

Efficient Bregman Iteration in fully 3D PET

László Szirmay-Kalos¹, Balázs Tóth¹, and Gábor Jakab²

Abstract—Positron Emission Tomography reconstruction is ill posed. The result obtained with the iterative ML-EM algorithm is often noisy, which can be controlled by *regularization*. Common regularization methods penalize high frequency features or the total variation, thus they compromise even valid solutions that have such properties. Bregman iteration offers a better choice enforcing regularization only where needed by the noisy data. Bregman iteration requires a nested optimization, which poses problems when the algorithm is implemented on the GPU where storage space is limited and data transfer is slow. Another problem is that the strength of the regularization is set by a single global parameter, which results in overregularization for voxels measured by fewer LORs. To handle these problems, we propose a modified scheme that merges the two optimization steps into one, eliminating the overhead of Bregman iteration. The algorithm is demonstrated for a 2D test scenario and also in fully 3D reconstruction. The benefits over TV regularization are particularly high if the data has higher variation and point like features. The proposed algorithm is built into the TeraTomoTM system.

I. INTRODUCTION

Tomography reconstruction is the *inverse problem* of particle transport, which requires the iteration of particle transport simulations and corrective back projections [10]. The inputs of the reconstruction are the measured values in *Lines of Responses* or *LORs*: $\mathbf{y} = (y_1, y_2, \dots, y_{N_{\text{LOR}}})$. The output of the reconstruction method is the *tracer density* function $x(\vec{v})$, which is approximated in a *finite function series* form:

$$x(\vec{v}) = \sum_{V=1}^{N_{\text{voxel}}} x_V b_V(\vec{v}), \quad (1)$$

where $\mathbf{x} = (x_1, x_2, \dots, x_{N_{\text{voxel}}})$ are the coefficients to be computed, and $b_V(\vec{v})$ ($V = 1, \dots, N_{\text{voxel}}$) are *basis functions*, which are typically defined on a *voxel grid*. As only non-negative tracer density makes sense, we impose non-negativity requirement $x(\vec{v}) \geq 0$ on the solution. If basis functions $b_V(\vec{v})$ are non-negative, this requirement can be formulated for the coefficients as well: $x_V \geq 0$.

The correspondence between tracer density coefficients x_V and the expected number of hits \tilde{y}_L in LOR L is described by *system matrix* A_{LV} , where element LV is the probability that an hit event occurs in LOR L given that a radioactive decay happened in voxel V :

$$\tilde{y}_L = \sum_{V=1}^{N_{\text{voxel}}} A_{LV} x_V. \quad (2)$$

Assuming that photon incidents in different LORs are independent random variables with Poisson distribution, the

Expectation Maximization (ML-EM) algorithm [11] should maximize the following likelihood function:

$$\log \mathcal{L}(x) = \log \left(\prod_{L=1}^{N_{\text{LOR}}} \frac{\tilde{y}_L^{y_L}}{y_L!} e^{-\tilde{y}_L} \right)$$

subject to $x_V \geq 0$.

The reconstruction process is ill-conditioned, which means that enforcing the maximization of the likelihood function may result in a noisy reconstruction. To recognize when the data is being fitted to noise, we can measure the *quality* of the approximation, i.e. how free the data is from unwanted high frequency characteristics. This measure is called the *regularization term* and is denoted by $R(x)$.

There are different possibilities to include regularization information in the reconstruction:

- 1) *Early termination* stops the iteration when the quality becomes degrading during ML-EM.
- 2) Enforcing the quality by voxel space or LOR space filtering during the reconstruction process. It means that in each iteration, we include a filtering step that improves the quality of the current approximation.
- 3) *Constrained optimization* is based on the recognition that we have two optimization criteria, the likelihood and the quality of the data term, so we can take one of them as an optimization objective while the other as a constraint. However, there are two problems. Firstly, the constraint cannot be well defined since it would require either the likelihood or the quality of the true solution, which is not available. Secondly, constrained optimization is more difficult computationally than unconstrained optimization.
- 4) Merging the data term and the regularization term into a single objective function where poor quality solutions are penalized by the regularization term.

In this paper, we investigate the last option, and add penalty or regularization term $R(x)$ to the negative likelihood. The penalty term should be high for unacceptable solutions and small for acceptable ones. In the objective function penalty term $R(x)$ is scaled by regularization parameter λ . We minimize

$$E(x) = -\log P(\mathbf{y}|\mathbf{x}) + \lambda R(\mathbf{x}). \quad (3)$$

To minimize the multi-variate objective function with inequality constraints of non-negativity, we can use the Kuhn-Tucker conditions, which lead to:

$$x_V \frac{\partial E}{\partial x_V} = x_V \left(\lambda \frac{\partial R}{\partial x_V} - \sum_{L=1}^{N_{\text{LOR}}} \mathbf{A}_{LV} \frac{y_L}{\tilde{y}_L} + \sum_{L=1}^{N_{\text{LOR}}} \mathbf{A}_{LV} \right) = 0$$

¹: Budapest University of Technology and Economics (e-mail: szirmay@it.bme.hu).

