SIAM J. SCI. COMPUT.
Vol. 14, No. 6, pp. 1487–1503, November 1993

012

THE USE OF THE L-CURVE IN THE REGULARIZATION OF DISCRETE ILL-POSED PROBLEMS*

PER CHRISTIAN HANSEN† AND DIANNE PROST O'LEARY‡

Abstract. Regularization algorithms are often used to produce reasonable solutions to ill-posed problems. The L-curve is a plot—for all valid regularization parameters—of the size of the regularized solution versus the size of the corresponding residual. Two main results are established. First a unifying characterization of various regularization methods is given and it is shown that the measurement of "size" is dependent on the particular regularization method chosen. For example, the 2-norm is appropriate for Tikhonov regularization, but a 1-norm in the coordinate system of the singular value decomposition (SVD) is relevant to truncated SVD regularization. Second, a new method is proposed for choosing the regularization parameter based on the L-curve, and it is shown how this method can be implemented efficiently. The method is compared to generalized cross validation and this new method is shown to be more robust in the presence of correlated errors.

Key words. ill-posed problems, regularization, L-curve, parameter choice, generalized cross validation, discrepancy principle

AMS subject classifications. 65R30, 65F20

1. Introduction.

In many applications such as spectroscopy [1], seismography [13], and medical imaging [11], data are gathered by convolution of a noisy signal with a detector. A linear model of this process leads to an integral equation of the first kind:

$$\int_0^1 k(s,t)\,x(t)\,dt = y_0(s) + e(s). \tag{1}$$

Here, $y_0(s) + e(s)$ is the measured signal, $y_0(s)$ is the true signal, $e(s)$ is the unknown noise, and the *kernel function* $k(s,t)$ is the instrument response function.

Since the measured signal is usually available only at a finite number of values of s, the continuous model (1) is replaced by a discrete linear model equation

$$Kx = y_0 + e \equiv y, \tag{2}$$

where K is a matrix of dimension $m \times n$ and we assume that $m \geq n$. In all but trivial deconvolution problems, the continuous problem is *ill posed* in the sense that small changes in the data can cause arbitrarily large changes in the solution, and this is reflected in ill conditioning of the matrix K of the discrete model, increasing as the dimension of the problem increases. Thus attempts to solve (2) directly yield solution vectors that are hopelessly contaminated with noise.

Hence some sort of *regularization* of the problem is required to filter out the influence of the noise. Well-known regularization methods are Tikhonov regularization and the truncated singular value decomposition (SVD). A common feature of these regularization methods is that they depend on some regularization parameter that controls how much filtering is introduced by the regularization. Often the key issue in connection with

* Received by the editors October 23, 1991; accepted for publication (in revised form) December 31, 1992.

† UNI•C (Danish Computing Center for Research and Education), Building 305, Technical University of Denmark, DK-2800 Lyngby, Denmark (unipch@wuli.uni-c.dk). The work of this author was partially supported by a travel grant from the Reinholdt W. Jorck og Hustrus Fond.

‡ Computer Science Department and Institute for Advanced Computer Studies, University of Maryland, College Park, Maryland 20742 (oleary@cs.umd.edu). The work of this author was supported by the Air Force Office of Scientific Research grant AFOSR-87-0158.

these methods is to find a regularization parameter that gives a good balance, filtering out enough noise without losing too much information in the computed solution.

The purpose of this paper is to propose new methods for the choice of the regularization parameter through use of the the *L-curve*. The L-curve is a plot—for all valid regularization parameters—of the size of the regularized solution versus the size of the corresponding residual. It was used by Lawson and Hanson [10] and further studied by Hansen [9]. In this work we establish two main results. First we give a unifying characterization of various regularization methods and show that the measurement of "size" is dependent on the particular regularization method chosen; for example, the 2-norm is appropriate for Tikhonov regularization, but a 1-norm in the coordinate system of the SVD is relevant to truncated SVD regularization. Second, we propose a systematic a posteriori method for choosing the regularization parameter based on this L-curve and show how this method can be implemented efficiently. We compare the method to generalized cross validation and the discrepancy principle.

Our analysis differs from the "asymptotic theory of filtering" [6] where the problem size (m and n) goes to infinity, in that we consider problems where the problem size is typically *fixed*, e.g., by the particular measurement setup. Thus we are ignoring the very important questions of convergence of the estimates as the model converges to the continuous problem or as the error converges to zero.

We give a unified survey of regularization methods in §2 and of algorithms for choosing the regularization parameter in §3. We investigate various important properties of the L-curve in §4 and demonstrate how a good regularization parameter can actually be computed from the L-curve. In §5 we discuss several important computational aspects of our method, and, finally, in §6 we illustrate the new method by numerical examples.

2. Regularization methods. Practical methods for solving the discretized problem (2) must diminish the influence of noise. They differ only in the way that they determine the filtering function for the noise. We illustrate this by discussing several of these methods in a common framework, using the SVD of the matrix K. Let

$$K = \sum_{i=1}^{n} \sigma_i \, u_i \, v_i^T. \tag{3}$$

Here, the left and right *singular vectors* u_i and v_i are orthonormal, and the *singular values* σ_i are nonnegative and nonincreasing numbers, i.e., $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$. Common for all discrete ill-posed problems is that the matrix K has a cluster of singular values at zero and that the size of this cluster increases when the dimension m or n is increased.

Using the SVD of K, it is straightforward to show that the ordinary least squares solution to (2), the one characterized by solving the unconstrained problem

$$\min_x \|Kx - y\|_2, \tag{4}$$

can be written as

$$x_{\mathrm{LSQ}} = \sum_{i=1}^{n} \frac{\alpha_i}{\sigma_i} v_i, \tag{5}$$

where $\alpha_i = u_i^T y$.

The trouble with using the least squares solution x_{LSQ} is that error in the directions corresponding to small singular values is greatly magnified and overwhelms the information contained in the directions corresponding to larger singular values. Any practical

method must therefore incorporate filter factors f_i, changing the computed solution to

$$x_{\text{filtered}} = \sum_{i=1}^{n} f_i \frac{\alpha_i}{\sigma_i} v_i, \tag{6}$$

where usually we take $0 \le f_i \le 1$. If each filter factor is equal to one, we have the least squares solution x_{LSQ}. The filtered residual vector corresponding to x_{filtered} is

$$r_{\text{filtered}} = \sum_{i=1}^{n} (1 - f_i)\, \alpha_i u_i + r_\perp,$$

where $r_\perp$ is the least squares residual, i.e., the component of y orthogonal to the vectors $u_1, \ldots, u_n$. Regularization methods differ only in how they choose the filter factors.

Perhaps the best known regularization method is the one due to Tikhonov [16], which chooses the solution x_λ that solves the minimization problem

$$\min_x \{ \|Kx - y\|_2^2 + \lambda^2 \|x\|_2^2 \}. \tag{7}$$

Here, the parameter λ controls how much weight is given to minimization of $\|x\|_2$ relative to minimization of the residual norm. In some applications it is not appropriate to minimize the 2-norm of the solution, but rather a seminorm $\|L\,x\|_2$ where L typically is a discrete approximation to some derivative operator. However, Eldén [2] has shown that it is always possible to transform such problems into a form where the 2-norm is minimized.

