Total Variation Regularization in PET Reconstruction

Milán Magdics, Balázs Tóth, Balázs Kovács, and László Szirmay-Kalos

BME IIT, Hungary
magdics@iit.bme.hu

Abstract. Positron Emission Tomography reconstruction is ill posed. The result obtained with iterative maximum likelihood estimation has maximum probability, but is often unrealistic and has noisy behavior. The introduction of additional knowledge in the solution process is called *regularization*. Common regularization methods enforce continuity, smoothness or finite band-limits, but these are inappropriate in PET since typical solutions may have sharp features. Total variation regularization does not impose smoothness requirements and preserves edges and boundary surfaces, thus we incorporate this scheme in the reconstruction process. However, total variation involves the absolute value of the derivative, and the absolute value function cannot be differentiated at zero. We investigate different options to handle this problem, including the addition of a small constant to avoid singularity and the primal-dual algorithm, and also address the efficient evaluation of the total variation on the massively parallel GPU architecture.

1 Introduction

In positron emission tomography (PET) we need to find the spatial intensity distribution of positron–electron annihilations. During an annihilation event, two oppositely directed photons are produced [1], which may be detected by detectors. We collect the number of simultaneous photon incidents in detector pairs, also called *Lines Of Responses* or *LORs* (Figure 1).

The inputs of the reconstruction algorithm are the measured responses of LORs: $(y_1, y_2, \ldots, y_{N_{LOR}})$. The required output of the reconstruction method is the *emission density* function $x(\boldsymbol{v})$ that describes the number of photon pairs (i.e. the annihilation events) born in a unit volume around point $\boldsymbol{v}$. To represent the unknown function with finite data, it is approximated in a *finite element* form:

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

where $x_1, x_2, \ldots, x_{N_{voxel}}$ are the unknown coefficients and $b_V(\boldsymbol{v})$ $(V = 1, \ldots, N_{voxel})$ are pre-defined basis functions. If $b_V(\boldsymbol{v})$ is tri-linearly decreasing from a voxel to the neighboring voxels, then we work with tri-linear approximation, which

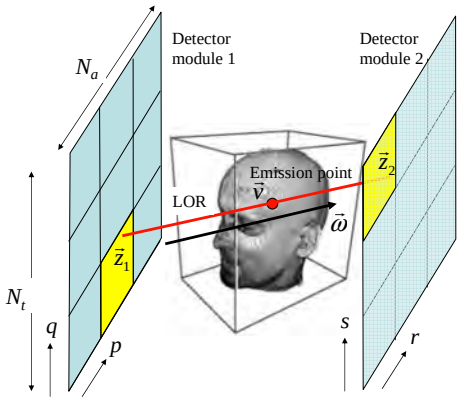


Fig. 1. Positron Emission Tomography. The pair of photons emitted at emission point $\boldsymbol{v}$ is detected by a pair of detectors at $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$, respectively. Detectors are organized into planar detector modules, which form a ring around the measured object.

is directly supported by the graphics hardware at no additional computational cost.

The correspondence between the coefficients of the emission function (voxel intensities) and the detector responses is built of the elemental conditional probability density that a photon pair lands at the two detectors of LOR L given that they are emitted at point $\boldsymbol{v}$ in directions $\boldsymbol{\omega}$ and $-\boldsymbol{\omega}$, which is denoted by $P(\boldsymbol{v}, \boldsymbol{\omega} \rightarrow L)$. This probability depends on the level of accuracy on which the physical phenomena are simulated. If we execute scatter, random, and attenuation compensation separately, and assume that detectors are ideally black, then only the geometry needs to be considered. In this case, the probability density is 1 if the line of place vector $\boldsymbol{v}$ and direction $\boldsymbol{\omega}$ intersects both detector surfaces of LOR L and zero otherwise (Figure 1).

If a photon pair is isotropically emitted from point $\boldsymbol{v}$, then the expectation of the photon incidents in LOR L is:

$$\tilde{y}_L = \sum_{V=1}^{N_{voxel}} x_V \int_{\mathcal{V}} \int_{\Omega_H} b_V(\boldsymbol{v}) P(\boldsymbol{v}, \boldsymbol{\omega} \rightarrow L) \frac{\mathrm{d}\omega}{2\pi} \mathrm{d}v = \sum_{V=1}^{N_{voxel}} \mathbf{A}_{LV} x_V \qquad (1)$$

where $\mathcal{V}$ is the volume of interest and Ω_H is the directional set of a hemisphere. Note that we should integrate only on the hemisphere since the photons of the emitted pair may be exchanged. We shall denote the probability that a photon pair born in voxel V contributes to LOR L by $\mathbf{A}_{LV}$:

$$\int_{\mathcal{V}} \int_{\Omega_H} b_V(\boldsymbol{v}) P(\boldsymbol{v}, \boldsymbol{\omega} \rightarrow L) \frac{\mathrm{d}\omega}{2\pi} \mathrm{d}v = \mathbf{A}_{LV}.$$

