Journal of Computational and Applied Mathematics 231 (2009) 914–924

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

A new variant of L-curve for Tikhonov regularization

Mansoor Rezghi, S. Mohammad Hosseini *

Department of Mathematics, Tarbiat Modares University, P.O. Box 14115-175, Tehran, Iran

ARTICLE INFO

Article history:
Received 17 June 2008
Received in revised form 16 December 2008

Keywords:
Tikhonov regularization
L-curve
Residual L-curve

ABSTRACT

In this paper we introduce a new variant of L-curve to estimate the Tikhonov regularization parameter for the regularization of discrete ill-posed problems. This method uses the solution norm versus the regularization parameter. The numerical efficiency of this new method is also discussed by considering some test problems.

1. Introduction

This paper is concerned with the computation of a meaningful approximate solution of linear systems of equations

$$Ax = b, \quad A \in \mathbb{R}^{m\times n}, x \in \mathbb{R}^n, b \in \mathbb{R}^m. \tag{1.1}$$

The singular values of the matrix accumulate at the origin and decay gradually to zero, and this makes the matrix severely ill-conditioned. Such systems are often referred to as linear discrete ill-posed problems.

Many linear discrete ill-posed problems that arise in applications have a right-hand side b that is contaminated by an error $e \in \mathbb{R}^m$ which may stem from discretization or measurement inaccuracies. Let $\bar{b}$ denote the unavailable error-free representative of the right-hand side vector b, i.e.

$$b = \bar{b} + e. \tag{1.2}$$

We would like to determine a solution $\bar{x}$ of an error-free linear system of equation

$$A\bar{x} = \bar{b}. \tag{1.3}$$

Since the right-hand side $\bar{b}$ is not available, we seek to determine an approximation of $\bar{x}$ by computing an approximate solution $\dot{x}$ of the available linear system (1.1). Note that, the solution $\dot{x}$ of (1.1) itself is possibly not a meaningful approximation of $\bar{x}$, due to the error e and ill-conditioning character of A.

In order to determine a meaningful approximation of $\bar{x}$, one typically replaces the linear system (1.1) by a nearby system that is less sensitive to perturbations of the right-hand side, and considers the solution of this system as an approximation of $\bar{x}$. This replacement is commonly referred to as regularization. Tikhonov regularization and truncated singular value decomposition(TSVD) are the most popular regularization methods, see Engl [1], Groetsch [2], Hanke [3], and Hansen [4] for detailed discussions of these methods.

In its simplest form, Tikhonov regularization replaces the solution of the linear system (1.1) by the minimization problem

$$\min_{x\in\mathbb{R}^n}\{\|Ax - b\|^2 + \lambda^2\|x\|^2\} \tag{1.4}$$

* Corresponding author.
E-mail addresses: Rezghi@modares.ac.ir (M. Rezghi), hossei_m@modares.ac.ir (S.M. Hosseini).

0377-0427/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.cam.2009.05.016

for a suitable positive value of the regularization parameter λ, where here and henceforth $\|.\|$ refers to the 2-norm of a vector or matrix. We remark that for any $\lambda > 0$, the minimization problem (1.4) has the unique solution

$$x_\lambda := (A^TA + \lambda^2 I)^{-1}A^Tb. \tag{1.5}$$

Let $A = U\Sigma V^T$ be the singular value decomposition of A, then the regularized solution x_λ is given by

$$x_\lambda = \sum_{i=1}^{n} f_i \frac{(u_i^Tb)}{\sigma_i} v_i, \qquad f_i = \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2}$$

and the solution and residual norms for x_λ are given by [5]

$$\|x_\lambda\|^2 = \sum_{i=1}^{n} f_i^2 \frac{(u_i^Tb)^2}{\sigma_i^2} \tag{1.6}$$

$$\|r_\lambda\|^2 = \|b - Ax_\lambda\|^2 = \sum_{i=1}^{n} (1 - f_i)^2 (u_i^Tb)^2. \tag{1.7}$$

In fact, the Tikhonov regularization seeks for some x_λ, so that at the same time provides a small residual $\|r_\lambda\|$ and a moderate value of the penalty term $\|x_\lambda\|$. But it is easy to see that $\|r_\lambda\|$ and $\|x_\lambda\|$ are increasing and decreasing functions of λ respectively. If the regularization parameter is chosen too small then $\|r_\lambda\|$ is small but $\|x_\lambda\|$ can be very large in comparison with $\|\bar{x}\|$ causing the solution to be very close to the original ill-posed solution $\dot{x}$, and hence one must expect instabilities. If λ becomes too large then $\|r_\lambda\|$ is large and $\|x_\lambda\|$ can be small and the problem we solve, has only a little connection with the original equation. Hence, the parameter λ controls how much weight is given to minimization of $\|x_\lambda\|$ relative to minimization of the residual. If we find a good balance between these two terms, via a suitable value of λ, it is then expected that the regularized solution be a good approximation of the exact solution. Thus finding the optimal parameter for regularization is an important and tough problem.

Methods for determining a suitable regularization parameter can be divided into two main classes such as the methods based on knowledge, or a good estimate, of error norm and methods that do not require any knowledge of $\|e\|$. The discrepancy principle is an example of the first class and the Cross-Validation and L-curve are examples from the second class.

Although, according to [6], those methods that do not use the level of noise explicitly may produce unsatisfactory results for $\|e\| \to 0$, but in many important applications the norm of the error in the given right-hand side is not explicitly known [4]. In this case the L-curve is a popular method for selection of a suitable regularization parameter. In recent years, some other versions of the L-curve have also been introduced. As two new versions, one can refer to a computationally less expensive strategy based on computing an L-ribbon [7,8] that contains the L-curve and its interior and also the other new method known as "residual L-curve" [9] that only uses the information of residual of the solution to pick the truncation parameter k in TSVD regularization. In residual L-curve the corner of $(k, \log(\|r_k\|))$ is proposed as an approximation of the optimal regularization parameter k.

In this paper we show that $(\lambda^2, \|x_\lambda\|^2)$ provides an L-shape curve on which its corner, can be used as a suitable regularization parameter in the Tikhonov regularization method. In our method, it is illustrated that before the corner of the curve, the solution is dominated by the error term and starts to decrease significantly after the corner point. Numerical results show that this method has at least the same performance as that of the L-curve but has a better performance in problems with very smooth solutions.

The organization of this paper is as follows. In Section 2, a brief introduction of the L-curve is given. In Section 3, our new method $(\lambda^2, \|x_\lambda\|^2)$ is discussed. In Section 4, some numerical results are reported.