Another regularizing method is *truncated* SVD [7], [17], where one simply truncates the summation in (5) at an upper limit $k < n$, before the small singular values start to dominate.

Certain iterative methods for solving the least squares problem (4) have minimization properties. The *conjugate gradient* family of methods minimizes the least squares function over an expanding sequence of subspaces

$$\mathcal{K}_k = \text{span}\{K^T y, (K^T K)K^T y, \ldots, (K^T K)^{k-1} K^T y\},$$

keeping $\|x\|_2$ as small as possible. One formulation particularly suited to ill-posed problems is the LSQR implementation of Paige and Saunders [12].

Another popular regularizing technique is the *maximum entropy* principle; see, for example, [15]. It is based on the idea that since y contains error, it is not reasonable to ask that x reproduce it exactly, but only that it approximate it within the expected value of the norm of the error in y. Among all of the x vectors that satisfy

$$\|Kx - y\|_2 \le \lambda,$$

the maximum entropy method chooses the vector whose entropy

$$s(x) = -\sum_{i=1}^{n} x_i \ln(x_i / x_i^0) - (x_i - x_i^0)$$

is the largest, where x^0 is a nonnegative initial approximation and the admissible points x are also nonnegative.

Like the least squares formulation (4), these regularization methods produce solutions that can be characterized as solutions to minimization problems. We can think of

them as minimizing the size of the solution x subject to keeping the size of the residual $r = Kx - y$ less than some fixed value, or in a dual sense as minimizing the size of r subject to keeping the size of x less than some value $M(\lambda)$. The way that "size" is measured varies from method to method, but many of these methods are defined in terms of the norms induced by the bases of the SVD. Although the 2-norm is invariant with respect to the choice of an orthonormal basis, other norms do not share this property. We will denote the 1-norm in the SVD basis by $\|\cdot\|_{\underline{1}}$, defined by

$$\|x\|_{\underline{1}} = \|\beta\|_1 \quad \text{if} \quad x = \sum_{i=1}^{n} \beta_i v_i,$$

$$\|r\|_{\underline{1}} = \|\gamma\|_1 + \|r_\perp\|_1 \quad \text{if} \quad r = \sum_{i=1}^{n} \gamma_i u_i + r_\perp.$$

We make a similar definition for the p-norms in the SVD basis, $p = 3, \ldots, \infty$. Moreover, we denote by $M(\lambda)$ the norm of the solution vector for regularization parameter λ.

Using this notation it is easy to verify that the regularization methods mentioned above have the characterizations shown in Table 1. For instance, for the truncated SVD, we have the relations

$$\|x\|_{\underline{1}} = \sum_{i=1}^{n} f_i \left| \frac{\alpha_i}{\sigma_i} \right|,$$

$$\|r\|_{\underline{1}} = \sum_{i=1}^{n} (1 - f_i)\, |\alpha_i| + \|r_\perp\|_1.$$

For $M(\lambda) = 0$, $x = 0$ and all filter factors are zero. As M increases, we want to increase the size of x as little as possible for a given decrease in the size of r. When the size of r has been reduced by by δ, the size of x is

$$\sum_{i=1}^{n} \frac{|\delta_i|}{\sigma_i},$$

with

$$\sum_{i=1}^{n} |\delta_i| = \delta.$$

Thus the minimal size of x is achieved by finding the largest integer k so that

$$\sum_{i=1}^{k} |\alpha_i| \le \delta,$$

and then setting

$$\delta_i = \begin{cases} \alpha_i, & i = 1, \ldots, k, \\ (M - \sum_{i=1}^{k} |\alpha_i|)\mathrm{sgn}(\alpha_i), & i = k+1, \\ 0, & i = k+2, \ldots, n. \end{cases}$$

TABLE 1
Various regularization methods and their characterizations.

Method	Minimizes	Domain	Filter factors
Tikhonov	$\|r\|_2$	$\{x : \|x\|_2 \le M(\lambda)\}$	$f_i = \frac{\sigma_i^2}{\lambda^2+\sigma_i^2}$
Truncated SVD	$\|r\|_1$	$\{x : \|x\|_1 \le M(\lambda)\}$	$f_i = 1, i = 1, \ldots, k(M)$, $f_{k(M)+1} = \frac{\sigma_{k(M)+1}}{\alpha_{k(M)+1}} \left(M - \sum_{i=1}^{k(M)} \frac{\|\alpha_i\|}{\sigma_i} \right)$ $f_i = 0, i = k(M)+2, \ldots, n$
l_∞	$\|r\|_\infty$	$\{x : \|x\|_\infty \le M(\lambda)\}$	$f_i = \min(1, \frac{M\sigma_i}{\|\alpha_i\|})$
LSQR	$\|r\|_2$	$\{x \in \mathcal{K}_k : \|x\|_2 \le M(\lambda)\}$	No simple formula
Maximum entropy	$\|r\|_2$	$\{x \ge 0 : s(x) \ge M(\lambda)\}$	No simple formula

There is no *simple formula* for the filter factors for LSQR; but we will show in forthcoming work that they can be computed easily from intermediate quantities in the LSQR algorithm. For maximum entropy, we are not aware of an algorithm for computing the filter factors.

To unify notation, we will denote the regularization parameter by λ, even for methods such as truncated SVD and LSQR in which the regularization is determined by a discrete value k. We will denote the function minimized by a regularization method by $\rho(\lambda)$ and the norm or function associated with the regularized solution vector x by $\eta(\lambda)$. As an example, for Tikhonov regularization we have $\rho(\lambda) = \|Kx - y\|_2$ and $\eta(\lambda) = \|x\|_2$.

The method l_∞ in the table is one of a family of methods, based on the l_p norms using the singular vector basis. For a comparable value of the regularization parameter, the l_∞ method produces a solution that is much "less smooth" than that of the truncated SVD or the Tikhonov method. The truncated SVD (l_1) solution has no components in directions corresponding to small singular values. The Tikhonov (l_2) solution has small components in these directions. The l_p ($p > 2$) solutions have larger components, and the l_∞ solution has components of size comparable to those in the directions corresponding to large singular values. We note that these methods can also be generalized to weighted l_p norms.

From this discussion we see that *the choice of regularization method is a choice of an appropriate pair of functions ρ and η. The proper choice of the regularization parameter is a matter of choosing the right cutoff for the filter factors f_i, i.e., the breakpoint in the singular value spectrum where one wants the damping to set in.* Algorithms for choosing the regularization parameter are still a subject of research. In the next section we survey two proposals for choosing the regularization parameter. Their shortcomings lead us to propose choosing the parameter based on the behavior of the L-curve, a plot of $\eta(\lambda)$ vs. $\rho(\lambda)$. The remainder of the paper is devoted to a discussion of the properties of the L-curve, numerical issues in using it to choose a regularization parameter, and examples of its performance compared with other methods.