The task of the reconstruction is to find voxel intensities of $\mathbf{x} = (x_1, x_2, \ldots, x_{N_{voxel}})$ based on the measured LOR incidents in $\mathbf{y} = (y_1, y_2, \ldots, y_{N_{LOR}})$. Assuming that photon incidents in different LORs are independent random variables with Poisson distribution, the *expectation maximization* (EM) [7] algorithm should maximize the following likelihood function:

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

where c is independent of the voxel intensities, and thus can be ignored during optimization.

2 Regularization

The Expectation Maximization 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 (Figure 2).

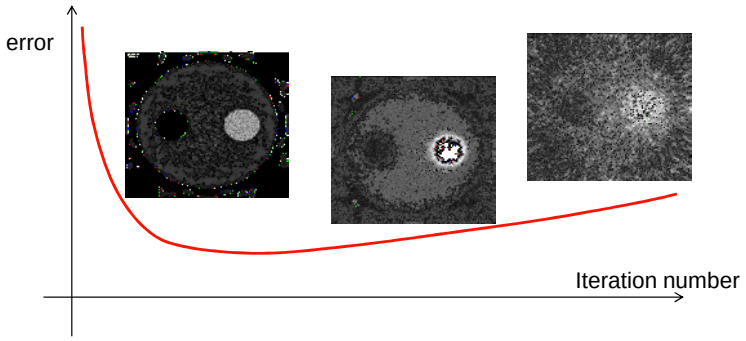


Fig. 2. The ill-conditioning of the EM algorithm. The reconstruction error first decreases then it starts to increase while noise levels get larger.

This problem can be attacked by *regularization methods* that include additional information in the objective function as a penalty term. The penalty term should be high for unacceptable solutions and small for acceptable ones. Standard regularization methods like Tikhonov regularization and Truncated Singular Value Decomposition (TSVD) assume 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. An appropriate penalty term is the *total variation* (TV) of the solution [5]. The

total variation measures the length of the path traveled by the function value while its parameter runs over the domain. In one-dimension and for differentiable functions, the total variation is the integral of the absolute value of the function's derivative. In higher dimensions, the total variation can be defined as the integral of the absolute value of the gradient:

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

Total variation regularization means the inclusion of the TV norm in the objective function that is optimized. In PET reconstruction, we find where the sum of the negative likelihood and a term that is proportional to the total variation has its minimum:

$$E(x) = -\log \mathcal{L}(x) + \lambda TV(x). \tag{2}$$

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

The objective can be regarded as a functional of emission function $x(\boldsymbol{v})$, or alternatively, having applied finite element decomposition to $x(\boldsymbol{v})$ as a function of voxel values $\mathbf{x} = (x_1, \ldots, x_{N_{voxel}})$. In our first approach, we adopt the second view. So, to minimize the objective function with respect to the voxel values, all partial derivatives must be zero:

$$\frac{\partial E(\mathbf{x})}{\partial x_V} = 0, \qquad V = 1, \ldots, N_{voxel}.$$

The derivative of the objective function is:

$$\frac{\partial E(\mathbf{x})}{\partial x_V} = -\sum_{L=1}^{N_{LOR}} \frac{\partial \tilde{y}_L}{\partial x_V} \frac{y_L}{\tilde{y}_L} + \sum_{L=1}^{N_{LOR}} \frac{\partial \tilde{y}_L}{\partial x_V} + \lambda \frac{\partial TV(x(\boldsymbol{v}))}{\partial x_V}.$$

At the optimal point these partial derivatives must be zero, thus we have

$$\sum_{L=1}^{N_{LOR}} \frac{\partial \tilde{y}_L}{\partial x_V} + \lambda \frac{\partial TV(x(\boldsymbol{v}))}{\partial x_V} = \sum_{L=1}^{N_{LOR}} \frac{\partial \tilde{y}_L}{\partial x_V} \frac{y_L}{\tilde{y}_L}.$$

which leads to the following iteration scheme to aim at the optimum:

$$x_V^{(n+1)} = \frac{x_V^{(n)} \cdot \sum_{L=1}^{N_{LOR}} \frac{\partial \tilde{y}_L}{\partial x_V} \frac{y_L}{\tilde{y}_L}}{\sum_{L=1}^{N_{LOR}} \frac{\partial \tilde{y}_L}{\partial x_V} + \lambda \frac{\partial TV(x^{(n)}(\boldsymbol{v}))}{\partial x_V}}.$$

