J. Korean Math. Soc. **54** (2017), No. 5, pp. 1557–1571
https://doi.org/10.4134/JKMS.j160566
pISSN: 0304-9914 / eISSN: 2234-3008

EFFICIENT ESTIMATION OF THE REGULARIZATION PARAMETERS VIA L-CURVE METHOD FOR TOTAL LEAST SQUARES PROBLEMS

Geunseop Lee

Abstract. The L-curve method is a parametric plot of interrelation between the residual norm of the least squares problem and the solution norm. However, the L-curve method may be hard to apply to the total least squares problem due to its no closed form solution of the regularized total least squares problems. Thus the sequence of the solution norm under the fixed regularization parameter and its corresponding residual need to be found with an efficient manner. In this paper, we suggest an efficient algorithm to find the sequence of the solutions and its residual in order to plot the L-curve for the total least squares problems. In the numerical experiments, we present that the proposed algorithm successfully and efficiently plots fairly 'L' like shape for some practical regularized total least squares problems.

1. Introduction

We consider the problem of finding an approximate solution to the overdetermined linear system

$$X\mathbf{y} \approx \mathbf{b}, \tag{1.1}$$

with the matrix $X \in \mathbf{R}^{m\times n}$ where $m \leq n$, the observed data $\mathbf{b} \in \mathbf{R}^m$, and the solution $\mathbf{y} \in \mathbf{R}^n$. One possible solution to (1.1) is to find the least squares solution $\mathbf{y}_{LS}$ from the standard Gauss-Markov linear model, which is given by

$$\mathbf{y}_{LS} = \mathbf{arg\ min_y} \|X\mathbf{y} - \mathbf{b}\|_2. \tag{1.2}$$

Unfortunately in many cases, the data matrix X arises from the discretization of an ill-posed problem, therefore the least squares solution $\mathbf{y}_{LS}$ of (1.2) is noise-contaminated, thereby meaningless. In particular, if the matrix X is ill-conditioned, then the smallest singular value of X gradually decays to zero, thus even a small perturbation in the data leads to large perturbation at the solution $\mathbf{y}_{LS}$. Therefore, the solution becomes very sensitive to noise.

Received August 28, 2016; Accepted December 26, 2016.
2010 *Mathematics Subject Classification.* Primary 65F10, 65F22.
Key words and phrases. total least squares method, L-curve.

To obtain a meaningful and stable solution to (1.1) despite noise, the solution has to be obtained with a prior information. It follows to consider a regularization technique that enables to filter out such noise and generate smoothness to the solution. One approach is Tikhonov regularization [20], which picks a linear operator L so that the solution $\mathbf{y}$ is constrained to satisfy the condition $\|L\mathbf{y}\|_2 \leq \delta$. Here L is, in general, an identity, a first-order derivative, or a second-order derivative operator. The parameter δ, which controls the size and smoothness of the solution, comes from knowledge of the underlying physical model. Therefore, the least squares solution with Tikhonov regularization is defined as

$$\mathbf{y}_{LS}(\delta) = \mathbf{arg\ min}_{\|L\mathbf{y}\|_2 \leq \delta} \|X\mathbf{y} - \mathbf{b}\|_2. \tag{1.3}$$

There are many methods that solve (1.3) either directly [3, 6] or with iteration-based approaches [16, 19]. One of the most well-known analysis tools of the regularized least squares problems in (1.3), or even can be used to find the optimal regularization parameter δ, is L-curve, which plots two quantities

$$(\|X\mathbf{y}(\delta) - \mathbf{b}\|_2, \|L\mathbf{y}(\delta)\|_2)$$

with parameterized by δ. Specifically, L-curve is kind a 'trade-off' curve between two quantities in which its corner of the curve indicates the optimized solution of (1.3).

However, the least squares model has a limitation. If error exists in both the right hand side (observed data $\mathbf{b}$) and the data matrix X, it is more appropriate to adopt an errors-in-variables model called the Total Least Squares (TLS) method than the ordinary least squares model (1.2). The TLS approach finds an error matrix E_{TLS} and an error vector $\mathbf{r}$ subject to satisfy the following minimization problem

$$\begin{pmatrix} E_{TLS} & \mathbf{r}_{TLS} \end{pmatrix} = \mathbf{arg\ min} \| \begin{pmatrix} E & \mathbf{r} \end{pmatrix} \|_F \tag{1.4}$$

so that for some $\mathbf{y} \in \mathbf{R}^n$,

$$(X + E)\mathbf{y} = \mathbf{b} + \mathbf{r}.$$

Still a regularization technique is required since the solution $\mathbf{y}_{TLS}$ from (1.4) is even more sensitive to noise than the solution $\mathbf{y}_{LS}$ from (1.2). Thus a model of the TLS problem with Tikhonov regularization is given by

$$(\mathbf{y}_{TLS}, E_{TLS}(\delta)) = \mathbf{arg\ min}_{\|L\mathbf{y}\|_2 \leq \delta} \|(X + E)\mathbf{y} - \mathbf{b}\|_2^2 + \|E\|_F^2. \tag{1.5}$$

In this paper, we investigate the efficient method of plotting L-curve for the TLS problems with Tikhonov Regularization in (1.5). The rest of this paper is organized as follows: in Section 2, we briefly discuss some issues to plot L-curve for the TLS problems, in Section 3 we suggest the algorithm of efficient quantization of two quantities for L-curve, in Section 4, we shows some examples of L-curve plots in practical TLS problems, and in Section 5 we give a conclusion.

