amcs

REGULARIZATION PARAMETER SELECTION IN DISCRETE ILL–POSED PROBLEMS — THE USE OF THE U–CURVE

Dorota KRAWCZYK-STAŃDO [*], Marek RUDNICKI [**]

[*] Center of Mathematics and Physics, Technical University of Łódź
ul. Al. Politechniki 11, 90–924 Łódź, Poland
e-mail: `krawczyk@p.lodz.pl`

[**] Institute of Computer Science
Technical University of Łódź
ul. Wólczańska 215, 90–924 Łódź, Poland
e-mail: `rudnicki@ics.p.lodz.pl`

To obtain smooth solutions to ill-posed problems, the standard Tikhonov regularization method is most often used. For the practical choice of the regularization parameter α we can then employ the well-known L-curve criterion, based on the L-curve which is a plot of the norm of the regularized solution versus the norm of the corresponding residual for all valid regularization parameters. This paper proposes a new criterion for choosing the regularization parameter α, based on the so-called U-curve. A comparison of the two methods made on numerical examples is additionally included.

Keywords: ill-posed problems, Tikhonov regularization, regularization parameter, L-curve, U-curve

1. Introduction

Ill-posed problems are frequently encountered in science and engineering. The term itself has its origins in the early 20-th century. It was introduced by Hadamard who investigated problems in mathematical physics. According to his beliefs, ill-posed problems did not model real-world problems, but later it appeared how wrong he was. Hadamard defined a linear problem to be *well posed* if it satisfies the following three requirements: (a) existence, (b) uniqueness, and (c) stability. A problem is said to be *ill-posed* if one or more of these requirements are not satisfied. A classical example of an ill-posed problem is a linear integral equation of the first kind in $L^2(I)$ with a smooth kernel. A solution to this equation, if it exists, does not continuously depend on the right-hand side and may not be unique. When a discretization of the problem is performed, we obtain a matrix equation in C^m,

$$Ku = f, \tag{1}$$

where K is an $m \times n$ matrix with a large condition number, $m \geq n$. A linear least-squares solution of the system (1) is a solution to the problem

$$\min_{u \in C^n} \|Ku - f\|^2, \tag{2}$$

where the Euclidean vector norm in C^m is used. We say that the algebraic problems (1) and (2) are discrete ill-posed problems.

The numerical methods for solving discrete ill-posed problems in function spaces and for solving discrete ill-posed problems have been presented in many papers. These methods are based on the so-called regularization methods. The main objective of regularization is to incorporate more information about the desired solution in order to stabilize the problem and find a useful and stable solution. The most common and well-known form of regularization is that of Tikhonov (Groetsch, 1984). It consists in replacing the least-squares problem (2) by that with a suitably chosen Tikhonov functional. The most basic version of this method can be presented as

$$\min_{u \in \mathbb{C}^n} \left\{ \|Ku - f\|^2 + \alpha^2 \|u\|^2 \right\}, \tag{3}$$

where $\alpha \in \mathbb{R}$ is called the regularization parameter. The Tikhonov regularization is a method in which the regularized solution is sought as a minimizer of a weighted combination of the residual norm and a side constraint. The regularization parameter controls the weight given to the minimization of the side constraint. Thus, the qual-

ity of the regularized solution is controlled by the regularization parameter. An optimal regularization parameter should fairly balance the perturbation error and the regularization error in the regularized solution.

A suitable choice of the regularization parameter is still a current and vital problem. There are several possible strategies that depend on additional information referring to the analysed problem and its solution, e.g., the discrepancy principle and the generalized cross-validation method. The discrepancy principle is an *a-posteriori* strategy for choosing α as a function of an error level (the input error level must be known). The generalized cross-validation method is based on *a-priori* knowledge of a structure of the input error, which means that the errors in f can be considered to be uncorrelated zero-mean random variables with a common variance, i.e., white noise.

Another practical method for choosing α when data are noisy is the L-curve criterion (Hansen, 1992; Hansen and O'Leary, 1993). The method is based on the plot of the norm of the regularized solution versus the norm of the corresponding residual. The practical use of such a plot was first introduced by Lawson and Hanson (1974). The idea of the L-curve criterion is to choose a regularization parameter related to the characteristic L-shaped "corner" of the graph.