The partial derivative of expected LOR value $\tilde{y}_L$ is the probability that an annihilation in the considered voxel contributes to this LOR:

$$\frac{\partial \tilde{y}_L}{\partial x_V} = \mathbf{A}_{LV}.$$

The partial derivatives of the total variation functional are:

$$\frac{\partial TV(x(\boldsymbol{v}))}{\partial x_V} = \int\limits_{\mathcal{V}} \frac{\partial \boldsymbol{\nabla} x(\boldsymbol{v})/\partial x_V}{\sqrt{|\boldsymbol{\nabla} x(\boldsymbol{v})|^2}} \mathrm{d}v. \tag{3}$$

The integrand of this formula has a singularity where the gradient is zero, which needs to be addressed. We consider two major ways to solve this problem, and several alternatives for the evaluation of the partial derivatives of the total variation.

2.1 Minimization with singularity elimination

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

$$TV(x) = \int\limits_{\mathcal{V}} \sqrt{|\boldsymbol{\nabla} x(\boldsymbol{v})|^2 + \beta} \mathrm{d}v$$

Note, however, that this modified TV term will not be invariant to sharp features anymore, and introduces some blurring.

One option to approximate the derivative formula is to use first a simple quadrature for the integral assuming that voxels are unit cubes:

$$TV(x) \approx \sum_{i,j,k} t(i,j,k)$$

where

$$t(i,j,k) = \sqrt{(x_{i+1,j,k} - x_{i,j,k})^2 + (x_{i,j+1,k} - x_{i,j,k})^2 + (x_{i,j,k+1} - x_{i,j,k})^2 + \beta}.$$

Derivating this approximation with respect to $x_V = x_{i,j,k}$, we obtain:

$$\frac{\partial TV(x(\boldsymbol{v}))}{\partial x_V} \approx \frac{3x_{i,j,k} - x_{i+1,j,k} - x_{i,j+1,k} - x_{i,j,k+1}}{t(i,j,k)} +$$

$$\frac{x_{i,j,k} - x_{i-1,j,k}}{t(i-1,j,k)} + \frac{x_{i,j,k} - x_{i,j-1,k}}{t(i,j-1,k)} - \frac{x_{i,j,k} - x_{i,j,k-1}}{t(i,j,k-1)}. \tag{4}$$

Better approximation schemes can be built by considering the actual finite element representation. Substituting the finite element representation of x, we obtain

$$\frac{\partial TV(x(\boldsymbol{v}))}{\partial x_V} = \sum_{V'=1}^{N_{voxels}} \int\limits_{\mathcal{V}} \frac{\boldsymbol{\nabla} b_V(\boldsymbol{v}) \cdot \boldsymbol{\nabla} b_{V'}(\boldsymbol{v})}{\sqrt{|\boldsymbol{\nabla} x(\boldsymbol{v})|^2 + \beta}} x_{V'} \mathrm{d}v. \tag{5}$$

If we consider the sample points as corner points of a 3D grid, then basis function $b_V(\boldsymbol{v})$ tri-linearly changes in a cell of this grid from corner point V to all neighboring corner points, where the X, Y, Z coordinates may differ by 1 voxel unit. As basis function $b_{V'}(\boldsymbol{v})$ has a similar property, the integrand can

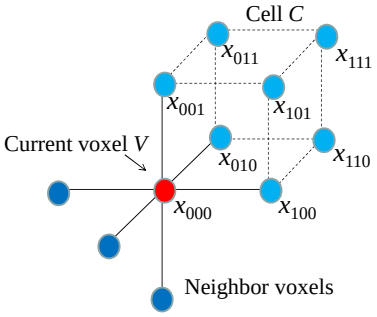
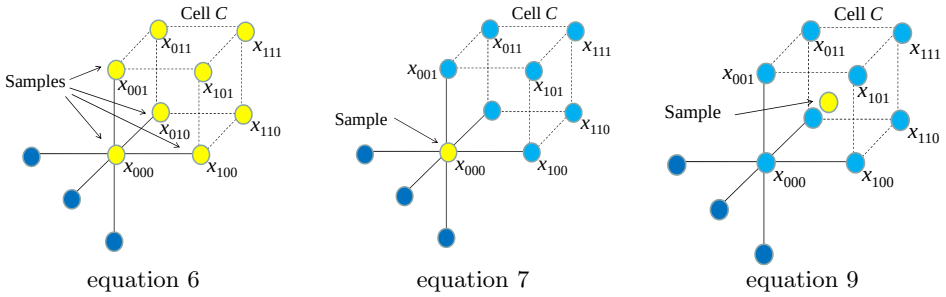