2. L-curve method

The L-curve for regularized least squares (RLS) problems, developed in [12, 14] and further analyzed in [7, 9, 10], is a parametric plot of size of the residual $\|X\mathbf{y} - \mathbf{b}\|_2^2$ versus size of the regularized solution $\|L\mathbf{y}\|_2$. The name 'L-curve' comes from its distinct corner and the shape of the curve which is appeared like 'L' letter when the error of the linear system is not too large and the discrete Picard condition [9, p. 82] holds. It is primarily used for choosing the optimal regularization parameter which is appeared in the corner of the L-curve plot in usual. Qualitatively, it is reasonable since the solution corresponding to the parameter on the corner yields both a small residual norm and the small regularized solution norm. Also, because of its properties, the L-curve is often used as a graphical analysis tool for the discrete ill-posed problems. In order to pick the corner of L-curve, one possibility is to rotate the curve so that the corner of L-curve might have minimum rotation [17]. Another idea suggests to determine the maximal curvature of the L-curve. Specifically, it can be fitted to a cubic spline curve [10].

The L-curve analysis is primarily for the ordinary least squares problem. Thus some small modifications are required to be adapted to TLS model. One possible modification is to use $\frac{\|X\mathbf{y}-\mathbf{b}\|_2^2}{1+\|\mathbf{y}\|_2^2}$ instead of $\|X\mathbf{y} - \mathbf{b}\|_2^2$. However, the L-curve cannot guarantee to generate a plot of the 'L' like shape in TLS model, since the discrete Picard condition is no longer applicable. Fortunately in many cases, 'L' shape still appears, thus the L-curve method is effective in analyzing the TLS problems [11].

3. L-curve method for the RTLS problem

As described in Section 2, the L-curve method is a useful tool for the graphical analysis of the RTLS problems by investigating the interrelation between the residual norm $\|X\mathbf{y} - \mathbf{b}\|_2^2$ and the solution norm $\|L\mathbf{y}\|_2$. However, there are two issues in the L-curve method with the RTLS problem. First, as we stated in §2, the L-curve method with the RTLS problem is not guaranteed to generate an 'L' letter like shape since it is hard to apply the discrete Picard condition directly to the RTLS problems. Second, since there is no direct or close form solution in the RTLS problems, the continuous plot of L-curve is prohibited. It leads us to require a quantization or choosing a sequence of parameters for plotting the discrete L-curve. Thus finding the sequence of the solution $\mathbf{y}_{TLS}$ under the fixed regularization parameter δ and corresponding residual μ needs a several trials of the RTLS algorithm. Regarding the first issue, a fully developed necessary or sufficient conditions are not yet known, however fortunately, many L-curve shapes of the RTLS problems still appear to have an 'L' letter, thus the L-curve method is viable to analyze the RTLS problem typically. In this section, thus we focus on the second issue of the L-curve method, constructing the solution $\mathbf{y}_{TLS}$ with a fixed parameter δ in an efficient manner. Here we use the sequence of the parameters $\lambda_i, 1 \le i \le n$, where n is the number of L-curve points, instead of δ based on the equations shown in [4]

$$\gamma = \frac{\mathbf{b}^T\tilde{\mathbf{r}} - \|\tilde{\mathbf{r}}\|_2^2}{\delta^2},$$

where $\tilde{\mathbf{r}}$ is obtained from

$$\tilde{\mathbf{r}}(1 + \|\mathbf{y}\|_2^2) = \mathbf{b} - X\mathbf{y}$$

and by the definition (3.4), $\lambda = \gamma(1 + \|\mathbf{y}\|_2^2)$.

Therefore, we plot the L-curve such that

$$\left(\log \frac{\|X\mathbf{y}(\lambda) - \mathbf{b}\|_2}{1 + \|\mathbf{y}\|_2^2}, \|L\mathbf{y}(\lambda)\|_2^2\right), \quad 0 \le \lambda < \infty. \tag{3.1}$$

As we stated earlier, finding the solution $\mathbf{y}(\lambda_i)$ with fixed λ_i in the sequence $[\lambda_1, \lambda_2, \ldots, \lambda_n]$ and corresponding RTLS residual μ is required to plot the L-curve in (3.1). To find the proper value of μ with given λ_i, we use the two parameter normal equation to which the solution of (1.5) is equivalent [4], where the two parameter normal equation is given by

$$(X^TX + \lambda_i L^TL - \mu I)\mathbf{y}(\lambda_i, \mu) = X^T\mathbf{b}, \tag{3.2}$$

$$\mu^* = \frac{\|X\mathbf{y}(\lambda^*, \mu^*) - \mathbf{b}\|_2^2}{1 + \|\mathbf{y}(\lambda^*, \mu^*)\|_2^2}, \tag{3.3}$$

$$\lambda^* = \gamma(1 + \|\mathbf{y}(\lambda^*, \mu^*)^2\|_2^2), \tag{3.4}$$

where the optimal parameter pair (λ^*, μ^*) satisfies $\mathbf{y}_{TLS} = \mathbf{y}(\lambda^*, \mu^*)$. The equation (3.2) was extended by Guo and Renaut [5] by constructing the augmented linear system such that

$$\begin{pmatrix} X^TX + \lambda_i L^TL & X^T\mathbf{b} \\ \mathbf{b}^TX & \mathbf{b}^T\mathbf{b} - \lambda_i\delta^2 \end{pmatrix} \begin{pmatrix} \mathbf{y}(\lambda_i, \mu) \\ -1 \end{pmatrix} = \mu \begin{pmatrix} \mathbf{y}(\lambda_i, \mu) \\ -1 \end{pmatrix}. \tag{3.5}$$

