Electronic Transactions on Numerical Analysis.
Volume 16, pp. 107-128, 2003.

ISSN 1068-9613.

ETNA
Kent State University
etna@mcs.kent.edu

A PARAMETER CHOICE METHOD FOR TIKHONOV REGULARIZATION*

LIMIN WU†

Abstract. A new parameter choice method for Tikhonov regularization of discrete ill-posed problems is presented. Some of the regularized solutions of a discrete ill-posed problem are less sensitive than others to the perturbations in the right-hand side vector. This method chooses one of the insensitive regularized solutions using a certain criterion. Numerical experiments show that the new method is competitive with the popular L-curve method. An analysis of the new method is given for a model problem, which explains how this method works.

Key words. discrete ill-posed problems, discrete Picard condition, Tikhonov regularization.

AMS subject classifications. 65F22.

1. Introduction.

This paper is concerned with solving the following algebraic least squares problem

$$\min_x \|Ax - b\|,\ A \in \mathbb{R}^{m\times n},\ m \geq n,\ b \in \mathbb{R}^m, \tag{1.1}$$

when it is ill-posed, that is, when the condition number of A is large and the singular values of A decay gradually towards zero. Here and henceforth, $\|\cdot\|$ is used to denote the 2-norm of a vector or a function. Ill-posed algebraic least squares problems, which are also called discrete ill-posed problems, arise frequently from the discretization of ill-posed problems such as Fredholm integral equations of the first kind. Let A have the singular value decomposition $A = \sum_{i=1}^n \sigma_i u_i v_i^T$. Then the least squares solution to (1.1) and the 2-norm of the solution are given by, respectively,

$$x_{LS} = \sum_{i=1}^n \frac{u_i^T b}{\sigma_i} v_i, \tag{1.2}$$

and

$$\|x_{LS}\| = \left(\sum_{i=1}^n \left(\frac{u_i^T b}{\sigma_i}\right)^2\right)^{1/2}. \tag{1.3}$$

Our goal is to compute a good estimate of the solution to (1.1). For discrete ill-posed problems, the formula in (1.2) may not be used for this purpose. The reason is that x_{LS} is too sensitive to the perturbations in b. That is, a small change in b can produce a large change in x_{LS}. To deal with ill-posedness in the presence of perturbations in b, various methods of regularization have been introduced. A comprehensive presentation of such methods can be found in [11]. One of these methods is the well-known Tikhonov regularization. This method attempts to provide a good estimate of the solution to (1.1) by a solution x_λ of the problem

$$\min_x \left\{\|Ax - b\|^2 + \lambda^2\|x\|^2\right\}, \tag{1.4}$$

for some positive λ value chosen in such a way that both the residual norm $\|Ax_\lambda - b\|$ and the solution norm $\|x_\lambda\|$ are made small simultaneously. In the literature, λ is called the

*Received June 24, 2002. Accepted for publication February 20, 2003. Recommended by Daniela Calvetti.
†Department of Applied Mathematics, Florida Institute of Technology, Melbourne, FL 32901-6988, U.S.A. E-mail: limin.wu@noaa.gov

ETNA
Kent State University
etna@mcs.kent.edu

regularization parameter and x_λ the regularized solution. The success of Tikhonov regularization of a discrete ill-posed problem depends on making a good choice of the regularization parameter. The above minimization problem is equivalent to

$$\min_x \left\| \begin{bmatrix} A \\ \lambda I \end{bmatrix} x - \begin{bmatrix} b \\ 0 \end{bmatrix} \right\|^2, \tag{1.5}$$

which has the least squares solution

$$x_\lambda = \sum_{i=1}^{n} \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2} \frac{u_i^T b}{\sigma_i} v_i. \tag{1.6}$$

From (1.2) and (1.6) it is obvious that $x_\lambda \to x_{LS}$ as $\lambda \to 0$.

2. Motivation. In this paper, I propose a new method for choosing a good regularization parameter λ. To motivate the method, we shall take a look at the Picard condition for Fredholm integral equations of the first kind, and then the discrete Picard condition for discrete least squares problems.

A Fredholm integral equation of the first kind is of the form

$$\int_c^d k(x,y) f(y) dy = g(x), \ \ a \le x \le b, \tag{2.1}$$

where k and g are given functions, and f is an unknown function. The function k is called the kernel of equation (2.1). This equation can be written as $Kf = g$, with K the integral operator mapping f to g. Assume that the kernel is square integrable over $[a,b] \times [c,d]$. Then it is a classical result that K is a compact operator from Hilbert space $L^2[c,d]$ into Hilbert space $L^2[a,b]$. Also, the kernel k has the singular value expansion $k(x,y) = \sum_{i=1}^{\infty} \eta_i p_i(x) q_i(x)$. According to Picard's theorem [3, Theorem 1.2.6], in order for (2.1) to have a solution $f \in L^2[c,d]$, it is necessary and sufficient that $g \in \overline{Range(K)}$ and

$$\sum_{i=1}^{\infty} \frac{|(p_i, g)|^2}{\eta_i^2} < \infty, \tag{2.2}$$

where $(\cdot,\cdot)$ denotes the inner product associated with $L^2[a,b]$. The condition given in (2.2) is known as the Picard condition. When $g \in Range(K)$, the Picard condition is necessarily satisfied and a solution of (2.1) is given by

$$\sum_{i=1}^{\infty} \frac{(p_i, g)}{\eta_i} q_i. \tag{2.3}$$

If $Range(K)$ is not dense in $L^2[a,b]$, and $g = g_1 + g_2$ with $g_1 \in Range(K)$ and $g_2 \neq 0$ in the orthogonal complement of $Range(K)$, then equation (2.1) has least squares solutions [3, Theorem 1.3.1]. Let f_n denote $\sum_{i=1}^{n} \frac{(p_i,g)}{\eta_i} q_i$ and s_n denote $\sum_{i=1}^{n} \frac{|(p_i,g)|^2}{\eta_i^2}$. Under the assumptions on g, the minimum norm least squares solution f_{LS} of equation (2.1) is given by

$$f_{LS} = \lim_{n\to\infty} f_n; \tag{2.4}$$

see [3, Theorem 1.3.4]. Since $\{q_i\}$ is an orthonormal sequence in $L^2[c,d]$ and $f_{LS} \in L^2[c,d]$, it follows that

$$\|f_{LS}\|^2 = \lim_{n\to\infty} s_n. \tag{2.5}$$

If we draw a piece-wise linear curve connecting the points $\{(n, s_n), n = 1, 2, ...\}$, the following is evident: $\|f_{LS}\|^2$ is the horizontal asymptote for the curve and the curve becomes flatter as n gets larger. In particular, since $\{s_n\}$ is a Cauchy sequence, given any $\epsilon > 0$, there exists an integer $M(\epsilon)$ such that for any $k > M(\epsilon)$, we have $s_{k+1} - s_k < \epsilon$. Note that the slope of the line segment connecting (n, s_n) and $(n + 1, s_{n+1})$ is given by $(s_{n+1} - s_n)/1$. Now we see that a flatter point on the curve corresponds to an f_n closer to f_{LS}.

The analog of the Picard condition for discrete least squares problems is the discrete Picard condition (DPC), recognized by Varah [14] and analyzed by Hansen [6, 7].

DEFINITION 2.1. *We say that the right-hand side vector b of the problem in (1.1) satisfies the DPC if the Fourier coefficients $|u_i^T b|$ decay, on the average, to zero faster than the singular values σ_i; that is, the sequence*

$$|u_1^T b|/\sigma_1, |u_2^T b|/\sigma_2, \ldots, |u_n^T b|/\sigma_n, \tag{2.6}$$

generally decreases (occasional exceptions allowed).

The above definition of the DPC represents an attempt to place a requirement on the right-hand side vector b in (1.1) which is somewhat analogous in finite-dimensional space to the requirement on the right-hand side function g in the above Picard condition (2.2) in infinite-dimensional Hilbert space. Of course, the infinite sum in (2.2) must converge, which means that the terms in the sum must decay to zero; so if problem (1.1) is obtained from the Fredholm integral equation (2.1) by some method of discretization, it makes sense that the corresponding discretized quantities in (2.6) should have the kind of behavior that would mimic the decay expressed in (2.2).

The method of truncated singular value decomposition (TSVD) can be used to regularize the discrete ill-posed problems; see [8]. This method amounts to truncating the singular value decomposition of the coefficient matrix A in such a way that the smallest singular values of A are discarded, and then solving the modified least squares problem. A TSVD solution of (1.1) can be written as

$$x^{(k)} = \sum_{i=1}^{k} \frac{u_i^T b}{\sigma_i} v_i, \; 1 \le k \le n, \tag{2.7}$$

and

$$\|x^{(k)}\|^2 = \sum_{i=1}^{k} (u_i^T b/\sigma_i)^2. \tag{2.8}$$

The integer k is called the truncation parameter. For a discrete least squares problem that satisfies the DPC, the piece-wise linear curve connecting the points $\{(k, \|x^{(k)}\|^2)\}$ is in general flatter for larger k, since most of the time, we have

$$\|x^{(m)}\|^2 - \|x^{(m-1)}\|^2 = (u_m^T b/\sigma_m)^2 < \|x^{(l)}\|^2 - \|x^{(l-1)}\|^2 = (u_l^T b/\sigma_l)^2,$$

if $m > l$. Note also that $x^{(k)}$ gets closer to x_{LS} as k gets larger since $\|x_{LS} - x^{(k)}\|^2 = \sum_{i=k+1}^{n} (u_i^T b/\sigma_i)^2$. In the TSVD method, when b satisfies the DPC, the situation is therefore similar to that in the infinite-dimensional Hilbert space expressed in (2.5); namely, most of the time, a flatter point on the curve corresponds to an $x^{(k)}$ closer to x_{LS}.

We now ask the following question: If b satisfies the DPC, what would be a proper curve to be chosen in the Tikhonov regularization that would promise "the flatter a point on the curve, the closer its corresponding x_λ to x_{LS}"? If we have such a curve, the above discussion suggests that we should pick a value of the regularization parameter corresponding to a point

ETNA
Kent State University
etna@mcs.kent.edu

on the flattest portion of the curve. In real problems, typically, the DPC is only partially satisfied, that is, the sequence $\{|u_i^T b|/\sigma_i\}$ decreases for a while and then begins to increase after some point $i = p < n$. This increase apparently represents "noise" resulting from the fact that the discretization process, or the measurement process, has produced a b-vector that does not entirely satisfy the DPC. This situation raises the most essential question: How should we choose the regularization parameter for problems that only partially satisfy the DPC?

3. The DPC at work. Two examples are given here to show the critical role the DPC plays for a discrete ill-posed problem to have a reasonable solution. In this section, "exact solution" is used to refer to an actual solution of an unperturbed algebraic problem, $Ax = b$.

We shall use the following as a definition of optimal regularization parameter.

DEFINITION 3.1. *For discrete ill-posed problems which have some solution that can be regarded as more or less "exact", we define the optimal value of the regularization parameter λ to be the one for which the computed relative error $\|x_\lambda - x^{exact}\|/\|x^{exact}\|$ is the smallest.*

This definition seems to conform with the common practice in the study of regularization problems; see Hansen [10]. Clearly, the aim of any parameter choice method ought to be able to select the optimal value of λ, if the problem has an acceptable exact solution.