In Section 2, we recall the L-curve criterion for choosing the regularization parameter α and some properties of the L-curve. In Section 3, we formulate a new criterion for choosing the regularization parameter α, based on the U-curve, and we give some of its properties. Section 4 presents numerical results obtained using the new U-curve criterion and compares them with those resulting from the application of the L-curve criterion.

The present work constitutes an extension of our previous investigations on this research topic (Krawczyk-Stańdo and Rudnicki, 2005; Neittaanmaki *et al.*, 1996; Stańdo *et al.*, 2003).

2. L-Curve Criterion for Choosing the Regularization Parameter

If $K \in \mathbb{C}^{m,n}$ is a matrix of rank r, then there exist unitary matrices $U \in \mathbb{C}^{m,m}$ and $V \in \mathbb{C}^{n,n}$ such that

$$K = U\Sigma V^*, \quad \Sigma = \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix},$$

where $\Sigma \in R^{m,n}$, $\Sigma_r = \mathrm{diag}\,(\sigma_1, \ldots, \sigma_r)$, and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$. The σ_is are called the singular values of K and the i-th column vectors u_i, v_i of U and V, respectively, are the left and right singular vectors corresponding to σ_i, $i = 1, \ldots, r$. The singular value decomposition (SVD) for the matrix K is a well-known approach to least-squares problems (Wahba, 1977).

The least-squares minimum-norm solution to (1) is the solution of the normal equation $K^* K u = K^* f$, and thus if

$$f = \sum_{i=1}^{m} f_i u_i, \tag{4}$$

where $f_i = u_i^* f$, $i = 1, \ldots, m$, then

$$u = \sum_{i=1}^{r} \frac{f_i}{\sigma_i} v_i. \tag{5}$$

For a discrete ill-posed problem the singular values σ_i tend rapidly to zero. Due to the errors on the right-hand side (we may not assume that $u_i^* e$, $i = 1, \ldots, r$, tend to zero faster than σ_i), the solution u is perturbed by the contributions corresponding to small singular values.

The regularized solution to (3)

$$u_\alpha = \sum_{i=1}^{r} \frac{\sigma_i f_i}{\sigma_i^2 + \alpha^2} v_i \tag{6}$$

satisfies the normal equation $K^* K u + \alpha^2 u = K^* f$. Since $\alpha > 0$, the problem of computing u_α becomes less ill-conditioned than that of computing u (the influence of the errors corresponding to small singular values becomes smaller).

It is easily found (Hansen and O'Leary, 1993; Regińska, 1996) that

$$x(\alpha) = \|Ku_\alpha - f\|^2 = \sum_{i=1}^{r} \frac{\alpha^4 f_i^2}{\left(\sigma_i^2 + \alpha^2\right)^2} + \|f_\perp\|^2, \tag{7}$$

where $f_\perp = \sum\limits_{i=r+1}^{m} f_i u_i$, and

$$y(\alpha) = \|u_\alpha\|^2 = \sum_{i=1}^{r} \frac{\sigma_i^2 f_i^2}{\left(\sigma_i^2 + \alpha^2\right)^2}. \tag{8}$$

In (Hansen and O'Leary, 1993; Regińska, 1996), it is shown that y as a function of x is decreasing and strictly convex.

A good method of choosing the regularization parameter for discrete ill-posed problems must incorporate information about the solution size in addition to using information about the residual size. This is indeed quite natural, because we are seeking a fair balance in keeping both of these values small.

A more recent method of choosing the regularization parameter makes use of the so-called L-curve, see (Hansen, 1992; Hansen and O'Leary, 1993). For Tikhonov regularization the L-curve is a parametric plot of $(x(\alpha), y(\alpha))$, where $x(\alpha)$ and $y(\alpha)$ measure the size of the regularized solution and the corresponding residual, respectively, for all $\alpha > 0$. The work (Hansen and O'Leary, 1993) contains many properties of the L-curve for Tikhonov regularization. In particular, whenever

(a) the discrete Picard condition is satisfied,

(b) the errors on the right-hand side are essentially "white noise",

(c) the signal-to-noise ratio is reasonably large,

the L-curve $(x(\alpha), y(\alpha))$ for Tikhonov regularization has two characteristic parts, namely, a "flat" part and an almost "vertical" part.

