JOURNAL OF APPLIED
COMPUTER SCIENCE
Vol. 19 No. 1 (2011), pp. 51-63

Modern Regularization Techniques for Inverse Modelling: a Comparative Study

Dorota Krawczyk-Stańdo[1], Marek Rudnicki[2], Jacek Stańdo[1]

[1]*Technical University of Łódź*
Centre of Mathematics and Physics
90-924 Łódź, al. Politechniki 11, Poland
krawczyk@p.lodz.pl

[2]*Technical University of Łódź*
Institute of Information Technology
90-924 Łódź, Wólczańska 215, Poland
rudnicki@ics.p.lodz.pl

Abstract. *Regularization techniques are used for computing stable solutions to ill-posed problems. The well-known form of regularization is that of Tikhonov in which the regularized solution is searched as a minimiser of the weighted combination of the residual norm and a side constraint-controlled by the regularization parameter. For the practical choice of regularization parameter we can use the L-curve approach, U-curve criterion introduced by us [1] and empirical risk method [2]. We present a comparative study of different strategies for the regularization parameter choice on examples of function approximation by radial basis neural networks. Such networks are universal approximators and can learn any nonlinear mapping. e.g. representing an magnetic inverse problem. Some integral equations of the first kind are considered as well.*

Keywords: *regularization parameter, L-curve, U-curve, discrepancy principle, empirical risk, inverse problem.*

1. Introduction

The regularisation method serves as a tool for the solution to ill-posed problems [3]. According to Hadamard, the originator of the concept of ill-posed problems, a problem can be defined as ill-posed if the solution is non-unique or if it is a discontinuous function of the data. The 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 of this equation, if it exists, does not continuously depend on the right-hand side and may not be unique. When the discretization of the problem is carried out we receive a matrix equation in C^m

$$Ax = b \tag{1}$$

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

$$\min \min_{x \in C^m} ||Ax - b||^2 \tag{2}$$

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

However, even ill-conditioned problems may have a meaningful solution, which can be found by regularization. Making use of the Tikhonov regularization, the regularized solution to a linear problem can be found as a minimiser of the following functional:

$$x_\alpha = \arg\min\{||Ax - b||_2^2 + \alpha^2\, ||Lx||_2^2\} \tag{3}$$

where α is the regularization parameter and L approximates a derivative operator.

Regularised solution is searched as a minimiser of the weighted combination of the residual norm and a side constraint. It can be stated that all regularisation methods for computing stable solutions to inverse problems involve a trade-off between the “size” of the regularised solution and the quality of the fit that it provides to the given data. What distinguishes the various regularisation methods is how they measure these quantities and how they decide on the optimal trade-off between the two quantities. The weight given to the minimisation of the side constraint is controlled by the regularisation parameter. Thus, the regularisation parameter is an important factor that controls the quality of the regularised solution. A good regularisation parameter should fairly balance the perturbation error and the regularisation error in the regularised solution.

2. Methods for choosing regularization parameter

There are several methods for choosing regularisation parameter. In the paper we examine three of them.

2.1. L-curve method

The L-curve is a log-log plot – for all valid regularisation parameters – of the norm $||Lx_\alpha||_2$ of the regularised solution versus the norm of the corresponding residual norm $||Ax_\alpha - b||_2$ [4]. We can write the solution and residual norms in terms of the singular value decomposition (SVD):

$$||x_\alpha||_2^2 = \sum_{i=1}^{n} \left(f_i \frac{u_i^T b}{\sigma_i} \right)^2 \tag{4}$$

$$||Ax_\alpha - b||_2 = \sum_{i=1}^{n} \left((I - f_i)\, u_i^T B \right)^2 \tag{5}$$

In this way, the L-curve clearly displays the compromise between the minimisations of these two quantities, which is the heart of any regularisation method. For discrete ill-posed problems, the L-curve has a characteristic L-shaped appearance with a distinct corner separating the horizontal and the vertical parts of the curve, where the solution is dominated by regularisation error and perturbation errors, respectively. Two meanings of the "corner" were suggested by Hansen and O'Leary [4]. The first one is the point where the curve closest to the origin, the second one is the point where the curvature is maximum. L-curve for Tikhonov regularization is important in the analysis of discrete ill-posed problems. Figure 1 shows an example of a typical L-curve.