Example 1. Consider the least squares problem that has $A = diag(1, 10^{-5}, 10^{-10})$ and $b = (1, 10^{-4}, 10^{-8})^T$. The system has the exact solution $x^{exact} = (1, 10, 100)^T$, that is, x^{exact} satisfies $Ax^{exact} = b$. The sequence in (2.6) for this problem is $1, 10, 100$, so the problem does not satisfy the DPC. Now we add a small perturbation $e = (10^{-6}, 10^{-6}, 10^{-6})^T$ to b ($\|e\| = \sqrt{3} \cdot 10^{-6}, \|b\| \approx 1$) to obtain a perturbed problem. The sequence in (2.6) for the perturbed problem is $1 + 10^{-6}, 10.1, 10100$, so the perturbed problem does not satisfy the DPC either. The TSVD solutions of the perturbed problem are:

$$x^{(1)} = (1 + 10^{-6}, 0, 0)^T, x^{(2)} = (1 + 10^{-6}, 10.1, 0)^T, x^{(3)} = (1 + 10^{-6}, 10.1, 10100)^T.$$

Their relative errors are:

$$\frac{\|x^{(1)} - x^{exact}\|}{\|x^{exact}\|} \approx 1.00, \frac{\|x^{(2)} - x^{exact}\|}{\|x^{exact}\|} \approx 0.995, \frac{\|x^{(3)} - x^{exact}\|}{\|x^{exact}\|} \approx 99.5.$$

We see that all three TSVD solutions have large relative error. This example shows that for a discrete ill-posed problem that does not satisfy the DPC, its actual solution may not be obtainable from its perturbed problem.

Example 2. Consider the least squares problem that has $A = diag(1, 10^{-5}, 10^{-10})$ and $b = (1, 10^{-7}, 10^{-14})^T$. The exact solution of the system is $x^{exact} = (1, 10^{-2}, 10^{-4})^T$. The sequence in (2.6) for this problem is $1, 10^{-2}, 10^{-4}$, so the problem satisfies the DPC. Again we add a small perturbation $e = (10^{-7}, 10^{-7}, 10^{-7})^T$ to b ($\|e\| = \sqrt{3} \cdot 10^{-7}, \|b\| \approx 1$) to obtain a perturbed problem. The sequence in (2.6) for the perturbed problem is $1 + 10^{-7}, 0.02, 10^{-4} + 10^3$, which says that the perturbed problem partially satisfies the DPC. The TSVD solutions of the perturbed problem are:

$$x^{(1)} = (1+10^{-7}, 0, 0)^T, x^{(2)} = (1+10^{-7}, 0.02, 0)^T, x^{(3)} = (1+10^{-7}, 0.02, 10^{-4}+10^3)^T.$$

Their relative errors are:

$$\frac{\|x^{(1)} - x^{exact}\|}{\|x^{exact}\|} \approx 0.01, \frac{\|x^{(2)} - x^{exact}\|}{\|x^{exact}\|} \approx 0.01, \frac{\|x^{(3)} - x^{exact}\|}{\|x^{exact}\|} \approx 1000.$$

ETNA
Kent State University
etna@mcs.kent.edu

We see that the first two TSVD solutions have small relative error but the third one has large relative error. In this example, we are able to compute good estimates of the solution to the original system from the perturbed system, when the DPC is satisfied by the original system and only partially satisfied by the perturbed system. This situation contrasts with that of example 1 where reasonable TSVD solutions for the perturbed problem are not available.

4. Experimenting on curve $(\ln(1/\lambda), \|x_\lambda\|)$**.** For the Tikhonov regularization, we seek to obtain a curve similar to those discussed in §2 that would have a flat portion where a good value of the regularization parameter can be located. After trying several possible curves on the 11 problems in Hansen's Regularization Tools package [10], I found that the plot of $(\ln(1/\lambda), \|x_\lambda\|)$ provided a good choice for use on all of the problems. It is possible that different choices of scaling of the λ and $\|x_\lambda\|$ variables would work better for this set of problems or problems from other sources. I shall examine the behavior of the curve through the study of two examples, with a view to locating a good λ value.

Example 1. Let A be the 20×20 Hilbert matrix, that is,

$$A = \begin{bmatrix} 1/1 & 1/2 & \cdots & 1/n \\ 1/2 & 1/3 & \cdots & 1/(n+1) \\ \vdots & \vdots & & \vdots \\ 1/n & 1/(n+1) & \cdots & 1/(2n-1) \end{bmatrix}.$$

Let x^{exact} be the vector of 20 elements with its ith element $x^{exact}(i) = \sqrt{0.5i}$, $i = 1, 2, \ldots, 20$. Consider the problem

$$Ax = b, \tag{4.1}$$

where b is produced by $b = Ax^{exact}$ in double precision. It is known that the singular value spectrum of a Hilbert matrix decays toward zero [12]. The computed condition number of A is on the order of 10^{+18}, which is very large with respect to the double precision used in the computation. We therefore have a discrete ill-posed problem. We now study the qualitative behavior of the curves given by

$$\left\{(k, \|x^{(k)}\|); k = 1, 2, \ldots, 20\right\} \quad \text{and} \quad \{(\ln(1/\lambda), \|x_\lambda\|); \lambda > 0\},$$

where $x^{(k)}$ is the kth TSVD solution of (4.1), and x_λ is the regularized solution of (4.1) as a function of λ.

In Figure 1, the graph of $\|x^{(k)}\|$ versus the truncation parameter k is plotted using the MATLAB "plot" command. Rising a little around $k = 2$, this curve quickly levels off as it moves to the right. But at about $k = 14$, it begins to grow very rapidly. At $k = 14$, just before the steep rise, $\|x^{14}\| \approx 10.263$, which is slightly larger than $\|x^{exact}\| \approx 10.247$. The optimal value of the truncation parameter k is found to be 11, at which $\|x^{(11)}\| \approx 10.247$ and the relative error is about 4.3×10^{-6}, the smallest among all the relative errors computed by $\|x^{(k)} - x^{exact}\|/\|x^{exact}\|$, $k = 1, \ldots, 20$..

The graph of $(\ln(1/\lambda), \|x_\lambda\|)$ in Figure 2 is plotted using the same MATLAB command "plot" over a mesh grid of 100 points. The mesh grid is constructed using the method employed in the MATLAB code $l\_curve.m$ of Hansen [10]. The method can be described as follows. Let $\lambda_1 = \sigma_{max}(A)$ and λ_l be positive but smaller than $\max\{\sigma_{min}(A),\ \sigma_{max}(A) \cdot \epsilon\}$, where ϵ is the machine roundoff unit, $\sigma_{max}(A)$ and $\sigma_{min}(A)$ are the largest and smallest singular values of A. We want to fill $l - 2$ numbers $\lambda_2, \ldots, \lambda_{l-l}$ between λ_1 and λ_l. Let $\mu_i = \ln(1/\lambda_i)$, $i = 1, 2, \ldots, l$. Since $\ln x$ is an increasing function of x,

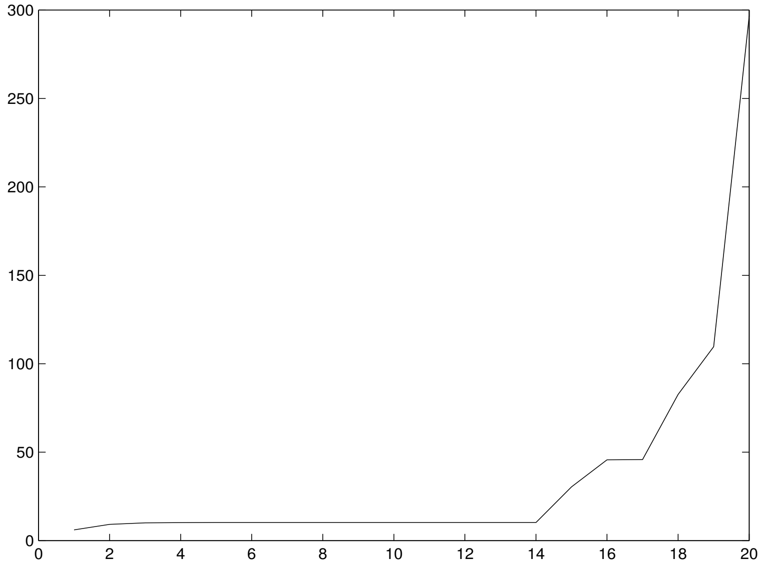

FIG. 1. *Truncation Parameter k versus* $\|x^{(k)}\|$

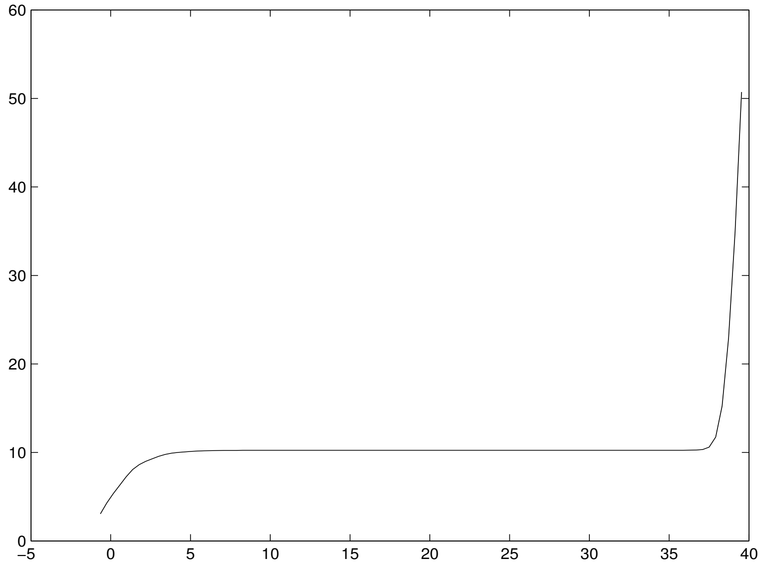

FIG. 2. *The graph of* $(\ln(1/\lambda), \|x_\lambda\|)$

we have $\mu_1 < \mu_2 < \cdots < \mu_l$. Let $h = (\mu_l - \mu_1)/(l-1) = \ln{(\lambda_1/\lambda_l)}^{l-1}$. Put $\mu_i = \mu_1 + (i-1)h,\ i = 1, \ldots, l$. Then μ_i's form a uniform mesh grid. Converting μ_i back to λ_i, we obtain a mesh grid for λ: $\lambda_1 > \lambda_2 > \cdots > \lambda_l$, with

$$\lambda_i = \lambda_1 \left(\frac{\lambda_l}{\lambda_1}\right)^{(i-1)/(l-1)},\ i = 1, 2, \ldots, l.$$

Table 1 gives some of the 100 data points used for the plot in Figure 2 along with some other quantities. Here, the fourth column lists $\|x_{\lambda_i} - x^{exact}\|/\|x^{exact}\|$, and the fifth column lists $\|x_{\lambda_i}\| - \|x_{\lambda_{i-1}}\|$. Notice that $(\|x_{\lambda_i}\| - \|x_{\lambda_{i-1}}\|)/h$ gives an approximate slope of the curve in Figure 2 at μ_i. It can be observed that the smallest difference and therefore the minimum slope over the mesh grid occurs at index $i = 67$, at which $\lambda_{67} \approx 4.4 \times 10^{-12}$. It can be seen that the relative error reaches its minimum value 6.15×10^{-6} at $\lambda_{66} \approx 6.7 \times 10^{-12}$. For this problem, we may regard λ_{66} as the optimal value of the regularization parameter for this choice of the λ grid.