3. Choosing the regularization parameter. We survey the discrepancy principle and generalized cross validation, and then we propose the new method based on the L-curve.

3.1. The discrepancy principle. Perhaps the simplest rule is to choose the regularization parameter to set the residual norm equal to some upper bound for the norm $\|e\|_2$ of the errors in the right-hand side. In connection with discrete ill-posed problems this is called the *discrepancy principle* [5, §3.3]. There is also a generalized discrepancy

principle that takes errors in the matrix K into account [9]. A major disadvantage of this method—apart from the fact that often a close bound on $\|e\|_2$ is not known—is the generally accepted fact that the discrepancy principle "oversmooths" the solution: i.e., it will choose the Tikhonov parameter λ too large, will drop too many singular values in the truncated SVD, or will halt LSQR at too early a stage. Thus we will not recover all the information actually present in the given right-hand side y.

3.2. Generalized cross-validation. A more promising rule is *generalized cross-validation* (GCV) [4], [18]. The basic idea in cross-validation is the following: if any data point y_i is left out and a solution $x_{\lambda,i}$ is computed to the reduced problem of dimension $(m-1)\times n$, then the estimate of y_i computed from $x_{\lambda,i}$ must be a good estimate. While ordinary cross-validation depends on the particular ordering of the data, generalized cross-validation is invariant to orthogonal transformation (including permutations) of the data vector y.

The GCV function to be minimized in this method is defined by

$$\mathcal{G}(\lambda) \equiv \frac{\|K\,x(\lambda) - y\|_2^2}{(\mathrm{trace}(I - K\,K(\lambda)^I))^2},$$

where $K(\lambda)^I$ is any matrix that maps the right-hand side y onto the solution $x(\lambda)$, i.e., $x(\lambda) = K(\lambda)^I y$.

Although GCV works well for many problems, there are some situations in which GCV has difficulty finding a good regularization parameter. One difficulty is that the GCV function can have a very flat minimum and hence the minimum itself may be difficult to localize numerically. This is illustrated in [17].

Another difficulty is that GCV can sometimes mistake correlated noise for a signal. The underlying assumption when deriving GCV, cf. [4], [18], is that the errors in the right-hand side are normally distributed with zero mean and covariance matrix $\sigma^2 I$. We state from [18, p. 65] that GCV "is fairly robust against nonhomogenity of variance and non-Gaussian errors.... However, the method is quite likely to give unsatisfactory results if the errors are highly correlated." We illustrate this difficulty with a numerical example in §6.

3.3. The L-curve method. Another, more recent, alternative is to base the regularization parameter on the so-called *L-curve* [9]. The L-curve is a parametric plot of $(\rho(\lambda), \eta(\lambda))$, where $\eta(\lambda)$ and $\rho(\lambda)$ measure the size of the regularized solution and the corresponding residual [10]. The underlying idea is that a good method for choosing the regularization parameter for discrete ill-posed problems must incorporate information about the solution size in addition to using information about the residual size. This is indeed quite natural, because we are seeking a fair balance in keeping both of these values small. The L-curve has a distinct L-shaped corner located exactly where the solution x_λ changes in nature from being dominated by regularization errors (i.e., by oversmoothing) to being dominated by the errors in the right-hand side. Hence the corner of the L-curve corresponds to a good balance between minimization of the sizes, and the corresponding regularization parameter λ is a good one.

A feature of the L-curve that has not previously been considered is that the 2-norm is not always the appropriate measure of the size of the solution and residual vectors. The natural way to measure size is induced by the choice of the regularization method. Referring to Table 1, we conclude that the 2-norm is natural for Tikhonov regularization, for example, while the l_1 norm should be used for the truncated SVD, since *that is the norm in which it is optimal.*

The idea of using the corner of the L-curve as a means for computing a good regularization parameter was originally proposed in [9], where it is also demonstrated that under certain assumptions that this criterion is indeed similar to both GCV and the discrepancy principle. Experiments confirm that whenever GCV finds a good regularization parameter, the corresponding solution is located at the corner of the L-curve.

The L-curve method for choosing the regularization parameter has advantages over GCV: computation of the corner is a well-defined numerical problem, and the method is rarely "fooled" by correlated errors. Even highly correlated errors will make the size of the solution grow once the regularization parameter λ becomes too small, thus producing a corner on the L-curve. We make these statements more precise in the next section.

4. Properties of the L-curve.

4.1. The shape of the curve. Many properties of the L-curve for Tikhonov regularization are investigated in [9]. In particular, it is shown that under certain assumptions the L-curve (ρ, η) for Tikhonov regularization 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 errors. The three assumptions made in [9] are:

1. The discrete Picard condition is satisfied, i.e., the coefficients $|\alpha_i|$ on average decay to zero faster than the singular values σ_i.
2. The errors in the right-hand side are essentially "white noise."
3. The signal-to-noise ratio is reasonably large.

It was also shown in [9], under assumptions 1–3, that for any method whose filter factors behave quantitatively like those for Tikhonov regularization, its 2-norm L-curve (discrete or continuous) will be close to that of Tikhonov regularization.

It is, however, possible to show that the L-curve will *always* have an L-shaped appearance. Assume that the right-hand side has a component in each singular direction. (If not, reduce the problem dimension by dropping the corresponding component in the SVD.) The only other assumption we need to make is that the desired solution vector is bounded in size by some number $\bar{M}$ that is less than the size of the least squares solution (5). Such an assumption is realistic in all practical problems; perhaps all we know is that $\bar{M}$ is less than 10^{10}, but that is all we assume for now.

Consider first the L-curve (ρ^2, η^2) for Tikhonov regularization. We know that

$$\eta^2(\lambda) = \|x(\lambda)\|_2^2 = \sum_{i=1}^{n} \frac{\sigma_i^2 \alpha_i^2}{(\lambda^2 + \sigma_i^2)^2}$$

and

$$\rho^2(\lambda) = \|r(\lambda)\|_2^2 = \sum_{i=1}^{n} \frac{\lambda^4 \alpha_i^2}{(\lambda^2 + \sigma_i^2)^2} + \|r_\perp\|_2^2.$$

Thus

$$\frac{d(\eta^2(\lambda))}{d\lambda} = -4\lambda \sum_{i=1}^{n} \frac{\sigma_i^2 \alpha_i^2}{(\lambda^2 + \sigma_i^2)^3}, \qquad \frac{d(\rho^2(\lambda))}{d\lambda} = 4\lambda^3 \sum_{i=1}^{n} \frac{\sigma_i^2 \alpha_i^2}{(\lambda^2 + \sigma_i^2)^3}.$$

Therefore $d(\eta^2)/d(\rho^2) = -\lambda^{-2}$. Evaluation of the second derivative shows that the L-curve is convex and becomes steeper as the parameter λ approaches the smallest singular value.

The truncated SVD solutions yield a piecewise linear L-curve (ρ, η) using the l_1 norm measure of size. On the ith segment, the size of the residual changes by $|\alpha_i|$, while the size of the solution changes by $-|\alpha_i|/\sigma_i$. Thus the slope of the ith segment is $-1/\sigma_i$, and the curve becomes steeper as the size of the residual decreases.