Theorem 3.1. *The solution $\mathbf{y}_{TLS}$ of the regularized problem* (1.5) *for which the constraint is active, satisfies the augmented eigenvalue problem*

$$A(\lambda^*, 0) \begin{pmatrix} \mathbf{y}(\lambda^*, \mu^*) \\ -1 \end{pmatrix} = \mu^* \begin{pmatrix} \mathbf{y}(\lambda^*, \mu^*) \\ -1 \end{pmatrix}, \tag{3.6}$$

where $A(\lambda^*, 0) = \begin{pmatrix} X^TX+\lambda_i L^TL & X^T\mathbf{b} \\ \mathbf{b}^TX & \mathbf{b}^T\mathbf{b}-\lambda_i\delta^2 \end{pmatrix}$ with optimal value λ^*. Conversely, if $\begin{pmatrix} \mathbf{y}(\lambda^*,\mu^*) \\ -1 \end{pmatrix}$ and μ^* is an eigenpair of $A(\lambda^*, 0)$, then $\mathbf{y}(\lambda^*.\mu^*) = \mathbf{y}_{TLS}$ and $\mu^* = f(\mathbf{y}_{TLS})$.

Guo and Renaut also proved that the smallest eigenvalue of $A(\lambda^*, 0)$ is simple under some assumptions ($X^T\mathbf{b} \ne 0$ and $\mathcal{N}(X) \cup \mathcal{N}(L) = 0$, where $\mathcal{N}(M)$ indicates the null space of the matrix M). Additionally, we assume that μ, the smallest eigenvalue of $A(\lambda, 0)$, is not an eigenvalue of $M(\lambda, 0)$. We justify that assumption in the following theorem.

Theorem 3.2. *Let* $X \in \mathbf{R}^{m\times n}$ *be such that* $\mathcal{N}(X) \cap \mathcal{N}(L) = \{0\}$, *where* $\mathcal{N}(M)$ *indicates the null space of the matrix* M. *Suppose that the smallest eigenvalue* τ *of* $M(\lambda, 0) = X^T X + \lambda L^T L$ *has the eigenspace spanned by the left orthogonal matrix* $Z(\lambda) \in \mathbf{R}^{n\times k}$. *If*

$$Z(\lambda)^T X^T \mathbf{b} \neq 0, \tag{3.7}$$

then μ, *the smallest eigenvalue of* $A(\lambda, 0)$ *satisfies* $\mu < \tau$ *and* $M(\lambda, \mu)$ *is positive definite.*

Proof. Let $V(\lambda)$ be an eigenvector matrix for $M(\lambda, 0)$ such that

$$V(\lambda) = \begin{pmatrix} \overset{j}{\hat{V}(\lambda)} & \overset{k}{Z(\lambda)} \end{pmatrix}.$$

Since $Z(\lambda)$ spans a subspace, we can choose it so that

$$Z(\lambda)^T X^T \mathbf{b} = \beta \mathbf{e}_k,$$

where $\beta > 0$. Thus $A(\lambda, 0)$ may be written

$$A(\lambda, 0) = \begin{pmatrix} V(\lambda) & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} D_1(\lambda) & 0 & \mathbf{z}_1 \\ 0 & \tau I_k & \beta \mathbf{e}_k \\ \mathbf{z}_1^T & \beta \mathbf{e}_k^T & \alpha(\lambda) \end{pmatrix} \begin{pmatrix} V(\lambda)^T & 0 \\ 0 & 1 \end{pmatrix},$$

where $\alpha(\lambda) = \mathbf{b}^T\mathbf{b} - \lambda\delta^2$. The smallest eigenvalue of $A(\lambda, 0)$, μ, is the smallest eigenvalue of

$$\begin{pmatrix} D_1(\lambda) & 0 & \mathbf{z}_1 \\ 0 & \tau & \beta \\ \mathbf{z}_1^T & \beta & \alpha(\lambda) \end{pmatrix}.$$

From Boley and Golub [2], that eigenvalue μ must be smaller than τ. Thus $M(\lambda, \mu) = M(\lambda, 0) - \mu I$ is positive definite and nonsingular. □

Theorem 3.2 establishes that, under reasonable circumstances, in a neighborhood of (λ^*, μ^*), $M(\lambda, \mu)$ is nonsingular and symmetric, positive definite, thus we may factor $A(\lambda, \mu)$ into

$$A(\lambda, \mu) = \begin{pmatrix} I & 0 \\ \mathbf{b}^T X M(\lambda, \mu)^{-1} & 1 \end{pmatrix} \begin{pmatrix} M(\lambda, \mu) & X^T\mathbf{b} \\ 0 & f_1(\lambda, \mu) \end{pmatrix}, \tag{3.8}$$

where

$$\begin{aligned} f(\lambda, \mu) &= \mathbf{b}^T\mathbf{b} - \mu - \lambda\delta^2 - \mathbf{b}^T X M(\lambda, \mu)^{-1} X^T \mathbf{b} \\ &= \mathbf{b}^T\mathbf{b} - \mu - \lambda\delta^2 - \mathbf{b}^T X \mathbf{y}(\lambda, \mu). \end{aligned} \tag{3.9}$$

Since $A(\lambda, \mu)$ is singular, and the first matrix in (3.8) is nonsingular, $f(\lambda, \mu) = 0$.