2. L-curve

From previous section it is clear that a suitable regularization parameter should properly balance between $\|r_\lambda\|$ and $\|x_\lambda\|$ quantities. So it is natural to use the $(\|r_\lambda\|, \|x_\lambda\|)$ curve, that is named as the L-curve, for a tradeoff between these two quantities that should be controlled.

The name L-curve derives from the characteristic shape of this curve. Especially, for the case when λ is very large (over-regularization), the residual norm is very sensitive to small changes in λ while the solution norm is relatively constant, so the curve is essentially a horizontal line. Conversely, when λ is very small (under-regularization), changes in the solution norm occur much faster than changes in the residual norm, and the curve is essentially a vertical line. Thus such a plot has a characteristic "L" shape. The transition between these two regions of under- and over-regularizations corresponds to the "corner" of the L-curve and the associated value of λ at this corner has been proposed as the optimal value of the regularization parameter [10].

The use of such plots goes back to Miller [11] and Lawson and Hanson [12], while the term "L-curve" for this plot was introduced much later, where the properties of the "L-curve" were studied in details [10].

Nevertheless, computational experience shows the validity of L-curve criterion to give suitable values of regularization parameter for many problems, but it has its limitations. For example it is likely to have poor performance in problems where the solution is smooth [5,13,14].

3. New L-curve method

In this section it is shown that, by using the properties of the solution norm with respect to different values of λ^2 it is possible to obtain suitable value of regularization parameter for Tikhonov regularization.

In the following, we investigate the properties of the $(\lambda^2, \|x_\lambda\|^2)$ curve and show that this curve is decreasing and convex and has an L-shape.

Also it is shown that this curve, similar to the L-curve, has two characteristic parts, namely, a "flat" part where the regularized solution x_λ is dominated by regularization errors and an almost "vertical" part where x_λ is dominated by the noise. So the "corner" of this curve which separates these two parts can be used to locate the regularization parameter.

The solution x_λ of the regularized system (1.5) can be written as

$$x_\lambda = \bar{x}_\lambda + x_\lambda^e, \tag{3.1}$$

where $\bar{x}_\lambda = (A^TA + \lambda^2 I)^{-1}A^T\bar{b}$ is the regularized version of the exact solution $\bar{x}$, and $x_\lambda^e = (A^TA + \lambda^2 I)^{-1}A^Te$ is the effect of pure noise in the regularized solution.

To explore the main properties of $(\lambda^2, \|x_\lambda\|^2)$ and the correlated effects of errors and regularization in the behavior of this curve, it can be easily shown that

$$\eta(\lambda) = \|x_\lambda\|^2 = \sum_{i=1}^{n} f_i^2 \frac{(u_i^Tb)^2}{\sigma_i^2}, \tag{3.2}$$

and

$$\frac{d}{d\lambda^2}(\eta) = -2\sum_{i=1}^{n} f_i^3 \frac{(u_i^Tb)^2}{\sigma_i^4}, \tag{3.3}$$

$$\frac{d}{d\lambda^2}\left(\frac{d}{d\lambda^2}\right)(\eta) = 6\sum_{i=1}^{n} f_i^4 \frac{(u_i^Tb)^2}{\sigma_i^6}. \tag{3.4}$$

So, $(\lambda^2, \eta(\lambda))$ is a decreasing and convex curve, and moreover satisfies the following relations

$$\lim_{\lambda^2\to 0} \eta(\lambda) = \|\dot{x}\|^2 \qquad \lim_{\lambda^2\to\infty} \eta(\lambda) = 0. \tag{3.5}$$

The decreasing and convexity properties of the $(\lambda^2, \eta(\lambda))$ hold for any right-hand side vector b. Also, according to (3.1), to gain insight into the properties of the $(\lambda^2, \|x_\lambda\|^2)$ curve and the effect of noise and regularization errors on the solution, it is also necessary to analyze and compare the behavior of $(\lambda^2, \|\bar{x}_\lambda\|^2)$ and $(\lambda^2, \|x_\lambda^e\|^2)$. To simplify the analysis, without losing generality, we can assume that $\sigma_1 = 1, 0 \le \lambda \le 1$ and $(u_i^T\bar{b}) \approx \sigma_i^{p+1}$, where p is a real number that controls the behavior of the right hand-side. This assumption is reasonable and it is the basis of many studies of model problems in ill-posed problems[5]. The case $p > 0$ corresponds to a right-hand side that satisfies the discrete Picard condition (DPC), while $p < 0$ corresponds to a right-hand side that does not satisfy the discrete Picard condition [5].

In particular, $p = -1$ corresponds to a right-hand side e which consisting of white noise.

$$(u_i^Te) \approx 1. \tag{3.6}$$

Since $\bar{b}$ satisfies the DPC, then p must be positive. From these assumptions, it is clear that

$$\bar{\eta}(\lambda) = \|\bar{x}_\lambda\|^2 \approx \sum_{i=1}^{n} \frac{\sigma_i^{2(p+2)}}{(\sigma_i^2 + \lambda^2)^2}, \tag{3.7}$$

$$\eta^e(\lambda) = \|x_\lambda^e\|^2 \approx \sum_{i=1}^{n} \frac{\sigma_i^2}{(\sigma_i^2 + \lambda^2)^2}, \tag{3.8}$$

$$\eta(\lambda) = \|x_\lambda\|^2 \approx \sum_{i=1}^{n} \frac{\sigma_i^2(\sigma_i^{(p+1)} + 1)^2}{(\sigma_i^2 + \lambda^2)^2}. \tag{3.9}$$

For comparison the behavior of these curves, suppose that $\lambda = \sigma_k$, then

$$\eta^e(\lambda) \approx \sum_{i=1}^{k} \sigma_i^{-2} + \lambda^{-4}\sum_{i=k+1}^{n} \sigma_i^2, \tag{3.10}$$

where, the first term $\sum_{i=1}^{k} \sigma_i^{-2}$ is dominated by $\sigma_k^{-2} = \lambda^{-2}$ and $\sum_{i=k+1}^{n} \sigma_i^2$ is small quantity which is dominated by $\sigma_k^2 = \lambda^2$. So the above equation can be approximated by the following expression