The L-curve is basically made up of two parts. The more horizontal part corresponds to the solutions where the regularization parameter is too large and the solution is dominated by the regularization errors. The vertical part corresponds to the solutions where the regularization parameter is too small and the solution is dominated by the right-hand errors, magnified by the division by small singular values. This behavior does not rely on any additional properties of the problem, e.g., a statistical distribution of the errors, the discrete Picard condition, etc. It should be taken into account that the vertical and horizontal parts correspond to the solutions that are under- and over-smoothed, respectively.

It is difficult to inspect the features of the L-curve when it is plotted in linear scale due to the large range of values for the two norms. As shown in (Hansen and O'Leary, 1993), the features become more visible (and easier to inspect) when the curve is plotted in the double-logarithmic scale. The log-log scale actually emphasizes the corner of the L-curve. One more advantage of the log-log scale is that particular scalings of the right-hand side and the solution simply shift the L-curve horizontally and vertically. So, in many cases it is better to analyze the L-curve $(x(\alpha), y(\alpha))$ in the log-log scale. There is a strong intuitive justification for this. Since the singular values typically span several orders of magnitude, we carry out all our computations related to curvature on $(\log x(\alpha), \log(y))$. The log-log transformation has a theoretical justification, see (Hansen and O'Leary, 1993). Some properties of the L-curve in other scales are shown in (Regińska, 1996).

The L-curve is of an interest because it shows how the regularized solution changes as the regularization parameter α changes. A distinct l-shaped corner of the L-curve is located exactly where the solution x_α changes, from being dominated by the regularization errors to being dominated by the errors on right-hand side. That is why the corner of the L-curve corresponds to a good balance between the minimization of the sizes, and the corresponding regularization parameter α is a good one.

Two meanings of the "corner" were suggested by Hansen and O'Leary (1993). First – it is the point where the curve is closest to the origin, second – it is the point where the curvature is maximum.

The L-curve for Tikhonov regularization is important in the analysis of discrete ill-posed problems. Fig. 1 shows a typical L-curve.
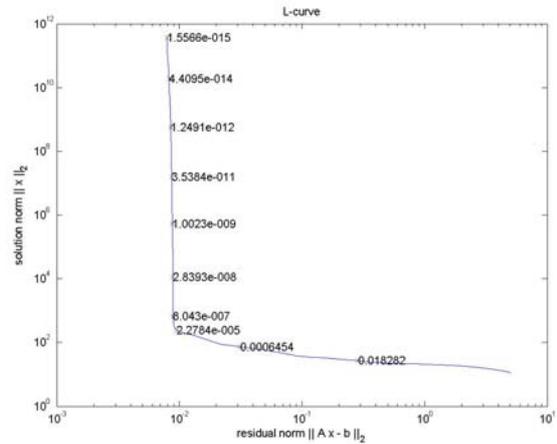

Fig. 1. A typical L-curve for Tikhonov regularization.

3. U-Curve Criterion for Choosing the Regularization Parameter

Consider the discrete ill-posed problem (2), its solution u (5) and the regularized solution (6) obtained by means of Tikhonov regularization (3).

The right-hand side f in (1) is assumed to be contaminated with measurement errors. The perturbation vector $f - \overline{f}$ does not need to meet the discrete Picard condition when the unperturbed right-hand side $\overline{f}$ satisfies it. That is why there is a large influence of the errors $f_i - \overline{f_i}$ corresponding to small singular values on the solution norm. Any $\alpha > 0$ reduces the norm of u. The problem is how to decide on an appropriate regularization parameter α for which $y(\alpha)$ is not overly large and it is a small norm of the residual.

3.1. U-Curve and Its Properties. Define

$$U(\alpha) = \frac{1}{x(\alpha)} + \frac{1}{y(\alpha)}, \qquad (9)$$

where $\alpha > 0$ and $x(\alpha)$ and $y(\alpha)$ are defined by (7) and (8), respectively.

Definition 1. We define the U-curve to be the plot of $U(\alpha)$, i.e., the plot of the sum of the reciprocals of the regularized solution norm and the corresponding residual norm, for $\alpha > 0$.

The U-curve consists of three characteristic parts, namely

- on the left and right sides, it is almost "vertical," and

- in the middle it is almost "horizontal."

The vertical parts correspond to the regularization parameter for which the solution norm and the residual norm

are dominated by each other. The more horizontal part corresponds to the regularization parameter for which the solution norm and the residual norm are close to each other.