TABLE 1
Regularization Data for Hilbert Matrix

i	λ_i	$\|x_{\lambda_i}\|$	Re. Error	Diff
57	2.57e-10	1.0246951e+01	2.69e-05	1.25e-08
58	1.71e-10	1.0246951e+01	2.28e-05	8.08e-09
59	1.14e-10	1.0246951e+01	2.14e-05	4.33e-09
60	7.60e-11	1.0246951e+01	2.04e-05	2.11e-09
61	5.06e-11	1.0246951e+01	1.87e-05	1.01e-09
62	3.37e-11	1.0246951e+01	1.59e-05	5.75e-10
63	2.25e-11	1.0246951e+01	1.22e-05	5.17e-10
64	1.50e-11	1.0246951e+01	8.74e-06	5.75e-10
65	9.99e-12	1.0246951e+01	6.57e-06	5.16e-10
66	6.66e-12	1.0246951e+01	6.15e-06	3.53e-10
67	4.44e-12	1.0246951e+01	7.40e-06	2.27e-10
68	2.95e-12	1.0246951e+01	1.09e-05	2.75e-10
69	1.97e-12	1.0246951e+01	1.83e-05	8.49e-10
70	1.31e-12	1.0246951e+01	3.24e-05	3.09e-09
71	8.75e-13	1.0246951e+01	5.46e-05	9.01e-09
72	5.83e-13	1.0246951e+01	8.18e-05	1.78e-08
73	3.88e-13	1.0246951e+01	1.06e-04	2.23e-08
74	2.59e-13	1.0246951e+01	1.22e-04	1.86e-08

The above two graphs look very much alike and the one in Figure 2 is smoother. In both graphs, a corner divides the curve into two pieces. The left piece is flat and more or less concave down; the right piece is very steep. The flat portion of the curves gets flatter and flatter as k or λ approaches the optimal truncation parameter or the optimal regularization parameter (for this particular λ grid), respectively.

Example 2. This example is the inverse heat equation

$$y(s) = \int_0^s k(s-t)x(t)dt, \ 0 < s \leq 1,$$

taken from [10]. The mathematical background associated with this problem is given in [2]. This is a Volterra integral equation of the first kind with $[0, 1]$ as the integration interval. The kernel is $k(s-t)$ with

$$k(t) = \frac{t^{-3/2}}{2\sqrt{\pi}} \exp(\frac{-1}{4t}).$$

The integral equation is discretized by means of simple collocation and the midpoint rule to obtain the coefficient matrix A. The size of A is taken to be 100×100. An exact solution, x^{exact}, is constructed, and then the right-hand side is produced as $b = Ax^{exact}$ in double precision. The matrix A, the exact solution x^{exact} and the right-hand side b used for the testing are generated by running Hansen's MATLAB code heat$(100, 1)$.

To see how perturbations on the right-hand side influence the regularized solutions, the right-hand side b has been modified by adding a normally distributed random vector e^δ to produce $b^\delta = b + e^\delta$. Here δ is used to indicate perturbation level in b, which is given by $\|e^\delta\|/\|b\|$. Each e^δ was generated by a call to MATLAB routine "randn" with seed = 15. The components of e^δ are normally distributed with a mean of zero and a standard deviation of one. The systems solved are $Ax = b^\delta$.

In Figure 3, the curve $(\ln(1/\lambda), \ln\|x_\lambda(b^\delta)\|)$ is plotted over a mesh grid of 100 points ($\lambda_1 \approx 3.6 \times 10^{-1}, \lambda_{100} \approx 1.3 \times 10^{-18}$) for δ at 10^{-2} and 10^{-3}, where $x_\lambda(b^\delta)$ is the regularized solution of $Ax = b^\delta$. Note that log-scale on the $y-$axis is used so that the different behavior of the two curves can be clearly displayed over the whole mesh grid. Both curves exhibit two flat portions. The left flat portion is less sensitive to the perturbation than the right one in that these two curves almost overlap for the left flat portion while they differ noticeably for the right flat portion. For each of these two curves an optimal value of λ is sought to produce a regularized solution and we wish the "optimal" regularized solutions computed from these two perturbed systems would be both close to the exact solution of the unperturbed system; it is therefore reasonable to choose a λ_i, for each system, that corresponds to a point on the left flat portion of the curve.

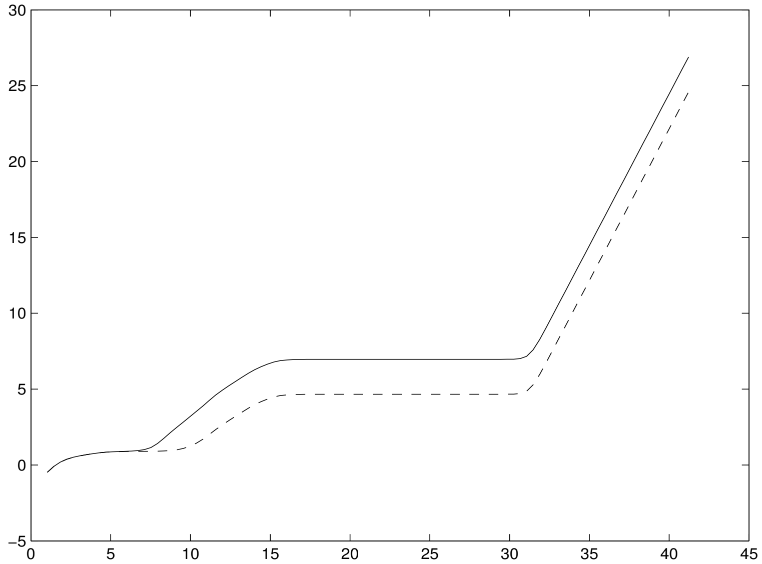

FIG. 3. *The graph of* $(\ln(1/\lambda), \ln(\|x_\lambda\|))$ *for problem heat(100,1) with perturbation level* δ *at* 0.01 *(solid line) and* 0.001 *(dashed line).*

5. The flattest slope method. From example 2 in §4, we see that x_λ is not equally sensitive to the changes in the right-hand side vector for all λ values. I shall loosely call the part of the curve $(\ln(1/\lambda), \|x_\lambda\|)$ that varies little under small perturbations to the right-hand side the insensitive part of the curve.

5.1. The Method. The proper choice of λ is critical to the success of the Tikhonov regularization method. A good λ value should be the one that suppresses as much as possible the influence of the errors in the right-hand side, and at the same time, gives a regularized solution x_λ that is as close as possible to the exact solution. The examples discussed in the previous sections suggest that we can locate a good λ somewhere corresponding to a point on the insensitive portion, which should look flat, of the curve $(\ln(1/\lambda), \|x_\lambda\|)$ just before the rapid growth. The new method, first given by the author in [16], for choosing a good value for the regularization parameter, is as follows.

1. Detect the insensitive portion, that is adjacent to the sensitive portion, of the curve $(\ln(1/\lambda), \|x_\lambda\|)$.
2. Choose a value of λ that corresponds to the point on the insensitive portion of the curve at which the slope is the flattest.

For convenience, I shall call the above parameter choice method "flattest slope method" or "f-slope method" hereafter.

5.2. Practical considerations. To find a point with a slope close to the flattest slope on the insensitive portion of the curve, we may use the method given in §4. The insensitive portion of the curve may be detected graphically, as the plots in §4 suggest. I shall consider two cases. If there is only one flat portion on the curve (eventually, $\|x_\lambda\|$ will level off as λ approaches 0, but that flat portion will be ignored), I simply accept it as an insensitive portion. If there are more than one flat portions on the curve, more information is needed to rule out the sensitive pieces. To do so, the problem can be altered a little by adding an artificial error vector e^δ to b, with error level $\|e^\delta\|/\|b\|$ slightly greater than the error level that comes with the problem, and then plot $(\ln(1/\lambda), \|x_\lambda\|)$ for b and $b+e^\delta$ against each other for comparison to determine the insensitive portion. Or, if the right-hand side b can be sampled twice for the same problem, we can plot $(\ln(1/\lambda), \|x_\lambda\|)$ for the two measurements against each other.

An implementation of the method can be summarized as follows: Compute all the differences $\|x_{\lambda_i}\| - \|x_{\lambda_{i-1}}\|$ over a specified mesh grid, identify the insensitive portion of the curve $(\ln(1/\lambda), \|x_\lambda\|)$, and choose the λ_i on the insensitive portion where the minimum difference (local to the insensitive portion) occurs.

6. Numerical examples. A popular method for choosing a proper regularization parameter is the L-curve criterion given by Hansen and O'Leary [13]. The method attempts to balance decreasing values of $\|Ax_\lambda - b\|$ against increasing values of $\|x_\lambda\|$ as λ tends to zero. For a discrete ill-posed problem it turns out that the plot of $\ln(\|Ax_\lambda - b\|)$ against $\ln(\|x_\lambda\|)$ has the shape of the letter L, with a distinct corner separating the vertical and horizontal parts of the curve. The L-curve method chooses the regularization parameter corresponding to the point on the curve with maximum curvature. For background material and underlying mathematics associated with the L-curve method, see also [9].

In this section, numerical results are given to compare the flattest slope method with the L-curve method against the optimal choice of the regularization parameter on several problems. Some MATLAB M-files for the computations are available for download at website

`http://my.fit.edu/beowulf`

under the Software link. Hansen generated a comprehensive MATLAB package implementing several regularization methods in [10]. I rely on this work for the numerical tests.

6.1. Test problems from Hansen's package. Table 2 displays data for the nine problems from Hansen's MATLAB package for which the exact solutions are known. In each case, the size of the coefficient matrix A is taken as 20×20. For problems **heat** and **shaw**, the exact solution x^{exact} given by Hansen is used and the right-hand side vector b is generated by $b = Ax^{exact}$. For problems **baart**, **deriv2**, **foxgood**, **ilaplace**, **phillips**, and **wing**, the choice of the exact solution x^{exact} is generally some discretized version of an exact solution for an integral equation, and in these cases both A and b are taken as discretized versions of an infinite-dimensional problem. For all the above test problems, a mesh grid of 100 points is used. Comparisons are made for the regularized solutions chosen by the L-curve method and the f-slope method with the optimal regularized solution that produces the minimum relative error over the mesh grid.