$$\eta^e(\lambda) \approx c/\lambda^2, \tag{3.11}$$

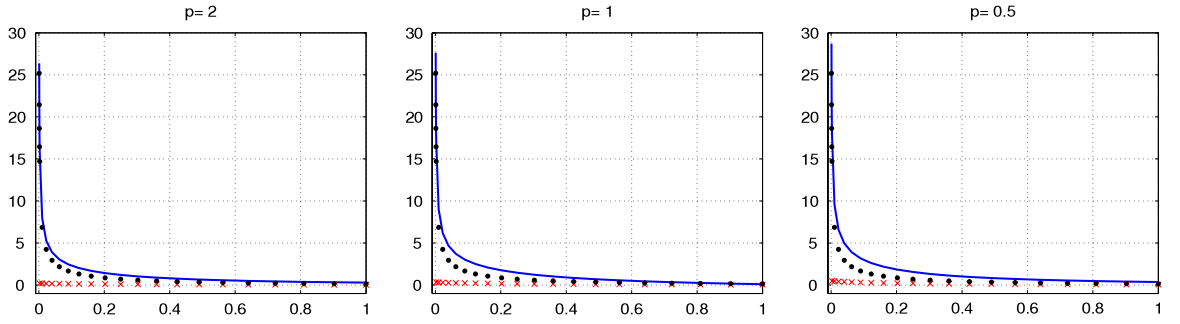

Fig. 1. Results for model problem with $p = 2, 1, 0.5$, solid (blue) for $(\lambda^2, \Gamma(\lambda))$, point(black) for $(\lambda^2, \Gamma^e(\lambda))$ and cross(red) for $(\lambda^2, \bar{\Gamma}(\lambda))$.

where c is dependent on λ and can be considered as constant, as it has very little variation with λ in comparison with $1/\lambda^2$. This relation shows that when λ approaches the origin the $\eta^e(\lambda)$ is very large. A similar approximation for $\bar{\eta}(\lambda)$ can be obtained

$$\bar{\eta}(\lambda) \approx \sum_{i=1}^{k} \sigma_i^{2p} + \lambda^{-4} \sum_{i=k+1}^{n} \sigma_i^{2(p+2)}. \tag{3.12}$$

The first term $\sum_{i=1}^{k} \sigma_i^{2p}$ is dominated by σ_1^{2p} and $\sum_{i=k+1}^{n} \sigma_i^{2(p+2)}$ is dominated by $\sigma_k^{2(p+2)} = \lambda^{2(p+2)}$. So $\bar{\eta}(\lambda)$ can be approximated as

$$\bar{\eta}(\lambda) \approx c_1 + c_2\lambda^{2p}, \tag{3.13}$$

where c_1 and c_2 can also be considered as constants. The above relation shows that $\bar{\eta}(\lambda)$ is very small in comparison with the $\eta^e(\lambda)$. A similar approximation can be obtained for $\eta(\lambda)$ as follows

$$\eta(\lambda) \approx \frac{c_3}{\lambda^2} + c_4\lambda^{2p} \tag{3.14}$$

where c_3, c_4 are some constants.

From these estimates it is evident that as $\lambda \to 0$ the behaviors of $(\lambda^2, \eta^e(\lambda))$ and $(\lambda^2, \eta(\lambda))$ are the same and very different from $(\lambda^2, \bar{\eta}(\lambda))$.

Any small variation of λ near zero causes very large changes in $(\lambda^2, \eta^e(\lambda))$ and $(\lambda^2, \eta(\lambda))$. (3.14) shows that as λ approaches zero, $\eta(\lambda)$ is dominated by a noise factor $\frac{1}{\lambda^2}$ and for large values of λ this effect is small. Although the values of p have an effect on $\bar{\eta}(\lambda)$, this effect is very small in comparison with $\eta^e(\lambda)$.

For a further analysis, we consider their infinite(continuous) versions and replace the discrete variable σ_i with continuous variable σ, and the summation with an integral. If the continuous versions of $\bar{\eta}(\lambda)$, $\eta^e(\lambda)$ and $\eta(\lambda)$ are denoted by $\bar{\Gamma}(\lambda)$, $\Gamma^e(\lambda)$ and $\Gamma(\lambda)$ respectively, then

$$\bar{\Gamma}(\lambda) = \int_0^1 \frac{\sigma^{2(p+2)}}{(\sigma^2 + \lambda^2)^2} \mathrm{d}\sigma, \tag{3.15}$$

$$\Gamma^e(\lambda) = \int_0^1 \frac{\sigma^2}{(\sigma^2 + \lambda^2)^2} \mathrm{d}\sigma, \tag{3.16}$$

$$\Gamma(\lambda) = \int_0^1 \frac{\sigma^2(\sigma^{(p+1)} + 1)^2}{(\sigma^2 + \lambda^2)^2} \mathrm{d}\sigma. \tag{3.17}$$

We plot $(\lambda^2, \bar{\Gamma}(\lambda))$, $(\lambda^2, \Gamma^e(\lambda))$ and $(\lambda^2, \Gamma(\lambda))$ for some values of p, for example, 2, 1 and 0.5, for $0 < \lambda < 1$ to provide an insight into the way that they work, which helps to interpret the behavior of these curves and the effect of error in the solution, see Fig. 1. From Fig. 1, the following conclusions are obviously reached that confirm the approximations (3.11), (3.13) and (3.14):

•The $(\lambda^2, \bar{\Gamma}(\lambda))$ curve is flat, and small in comparison with $(\lambda^2, \Gamma^e(\lambda))$. This is natural because $\bar{b}$ satisfies the discrete Picard condition, so the solution coefficients $|v_i^{\mathrm{T}}\bar{x}_\lambda| = f_i|\frac{u_i^{\mathrm{T}}\bar{b}}{\sigma_i}|$ also decay. This ensures that $\bar{x}_\lambda$ does not have a large norm for any λ.

•The $(\lambda^2, \Gamma^e(\lambda))$ has two main parts: For small λ the value of $\Gamma^e(\lambda)$ is very large and it looks vertical and for large λ this curve is approximately horizontal. So, this behavior of $\Gamma^e(\lambda)$ is similar to the estimate (3.11) of $\eta^e(\lambda)$.