Fig. 3. The structure of the voxel grid. A voxel sample V has 27 neighbors and is a corner of 8 cells.

be non-zero only in those eight cells that share corner point V. Let us consider just a single cell $\mathcal{C}$ of eight corner points $x_{V'} = x_{000}, x_{100}, \ldots, x_{111}$ where x_{ijk} is the corner point, i.e. voxel, that can be obtained translating voxel V by vector (i, j, k). The other seven cells can be handled similarly and the derivative of the total variation functional will be the sum of eight terms.

In a single cell, discrete sample points are generated and the integral is estimated as a numerical quadrature. In order to compute the integral of equation 5 on the voxel data, we consider four different approximation schemes (Figure 4).



equation 6 equation 7 equation 9

Fig. 4. When we integrate in a cell $\mathcal{C}$ where voxel V is a corner, the samples of the integral quadrature may be at the corners of the cell (left), at V (middle), and in the center of the cell (right).

Sampling voxel corners for quadrature: In this first option, the samples for the numerical quadrature are the same where voxel values are represented and are the corners of the considered cell $\mathcal{V}$. As the basis functions are tri-linearly changing and are either 0 or 1 at these points, $\boldsymbol{\nabla} b_V(\boldsymbol{v})$ and $\boldsymbol{\nabla} b_{V'}(\boldsymbol{v})$ can be

simultaneously non zero only if V and V' differ at most only in one coordinate. When $\boldsymbol{v}$ is in V, then scalar product $\boldsymbol{\nabla} b_V(\boldsymbol{v}) \cdot \boldsymbol{\nabla} b_{V'}(\boldsymbol{v})$ equals to 3 if $V = V'$, it is -1 if V and V' differ in exactly one coordinate by 1, and zero in all other cases. When $\boldsymbol{v}$ is in neighbor voxel V', then scalar product $\boldsymbol{\nabla} b_V(\boldsymbol{v}) \cdot \boldsymbol{\nabla} b_{V'}(\boldsymbol{v})$ equals to 1 if $V = V'$, it is -1 if V and V' differ in exactly one coordinate by 1, and zero in all other cases.

Thus the integral quadrature using the cell corners as samples is:

$$\int_{\mathcal{C}} \frac{\partial TV(x(\boldsymbol{v}))}{\partial x_V} \mathrm{d}v \approx \frac{1}{8} \left(\frac{3x_{000} - x_{100} - x_{010} - x_{001}}{\sqrt{(x_{000} - x_{100})^2 + (x_{000} - x_{010})^2 + (x_{000} - x_{001})^2 + \beta}} + \right.$$
$$\frac{x_{000} - x_{100}}{\sqrt{(x_{100} - x_{000})^2 + (x_{100} - x_{110})^2 + (x_{100} - x_{101})^2 + \beta}} +$$
$$\frac{x_{000} - x_{010}}{\sqrt{(x_{010} - x_{110})^2 + (x_{010} - x_{000})^2 + (x_{010} - x_{011})^2 + \beta}} +$$
$$\left. \frac{x_{000} - x_{001}}{\sqrt{(x_{001} - x_{101})^2 + (x_{001} - x_{011})^2 + (x_{001} - x_{000})^2 + \beta}} \right). \tag{6}$$

This equation can be simplified by taking just a single quadrature sample at the lower left corner of the cell [6]:

$$\int_{\mathcal{C}} \frac{\partial TV(x(\boldsymbol{v}))}{\partial x_V} \mathrm{d}v \approx \frac{3x_{000} - x_{100} - x_{010} - x_{001}}{4\sqrt{(x_{000} - x_{100})^2 + (x_{000} - x_{010})^2 + (x_{000} - x_{001})^2 + \beta}}. \tag{7}$$

Keeping only the first term of the denominator when $x_{000} - x_{100}$ is divided, the second term when $x_{000} - x_{010}$ is divided, and the third when $x_{000} - x_{001}$, we obtain a particularly simple, but rather crude approximation:

$$\int_{\mathcal{C}} \frac{\partial TV(x(\boldsymbol{v}))}{\partial x_V} \mathrm{d}v \approx \frac{1}{4} \left(\mathrm{sign}(x_{000} - x_{100}) + \mathrm{sign}(x_{000} - x_{010}) + \mathrm{sign}(x_{000} - x_{001}) \right) \tag{8}$$

where sign is the signum function.

