Non-convergence of the L-Curve Regularization Parameter Selection Method

C. R. Vogel

Department of Mathematical Sciences, Montana State University, Bozeman MT 59717-0240, USA, e-mail: vogel@math.montana.edu. Research was supported in part by the NSF under Grant DMS-9303222.

Abstract. The L-curve method was developed for the selection of regularization parameters in the solution of discrete systems obtained from ill-posed problems. An analysis of this method is given for selecting a parameter for Tikhonov regularization. This analysis, which is carried out in a semi-discrete, semi-stochastic setting, shows that the L-curve approach yields regularized solutions which fail to converge for a certain class of problems. A numerical example is also presented which indicates that this lack of convergence can arise in practical applications.

1. Introduction

Of vital importance in the implementation of regularization methods for ill-posed problems is an appropriate choice of the regularization parameter. Recently a parameter selection technique known as the L-curve method has gained attention (see [7, 8, 9]). We will consider the implementation of this method for ill-posed linear operator equations

$$Af = z \tag{1.1}$$

in a Hilbert space setting. An example of such an operator is the Fredholm first kind integral,

$$[Af](x) = \int_\Omega a(x,y)f(y)\,dy, \quad x \in \Omega, \tag{1.2}$$

which is a compact operator on $L^2(\Omega)$, provided Ω is a bounded region in R^m and the kernel $a \in L^2(\Omega \times \Omega)$. Fredholm first kind integral equations occur in a number of important applications (see [4] and the references therein). Suppose a regularization method is applied to (1.1), let α denote the regularization parameter, and let f_α denote the regularized solution. One plots, on a log-log scale, the parameterized curve $(||Af_\alpha - z||, ||f_\alpha||)$ for a range of values of α. This curve typically has a characteristic L-shape. The L-curve criterion for the selection of the regularization parameter is to

pick α corresponding to the "corner" of this curve. Although the L-curve method has been formulated in a continuous setting (see [13, 6]), the method was originally applied to so-called "discrete ill-posed problems". By this we mean the highly ill-conditioned systems arising from the discretization of a (continuous) ill-posed problem.

Engl and Grever [1] and Hanke [6] have recently carried out analyses of the L-curve method in a continuous, deterministic setting. These papers make use of tools developed by researchers in the inverse problems community (see [3]). The analysis in this paper uses techniques developed by Wahba and others (see [15] and [16] and the references therein) in the statistics community. The primary difference between the two approaches is in the treatment of error in the data. With the first approach, one assumes (continuous) data

$$z = Af^* + \eta, \tag{1.3}$$

where A is an operator between Hilbert spaces $\mathcal{H}$ and $\mathcal{Z}$. $f^* \in \mathcal{H}$ is the desired true solution. The error η is assumed to lie in $\mathcal{Z}$. An important parameter is an upper bound for the error, say $||\eta|| \leq \delta$. Let $f_{\alpha,\delta}$ denote the approximate solution to (1.3) obtained when a regularization method is applied with parameter α. In this continuous, deterministic setting, a regularization parameter selection method is said to be *convergent* if it yields a parameter $\alpha(\delta)$ for which

$$||f_{\alpha(\delta),\delta} - f^*|| \to 0 \quad \text{as} \quad \delta \to 0. \tag{1.4}$$

On the other hand, the statistical approach of Wahba et al makes use of a semi-discrete, semi-stochastic data model

$$z_n = A_n f^* + \eta_n. \tag{1.5}$$

In this case, A_n is an operator from the Hilbert space $\mathcal{H}$ into R^n. For example, if the kernel in (1.2) is smooth, applying "moment discretization" yields

$$[A_n f]_i = \int_\Omega a(x_i, y) f(y)\, dy, \quad i = 1, \ldots, n. \tag{1.6}$$

This is a mathematical model for discrete observations of a physical process described by the Fredholm first kind integral operator (1.2). In (1.5), η_n is an n-vector, whose components are each a random variable. Typically, these components are assumed to be uncorrelated with zero mean and common variance σ^2. The error is then referred to as *discrete white noise*. The exact solution f^* is again assumed to be deterministic. In many practical applications, one can control the size of n, e.g., by taking more measurements. However, one often has no control over the size of the individual errors, so σ is assumed to be fixed. In this context, the regularized solution will be denoted by $f_{\alpha,n}$. Note that $f_{\alpha,n}$ is now a random function, or stochastic process. A regularization

parameter selection method is then said to be convergent if it yields a parameter $\alpha(n)$ for which

$$\mathcal{E}[||f_{\alpha(n),n} - f^*||^2] \to 0 \quad \text{as} \quad n \to \infty. \tag{1.7}$$

Here $\mathcal{E}$ denotes the mathematical expectation operator. This type of convergence is often referred to as *convergence in mean square*.

The model (1.5) was first used in the analysis of regularization parameter selection methods by Wahba [15]. She used it to prove convergence of the method of generalized cross validation (GCV) for the selection of the parameter in Tikhonov regularization. The author [14] has also employed this approach in the analysis of GCV for the selection of a truncation level for the truncated singular value decomposition (TSVD) method.

It should be observed that there is no nice continuous analogue of discrete white noise. The obvious approach is to assume a continuous stochastic process $\eta(x)$ for which

$$\begin{aligned} \mathcal{E}[\eta(x)] &= 0, \\ \mathcal{E}[\eta(x)\eta(y)] &= \sigma^2\delta(x-y), \end{aligned} \tag{1.8}$$

where δ denotes the Dirac Delta distribution. One cannot simply embed the error in a Hilbert space, since the Dirac Delta is not an L^2 function. In addition, one should be aware of the existence of "fully stochastic" models, where both the error η and the solution f^* are assumed to be stochastic. See for example [2].