It is easy to show that the l_∞ method has a similar property: the slope of each segment is again $-1/\sigma_i$ for some value of i.

Thus we have shown that for Tikhonov regularization, truncated SVD, and the l_∞ method, the L-curves become vertical as the size of the residual approaches its lower limit. Note that the slopes in each of these three cases are determined by K alone, independent of the right-hand side.

In forthcoming work with G. W. Stewart it is shown that the L-curve for LSQR has similar behavior if the right-hand-side coefficients decay sufficiently rapidly. The behavior of the curve for the maximum entropy criterion is a topic for future research.

Thus the L-curve basically consists of a vertical part for values of $\eta(\lambda)$ near the maximum value and an adjacent part with smaller slope. The more horizontal part corresponds to solutions where the regularization parameter is too large and the solution is dominated by regularization errors. The vertical part corresponds to solutions where the regularization parameter is too small and the solution is dominated by right-hand-side errors magnified by the division by small singular values. This behavior does not rely on any additional properties of the problem, e.g., statistical distribution of the errors, the discrete Picard condition, etc.

The idea of the L-curve criterion for choosing the regularization parameter is to choose a point on this curve that is at the "corner" of the vertical piece. Having an upper bound $\bar{M}$ on the size of x prevents us from being fooled by any other corners that the L-curve may have; in the absence of other information, we seek the leftmost corner consistent with the bound $\bar{M}$. The following are two ways of viewing the problem of corner location.

1. We could seek the point on the curve closest to the origin. The definition of "closest" can vary from method to method. For example, Tikhonov regularization measures distance as $\rho + \lambda^2\eta$.

2. We could choose the point on the L-curve where the curvature is maximum. The curvature is a purely geometrical quantity that is independent of transformations of the regularization parameter. We discuss implementation of this idea in §5.

The rationale behind using the corner to find a regularization parameter λ is that the corner corresponds to a solution in which there is a fair balance between the regularization and perturbation errors—because the corner separates the horizontal part of the curve from the more vertical part. This choice of λ may lead to a slightly underregularized solution because the influence of the perturbation errors must become apparent before the corner appears.

We stress numerically reliable methods but emphasize the fact that the L-curve picture gives a check on any method for locating the corner, as well as further insight into problem behavior, and that the user should not fail to look at it. There is good reason to use a set of routines that provide reliable numerical methods as well as good graphics, such as the Matlab-based code of Hansen [8].

4.2. Distinguishing signal from noise. In our experiments we have found that in many cases it is advantageous to consider the L-curve (ρ, η) in a log-log scale. There is strong intuitive justification for this. Since the singular values typically span several orders of magnitude, the behavior of the L-curve is more easily seen in such a log-log

scale. In addition, the log-log scale emphasizes "flat" parts of the L-curve where the variation in either ρ or η is small compared to the variation in the other variable. These parts of the L-curve are often "squeezed" close to the axes in a lin-lin scale. Hence the log-log scale actually emphasizes the corner of the L-curve. One more advantage of the log-log scale is that particular scalings of the right-hand side and the solution simply shift the L-curve horizontally and vertically. Thus we do all of our computations related to curvature on $(\log \rho, \log \eta)$.

The log-log transformation has a theoretical justification as well. Consider the (ρ, η) curve for the truncated SVD algorithm. Recall that the ρ is the l_1 norm of the residual, while η is the l_1 norm of the solution vector, and the curve consists of the points produced by the truncated SVD algorithm for various numbers of retained singular values, $1 \leq k \leq n$. Using (5) we see that as k is increased by 1, the change in ρ is $|\alpha_k|$, while the change in η is $|\alpha_k/\sigma_k|$. Thus the slope of the kth segment of the piecewise linear interpolant is $1/\sigma_k$, independent of the right-hand side for the problem. Therefore, there is no hope of distinguishing signal from noise by examining properties of the L-curve in the lin-lin scale.

In a log-log scale, however, the slope of the kth segment is the *relative* change in η divided by the *relative* change in ρ, and these behave quite differently for signal and noise. A noiseless signal for which the discrete Picard condition is satisfied has the property that the sequences $|\alpha_k|$ and $|\alpha_k/\sigma_k|$ both approach zero. Thus the relative change in η approaches zero, while the relative change in ρ is finite and nonzero. Therefore, for a signal, the L-curve in log-log coordinates becomes flat as k is increased.

Pure noise gives a quite different L-curve in log-log scale. If we assume that the error components $|\alpha_k|$ are roughly a constant value ϵ, then $\eta_k \approx \epsilon/\sigma_k$ and the relative change in η_k is approximately σ_{k-1}/σ_k. The relative change in ρ_k is $1/(m-k)$, so the slope of the piecewise linear interpolant is $(m-k)\sigma_{k-1}/\sigma_k$. The L-curve for noise in log-log scale therefore has a steep slope as k increases, unlike the flat curve of the signal.

The same conclusion holds for the l_∞ regularization method. Suppose we increase the norm of x from γ to $\hat{\gamma}$. Then the norm of the residual changes from $\max_i |\beta_i| - |\gamma\sigma_i|$ to $\max_i |\beta_i| - |\hat{\gamma}\sigma_i|$. The slope of the L-curve in lin-lin coordinates is not strongly dependent on the right-hand-side coefficients β_i. The picture is different in log-log coordinates, though. The relative change in the norm of x is $(\hat{\gamma}-\gamma)/\gamma$, and the relative change in the norm of r is $(\hat{\gamma}-\gamma)\sigma_i/(|\beta_i| - |\gamma\sigma_i|)$ for some value of i. For pure signal, as γ increases, the value of i will be small: since $\alpha_i/\sigma_i \to 0$ by the discrete Picard condition, the components corresponding to small singular values are not changed by further increase in the norm of x. Thus the L-curve will have moderate slope. For pure noise, $i = n$, and the L-curve will be quite steep as γ increases.

The L-curves for pure signal and pure noise in Tikhonov regularization have similar characters: both curves are steep in lin-lin scale as $\lambda \to 0$, but only the noise curve is steep in log-log scale. This can be shown either by appealing to the closeness of the truncated SVD and the Tikhonov solutions and residuals when both are measured in the 2-norm, or by rather tedious computations with the 2-norm.

5. Numerical issues in locating the corner of the L-curve. Although the L-curve is easily defined and quite satisfying intuitively, computing the point of maximum curvature in a numerically reliable way is not as easy as it might seem. Below, we discuss three cases of increasing difficulty. Throughout this section we use the notation $(\hat{\rho}, \hat{\eta})$ for the chosen measures of the size of the residual vector and the size of the solution vector. In practical computation these would probably be taken to be the *log*s of the 1, 2, or p norms of these vectors, or some weighted versions of these norms.

