Homework III
Regularization and ill-posed problems

Moritz Wolter

November 15, 2015

1 Abstract

This report deals with methods to handle perturbations to the right hand side of $\mathbf{Ax} = \mathbf{b}$. Precisely an additional noise contribution is added to $\mathbf{b}$ such that $\mathbf{b} = \mathbf{b_{exact}} + \alpha \cdot \mathbf{b_{noise}}$. Where $\mathbf{b_{noise}}$ is a vector with entries drawn from the standard normal distribution $\mathcal{N}(0,1)$. The process of recovering $\mathbf{x_{reg}} \approx \mathbf{A}^{-1}\mathbf{b}_{exact}$ as precisely as possible is called regularization. Tikhonov regularization, TSVD and conjugate gradient methods will be applied and their regularization properties compared. The first two methods require a regularization parameter λ. It can be obtained using for example the L-curve criterion or generalized cross-validation (GCV). Both methods will be explored further.

2 Results

2.1 Tikhonov regularization

Tikhonov regularization can be implemented using the singular value decomposition of $\mathbf{A}$. The idea is to filter out the very small singular values, which are greatly influenced by the noise contribution using

$$\mathbf{x}_{reg} = \sum_{i=1}^{n} f_i \frac{\mathbf{u}_i^T b}{\sigma_i} \mathbf{v}_i. \tag{1}$$

The filter factors f_i are computed using

$$f_i = \sigma_i^2 / (\sigma_i^2 + \lambda). \tag{2}$$

Where filtering takes place if $\sigma_i < \lambda$. On the other hand a singular value σ remains unfiltered if $\sigma_i > \lambda$. For this method to work it is essential to choose a good value for λ.

2.1.1 The L-curve

The L-curve is a plot of the residual norm $\|\mathbf{Ax} - \mathbf{b}\|$ and the norm of the regularized solution $\|\mathbf{Lx}\|$. It displays the trade off between the fit to regularized data and the size of
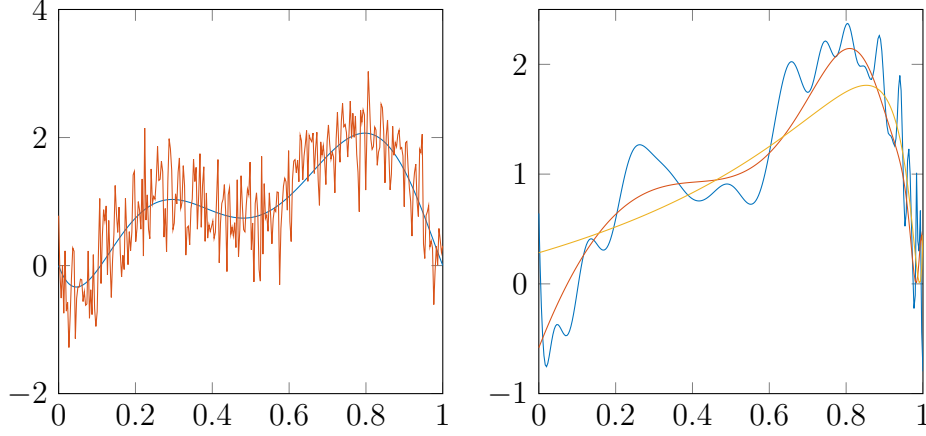
Figure 1: The l-curve in relation to curve fitting problems. On the left a known curve $\mathbf{b}_{true}$ is shown in blue. $\mathbf{b} = \mathbf{b}_{true} + \mathbf{b}_{noise}$ is shown in red. The right plot shows efforts to recover the noise free solution. The effects of no filtering (blue, $\|\mathbf{Ax}\| = 21.77$, $\|\mathbf{Ax} - \mathbf{b}_{true}\| = 7.59$), optimal filtering (red, $\|\mathbf{Ax}\| = 19.44$, $\|\mathbf{Ax} - \mathbf{b}_{true}\| = 3.33$) and over-damping (yellow, $\|\mathbf{Ax}\| = 18.3362$, $\|\mathbf{Ax} - \mathbf{b}_{true}\| = 5.08$) are shown.