The model (1.5) seems appropriate for "discrete ill-posed problems" arising in many applications. It takes into consideration the discrete nature of observed data. A stochastic model for discrete measurement error seems reasonable, given the non-reproducible, highly varying nature of such error. The use of the expectation operator in the analysis is also reasonable. The Fourier coefficients of regularized solutions depend on linear combinations of the data, typically through inner products. As the number of terms in these linear combinations becomes large, as a result of the "law of large numbers", one sees convergence of the linear combinations to their expected values. Finally, for many schemes for discretization in the variable y in (1.6), the spectra of the fully discrete operators closely resemble the spectra of the semi-discrete operators.

This paper is organized as follows. The next section deals with operator approximations and their spectral properties. In Section 3, the stochastic error model (1.5) and its ramifications are discussed in detail. In section 4, tools presented in Sections 2 and 3 are applied to analyze the L-curve method. Following Hansen and O'Leary [9], we characterize the "corner" of the L-curve as the point of maximum curvature. With this characterization and under a few well-motivated assumptions, we rigorously prove that the L-curve method fails to converge. Finally, Section 5 contains a numerical example to illustrate the analysis and to corroborate the nonconvergence result.

2. Operator Approximations

Let A be a compact operator between separable Hilbert spaces $\mathcal{H}$ and $\mathcal{Z}$. The inner products on $\mathcal{H}$ and on $\mathcal{Z}$ will be denoted by $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ and by $\langle \cdot, \cdot \rangle_{\mathcal{Z}}$, respectively, and similarly for the Hilbert space norms $|| \cdot ||_{\mathcal{H}}$, $|| \cdot ||_{\mathcal{Z}}$. Assume that A is injective and has a singular value decomposition (SVD) $\{u_j, s_j, v_j, j = 1, 2, \ldots\}$. This means that $\{v_j\}$ and $\{u_j\}$ form orthonormal bases for $\mathcal{H}$ and for the range of A in $\mathcal{Z}$, respectively. In addition, the s_j are decreasing positive numbers, and

$$Av_j = s_j u_j, \quad A^* u_j = s_j v_j, \quad j = 1, 2, \ldots .$$

Here $A^* : \mathcal{Z} \to \mathcal{H}$ denotes the adjoint of A. We shall assume that

$$\sum_{j=1}^{\infty} \lambda_j < \infty, \tag{2.1}$$

where

$$\lambda_j \stackrel{\text{def}}{=} s_j^2. \tag{2.2}$$

This holds, for example, when the kernel in (1.2) is smooth.

Consider a family of semi-discrete approximate linear operators A_n mapping $\mathcal{H}$ into R^n. Let $|| \cdot ||_n$ denote the usual Euclidean (ℓ^2) norm on R^n. Assume each A_n has "full rank", i.e., its range is all of R^n. Denote the SVD of A_n by $\{u_{jn}, s_{jn}, v_{jn}, j = 1, \ldots, n\}$. In this case, $\{v_{jn}\}$ and $\{u_{jn}\}$ form orthonormal bases for the orthogonal complement of the null space of A_n in $\mathcal{H}$ and for R^n, respectively, and the s_{jn} are all strictly positive. The pseudo-inverse operator $A_n^\dagger : R^n \to \mathcal{H}$ has a representation

$$A_n^\dagger z_n = \sum_{j=1}^{n} \frac{\hat{z}_{jn}}{s_{jn}} v_{jn}, \tag{2.3}$$

where

$$\hat{z}_{jn} \stackrel{\text{def}}{=} z_n^T u_{jn} = \sum_{i=1}^{n} [z_n]_i [u_{jn}]_i. \tag{2.4}$$

The SVD of A_n induces a projection on $\mathcal{H}$,

$$P_n f = A_n^\dagger A_n f = \sum_{j=1}^{n} \hat{f}_{jn} v_{jn}, \tag{2.5}$$

where

$$\hat{f}_{jn} \stackrel{\text{def}}{=} \langle f, v_{jn} \rangle_{\mathcal{H}}. \tag{2.6}$$

With the moment discretization (1.6) in mind, we make several assumptions below relating the operator A and the approximations A_n. These can be loosely summarized by

$$\frac{1}{\sqrt{n}} A_n \approx A \quad \text{for large} \quad n. \tag{2.7}$$

The scaling factor $\frac{1}{\sqrt{n}}$ arises from

$$||g||_{\mathcal{Z}}^2 = \int_\Omega g^2(x)\,dx \approx \frac{1}{n}\sum_{j=1}^{n} g^2(x_i).$$

This approximation is accurate provided that g is smooth, the points x_i are not clustered, and the region Ω has been scaled so that it has unit volume. This induces the following connection between the spectrum of A and that of A_n:

$$\lambda_{jn} \approx \lambda_j, \quad j = 1, \ldots, n, \tag{2.8}$$

where

$$\lambda_{jn} \stackrel{\text{def}}{=} \frac{s_{jn}^2}{n}, \tag{2.9}$$

as well as a connection between the generalized Fourier coefficients of any $f \in \mathcal{H}$,

$$f_{jn} \approx f_j, \tag{2.10}$$

where

$$f_j \stackrel{\text{def}}{=} \langle f, v_j \rangle_{\mathcal{H}}. \tag{2.11}$$

For a heuristic explanation, see [14]. This nonrigorous discussion motivates the following assumptions:

Assumption 1. For any $f \in \mathcal{H}$,

$$\lim_{n\to\infty} \frac{1}{\sqrt{n}}||A_n f||_n = ||Af||_{\mathcal{Z}}. \tag{2.12}$$