²: Mediso Ltd. (e-mail: gabor.jakab@mediso.hu).

for $V = 1, 2, \dots, N_{\text{voxel}}$. Rearranging the terms, we obtain the following equation for the optimum:

$$x_V \left(\lambda \frac{\partial R}{\partial x_V} + \sum_{L=1}^{N_{\text{LOR}}} \mathbf{A}_{LV} \right) = x_V \sum_{L=1}^{N_{\text{LOR}}} \mathbf{A}_{LV} \frac{y_L}{\tilde{y}_L}.$$

There are many possibilities to establish an iteration scheme that has a fix point satisfying this equation. A provably converging backward scheme would solve the following equation in every step:

$$x_V^{(n+1)} \left(\lambda \frac{\partial R(x^{(n+1)})}{\partial x_V} + \sum_{L=1}^{N_{\text{LOR}}} \mathbf{A}_{LV} \right) = x_V^{(n)} \sum_{L=1}^{N_{\text{LOR}}} \mathbf{A}_{LV} \frac{y_L}{\tilde{y}_L^{(n)}}$$

where

$$\tilde{y}_L^{(n)} = \sum_{V=1}^{N_{\text{voxel}}} A_{LV} x_V^{(n)}$$

is the expected number of hits in LOR L computed from the activity distribution available in iteration step n .

A. Analysis of TV regularization

Classical regularization terms measure the distance between the constant function and the actual estimate. Tychonoff regularization is based on the L_2 distance. As a result, Tychonoff regularization produces blurred, oversmoothed edges. In PET data there are sharp features that should not be smoothed with the regularization method. We need a penalty term that minimizes the unjustified oscillation without blurring sharp features. An appropriate functional is the *Total Variation* (TV) of the solution [9],[6]. In one dimension the total variation measures the length of the path traveled by the function value while its parameter runs over the domain. For differentiable functions, the total variation is the integral of the absolute value of the function's derivative. If the basis functions are piece-wise linear tent-like functions, then the TV in one dimension is

$$TV(x) = \sum_{V=1}^{N_{\text{voxel}}} |x_V - x_{V-1}|. \quad (4)$$

In higher dimensions, the total variation can be defined as the integral of the absolute value of the gradient over the total volume \mathcal{V} :

$$TV(x) = \int_{\mathcal{V}} |\nabla x(\vec{v})| dv. \quad (5)$$

In regularization, the variation of the true solution is also penalized, so the optimum is be modified. Total Variation (TV) regularization gives reduced contrast solutions having stair-case artifacts, which can be controlled by regularization parameter λ . The optimal regularization parameter can be obtained with *Hansen's L-curves* [4], which states that the optimal λ is where the $(-\log \mathcal{L}(x_\lambda), TV(x_\lambda))$ parametric curve has maximum curvature (note that in the final solution x depends on λ , so both the likelihood and the total variation will be functions of the regularization parameter). However, the algorithm developed to locate the maximum curvature points assumes Tychonoff regularization, and there is no practically feasible generalization to large scale problems based on TV.

To examine these artifacts formally, let us consider the one dimensional case when the TV is defined by Equ. 4. The partial derivatives of the TV functional is

$$\frac{\partial TV(x(\vec{v}))}{\partial x_V} = \frac{\partial \sum_{V=1}^{N_{\text{voxel}}} |x_V - x_{V-1}|}{\partial x_V} =$$

$$\begin{cases} 2 & \text{if } x_V > x_{V-1} \text{ and } x_V > x_{V+1}, \\ -2 & \text{if } x_V < x_{V-1} \text{ and } x_V < x_{V+1}, \\ 0 & \text{otherwise.} \end{cases}$$

Thus the iteration formula becomes independent of the regularization when $x_V^{(n)}$ is not a local extremum:

$$x_V^{(n+1)} = \frac{x_V^{(n)} \cdot \sum_{L=1}^{N_{\text{LOR}}} \mathbf{A}_{LV} \frac{y_L}{\tilde{y}_L^{(n)}}}{\sum_{L=1}^{N_{\text{LOR}}} \mathbf{A}_{LV}}.$$

When $x_V^{(n)}$ is a local maximum, regularization makes it smaller by increasing the denominator by 2λ :

$$x_V^{(n+1)} = \frac{x_V^{(n)} \cdot \sum_{L=1}^{N_{\text{LOR}}} \mathbf{A}_{LV} \frac{y_L}{\tilde{y}_L^{(n)}}}{\sum_{L=1}^{N_{\text{LOR}}} \mathbf{A}_{LV} + 2\lambda}.$$

Similarly, when $x_V^{(n)}$ is a local minimum, regularization makes it greater by decreasing the denominator by 2λ .

