Seventh Hungarian Conference on Computer Graphics and Geometry, Budapest, 2014

Analysis of Bregman Iteration in PET reconstruction

László Szirmay-Kalos[1] and Gábor Jakab[2]

[1]: Department of Control Engineering and Information Technology, Budapest University of Technology and Economics, Hungary
[2]: Mediso Ltd.

Abstract
Positron Emission Tomography reconstruction is ill posed. The result obtained with iterative maximum likelihood estimation is often unrealistic and has noisy behavior. The introduction of additional knowledge in the solution process is called regularization. *Common regularization methods penalize high frequency features or the total variation, thus compromise even valid solutions that have such properties. Another problem is the determination of the strength of regularization, for which no practically useful approach is available in complex problems like PET reconstruction. Bregman iteration offers a better choice enforcing regularization only where needed by the noisy data, thus we incorporate this strategy into our reconstruction algorithm. This paper analyzes Bregman iteration from the point of view of GPU-based PET reconstruction.*

1. Introduction

Tomography reconstruction is the *inverse problem* of particle transport, which requires the iteration of particle transport simulations and corrective back projections [6]. The inputs of the reconstruction are the measured values in *Lines of Responses* or *LORs*: $\mathbf{y} = (y_1, y_2, \ldots, y_{N_{\rm LOR}})$ The output of the reconstruction method is the *tracer density* function $x(\vec{v})$, which is approximated in a *finite function series* form:

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

where $\mathbf{x} = (x_1, x_2, \ldots, x_{N_{\rm voxel}})$ are the coefficients to be computed, and $b_V(\vec{v})$ $(V = 1, \ldots, N_{\rm voxel})$ 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 positron density $x(\vec{v})$ and the expected number of hits $\tilde{y}_L$ in LOR L is described by *scanner sensitivity*[8] $\mathcal{T}(\vec{v} \to L)$ that expresses the probability of generating an event in LOR L given that a positron is emitted in point $\vec{v}$ of volume $\mathcal{V}$:

$$\tilde{y}_L = \int_{\vec{v} \in \mathcal{V}} x(\vec{v}) \mathcal{T}(\vec{v} \to L) \mathrm{d}v = \sum_{V=1}^{N_{\rm voxel}} A_{LV} x_V. \qquad (2)$$

where A_{LV} is the *System Matrix* (*SM*):

$$A_{LV} = \int_{\mathcal{V}} b_V(\vec{v}) T(\vec{v} \to L) \mathrm{d}v. \qquad (3)$$

A system matrix element is the probability of that a positron is born in $\vec{v}$ with probability density $b_V(\vec{v})$ and generates an event in LOR L. Assuming that photon incidents in different LORs are independent random variables with Poisson distribution, the *Expectation Maximization* (ML-EM) algorithm[7] should maximize the following likelihood function:

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

$$\sum_{L=1}^{N_{LOR}} (y_L \log \tilde{y}_L - \tilde{y}_L) - \log y_L!.$$

subject to $x_V \geq 0$. Here $\log y_L!$ is independent of the voxel intensities, and thus can be ignored during optimization.

2. Regularization

The ML-EM algorithm, as the solution of inverse problems in general, is known to be ill-conditioned, which means that enforcing the maximization of the likelihood function may result in a solution with drastic oscillations and noisy behavior (Fig. 1). To handle this problem, we have to recognize

ground truth $n=10$ $n=30$ $n=50$ $n=70$ $n=90$

Figure 1: *Problem of overfitting to noise in ML-EM iteration when the Three Squares phantom is reconstructed.*

the cases caused by overfitting and exclude them from the solutions.