•The $(\lambda^2, \Gamma(\lambda))$ has a behavior similar to that of $(\lambda \cdot \Gamma^e(\lambda))$ for very small λ's, i.e., this curve for small λ is dominated by a pure-noise curve, but such an effect becomes significantly small as λ increases. Somewhere in-between, there is a range of λ-values, for which the transition between these two parts occurs. So choosing the regularization parameter in this part – that we call it corner of the curve – is reasonable. Similar to the L-curve there are several methods to choose this corner. For

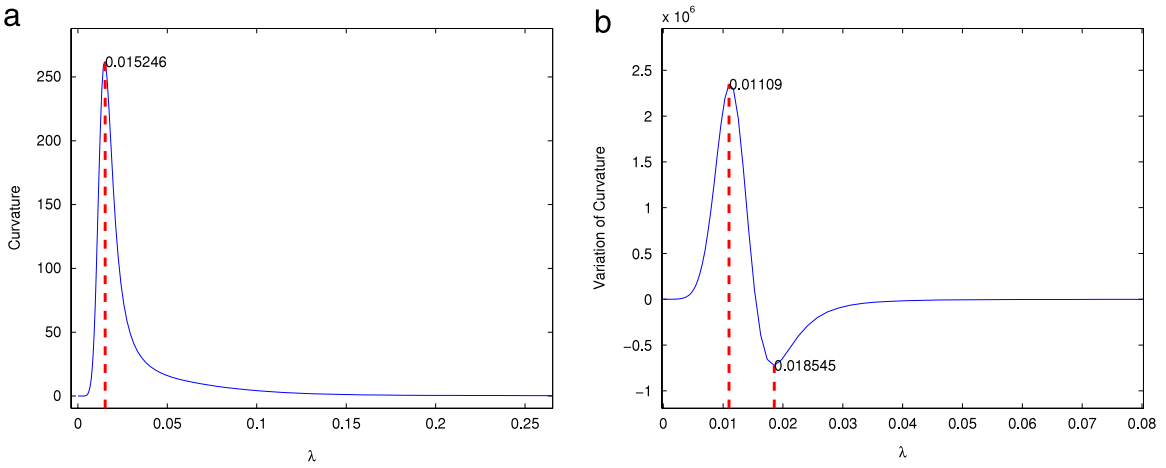

Fig. 2. (a) Curvature and (b) Variations of curvature, for **phillips** test problem.

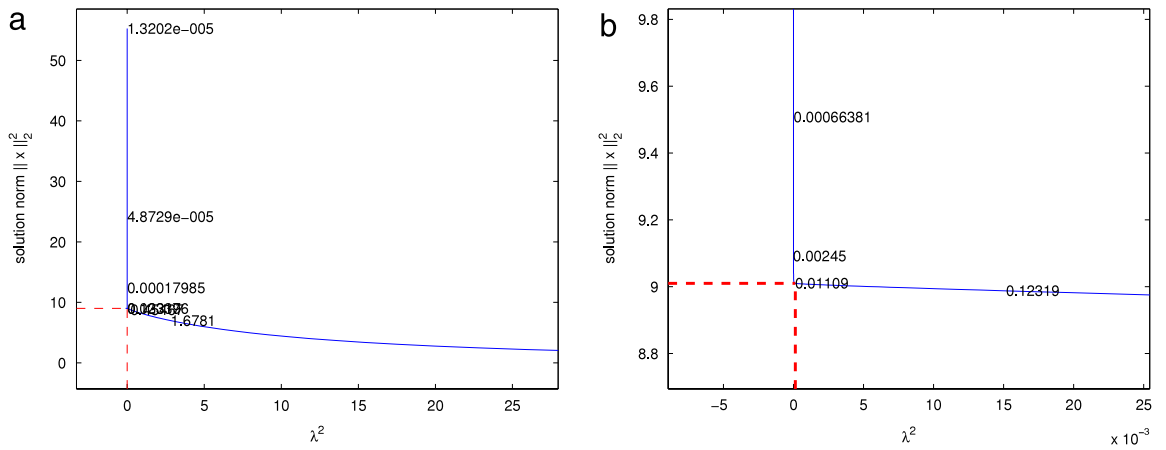

Fig. 3. $(\lambda^2, \|x_\lambda\|^2)$ curve and its corner for **phillips**, (b) closer view of (a) to corner.

example, using maximum curvature criteria is one of the possibilities. But as our experiment shows, if k_i is the curvature of the $(\lambda^2, \|x_\lambda\|^2)$ evaluated at λ_i, curvature variations of $(\lambda^2, \|x_\lambda\|^2)$ (i.e $v_i = \frac{k_{i+1}-k_i}{\lambda_{i+1}^2-\lambda_i^2}$) provides a better means of determining the regularization parameter. To this end, we detect a point with maximum $|v_i|$ as the regularization parameter.

Experiments show that the result of this strategy, i.e. maximum variation of curvature criteria, is better overall. So, in numerical section we only report the results of the maximum variation of curvature criteria.

Fig. 2 shows the curvature and curvature variations of $(\lambda^2, \|x_\lambda\|^2)$ for **phillips**(64) test problem [15] where $\|e\| = 10^{-4}$.

Also Fig. 3 shows the $(\lambda^2, \|x_\lambda\|^2)$ and corner of this curve that obtained by the maximum variation of curvature method for the above test problem. This Figure shows that $(\lambda^2, \|x_\lambda\|^2)$ has L-shape.

In the L-curve method, using a log-log scale is necessary to distinguish between noise and signal, because the slope of the L-curve is $1/\lambda^2$ and is independent of the right-hand side and so we hardly distinguish signal from noise if plot it on a line-line scale. In a log-log scale the slope of the L-curve depends on the right-hand side, and presents different behaviors for noise and signal, see [13,16] for details.

The slope of the curve $(\lambda^2, \|x_\lambda\|^2)$, depends on the right-hand side and so it behaves differently for signal and noise. This behavior can be observed from estimates (3.11), (3.13) and (3.14) and Fig. 1.

4. Numerical results

In this section we investigate the performance of our method. We applied this method to several test problems and compared the results with those of the L-curve, for which we have used Hansen's Regularization Tools package [15]. In our computations we have used maximum variation of curvature criteria. Although we have only analyzed the curve $(\lambda^2, \|x_\lambda\|^2)$, we have also applied the mentioned strategies to $(\lambda, \|x_\lambda\|)$ accordingly and reported its numerical results for comparison. All the computations were carried out in Matlab. In the following examples the contaminated vector b and the error term e

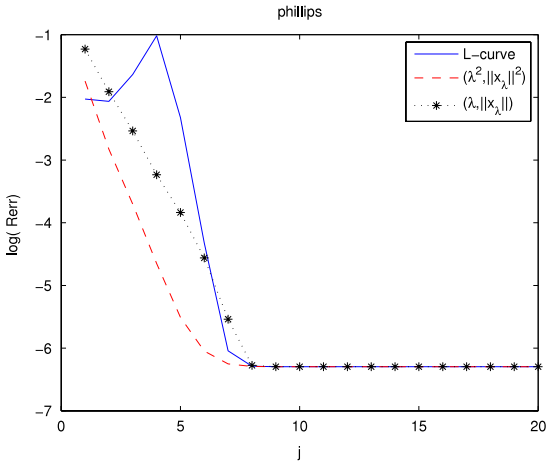

Fig. 4. Relative errors of proposed methods and L-curve for $j = 1, \ldots, 20$ where $e = 10^{-j}\|b\|\frac{w}{\|w\|}$.

are considered as:

$$b = \bar{b} + e. \tag{4.1}$$

We used a broad range of error levels $e = 10^{-j}\|\bar{b}\|\frac{w}{\|w\|}$, $j = 1, \ldots, 20$, where w is a vector with normally distributed entries with zero mean and unit variance. For the comparison of the validity of L-curve method and the new proposed methods, we have generated 20 sample problems for each noise level. Thus for each model problem we have solved 400 regularization sample problems.