Sampling voxel centers for quadrature: In this option, we select the centers of the cells as sample points. The integral of equation (5) in cell $\mathcal{C}$ is

$$\int_{\mathcal{C}} \frac{\partial TV(x(\boldsymbol{v}))}{\partial x_V} \mathrm{d}v \approx \frac{3x_{000} + x_{100} + x_{010} + x_{001} - x_{101} - x_{011} - x_{110} - 3x_{111}}{4\sqrt{\Delta_X^2 + \Delta_Y^2 + \Delta_Z^2 + \beta}}, \tag{9}$$

where

$$\Delta_X = \sum_{jk} (x_{1jk} - x_{0jk}), \quad \Delta_Y = \sum_{ik} (x_{i1k} - x_{i0k}), \quad \Delta_Z = \sum_{ij} (x_{ij1} - x_{ij0}).$$

2.2 Primal-dual method

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]. They define a *dual variable* $\boldsymbol{p}$ that is supposed to converge to $\boldsymbol{\nabla} x/|\boldsymbol{\nabla} x|$. The relation between *primal variable* x and dual variable $\boldsymbol{p}$ is defined by the following extremal property:

$$|\boldsymbol{\nabla} x| = \sup_{|\boldsymbol{p}|\leq 1} (\boldsymbol{p} \cdot \boldsymbol{\nabla} x).$$

This extremal property is justified by the fact that the scalar product with a free vector $\boldsymbol{p}$ is maximum if this vector is parallel to the other vector $\boldsymbol{\nabla} x$ and is as long as possible, i.e. $\boldsymbol{p}$ is a unit vector because of the constraint. This vector is $\boldsymbol{p} = \boldsymbol{\nabla} x/|\boldsymbol{\nabla} x|$.

The objective function with the dual variable is converted to the following form:

$$E(x, \boldsymbol{p}) = -\log \mathcal{L}(x) + \lambda \int_{\mathcal{V}} \boldsymbol{p} \cdot \boldsymbol{\nabla} x \mathrm{d}v.$$

The reconstruction aims at finding the saddle point of this objective function:

$$(x, \boldsymbol{p}) = \min_{x} \sup_{|\boldsymbol{p}|\leq 1} E(x, \boldsymbol{p}).$$

The saddle point can be found by alternating minimization with respect to x and maximization with respect to $\boldsymbol{p}$ under constraint $|\boldsymbol{p}| \leq 1$.

In order to minimize $E(x, \boldsymbol{p})$ with respect to emission density function x, we force the first variation of the objective functional to be zero. The first variation is computed as follows:

$$\delta_E = \left.\frac{\partial E(x + \epsilon\delta_x)}{\partial \epsilon}\right|_{\epsilon=0} = -\sum_{L=1}^{N_{LOR}} \left(\frac{y_L}{\tilde{y}_L} - 1\right) \frac{\partial \tilde{y}_L(x + \epsilon\delta_x)}{\partial \epsilon} + \lambda \int_{\mathcal{V}} \boldsymbol{p} \cdot \frac{\partial \boldsymbol{\nabla}(x + \epsilon\delta_x)}{\partial \epsilon} \, \mathrm{d}v$$

where $\delta_x(\boldsymbol{v})$ is an arbitrary perturbation function which is zero at the volume boundary. Substituting the interpretation of the expected LOR incidents $\tilde{y}_L$ from equation 1, we get:

$$\frac{\partial \tilde{y}_L(x + \epsilon\delta_x)}{\partial \epsilon} = \frac{\partial}{\partial \epsilon} \int_{\mathcal{V}} (x + \epsilon\delta_x) \int_{\Omega_H} P(\boldsymbol{v}, \boldsymbol{\omega} \to L) \frac{\mathrm{d}\omega}{2\pi} \, \mathrm{d}v = \int_{\mathcal{V}} \delta_x \int_{\Omega_H} P(\boldsymbol{v}, \boldsymbol{\omega} \to L) \frac{\mathrm{d}\omega}{2\pi} \, \mathrm{d}v.$$