Assumption 2. There exists a constant c_2 for which

$$\sum_{j=1}^{n} \lambda_{jn} \le c_2. \tag{2.13}$$

Assumption 3. Given any $\tau > 0$, there exist positive integers N, K for which

$$|\lambda_{jn} - \lambda_j| < \tau \quad \text{whenever} \quad j \ge J, \;\; n \ge N, \;\; j \le n, \tag{2.14}$$

and

$$J,\; N \to \infty \quad \text{as} \quad \tau \to 0. \tag{2.15}$$

From (2.12) and the principle of uniform boundedness, there exists a constant c_1 for which

$$\frac{1}{\sqrt{n}}||A_n|| \le c_1. \tag{2.16}$$

Here $||\cdot||$ denotes the operator (spectral) norm. Equations (2.1) and (2.14)-(2.15) imply that

$$\lambda_{jn} \to 0 \quad \text{as} \quad j, n \to \infty. \tag{2.17}$$

3. Stochastic Model for Data Error

We assume the semi-discrete, semi-stochastic data model

$$[z_n]_i = [A_n f^*]_i + \sigma[\epsilon_n]_i, \quad i = 1, \ldots, n. \tag{3.1}$$

Here the desired exact solution $f^* \in \mathcal{H}$ is deterministic, but the error is stochastic in that the ϵ_n are discrete white noise vectors with unit covariance. This means

$$\begin{aligned} &\mathcal{E}([\epsilon_n]_i) = 0, && i = 1, \ldots, n, \\ &\mathcal{E}([\epsilon_n]_i[\epsilon_n]_j) = \delta_{ij}, && i, j = 1, \ldots, n, \end{aligned} \tag{3.2}$$

where δ_{ij} denotes the Kronecker Delta. The positive parameter σ is fixed.

Because of (2.17), the condition number of A_n, $\max_j s_{jn} / \min_j s_{jn}$, becomes arbitrarily large as n increases. Hence, regularization is needed to obtain accurate approximations to the exact solution f^*. We will analyze only Tikhonov regularization. This regularization method yields approximations to f^* (based on the operator A_n and the data z_n) with the representations

$$\begin{aligned} f_{\alpha,n} &= \arg\min_{f \in \mathcal{H}} \frac{1}{n} ||A_n f - z_n||_n^2 + \alpha ||f||_{\mathcal{H}}^2 \\ &= \sum_{j=1}^{n} \frac{s_{jn}}{s_{jn}^2 + n\alpha} \, \hat{z}_{jn} v_{jn} \\ &= \sum_{j=1}^{n} \frac{\lambda_{jn}}{\lambda_{jn} + \alpha} \, \frac{\hat{z}_{jn}}{s_{jn}} v_{jn}, \end{aligned} \tag{3.3}$$

where $\hat{z}_{jn}$ and λ_{jn} are defined in (2.4) and (2.9). Other regularization methods (e.g., TSVD and Landweber iteration) yield approximate solutions with similar representations,

$$f_{\alpha,n} = \sum_{j=1}^{n} w(\lambda_{jn}; \alpha) \frac{\hat{z}_{jn}}{s_{jn}} v_{jn}.$$

The weighting function w serves to filter out singular components corresponding to small singular values (See for example [5] for further discussion). The analysis presented here is amenable to these other regularization methods, as well as to other regularization parameter selection methods.

Note that $f_{\alpha,n}$ is a random function. Given an orthonormal sequence $\{v_j\}$ in $\mathcal{H}$ and a square summable sequence $\{\hat{f}_j\}$ of random coefficients, the "mean square" norm on $\mathcal{H}$ is defined by

$$|||\sum_j \hat{f}_j v_j|||^2 \stackrel{\text{def}}{=} \mathcal{E}[||\sum_j \hat{f}_j v_j||_{\mathcal{H}}^2] = \sum_j \mathcal{E}[\hat{f}_j^2]. \tag{3.4}$$

When the coefficients $\hat{f}_j$ are deterministic, the right hand side (r.h.s) reduces to the usual norm in $\mathcal{H}$.

In this context, we say that a regularization parameter selection method is *convergent* if it yields a parameter $\alpha = \alpha(n)$ for which

$$|||f_{\alpha(n),n} - f^*||| \to 0 \quad \text{as} \quad n \to \infty. \tag{3.5}$$

From the triangle inequality, and the fact that f^* is assumed to be deterministic,

$$|||f_{\alpha,n} - f^*||| \leq |||f_{\alpha,n} - P_n f^*||| + ||P_n f^* - f^*||_{\mathcal{H}}. \tag{3.6}$$

Nashed and Wahba [12] have shown that when moment discretization is applied, $||P_n f^* - f^*||_{\mathcal{H}} \to 0$ as $n \to \infty$. Hence, a necessary and sufficient condition for convergence is

$$\begin{aligned} |||f_{\alpha,n} - P_n f^*|||^2 = & \sum_{j=1}^{n} \frac{\alpha^2 \, (\hat{f}_{jn}^*)^2}{(\lambda_{jn} + \alpha)^2} + \frac{\sigma^2}{n} \sum_{j=1}^{n} \frac{\lambda_{jn}}{(\lambda_{jn} + \alpha)^2} \\ & \to 0 \quad \text{as} \quad n \to \infty. \end{aligned} \tag{3.7}$$

Note that the r.h.s. of (3.7) is a consequence of (3.1)-(3.3), the fact that f^* is deterministic, and

$$\hat{z}_{jn} = s_{jn} \hat{f}_{jn}^* + \sigma \hat{\epsilon}_{jn}, \tag{3.8}$$

where

$$\hat{\epsilon}_{jn} = \epsilon_n^T u_{jn} = \sum_{i=1}^{n} [\epsilon_n]_i [u_{jn}]_i.$$

See [14] for details.