the solution and gives insight into the properties of the underlying regularization method.[1] Using Tikhonov regularization as described in equation 1 and the filter of equation 2. The plots in figure 2 have been computed. We are interested in obtaining a result as close to the noise free example as possible. The trade off shown in the L-curve above is illustrated in another way in Figure 1.[2] When choosing a good filter parameter it is important to remove noise contributions, which is done by minimizing $\|\mathbf{Lx}\|$, while at the same time making sure that the data points are still being followed in a satisfactory way, which is done by keeping $\|\mathbf{Ax} - \mathbf{b}\|$ as small as possible. In practice $\mathbf{b}_{true}$ is unknown, it has been used in the experiment shown in figure 1 only to verify that the l-curve criterion which as been computed using the noisy $\mathbf{b}$ is indeed working. The optimal point can be found at the pointy edge of the L-curve, in most cases the curvature will display a maximal value here as well. Following Hansen once more the curvature κ has been computed using[3]

$$\kappa = 2 \cdot \frac{\hat{\rho}'\hat{\eta}'' - \hat{\rho}''\hat{\eta}'}{((\hat{\rho}')^2) + (\hat{\eta}')^2)^{3/2}}. \tag{3}$$

with $\hat{\eta} = log(\|x_\lambda\|_2^2)$ and $\hat{\rho} = log(\|Ax_\lambda - b\|_2^2)$. The derivatives have been implemented by simply using right forward differences

$$\hat{\rho}' = \frac{(\hat{\rho}_{k+1} - \hat{\rho}_k)}{(\lambda_{k+1} - \lambda_k)}. \tag{4}$$

For the second derivatives the same process has simply been applied again. $\hat{\eta}'$ and $\hat{\eta}''$ have been computed in the same manner. The plot of the curvature which has been obtained by using the formulas above is shown in figure 2 in the bottom left. In all plots

---

[1] The L-curve and its use in the numerical treatment of inverse problems, P. C. Hansen, Department of Mathematical Modelling, Technical University of Denmark, DK-2800 Lyngby, Denmark

[2] following a similar example in: The L-curve and its use in the numerical treatment of inverse problems, P. C. Hansen

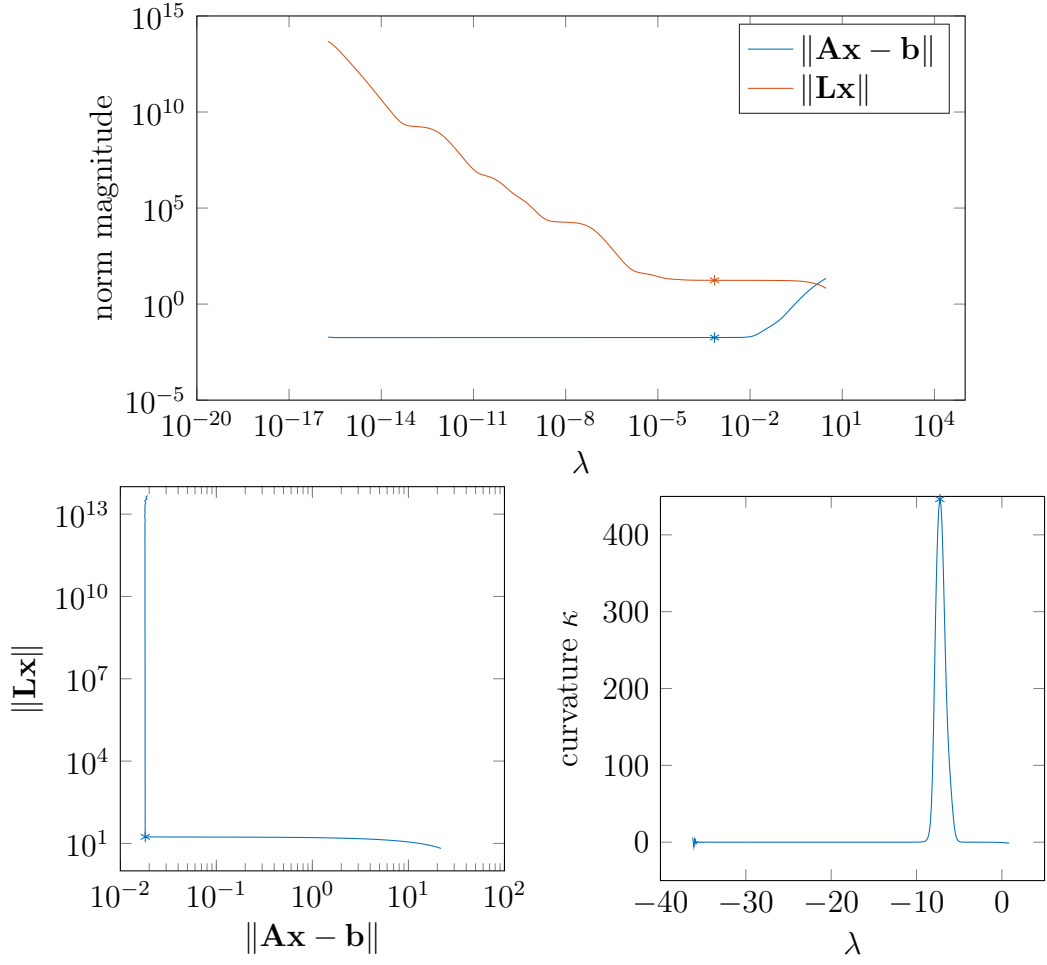[3] The L-curve and its use in the numerical treatment of inverse problems, P. C. Hansen