5.1. The ideal situation: The L-curve defined by a smooth, computable formula. If the functions $\hat{\rho}$ and $\hat{\eta}$ are defined by some computable formulas, and if the L-curve is twice continuously differentiable, then it is straightforward to compute the curvature $\kappa(\lambda)$ of the L-curve by means of the formula

$$\kappa(\lambda) = \frac{\hat{\rho}'\hat{\eta}'' - \hat{\rho}''\hat{\eta}'}{((\hat{\rho}')^2 + (\hat{\eta}')^2)^{3/2}}. \tag{8}$$

Here, $'$ denotes differentiation with respect to the regularization parameter λ. Any one-dimensional optimization routine can be used to locate the value of λ that corresponds to maximum curvature.

This situation arises when using Tikhonov regularization on a problem for which the singular values of the matrix K are known. It is practical computationally, since the effort involved in such a minimization is much smaller than that for computing the SVD.

5.2. Lacking a smooth, computable function defining the L-curve. In many situations we are limited to knowing only a finite set of points on the L-curve. This is the case, for example, for the truncated SVD and LSQR algorithms, and in these and other cases the underlying curve is not differentiable. Thus the curvature (8) cannot be computed and in fact may fail to exist. The same may be the case for problems where the regularized solution results from some black box routine.

In a computational sense, the L-curve then consists of a number of discrete points corresponding to different values of the regularization parameter at which we have evaluated $\hat{\rho}$ and $\hat{\eta}$. In many cases, these points are clustered, giving the L-curve fine-grained details that are not relevant for our considerations. For example, if there is a cluster of small singular values σ_{i_1} through σ_{i_2} with right-hand-side coefficients even smaller, then the L-curve for the truncated SVD will have a cluster of points for values of k from i_1 to i_2. This situation does not occur for Tikhonov regularization because all the components in the solution come in gradually as the filter factors change from zero to one.

We must define a differentiable, smooth curve associated with the discrete points in such a way that fine-grained details are discarded while the overall shape of the L-curve is maintained; i.e., we want the approximating curve to achieve local averaging while retaining the overall shape of the curve. A reasonable approach is therefore to base the approximating smoothing curve on cubic splines. If we fit a pair of cubic splines to $\hat{\rho}(\lambda)$ and $\hat{\eta}(\lambda)$, or if we fit a cubic spline to $\hat{\eta}(\hat{\rho})$, then we have difficulty with approximating the corner well because dense knots are required here. This conflicts with the purpose of the fit, namely, to locate the corner.

Instead, we propose fitting a *cubic spline curve* to the discrete points of the L-curve. Such a curve has several favorable features in connection with our problem: it is twice differentiable, it can be differentiated in a numerically stable way, and it has local shape-preserving features [3]. Yet we must be careful not to approximate the fine-grained details of clusters of points too well. Since a cubic spline curve does not intrinsically have the desired local smoothing property, we propose the following two-step algorithm for computing a cubic spline-curve approximation to a discrete L-curve.

ALGORITHM FITCURVE

1. Perform a local smoothing of the L-curve points, in which each point is replaced by a new point obtained by fitting a low-degree polynomial to a few neighboring points.
2. Use the new smoothed points as control points for a cubic spline curve with knots $1, \ldots, N + 4$, where N is the number of L-curve points.

Step 1 essentially controls the level of fine-grained details that are ignored. We have good experience with fitting a straight line in the least squares sense to five points centered at the point to be smoothed (a 1-norm fit may also work well, but is more difficult to compute). We illustrate the use of this algorithm in §6 .

In connection with using this algorithm as a stopping criterion for LSQR or any other iterative method, we stress that it is our belief that any sophisticated stopping rule for regularizing iterative methods (GCV, locating the point closest to the origin, finding the point of maximum curvature, etc.) must go a few iterations too far in order to determine the corner of the L-curve.

5.3. Limiting the number of L-curve points. In many cases, evaluating points on the L-curve is computationally very demanding and one would prefer to compute as few points as possible. For such problems, with differentiable as well as nondifferentiable L-curves, we need an algorithm that tries to locate the corner of the L-curve efficiently.

Assume that one knows a few points on each side of the corner. Then the ideas from the previous section can be used to derive an algorithm that seeks to compute a sequence of new regularized solutions whose associated points on the L-curve (hopefully) approach its corner. The algorithm is as follows.

ALGORITHM FINDCORNER

1. Start with a few points $(\hat{\rho}_i, \hat{\eta}_i)$ on each side of the corner.
2. Use the ideas in Algorithm FITCURVE to find an approximating three-dimensional cubic spline curve S for the points $(\hat{\rho}_i, \hat{\eta}_i, \lambda_i)$, where λ_i is the regularization parameter that corresponds to $(\hat{\rho}_i, \hat{\eta}_i)$.
3. Let S_2 denote the first two coordinates of S, such that S_2 approximates the L-curve.
4. Compute the point on S_2 with maximum curvature, and find the corresponding λ_0 from the third coordinate of S.
5. Solve the regularization problem for λ_0 and add the new point $(\hat{\rho}(\lambda_0), \hat{\eta}(\lambda_0))$ to the L-curve.
6. Repeat from Step 2 until convergence.

In step 2, it is necessary to introduce λ_i as a third coordinate of S because we need to associate a regularization parameter with every point on S_2. A two-dimensional spline curve with λ_i as knots does not provide this feature. We stress again that it is not suitable to fit individual splines to $\hat{\rho}_i$ and $\hat{\eta}_i$.

Initial points for step 1 can be generated by choosing very "large" and very "small" regularization parameters, for example, λ equal to σ_1, $\frac{1}{10}\sigma_1$, $10\sigma_n$, and σ_n. Since these initial points may be far from the corner, we found it convenient to introduce an artificial temporary point $(\min_i(\hat{\rho}_i), \min_i(\hat{\eta}_i))$ between the points corresponding to "large" and "small" λ. This temporary point is *replaced* by the first L-curve point $(\hat{\rho}(\lambda_0), \hat{\eta}(\lambda_0))$ computed in the first iteration.

6. Numerical examples. In this section we illustrate the theory from the previous sections with numerical examples. We consider a first-kind Fredholm integral equation which is a one-dimensional model problem in image reconstruction from [14]. In this model, the unknown function x is the original signal, the kernel function $k(s,t)$ is the point spread function of an infinitely long slit, and the right-hand side y is the measured signal: i.e., y consists of the original signal x integrated with $k(s,t)$ plus additional noise e. The kernel is given by

$$(9)\quad k(s,t) = (\cos s + \cos t)\left(\frac{\sin u}{u}\right)^2, \quad u = \pi(\sin s + \sin t), \quad s,t \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right].$$

Discretization is performed by means of simple collocation with delta functions as basis functions. Hence the vectors x and y are simply samples of the underlying functions. Throughout, the order of the matrix K is $m = n = 64$.

We consider two different right-hand sides, both generated by multiplying the matrix K times the corresponding true solution vector x. The first right-hand side, y_1, satisfies the discrete Picard condition; i.e., the Fourier coefficients $\alpha_i = u_i^T y_1$ decay to zero faster than the singular values σ_i, so the solution coefficients α_i/σ_i also decay to zero. This right-hand side y_1 corresponds to a solution x_1 with two "humps" given by

$$(10)\qquad x_1(t) = 2\exp(-6(t-0.8)^2) + \exp(-2(t+0.5)^2).$$

The second right-hand side y_2 only marginally satisfies the discrete Picard condition: the right-hand-side coefficients are artificially generated so that all the solution coefficients are of approximately the same size. This problem is harder to solve numerically than the first one. The norms of the two right-hand sides are $\|y_1\|_2 = 18.6$ and $\|y_2\|_2 = 19.3$.