The convergence (3.5) has several important consequences. From (2.16), (3.1), and the fact that $\mathcal{E}[||\epsilon_n||_n^2] = n$,

$$\frac{1}{n} \mathcal{E}[||A_n f_{\alpha,n} - z||_n^2] \to \sigma^2 \quad \text{as} \quad n \to \infty. \tag{3.9}$$

Also, because the first term on the r.h.s. of (3.7) must tend to zero, it is necessary for convergence that

$$\alpha(n) \to 0 \quad \text{as} \quad n \to \infty. \tag{3.10}$$

However, since the second term on the r.h.s. of (3.7) also must tend to zero, $\alpha(n)$ cannot go to zero too rapidly.

4. L-curve Analysis

We first examine the expected L-curve components. From (3.3), the expected value of the squared solution norm is

$$S_n(\alpha) \stackrel{\text{def}}{=} |||f_{\alpha,n}|||^2 = \sum_{j=1}^{n} \left(\frac{\lambda_{jn}}{\lambda_{jn} + \alpha} \right)^2 \frac{\mathcal{E}[\hat{z}_{jn}^2]}{n \lambda_{jn}}. \tag{4.1}$$

One obtains from (3.8) that

$$\mathcal{E}[\hat{z}_{jn}^2] = n\lambda_{jn}(\hat{f}_{jn}^*)^2 + \sigma^2,$$

and hence,

$$S_n(\alpha) = \sum_{j=1}^{n} \frac{\lambda_{jn}^2(\hat{f}_{jn}^*)^2}{(\lambda_{jn}+\alpha)^2} + \frac{\sigma^2}{n}\sum_{j=1}^{n}\frac{\lambda_{jn}}{(\lambda_{jn}+\alpha)^2}. \tag{4.2}$$

Similarly, define

$$\begin{aligned} R_n(\alpha) &\stackrel{\text{def}}{=} \frac{1}{n}\mathcal{E}[||A_n f_{\alpha,n} - z_n||_n^2] \\ &= \sum_{j=1}^{n} \frac{\alpha^2\lambda_{jn}(\hat{f}_{jn}^*)^2}{(\lambda_{jn}+\alpha)^2} + \frac{\sigma^2}{n}\sum_{j=1}^{n}\frac{\alpha^2}{(\lambda_{jn}+\alpha)^2} \end{aligned} \tag{4.3}$$

This is the expected value of the squared norm of the residual, scaled by $1/n$.

The "expected L-curve components" are

$$X(\alpha) \stackrel{\text{def}}{=} \log R_n(\alpha) \tag{4.4}$$

$$Y(\alpha) \stackrel{\text{def}}{=} \log S_n(\alpha) \tag{4.5}$$

and the "expected curvature" is given by

$$\kappa_n(\alpha) \stackrel{\text{def}}{=} \frac{X'(\alpha)Y''(\alpha) - X''(\alpha)Y'(\alpha)}{(X'(\alpha)^2 + Y'(\alpha)^2)^{3/2}}, \tag{4.6}$$

where the prime ($'$) denotes differentiation with respect to α. These are not expected values; rather they are nonlinear functions of $R_n(\alpha)$ and $S_n(\alpha)$, which are expected values. Note that

$$S_n'(\alpha) = -2\left(\sum_{j=1}^{n}\frac{\lambda_{jn}^2(\hat{f}_{jn}^*)^2}{(\lambda_{jn}+\alpha)^3} + \frac{\sigma^2}{n}\sum_{j=1}^{n}\frac{\lambda_{jn}}{(\lambda_{jn}+\alpha)^3}\right) \tag{4.7}$$

and

$$R_n'(\alpha) = -\alpha S_n'(\alpha).$$

Consequently,

$$\kappa_n(\alpha) = -\left(\frac{R_n S_n}{S_n'}\right)\frac{\alpha R_n S_n' + \alpha^2 S_n S_n' + R_n S_n}{(R_n^2 + \alpha^2 S_n^2)^{3/2}} \tag{4.8}$$

$$= -\frac{R_n S_n(\alpha R_n + \alpha^2 S_n) + R_n^2 S_n^2/S_n'}{(R_n^2 + \alpha^2 S_n^2)^{3/2}} \tag{4.9}$$

We now examine the behavior of κ_n under Assumptions 1-3 and the necessary conditions for convergence, cf., (3.7), (3.9)-(3.10). From (3.9), there exist positive constants R_{min}, R_{max} for which

$$R_{min} \le R_n(\alpha(n)) \le R_{max}. \tag{4.10}$$

Hence, the denominator in (4.9) is bounded below by R_{min}^3. Note that the right hand sides of (3.7) and (4.2) share the same second term, which must tend to 0 as $n \to \infty$. Moreover, the first term on the right hand side of (4.2) is bounded above by $||f^*||_{\mathcal{H}}^2$. Consequently, there exists a constant S_{max} for which

$$S_n(\alpha(n)) \leq S_{max}. \tag{4.11}$$

Combining (4.11) with (4.10) and (4.9),

$$|\kappa_n(\alpha(n))| \leq \frac{\mathcal{O}(\alpha(n)) + R_{max}^2 S_{max}^2/|S_n'(\alpha(n))|}{R_{min}^3}. \tag{4.12}$$

We make an additional assumption, which relates the decay of the singular values of A_n and the generalized Fourier coefficients of f^*. This may be interpreted as an assumption that f^* is "rough".