TABLE 2
Regularization Data for Nine Problems from Hansen's MATLAB Package

Problem	Method	Rerr	Res	λ
Baart(20)	L-curve	2.69e-01	2.20e-15	2.24e-15
	f-slope	1.57e-02	2.79e-15	4.39e-13
	Optimal	1.10e-02	2.55e-15	5.76e-14
deriv2(20,3)	L-curve	6.01e-03	4.61e-13	2.08e-07
	f-slope	6.01e-03	4.61e-13	2.08e-07
	Optimal	3.83e-03	1.22e-06	3.91e-04
foxgood(20)	L-curve	2.00e+01	4.16e-14	8.45e-16
	f-slope	6.46e-03	4.83e-08	6.27e-06
	Optimal	2.31e-03	6.59e-06	5.45e-04
heat(20,1)	L-curve	1.08e-14	1.36e-16	1.44e-11
	f-slope	1.04e-14	1.46e-16	3.69e-10
	Optimal	5.91e-15	2.49e-14	1.87e-09
ilaplace(20,1)	L-curve	3.08e-01	2.38e-04	1.39e-04
	f-slope	8.50e-03	3.11e-04	5.36e-03
	Optimal	4.80e-03	7.46e-04	1.81e-02
phillips(20)	L-curve	1.87e-02	7.39e-12	1.46e-06
	f-slope	1.87e-02	7.39e-12	1.46e-06
	Optimal	1.84e-02	8.81e-04	2.70e-02
shaw(20)	L-curve	6.85e-02	3.36e-15	9.24e-16
	f-slope	1.24e-04	3.31e-15	4.07e-13
	Optimal	1.03e-04	2.94e-15	3.10e-12
spikes(20,5)	L-curve	1.62e-01	1.49e-14	7.34e-16
	f-slope	1.58e-01	3.53e-14	2.15e-13
	Optimal	1.57e-01	1.49e-14	1.65e-15
wing(20,1/3,2/3)	L-curve	3.26e-01	7.60e-17	2.07e-16
	f-slope	3.17e-01	8.89e-17	1.80e-14
	Optimal	3.16e-01	8.30e-17	3.55e-15

The third column in Table 2 gives the relative error Rerr = $\|x_\lambda - x^{exact}\|/\|x^{exact}\|$, the fourth column gives the residual norm Res = $\|Ax_\lambda - b\|$, and the last column gives the λ value for each method and for the optimal choice. For problems baart(20), foxgood(20), ilaplace(20,1), and shaw(20), the f-slope method performs significantly better. For the other five problems, the L-curve and f-slope methods produce almost the same relative errors.

6.2. The Hilbert matrix. This example consists of the 20×20 Hilbert matrix A with five different choices of the right-hand side vector b. The exact solution x^{exact} is generated by some mathematical formula as $x^{exact}(i) = f(t_i), t_i = 0.5i$. The right-hand side b is then produced as $b = Ax^{exact}$. Each test problem is run using the L-curve method and the f-slope method over a mesh grid of 50 points for comparison. The λ value from the L-curve method is generated by runs of the MATLAB codes in [10]. The results listed in Table 3 show that higher accuracy (smaller relative error) is obtained by the f-slope method for these problems.

TABLE 3
Regularization Data for Five Problems related to the 20×20 *Hilbert Matrix*

$x_{ex}(i)$	Method	Rerr	Res	λ
$t_i^{1/2}$	L-curve	5.56e-02	4.09e-15	1.80e-16
	f-slope	9.43e-06	3.95e-15	3.38e-12
	Optimal	6.11e-06	4.68e-15	7.67e-12
t_i^2	L-curve	2.68e-02	5.62e-14	1.80e-16
	f-slope	8.69e-05	5.74e-14	1.27e-13
	Optimal	8.26e-05	7.04e-14	3.38e-12
$1/(t_i^3 + t_i^2 + t_i + 1)$	L-curve	3.49e-02	2.58e-16	1.80e-16
	f-slope	5.12e-07	2.02e-16	1.74e-11
	Optimal	4.89e-07	2.44e-16	3.95e-11
$\sin(t_i)$	L-curve	6.65e-02	5.38e-16	7.93e-17
	f-slope	2.69e-03	5.18e-16	2.11e-15
	Optimal	2.05e-03	6.40e-16	4.78e-15
$\exp(t_i)$	L-curve	1.42e-02	5.01e-12	1.80e-16
	f-slope	4.65e-03	4.81e-12	9.28e-16
	Optimal	4.03e-03	5.01e-12	4.78e-15

It will be observed that in all five problems the relative errors for the f-slope method are smaller than those for the L-curve method, while the residual norms remain at about the same level for both methods. The relative errors obtained from the L-curve method remain at about 10^{-2} for all choices of the right-hand side vector, while those from the f-slope method have a wider variation from 2.7×10^{-3} in the fourth problem to 5.1×10^{-7} in the third problem. We also observe that those exact solutions with less variation in $x^{exact}(i)$ (e.g., the first three in Table 2) allow smaller relative error in the computed regularized solution.

7. Perturbation experiments. In this section, comparisons are made for the performance of the f-slope method with the L-curve method for various perturbation levels in the right-hand side b. For each test problem, the b-vector is modified by adding a normally distributed random vector e^δ to produce $b^\delta = b + e^\delta$. Ten different perturbation vectors are generated with decreasing perturbation levels in such a way that

$$\|e^\delta\|/\|b\| = 10^{-2}, 10^{-3}, \ldots, 10^{-10}, 10^{-12}. \tag{7.1}$$

The perturbations e^δ are generated by calls to MATLAB routine "randn" with seed = 15. The systems solved are $Ax = b^\delta$.

Example 1. The first example is the problem **heat** considered in §4. Recall that the size of A is taken to be 100×100 and the matrix A, the right-hand side b and the exact solution x^{exact} are generated by running Hansen's MATLAB code heat$(100, 1)$. Complete test results are given in Table 4. The data for the f-slope method reported in the table corresponds to the choice of the left flat portion of the curve $(\ln(1/\lambda), \ln\|x_\lambda\|)$ for perturbation levels $10^{-2}, 10^{-3}, \ldots$, and 10^{-7}. For smaller perturbation levels $10^{-8}, 10^{-9}, 10^{-10}$, and 10^{-12}, only one flat portion is identified from the curve.

As we can see from the table, when δ takes on the first four error levels $10^{-2}, 10^{-3}$, 10^{-4}, and 10^{-5} in decreasing order, the optimal relative error decreases strictly though

ETNA
Kent State University
etna@mcs.kent.edu

TABLE 4
Perturbation Experiment On Problem heat(100,1)

δ	L-curve Rerr	L-curve Res	L-curve λ	f-slope Rerr	f-slope Res	f-slope λ	Optimal Rerr	Optimal λ
10^{-2}	1.9e-01	3.9e-03	1.8e-03	1.3e-01	7.6e-04	2.7e-03	1.1e-01	4.1e-03
10^{-3}	3.3e-01	2.9e-04	1.6e-04	5.2e-02	7.6e-05	8.1e-04	4.0e-02	1.2e-03
10^{-4}	5.5e-01	1.3e-05	6.2e-06	2.7e-02	7.6e-06	2.4e-04	1.9e-02	5.4e-04
10^{-5}	3.9e-01	7.6e-07	1.6e-07	1.4e-02	7.6e-07	4.7e-05	1.2e-02	1.1e-04
10^{-6}	4.2e-02	7.6e-08	1.9e-09	6.7e-03	7.6e-08	9.3e-06	6.1e-03	2.1e-05
10^{-7}	4.2e-03	7.6e-09	1.1e-10	3.1e-03	7.6e-09	5.4e-07	2.9e-03	8.2e-07
10^{-8}	4.2e-04	7.6e-10	3.2e-11	4.2e-04	7.6e-10	9.5e-12	4.2e-04	1.4e-11
10^{-9}	4.2e-05	7.6e-11	9.5e-12	4.2e-05	7.6e-11	6.3e-12	4.2e-05	9.5e-12
10^{-10}	4.6e-06	7.6e-12	1.9e-12	4.2e-06	7.6e-12	2.8e-12	4.2e-06	6.3e-12
10^{-12}	4.3e-08	7.6e-14	1.6e-13	4.2e-08	7.6e-14	2.8e-12	4.2e-08	2.8e-12

slowly. The relative error corresponding to the f-slope method behave similarly with slightly larger values. On the other hand, for these four error levels, the relative error obtained from the L-curve method actually increases before they begin to decrease. Similar results for the present problem have also been reported by M. Hanke [4, p.299, Table 6.1]. As δ goes further down from 10^{-6} to 10^{-12}, the relative errors of the two methods that decrease and remain very close. We notice that the L-curve method tends to choose a smaller λ value than the optimal one.

Example 2. The second example is the problem **shaw** considered in the previous section. The A matrix is 100×100. The coefficient matrix A, the right-hand side b and the exact solution x^{exact} used for the test are generated by running Hansen's MATLAB code shaw(100). The mesh grid used for this test contains 100 points. Table 5 gives the numerical results for this problem. It shows that the overall performance of the f-slope method is better than that of the L-curve method. The relative error given by the f-slope method decreases strictly as δ tends to zero. The relative error given by the L-curve criterion, however, bounces up for δ at 10^{-6} and 10^{-10}.

TABLE 5
Perturbation Experiment On Problem shaw(100)

δ	L-curve Rerr	L-curve Res	L-curve λ	f-slope Rerr	f-slope Res	f-slope λ	Optimal Rerr	Optimal λ
10^{-2}	8.2e-02	2.1e-01	1.5e-02	1.0e-01	2.2e-01	5.2e-02	5.6e-02	2.3e-02
10^{-3}	4.9e-02	2.1e-02	2.0e-03	4.9e-02	2.1e-02	2.0e-03	4.8e-02	6.8e-03
10^{-4}	4.5e-02	2.1e-03	2.6e-04	3.8e-02	2.1e-03	4.0e-04	3.5e-02	8.9e-04
10^{-5}	2.9e-02	2.1e-04	5.2e-05	3.1e-02	2.1e-04	1.2e-04	2.9e-02	5.2e-05
10^{-6}	7.6e-02	2.1e-05	2.0e-06	2.7e-02	2.5e-05	1.2e-04	1.7e-02	1.5e-05
10^{-7}	1.8e-02	2.1e-06	1.8e-07	1.5e-02	2.2e-06	6.9e-06	7.0e-03	2.0e-06
10^{-8}	5.1e-03	2.1e-07	2.3e-08	4.6e-03	2.1e-07	3.5e-08	3.8e-03	2.7e-07
10^{-9}	4.9e-03	2.1e-08	4.6e-09	3.5e-03	2.1e-08	2.3e-08	3.0e-03	6.9e-09
10^{-10}	6.3e-02	2.1e-09	1.2e-10	3.2e-03	2.1e-09	1.6e-08	1.1e-03	3.1e-09
10^{-12}	5.3e-03	2.1e-11	7.0e-12	8.2e-04	2.1e-11	5.3e-11	7.3e-04	2.7e-10