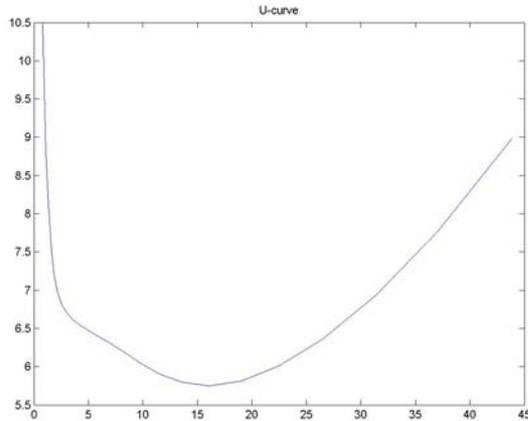Figure 2 shows an example of a typical U-curve.

Fig. 2. A typical U-curve for Tikhonov regularization.

Theorem 1. *The function $U(\alpha)$ is strictly decreasing on the interval $\alpha \in \left(0, \sigma_r^{\frac{2}{3}}\right)$ and strictly increasing on the interval $\alpha \in \left(\sigma_1^{\frac{2}{3}}, \infty\right)$, where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ are the singular values.*

Proof. For simplicity, assume that $x = x(\alpha)$ and $y = y(\alpha)$, so that

$$U(\alpha) = \frac{1}{x} + \frac{1}{y}$$

and its first derivative is

$$U'(\alpha) = \frac{-x'}{x^2} + \frac{-y'}{y^2}.$$

From (Hansen and O'Leary, 1993; Regińska, 1996), we know that $x' = -\alpha^2 y'$, and thus

$$U'(\alpha) = \frac{-y'(x + \alpha y)(x - \alpha y)}{(xy)^2}.$$

To analyze the sign of $U'(\alpha)$, we consider the factor $x - \alpha y$ because

$$\frac{-y'(x + \alpha y)}{(xy)^2} > 0.$$

Recall that

$$x = x(\alpha) = \sum_{i=1}^r \frac{\alpha^4 f_i^2}{\left(\sigma_i^2 + \alpha^2\right)^2}$$

and

$$y = y(\alpha) = \sum_{i=1}^r \frac{\sigma_i^2 f_i^2}{\left(\sigma_i^2 + \alpha^2\right)^2}.$$

Thus

$$x - \alpha y = \sum_{i=1}^r \frac{\alpha^4 f_i^2}{\left(\sigma_i^2 + \alpha^2\right)^2} - \alpha \sum_{i=1}^r \frac{\sigma_i^2 f_i^2}{\left(\sigma_i^2 + \alpha^2\right)^2}$$
$$= \sum_{i=1}^r \frac{\alpha f_i^2 \left(\alpha^3 - \sigma_i^2\right)}{\left(\sigma_i^2 + \alpha^2\right)^2}.$$

Since

$$\sum_{i=1}^r \frac{\alpha f_i^2}{\left(\sigma_i^2 + \alpha^2\right)^2} > 0,$$

we consider only the factor $\alpha^3 - \sigma_i^2$. Consequently,

$$\alpha^3 - \sigma_i^2 > 0 \Leftrightarrow \alpha > \sigma_i^{\frac{2}{3}}$$

and

$$\alpha^3 - \sigma_i^2 < 0 \Leftrightarrow \alpha < \sigma_i^{\frac{2}{3}}.$$

To generalize, we can deduce that

$$\alpha \in \left(0, \sigma_r^{\frac{2}{3}}\right) \Rightarrow U'(\alpha) < 0$$

and

$$\alpha \in \left(\sigma_1^{\frac{2}{3}}, \infty\right) \Rightarrow U'(\alpha) > 0.$$

It follows that the function $U(\alpha)$ is strictly decreasing on the interval $\alpha \in \left(0, \sigma_r^{\frac{2}{3}}\right)$ and strictly increasing on the interval $\alpha \in \left(\sigma_1^{\frac{2}{3}}, \infty\right)$. ■

Lemma 1. *For the function $U(\alpha)$ we have the following:*

(i) $\lim_{\alpha \to 0^+} U(\alpha) = +\infty$,

(ii) $\lim_{\alpha \to +\infty} U(\alpha) = +\infty$.