This behavior is advantageous when local maxima and minima are due to the noise and overfitting. Note that the regularization has a discontinuity when x_V is similar in one of the neighbors of voxel V , so iteration is likely to stop here, resulting in staircase-like reconstructed signals. On the other hand, when the true data really has a local extremum, regularization decreases its amplitude, resulting in contrast reduction. Assume, for example, that the true data is a point source like feature that is non-zero only in a single voxel V . For this value, the ratio of the reconstruction with and without regularization, i.e. the contrast reduction is

$$C(\lambda) = \frac{\sum_{L=1}^{N_{\text{LOR}}} \mathbf{A}_{LV}}{\sum_{L=1}^{N_{\text{LOR}}} \mathbf{A}_{LV} + 2\lambda} \approx 1 - \frac{2\lambda}{\sum_{L=1}^{N_{\text{LOR}}} \mathbf{A}_{LV}}.$$

Note that contrast reduction is not uniform, but is higher where sensitivity $\sum_{L=1}^{N_{\text{LOR}}} \mathbf{A}_{LV}$ is low. These parts of the volume are close to the entry or exit of the gantry where a voxel is seen just by a few LORs. Here the regularization can significantly modify the data.

B. Bregman iteration

An optimal regularization term would have its minimum at the ground truth solution when $x = x_{\text{true}}$, and would measure the “distance” $D(x, x_{\text{true}})$ between x and x_{true} . An appropriate distance function is the *Bregman distance* [1], [15], [14] that can be based on an arbitrary convex penalty term, for example, on the total variation of the activity distribution, denoted by $TV(x)$:

$$D(x, x_{\text{true}}) = TV(x) - TV(x_{\text{true}}) - \langle \mathbf{p}, x - x_{\text{true}} \rangle$$

where \mathbf{p} is the gradient of $TV(x)$ at x_{true} , i.e. $\mathbf{p}_V = \partial TV / \partial x_V$ if it exists.

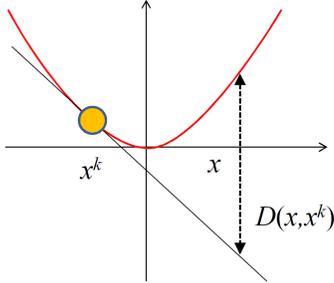


Fig. 1. Bregman distance in 1D.

In practice, we do not know the true solution, so it is replaced by an earlier estimate $x^{(k)}$. Note that if the regularization term is linear between the true solution and $x^{(k)}$, then this approximation is precise since

$$D(x, x_{\text{true}}) = D(x, x^{(k)}).$$

Total variation is based on the absolute value function, which results in piece-wise linear regularization term, making it particularly attractive for Bregman iteration.

Inserting the Bregman distance into the goal of the optimization, we get

$$E(\mathbf{x}) = -\log \mathcal{L}(\mathbf{x}) + \lambda D(\mathbf{x}, \mathbf{x}^{(k)}). \quad (6)$$

II. THE NEW METHOD: MODIFIED BREGMAN ITERATION

There are two problems with the classical Bregman iteration. The new activity $x_V^{(n+1)}$ shows up at the left side and also in the right side as an input of the Bregman distance, so this equation should also be solved by an embedded iteration. To attack this problem, we use a *one step late* option [3], which means that the Bregman distance is computed not from the new but from the actual estimate $x_V^{(n)}$, which significantly simplifies the algorithm, resulting the following scheme:

$$x_V^{(n+1)} = \frac{x_V^{(n)} \cdot \sum_{L=1}^{N_{\text{LOR}}} \mathbf{A}_{LV} \frac{y_L}{\tilde{y}_L}}{\sum_{L=1}^{N_{\text{LOR}}} \mathbf{A}_{LV} + \lambda \left(\frac{\partial TV(x^{(n+1)})}{\partial x_V} - \mathbf{p}_V^{(k)} \right)}.$$

Another problem is that regularization is not uniformly strong for different voxels but depends on *voxel sensitivity* $\sum_{L=1}^{N_{\text{LOR}}} \mathbf{A}_{LV}$, which expresses the probability that a positron

born with in voxel V is detected by the tomograph. The sensitivity is high in the center of a fully 3D PET but can be very small close to the exits of the gantry, causing over-regularization here. On the other hand, regularization strength also depends on the size of the LOR subset in OSEM iteration. To equalize the strength of regularization, we scale the global regularization parameter λ by the local voxel sensitivity. Putting these together, we can establish the following *Equalized One Step Late* scheme:

$$x_V^{(n+1)} = \frac{x_V^{(n)} \cdot \sum_{L=1}^{N_{\text{LOR}}} \mathbf{A}_{LV} \frac{y_L}{\tilde{y}_L}}{\sum_{L=1}^{N_{\text{LOR}}} \mathbf{A}_{LV} \left(1 + \lambda \left(\frac{\partial TV(x^{(n)}(\vec{v}))}{\partial x_V} - \mathbf{p}_V^{(k)} \right) \right)}.$$