For comparison, we plot the averages of relative errors for each level of noise.

Example 4.1. Consider the Fredholm integral equation of the first kind,

$$\int_{-6}^{+6} k(s,t)x(t)\mathrm{d}t = y(s), \quad -6 \le s \le 6 \tag{4.2}$$

discussed by **phillips** [15]. Its solution, kernel, and right-hand side are given by

$$x(t) = \begin{cases} 1 + \cos\left(\frac{\pi}{3}t\right), & \text{if } |t| < 3, \\ 0, & \text{otherwise} \end{cases}$$
$$k(s,t) = x(s-t),$$
$$y(s) = (6 - |s|)\left(1 + \frac{1}{2}\cos\left(\frac{\pi}{3}s\right)\right) + \frac{9}{2\pi}\sin\left(\frac{\pi}{3}|s|\right).$$

We use the code **phillips** from [15] to discretize (4.2) by the Galerkin method to obtain the symmetric indefinite matrix $A \in \mathbb{R}^{64\times 64}$ and vector solution $\bar{x}$. We determine the error-free right-hand side from $\bar{b} = A\bar{x}$, and contaminated right-hand sides b are computed according to (4.1). We solve 400 sample problems as mentioned above, and plot the logarithm of the average of the relative errors for each level in Fig. 4. In Table 1 it is possible to see one of the sample results explicitly for $e = 10^{-j}\|\bar{b}\|\frac{w}{\|w\|}$, $j = 1, \ldots 10$.

Fig. 4 shows that from 10^{-2} to 10^{-8}, $(\lambda^2, \|x_\lambda\|^2)$ has a better results in comparison with those of the L-curve and $(\lambda, \|x_\lambda\|)$. From 10^{-9} to 10^{-20} their relative errors are almost constant and so have the same performance.

Example 4.2. The inverse Laplace transform

$$\int_0^{\infty} \exp(-st)x(t)\mathrm{d}t = \frac{16}{(2s+1)^3} \quad s \ge 0 \tag{4.3}$$

has the solution $x(t) = t^2\exp(-t/2)$. We discretize (4.3), using the function **ilaplace** [15], to obtain the matrix $A \in \mathbb{R}^{64\times 64}$ and the vectors $\bar{x}$, $\bar{b}$.

Fig. 5 shows that the performance of $(\lambda^2, \|x_\lambda\|^2)$ is slightly better than the L-curve. It also can be seen, that the relative error of $(\lambda^2, \|x_\lambda\|^2)$ decreases with the level of noise, while the L-curve method does not show this property and even from $j = 11$ its relative error starts to increase. Table 2 shows the details of one its samples.

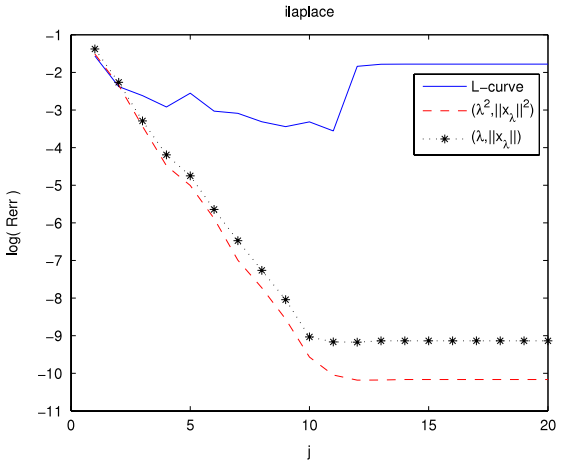

Fig. 5. Relative errors versus j for which 10^{-j} is the level of noise.

Table 1
Numerical results for **phillips** for $e = 10^{-j}\|\bar{b}\|\frac{w}{\|w\|}, j = 1, \ldots, 10$.

j	L-Curve		$(\lambda^2, \|x_\lambda\|^2)$		$(\lambda, \|x_\lambda\|)$	
	λ	Rerr	λ	Rerr	λ	Rerr
1	3.91e−1	1.03e−1	2.88e−1	1.56e−1	1.98e−1	2.66e−1
2	6.12e−2	1.25e−1	1.08e−1	6.69e−2	5.17e−2	1.53e−1
3	6.99e−3	1.68e−1	3.56e−2	2.53e−2	1.35e−2	7.15e−2
4	1.77e−4	6.17e−1	1.11e−2	9.07e−3	3.66e−3	3.98e−2
5	9.24e−5	9.01e−2	4.09e−3	4.02e−3	6.59e−4	2.37e−2
6	9.24e−5	9.18e−3	1.07e−3	2.39e−3	7.28e−5	1.05e−2
7	9.24e−5	2.04e−3	1.46e−4	1.96e−3	1.32e−5	2.93e−3
8	1.32e−5	1.86e−3	4.50e−5	1.85e−3	1.32e−5	1.85e−3
9	1.32e−5	1.84e−3	2.50e−5	1.84e−3	1.32e−5	1.84e−3
10	1.32e−5	1.84e−3	2.50e−5	1.84e−3	1.32e−5	1.84e−3

Table 2
Numerical results for **ilaplace** for $e = 10^{-j}\|\bar{b}\|\frac{w}{\|w\|}, j = 1, \ldots, 10$.

j	L-Curve		$(\lambda^2, \|x_\lambda\|^2)$		$(\lambda, \|x_\lambda\|)$	
	λ	Rerr	λ	Rerr	λ	Rerr
1	1.35e−1	2.05e−1	1.31e−1	2.04e−1	1.02e−1	1.98e−1
2	1.38e−2	7.83e−2	1.25e−2	8.03e−2	1.15e−2	8.34e−2
3	1.73e−3	3.87e−2	1.95e−3	3.61e−2	1.74e−3	3.84e−2
4	1.59e−4	9.18e−2	9.89e−4	7.72e−3	7.67e−4	8.05e−3
5	1.09e−5	1.20e−1	3.10e−4	2.29e−3	2.30e−4	3.90e−3
6	2.05e−6	1.97e−2	5.82e−5	2.58e−3	3.55e−5	4.62e−3
7	1.67e−7	2.76e−2	4.01e−6	1.62e−3	2.94e−6	1.68e−3
8	2.25e−8	3.88e−2	1.47e−6	1.88e−4	7.25e−7	4.07e−4
9	1.83e−9	6.15e−2	3.86e−7	4.65e−5	2.09e−7	1.59e−4
10	1.76e−10	2.79e−2	8.57e−8	2.22e−5	6.01e−8	6.39e−5