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. *Tychonoff regularization*[10] assumes the data set to be smooth and continuous, and thus enforce these properties during reconstruction. However, the typical data in PET reconstruction are different, 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. A better functional is the *Total Variation* (TV) of the solution[5, 3, 4]. 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

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

In higher dimensions, the total variation can be defined as the integral of the absolute value of the gradient:

$$TV(x) = \int_{\mathcal{V}} |\nabla x(\vec{v})| \mathrm{d}v. \tag{5}$$

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. *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 are not available. Secondly, constrained optimization is more difficult computationally than unconstrained optimization.
3. 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 third option, the merging of a penalty or regularization term $R(x)$ and the likelihood. The penalty term should be high for unacceptable solutions and small for acceptable ones. In PET reconstruction, under positivity constraint $x(\vec{v}) \geq 0$, we find where the sum of the negative likelihood and a term that is proportional to the total variation has its minimum:

$$E(x) = -\log \mathcal{L}(x) + \lambda R(x). \tag{6}$$

Here, λ is the *regularization parameter* that expresses the strength of the regularizing penalty term.

The objective can be regarded as a functional of tracer density function $x(\vec{v})$, or alternatively, having applied finite element decomposition to $x(\vec{v})$, as a function of voxel values $\mathbf{x} = (x_1, \ldots, x_{N_{voxel}})$:

$$E(\mathbf{x}) = -\log \mathcal{L}(\mathbf{x}) + \lambda R(\mathbf{x}). \tag{7}$$

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

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

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

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

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:

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

where

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

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

On the other hand, the forward scheme, also called *One Step Late* (OSL) scheme, does not need to solve any equation in a single iteration step:

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

However, the convergence of this scheme is proven only if $\lambda \partial R(x^{(n)}(\vec{v}))/\partial x_V$ is constant[7].

In this paper, we use this One Step Late option with Total Variation regularization and demonstrate that it is stable in practical situations. The partial derivatives of the total variation functional are:

$$\frac{\partial TV(x(\vec{v}))}{\partial x_V} = \int_{\mathcal{V}} \frac{\partial \vec{\nabla} x(\vec{v})/\partial x_V}{\sqrt{|\vec{\nabla} x(\vec{v})|^2}} \mathrm{d}v. \qquad (8)$$

The integrand of this formula has a singularity where the gradient is zero, which needs to be addressed.

The singularity can be avoided by re-defining the TV term by adding a small positive constant β:

$$TV(x) = \int_{\mathcal{V}} \sqrt{|\vec{\nabla} x(\vec{v})|^2 + \beta} \mathrm{d}v.$$

Note, however, that this modified TV term will not be invariant to sharp features anymore, and introduces some blurring. Instead of adding a small constant to the gradient, which eventually introduces blurring, primal-dual methods increase the free variables of the search to solve the problem of the singularity of the TV term[3,4].

3. Analysis of TV regularization

If both the negative likelihood and the regularization term are convex, then the objective function has a global minimum, which moves closer to the minimum of the regularization term if λ is increased. However, this means that the optimum of the likelihood is modified even if the noise level is small and thus no regularization is needed. Classical regularization terms measure the distance between the constant function and the actual estimate. The variation of the true solution is also penalized, so the optimum would be modified. As a result, Tychonoff regularization produces blurred, oversmoothed edges, TV regularization reduced contrast solutions having stair-case artifacts. The optimal weight can be obtained with *Hansen's L-curves*[2], 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

$$\frac{\partial TV(x(\vec{v}))}{\partial x_V} =$$

$$\frac{\partial \sum_{V=1}^{N_{voxel}} |x_V - x_{V-1}|}{\partial x_V} = \begin{cases} 2 & \text{if } x_V > x_{V-1} \text{ and } x_V > x_{V+1}, \\ -2 & \text{if } x_V < x_{V-1} \text{ and } x_V < x_{V+1}, \\ 0 & \text{otherwise.} \end{cases}$$

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

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

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

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

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 equal to one of its neighbors 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 point V. For this value, the ratio of the reconstruction with and without regularization, i.e. the contrast reduction is

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

Note that contrast reduction is not uniform for different voxels but depends on *sensitivity* $\sum_{L=1}^{N_{LOR}} \mathbf{A}_{LV}$, which expresses the probability that a positron born with probability density of basis function b_V is detected by LOR L. 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. To attack this problem, we propose an *Equalized One Step Late* (EOSL) scheme:

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

4. Bregman iteration

An optimal regularization term would have its minimum at the ground truth solution when $x = x_{\text{true}}$. In this case, regularization would not compromise the solution when regularization is not needed, and would become larger when the $R(x)$ is significantly different from $R(x_{\text{true}})$. Thus, an optimal regularization term would measure the "distance" $D(x, x_{\text{true}})$ between x and x_{true}. An appropriate distance function is the *Bregman distance*[1, 12, 11] that can be based on an arbitrary convex penalty term $R(x)$:

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

where $\mathbf{p}$ is the gradient of $R(x)$ at x_{true} if it exists:

$$\mathbf{p}_V = \frac{\partial R(x(\vec{v}))}{\partial x_V}.$$

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.