When k is incremented, the gradient vector of the total variation, \mathbf{p} , should be updated. This can be done directly considering the criterion of optimality:

$$\mathbf{p}_V^{(k+1)} = \mathbf{p}_V^{(k)} + \frac{1}{\lambda} \sum_{L=1}^{N_{\text{LOR}}} \left(\frac{\mathbf{A}_{LV} \frac{y_L}{\tilde{y}_L}}{\mathbf{A}_{LV}} - 1 \right).$$

The number of k increments should be small, otherwise the noise is reintroduced into the solution.

The proposed reconstruction algorithm is:

```

for  $k = 1$  to  $K$  do
  for  $n = (k - 1)K$  to  $kK - 1$  do
     $x_V^{(n+1)} = x_V^{(n)} \cdot \frac{\sum_L \mathbf{A}_{LV} \frac{y_L}{\tilde{y}_L}}{\sum_L \mathbf{A}_{LV} (1 + \lambda (\frac{\partial TV}{\partial x_V} - \mathbf{p}_V^{(k)}))}.$ 
  endfor
   $\mathbf{p}_V^{(k+1)} = \mathbf{p}_V^{(k)} + \frac{1}{\lambda} \sum_L \left( \frac{\mathbf{A}_{LV} \frac{y_L}{\tilde{y}_L}}{\mathbf{A}_{LV}} - 1 \right).$ 
endfor

```

Both TV regularization and Bregman iteration require the computation of the derivative of the total variation with respect to each coefficient in the finite element representation of the function to be reconstructed. If the finite element basis functions have local support, then the derivative with respect to a single coefficient depends just on its own and its neighbors’ values. Thus, this operation becomes similar to a image filtering or convolution step, which can be very effectively computed on a parallel machine, like the GPU [5], [13].

A. Bregman iteration in OSEM

In OSEM the set of LORs are partitioned into subsets and only a subset is evaluated in a single iteration. The set of iterations providing a covering for the complete LOR set is called OSEM cycle. In a Bregman phase where $\mathbf{p}^{(k)}$ is constant, no modification is required, we can use the same scheme independently of whether EM or OSEM iterations are executed. However, the update of gradient $\mathbf{p}^{(k)}$ should be modified since in a single step we have just partial information about the system matrix elements. Denoting the system matrix element in OSEM iteration s by $\mathbf{A}_{LV}^{(s)}$, we can express the

system matrix element as the sum of OSEM estimates:

$$\mathbf{A}_{LV} = \sum_{s=1}^S \mathbf{A}_{LV}^{(s)}$$

Thus the update of the gradient is

$$\mathbf{p}_V^{(k+1)} = \mathbf{p}_V^{(k)} + \frac{1}{\lambda} \sum_{s=1}^S \sum_{L=1}^{N_{\text{LOR}}} \left(\frac{\mathbf{A}_{LV}^{(s)} \frac{y_L}{\hat{y}_L^{n-S+s}}}{\mathbf{A}_{LV}^{(s)}} - 1 \right).$$

III. RESULTS

To demonstrate the results we run experiments on a simple 2D tomograph model, on a realistic fully 3D model of the Mediso's *AnyScan human PET/CT* [7], and finally reconstructed measured data as well.

A. 2D tomograph model

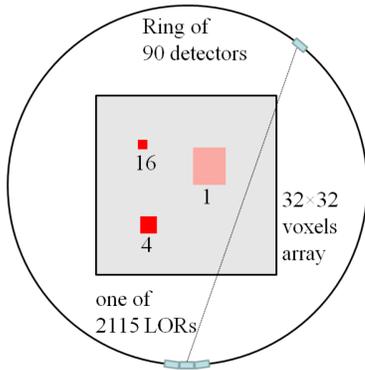


Fig. 2. A simple 2D tomograph model used in the experiments of this section (left). The detector ring contains 90 detector crystals and each of them is of size 2.2 in voxel units and participates in 47 LORs connecting this crystal to crystals being in the opposite half circle, thus the total number of LORs is $90 \times 47/2 = 2115$. The voxel array to be reconstructed is in the middle of the ring and has 32×32 resolution, i.e. 1024 voxels.