Proof. Consider the function

$$U(\alpha) = \frac{1}{x} + \frac{1}{y}.$$

From (7) and (8), we obtain

$$U(\alpha) = \sum_{i=1}^r \frac{\left(\sigma_i^2 + \alpha^2\right)^2 \left(\sigma_i^2 + \alpha^4\right)}{f_i^2 \alpha^4 \sigma_i^2}$$
$$= \sum_{i=1}^r \frac{\left(\sigma_i^2/\alpha^2 + 1\right)^2 + \sigma_i^2 + \alpha^2 + \alpha^4/\sigma_i^2}{f_i^2}.$$

It is immediate that

$$\lim_{\alpha \to 0^+} U(\alpha) = +\infty$$

and

$$\lim_{\alpha \to +\infty} U(\alpha) = +\infty.$$

■

Remark 1. The function $U(\alpha)$ certainly has a local minimum in the interval $\alpha \in \left(\sigma_r^{\frac{2}{3}}, \sigma_1^{\frac{2}{3}} \right)$.

Remark 2. If in the SVD there is only one non-zero value (it may be multiple, too) we can analytically calculate a unique α for which the U-function will reach the minimum (the U-function then has only one minimum).

The objective of the U-curve criterion for selecting the regularization parameter is to choose a parameter for which the curvature attains a local maximum close to the left vertical part of the U-curve.

The regularization parameter appropriate for the U-curve criterion is calculated numerically by applying routines available in the Matlab Regularization Tools package.

4. Numerical Examples

A classical example of an ill-posed problem is a Fredholm integral equation of the first kind (Groetsch, 1984) with a square integrable kernel,

$$\int_a^b K(s,t) u(t) \, \mathrm{d}t = f(s), \quad c \le s \le d,$$

where the right-hand side f and the kernel K are given, and u is unknown.

Example 1. Consider the test problem 'shaw' (Hansen, 1993), in which the integral equation is a one-dimensional model of an image reconstruction problem with $[-\pi/2, \pi/2]$ as both integration intervals. The kernel K and the solution u are respectively given by

$$K(s,t) = \Big(\cos(s) + \cos(t) \Big) \left(\frac{\sin(l)}{l} \right)^2,$$

where

$$l = \pi \Big(\sin(s) + \sin(t) \Big),$$

and

$$u(t) = 2 \exp\Big(-6 (t-0.8)^2\Big) + \exp\Big(-2 (t+0.5)^2\Big).$$

The numerical results are shown in Figs. 3–7. We get

$$\|u_{\text{exact}} - u_{\text{L-curve}}\| = 0.012633,$$
$$\|u_{\text{exact}} - u_{\text{U-curve}}\| = 0.010521.$$

Example 2. Consider now the test problem 'heat' from (Hansen, 1993). Our numerical results are shown in Figs. 8–12.

We get

$$\|u_{\text{exact}} - u_{\text{L-curve}}\| = 0.007084,$$
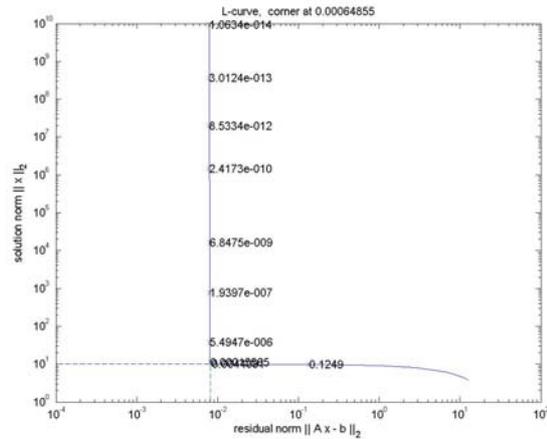$$\|u_{\text{exact}} - u_{\text{U-curve}}\| = 0.006962.$$

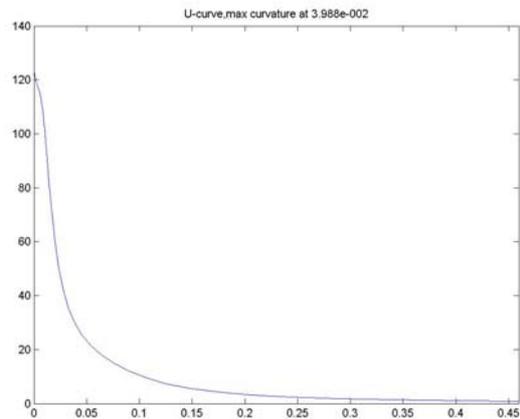Fig. 3. The L-curve and the selected regularization parameter.