Figure 2: Plots of $\|\mathbf{Ax} - \mathbf{b}\|$, $\|\mathbf{Lx}\|$ and the curvature of the L-curve κ. For the given `A1`, `berr1` pair.
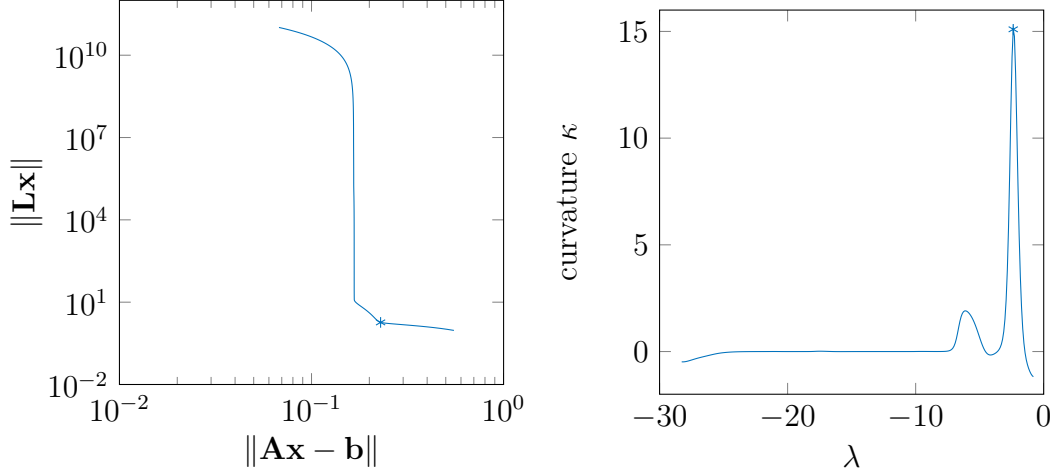
Figure 3: Tikhonov regularization L-curve and curvature for the matrix $\mathbf{A}_2$ and vector $\mathbf{b}_{err2}$.
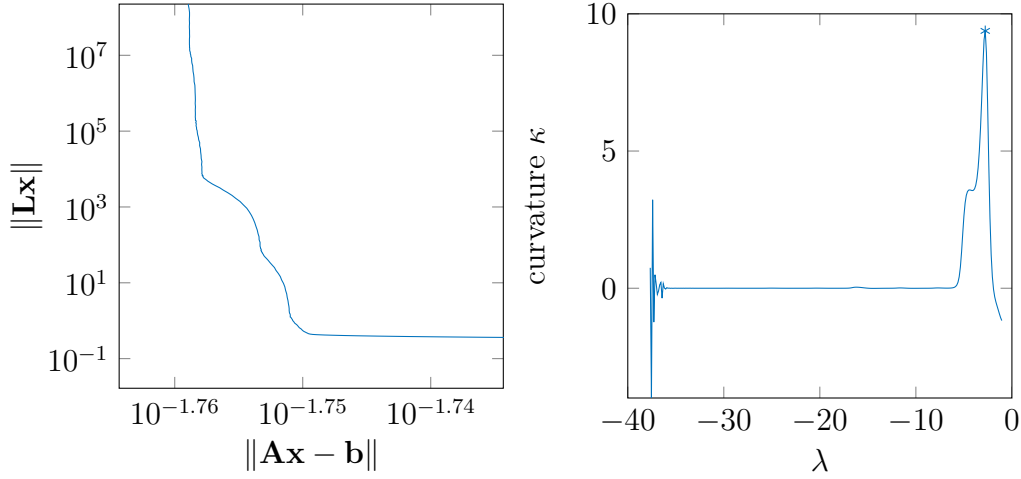