ETNA
Kent State University
etna@mcs.kent.edu

TABLE 6
Perturbation Experiment for the Problem Associated with Hilbert Matrix

δ	L-curve			f-slope			Optimal	
	Rerr	Res	λ	Rerr	Res	λ	Rerr	λ
10^{-2}	1.4e-01	5.5e-01	1.1e-02	1.5e-01	5.6e-01	1.7e-02	1.4e-01	1.1e-02
10^{-3}	1.1e-01	5.3e-02	9.8e-04	1.1e-01	5.3e-02	9.8e-04	9.8e-02	5.0e-03
10^{-4}	5.2e-02	5.3e-03	1.3e-04	2.5e-02	5.3e-03	4.3e-04	1.5e-02	2.9e-04
10^{-5}	7.0e-02	5.1e-04	1.1e-05	6.7e-03	5.3e-04	8.6e-05	6.7e-03	8.6e-05
10^{-6}	4.5e-02	5.1e-05	9.9e-07	4.4e-03	5.2e-05	1.1e-05	4.4e-03	1.1e-05
10^{-7}	3.6e-02	5.1e-06	8.6e-08	2.1e-03	5.1e-06	2.2e-06	2.1e-03	2.2e-06
10^{-8}	2.0e-02	5.1e-07	1.1e-08	2.7e-03	5.1e-07	1.9e-07	1.6e-03	6.6e-07
10^{-9}	2.3e-02	5.1e-08	1.5e-09	8.8e-04	5.1e-08	5.8e-08	8.4e-04	8.6e-08
10^{-10}	4.9e-02	5.0e-09	1.3e-10	2.9e-04	5.1e-09	1.1e-08	2.8e-04	7.6e-09
10^{-12}	5.7e-02	4.9e-11	1.5e-12	1.3e-04	5.0e-11	6.6e-10	9.2e-05	1.5e-09

Example 3. In this example, the coefficient matrix A is the 100×100 Hilbert matrix. The right-hand side vector b is produced by $b = Ax^{exact}$, where $x^{exact}(i) = \sqrt{0.5i}, \; i = 1, 2, \ldots, 100$. The numerical results are reported in Table 6. We can see that as δ goes down from 10^{-2} to 10^{-12}, the relative error given by the f-slope method decreases from $1.5 \cdot 10^{-1}$ to $1.3 \cdot 10^{-4}$, while the relative error given by the L-curve criterion stagnates at the level of 10^{-2}.

8. An analysis of the f-slope method. The f-slope method works well for the test problems discussed in the previous sections. This section attempts to explain how the method works through an analysis of a model problem. The slope of $(\mu, \|x_\mu\|^2)$ will be studied, instead of the slope of $(\mu, \|x_\mu\|)$, which is used in the numerical experiments. A simple relation between $(\|x_\mu\|^2)'$ and $\|x_\mu\|'$ is $(\|x_\mu\|^2)' = 2\|x_\mu\|\|x_\mu\|'$. If $\|x_\mu\|$ changes little over a range for μ, then the two quantities should vary in a similar way over the range. In the case that a problem partially satisfying the DPC, the $\|x_\mu\|$ would stay almost the same over the range where the optimal μ value is located, as we have seen in the test problems. This relationship then implies that the flattest point on both curves would occur approximately at the same μ value. Therefore both curves may serve the purpose of selecting a proper regularization parameter value in the f-slope method.

8.1. The general DPC assumption. We start with a problem defined by (1.1) that partially satisfies the DPC. Let $\gamma_i = u_i^T b/\sigma_i$. Then on the average, $|\gamma_i|$ decreases until some point γ_p, where it starts to increase. Write b as $b = b^{exact} + b^{error}$, where b^{exact} is the exact right-hand side vector for the problem and b^{error} is a perturbation vector. Define $\gamma_i^{exact} = u_i^T b^{exact}/\sigma_i$ and $\gamma_i^{error} = u_i^T b^{error}/\sigma_i$. Then

$$\gamma_i = \gamma_i^{exact} + \gamma_i^{error}.$$

Additionally, we assume that, on the average, $|\gamma_i^{exact}|$ decreases toward zero and $|\gamma_i^{error}|$ increases from zero. That is, we assume that the exact problem $Ax = b^{exact}$ satisfies the DPC completely and the error problem $Ax = b^{error}$ does not satisfy the DPC at all. With these assumptions, we see that for the decreasing part of the γ_i, γ_i^{exact} is the dominant component; and γ_i^{error} is the dominant one for the increasing part.

Now we make the following assumptions for the singular values of A:

$$\sigma_1 \geq \cdots \geq \sigma_k \geq \cdots \geq \sigma_p \geq \cdots \geq \sigma_l \geq \cdots \geq \sigma_n > 0,$$

ETNA
Kent State University
etna@mcs.kent.edu

and

$$\sigma_{k-1} \gg \sigma_p \gg \sigma_{l+1}.$$

Here, as usual, notation $\gg$ means "is much larger than".

The magnitude of the error $\|x^{exact} - x_\lambda\|$, where x^{exact} denotes the solution to the exact problem, depends on λ. This error can be made small if a λ value close to σ_p is selected, as shown below. Note that σ_p corresponds to the "turning point" $|\gamma_p|$ of $\{|\gamma_i|\}$.

Replacing b in (1.2) by b^{exact} and using the expression in (1.6) for x_λ, we have

$$x^{exact} - x_\lambda = \sum_{i=1}^{n} \left(\frac{\lambda^2}{\sigma_i^2 + \lambda^2} \gamma_i^{exact} - \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2} \gamma_i^{error} \right) v_i. \tag{8.1}$$

Hence,

$$\|x^{exact} - x_\lambda\|^2 = \sum_{i=1}^{n} \left(\frac{\lambda^2}{\sigma_i^2 + \lambda^2} \gamma_i^{exact} - \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2} \gamma_i^{error} \right)^2. \tag{8.2}$$

Let

$$a_i(\lambda) = \frac{\lambda^2}{\sigma_i^2 + \lambda^2} \quad \text{and} \quad b_i(\lambda) = \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2}.$$

Now suppose $\lambda \approx \sigma_p$. Then,

$$a_i(\lambda) \approx 0 \quad \text{and} \quad b_i(\lambda) \approx 1 \text{ for } 1 \le i \le k-1 \quad \text{since} \quad \sigma_i \gg \lambda;$$

and

$$a_i(\lambda) \approx 1 \quad \text{and} \quad b_i(\lambda) \approx 0 \text{ for } l+1 \le i \le n \quad \text{since} \quad \lambda \gg \sigma_i.$$

Thus,

$$\begin{aligned} &\|x^{exact} - x_\lambda\|^2 \\ &\approx \sum_{i=1}^{k-1} (\gamma_i^{error})^2 + \sum_{i=k}^{l} \left(a_i(\lambda)\gamma_i^{exact} - b_i(\lambda)\gamma_i^{error} \right)^2 + \sum_{i=l+1}^{n} (\gamma_i^{exact})^2. \end{aligned} \tag{8.3}$$

Note for $\lambda > 0$, $0 < a_i(\lambda) < 1$ and $0 < b_i(\lambda) < 1$. By the assumptions on γ_i^{exact} and γ_i^{error}, the three summations on the right-hand side in (8.3) would be small and so would $\|x^{exact} - x_\lambda\|$.

The following shows that a λ value close to σ_p is indeed what the f-slope method looks for. ¿From the expression for x_λ in (1.6) and by noting that $\lambda = e^{-\mu}$, we see that

$$\|x_\mu\|^2 = \sum_{i=1}^{n} \gamma_i^2 \left(\frac{\sigma_i^2}{\sigma_i^2 + e^{-2\mu}} \right)^2. \tag{8.4}$$

Differentiating $\|x_\mu\|^2$ with respect to μ gives

$$E(\mu) := (\|x_\mu\|^2)' = 4 \sum_{i=1}^{n} \gamma_i^2 \frac{\sigma_i^4 e^{-2\mu}}{(\sigma_i^2 + e^{-2\mu})^3}. \tag{8.5}$$

Define

$$H(\lambda) := \frac{1}{4}E(\mu(\lambda)) = \sum_{i=1}^{n} \gamma_i^2 \frac{\sigma_i^4 \lambda^2}{(\sigma_i^2+\lambda^2)^3} = \sum_{i=1}^{n} \gamma_i^2 \frac{(\sigma_i/\lambda)^4}{((\sigma_i/\lambda)^2+1)^3}. \tag{8.6}$$

The f-slope method finds a μ value that minimizes the E function over some interval $[c,d]$. If μ_{min} is such a value, then it is easy to see that $\lambda_{min} = e^{-\mu_{min}}$ minimizes the H function over $[e^{-d}, e^{-c}]$. We shall analyze function H in stead of function E.

The function form

$$s(t) = \frac{t^4}{(t^2+1)^3} \tag{8.7}$$

is common to all the terms in the right summation in (8.6). If $t \ll 1$, then $s \approx t^4$; if $t \gg 1$, then $s \approx 1/t^2$. So s rapidly diminishes as t goes away from 1 on either side. We can see that for a fixed λ value, the contribution of those terms in the summation in (8.6) with σ_i/λ very small or large will get suppressed. If for some $\lambda \approx \sigma_j$, s effectively filters out the influence of all γ_i's, $i \neq j$, then $H(\lambda) \approx \gamma_j^2/8$. This dependence of H on γ_i's indicates that the rise and fall of the γ_i may well be reflected in the change of H as λ moves down from σ_1 to σ_n. Therefore, the λ_{min} given by the f-slope method may fall close to σ_p.

Remark. The quasi-optimality criterion is another parameter choice method in Tikhonov regularization. This criterion is based on an error estimate of $\|x^{exact} - x_\lambda\|$ [5]. Minimization of the error estimate leads to the problem of minimizing the function

$$Q(\lambda) = \left(\sum_{i=1}^{n} (f_i(1-f_i)\gamma_i)^2\right)^{1/2} = \left(\sum_{i=1}^{n} \gamma_i^2 \frac{(\sigma_i/\lambda)^4}{\left((\sigma_i/\lambda)^2+1\right)^4}\right)^{1/2}, \tag{8.8}$$

where $f_i = \sigma_i^2/(\sigma_i^2+\lambda^2)$. We observe that $Q(\lambda)^2$ resembles $H(\lambda)$ in (8.6).