2.2. U-curve method

Consider the following function

$$U(\alpha) = \frac{1}{x(\alpha)} + \frac{1}{y(\alpha)},\ for\ \alpha > 0 \tag{6}$$

where $x(\alpha)$ and $y(\alpha)$ are defined by

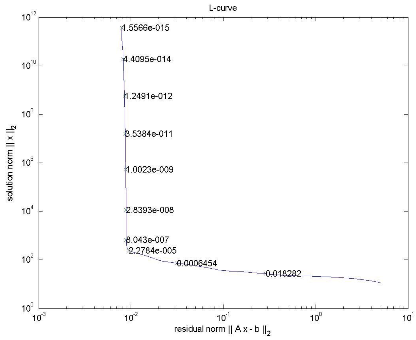

Figure 1. L-curve

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

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

By U-curve we understand the plot of $U(\alpha)$, i.e. the plot of the sum of the reverse both of the regularized solution norm and corresponding residual norm, for $\alpha > 0$ [1]. Figure 2 shows an example of a typical U-curve.
The U-curve consists of three characteristic parts, namely: on the left and right side, almost "vertical" parts, in the middle almost, "horizontal" part. The vertical parts correspond to the regularization parameter, for which the solution norm and the residual norm are dominated by each other respectively. The more horizontal part corresponds to the regularization parameter, for which the solution norm and the residual norm are close to each other. The objective of the U-curve criterion for selecting the regularization parameter is to choose a parameter where the curvature attains a local maximum close to the left vertical part of the U-curve.

2.3. Vapnik method of empirical risk minimization

The details of the Vapnik's method can be found in [2]. The principle is the one of minimizing the empirical risk which is the empirical counterpart of the expected

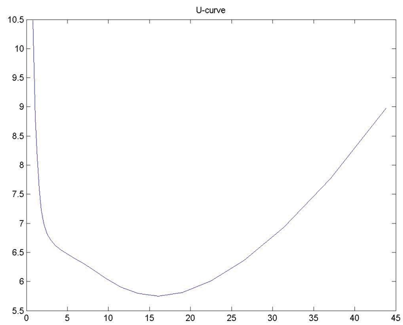

Figure 2. U-curve

risk. The starting point is that of Tikhonov functional minimization (formula 3) with L being an identity operator. Hence, the problem is reduced to the solution of the regularized set of normal equations:

$$\left(A^T A + \alpha I\right) x = A^T b \tag{9}$$

Following Vapnik, we introduce a topological ε-net in the space of admissible solutions to (9). The set consists of vectors of the form

$$V = \sum_{j=1}^{n} \frac{\varepsilon \mu_j}{\lambda_j} \vartheta_j \tag{10}$$

where λ_j, ϑ_j are, respectively, eigenvalues and eigenvectors of the normal matrix $A^T A, \mu_j$, are natural numbers and ε is the step of the ε-net. The normal set of equations is then solved for given α to find a node of the ε-net which is closest to found solution x^α. The node can be found from the formula

$$V_\alpha = \sum_{j=1}^{n} \left[\frac{x^\alpha \vartheta_j}{\varepsilon} \lambda_j + \frac{1}{2} \right] \frac{\varepsilon}{\mu_j} \vartheta_j \tag{11}$$

where the expression in square brackets stands for the entier function. The quality of the solution is then estimated by minimizing the empirical risk functional with respect to α and ε. As a result, we can find the optimal regularization parameter α for which the solution x^α is of best quality [2]. The Vapnik's method is closely related to the ridge estimates approach.

3. RBF network

Neural networks are composed of simple elements operating in parallel. These elements are inspired by biological nervous systems. As in nature, the network function is determined largely by the connections between elements. We can train a neural network to perform a particular function by adjusting the values of the connections (weights) between elements. Today neural networks can be trained to solve problems that are difficult for conventional computers or human beings-they are used in engineering, financial and other practical applications [5]. Typically many input-output pairs are used, in supervised training, to train a network.