Assumption 4. There exists a positive constant c_0 for which

$$\frac{(\hat{f}_{jn}^*)^2}{\lambda_{jn}} \geq c_0, \quad j = 1, \ldots, n. \tag{4.13}$$

Then from (4.7),

$$|S_n'(\alpha(n))| \geq 2c_0 \sum_{j=1}^{n} \left(\frac{\lambda_{jn}}{\lambda_{jn} + \alpha(n)} \right)^3 \tag{4.14}$$

$$\geq 2c_0 \sum_{\lambda_{jn} \leq \alpha(n)} \left(\frac{\lambda_{jn}}{2\alpha(n)} \right)^3 + 2c_0 \sum_{\lambda_{jn} > \alpha(n)} \left(\frac{1}{2} \right)^3. \tag{4.15}$$

Thus, from (3.10) and (2.14)-(2.15),

$$|S_n'(\alpha(n))| \to \infty \quad \text{as} \quad n \to \infty. \tag{4.16}$$

Combining this with (4.12) yields

Lemma 1. If $\alpha(n)$ is chosen so that $|||f_{\alpha(n),n} - f^*||| \to 0$ and Assumptions 1-4 hold, then

$$\kappa_n(\alpha(n)) \to 0 \quad \text{as} \quad n \to \infty. \tag{4.17}$$

We next establish that there exists a constant $\overline{\alpha} > 0$ for which $\kappa_n(\overline{\alpha})$ remains positive and bounded away from zero for large n. To this end, take any fixed $\alpha_0 > 0$. From (2.13), we obtain a bound for the second term on the r.h.s. of (4.2),

$$\frac{\sigma^2}{n} \sum_{j=1}^{n} \frac{\lambda_{jn}}{(\lambda_{jn} + \alpha)^2} \leq \frac{\sigma^2 c_1}{n\alpha_0^2}.$$

Hence, this term converges to zero as $n \to \infty$, uniformly for $\alpha \geq \alpha_0$. The same holds for the second term on the r.h.s. of (4.7). Similarly from (2.14)-(2.15), the second term on

the r.h.s. of (4.3) converges to σ^2, uniformly for $\alpha \geq \alpha_0$. Motivated by this and (2.8), (2.10), we make one final assumption:

Assumption 5. For each fixed $\alpha > 0$,

$$\lim_{n\to\infty} R_n(\alpha) = R_\infty(\alpha) \stackrel{\text{def}}{=} \sum_{j=1}^{\infty} \frac{\alpha^2 \lambda_j (\hat{f}_j^*)^2}{(\lambda_j + \alpha)^2} + \sigma^2, \tag{4.18}$$

$$\lim_{n\to\infty} S_n(\alpha) = S_\infty(\alpha) \stackrel{\text{def}}{=} \sum_{j=1}^{\infty} \frac{\lambda_j^2 (\hat{f}_j^*)^2}{(\lambda_j + \alpha)^2} \tag{4.19}$$

$$\lim_{n\to\infty} S_n'(\alpha) = S_\infty'(\alpha) = -2\sum_{j=1}^{\infty} \frac{\lambda_j^2 (\hat{f}_j^*)^2}{(\lambda_j + \alpha)^3} \tag{4.20}$$

Convergence is uniform whenever $\alpha \geq \alpha_0$.

The following properties can be easily verified:

$$R_\infty'(\alpha) = -\alpha S_\infty'(\alpha), \tag{4.21}$$

$$R_\infty(\alpha) \to \sigma^2 \qquad \text{as} \quad \alpha \to 0+, \tag{4.22}$$

$$S_\infty(\alpha) \to ||f^*||_{\mathcal{H}}^2 \qquad \text{as} \quad \alpha \to 0+, \tag{4.23}$$

$$S_\infty'(\alpha) \to -\infty \qquad \text{as} \quad \alpha \to 0+, \tag{4.24}$$

$$R_\infty(\alpha) \to ||Af^*||_{\tilde{z}}^2 + \sigma^2 \qquad \text{as} \quad \alpha \to \infty, \tag{4.25}$$

$$S_\infty(\alpha) \sim \frac{1}{\alpha^2} \qquad \text{as} \quad \alpha \to \infty, \tag{4.26}$$

$$S_\infty'(\alpha) \sim \frac{1}{\alpha^3} \qquad \text{as} \quad \alpha \to \infty. \tag{4.27}$$

Now define

$$X_\infty(\alpha) = \log R_\infty(\alpha), \tag{4.28}$$

$$Y_\infty(\alpha) = \log S_\infty(\alpha). \tag{4.29}$$

From this and (4.21), the slope of the curve $(X_\infty(\alpha), Y_\infty(\alpha))$ is given by

$$m_\infty(\alpha) = \frac{Y_\infty'(\alpha)}{X_\infty'(\alpha)} = -\frac{1}{\alpha}\frac{R_\infty(\alpha)}{S_\infty(\alpha)}. \tag{4.30}$$

Its curvature is given by

$$\begin{aligned} \kappa_\infty(\alpha) &= \frac{X_\infty'(\alpha)Y_\infty''(\alpha) - X_\infty''(\alpha)Y_\infty'(\alpha)}{(X_\infty'(\alpha)^2 + Y_\infty'(\alpha)^2)^{3/2}}. \\ &= -\frac{R_\infty S_\infty(\alpha R_\infty + \alpha^2 S_\infty) + R_\infty^2 S_\infty^2/S_\infty'}{(R_\infty^2 + \alpha^2 S_\infty^2)^{3/2}}. \end{aligned} \tag{4.31}$$

