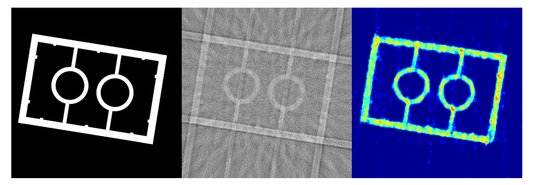
X-ray tomography on a Lego block

Santeri Kaupinmäki & Thomas Malcher

santeri.kaupinmaki@helsinki.fi — thomas.malcher@helsinki.fi



Introduction

In this project we performed X-ray tomography on a 3×2 Lego block using an X-ray with a resolution of 50 micrometers and an acceleration voltage of 50 kV. The Lego block includes hard edges, curved features, and hollow sections, providing for an interesting interior structure that needs to be reconstructed. The measured data $\mathbf{m}$ can be modeled as $\mathbf{m} = A\mathbf{f} + \epsilon$ where $\mathbf{f}$ is the discretized object slice that gets projected onto the detector along a line and $\mathbf{m}$ is the vectorized set of these projection lines, each from a different angle (Figure 1).

To solve the inverse problem, that is from a given set of projection lines $\mathbf{m}$ find the 2D object slice $\mathbf{f}$, we had to use a regularized approach due to the ill possessedness of the problem. As our object has large homogeneous areas and sharp edges we chose a regularization penalty term that boosts solutions with such features. We used matrix-free (approximate) total variation regularization implemented in iterative form using the Barzilai-Borwein method.

The choice of regularization parameter was made using a combination of the S-curve method, Morozov's discrepancy principle, as well as percent difference between the reconstruction and a binary photograph of the bottom of the Lego block. The photograph was also used in the S-curve method to determine the a priori sparsity.

The stopping criteria for the iteration was a combination of a threshold limit based on the 2-norm proximity of successive iterations, and a hard limit to avoid a non-terminating computation.

Materials and Methods

In this work we assumed a parallel x-ray beam model (Figure 1). Although the x-rays in the experiment actually have a fan-like profile, the parallel approximation is justified by having a small imaging object that is relatively distant from the x-ray source.

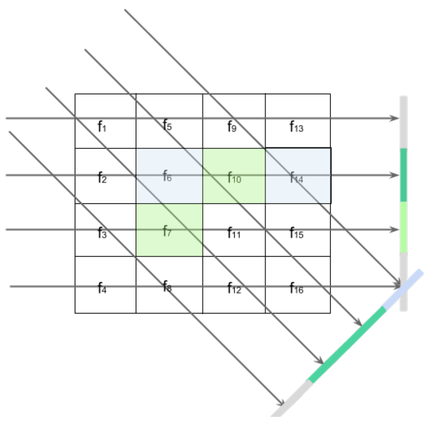


Figure 1: Parallel beam model

Total Variation Regularization

In total variation (TV) regularization we consider the minimization problem (1) to solve the inverse problem.

$$T_\alpha(\mathbf{m}) = \arg\min_{\mathbf{f}\in\mathbb{R}^{n^2}} \left\{ \|A\mathbf{f} - \mathbf{m}\|_2^2 + \alpha\|L_V\mathbf{f}\|_1 + \alpha\|L_H\mathbf{f}\|_1 \right\} \tag{1}$$

Where $\|L_H\mathbf{f}\|_1$ and $\|L_V\mathbf{f}\|_1$ are the regularization terms and α the regularization parameter.

Barzilai-Borwein Gradient Descent

For the gradient descent implementation of TV regularization, we need the expression in (1) to be smooth, and we take the discrete differentiation matrix L, so we approximate it by (2)

$$G_\beta(\mathbf{f}) = \arg\min_{\mathbf{f}\in\mathbb{R}^{n^2}} \left\{ \|A\mathbf{f} - \mathbf{m}\|_2^2 + \alpha\sum_{j=1}^{n^2} |f_j - f_{j-1}|_\beta + \alpha\sum_{j=1}^{n^2} |f_j - f_{j-n}|_\beta \right\} \tag{2}$$

$$f_j = 0 \text{ for } j < 1 \text{ and } j > n^2 \tag{3}$$

where $|t|_\beta := \sqrt{t^2 + \beta}$ for some small $\beta > 0$. This modified absolute value is smooth, and so the gradient of $G_\beta(\mathbf{f})$ is well-defined. With this, we define the Barzilai-Borwein gradient descent method,

$$\mathbf{f}^{(\ell+1)} = \mathbf{f}^{(\ell)} - \delta_\ell \nabla G_\beta(\mathbf{f}^{(\ell)}) \tag{4}$$

$$\delta_\ell = \frac{(\mathbf{f}^{(\ell)} - \mathbf{f}^{(\ell-1)})^T(\mathbf{f}^{(\ell)} - \mathbf{f}^{(\ell-1)})}{(\mathbf{f}^{(\ell)} - \mathbf{f}^{(\ell-1)})^T(\nabla G_\beta(\mathbf{f}^{(\ell)}) - \nabla G_\beta(\mathbf{f}^{(\ell-1)}))} \tag{5}$$

The initial step length δ_1 was chosen to be the small number, $\frac{1}{10000}$. We chose the parameter β by experimenting on the data.