8.2. A model for the γ_i. To better understand how the f-slope method works, we then consider the following model for the γ_i:

$$\gamma_i = \begin{cases} \sigma_i^\alpha, & i = 1, ..., p, \\ \sigma_p^{\alpha+1}/\sigma_i, & i = p+1, ..., n, \end{cases} \tag{8.9}$$

where $\alpha > 0$. This model is a variant of the models assumed in [7] and [8] in obtaining theoretical results and satisfies

$$\gamma_1 \geq \cdots \geq \gamma_p \leq \gamma_{p+1} \leq \cdots \leq \gamma_n.$$

Further, we assume

$$\gamma_i^{exact} = \begin{cases} \gamma_i, & i = 1, ..., p, \\ 0, & i = p+1, ..., n, \end{cases} \quad \text{and} \quad \gamma_i^{error} = \begin{cases} 0, & i = 1, ..., p, \\ \gamma_i, & i = p+1, ..., n. \end{cases} \tag{8.10}$$

We shall first derive some bounds for $\|x^{exact} - x_\lambda\|^2$ and $H(\lambda)$ using the model. The expression in (8.2) together with (8.9) and (8.10) yields

$$\|x^{exact} - x_\lambda\|^2 = \sum_{i=1}^{p} \left(\frac{\lambda^2}{\sigma_i^2+\lambda^2}\sigma_i^\alpha\right)^2 + \sum_{i=p+1}^{n} \left(\frac{\sigma_i^2}{\sigma_i^2+\lambda^2}\frac{\sigma_p^{\alpha+1}}{\sigma_i}\right)^2. \tag{8.11}$$

ETNA
Kent State University
etna@mcs.kent.edu

LEMMA 8.1. *For* $\sigma_1 \geq \cdots \geq \sigma_k \geq \lambda \geq \sigma_{k+1} \geq \cdots \geq \sigma_p$,

$$\|x^{exact} - x_\lambda\|^2 < \sigma_k^{2\alpha}\theta_{k,1} = \sigma_k^4 \sigma_1^{2(\alpha-2)}\kappa_{k,1} \tag{8.12}$$

and

$$\|x^{exact} - x_\lambda\|^2 \geq \frac{1}{4}\sigma_{k+1}^{2\alpha}\tau_{k,1} = \frac{1}{4}\sigma_{k+1}^4 \sigma_1^{2(\alpha-2)}\chi_{k,1}, \tag{8.13}$$

where $\theta_{k,1}, \tau_{k,1}, \kappa_{k,1}, and\ \chi_{k,1}$ *satisfy*

$$1 < \theta_{k,1}, \tau_{k,1} \leq n\ for\ 0 < \alpha \leq 2 \quad and \quad 1 < \kappa_{k,1}, \chi_{k,1} \leq n\ for\ \alpha > 2.$$

Proof. Let

$$S_1 = \sum_{i=1}^{p} \left(\frac{\lambda^2}{\sigma_i^2 + \lambda^2}\sigma_i^\alpha \right)^2 \quad \text{and} \quad S_2 = \sum_{i=p+1}^{n} \left(\frac{\sigma_i^2}{\sigma_i^2 + \lambda^2}\frac{\sigma_p^{\alpha+1}}{\sigma_i} \right)^2.$$

Then

$$\|x^{exact} - x_\lambda\|^2 = S_1 + S_2.$$

Write S_1 as

$$\sum_{i=1}^{k} \left(\frac{\lambda^2}{\sigma_i^2 + \lambda^2}\sigma_i^\alpha \right)^2 + \sum_{i=k+1}^{p} \left(\frac{\lambda^2}{\sigma_i^2 + \lambda^2}\sigma_i^\alpha \right)^2.$$

We derive the upper bound first.

$$\begin{aligned} S_1 &< \sum_{i=1}^{k} \left(\frac{\lambda^2}{\sigma_i^2}\sigma_i^\alpha \right)^2 + \sum_{i=k+1}^{p} \sigma_i^{2\alpha} \\ &\leq \sigma_k^4 \sum_{i=1}^{k} \sigma_i^{2(\alpha-2)} + \sum_{i=k+1}^{p} \sigma_i^{2\alpha} \\ &= \sigma_k^{2\alpha} \sum_{i=1}^{k} \left(\frac{\sigma_k}{\sigma_i} \right)^{2(2-\alpha)} + \sum_{i=k+1}^{p} \sigma_i^{2\alpha}, \end{aligned}$$

and

$$S_2 < \sum_{i=p+1}^{n} \left(\frac{\sigma_i^2}{\lambda^2}\frac{\sigma_p^{\alpha+1}}{\sigma_i} \right)^2 \leq \sigma_p^{2\alpha} \left(\frac{\sigma_p}{\sigma_{k+1}} \right)^2 \sum_{i=p+1}^{n} \left(\frac{\sigma_i}{\sigma_{k+1}} \right)^2.$$

Hence,

$$\begin{aligned} &\|x^{exact} - x_\lambda\|^2 \\ &< \sigma_k^{2\alpha} \left(\sum_{i=1}^{k} \left(\frac{\sigma_k}{\sigma_i} \right)^{2(2-\alpha)} + \sum_{i=k+1}^{p} \left(\frac{\sigma_i}{\sigma_k} \right)^{2\alpha} + \left(\frac{\sigma_p}{\sigma_k} \right)^{2\alpha} \left(\frac{\sigma_p}{\sigma_{k+1}} \right)^2 \sum_{i=p+1}^{n} \left(\frac{\sigma_i}{\sigma_{k+1}} \right)^2 \right) \\ &= \sigma_k^4 \sigma_1^{2(\alpha-2)} \cdot \\ &\left(\sum_{i=1}^{k} \left(\frac{\sigma_i}{\sigma_1} \right)^{2(\alpha-2)} + \sum_{i=k+1}^{p} \left(\frac{\sigma_i}{\sigma_k} \right)^4 \left(\frac{\sigma_i}{\sigma_1} \right)^{2(\alpha-2)} + \left(\frac{\sigma_p}{\sigma_k} \right)^4 \left(\frac{\sigma_p}{\sigma_{k+1}} \right)^2 \left(\frac{\sigma_p}{\sigma_1} \right)^{2(\alpha-2)} \sum_{i=p+1}^{n} \left(\frac{\sigma_i}{\sigma_{k+1}} \right)^2 \right). \end{aligned}$$

The result follows.

Next, we derive the lower bound.

$$S_1 \geq \sum_{i=1}^{k}\left(\frac{\lambda^2}{2\sigma_i^2}\sigma_i^\alpha\right)^2 + \sum_{i=k+1}^{p}\left(\frac{\lambda^2}{2\lambda^2}\sigma_i^\alpha\right)^2$$
$$= \frac{1}{4}\left(\lambda^4\sum_{i=1}^{k}\sigma_i^{2(\alpha-2)} + \sum_{i=k+1}^{p}\sigma_i^{2\alpha}\right)$$
$$\geq \frac{1}{4}\left(\sigma_{k+1}^4\sum_{i=1}^{k}\sigma_i^{2(\alpha-2)} + \sum_{i=k+1}^{p}\sigma_i^{2\alpha}\right)$$
$$= \frac{1}{4}\left(\sigma_{k+1}^{2\alpha}\sum_{i=1}^{k}\left(\frac{\sigma_{k+1}}{\sigma_i}\right)^{2(2-\alpha)} + \sum_{i=k+1}^{p}\sigma_i^{2\alpha}\right),$$

and

$$S_2 \geq \sum_{i=p+1}^{n}\left(\frac{\sigma_i^2}{2\lambda^2}\frac{\sigma_p^{\alpha+1}}{\sigma_i}\right)^2 \geq \frac{1}{4}\sum_{i=p+1}^{n}\left(\frac{\sigma_i^2}{\sigma_k^2}\frac{\sigma_p^{\alpha+1}}{\sigma_i}\right)^2 = \frac{1}{4}\sigma_p^{2\alpha}\left(\frac{\sigma_p}{\sigma_k}\right)^2\sum_{i=p+1}^{n}\left(\frac{\sigma_i}{\sigma_k}\right)^2.$$

Hence,

$$\|x^{exact} - x_\lambda\|^2$$
$$\geq \frac{1}{4}\sigma_{k+1}^{2\alpha}\left(\sum_{i=1}^{k}\left(\frac{\sigma_{k+1}}{\sigma_i}\right)^{2(2-\alpha)} + \sum_{i=k+1}^{p}\left(\frac{\sigma_i}{\sigma_{k+1}}\right)^{2\alpha} + \left(\frac{\sigma_p}{\sigma_{k+1}}\right)^{2\alpha}\left(\frac{\sigma_p}{\sigma_k}\right)^2\sum_{i=p+1}^{n}\left(\frac{\sigma_i}{\sigma_k}\right)^2\right)$$
$$= \frac{1}{4}\sigma_{k+1}^4\sigma_1^{2(\alpha-2)}\cdot$$
$$\left(\sum_{i=1}^{k}\left(\frac{\sigma_i}{\sigma_1}\right)^{2(\alpha-2)} + \sum_{i=k+1}^{p}\left(\frac{\sigma_i}{\sigma_{k+1}}\right)^4\left(\frac{\sigma_i}{\sigma_1}\right)^{2(\alpha-2)} + \left(\frac{\sigma_p}{\sigma_{k+1}}\right)^4\left(\frac{\sigma_p}{\sigma_k}\right)^2\left(\frac{\sigma_p}{\sigma_1}\right)^{2(\alpha-2)}\sum_{i=p+1}^{n}\left(\frac{\sigma_i}{\sigma_k}\right)^2\right)$$

The result follows. $\square$

LEMMA 8.2. *For* $\sigma_p \geq \cdots \geq \sigma_l \geq \lambda \geq \sigma_{l+1} \geq \cdots \geq \sigma_n$,

$$\|x^{exact} - x_\lambda\|^2 < \sigma_p^{2\alpha}\left(\frac{\sigma_p}{\sigma_{l+1}}\right)^2\theta_{l,2} = \sigma_p^4\sigma_1^{2(\alpha-2)}\left(\frac{\sigma_p}{\sigma_{l+1}}\right)^2\kappa_{l,2} \tag{8.14}$$

and

$$\|x^{exact} - x_\lambda\|^2 \geq \frac{1}{4}\sigma_p^{2\alpha}\left(\frac{\sigma_p}{\sigma_l}\right)^2\tau_{l,2} = \frac{1}{4}\sigma_p^4\sigma_1^{2(\alpha-2)}\left(\frac{\sigma_p}{\sigma_l}\right)^2\chi_{l,2}, \tag{8.15}$$

where $\theta_{l,2}, \tau_{l,2}, \kappa_{l,2}, and\ \chi_{l,2}$ *satisfy*

$$1 < \theta_{l,2} \leq n,$$

$$n \geq \tau_{l,2} > \begin{cases} 1, & l \geq p+1, \\ \sigma_{p+1}/\sigma_p, & l = p, \end{cases}$$

for $0 < \alpha \leq 2$*; and*

$$\left(\frac{\sigma_l}{\sigma_p}\right)^4\left(\frac{\sigma_{l+1}}{\sigma_p}\right)^2 + \left(\frac{\sigma_p}{\sigma_1}\right)^{2(\alpha-2)} < \kappa_{l,2} \leq n,$$

$$n \geq \chi_{l,2} > \begin{cases} \left(\frac{\sigma_{l+1}}{\sigma_p}\right)^4\left(\frac{\sigma_l}{\sigma_p}\right)^2 + \left(\frac{\sigma_p}{\sigma_1}\right)^{2(\alpha-2)}, & l \geq p+1, \\ \left(\frac{\sigma_{p+1}}{\sigma_p}\right)^4 + \left(\frac{\sigma_p}{\sigma_1}\right)^{2(\alpha-2)}\left(\frac{\sigma_{p+1}}{\sigma_p}\right)^2, & l = p, \end{cases}$$

for $\alpha > 2$.