Figure 4: Tikhonov regularization, L-curve and curvature for the matrix $\mathbf{A}_3$ and vector $\mathbf{b}_{err3}$.

in the figure the location of the maximum of κ has been indicated with stars(*). It is important to notice that λ increases when following the L-curve from the top left to the bottom right. Figure 3 shows a more dubious case. For the given matrix $\mathbf{A}_2$ the L-curve displays two edges. The selection criterion now chooses the edge at the bottom as it curvier then the previous one. However maybe the emphasis should have been to reduce the residual instead of the solution norm. In that case the previous edge should have been selected. This example shows that selection by curvature is not free of flaws, it tends to oversmooth the solution in this case. Figures 5, 6 and 7 show more examples where the L-curve criterion performs very well. However there is an absentee. Matrix $\mathbf{A}_3$ is missing. Here the L-curve's edged is not found exactly by the curvature based selection criterion, this example will be solved using GCV, which as explained in the next section.
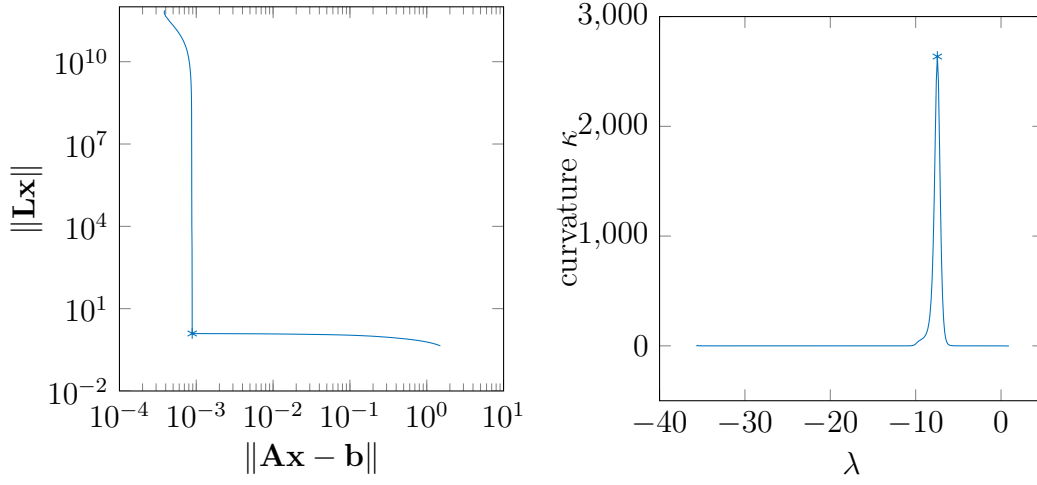
Figure 5: Tikhonov regularization, L-curve and curvature for the matrix $\mathbf{A}_4$ and vector $\mathbf{b}_{err4}$.
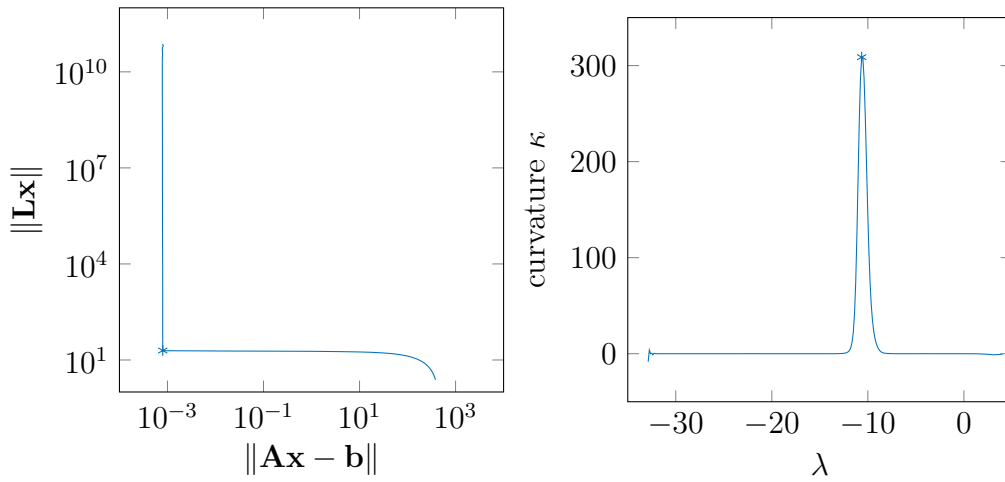