Replacing the TV functional by the Bregman distance induced by the TV functional, the goal of the optimization is

$$E(\mathbf{x}) = -\log \mathcal{L}(\mathbf{x}) + \lambda D(\mathbf{x}, \mathbf{x}^{(k)}). \tag{9}$$

Using the result of an arbitrary regularization for the special case of the Bregman distance, we obtain the following forward iteration:

$$x_V^{(n+1)} = \frac{x_V^{(n)} \cdot \sum_{L=1}^{N_{LOR}} \mathbf{A}_{LV} \frac{y_L}{\tilde{y}_L}}{\sum_{L=1}^{N_{LOR}} \mathbf{A}_{LV} + \lambda \frac{\partial D(x^{(n)}(\vec{v}), x^{(k)}(\vec{v}))}{\partial x_V}}.$$

where the partial derivatives of the Bregman distance are:

$$\frac{\partial D(x(\vec{v}), x^{(k)}(\vec{v}))}{\partial x_V} = \frac{\partial TV(x(\vec{v}))}{\partial x_V} - \mathbf{p}_V^{(k)}. \tag{10}$$

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. Iteration arrives at the optimum when the derivative of the objective function is zero:

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

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

Setting $\mathbf{p}^{(k+1)}$ to the gradient of the TV term and expressing it from the equation, we get

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

However, this step size may to too big and may make the iteration unstable, so we scale it and use the following update formula:

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

where δ is a controllable step size. The reconstruction algorithm based on Bregman iteration can be summarized as:

for $k = 1$ **to** K **do** *// Bregman iterations*
  for $n = (k-1)K$ **to** $kK - 1$ **do** *// subiterations*
    $x_V^{(n+1)} = x_V^{(n)} \cdot \frac{\sum_L \mathbf{A}_{LV} \frac{y_L}{\tilde{y}_L}}{\sum_L \mathbf{A}_{LV} + \lambda\left(\frac{\partial TV}{\partial x_V} - \mathbf{p}_V^{(k)}\right)}.$
  endfor
  $\mathbf{p}_V^{(k+1)} = \mathbf{p}_V^{(k)} + \delta \sum_L \left(\mathbf{A}_{LV} \frac{y_L}{\tilde{y}_L} - \mathbf{A}_{LV}\right).$
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[9].

5. Results

We examine a simple 2D PET model where a SM of dimensions $N_{\text{LOR}} = 2115$ and $N_{\text{voxel}} = 1024$ (Fig. 2).

We considered three phantoms, the *Three Squares* where each square has 64 Bq activity, the *Point Source* of 20 Bq activity, the *Homogeneity* of $2 \cdot 10^4$ Bq activity and using a Monte Carlo particle transport method, we simulated a 5 sec long measurement for all these phantoms (Fig. 3). It means that the Three Squares phantom is projected with 1000 photon pairs, the Point source with 100 photon pairs, and the

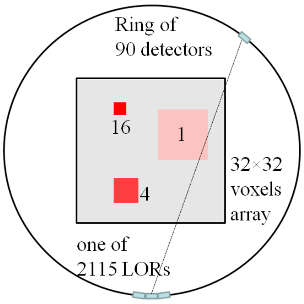


Figure 2: *2D tomograph model: The detector ring contains 90 detector crystals and each of them is of size 2.2 in voxel units and participates in 47 LORs connecting this crystal to crystals being in the opposite half circle, thus the total number of LORs is* $90 \times 47/2 = 2115$. *The voxel array to be reconstructed is in the middle of the ring and has* 32×32 *resolution, i.e. 1024 voxels. The ground truth voxel array of the Three Squares phantom has three hot squares of activity densities 1, 4, and 16 and of sizes* 8^2, 4^2, *and* 2^2.