The construction of a radial-basis function (RBF) network in its most basic form involves three entirely different layers. The input layer is made up of source nodes (sensory units). The second layer is a hidden layer of high enough dimension, which serves a different purpose from that in a multilayer perceptron. The output layer supplies the response of the network to the activation patterns applied to the input layer. The transformation from the input space to the hidden-unit space is nonlinear, whereas the transformation from the hidden-unit space to the output space is linear. From the output space we get:

$$F(x) = \sum_i w_i \varphi(\|x - c_i\|) \tag{12}$$

where $(\|..\|)$ is, for the most part, Euclidean norm, are weights, $\varphi(\|x - c_i\|)$ are radial basis functions the values of which vary in a radial way around the centre.

Let the set of input-output data available for approximation is described by: input signal: $x_i \in R^p$,$i = 1, 2, ..., N$ and desired response: $d_i \in R$, $i = 1, 2, ..., N$,. Let the approximating function be denoted by $F(x_i) = d_i$, where we have omitted the weight vector w of the network from the argument of the function F. According to Tikhonov regularization theory, the function F is determined by minimizing a cost functional

$$E(F) = \frac{1}{2}\sum_{i=1}^{N} [d_i - F(x_i)]^2 + \frac{1}{2}\alpha \|PF\|^2 \tag{13}$$

The α is a positive real number called the regularization parameter. In [6] Haykin presented the solution to the regularization problem. Following him, we can write down the formula (3.1) as follows:

$$F(x_i) = \sum_{i=1i}^{N} w_i G(x_j, x_i), \quad j = 1, 2, 3, ...N \tag{14}$$

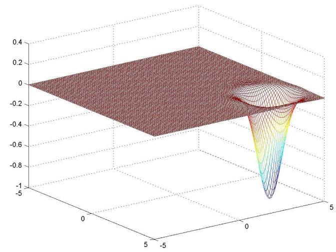

Figure 3. The plot of exact function

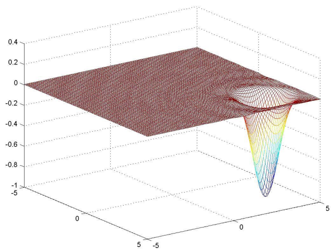

Figure 4. The plot of approximated function; U-curve

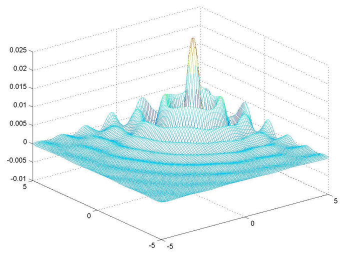

Figure 5. The plot of error

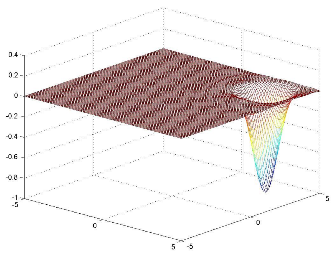

Figure 6. The plot of approximated function; L-curve

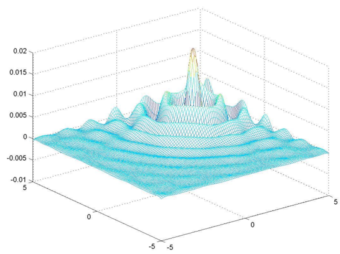

Figure 7. The plot of error

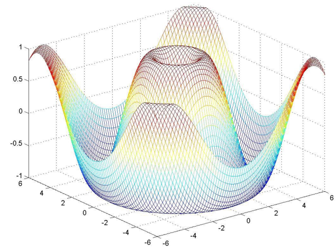

Figure 8. The plot of exact function

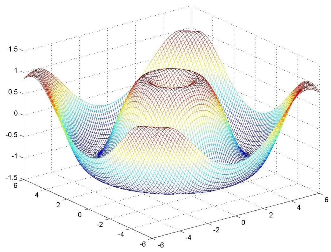

Figure 9. The plot of approximated function; U-curve

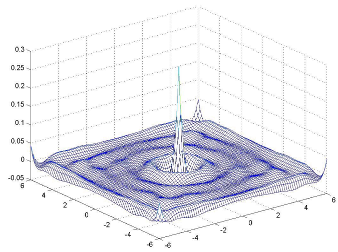

Figure 10. The plot of error

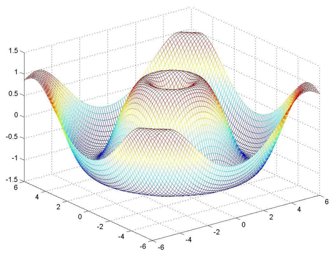

Figure 11. The plot of approximated function; L-curve

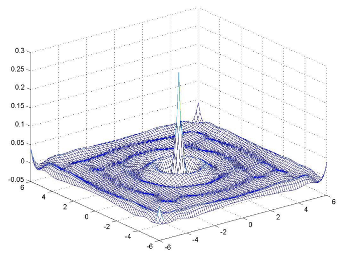

Figure 12. The plot of error

And similarly introduce the following:

$$F = [F(x_1), ..., F(x_N)]^T\,, \qquad\qquad d = [d_1, ..., d_N]^T\,,$$

$$G = \begin{bmatrix} G(x_1, x_1) & .... & G(x_1, x_N) \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ G(x_N, x_1) & .... & G(x_N, x_N) \end{bmatrix}, \quad w = [w_1, ..., w_N]^T\,, \tag{15}$$

we get

$$F = Gw \tag{16}$$

For $w = \frac{1}{\lambda}(d - F)$, we get

$$(G + \lambda I)w = d \tag{17}$$

where I is the $N\!-\!by\!-\!N$ identity matrix. We call the matrix G the Green's matrix. In practice, we may always choose λ sufficiently large to ensure that $G+\lambda I$ is positive defined and, therefore, invertible. This, in turn, means that the linear system of equations (17) will have a unique solution given by [6]

$$w = (G + \lambda I)^{-1}\, d \tag{18}$$

We may use equation (18) to obtain the weight vector for a specified desired response vector d and an appropriate value of regularization parameter α.

4. Numerical examples

Example 1
The Easom function: $F(x, y) = -\cos(x)\cos(y)\exp((x - \pi)^2 + (y - \pi)^2)$

$D = \{(x, y) : -5 \le x \le 5, -5 \le y \le 5\}$

Learning patterns: 1000.
Error norm $1.3311e - 005$-*U-curve, parameter* 0.0256*, minimum* -0.9778
Error norm $1.1302e - 005$-*L-curve, parameter* 0.0126*, minimum* -0.9801

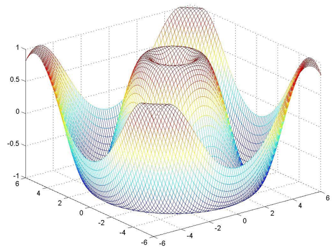

Figure 13. The plot of exact function

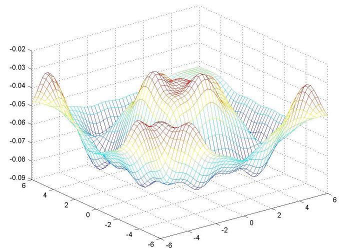

Figure 14. The plot of approximated function; U-curve

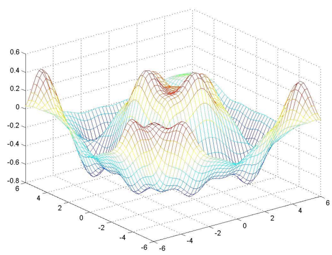

Figure 15. The plot of approximated function; L-curve

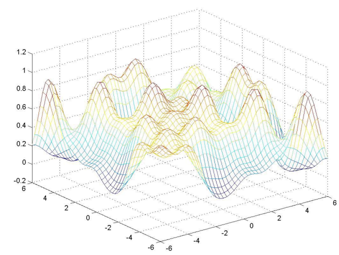

Figure 16. The plot of approximated function; Vapnik

Example 2
The function: $F(x, y) = \sin\sqrt{x^2 + y^2}$

$D = \{(x, y) : -6 \leq x \leq 6, -6 \leq y \leq 6\}$

Learning patterns: 6561.
Error norm $7.2635e - 005$-*U- curve, parameter* 0.3488
Error norm $7.3011e - 005$-*L- curve, parameter* 0.2314