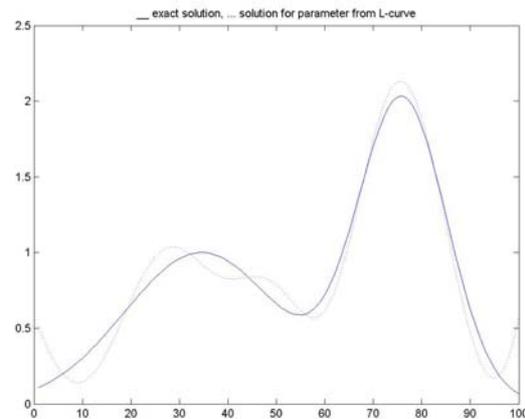Fig. 4. The U-curve and the selected regularization parameter.

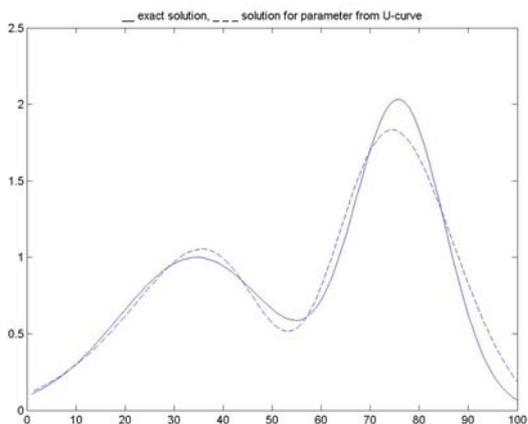Fig. 5. The exact solution and the solution for the parameter from the L-curve.

Fig. 6. The exact solution and the solution for
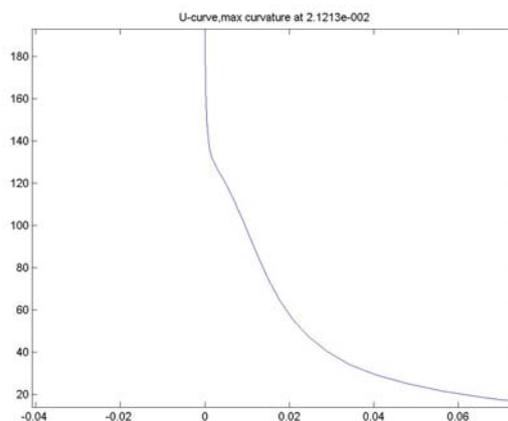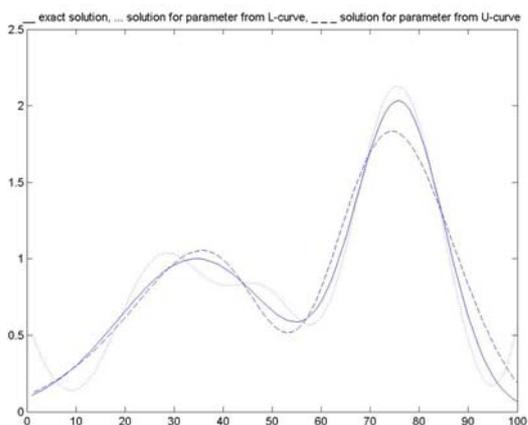the parameter from the U-curve.

Fig. 7. The exact solution the solution for the parameter
from the U-curve and the solution for the parame-
ter from the L-curve.