To each of these right-hand sides we add perturbation error consisting of normally distributed numbers with zero mean and standard deviation 10^{-2} so that $\|e\|_2 \approx 8 \cdot 10^{-2}$.

The L-curves associated with truncated SVD, Tikhonov regularization, and the $\ell_{\underline{\infty}}$ methods (see §2) are shown in Fig. 1. In lin-lin scale the truncated SVD and $\ell_{\underline{\infty}}$ curves are piecewise linear, but these segments appear curved in the log-log scale. For both model problems and all three methods, the L-shaped appearance of the curves is very distinct. In particular, we notice the flat parts of the curves, corresponding to domination by the regularization error, and the vertical parts, corresponding to domination by perturbation errors. We also notice that even though the discrete Picard condition is barely satisfied for the second right-hand side y_2, the corresponding L-curves still have a distinct flat part.

The rounded corner on the L-curve for truncated SVD shows the need for a rigorous definition of the "corner" of the L-curve—but in fact the other L-curves also have a rounded "corner" on a finer scale.

For the first model problem and for Tikhonov regularization, let us now consider the two types of errors in the regularized solution $x(\lambda)$. To this end, let $\hat{x}(\lambda)$ denote the part of $x(\lambda)$ solely from the unperturbed part of the right-hand side, such that $x(\lambda) - \hat{x}(\lambda)$ is the perturbation component of $x(\lambda)$. We want to compare the regularization error $\|x_1 - \hat{x}(\lambda)\|_2$ with the perturbation error $\|\hat{x}(\lambda) - x(\lambda)\|_2$. This is done in the left part of Fig. 2. Obviously, the regularization error increases with λ while the perturbation error decreases. The regularization parameter λ for which the two error types are identical can be characterized as the optimal λ.

The two vertical lines in Fig. 2 represent the regularization parameters chosen by means of the L-curve criterion (dashed-dotted line) and GCV (dotted line). The right part of the figure shows the corresponding GCV function and its minimum. Two typical situations are shown. In the upper part both GCV and the L-curve criterion yield approximately the same regularization parameter. In the bottom part of the figure the GCV function is quite flat, and the regularization parameter is a factor 10 too small. Thus Fig. 2 illustrates the major difficulty with the GCV method, namely, that the minimum is not always so well-defined.

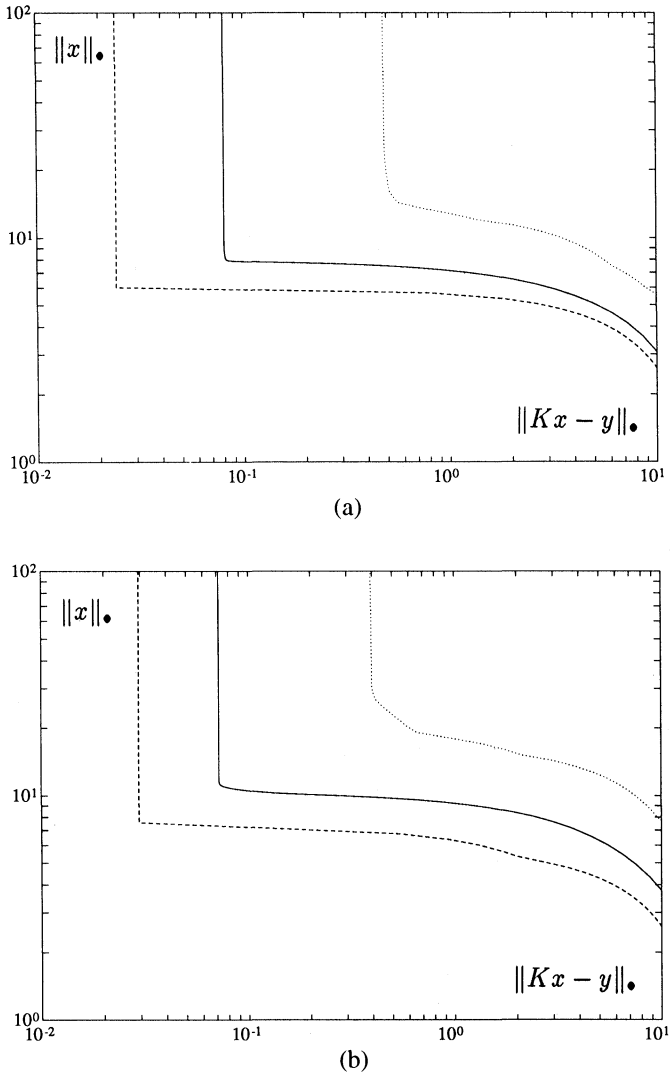

FIG. 1. *The L-curves for model problems* (a) *one and* (b) *two for the the* $\ell_{\underline{\infty}}$ *method* (*dashed line*), *Tikhonov regularization* (*solid line*), *and truncated SVD* (*dotted line*).

Let us now consider the robustness of the two competing methods in more detail. To do this, we compute the relative error in the regularized solution for both model problems:

$$\frac{\|x_i - x_\lambda\|_2}{\|x_i\|_2}, \qquad i = 1, 2,$$

computing the regularization parameter λ by both the L-curve criterion and by GCV. We used a broad range of error levels: $\|e\|_2/\|y\|_2 = 10^{-j}$, $j = 1, 2, \ldots, 8$, and for each error level we generated 25 error vectors. Thus for each model problem we solved 200 regularization problems. The results are shown in Figs. 3 and 4 as histograms with a logarithmic abscissa axis. The L-curve criterion rarely fails to compute a satisfactory