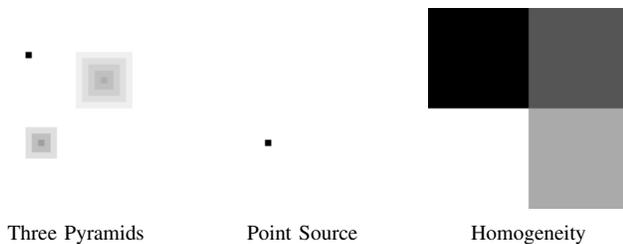


Fig. 3. 2D phantoms.

We first examine a simple 2D PET model where $N_{\text{LOR}} = 2115$ and $N_{\text{voxel}} = 1024$ [12]. We considered the *Three Pyramids* phantom that is projected with 2000 photon pairs and reconstructed with TV regularization and with the proposed method (Figs. 4 and 5). This is a low statistics measurement where regularization is necessary, making the ‘No TV’ method diverging. Note that Bregman iteration provides higher contrast and is significantly better than TV regularization if the

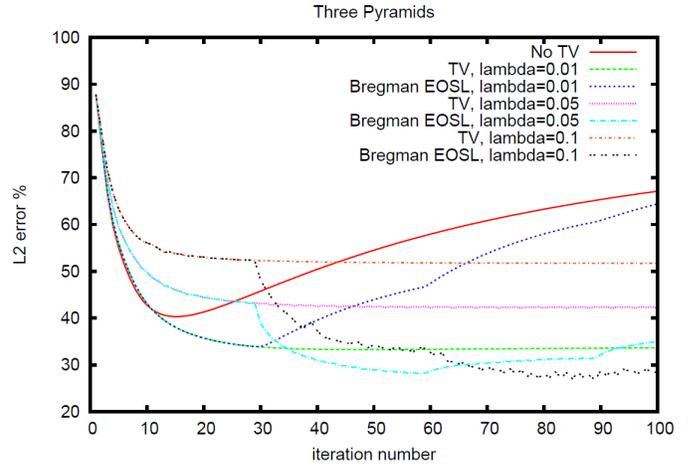


Fig. 4. L_2 error curves of the Three Pyramids phantom reconstruction.

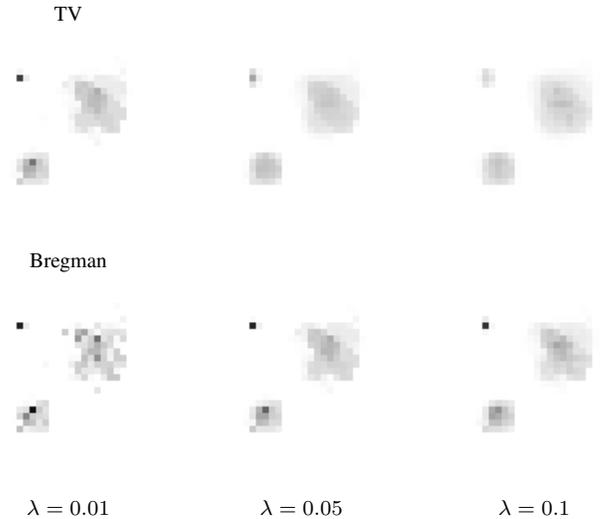


Fig. 5. Reconstructions of the Three Pyramids phantom.

regularization is strong enough. Too weak regularization, however, makes Bregman iteration similar to the non-regularized case.

We also considered two other 2D phantoms with extreme distributions (Fig. 3). The Point phantom has all activity concentrated in a single voxel while the Homogeneity phantom comprises four homogeneous squares. The reconstruction of the Point phantom can be evaluated in Figures 6 and 7. This measurement is of high statistics, so the L_2 error decreases even if no regularization is applied. On the other hand, the phantom has a high variation, so here regularization slows down the convergence and reduces the contrast. Total variation regularization stops the convergence on high error levels. Bregman iteration can help and make the process still converging.

When the Homogeneity is reconstructed (Figures 8 and 9), regularization is indeed needed since when there is no regularization (No TV), the error grows after an initial reduction due to overfitting. Here the best option is TV regularization, because the phantom itself has very low variation.

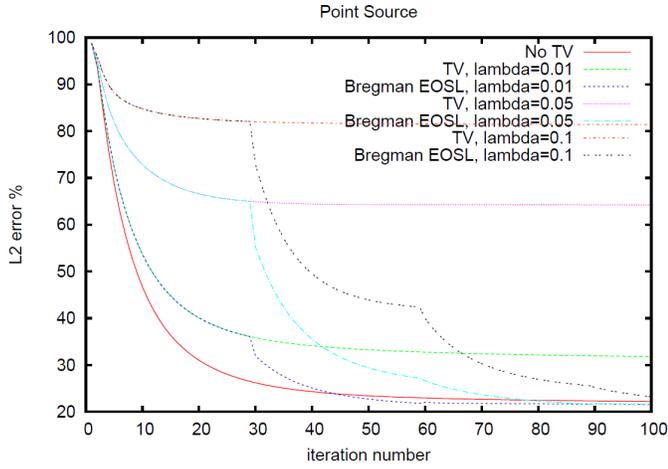


Fig. 6. The L_2 error curves of the Point phantom reconstruction.