The first derivative of (3.9) with respect to μ is

$$\frac{\partial f(\lambda, \mu)}{\partial \mu} = -1 - \mathbf{b}^T X \mathbf{y}_\mu,$$

Algorithm 1 Newton based RTLS Algorithm for L-curve

Input $X, L, \mathbf{b}, \lambda, tol$
1. Choose initial guesses μ_0 and $k = 0$
2. Solve $M(\lambda, \mu_0)\mathbf{y}_0 = X^T\mathbf{b}$
3. Solve $M(\lambda, \mu_0)\mathbf{y}_\mu = \mathbf{y}_0$
4. Set $\delta = \|L\mathbf{y}_0\|_2$
5. While $\epsilon \geq tol$
   Set $\phi(\lambda, \mu_k) = \mathbf{b}^T\mathbf{b} - \mu_k - \lambda\|L\mathbf{y}_k\|_2^2 - \mathbf{b}^T X\mathbf{y}_k$
   Compute the first derivative, $J(\lambda, \mu_k) = -1 - \|\mathbf{y}_k\|_2^2 - 2\lambda\mathbf{y}^T(\lambda, \mu)L^T L\mathbf{y}_\mu$.
   $\mu_{k+1} = \mu_k + f(\lambda, \mu_k)/J(\lambda, \mu_k)$,
   Solve $M(\lambda, \mu_{k+1})\mathbf{y}_{k+1} = X^T\mathbf{b}$
   Solve $M(\lambda, \mu_{k+1})\mathbf{y}_\mu = \mathbf{y}_{k+1}$
   $\epsilon = \left\|\mathbf{y}_{k+1} - \mathbf{y}_k\right\|_2 / \|\mathbf{y}_k\|_2$
   $k = k + 1$
6. return $\mathbf{y}_k$ and μ_k

where

$$M(\lambda, \mu)\mathbf{y}_\mu = \mathbf{y}(\lambda, \mu). \tag{3.10}$$

Since $\mathbf{y}^T M(\lambda, \mu)\mathbf{y}_\mu = \mathbf{b}^T X\mathbf{y}_\mu = \|\mathbf{y}(\lambda, \mu)\|_2^2$, the first derivative of (3.9) is further simplified as

$$\frac{\partial f(\lambda, \mu)}{\partial \mu} = -1 - \|\mathbf{y}(\lambda, \mu)\|_2^2. \tag{3.11}$$

Clearly, the first derivative $-1 - \|\mathbf{y}(\lambda, \mu)\|_2^2 < 0$, which implies that $f(\lambda, \mu)$ has an unique solution.

We also need for the constraint $\|L\mathbf{y}\|_2 = \delta$, thus it requires to satisfy

$$g(\lambda, \mu) = \delta^2 - \|L\mathbf{y}\|_2 = 0.$$

Thus our modification to $f(\lambda, \mu)$ for the simple computation is as follows:

$$\begin{aligned}\phi(\lambda, \mu) &= f(\lambda, \mu) + \lambda g(\lambda, \mu) \\ &= \mathbf{b}^T\mathbf{b} - \mu - \lambda\|L\mathbf{y}(\lambda, \mu)\|_2^2 - \mathbf{b}^T X\mathbf{y}(\lambda, \mu) \\ &= 0,\end{aligned} \tag{3.12}$$

and its first derivative respect to μ forms

$$\frac{\partial \phi(\lambda, \mu)}{\partial \mu} = -1 - \|\mathbf{y}(\lambda, \mu)\|_2^2 - 2\lambda\mathbf{y}^T(\lambda, \mu)L^T L\mathbf{y}_\mu. \tag{3.13}$$

Using (3.12), (3.13) and (3.10), we produce the Newton method which is summarized in Algorithm 1. Here the variable μ is updated with given λ, such that

$$\mu_{k+1} = \mu_k + f(\mu_k)/\frac{\partial f(\mu_k)}{\partial \mu}. \tag{3.14}$$

Remark 3.3. A Newton iteration requires to have an initial guess close to the solution for its convergence and less iteration steps. Since we make a call Algorithm 1 with a sequence of the constants λ_i, $0 \leq i \leq n$, we can save some iterations by setting the initial value μ_0 in step 1 from the previous result of Algorithm 1.

