_i}{\sum_i c_{ik}b_i \exp[-\sum_j c_{ij}\mu_j]} \right) \tag{A5}$$

where we used the fact that $\sum_j c_{ij} = N$ for an image of $N \times N$ pixels. In principle, the relaxation factor could be chosen to maximize the likelihood increase in every iteration (Kaufman 1987). We have set $\alpha = 2$ for all iterations, which we found to produce a monotonic increase in likelihood and fair convergence on a large set of images. The value of L is explicitly calculated, and, if required, α is decreased to ensure that each iteration step actually increases L.

References

Barrett H H, Yao J, Rolland J P and Kyle J M 1993 Model observers for assessment of image quality *Proc. Natl Acad. Sci. USA* **90** 9758–65

Fessler J A, Ficaro E P, Clinthorne N H and Lange K 1997 Grouped-coordinate ascent algorithms for penalized-likelihood transmission image reconstruction *IEEE Trans. Med. Imaging* **16** 166–75

Furuie S S, Herman G T, Narayan T K, Kinahan P E, Karp J S, Lewitt R M and Matej S 1994 A methodology for testing for statistically significant differences between fully 3D PET reconstruction algorithms *Phys. Med. Biol.* **39** 341–54

Geman S and McClure D E 1987 Statistical methods for tomographic image reconstruction *Bull. Int. Stat. Inst.* **52** 5–21

Guan H and Gordon R 1996 Computed tomography using algebraic reconstruction techniques with different projection access schemes: a comparison study under practical situations *Phys. Med. Biol.* **41** 1727–43

Hudson H M and Larkin R S 1994 Accelerated image reconstruction using ordered subsets of projection data *IEEE Trans. Med. Imaging* **13** 601–9

Kaufman L 1987 Implementing and accelerating the EM algorithm for positron emission tomography *IEEE Trans. Med. Imaging* **6** 37–51

Lange K and Carson R 1984 EM reconstruction algorithms for emission and transmission tomography *J. Comput. Assist. Tomogr.* **8** 306–16

Mumcuoğlu E, Leahy R and Cherry S R 1996 Bayesian reconstruction of PET images: methodology and performance analysis *Phys. Med. Biol.* **41** 1777–807

Ollinger 1994 Maximum-likelihood reconstruction of transmission images in emission computed tomography via the EM algorithm *IEEE Trans. Med. Imaging* **13** 89-101.

Pan T-S, Yagle A E, Clinthorne N H and Rogers W L 1993 Acceleration and filtering in the generalized Landweber iteration using a variable shaping matrix *IEEE Trans. Med. Imaging* **12** 278–86

Riddell C, Bendriem B, Bourguignon M and Kernevez J-P 1995 The approximate inverse and conjugate gradient: non-symmetrical algorithms for fast attenuation correction in SPECT *Phys. Med. Biol.* **40** 269–81

Shepp L A and Vardi Y 1982 Maximum likelihood reconstruction for emission tomography *IEEE Trans. Med. Imaging* **1** 113–22

Snyder D L and Miller M I 1985 The use of sieves to stabilize images produced with the EM algorithm for emission tomography *IEEE Trans. Nucl. Sci.* **32** 3864–72

Snyder D L, Miller M I, Lewis J T and Politte D G 1987 Noise and edge artifacts in maximum-likelihood reconstructions for emission tomography *IEEE Trans. Med. Imaging* **6** 228–38

Wang G, Snyder D L, O'Sullivan J A and Vannier M W 1996 Iterative deblurring for CT metal artifact reduction *IEEE Trans. Med. Imaging* **15** 657–64