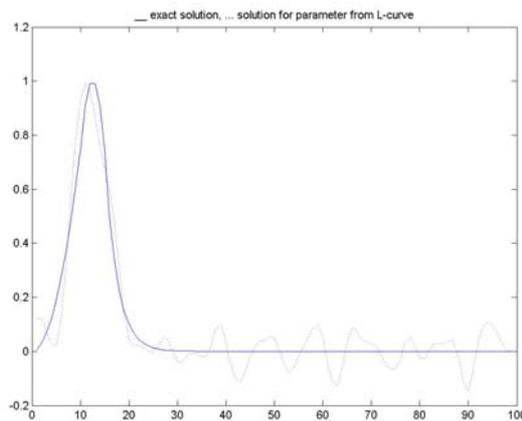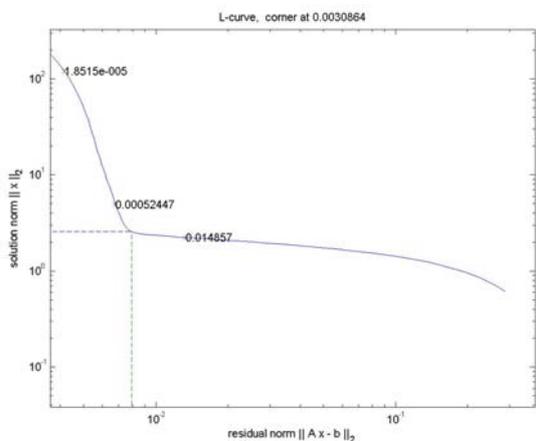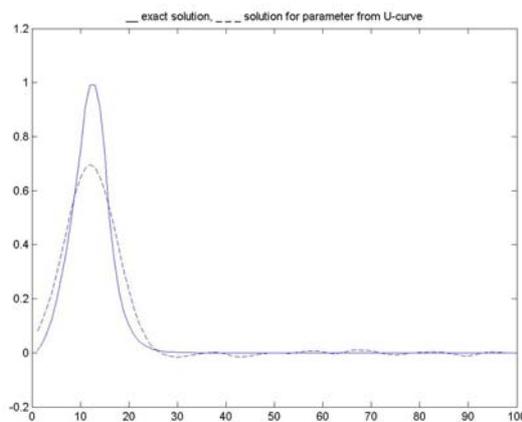Fig. 8. The L-curve and the selected regularization parameter

Fig. 9. The U-curve and the selected regularization parameter.

Fig. 10. The exact solution and the solution for
the parameter from the L-curve.

Fig. 11. The exact solution and the solution for
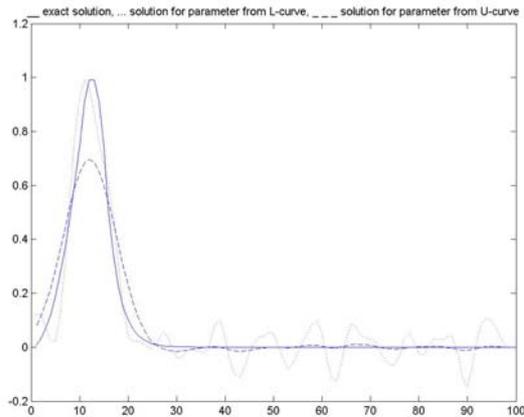the parameter from the U-curve.

Fig. 12. The exact solution the solution for the parameter from the U-curve and the solution for the parameter from the L-curve.

Example 3. Consider now an example of (Neittaanmaki *et al.*, 1996). The Fredholm integral equation of the first kind is

$$\int_a^b K\left(s,t\right)u\left(t\right)\mathrm{d}t = f\left(s\right), \quad c \leq s \leq d,$$

where the right-hand side $f\left(s\right) = 1$ and the kernel is

$$K(s,t) = \frac{1}{\pi\left[1 + \left(s-t\right)^2\right]^{\frac{3}{2}}}.$$

Here $0.2 \leq s \leq 0.8$ is given, and u is unknown. The numerical results are shown in Figs. 13–15.

Some of the numerical results were analyzed by us and presented in (Krawczyk-Stańdo and Rudnicki, 2005; Neittaanmaki *et al.*, 1996).

5. Conclusions

As can be seen, the results we obtained from both the L-curve and U-curve methods are comparable. In the first two examples the norm of the error for the U-curve criterion is smaller than that for the L-curve criterion. We cannot calculate the norm of the error in the third example because the exact solution is not known. However, the result we obtained is close to the results which we published in (Krawczyk-Stańdo and Rudnicki, 2005; Neittaanmaki *et al.*, 1996).

Obviously, we realize that an example might be produced for which the U-curve criterion will fail, but our feeling is that it works well in practice and that it is a useful method. We do hope that further work on this method can yield interesting results.
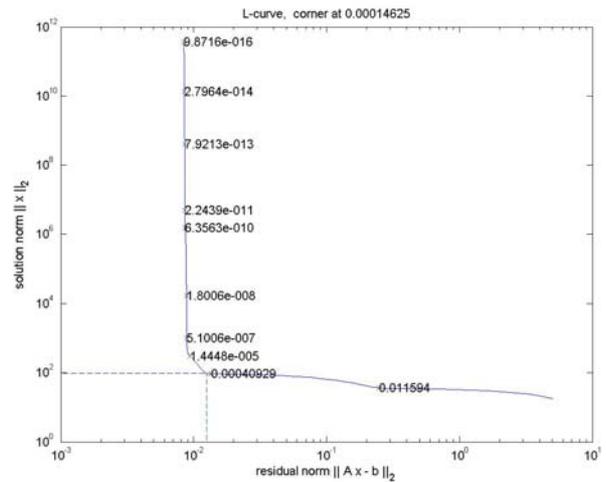