Remark 3.4. The computational bottleneck of Algorithm 1 may occur in steps 2, 3, and 5 which require to solve the linear system. Since the matrix L has a structure in practice, if the matrix X has a structure which has a special matrix-vector multiplication or X is a sparse matrix, we can save some computational resources when computing the solution of the inverse problems. However, if the matrix X is an ordinary dense-matrix, computational speed may decrease as the size of X increases. To save the execution time, one idea is to compute the eigendecomposition of $(X^TX + \lambda_i L^TL)$ before Algorithm 1, such that

$$X^TX + \lambda_i L^TL = V\Sigma^2V^T,$$

where V is an orthogonal matrix, and Σ is a diagonal matrix. Since λ_i maintains a constant in Algorithm 1, computation of the inverse problem in step 2 can be easily done with

$$\begin{aligned} \mathbf{y}_0 &= M(\lambda, \mu_0)^{-1}X^T\mathbf{b} \\ &= V(\Sigma^2 - \mu_0 I)^{-1}V^TX^T\mathbf{b}. \end{aligned} \tag{3.15}$$

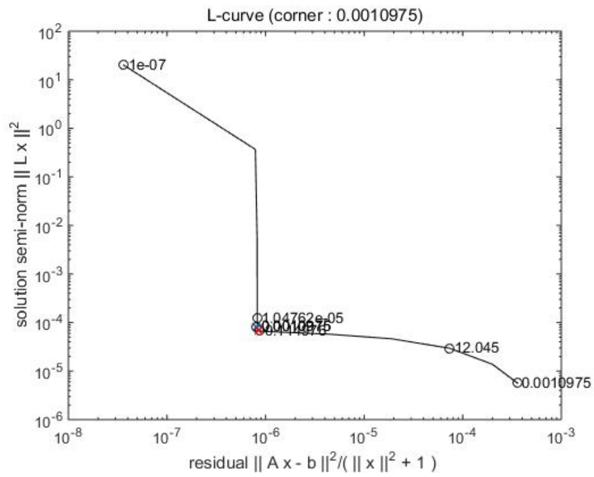
Likewise, we can compute $\mathbf{y}_\mu$ and $\mathbf{y}_i$ with V and Σ instead of computing the dense-matrix inverse problems. Therefore, the overall computational speed of Algorithm 1 takes $O(mn)$ flops.

4. Experimental results

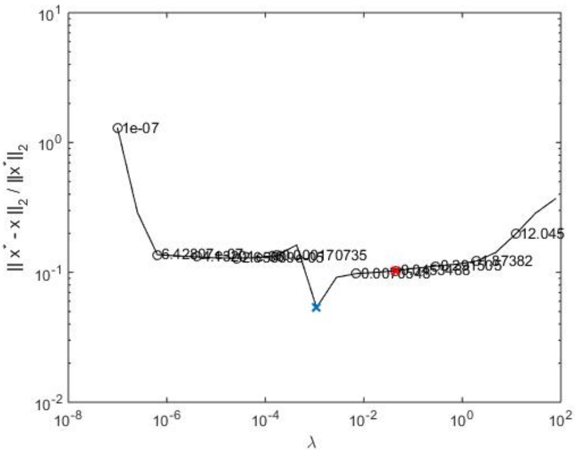
We now present the results of the experiments to demonstrate the applicability of the proposed algorithm to plot the L-curve of some practical applications. One of the most well-known RTLS applications arises in the discretization of a Fredholm integral equation. Thus we use a Fredholm integral equation of the first kind, which is defined as

$$\int_b^a k(s,t)f(t)dt = g(s), \quad s \in [a,b],$$

where $k(s,t)$ indicates the kernel function, $f(t)$ represents the solution, and $g(s)$ is the observed data, respectively. To generate the data matrix $X \in \mathbf{R}^{n\times n}$ as the discretized kernel function $k(s,t)$ we use *the regularization tools* from Hansen's regularization package functions [8]. The examples chosen from [8] are generated by BAART, DERIVE2, GRAVITY, HEAT, PHILLIPS, and SHAW functions. All lead to an ill-conditioned to X. To simulate the TLS model, we generate a noise-contaminated data matrix X_{noise} and a observed vector $\mathbf{b}$



(a) BAART

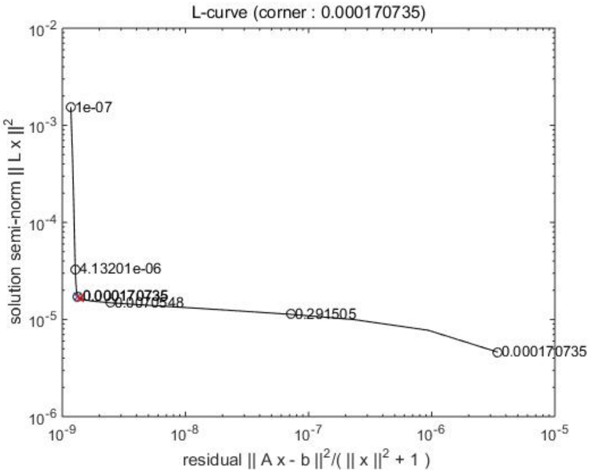


(b) Relative errors

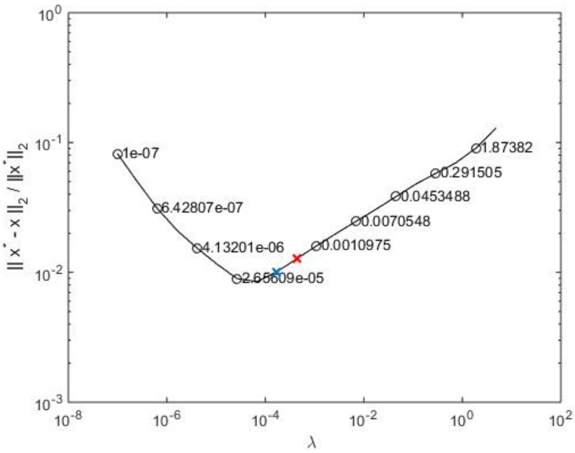
FIGURE 1. Plots of the L-curve when $X \in \mathbf{R}^{400\times 400}$ is generated from BAART function. Noise with $\sigma = 0.01$ is added.