Example 4.3. Let us consider the integral equation

$$\int_{-\pi/2}^{\pi/2} k(s,t)x(t)\mathrm{d}t = y(s), \quad -\frac{\pi}{2} \le s \le \frac{\pi}{2} \tag{4.4}$$

where

$$k(s,t) = (\cos(s) + \cos(t))^2 \left(\frac{\sin(\xi)}{\xi}\right)^2, \quad \xi = \pi(\sin(s) + \sin(t))$$

and the right-hand side $y(s)$ is chosen so that solution $x(t)$ is a sum of two Gaussian functions. We use the code **shaw** from [15] to discretize (4.4) by a quadrature rule with 64 nodes. As it is expected to see a poor performance of the L-curve method in the problem with a smooth solution [14], in this example, we will also show that when the solution is smooth, $(\lambda^2, \|x_\lambda\|^2)$ has significantly better performance. To this end, we consider the **shaw** problem in two cases.

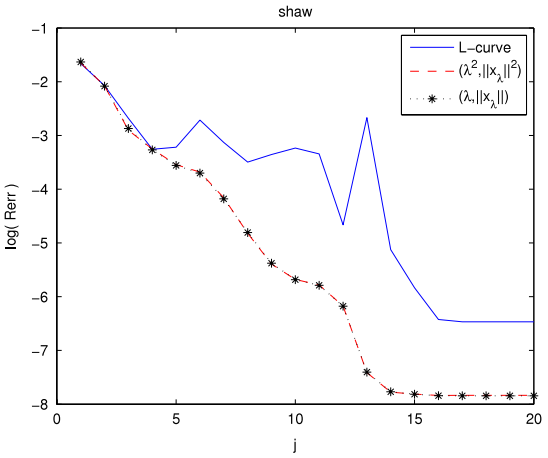

Fig. 6. Relative errors versus j for which 10^{-j} is the level of noise.

Table 3
Numerical results for **shaw** (Case 1) for $e = 10^{-j}\|\bar{b}\|\frac{w}{\|w\|}, j = 1, \ldots, 10$.

j	L-Curve		$(\lambda^2, \|x_\lambda\|^2)$		$(\lambda, \|x_\lambda\|)$	
	λ	Rerr	λ	Rerr	λ	Rerr
1	1.51e−1	1.54e−1	1.06e−1	1.60e−1	1.00e−1	1.62e−1
2	1.53e−2	9.91e−2	1.01e−2	1.09e−1	9.06e−3	1.13e−1
3	2.66e−3	7.33e−2	3.67e−3	6.21e−2	3.24e−3	6.58e−2
4	4.14e−4	2.95e−2	7.73e−4	3.67e−2	6.55e−4	3.51e−2
5	2.41e−5	4.76e−2	2.12e−4	2.55e−2	1.94e−4	2.46e−2
6	2.32e−6	5.92e−2	1.52e−4	2.72e−2	1.42e−4	2.68e−2
7	2.23e−7	4.83e−2	6.33e−6	1.53e−2	5.42e−6	1.44e−2
8	3.00e−8	2.75e−2	1.89e−7	6.60e−3	1.77e−7	6.65e−3
9	2.07e−9	6.94e−2	4.96e−8	3.10e−3	4.36e−8	2.96e−3
10	2.35e−10	6.49e−2	2.15e−8	2.98e−3	2.34e−8	3.06e−3

Table 4
Numerical results for **Modify-shaw** (Case 2) for $e = 10^{-j}\|\bar{b}\|\frac{w}{\|w\|}, j = 1, \ldots, 10$.

j	L-Curve		$(\lambda^2, \|x_\lambda\|^2)$		$(\lambda, \|x_\lambda\|)$	
	λ	Rerr	λ	Rerr	λ	Rerr
1	1.98e−1	8.88e−2	1.48e−1	1.11e−1	1.16e−1	1.38e−1
2	2.62e−2	6.05e−2	5.41e−2	3.15e−2	1.13e−2	9.52e−2
3	5.14e−3	1.85e−2	1.01e−2	1.01e−2	4.45e−3	2.27e−2
4	4.47e−4	2.90e−2	5.02e−3	1.91e−3	6.55e−4	2.21e−2
5	2.85e−5	6.03e−2	1.87e−3	8.51e−4	2.27e−4	6.90e−3
6	3.24e−6	5.54e−2	4.47e−4	2.90e−4	2.57e−5	6.34e−3
7	2.64e−7	5.55e−2	1.79e−4	1.01e−4	8.65e−5	1.62e−3
8	3.55e−8	3.10e−2	2.41e−5	6.55e−5	2.49e−6	7.52e−4
9	2.45e−9	8.24e−2	1.24e−5	1.10e−5	5.25e−7	3.41e−4
10	2.78e−10	7.20e−2	4.53e−6	3.61e−6	8.13e−8	1.29e−4

Case 1: The **shaw** problem with exact solution $\bar{x}$, $\bar{b} = A\bar{x}$, and $b = \bar{b} + e$.

Fig. 6, similar to the previous problems, confirms again the better performance of our proposed method. Also it can be seen that relative errors of $(\lambda^2, \|x_\lambda\|^2)$ is decreasing with level of noise, but L-curve does not have this property. Table 3 shows details of one of its samples.

Case 2: Consider the **shaw** problem with the exact solution $\bar{\bar{x}} = \sigma_1^{-2}A^{\mathrm{T}}A\bar{x}$ and the exact right-hand side obtained from $\bar{\bar{b}} = A\bar{\bar{x}}$, and $b = \bar{\bar{b}} + e$. In this case the solution is smoother than that of the **Case 1**, see Fig. 7 and [5].

In this case the result of the new method $(\lambda^2, \|x\|^2)$ is significantly better than the result of the L-curve and also other proposed method. Fig. 8 clearly illustrates their comparative performances. Table 4 also lists some related numerical results for $e = 10^{-j}\|\bar{\bar{b}}\|\frac{w}{\|w\|}, j = 1, \ldots, 10$.