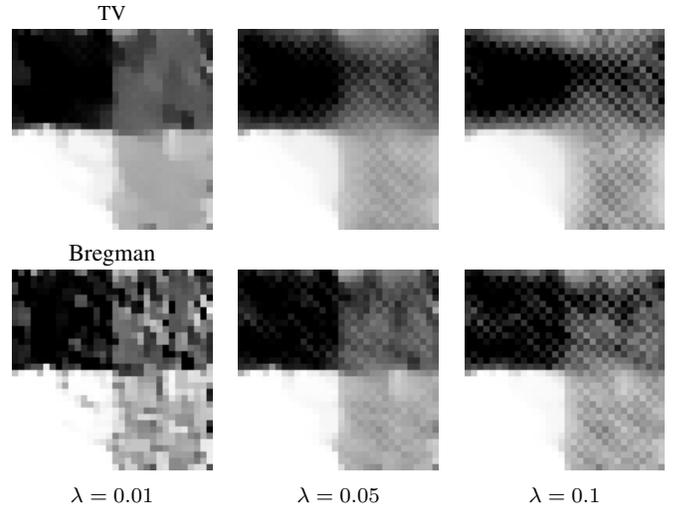


Fig. 9. Reconstructions of the Homogeneity phantom.

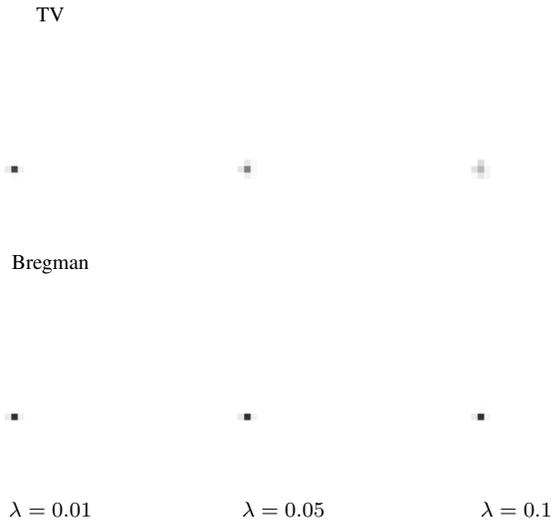


Fig. 7. Reconstructions of the Point Source phantom.

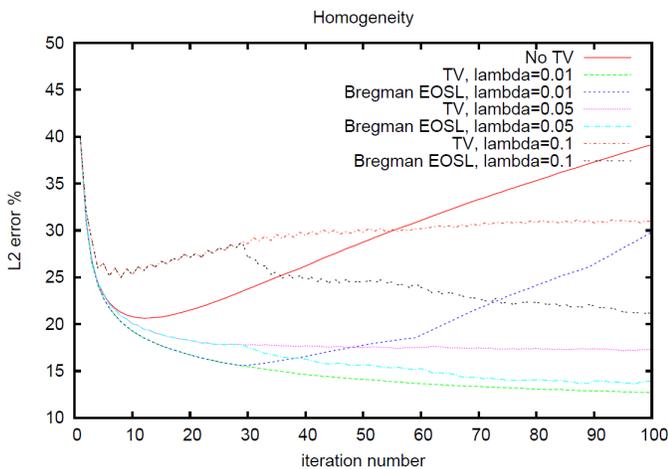


Fig. 8. The L_2 error curves of the Homogeneity phantom reconstruction.

B. 3D reconstructions

The proposed Bregman variant has been integrated into the TeraTomoTM fully-3D reconstruction system [2]. With this system we reconstructed a GATE simulated measurement of the NEMA NU2-2007 human IQ phantom, which again shows the superior accuracy and image quality of the proposed scheme over unregularized and TV-regularized methods (Figs. 10 and 12). The resolution of the reconstructed volume is $166 \times 166 \times 75$ voxels with edge size of 2 [mm]. We also tested the effect of the regularization parameter (Fig. 11). Note that the method is quite robust even to stronger regularization but textile patterns may show up in the image, which are due to the approximation error of the gradient.