with Gaussian white nose E and $\mathbf{r}$, such that

$$\begin{aligned} \tilde{X} &= X + \sigma \|E\|_F^{-1} E, \\ \mathbf{b} &= \tilde{X}\mathbf{y}_{true} + \sigma \|\mathbf{r}\|_2^{-1}\mathbf{r}. \end{aligned} \tag{4.1}$$



(a) DERIV2

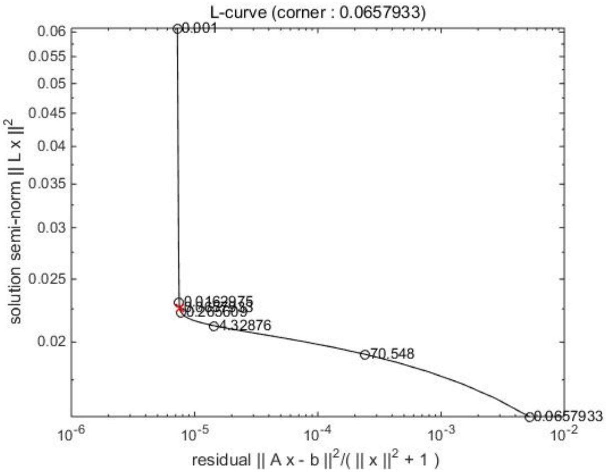


(b) Relative errors

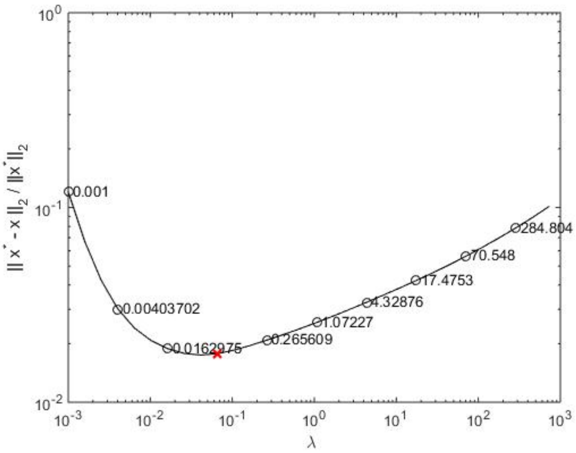
FIGURE 2. Plots of the L-curve when $X \in \mathbf{R}^{400\times 400}$ is generated from DERIV2 function. Noise with $\sigma = 0.01$ is added.

We have used an approximate first-order derivative operator for L

$$L = \begin{pmatrix} 1 & -1 & & & \\ & 1 & -1 & & \\ & & \ddots & \ddots & \\ & & & 1 & -1 \end{pmatrix}.$$



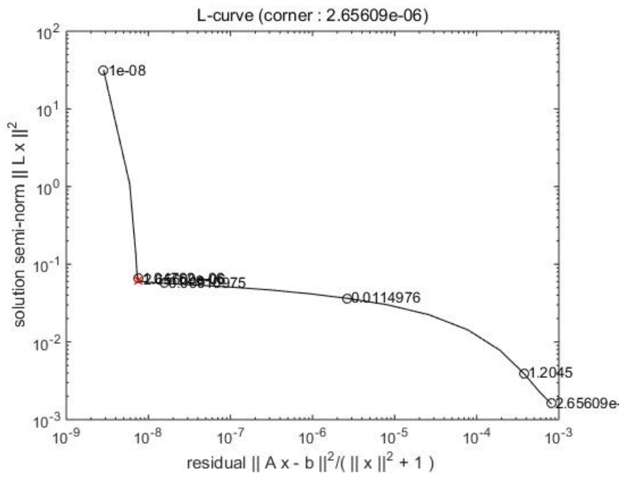
(a) GRAVITY



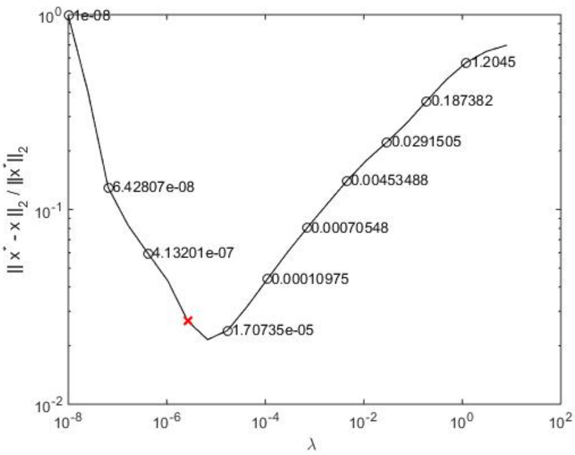
(b) Relative errors

FIGURE 3. Plots of the L-curve when $X \in \mathbf{R}^{400\times 400}$ is generated from GRAVITY function. Noise with $\sigma = 0.01$ is added.