We use partial integration for the regularization term:

$$\int_{\mathcal{V}} \boldsymbol{p} \cdot \frac{\partial \boldsymbol{\nabla}(x + \epsilon\delta_x)}{\partial \epsilon} \, \mathrm{d}v = \int_{\mathcal{V}} \boldsymbol{p} \cdot \boldsymbol{\nabla} \delta_x \, \mathrm{d}v = \int_{\partial\mathcal{V}} (\boldsymbol{p} \cdot \boldsymbol{n}) \delta_x \, \mathrm{d}v - \int_{\mathcal{V}} \boldsymbol{\nabla} \cdot \boldsymbol{p} \delta_x \, \mathrm{d}v.$$

where $\partial\mathcal{V}$ is the boundary of the volume and $\boldsymbol{n}$ is the normal vector of the boundary surface. The first term here is zero since we assume that the perturbation δ_x is zero at the boundary.

Putting all these together, we obtain first variation:

$$\delta_E = \int_{\mathcal{V}} \delta_x \left(- \sum_{L=1}^{N_{LOR}} \left(\frac{y_L}{\tilde{y}_L} - 1 \right) \int_{\Omega_H} P(\boldsymbol{v}, \boldsymbol{\omega} \to L) \frac{\mathrm{d}\omega}{2\pi} - \lambda \boldsymbol{\nabla} \cdot \boldsymbol{p} \right) \mathrm{d}v$$

As this variation must be zero for arbitrary perturbation δ_x that is zero at the boundary, the equation of the extremum is:

$$- \sum_{L=1}^{N_{LOR}} \left(\frac{y_L}{\tilde{y}_L} - 1 \right) \int_{\Omega_H} P(\boldsymbol{v}, \boldsymbol{\omega} \to L) \frac{\mathrm{d}\omega}{2\pi} - \lambda \boldsymbol{\nabla} \cdot \boldsymbol{p} = 0.$$

As we store the emission intensity in a finite element form, instead of requiring this function to be constantly zero, we expect its integrals in every voxel to be zero:

$$- \sum_{L=1}^{N_{LOR}} \left(\frac{y_L}{\tilde{y}_L} - 1 \right) A_{LV} - \lambda \int_V \boldsymbol{\nabla} \cdot \boldsymbol{p} \mathrm{d}v = 0.$$

The iteration formula targeting this optimum is:

$$x_V^{(n+1)} = \frac{x_V^{(n)} \cdot \sum_{L=1}^{N_{LOR}} \mathbf{A}_{LV} \frac{y_L}{\tilde{y}_L}}{\sum_{L=1}^{N_{LOR}} \mathbf{A}_{LV} - \lambda \int_V \boldsymbol{\nabla} \cdot \boldsymbol{p} \mathrm{d}v}.$$

To complete our primal-dual method, we have to update our dual variable in every iteration. We compute the weighted average of the current estimate and the gradient of the scalar field, and also enforce the requirement of unit length:

$$\boldsymbol{q} = \boldsymbol{p}^{(n)} + \tau^{(n)} \boldsymbol{\nabla} x^{(n)} \tag{10}$$

where $\tau^{(n)} = 0.3 + 0.02n$ is the weight, which should be chosen carefully because it strongly affects the convergence time. Since $\boldsymbol{p}$ should have maximum unit length:

$$\boldsymbol{p}^{(n+1)} = \frac{\boldsymbol{q}}{\max(1, |\boldsymbol{q}|)}. \tag{11}$$

3 Total variation control on the GPU

The iteration step including the total variation requires 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.

4 Results

The presented algorithm have been implemented in CUDA and run on nVidia GeForce 480 GFX GPUs. We have modeled the PET system of *nanoPET/CT* [4] consisting of twelve square detector modules organized into a ring, and the system measures LORs connecting a detector to three other detectors being at the opposite sides of the ring, which means that $12 \times 3/2 = 18$ module pairs need to be processed. A detector module consists of $N_{det} = 81 \times 39$ crystal detectors. The total number of LORs is $N_{LOR} = 18 \cdot (81 \times 39)^2 \approx 180$ million.

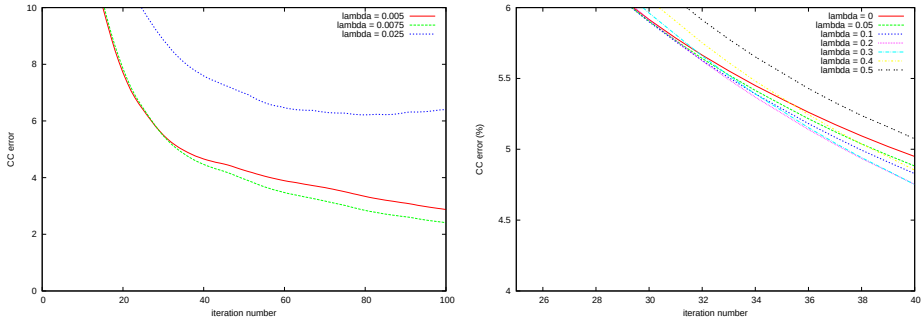


Fig. 5. Error curves with respect to weight λ of the TV regularization. The left image shows the results obtained with equation 9, the right the primal-dual algorithm.