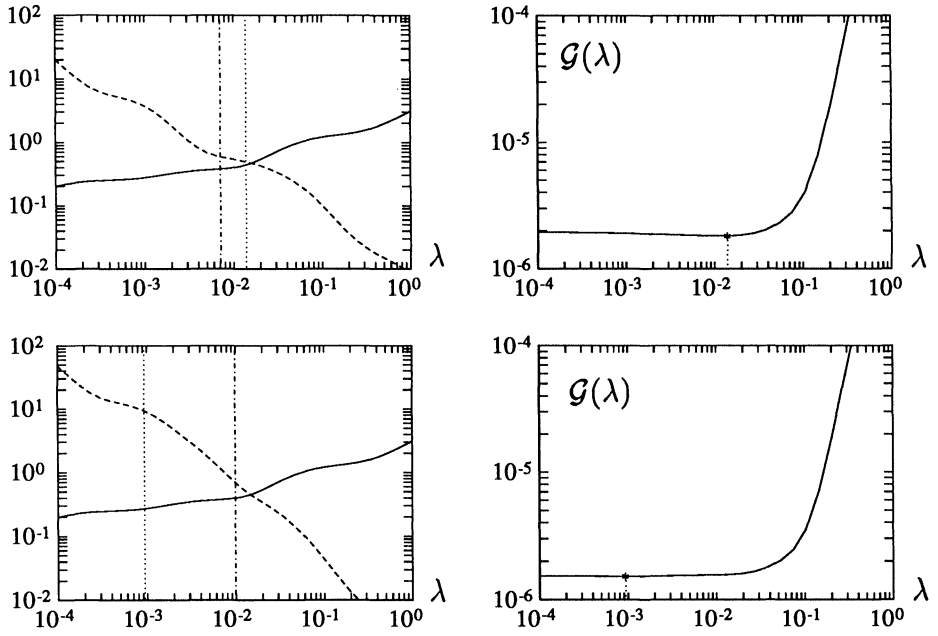

FIG. 2. *The left part shows the regularization error (solid line) and the perturbation error (dashed line) for model problem one, Tikhonov regularization, and two different random perturbations. The vertical lines represent the regularization parameters computed by means of the L-curve criterion (dashed-dotted line) and the* GCV *method (dotted line). The right part shows the corresponding* GCV *functions and their minima.*

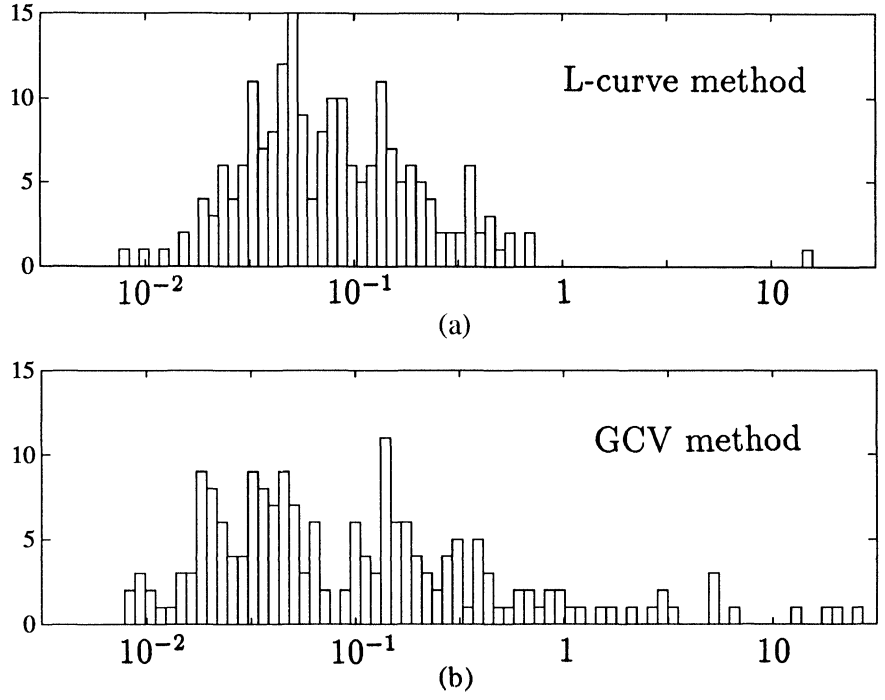

FIG. 3. *Histograms of* 200 *relative errors* $\|x_1 - x(\lambda)\|_2/\|x_1\|_2$ *for model problem one, Tikhonov regularization, and* λ *chosen by means of* (a) *the L-curve criterion and* (b) *the* GCV. *The error level* $\|e\|_2/\|y\|_2$ *varies between* 10^{-9} *and* 10^{-1}.

regularization parameter, while the GCV method fails quite often, due to the difficulties mentioned above. It is no surprise that the relative errors are typically smaller for the first model problem, because the satisfaction of the discrete Picard condition makes this problem somewhat easier to solve numerically than the second model problem.

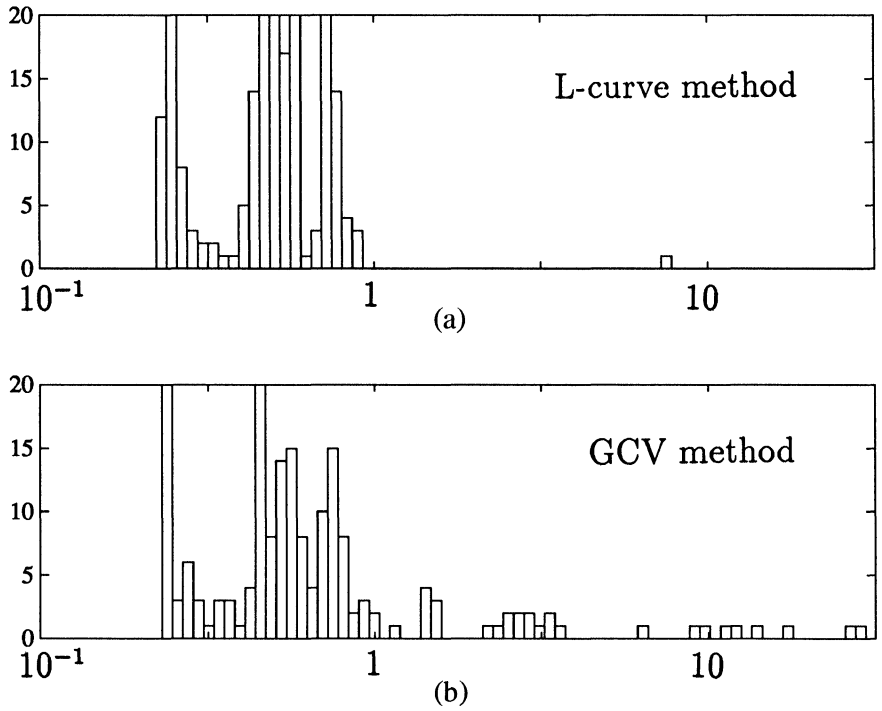

FIG. 4. *Histograms of* 200 *relative errors* $\|x_2 - x_\lambda\|_2 / \|x_2\|_2$ *for model problem two, Tikhonov regularization, and* λ *chosen by means of* (a) *the L-curve criterion and* (b) *the* GCV. *The error level* $\|e\|_2 / \|y\|_2$ *varies between* 10^{-9} *and* 10^{-1}.

Finally, let us consider problems with highly correlated errors. For this purpose we use the first problem, but now the perturbation e is generated as follows. Once the matrix K and the right-hand-side y_1 have been computed, we smooth their elements k_{ij} and $y_{1,i}$ by the following scheme:

$$\tilde{y}_{1,i} = y_{1,i} + \mu\,(y_{1,i-1} + y_{1,i+1}), \qquad i = 2, \ldots, n-1,$$

$$\tilde{k}_{ij} = k_{ij} + \mu\,(k_{i-1,j} + k_{i+1,j} + k_{i,j-1} + k_{i,j+1}), \qquad i, j = 2, \ldots, n-1.$$

Hence the right-hand-side errors are $e_i = \tilde{y}_{1,i} - y_{1,i}$, and similarly for the matrix. The parameter μ controls the amount of smoothing. These errors may, for example, represent sampling errors or the approximation errors involved in computing K and y by means of a Galerkin-type method where some "local" integration is performed. The noise is not "white" as in the first two model problems; rather, e has larger components along the singular vectors corresponding to the larger singular values.