Figures 1-6 show the examples of the L-curve plot of which the data matrix $X \in \mathbf{R}^{400\times 400}$ is generated from the functions BAART, DERIV2, GRAVITY, HEAT, PHILLIPS, and SHAW. We quantize $n = 100$ log-scale bins with the range of λ_i given in Table 1. We set the initial value μ_0 for Algorithm 1 as 1.0e-8. The tolerance value *tol* in Algorithm 1 is 1.0e-8. We also used the



(a) HEAT

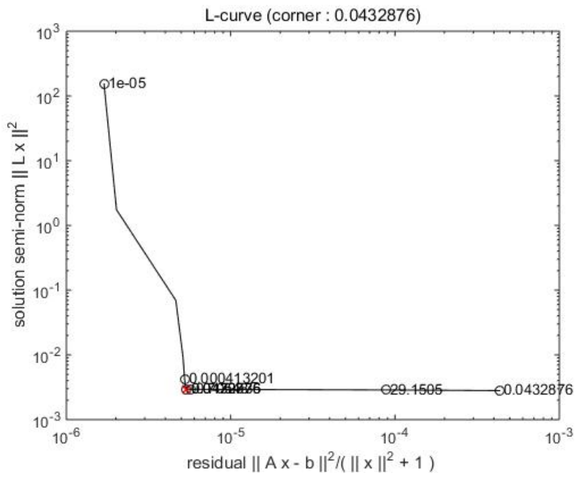


(b) Relative errors

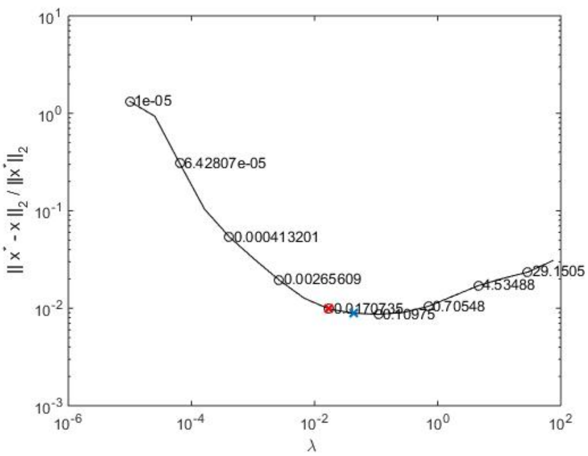
FIGURE 4. Plots of the L-curve when $X \in \mathbf{R}^{400\times 400}$ is generated from HEAT function. Noise with $\sigma = 0.01$ is added.

algorithms from [17] and [10] to detect the corner, which can be considered close to an optimal parameter λ^*. The red and blue crosses represent the corners computed from [17] and [10], respectively. All graphs have fairly 'L'-like shape with clear corners. We also plot the relative errors corresponding to λ_i, where the relative error r is defined as

$$r = \|\mathbf{y}(\lambda^*, \mu^*) - \mathbf{y}(\lambda_i, \mu)\|_2 / \|\mathbf{y}(\lambda^*, \mu^*)\|_2.$$



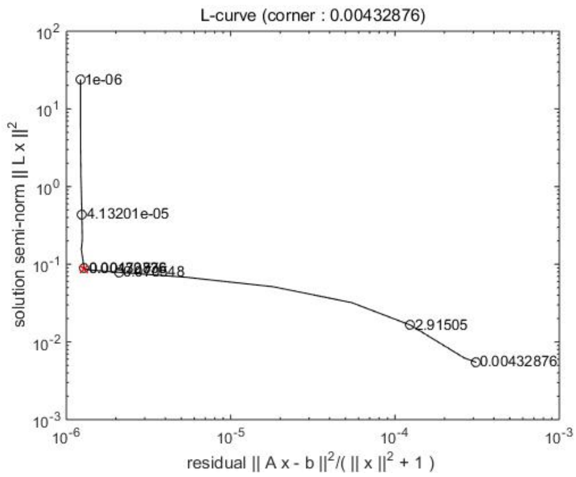
(a) PHILLIPS



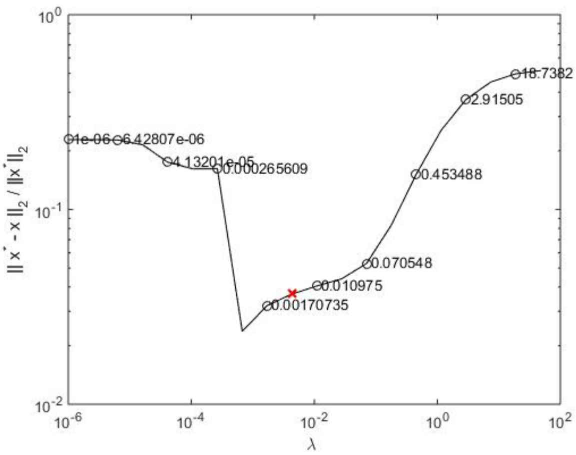
(b) Relative errors

FIGURE 5. Plots of the L-curve when $X \in \mathbf{R}^{400\times 400}$ is generated from PHILLIPS function. Noise with $\sigma = 0.01$ is added.