The error curves are shown in Figure 5. The CC error is computed as the difference of one and the correlation of the reconstructed and the reference voxel arrays. Note that the primal-dual algorithm is not better than the simpler TV regularization scheme. As its memory requirement is much higher since not only the actual scalar field but is gradient should also be stored, we concluded that the primal-dual method is not worth taking. On the other hand, the accuracy of gradient estimation should be maximized since it affects the convergence but its computational cost is negligible with respect to the other work (Figure 6). Thus, we should prefer equations 6, 7, 9, and 4 to equation 8.

In the error plot of the singularity elimination method, we can observe that total variation regularization slightly reduces the speed of the initial convergence, but makes the error keep decreasing even at higher iteration numbers (Figure 5). However, too high TV weight ($\lambda = 0.05$) causes blurring (Figure 7). With TV regularization using proper weight ($\lambda = 0.0075$), we can obtain very attractive, 2.4% error level after 100 iteration steps and 4% error level after 50 steps. 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

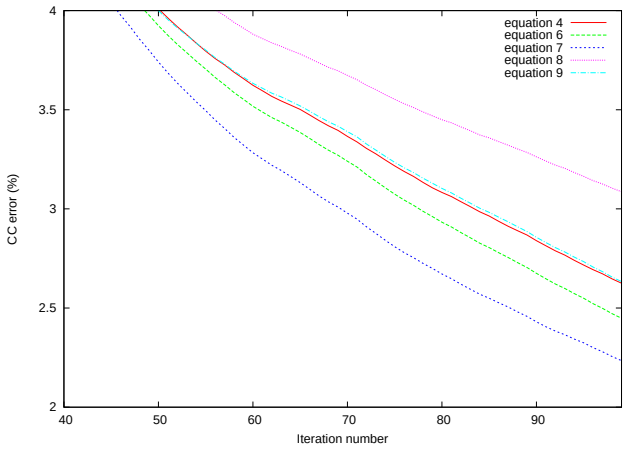
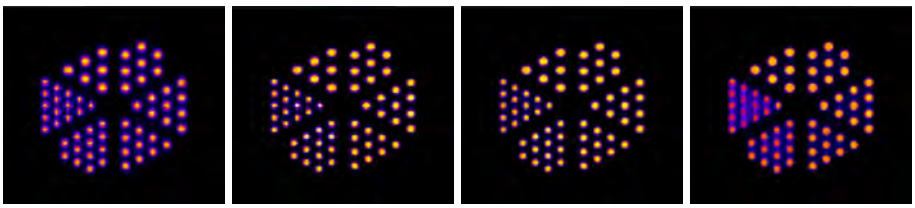


Fig. 6. Comparison of different gradient estimation formulae, including equation 4, 6, 7, 8, and 9.

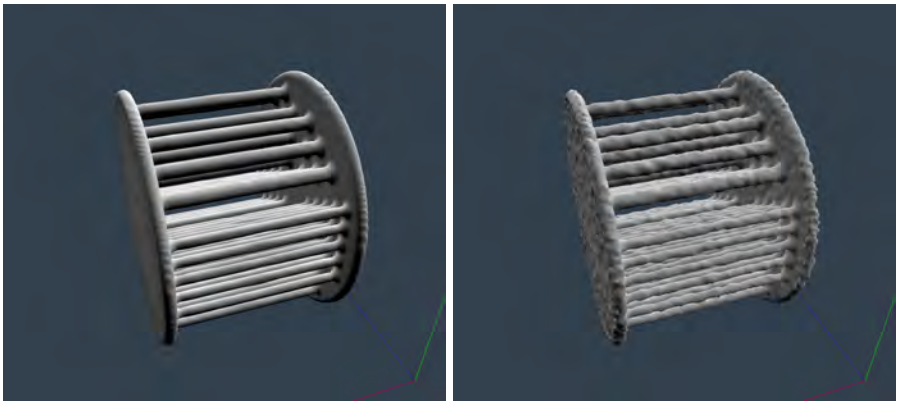
likelihood and the total variation will be functions of the regularization parameter. However, the algorithm developed to locate the maximum curvature points assumes Tikhonov regularization. Our next research objective is to generalize the L-curves method for the likelihood function and total variation term.