Proof. The proof for this lemma is similar to the one given to Lemma 8.1. Using the idea employed in the proof of Lemma 8.1, we can show the following:

$$\begin{aligned}
&\|x^{exact} - x_\lambda\|^2 \\
&< \sigma_p^{2\alpha}\left(\frac{\sigma_p}{\sigma_{l+1}}\right)^2\left(\left(\frac{\sigma_l}{\sigma_p}\right)^{2\alpha}\left(\frac{\sigma_{l+1}}{\sigma_p}\right)^2\sum_{i=1}^{p}\left(\frac{\sigma_l}{\sigma_i}\right)^{2(2-\alpha)} + \left(\frac{\sigma_{l+1}}{\sigma_l}\right)^2\sum_{i=p+1}^{l}\left(\frac{\sigma_l}{\sigma_i}\right)^2 + \sum_{i=l+1}^{n}\left(\frac{\sigma_i}{\sigma_{l+1}}\right)^2\right) \\
&= \sigma_p^4\sigma_1^{2(\alpha-2)}\left(\frac{\sigma_p}{\sigma_{l+1}}\right)^2 \cdot \\
&\quad \left(\left(\frac{\sigma_l}{\sigma_p}\right)^4\left(\frac{\sigma_{l+1}}{\sigma_p}\right)^2\sum_{i=1}^{p}\left(\frac{\sigma_i}{\sigma_1}\right)^{2(\alpha-2)} + \left(\frac{\sigma_p}{\sigma_1}\right)^{2(\alpha-2)}\left(\left(\frac{\sigma_{l+1}}{\sigma_l}\right)^2\sum_{i=p+1}^{l}\left(\frac{\sigma_l}{\sigma_i}\right)^2 + \sum_{i=l+1}^{n}\left(\frac{\sigma_i}{\sigma_{l+1}}\right)^2\right)\right)
\end{aligned}$$

and

$$\begin{aligned}
&\|x^{exact} - x_\lambda\|^2 \\
&\geq \frac{1}{4}\sigma_p^{2\alpha}\left(\frac{\sigma_p}{\sigma_l}\right)^2\left(\left(\frac{\sigma_{l+1}}{\sigma_p}\right)^{2\alpha}\left(\frac{\sigma_l}{\sigma_p}\right)^2\sum_{i=1}^{p}\left(\frac{\sigma_{l+1}}{\sigma_i}\right)^{2(2-\alpha)} + \sum_{i=p+1}^{l}\left(\frac{\sigma_l}{\sigma_i}\right)^2 + \sum_{i=l+1}^{n}\left(\frac{\sigma_i}{\sigma_l}\right)^2\right) \\
&= \frac{1}{4}\sigma_p^4\sigma_1^{2(\alpha-2)}\left(\frac{\sigma_p}{\sigma_l}\right)^2 \cdot \\
&\quad \left(\left(\frac{\sigma_{l+1}}{\sigma_p}\right)^4\left(\frac{\sigma_l}{\sigma_p}\right)^2\sum_{i=1}^{p}\left(\frac{\sigma_i}{\sigma_1}\right)^{2(\alpha-2)} + \left(\frac{\sigma_p}{\sigma_1}\right)^{2(\alpha-2)}\left(\sum_{i=p+1}^{l}\left(\frac{\sigma_l}{\sigma_i}\right)^2 + \sum_{i=l+1}^{n}\left(\frac{\sigma_i}{\sigma_l}\right)^2\right).\right)
\end{aligned}$$

The result follows. □

With γ_i being modeled as in (8.9), the function in (8.6) becomes:

$$H(\lambda) = \sum_{i=1}^{p}\frac{\sigma_i^{4+2\alpha}\lambda^2}{(\sigma_i^2+\lambda^2)^3} + \sigma_p^{2(\alpha+1)}\sum_{i=p+1}^{n}\frac{\sigma_i^2\lambda^2}{(\sigma_i^2+\lambda^2)^3}. \tag{8.16}$$

LEMMA 8.3. *For* $\sigma_1 \geq \cdots \geq \sigma_k \geq \lambda \geq \sigma_{k+1} \geq \cdots \geq \sigma_p$,

$$H(\lambda) < \sigma_k^{2\alpha}\theta_{k,3} = \sigma_k^2\sigma_1^{2(\alpha-1)}\kappa_{k,3} \tag{8.17}$$

and

$$H(\lambda) \geq \frac{1}{8}\sigma_k^{2\alpha}\left(\frac{\sigma_{k+1}}{\sigma_k}\right)^2\tau_{k,3} = \frac{1}{8}\sigma_{k+1}^2\sigma_1^{2(\alpha-1)}\chi_{k,3}, \tag{8.18}$$

where $\theta_{k,3}, \tau_{k,3}, \kappa_{k,3},$ *and* $\chi_{k,3}$ *satisfy*

$$1 < \theta_{k,3}, \tau_{k,3} \leq n \text{ for } 0 < \alpha \leq 1 \quad \text{and} \quad 1 < \kappa_{k,3}, \chi_{k,3} \leq n \text{ for } \alpha > 1.$$

Proof. The proof for this lemma is similar to the one given to Lemma 8.1. Following the steps applied in the proof of Lemma 8.1, we can show the following:

$$\begin{aligned}
&H(\lambda) \\
&< \sigma_k^{2\alpha}\left(\sum_{i=1}^{k}\left(\frac{\sigma_k}{\sigma_i}\right)^{2(1-\alpha)} + \left(\frac{\sigma_{k+1}}{\sigma_k}\right)^{2\alpha}\sum_{i=k+1}^{p}\left(\frac{\sigma_i}{\sigma_{k+1}}\right)^{4+2\alpha} + \left(\frac{\sigma_p}{\sigma_k}\right)^{2\alpha}\left(\frac{\sigma_p}{\sigma_{k+1}}\right)^2\sum_{i=p+1}^{n}\left(\frac{\sigma_i}{\sigma_{k+1}}\right)^2\right) \\
&= \sigma_k^2\sigma_1^{2(\alpha-1)} \cdot \left(\sum_{i=1}^{k}\left(\frac{\sigma_i}{\sigma_1}\right)^{2(\alpha-1)} + \left(\frac{\sigma_{k+1}}{\sigma_1}\right)^{2(\alpha-1)}\left(\frac{\sigma_{k+1}}{\sigma_k}\right)^2\sum_{i=k+1}^{p}\left(\frac{\sigma_i}{\sigma_{k+1}}\right)^{4+2\alpha} + \right. \\
&\quad \left.\left(\frac{\sigma_p}{\sigma_1}\right)^{2(\alpha-1)}\left(\frac{\sigma_p}{\sigma_k}\right)^2\left(\frac{\sigma_p}{\sigma_{k+1}}\right)^2\sum_{i=p+1}^{n}\left(\frac{\sigma_i}{\sigma_{k+1}}\right)^2\right)
\end{aligned}$$

ETNA
Kent State University
etna@mcs.kent.edu

and

$$
\begin{aligned}
&H(\lambda) \\
&\geq \frac{1}{8}\sigma_k^{2\alpha}\left(\frac{\sigma_{k+1}}{\sigma_k}\right)^2 \cdot \\
&\quad\left(\sum_{i=1}^{k}\left(\frac{\sigma_k}{\sigma_i}\right)^{2(1-\alpha)} + \left(\frac{\sigma_{k+1}}{\sigma_k}\right)^{2(\alpha+1)}\sum_{i=k+1}^{p}\left(\frac{\sigma_i}{\sigma_{k+1}}\right)^{4+2\alpha} + \left(\frac{\sigma_p}{\sigma_k}\right)^{2\alpha}\left(\frac{\sigma_p}{\sigma_{k+1}}\right)^2\sum_{i=p+1}^{n}\left(\frac{\sigma_i}{\sigma_k}\right)^2\right) \\
&= \frac{1}{8}\sigma_{k+1}^2\sigma_1^{2(\alpha-1)} \cdot \left(\sum_{i=1}^{k}\left(\frac{\sigma_i}{\sigma_1}\right)^{2(\alpha-1)} + \left(\frac{\sigma_k}{\sigma_1}\right)^{2(\alpha-1)}\left(\frac{\sigma_{k+1}}{\sigma_k}\right)^{2(\alpha+1)}\sum_{i=k+1}^{p}\left(\frac{\sigma_i}{\sigma_{k+1}}\right)^{4+2\alpha} + \right. \\
&\quad\left.\left(\frac{\sigma_p}{\sigma_1}\right)^{2(\alpha-1)}\left(\frac{\sigma_p}{\sigma_k}\right)^4\sum_{i=p+1}^{n}\left(\frac{\sigma_i}{\sigma_{k+1}}\right)^2\right).
\end{aligned}
$$

The result follows. $\square$

LEMMA 8.4. *For $\sigma_p \geq \cdots \geq \sigma_l \geq \lambda \geq \sigma_{l+1} \geq \cdots \geq \sigma_n$,*

$$
H(\lambda) < \sigma_p^{2\alpha}\left(\frac{\sigma_p}{\sigma_{l+1}}\right)^2\theta_{l,4} = \sigma_p^2\sigma_1^{2(\alpha-1)}\left(\frac{\sigma_p}{\sigma_{l+1}}\right)^2\kappa_{l,4} \tag{8.19}
$$

and

$$
H(\lambda) \geq \frac{1}{8}\sigma_p^{2\alpha}\left(\frac{\sigma_p}{\sigma_l}\right)^2\left(\frac{\sigma_{l+1}}{\sigma_l}\right)^2\tau_{l,4} = \frac{1}{8}\sigma_p^2\sigma_1^{2(\alpha-1)}\left(\frac{\sigma_p}{\sigma_l}\right)^2\left(\frac{\sigma_{l+1}}{\sigma_l}\right)^2\chi_{l,4}, \tag{8.20}
$$

where $\theta_{l,4}, \tau_{l,4}, \kappa_{l,4}$, and $\chi_{l,4}$ satisfy

$$
1 < \theta_{l,4}, \tau_{l,4} \leq n \ for \ 0 < \alpha \leq 1
$$

and

$$
\left(\frac{\sigma_l}{\sigma_p}\right)^2\left(\frac{\sigma_{l+1}}{\sigma_p}\right)^2 + \left(\frac{\sigma_p}{\sigma_1}\right)^{2(\alpha-1)} < \kappa_{l,4} \leq n,
$$

$$
\left(\frac{\sigma_l}{\sigma_p}\right)^4 + \left(\frac{\sigma_p}{\sigma_1}\right)^{2(\alpha-1)} < \chi_{l,4} \leq n,
$$

for $\alpha > 1$.