Example 3
The function:$F(x, y) = \sin\sqrt{x^2 + y^2}$

$D = \{(x, y) : -6 \leq x \leq 6, -6 \leq y \leq 6\}$

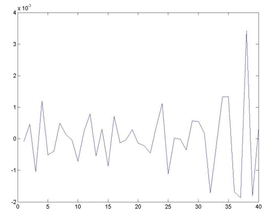

Figure 17. The plot of error; L-curve

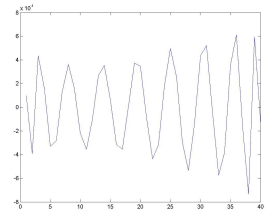

Figure 18. The plot of error; L-curve

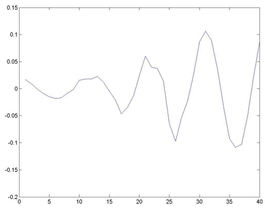

Figure 19. The plot of error; Vapnik

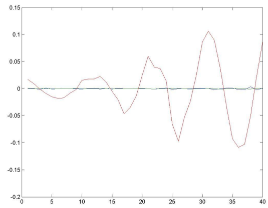

Figure 20. The plot of all error

Learning patterns: 36.
Error norm $1.3145e-002$-*U-curve, parameter* 6.1118
Error norm $7.0011e-003$-*L- curve, parameter* 0.8377
Error norm $1.6234e-002$-*Vapnik, parameter* 0.1201

Example 4
Equation Shaw [7].

Error norm $5.6101e-005$ -*U- curve, parameter* $4.3720e-012$
Error norm $1.5403e-004$ -*L- curve, parameter* $1.0634e-014$
Error norm 0.0080 *Vapnik, parameter* $0.100e-03$

5. Conclusions

Our study shows that the novel U-curve approach is fairly competitive with respect to the others. In principle, it can be used for finding a regularized solution to any ill-posed problem, including, e.g. Fredholm integral equations of the first

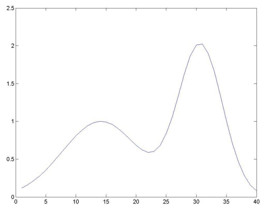

Figure 21. The plot of exact function

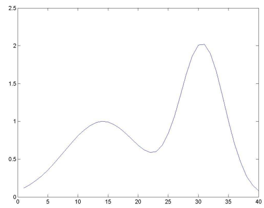

Figure 22. The plot of approximated function; L-curve

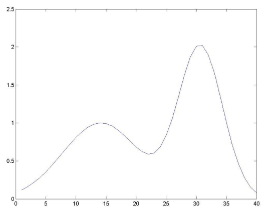

Figure 23. The plot of approximated function; U-curve

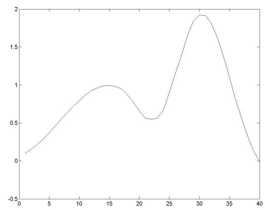

Figure 24. The plot of approximated function; Vapnik

kind. On the other hand, Vapnik's method based on regularised empirical risk minimization seems to be less reliable. The results obtained form L-curve method are comparable to those from the U-curve method. Therefore, it is a question of choice which method is best suited to the problem specific.

References

[1] Krawczyk-Stańdo, D. and Rudnicki, M., *Regularization parameter in discrete ill-posed problems-the use of the U-curve*, International Journal of Applied Mathematics and Computer Science, Vol. 17, No. 2, 2007, pp. 101–108.

[2] Vapnik, S., *Estimation of dependencies based on empirical data*, Verlag, Berli, 1982.

[3] Krawczyk-Stańdo, D. and Rudnicki, M., *Regularised synthesis of the magnetic field using the L-curve approach*, International Journal of Applied Electromagnetics and Mechanics, Vol. 22, No. 3,4, 2005, pp. 233–242.

[4] Hansen, P. C. and O'Leary, D. P., *The use of the L-curve in the regularization of discrete ill- posed problems*, SIAM J. Sci. Comput., Vol. 14, 1993, pp. 1487–1503.

[5] D. Krawczyk-Stańdo, M. R. and Stańdo, J., *Radial neural network learning using U-curve approach*, Polish Journal of Environmental Studies, Vol. 17s, No. 2A, 2008, pp. 42–46.

[6] Haykin, S., *Neural Networks*, a Comprehensive Foundation, Macmillan College Publishing Company, New York, 1994.

[7] Hansen, P. C., *Regularization Tools, a Matlab package for analysis and solution of discrete ill-posed problems; version 2.0 for Matlab 4.0*, Report UNIC-92-03, 1993.