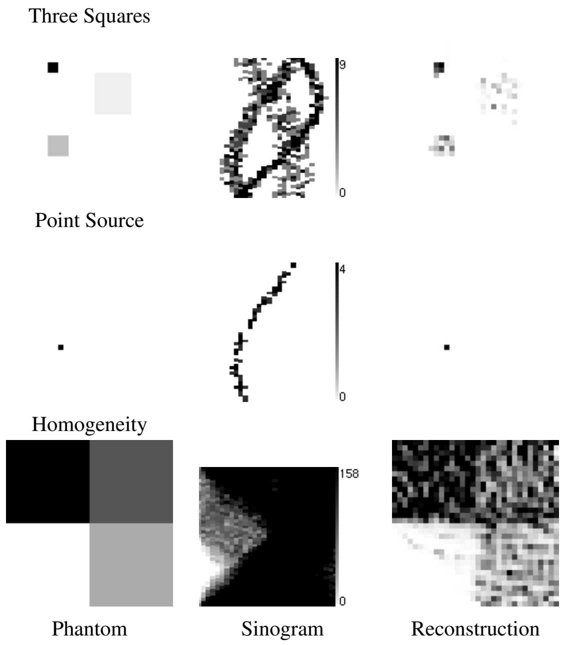


Figure 3: *The three phantoms used in the experiments, their random projections in sinogram parametrization, and the reconstructions without regularization.*

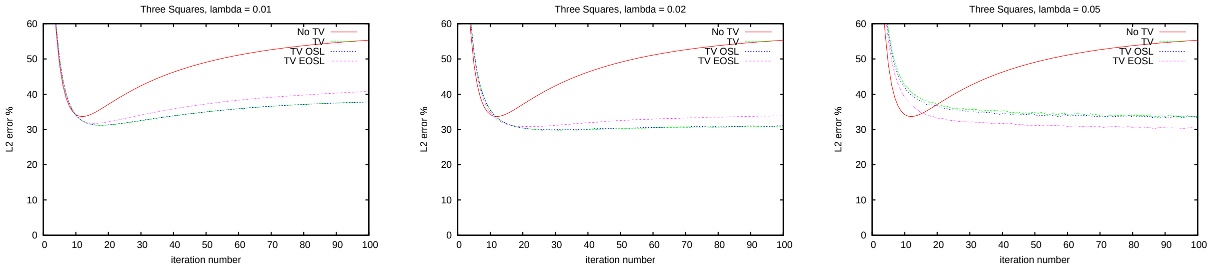
Homogeneity with 10^5 photon pairs, resulting in measured data having 1.21 Signal-to-Noise ratio (SNR) for the Three Squares, 1.07 SNR for the Point source and 1.69 SNR for the Homogeneity. Only geometric effects were simulated, we ignored scattering and absorbtion in the phantom. Three Squares is formed by three active squares of increasing size, and represents a realistic example between the other two extremes. Point Source and the Homogeneity represent two extreme cases. Point Source has a high variation since it has just a single voxel where the activity is non-zero and is well determined by the measurement. Thus, the reconstruction of Point Source would not need regularization, and regularization would just slow down the convergence. Homogeneity is formed by four constant activity squares, so the activity distribution is rather flat and the measurement is quite noisy. Such cases badly need regularization.

First we compared TV regularization strategies including the classical TV, One Step Late (OSL) and Equalized One Step Late (EOSL) for the Three Squares phantom. The L_2 error curves with respect to iteration number n are shown by Fig. 4. Note that there is no significant difference between the classical TV and the OSL options, while the OSL is much easier to compute. The Equalized version seems to be better when overregularization happens and poorer when the regularization is not strong enough, but the fact is that for the 2D tomograph model, the probability that a voxel event is detected is quite uniform thus the sensitivity is already equalized. The apparent difference is due to the fact that in the equalized option, the regularization parameter is scaled down by the sensitivity value, so similar results could be obtained if the global regularization parameter λ is scaled up in the equalized case.