Similar to the L-curve plot cases, the red and blue crosses represent the relative errors when the algorithms from [17] and [10] detect the corner points from the L-curves. The plots 1(b)-6(b) indicate that the regularization parameters λ from the corner points produce the solution close to the optimal solutions.



(a) SHAW



(b) Relative errors

FIGURE 6. Plots of the L-curve when $X \in \mathbf{R}^{400\times 400}$ is generated from SHAW function. Noise with $\sigma = 0.01$ is added.

5. Conclusion

The L-curve is a log-log plot which represents the correlation between a regularized solution and the corresponding residual norm. It aims to provide a graphical tool to display the trade-off relations between the size of a regularized solution and its fit to the given data. Another purpose of this plot is to estimate a good regularization parameter which is typically appeared in the corner at

Table 1. Range of the regularization parameter λ for each TLS problems

	λ_{min}	λ_{max}
BAART	1.0e-7	1.0e+2
DERIV2	1.0e-6	1.0e+1
GRAVITY	1.0e-3	1.0e+3
HEAT	1.0e-8	1.0e+1
PHILLIPS	1.0e-5	1.0e+2
SHAW	1.0e-6	1.0e+2

the 'L' shape plot. To apply this L-curve method to the total least squares problems, the sequence of the solution norm under the fixed regularization parameter and its corresponding residual norm must be found efficiently. In this paper, we showed an efficient algorithm to find the sequence of the solutions and its residual so that those can be used to plot the L-curve with total least squares problems. In the numerical experiments, we showed that the algorithm plotted fairly 'L' like shapes and its corner point produced the solution close to the optimal solution.

References

[1] G. Arfken, *Mathematical Methods for Physicists*, Third edition, Academic Press, Orlando, FL, USA, 1985.

[2] D. Boley and G. H. Golub, *Inverse eigenvalue problems for band matrices*, Technical Report STAN-CS-77-623, Department of Computer Science, Stanford University, Stanford. CA., 1977.

[3] L. Eldén, *Algorithms for the regularization of ill-conditioned least squares problems*, BIT **17** (1977), no. 2, 134–145.

[4] G. H. Golub, P. C. Hansen, and D. P. O'Leary, *Tikhonov regularization and total least squares*, SIAM J. Matrix Anal. Appl. **21** (1999), no. 1, 185–194.

[5] H. Guo and R. A. Renaut, *A regularized total least squares algorithm*, Total least squares and errors-in-variables modeling (Leuven, 2001), 57–66, Kluwer Acad. Publ., Dordrecht, 2002.

[6] P. C. Hansen, *Regularization, GSVD and truncated GSVD*, BIT **29** (1989), no. 3, 491–504.

[7] ______, *Analysis of discrete ill-posed problems by means of the L-curve*, SIAM Rev. **34** (1992), no. 4, 561–580.

[8] ______, *Regularization Tools: A Matlab package for analysis and solution of discrete ill-posed problems*, Numer. Algorithms **6** (1994), no. 1-2, 1–35.

[9] ______, *Rank-Deficient and Discrete Ill-Posed Problems. Numerical Aspects of Linear Inversion*, SIAM, Philadelphia, 1998.

[10] P. C. Hansen and D. P. O'Leary, *The use of the L-curve in the regularization of discrete ill-posed problems*, SIAM J. Sci. Comput. **14** (1993), no. 6, 1487–1503.

[11] J. Lampe, *Solving Regularized Total Least Squares Problems Based on Eigenproblems*, Ph.D. Thesis, Hamburg University of Technology, Institute of Numerical Simulation, 2010.

[12] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*, Solving Least Squares Problems, Classics in Applied Mathematics, Society for Industrial and Applied Mathematics, 1987.

[13] G. Lee, H. Fu, and J. L. Barlow, *Fast high-resolution image reconstruction using Tikhonov regularization based total least squares*, SIAM J. Sci. Comput. **35** (2013), no. 1, B275–B290.

[14] K. Miller, *Least squares methods for ill-posed problems with a prescribed bound*, SIAM J. Math. Anal. **1** (1970), no. 1, 52–74.

[15] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer-Verlag, Berlin, 2000.

[16] C. C. Paige and M. A. Saunders, *LSQR: An algorithm for sparse linear equations and sparse least squares*, ACM Trans. Math. Software **8** (1982), no. 1, 43–71.

[17] T. Reginska, *A regularization parameter in discrete ill-posed problems*, SIAM J. Sci. Comput. **17** (1996), no. 3, 740–749.

[18] R. Renaut and H. Guo, *Efficient algorithms for solution of regularized total least squares*, SIAM J. Matrix Anal. Appl. **26** (2004), no. 2, 457–476.

[19] J. D. Riley, *Solving systems of linear equations with a positive definite symmetric but possibly ill-conditioned matrix*, Math. Tables Aids Comput. **9** (1955), 96–101.

[20] A. N. Tikhonov and V. Y. Arsenin, *Solution of Ill-Posed Problems*, Winston, Washington D.C., 1977.

Geunseop Lee
Test and Package Center
Samsung Electronics Co. Ltd Asan 31487, Korea
E-mail address: geunseoplee.la@gmail.com