From (4.22)-(4.23),

$$\lim_{\alpha\to 0+} m_\infty(\alpha) = -\infty, \tag{4.32}$$

and from (4.25)-(4.26),

$$\lim_{\alpha\to\infty} m_\infty(\alpha) = -\infty. \tag{4.33}$$

Hence the curve $(X_\infty(\alpha), Y_\infty(\alpha))$ becomes vertical as $\alpha \to 0+$ and as $\alpha \to \infty$. Also, $X_\infty(\alpha)$ increases monotonically from $\log(\sigma^2)$ to $\log(||Af^*||_{\mathcal{Z}}^2+\sigma^2)$, while $Y_\infty(\alpha)$ decreases monotonically from $\log(||f^*||_{\mathcal{H}}^2)$ to minus infinity as α increases from zero to positive infinity. In addition, from (4.22)-(4.24),

$$\lim_{\alpha\to 0+} \kappa_\infty(\alpha) = 0.$$

From this and the geometry (see Figure 2 in the next section; see also the arguments in [13]), there must be a point $\overline{\alpha} > 0$ for which $\kappa_\infty(\overline{\alpha}) > 0$. As a consequence of this and the uniform convergence of the components of κ_n to those of κ_∞ for α larger than any fixed α_0, we obtain

Lemma 2. Under Assumptions 1-3 and Assumption 5, there exists positive constants $\overline{\alpha}, \kappa_0$ such that for n sufficiently large,

$$\kappa_n(\overline{\alpha}) \geq \kappa_0. \tag{4.34}$$

As an immediate consequence of Lemmas 1 and 2, we have

Theorem 1. Under Assumptions 1-5, the L-curve method does not converge.

5. A Numerical Example

Several assumptions and simplifications were made in order to carry out the L-curve analysis in the previous section. The purpose of this section is to demonstrate that this analysis captures the essential features of the the L-curve. Thus the conclusions should hold in practical computational settings.

Consider a model problem very similar to that in [14]. The operator is of Fredholm first kind integral type with a convolution kernel,

$$(Af)(x) \stackrel{\text{def}}{=} \int_0^1 a(x-y)\, f(y)\, dy. \tag{5.1}$$

The kernel is taken to be $a(x) = 30\, x\, (1-x)$, extended periodically. The exact solution is taken to be $f^*(x) = \pi(1-x)$. The Hilbert spaces are $\mathcal{H} = \mathcal{Z} = L^2(0,1)$. The singular values of the operator A are the magnitudes of the Fourier coefficients of the convolution kernel,

$$s_j = |\int_0^1 a(x)\, e^{-\imath 2\pi\, jx}\, dx|, \tag{5.2}$$

where $\imath = \sqrt{-1}$. One can show that the singular values decay at a rate proportional to j^{-2}. The corresponding left and right singular vectors v_j, u_j of A are the real and

imaginary parts of $\exp(\imath 2\pi jx)$, and the expansion coefficients for f^* are the (usual real-valued) Fourier coefficients. These decay at a rate proportional to j^{-1}. Consequently, the ratio $(\hat{f}_j^*)^2/\lambda_j$ is proportional to j^{-1}, and (4.13) should hold for most discretization schemes.

Equation (1.2) is discretized by evaluating x at equispaced points

$$x_i = (i-1)/n, \quad i = 1, 2, \ldots, n, \tag{5.3}$$

and by applying trapezoidal quadrature with these same points as quadrature nodes. This, along with the discrete data, yields a "discrete ill-posed problem". Because of the periodicity of the kernel, the resulting matrix $\mathbf{A}_n$ is circulant with entries

$$[\mathbf{A}_n]_{ij} = \frac{1}{n} a(x_i - x_j). \tag{5.4}$$

The Fast Fourier Transform (FFT) can be used to efficiently compute the SVD of $\mathbf{A}_n$ for very large values of n. The left and right singular vectors of $\mathbf{A}_n$ have components

$$[\mathbf{u}_{jn}]_k = \frac{u_j(x_k)}{\sqrt{n}}, \quad [\mathbf{v}_{jn}]_k = \frac{v_j(x_k)}{\sqrt{n}}, \quad j, k = 1, \ldots, n, \tag{5.5}$$

and due to the trapezoidal approximations to the integral in (5.2), the scaled, squared singular values $\lambda_{jn} \stackrel{\text{def}}{=} \mathbf{s}_{jn}^2/n$ satisfy for each fixed j,

$$\lim_{n\to\infty} \lambda_{jn} = \lambda_j. \tag{5.6}$$

Similarly for the Fourier coefficients of f^*,

$$\lim_{n\to\infty} \mathbf{f}_{jn}^* = f_j^*. \tag{5.7}$$

For systems of order $n = 2^p$, $p = 8, 10, 12, 14$, data was generated according to (3.1) using pseudo-random Gaussian white noise with $\sigma = 0.1$. All computations were performed using the MATLAB commercial software package [10]. Using singular components obtained from the FFT's, the solution error $||f_{\alpha,n} - f^*||$, the L-curve components, and the curvature $\kappa_n(\alpha)$ were computed numerically. The results are summarized in the table and figures below.

In Table 1 below, α_L denotes the computed maximizer of curvature function κ_n, while α_{opt} denotes the computed minimizer of the solution error. The analysis in [15] indicates that α_{opt} should converge to zero at a rate proportional to $(\sigma^2/n)^{b/(b+c)}$, where $b = 2$ and $c = 4$ are the powers in the (polynomial) decay rates for the λ_j and for $(\hat{f}_j^*)^2$, respectively. A log-log plot indicates that α_{opt} is proportional to $n^{-1/3}$, as predicted in [15]. On the other hand, the values of α_L obtained by the L-curve method initially decrease and then approach a fixed value, or "stagnate". This stagnation behavior is consistent with the analysis in Section 4.