We carried out several experiments with this third model problem for various values of μ. For all these experiments, the GCV method completely failed to compute a reasonable regularization parameter. Fig. 5 shows a typical GCV function for these experiments, for the particular choice $\mu = 0.05$. The GCV function is monotonically

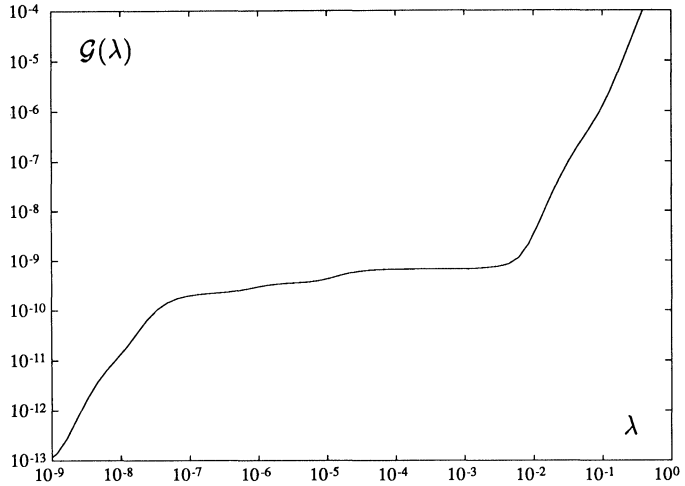

FIG. 5. (a) *shows the* GCV *function for a problem with correlated errors. The* GCV *function has no minimum for any reasonable value of* λ. (b) *shows the L-curve for Tikhonov regularization for the same problem as the corner computed by our algorithm. The numbers are the regularization parameters that correspond to the dots on the L-curve.*

increasing. (In fact, there is a minimum for λ of the order of the machine precision, but this is not a useful regularization parameter.)

The L-curve, on the other hand, always has an unmistakable corner. For the same value of $\mu = 0.05$ as above, this corner occurs at $\lambda \approx 3 \cdot 10^{-4}$, and this value of the regularization parameter indeed produces a regularized solution with optimal error. There is, in fact, one more corner for λ of the order of the machine precision outside of the plot. To get the correct corner, we used $\bar{M} = 10^4$ as a huge overestimate of the solution norm. For smaller values of μ the L-curve criterion sometimes leads to an underregularized solution, essentially because, as we mentioned in §4.1, the perturbation errors must become slightly apparent in the solution to produce the corner. Nevertheless, the computed λ is still fairly close to the optimal one.

7. Conclusions. We have shown that a number of regularization methods have naturally associated L-curves defined in terms of norms that are characteristic for the particular method. We have introduced new regularization methods, based on l_p norms in the coordinate system of the singular vectors of the matrix. Moreover, we have shown that, when plotted in a log-log scale, L-curves indeed have a characteristic L-shaped appearance and that the corner corresponds to a good choice of the regularization parameter.

Based on this characterization of the L-curves, we have proposed a new a posteriori scheme for computing the regularization parameter for a given problem. This scheme uses the parameter corresponding to a corner of the L-curve, a point of maximum curvature. We have also extended this idea to discrete L-curves such as those associated with truncated SVD and iterative methods.

Our numerical examples clearly illustrate the usefulness of the L-curve criterion for choosing the regularization parameter. Although the L-curve criterion sometimes fails to compute a reasonable regularization parameter, it seems to be much more robust than its main competitor, generalized cross-validation. Of course, one can always construct problems that will also "fool" the L-curve criterion; but it is our feeling that it works so well in practice that it is indeed a useful method. Further work is needed to determine circumstances under which the regularized solution converges to the true solution as the size of the error converges to zero.

REFERENCES

[1] M. BERTERO, C. DE MOL, AND E. R. PIKE, *Applied inverse problems in optics*, in Inverse and Ill-Posed Problems, Heinz W. Engl and C. W. Groetsch, eds., Academic Press, New York, 1987, pp. 291–313.

[2] L. ELDÉN, *Algorithms for regularization of ill-conditioned least squares problems*, BIT, 17 (1977), pp. 134–145.

[3] G. FARIN, *Curves and Surfaces for Computer Aided Geometric Design*, Academic Press, New York, 1988.

[4] G. H. GOLUB, M. HEATH, AND G. WAHBA, *Generalized cross-validation as a method for choosing a good ridge parameter*, Technometrics, 21 (1979), pp. 215–223.

[5] C. W. GROETSCH, *The Theory of Tikhonov Regularization for Fredholm Integral Equations of the First Kind*, Pitman, Boston, 1984.

[6] C. W. GROETSCH AND C. R. VOGEL, *Asymptotic theory of filtering for linear operator equations with discrete noisy data*, Math. Comput., 49 (1987), pp. 499–506.

[7] P. C. HANSEN, *The truncated SVD as a method for regularization*, BIT, 27 (1987), pp. 354–553.

[8] ———, *Regularization tools, a Matlab package for analysis of discrete regularization problems*, Tech. Report, Danish Computing Center for Research and Education, Lyngby, Denmark, 1991.

[9] ———, *Analysis of discrete ill-posed problems by means of the L-curve*, SIAM Rev., 34 (1992), pp. 561–580.

[10] C. L. LAWSON AND R. J. HANSON, *Solving Least Squares Problems*, Prentice-Hall, Englewood Cliffs, NJ, 1974.

[11] F. NATTERER, *The Mathematics of Computerized Tomography*, Wiley, New York, 1986.

[12] C. C. PAIGE AND M. A. SAUNDERS, *LSQR: An algorithm for sparse equations and sparse least squares*, ACM Trans. Math. Software, 8 (1982), pp. 43–71.

[13] J. A. SCALES AND A. GERSZTENKORN, *Robust methods in inverse theory*, Inverse Problems, 4 (1988), pp. 1071–1091.

[14] C. B. SHAW, JR., *Improvements of the resolution of an instrument by numerical solution of an integral equation*, J. Math. Anal. Appl., 37 (1972), pp. 83–112.

[15] J. SKILLING AND S. F. GULL, *Algorithms and applications*, in Maximum-Entropy and Bayesian Methods in Inverse Problems, C. R. Smith and W. T. Grandy, Jr., eds., D. Reidel Pub. Co., Boston, 1985, pp. 83–132.

[16] A. N. TIKHONOV AND V. Y. ARSENIN, *Solutions of Ill-Posed Problems*, Wiley, New York, 1977.

[17] J. M. VARAH, *Pitfalls in the numerical solution of linear ill-posed problems*, SIAM J. Sci. Statis. Comput., 4 (1983), pp. 164–176.

[18] G. WAHBA, *Spline Models for Observational Data*, CBMS-NSF Regional Conference Series in Applied Mathematics, Vol. 59, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1990.