Then, we considered the Bregman schemes and used only the TV-OSL as a reference. The L_2 error curves with respect to iteration number n are depicted by Figs. 5, 6, and 7, the reconstruction results by Figs. 8, 9, and 10. We set Bregman period K to 10 and step size δ to 1 in all tests.

We can make the following observations. Figs. 5 and 8 demonstrate the reconstruction of the Three Squares. This is a low statistics measurement where regularization is necessary. Total variation regularization and Bregman iteration with period 10 result in error level 30%, which remains the same for Bregman iteration even for strong regularization but gets slightly worse for total variation regularization. The One Step Late version of Bregman iteration outperforms all other methods despite the fact that it tends to oscillate when regularization is too strong.

The reconstruction of the Point phantom can be evaluated in Figs. 6 and 9. 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 error levels 13%, 25% and 60% when the parameter of regularization is $\lambda = 0.01$, $\lambda = 0.02$, and $\lambda = 0.05$, respectively, and there is no difference whether or not the One Step Late option is used. Bregman iteration can help and make



$\lambda = 0.01$ $\lambda = 0.02$ $\lambda = 0.05$

Figure 4: *L_2 error curves of the Three Squares reconstruction with TV regularization options.*

the process still converging, but the convergence is slower than without regularization.

When the Homogeneity is reconstructed (Figs. 7 and 10), regularization is indeed needed since when there is no regularization (No TV), the error grows after an initial reduction due to overfitting. TV-OSL and Bregman iteration perform similarly for this phantom. Regularization parameter $\lambda = 0.01$ seems to be too small since the error slightly increases. When $\lambda = 0.02$ the error converges to 21%, and with very strong regularization set by $\lambda = 0.05$, the error is also stable but at a higher level. The similarity of TV regularization and Bregman iteration in this case is explained by the fact that the phantom itself is flat, so its total variation is modest. The interesting fact is that One Step Late Bregman iteration performs better than both TV and Bregman regularization for lower regularization parameters, but gets the error to oscillate when the regularization is too strong. This oscillation is due to the fact that the step size of the optimization algorithm gets two large, so it steps over the minimum.

6. Conclusions

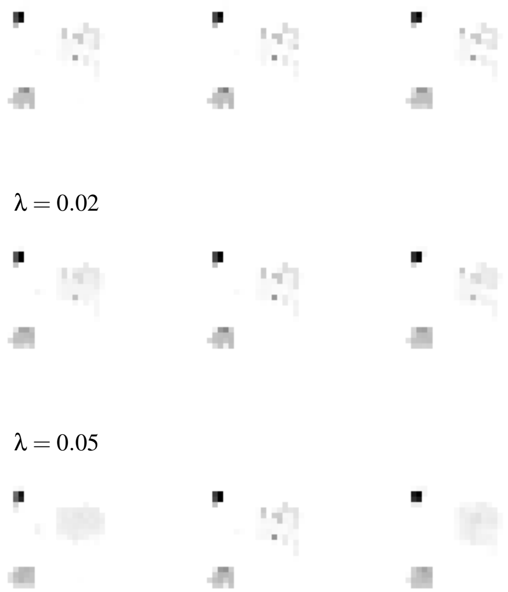
Based on the analysis and simulation results we can conclude that the One Step Late option works both for TV regularization and Bregman iteration, making their implementation fairly simple. Bregman iteration is a promising alternative to TV regularization and its one step late evaluation not only makes it more efficient to compute but also helps improving the error reduction. The only drawback of One Step Late Bregman iteration with respect to One Step Late TV regularization is that we should maintain another voxel array $\mathbf{p}_V$ in addition to activity values x_V. This not only doubles the storage space but also slows down the GPU implementation where the data transfer is the bottleneck.