Table 1. Tikhonov Regularization Parameters.

n	α_{opt}	α_L
2^8	9.1337×10^{-3}	7.0641×10^{-3}
2^{10}	2.5939×10^{-3}	5.8557×10^{-3}
2^{12}	1.1061×10^{-3}	4.3354×10^{-3}
2^{14}	8.2028×10^{-4}	4.7293×10^{-3}

Figures 1-3 provide qualitative information. Figure 1 shows the norm of the solution error as a function of the regularization parameter α for $n = 2^{10}$ and $n = 2^{14}$. Figure 2 shows L-curves obtained with $n = 2^{10}$ and $n = 2^{14}$, and Figure 3 shows the corresponding curvature functions $\kappa_n(\alpha)$. One can clearly see that the L-curves have well-defined corners at which the curvatures are maximized. One can also see in Figure 3 that there is very little shift to the left (corresponding to a decrease) in the maximizer of the κ_n as n becomes large. This suggests that in a neighborhood of the maximizer of curvature, the behavior of the L-curve is as predicted by the analysis immediately preceding Lemma 2, cf. (4.22)-4.27). Further evidence that this prediction is correct is given by the vertical slopes of the curves in Figure 2 for very small and very large values of α.

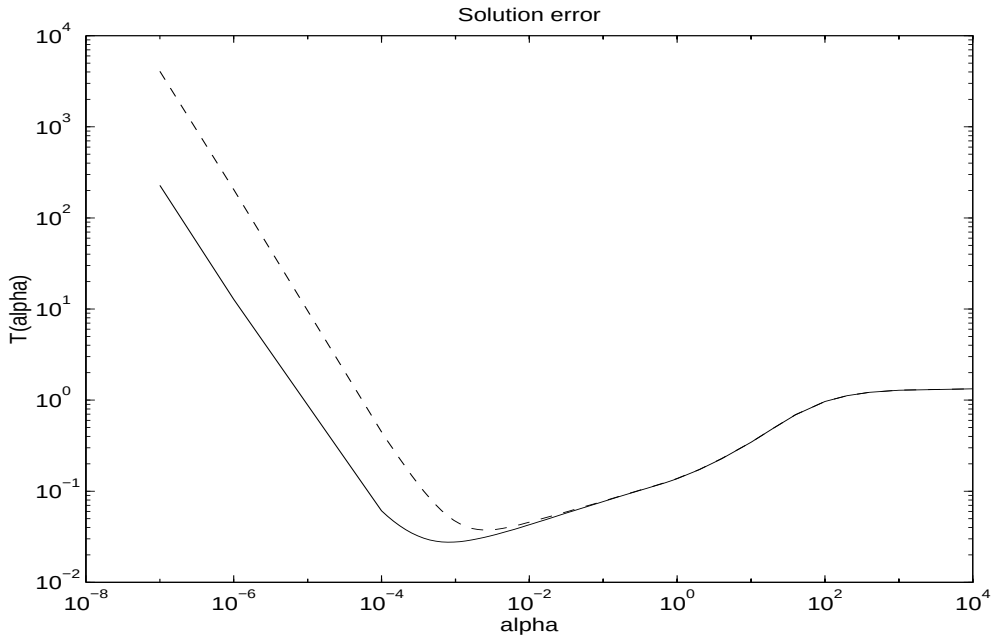

Figure 1. *Computed solution errors for $n = 2^{10}$ (dashed line) and for $n = 2^{14}$ (solid line).*

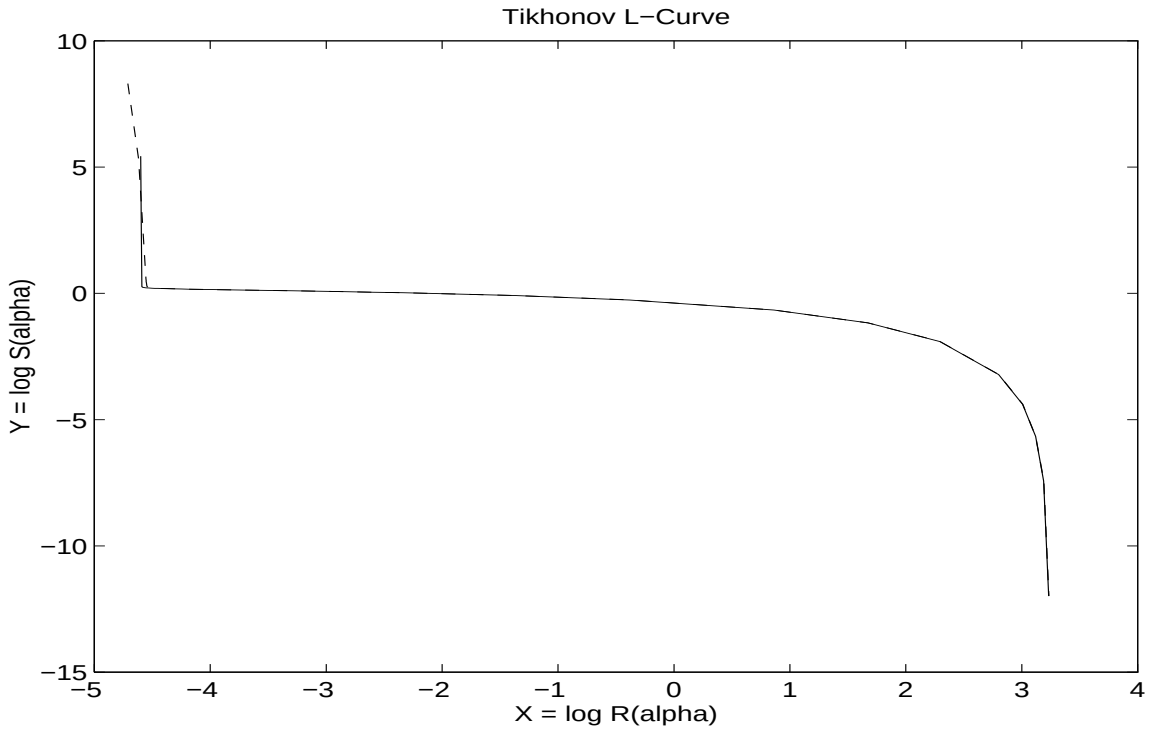

Figure 2. *Computed L-curves for* $n = 2^{10}$ *(dashed line) and for* $n = 2^{14}$ *(solid line). The regularization parameter* α *varies continuously between* 10^{-7} *and* 10^{4}*. The curves move from the upper left to the lower right as* α *increases.*