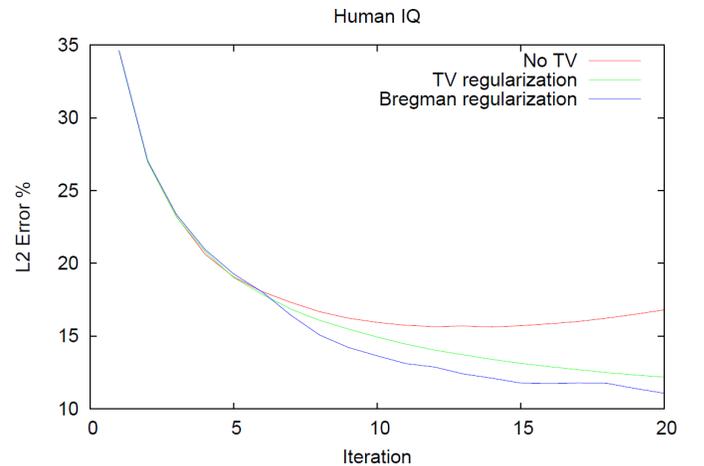


Fig. 12. L_2 error curve of the reconstructions of the Human IQ phantom.

We also reconstructed a Small animal IQ phantom projected with GATE simulating the NanoPET/CT pre-clinical PET-CT system [8] (Fig. 13). The resolution of the reconstructed volume is $200 \times 200 \times 200$ voxels with edge size of 0.4 [mm]. We can come to a similar conclusion concerning the regularization parameter, it can be set in a wider range but

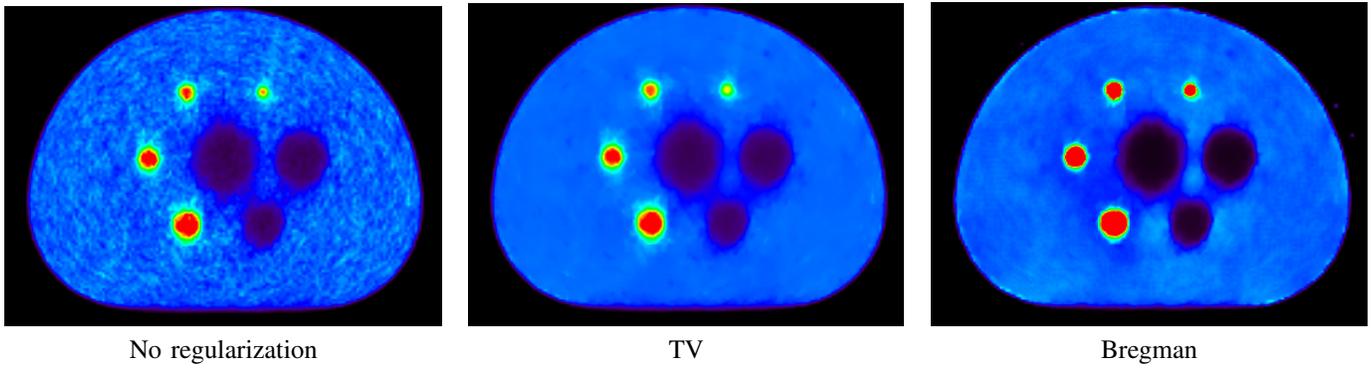


Fig. 10. Reconstructions of the Human IQ phantom.

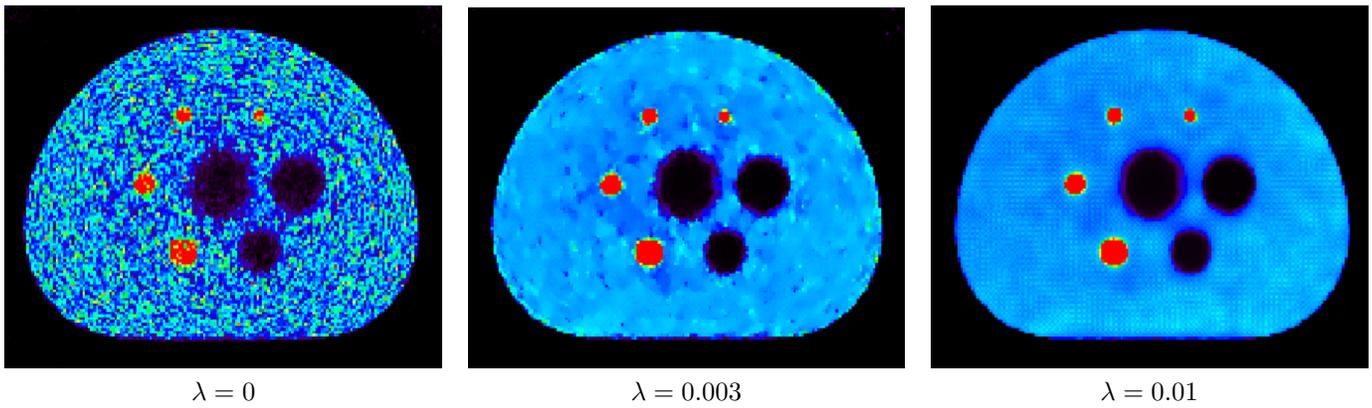


Fig. 11. Reconstructions of the Human IQ phantom with different regularization parameters.