Example 4.4. Now we consider here some other examples. We take **deriv2**, **foxgood** and **baart** from Regtool [15]. We also take the **Hilbert** matrix from Matlab such that $x_i = 1/(i^3 + i^2 + i + 1), i = 1, \ldots, 64$. The following is a report of the results

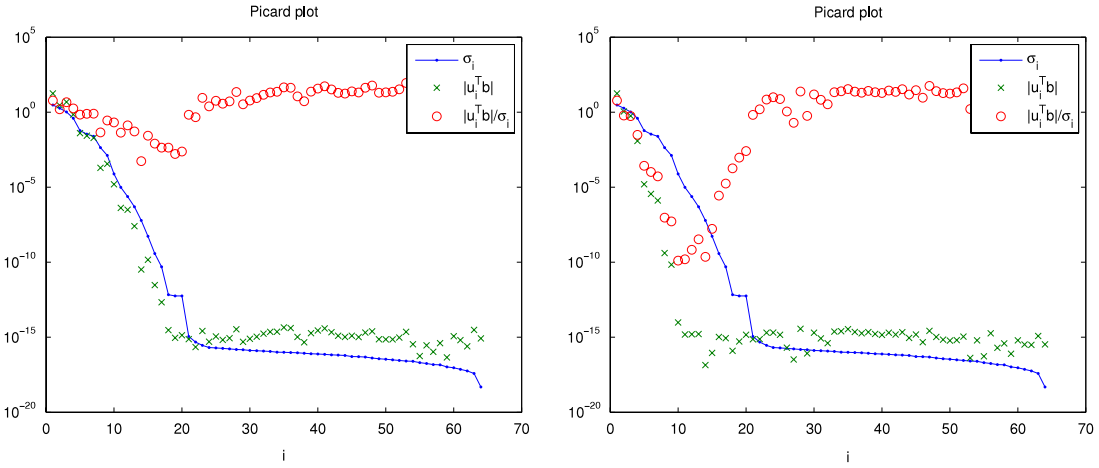

Fig. 7. Picard plots for **shaw** (left) and ***Modify — shaw*** (right).

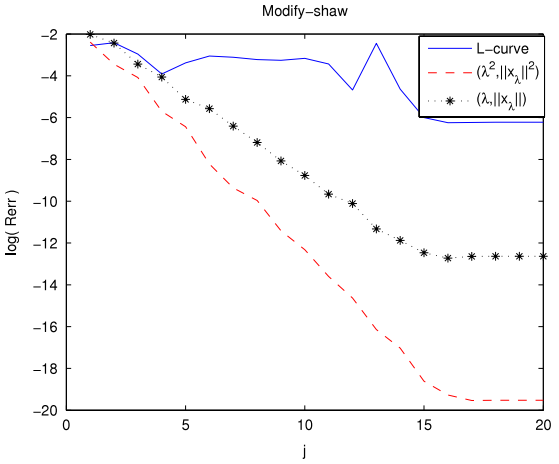

Fig. 8. Relative errors versus j for which 10^{-j} is the level of noise.

of the methods for these systems. From Fig. 9 it is clear that for deriv2 problem the results of $(\lambda^2, \|x_\lambda\|^2)$ are better than those of L-curve from $j = 2$ to $j = 4$ and they have similar validity afterwards. For other three problems it is clear that the performance of our new method is better than that of L-curve. Also it should be mentioned that for baart and foxgood problems the relative error of L-curve tends to increase.

A remark on the large scale applications

In order to use $(\lambda^2, \|x_\lambda\|^2)$ one needs to compute $\|x_\lambda\|$ for many λ's, and this can cause some difficulties in large scale problems. For this reason similar to [7] one can take the upper and lower approximation of $\|x_\lambda\|$, by using the partial Lanczos bi-diagonalization process, to obtain the regularization parameter. In [7] the authors show that by using the partial Lanczos process, it is possible to obtain the upper and lower bounds for $\|x_\lambda\|$ and $\|r_\lambda\|$, such that by increasing the dimension of bi-diagonal factor (increasing the iteration of Lanczos process),these bounds become closer to $\|x_\lambda\|$ and $\|r_\lambda\|$.

In this part we obtained the upper and lower bounds of $\|x_\lambda\|$ by using a similar algorithm as 5.1 of [7], and then used the average of these bounds to find the regularization parameter in the same way as for small scale problems. Similarly, the average of upper and lower bounds of $\|x_\lambda\|$ and $\|r_\lambda\|$ has been used to obtain the corner for the L-ribbon. In the following, we see the results on **phillips**(200) test problem for approximate and exact case.

Fig. 10(a) shows the approximation case, when $l = 9$ where l is the number of iterations in the Lanczos process. Fig. 10(b) shows the results of $(\lambda^2, \|x_\lambda\|^2)$ and the L-curve for $l = 9$, where l is the number of iteration in the Lanczos process. Fig. 10 shows that, similar to the exact case, the new method has a better results in comparison with the L-curve.

5. Conclusion

In this paper we used the curve $(\lambda^2, \|x_\lambda\|^2)$ to obtain a suitable regularization parameter for Tikhonov regularization. This new method has similar results as those of the L-curve, but for problems with a very smooth solution the results of this