Iteration Limit

As a first line of attack, we stopped the iteration of the Barzilai-Borwein method when $\|G_\beta(\mathbf{f}^{(\ell)}) - G_\beta(\mathbf{f}^{(\ell-1)})\| \leq 10^{-4}$. In case this condition was never met, at least in reasonable time, we further set a hard iteration limit of 200.

Choice of Regularization Parameter

The choice of regularization parameter was made using results from the S-curve method, Morozov's discrepancy principle, and the percent difference between the reconstruction and a binary photograph of the Lego block. To determine the a priori sparsity, $\hat{S}$, of the Lego block for the S-curve method, we used an optical photograph of the bottom of the Lego block and transformed the colours into white (Lego) and black (background). The image was then normalized so that the norm of its sinogram corresponded to the norm of our X-ray projection sinogram. The black and white photograph can be seen in the header, and we found that $\hat{S} = 102$. For Morozov discrepancy principle, we constructed a curve with respect to α of the form $\|AT_\alpha(\mathbf{m}) - \mathbf{m}\|$, and compared it to the constant curve $\delta = \|\varepsilon\|$. And for the percent difference, we simply took the same optical image we used in the S-curve method and calculated $\frac{\|T_\alpha(\mathbf{m}) - \mathbf{f}_{optical}\|}{\|\mathbf{f}_{optical}\|}$.

Results

We applied TV regularization with sparse data consisting of 18 x-ray projections which were evenly chosen from the range 0 to 180 degrees. A grid of 80 different parameter configurations was tested. For β we chose 4 values between 10^{-1} and 10^{-5}, and for α we chose 20 values between 10^{-4} and 10^{7}.

We found that 100 iterations are a reasonable tradeoff between time and reconstruction quality for our experiments; the difference between reconstructions done with 100 and 200 iterations was found to be in the order of 10^{-6}. It was found that all of our methods for choosing the regularization parameter were ineffective; as the three curves produced by these methods, as seen in Figure (3), show a nearly constant behaviour until we reach $\alpha \sim 10^4$, at which point the curves blow up.

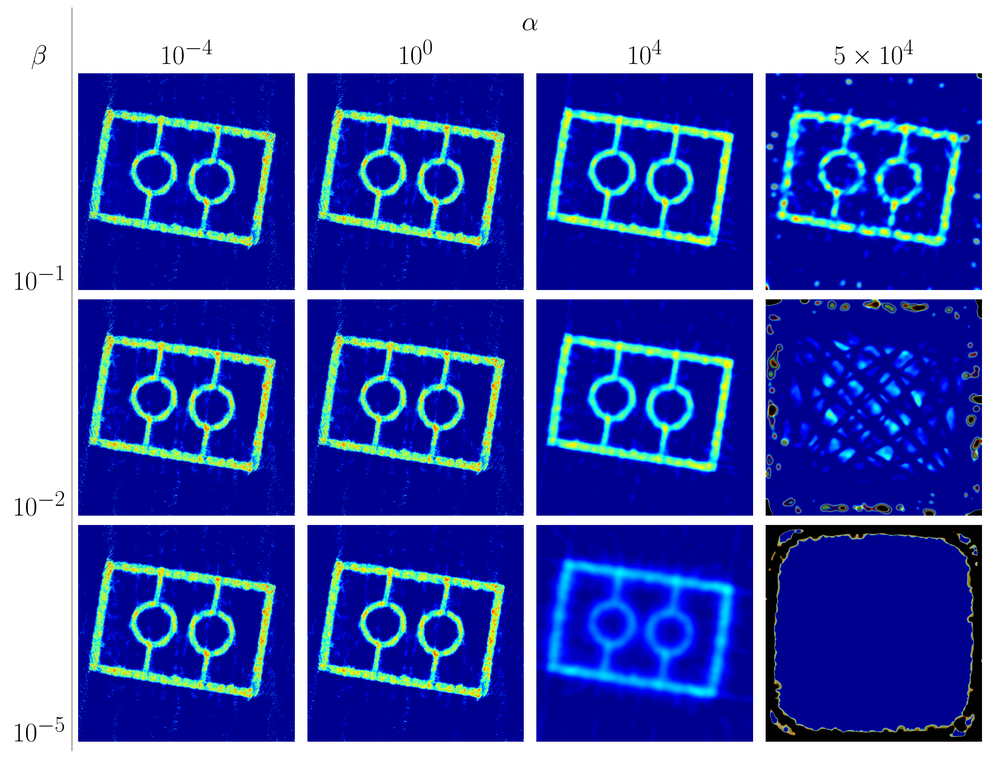


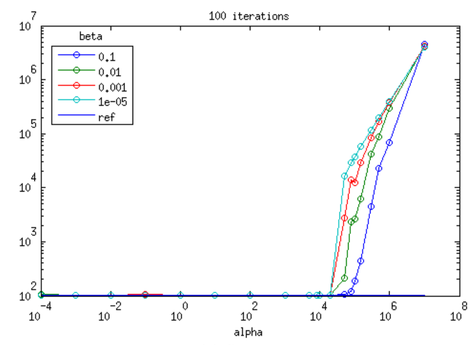
Figure 2: Reconstructions of the Lego block with different parameter configurations and 100 iterations

In Figure 2 we show a subset of the reconstructions we computed with different α and β combinations. For smaller β the reconstructions break down faster with increasing α, although this can be delayed by increasing the number of iterations.