Figure 6: Tikhonov regularization, L-curve and curvature for the matrix $\mathbf{A}_5$ and vector $\mathbf{b}_{err5}$.
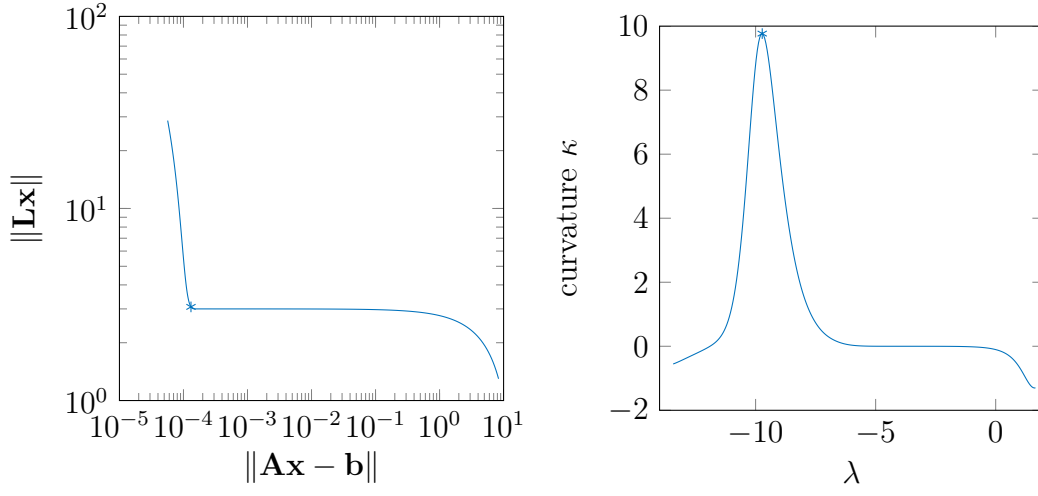
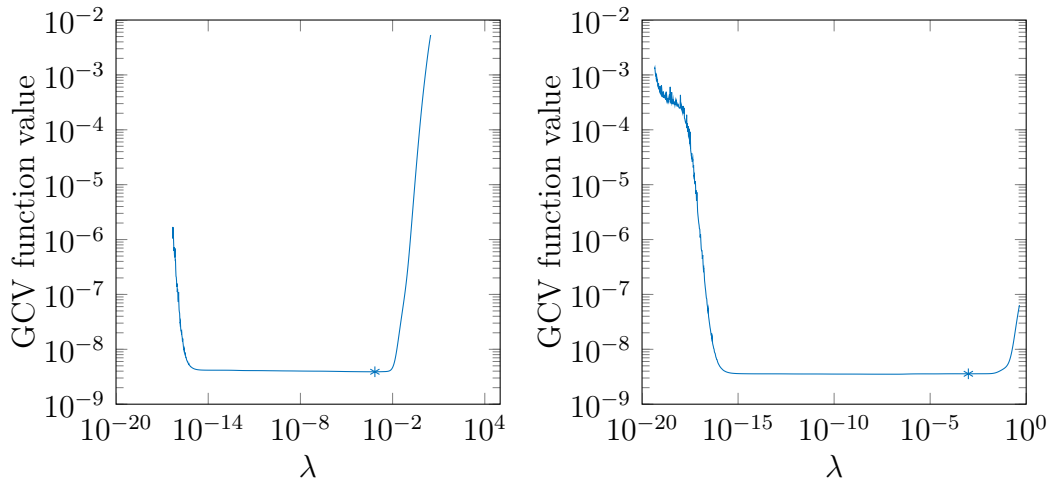Figure 7: Tikhonov regularization, L-curve and curvature for the matrix $\mathbf{A}_6$ and vector $\mathbf{b}_{err6}$.