There are two issues that need to be addressed in our future work. Firstly, in case of overregularization One Step Late Bregman iteration may have oscillating error curves which means that the process jumps over the minimum in each step. We plan to consider the cases when the derivative

$\lambda = 0.01$

$\lambda = 0.02$

$\lambda = 0.05$

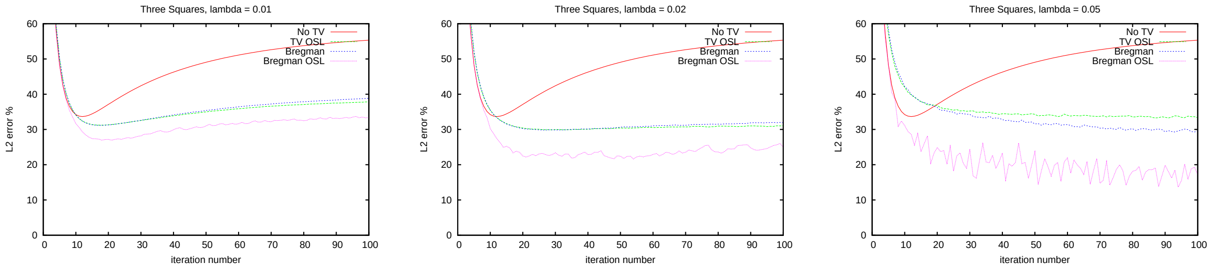


TV OSL Bregman Bregman OSL

Figure 8: *Reconstructions of the Three Squares phantom.*

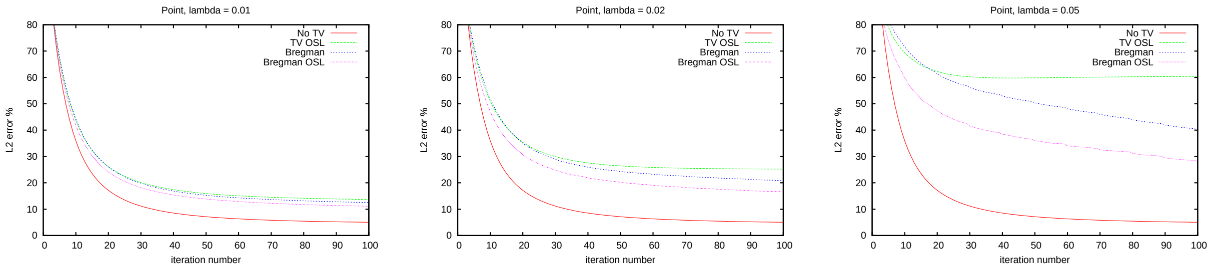
of the regularization term changes its sign and reduce the step size to enforce convergence.

Secondly, when the phantom has very high variation like in the case of Point phantom, even Bregman iteration slows down the convergence. This problem can be addressed by reducing the Bregman period, when the original convergence speed can be restored. However, in other cases, when regularization is badly needed, too low Bregman period cannot prevent the process from divergence. Thus our goal is to find



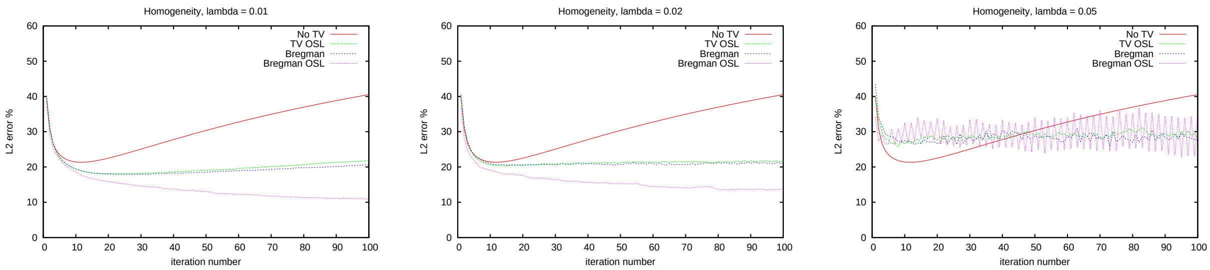
$\lambda = 0.01$ $\lambda = 0.02$ $\lambda = 0.05$