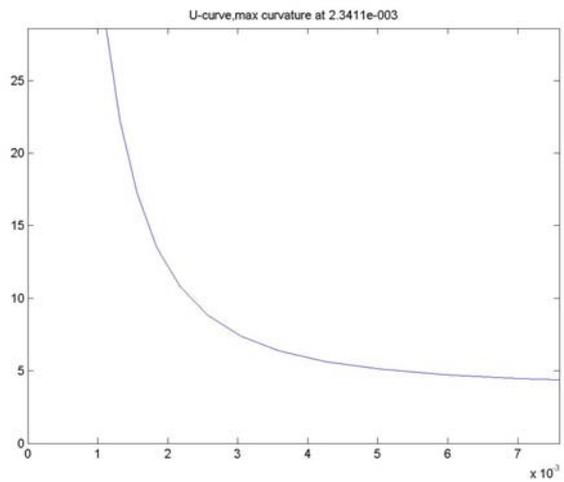Fig. 13. The L-curve and the selected regularization parameter.

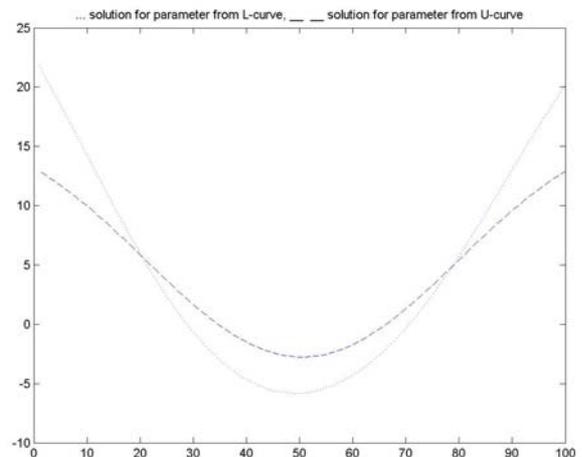Fig. 14. The u-curve and the selected regularization parameter.

Fig. 15. The solution for the parameter from the L-curve, the solution for the parameter from the U-curve.

References

Groetsch N. (1984): *The Theory of Tikhonov Regularization for Fredholm Integral Equations of the First Kind.* — London: Pitman.

Hansen P.C. (1992): *Analysis of discrete ill-posed problems by means of the L-curve.*— SIAM Rev., Vol. 34, No. 4, pp. 561–580.

Hansen P.C. and O'Leary D.P. (1993): *The use of the L-curve in the regularization of discrete ill- posed problems.* — SIAM J. Sci. Comput., Vol. 14, No. 6, pp. 487–1503.

Hansen P.C. (1993): *Regularization Tools, a Matlab package for analysis and solution of discrete ill-posed problems.* — Report UNIC–92–03

Krawczyk-Stańdo D. and Rudnicki M. (2005): *Regularized synthesis of the magnetic field using the L-curve approach.* — Int. J. Appl. Electromagnet. Mech., Vol. 22, No. 3–4, pp. 233–242.

Lawson C.L. and Hanson R.J. (1974): *Solving Least Squares Problems.* — Englewood Cliffs, NJ: Prentice-Hall.

Neittaanmaki P., Rudnicki M. and Savini A. (1996): *Inverse Problems and Optimal Design in Electrity and Magnetism.* — Oxford: Clarendon Press.

Regińska T. (1996): *A regularization parameter in discrete ill-posed problems.* — SIAM J. Sci. Comput., Vol. 17, No. 3, pp. 740–749.

Stańdo J., Korotow S., Rudnicki M., Krawczyk-Stańdo D. (2003): *The use of quasi-red and quasi-yellow nonobtuse refinements in the solution of 2-D electromagnetic, PDE's,* In: Optimization and inverse problems in electromagnetism (M. Rudnicki and S. Wiak, Ed.). — Dordrecht, Kluwer, pp. 113-124.

Wahba G. (1977): *Practical approximate solutions to linear operator equations when data are noisy.*— SIAM J. Numer. Anal., Vol. 14, No. 4, pp. 651–667.