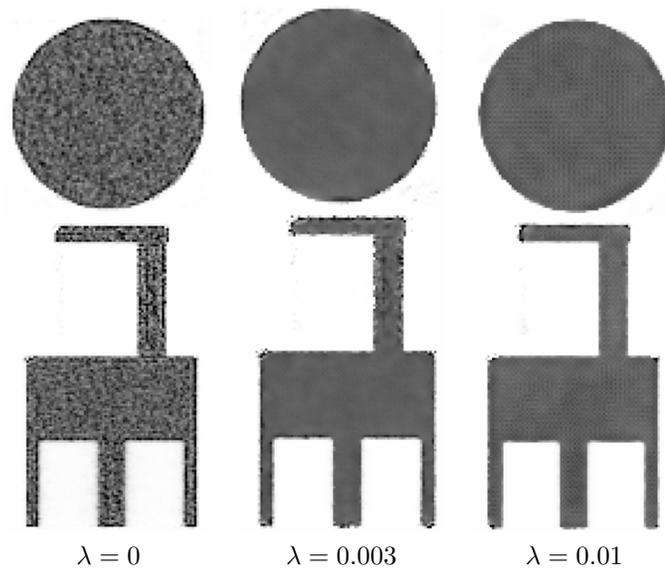


Fig. 13. Reconstructions of the Small Animal phantom with different regularization parameters.

too strong regularization may create textile patterns due to the approximation error of the gradient.

IV. CONCLUSIONS

In this paper we proposed a simple and efficient implementation of Bregman iteration for PET reconstruction. The proposed algorithm runs on GPUs. Based on the analysis and simulation results we can conclude that the Equalized One Step Late option works for Bregman iteration, making its implementation fairly simple. Bregman iteration is a promising alternative to TV regularization. Its only drawback is that we should maintain another voxel array p_V in addition to activity values x_V .

Acknowledgement

This work has been supported by OTKA K-104476.

REFERENCES

- [1] L. Bregman. The relaxation method for finding common points of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics*, 7:200–217, 1967.
- [2] M. Magdics et al. TeraTomo project: a fully 3D GPU based reconstruction code for exploiting the imaging capability of the NanoPET/CT system. In *World Molecular Imaging Congress*, 2010.
- [3] P. J. Green. Bayesian reconstructions from emission tomography data using a modified EM algorithm. *IEEE Trans. Med. Imaging*, 9(1):84–93, 1990.
- [4] P. C. Hansen. The L-curve and its use in the numerical treatment of inverse problems. In *Computational Inverse Problems in Electrocardiology*, ed. P. Johnston, *Advances in Computational Bioengineering*, pages 119–142. WIT Press, 2000.

- [5] M. Magdics, L. Szirmay-Kalos, B. Tóth, D. Légrády, Á. Cserkaszky, L. Balkay, B. Domonkos, D. Völgyes, G. Patay, P. Major, J. Lantos, and T. Bükki. Performance Evaluation of Scatter Modeling of the GPU-based “Tera-Tomo” 3D PET Reconstruction. In *IEEE Nuclear Science Symposium and Medical Imaging*, pages 4086–4088, 2011.
- [6] M. Magdics, B. Tóth, B. Kovács, and L. Szirmay-Kalos. Total variation regularization in PET reconstruction. In *KEPAF*, pages 40–53, 2011.
- [7] <http://www.mediso.com/products.php?fid=1,9&pid=73>.
- [8] <http://www.mediso.com/products.php?fid=2,11&pid=86>.
- [9] M. Persson, D. Bone, and H. Elmqvist. Total variation norm for three-dimensional iterative reconstruction in limited view angle tomography. *Physics in Medicine and Biology*, 46(3):853–866, 2001.
- [10] A. J. Reader and H. Zaidi. Advances in PET image reconstruction. *PET Clinics*, 2(2):173–190, 2007.
- [11] L. Shepp and Y. Vardi. Maximum likelihood reconstruction for emission tomography. *IEEE Trans. Med. Imaging*, 1:113–122, 1982.
- [12] L. Szirmay-Kalos, M. Magdics, B. Tóth, and T. Bükki. Averaging and metropolis iterations for positron emission tomography. *IEEE Trans. Med. Imaging*, 32(3):589–600, 2013.
- [13] L. Szirmay-Kalos, L. Szécsi, and M. Sbert. *GPU-Based Techniques for Global Illumination Effects*. Morgan and Claypool Publishers, San Rafael, USA, 2008.
- [14] M. Yan, A. Bui, J. Cong, and L.A. Vese. General convergent expectation maximization (EM)-type algorithms for image reconstruction. *Inverse Problems and Imaging*, 7:1007–1029, 2013.
- [15] W. Yin, S. Osher, J. Darbon, and D. Goldfarb. Bregman iterative algorithms for compressed sensing and related problems. *IAM Journal on Imaging Sciences*, 1(1):143–168, 2008.