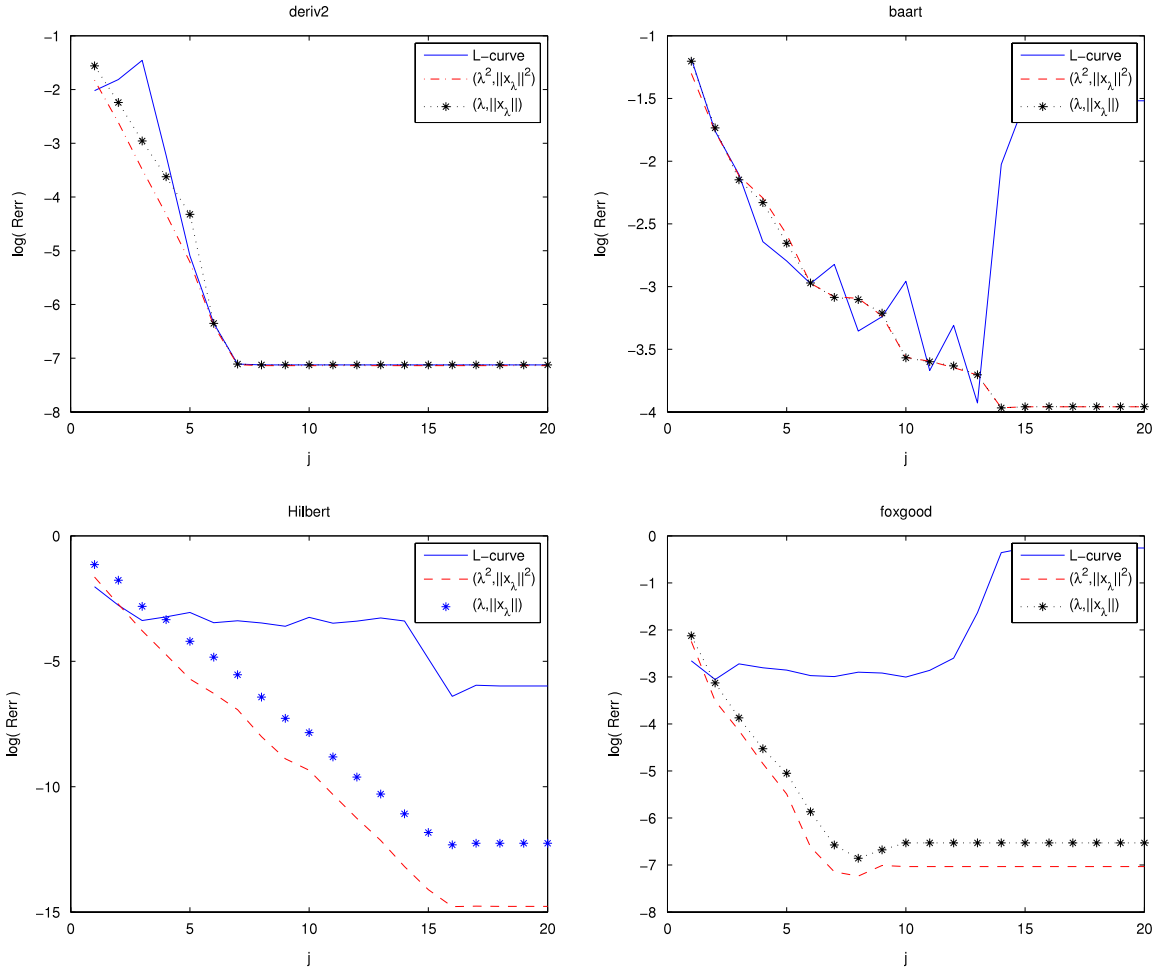

Fig. 9. Convergence results for deriv2, baart, Hilbert and foxgood.

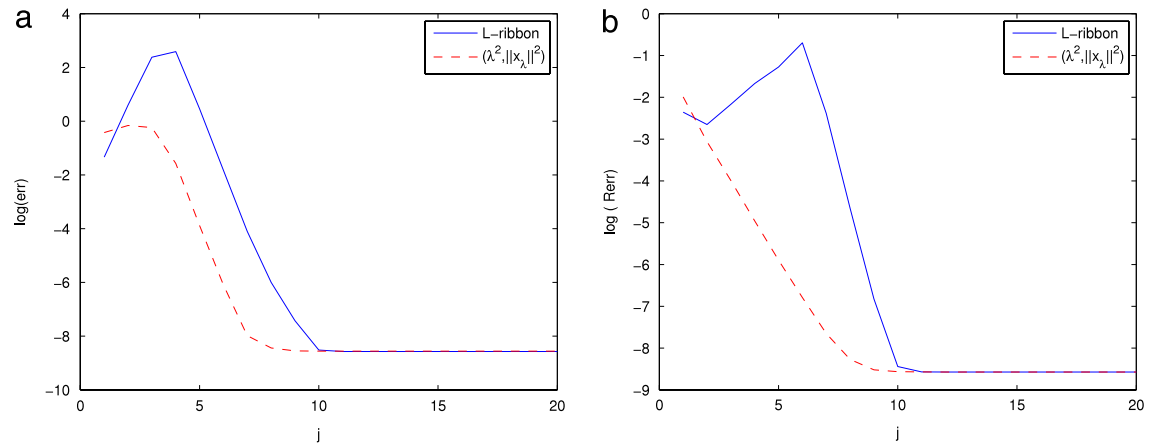

Fig. 10. Large scale for **phillips** (a) $l = 9$ (b) Exact.

new method are better. The reported numerical results confirm this comparison. It is also worth mentioning that for large problems the overall computational costs of this new method is sensibly less than that of the L-curve. As the numerical tests showed, the convergence behavior of the proposed method is in general superior to that of the L-curve method and the error of this new method is monotone decreasing.

References

[1] H.W. Engl, M. Hanke, Andreas Neubauer, Regularization of Inverse Problems, Kluwer, Dordrecht, 1996.
[2] C.W. Groetsch, Inverse problems in methematical sciences, in: Viewing, Wiesbadenm, Germany, 1993.
[3] M. Hanke, Conjugate gradient type methods for ill-posed problems, Pitman Research Notes in Mathematics, Longman, Harlow, UK, 1995.
[4] P.C. Hansen, Rank-Deficient and Discrete Ill-Posed Problems, SIAM, Philadelphia, 1998.
[5] P.H. Hansen, The L-curve and its use in the numerical treatment of inverse problems, invited chapter, in: P. Johnston (Ed.), Computational Inverse Problems in Electrocardiology, WIT Press, Southampton, 2001, pp. 119–142.
[6] A.B. Bakushinskii, Remarks on choosing a regularization parameter using the quasi-optimality and ratio criterion, USSR Comput. Math. Phys. 24 (4) (1984) 181–182.
[7] D. Calvetti, G. Golub, L. Reichel, Estimation of the L-curve via Lanczos bidiagonalization, BIT 39 (1999) 603–619.
[8] D. Calvetti, L. Reichel, A. Shuibi, L-curve and curvature bounds for Tikhonov regularization, Numer. Algorithms 35 (2004) 301–314.
[9] L. Reichel, H. Sadok, A new L-curve for ill-posed problems, J. Comput. Appl. Math. 219 (2008) 493–508.
[10] P.C. Hansen, Analysis of discrete Ill-posed problems by means of the L-curve, SIAM Review 34 (1992) 561–580.
[11] K. Miller, Least squares methods for ill-posed problems with a prescribed bound, SIAM J. Math. Anal. 1 (1970) 52–74.
[12] C.L. Lawson, R.J. Hanson, Solving Least Squares Problems, Prentice-Hall, Englewood Cliffs, NJ, 1974, reprinted by SIAM, Philadelphia, PA, 1995.
[13] M. Hanke, P.C. Hansen, Regularization methods for large-scale problems, Surv. Math. Ind. 3 (1993) 253–315.
[14] M. Hanke, Limitations of the L-curve method in ill-posed problems, BIT 36 (1996) 287–301.
[15] P.C. Hansen, Regularization tools: A MATLAB package for analysis and solution of discrete ill-posed problems, Numer. Algorithms 6 (1994) 1–35. Software is available in Netlib at http://www.netlib.org.
[16] P.C. Hansen, D.P. O'Leary, The use of the L-curve in the regularization of discrete il-posed problems, SIAM J. Sci. Comput. 14 (1993) 1486–1503.