no TV | TV with $\lambda = 0.005$ | TV with $\lambda = 0.0075$ | TV with $\lambda = 0.05$

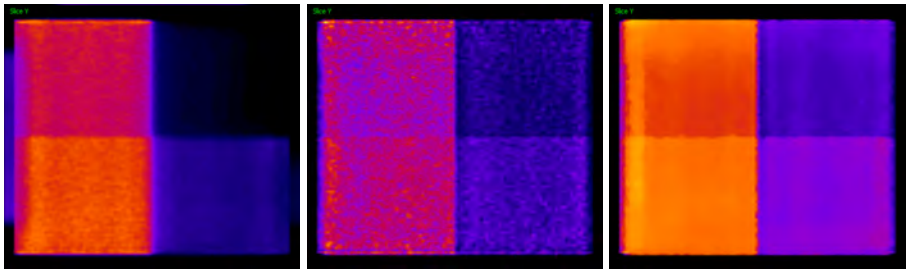
Fig. 7. Effects of TV regularization for $N_{detline} = 4$ lines per LOR. We executed 50 stochastic iteration steps in all experiments and turned filtered sampling off.

Figure 8 depicts an isosurface of the Derenzo phantom reconstructed at $128 \times 144 \times 144$ resolution using $N_{detline} = 1$ lines per LOR, $N_{march} = 36$ steps per line, and turning total variation regularization on with $\lambda = 0.0075$. We executed 10 iterations on reduced voxel resolution and another 40 iterations on the target resolution. On the full resolution, a single forward projection needed 2 seconds, a backprojection 16 seconds. During the initial steps, when the volume of half



320 MBq 16 MBq

Fig. 8. 3D isosurface renderings of Derenzo phantoms of 320 MBq and 16 MBq total activities, reconstructed in 3 minutes on a PC equipped with two GPU cards. The two cases corresponds to two significantly different noise levels. The higher the activity, the lower the relative standard deviation of the measured data, i.e. the lower the noise level.



no TV, 20 iterations | no TV, 100 iterations | TV with $\alpha = 0.0075$, 100 iterations

Fig. 9. Effects of TV regularization on the linearity-homogenity phantom simulated and reconstructed with realistic detector modeling. The phantom consists of eight homogeneous cubes with different activities. Without TV normalization, the high frequency noise increases as the iteration proceeds.

resolution ($64 \times 72 \times 72$) is iterated, an iteration cycle of both forward projection and backprojection required just 2 seconds. The complete reconstruction took 12 minutes on a single GPU and 6 minutes on two GPUs running in parallel.

5 Conclusions

This paper investigated total variation regularization schemes for GPU based algorithm for the reconstruction of PET measurements. As in a steepest descent search the derivative of the total variation functional is needed, which can be computed just from local data, this approach is very efficient on the massively parallel GPU architecture. With respect to the expected LOR value computation, the total variation regularization has a negligible overhead. Thus, the accuracy of the gradient estimation is worth setting to its maximum. Although the primal-dual method has theoretical advantages, it has not proven to be better than the simple elimination of the singularity in the absolute value function, and its four-fold memory requirements are not justified. Concerning the simple approach, we also showed that increasing the accuracy of the gradient estimation helps improving the convergence of the algorithm.

Acknowledgements

This work has been supported by the TeraTomo project of the National Office for Research and Technology, OTKA K-719922, and by the scientific program of the "Development of quality-oriented and harmonized R+D+I strategy and functional model at BME" (Project ID: TMOP-4.2.1/B-09/1/KMR-2010-0002). The authors are grateful to NVIDIA for donating the GeForce 480 GPU cards.

References

1. Geant. Physics reference manual, Geant4 9.1. Technical report, CERN, 2007.
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. Florian Knoll, Markus Unger, Clemens Diwoky, Christian Clason, Thomas Pock, and Rudolf 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, April 2010.
4. Mediso. http://www.bioscan.com/molecular-imaging/ nanopet-ct.
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. Mikael Perssona, Dianna Boneb, and Hakan Elmqvista. Three-dimensional total variation norm for SPECT reconstruction. *Nuclear Instruments and Methods in Physics Research*, pages 98–102, 2001.
7. L. Shepp and Y. Vardi. Maximum likelihood reconstruction for emission tomography. *IEEE Trans. Med. Imaging*, 1:113–122, 1982.