Figure 8: Plot of the value of the GCV-function for different values of λ for the $\mathbf{A}_1, \mathbf{b}_{err1}$ and the $\mathbf{A}_3, \mathbf{b}_{err3}$ pair.
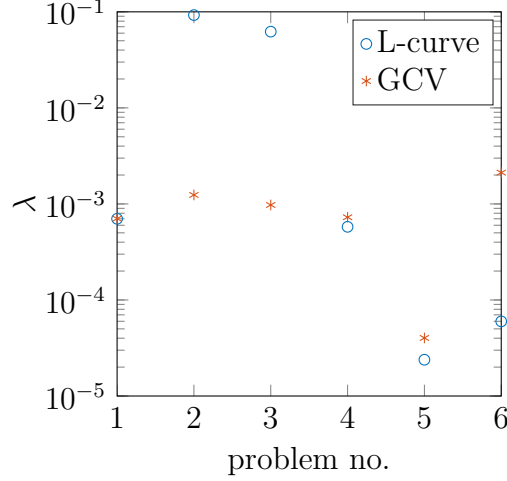
Figure 9: Plot of the lambdas selected by the L-curve curvature criterion and generalized cross validation.

2.2 Generalized Cross validation (GCV)

An other way to find a suitable regularization parameter is generalized cross validation. When this method is used the function

$$G = \frac{\|\mathbf{A}\mathbf{x}_{reg} - b\|_2^2}{(\text{trace}(\mathbf{I}_m - mathbf{A}\mathbf{A}^I))^2} \tag{5}$$

is minimized. With $\mathbf{x}_{reg}$ defined as in equation 1. The trace of the denominator can be simplified to:

$$\text{trace}(\mathbf{I}_m) - \text{trace}(\mathbf{A}\mathbf{A}^I) \quad \text{assuming symmetrie} \tag{6}$$
$$= n - \text{trace}(\mathbf{A}^T\mathbf{A}^I) \tag{7}$$
$$= n - \text{trace}(\mathbf{V}\mathbf{\Sigma}^T\mathbf{U}^T\mathbf{U}\mathbf{F}\mathbf{\Sigma}^{-1}\mathbf{V}^T) \quad \text{svd} \tag{8}$$
$$= n - \text{trace}(\mathbf{V}^T\mathbf{V}\mathbf{\Sigma}^T\mathbf{F}\mathbf{\Sigma}^{-1}) \quad \text{cyclic rotation} \tag{9}$$
$$= n - \text{trace}(\mathbf{\Sigma}^{-1}\mathbf{\Sigma}\mathbf{F}) \quad \text{cyclic rotation} \tag{10}$$
$$= n - \text{trace}(\mathbf{F}) \quad \text{F is diagonal} \tag{11}$$
$$= n - \text{sum}(\mathbf{F}) \tag{12}$$

With $\mathbf{F}$ being a matrix with the filter factor on the diagonal. The reasoning described above is used to compute the GCV function efficiently. A comparison of the λ values selected by the L-curve curvature criterion and GCV is given in figure 9. For the first problem both methods select almost the same value. In two an three the value selected by the L-curve curvature criterion is significantly larger then what is chosen by the GCV criterion and what a human would have selected. In these six examples this generally happens when the value of the maximum of the κ curve is not very large. In these cases GCV should be preferred.

3 Conclusion