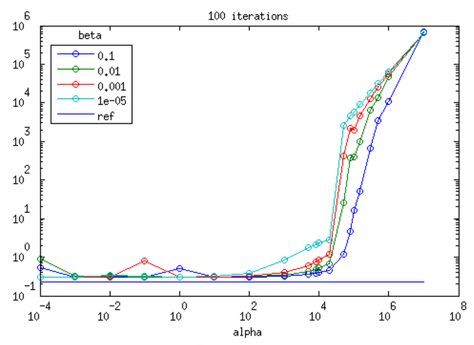
Using the formula $\sqrt{\frac{\|\mathbf{f}_{\alpha_1} - \mathbf{f}_{\alpha_2}\|^2}{\|\mathbf{f}_{\alpha_1}\|\|\mathbf{f}_{\alpha_2}\|}}$ we can calculate a percent difference between the reconstructions in Figure 2. Doing so for $\alpha_1 = 10^{-4}$ and $\alpha_2 = 10^4$ gives us Table (1).

β	% Difference
10^{-1}	25.4%
10^{-2}	29.3%
10^{-5}	69.8%

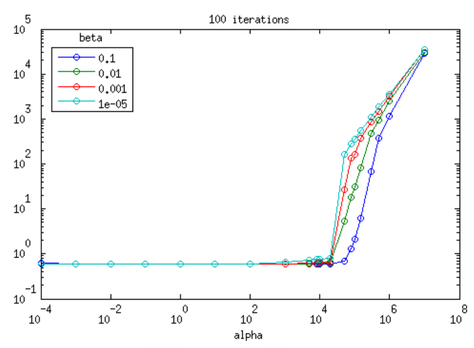
Table 1: Percent difference between $\alpha = 10^{-4}$ and $\alpha = 10^4$ reconstructions



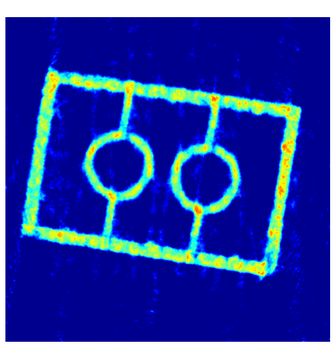
(a) S-curve



(b) Morozov



(c) Percentage difference



(d) Optimal Reconstruction

Figure 3: Different parameter evaluation methods and the optimal reconstruction

For our optimal reconstruction, which is shown in Figure 3(d), we used $\alpha = 10^2$ and $\beta = 10^{-3}$. We chose these parameters for producing the least percent difference in Figure 3(c).

Discussion

For small α values we got reconstructions with a lot of small changes or noise (figure 2 first column). The reason is that the regularization term in the objective function gets scaled down and therefore the method mainly tries to fit the reconstruction to the measurements regardless of the complexity of the image. The smaller β is the better the L_1 norm gets approximated. A small β value means that larger derivatives in the image get penalized more compared to small derivatives. So for smaller β, there are fewer changes in the image and more homogeneous regions for a given α value; as can be seen in Figure 2 in the first and third row of the third column. Therefore the reconstruction for small β gets flat even for smaller α values (seen in the third row of the fourth column). In summary for smaller β, a change of α has a higher impact. This can be also seen from table 1. Comparing the different evaluation methods for the parameter choice we can see that all of them show a similar behaviour (Figure 3). They all explode at an α value of about 10^4. Before that value all parameter choices seem to be quite reasonable. So we conclude that these 3 methods are interchangeable and none of them outplays the others, at least for our example. The methods fail due to the stability of the reconstructions with varying parameter choices. We also see that the behaviour for different values of β is the same but the curves are shifted along the α values. Therefore it seems that the effects of α and β are interchangeable to some extent, one can compensate the other.

References

[1] K. Niinimäki, M. Lassas, K. Hämäläinen, A. Kallonen, V. Kolehmainen, E. Niemi and S. Siltanen, *Multi-resolution parameter choice method for total variation regularized tomography*, 2014.

[2] K. Hämäläinen, A. Kallonen, V. Kolehmainen, M. Lassas, K. Niinimäki, and S. Siltanen, *Sparse tomography*, SIAM Journal on Scientific Computing, 35 (2013), pp. B644–B665.

[3] J. Mueller and S. Siltanen, *Linear and Nonlinear Inverse Problems with Practical Applications*, vol. 10 of Computational Science and Engineering, SIAM, 2012.