Figure 5: *L_2 error curves of the Three Squares reconstruction.*



$\lambda = 0.01$ $\lambda = 0.02$ $\lambda = 0.05$

Figure 6: *L_2 error curves of the Point phantom*



$\lambda = 0.01$ $\lambda = 0.02$ $\lambda = 0.05$

Figure 7: *L_2 error curves of the Homogeneity phantom*

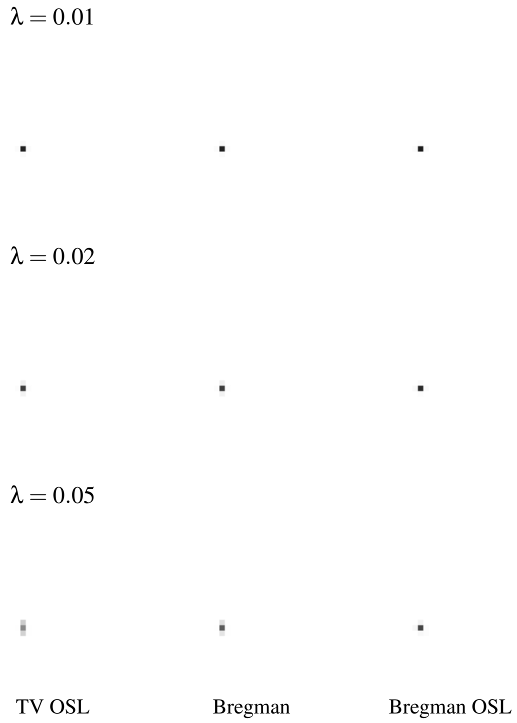


Figure 9: *Reconstructions of the Point phantom*

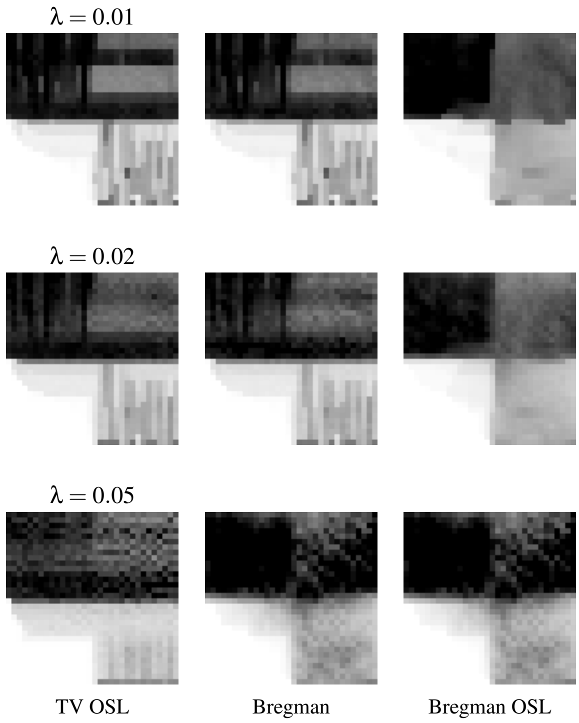


Figure 10: *Reconstructions of the Homogeneity phantom.*

an optimal Bregman period based on the properties of earlier iteration step.

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. 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.

3. F. Knoll, M. Unger, C. Diwoky, C. Clason, T. Pock, and R. Stollberger. Fast reduction of undersampling artifacts in radial MR angiography with 3D total variation on graphics hardware. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 23(2):103–114, 2010.

4. Milán Magdics, Balázs Tóth, Balázs Kovács, and László Szirmay-Kalos. Total variation regularization in PET reconstruction. In *KÉPAF*, pages 40–53, 2011.

5. M. Persson, D. Bone, and H. Elmqvist. Total variation norm for threedimensional iterative reconstruction in limited view angle tomography. *Physics in Medicine and Biology*, 46(3):853–866, 2001.

6. A. J. Reader and H. Zaidi. Advances in PET image reconstruction. *PET Clinics*, 2(2):173–190, 2007.

7. L. Shepp and Y. Vardi. Maximum likelihood reconstruction for emission tomography. *IEEE Trans. Med. Imaging*, 1:113–122, 1982.

8. 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.

9. L. Szirmay-Kalos, L. Szécsi, and M. Sbert. *GPU-Based Techniques for Global Illumination Effects*. Morgan and Claypool Publishers, San Rafael, USA, 2008.

10. A. N. Tychonoff and V. Y. Arsenin. *Solution of Ill-posed Problems*. Winston & Sons., Washington, 1977.

11. 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.

12. 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.