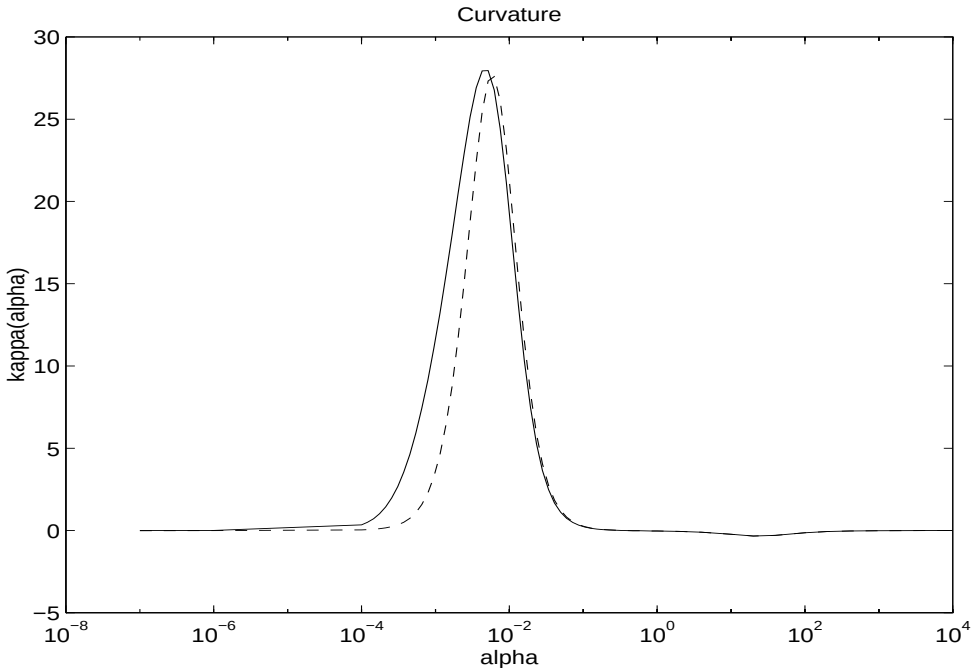

Figure 3. *Computed curvature functions* $\kappa_n(\alpha)$ *for* $n = 2^{10}$ *(dashed line) and for* $n = 2^{14}$ *(solid line).*

6. Conclusions

A partially discrete, partially stochastic model is used to analyze the L-curve method for the selection of a regularization parameter in ill-posed problems. It is demonstrated with this model that under certain conditions, the L-curve method is not convergent. This nonconvergence occurs when the solution is "rough", i.e., it's generalized Fourier coefficients decay at the same rate or less rapidly than the singular values of the operator. Numerical results are presented which corroborate the model and confirm the results of the analysis. Competing methods (e.g., GCV and the discrepancy method [11]) can be shown to converge under these circumstances (see [15, 16]).

The analysis does not apply directly to "discrete ill-posed problems", i.e., the ill-conditioned systems arising in the full discretization of ill-posed problems. However, this analysis and the numerics strongly suggests that the L-curve method may fail for discrete ill-posed problems as well.

References

[1] H. W. Engl and W. Grever, *Using the L-curve for determining optimal regularization parameters*, Numer. Math., vol. 69 (1994), pp. 25-31.

[2] J. N. Franklin, *Well-posed extensions of ill-posed linear problems*, J. Math. Anal. Appl., vol. 31 (1970), pp. 682-716.

[3] C. W. Groetsch, *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind*, Pitman, 1984.

[4] C. W. Groetsch, *Inverse Problems in the Mathematical Sciences*, Vieweg, 1993.

[5] C. W. Groetsch and C. R. Vogel, *Asymptotic theory of filtering for linear operator equations with discrete noisy data*, Mathematics of Computation, vol. 49 (1987), pp. 499-506.

[6] M. Hanke, *Limitations of the L-curve method in ill-posed problems*, preprint, 1995.

[7] M. Hanke and P. C. Hansen, *Regularization methods for large-scale problems*, Surveys Math. Indust. vol. 3 (1993), pp. 253-315.

[8] P. C. Hansen, *Analysis of discrete ill-posed problems by means of the L-curve*, SIAM Review, vol. 34 (1992), pp. 561-580.

[9] 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., vol. 14 (1993), pp. 1487-1503.

[10] MATLAB, The MathWorks, Inc., Cochituate Place, 24 Prime Park Way, South Natick, Massachusetts 01760.

[11] V. A. Morozov, *The error principle in the solution of operator equations by the regularization method*, USSR Comput. Math. and Math. Phys., vol. 8 (no. 2) (1968), pp. 63-87.

[12] N. Z. Nashed and G. Wahba, *Convergence rates of approximate least squares solutions of linear integral and operator equations of the first kind*, Math. Comp., vol. 28 (1974), pp. 69-80.

[13] T. Reginska, *A regularization parameter in discrete ill-posed problems*, SIAM J. Sci. Comput., to appear.

[14] C. R. Vogel, *Optimal choice of a truncation level for the Truncated SVD solution of linear first kind integral equations when the data are noisy*, SIAM J. Numer. Anal., vol. 23 (1986), pp. 109-117.

[15] G. Wahba, *Practical approximate solutions to linear operator equations when the data are noisy*, SIAM J. Numer. Anal., vol. 14 (1977), pp. 651-667.

[16] G. Wahba, *Spline Models for Observational Data*, SIAM, 1990.