Proof. The proof for this lemma is similar to the one given to Lemma 8.1. Following the idea employed in the proof of Lemma 8.1, we can show the following:

$$
\begin{aligned}
&H(\lambda) \\
&< \sigma_p^{2\alpha}\left(\frac{\sigma_p}{\sigma_{l+1}}\right)^2\left(\left(\frac{\sigma_l}{\sigma_p}\right)^{2\alpha}\left(\frac{\sigma_{l+1}}{\sigma_p}\right)^2\sum_{i=1}^{p}\left(\frac{\sigma_l}{\sigma_i}\right)^{2(1-\alpha)} + \sum_{i=p+1}^{l}\left(\frac{\sigma_l}{\sigma_i}\right)^2\left(\frac{\sigma_{l+1}}{\sigma_i}\right)^2 + \sum_{i=l+1}^{n}\left(\frac{\sigma_i}{\sigma_{l+1}}\right)^2\right) \\
&= \sigma_p^2\sigma_1^{2(\alpha-1)}\left(\frac{\sigma_p}{\sigma_{l+1}}\right)^2 \cdot \\
&\quad\left(\left(\frac{\sigma_l}{\sigma_p}\right)^2\left(\frac{\sigma_{l+1}}{\sigma_p}\right)^2\sum_{i=1}^{p}\left(\frac{\sigma_i}{\sigma_1}\right)^{2(\alpha-1)} + \left(\frac{\sigma_p}{\sigma_1}\right)^{2(\alpha-1)}\left(\sum_{i=p+1}^{l}\left(\frac{\sigma_l}{\sigma_i}\right)^2\left(\frac{\sigma_{l+1}}{\sigma_i}\right)^2 + \sum_{i=l+1}^{n}\left(\frac{\sigma_i}{\sigma_{l+1}}\right)^2\right)\right)
\end{aligned}
$$

and

$$
\begin{aligned}
&H(\lambda)\\
&\geq \frac{1}{8}\sigma_p^{2\alpha}\left(\frac{\sigma_p}{\sigma_l}\right)^2\left(\frac{\sigma_{l+1}}{\sigma_l}\right)^2\left(\left(\frac{\sigma_l}{\sigma_p}\right)^{2(\alpha+1)}\sum_{i=1}^{p}\left(\frac{\sigma_l}{\sigma_i}\right)^{2(1-\alpha)}+\sum_{i=p+1}^{l}\left(\frac{\sigma_l}{\sigma_i}\right)^4+\sum_{i=l+1}^{n}\left(\frac{\sigma_i}{\sigma_{l+1}}\right)^2\right)\\
&=\frac{1}{8}\sigma_p^2\sigma_1^{2(\alpha-1)}\left(\frac{\sigma_p}{\sigma_l}\right)^2\left(\frac{\sigma_{l+1}}{\sigma_l}\right)^2\cdot\\
&\quad\left(\left(\frac{\sigma_l}{\sigma_p}\right)^4\sum_{i=1}^{p}\left(\frac{\sigma_i}{\sigma_1}\right)^{2(\alpha-1)}+\left(\frac{\sigma_p}{\sigma_1}\right)^{2(\alpha-1)}\left(\sum_{i=p+1}^{l}\left(\frac{\sigma_l}{\sigma_i}\right)^4+\sum_{i=l+1}^{n}\left(\frac{\sigma_i}{\sigma_{l+1}}\right)^2\right)\right).
\end{aligned}
$$

The result follows. □

The following two theorems give bounds to the minimizers of $\|x^{exact}-x_\lambda\|^2$ and $H(\lambda)$ in terms of the singular values of A.

THEOREM 8.5. *Let λ_{min}^{error} be a λ value that minimizes $\|x^{exact}-x_\lambda\|^2$. Let*

$$\phi=\sum_{i=1}^{p}\left(\frac{\sigma_p}{\sigma_i}\right)^{2(2-\alpha)}+\sum_{i=p+1}^{n}\left(\frac{\sigma_i}{\sigma_p}\right)^2.$$

For any $0<\alpha\leq 2$,

$$if\quad (\sigma_{i+1}/\sigma_p)^\alpha > 2\sqrt{\phi/\tau_{i,1}},\quad i=1,...,k,\ k\leq p-1,$$

and

$$if\quad \sigma_p/\sigma_j > 2\sqrt{\phi/\tau_{j,2}},\quad j=l,...,n-1,\ l\geq p+1,$$

then

$$\sigma_l<\lambda_{min}^{error}<\sigma_{k+1}.$$

Proof. Using the idea employed in the proof of Lemma 8.1, we can show at $\lambda=\sigma_p$ that

$$\|x^{exact}-x_\sigma\|^2<\sigma_p^{2\alpha}\phi. \tag{8.21}$$

First,

$$(\sigma_{i+1}/\sigma_p)^\alpha>2\sqrt{\phi/\tau_{i,1}}\quad\Longleftrightarrow\quad\sigma_p^{2\alpha}\phi<\sigma_{i+1}^{2\alpha}\tau_{i,1}/4.$$

By (8.21) and Lemma 8.1, we have

$$\|x^{exact}-x_{\sigma_p}\|^2<\sigma_p^{2\alpha}\phi<\sigma_{i+1}^{2\alpha}\tau_{i,1}/4\leq\|x^{exact}-x_\lambda\|^2,$$

for $\sigma_{i+1}\leq\lambda\leq\sigma_i, i=1,..,k$. This implies $\lambda_{min}^{error}<\sigma_{k+1}$.

Second,

$$\sigma_p/\sigma_j>2\sqrt{\phi/\tau_{j,2}}\quad\Longleftrightarrow\quad\sigma_p^{2\alpha}\phi<(\sigma_p/\sigma_j)^2\,\sigma_p^{2\alpha}\tau_{j,2}/4.$$

By (8.21) and Lemma 8.2, we have

$$\|x^{exact}-x_{\sigma_p}\|^2<\sigma_p^{2\alpha}\phi<(\sigma_p/\sigma_j)^2\,\sigma_p^{2\alpha}\tau_{j,2}/4\leq\|x^{exact}-x_\lambda\|^2,$$

for $\sigma_{j+1}\leq\lambda\leq\sigma_j, j=l,..,n-1$, where $l\geq p+1$. This implies $\lambda_{min}^{error}>\sigma_l$. □

THEOREM 8.6. *Let λ_{min}^{H} be a λ value that minimizes $H(\lambda)$. Let*

$$\psi=\sum_{i=1}^{p}\left(\frac{\sigma_p}{\sigma_i}\right)^{2(1-\alpha)}+\sum_{i=p+1}^{n}\left(\frac{\sigma_i}{\sigma_p}\right)^2.$$

For any $0 < \alpha \leq 1$,

$$if \quad (\sigma_i/\sigma_p)^\alpha > \frac{\sigma_i}{\sigma_{i+1}}\sqrt{8\psi/\tau_{i,3}}, \quad i = 1, ..., k, k \leq p-1,$$

and

$$if \quad \sigma_p/\sigma_j > \frac{\sigma_j}{\sigma_{j+1}}\sqrt{8\psi/\tau_{j,4}}, \quad j = l, ..., n-1, \ l \geq p+1,$$

then

$$\sigma_l < \lambda_{min}^H < \sigma_{k+1}.$$

Proof. Using the idea employed in the proof of Lemma 8.1, we can show at $\lambda = \sigma_p$ that

$$H(\sigma_p) < \sigma_p^{2\alpha}\psi. \tag{8.22}$$

The rest of the proof is immediate by following the steps given in the proof of Theorem 8.5 with the use of Lemma 8.3 and Lemma 8.4. ☐

We see from Theorem 8.5 that if the σ_i tends to zero and if α is not too close to 0, then λ_{min}^{error} would fall close to σ_p. We can arrive at the same conclusion for λ_{min}^H from Theorem 8.6 if we assume additionally that the ratio σ_j/σ_{j+1} is not too large for $j = p+1, ..., n-1$. Notice that $\phi, \psi, \tau_{i,1}, \tau_{i,2}, \tau_{i,3}$ and $\tau_{i,4}$ in Theorem 8.5 and Theorem 8.6 lie between 1 and n. It should be noted that the ranges for α in Theorem 8.5 and Theorem 8.6 are different. For the model problem, these two theorems show that the f-slope method is able to find a λ value that is close to the minimizer of $\|x^{exact} - x_\lambda\|^2$ under certain conditions. Furthermore, we can see from Lemma 8.1 and Lemma 8.2 that a small error can be achieved as long as the minimizer of $H(\lambda)$ falls close to σ_p.

9. Conclusions. In Tikhonov regularization of discrete ill-posed problems, choosing an appropriate regularization parameter is crucial. In this paper, a new method, the flattest slope method, for choosing a good regularization parameter is proposed. Comparisons are made with the L-curve method. For all the test problems considered in this paper, the new method chooses a regularized solution that is as good as and frequently better than the regularized solution chosen by the L-curve method. Recently, there have been a few papers illustrating limitations of the L-curve method with some criticisms of the non-convergence it sometimes exhibits as the relative error level in the right-hand side vector approaches zero, see [1, 4, 15]. The perturbation experiments in §7 show that the f-slope method exhibits convergence as the perturbation level δ decays to zero, in contrast to the L-curve method which exhibits non-convergence for some problems. The analysis given in §8 explains how the f-slope method works for some model problems.

Acknowledgments. I wish to thank Professor Charles Fulton for many helpful discussions and suggestions. I would also like to thank the referee for useful comments on an earlier version of this paper.

REFERENCES

[1] H. W. ENGL AND W. GREVER, *Using the L-Curve for Determining Optimal Regularization Parameters*, Numer. Math., 69 (1994), pp. 25–31.

[2] L. ELDÉN, *The numerical solution of a non-characteristic Cauchy problem for a parabolic equation*, in Numerical Treatment of Inverse Problems in Differential and Integral Equations, P. Deuflhart & E. Hairer, Eds., Birkhäuser, Boston, 1983.

[3] C. W. GROETSCH, *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind*, Research Notes in Mathematics 105, Pitman, Boston, 1984.

[4] M. HANKE, *Limitations of the L-Curve Method in Ill-posed Problems*, BIT, 36 (1996), no. 2, pp. 287–301.
[5] M. HANKE AND T. RAUS, *A general heuristic for choosing the regularization parameter in ill-posed problems*, SIAM J. Sci. Comput., 17 (1996), pp. 956-972.
[6] P. C. HANSEN, *Computation of the Singular Value Expansion*, Computing, 40 (1988), pp. 185–199.
[7] P. C. HANSEN, *The Discrete Picard Condition for Discrete Ill-posed Problems*, BIT, 30 (1990), pp. 658–672.
[8] P. C. HANSEN, *Truncated Singular Value Decomposition Solutions to Discrete Ill-Posed Problems with Ill-determined Numerical Rank*, SIAM J. Sci. Statist. Comput., 11 (1990), no. 3, pp. 503–518.
[9] P. C. HANSEN, *Analysis of Discrete Ill-posed Problems by Means of the L-curve*, SIAM Rev., 34 (1992), pp. 561–580.
[10] P. C. HANSEN, *Regularization Tools, a Matlab Package for Analysis and Solution of Discrete Ill-posed Problems*, Version 3.0 for Matlab 5.2, Technical Report UNIC-92-03, Danish Computing Center for Research and Education, Technical University of Denmark, June 1992 (Revised June 1998). Available online from http://www.netlib.org/numeralgo/na4.
Version 2.1 for Matlab 4.2c is also available on line from
http://www.netlib.org/numeralgo/na4-matlab4.
[11] P. C. HANSEN, *Rank-Deficient and Discrete Ill-Posed Problems*, SIAM, Philadelphia, 1998.
[12] P. C. HANSEN, *Test Matrices for Regularization Methods*, SIAM J. Sci. Comput., 16 (1995), no. 2, pp. 506–512.
[13] P. C. HANSEN AND D. O'LEARY, *The Use of the L-Curve in the Regularization of Discrete Ill-Posed Problems*, SIAM J. Sci. Comput., 14 (1993), pp. 1487–1503.
[14] J. M. VARAH, *A Practical Examination of Some Numerical Methods for Linear Discrete Ill-posed Problems*, SIAM Rev., 21 (1979), pp. 100–111.
[15] C. R. VOGEL, *Non-convergence of the L-curve Regularization Parameter Selection Method*, Inverse Problems, 12 (1996), no. 4, pp. 535–547.
[16] L. WU, *Regularization Methods and Algorithms for Least Squares and Kronecker Product Least Squares Problems*, Ph.D. thesis, Department of Applied Mathematics, Florida Institute of Technology, Melbourne, FL, May 1997.