SIAM J. SCI. COMPUT.
Vol. 23, No. 4, pp. 1161–1171

NEAR-OPTIMAL PARAMETERS FOR TIKHONOV AND OTHER REGULARIZATION METHODS*

DIANNE P. O'LEARY†

Abstract. Choosing the regularization parameter for an ill-posed problem is an art based on good heuristics and prior knowledge of the noise in the observations. In this work, we propose choosing the parameter, without a priori information, by approximately minimizing the distance between the true solution to the discrete problem and the family of regularized solutions. We demonstrate the usefulness of this approach for Tikhonov regularization and for an alternate family of solutions. Further, we prove convergence of the regularization parameter to zero as the standard deviation of the noise goes to zero.

Key words. ill-posed problems, regularization, Tikhonov

AMS subject classifications. 65R30, 65F20

PII. S1064827599354147

1. Introduction.

Linear, discrete ill-posed problems of the form

$$\min_x \|Ax-b\|_2, \quad \text{or, equivalently,} \quad A^*Ax = A^*b \tag{1.1}$$

arise, for example, from the discretization of first-kind Fredholm integral equations and occur in a variety of applications. We shall assume the following:

1. The full-rank matrix A is $m \times n$ with $m \geq n$.
2. A is ill-conditioned with no significant gap in the singular value spectrum. (A gap would make the problem somewhat easier.) The problem is normalized so that the largest singular value is 1.
3. The right-hand side b consists of true data plus random noise: $b = b_t + e$, where the components of e are an independent sample from a probability distribution with mean 0 and standard deviation s.
4. The discretization error caused by making a finite dimensional approximation to the continuous operator is much smaller than the noise.
5. The system satisfies the *discrete Picard condition*, which we will define in section 2 after introducing some notation.

The noise in the measurements, in combination with the ill conditioning of A, means that the exact solution of (1.1) has little relationship to the noise-free solution and is worthless. Instead, we use a *regularization* method to determine a solution that approximates the noise-free solution. Regularization methods replace the original operator by a better-conditioned, but related, one in order to diminish the effects of noise in the data and produce a *regularized solution* to the original problem. In this work, we first consider Tikhonov regularization, in which the problem (1.1) is replaced by

$$\min_x \left(\|Ax-b\|_2^2 + \lambda\|Lx\|_2^2\right), \quad \text{or, equivalently,} \quad (A^*A + \lambda L^*L)x = A^*b, \tag{1.2}$$

*Received by the editors April 14, 1999; accepted for publication (in revised form) June 12, 2001; published electronically November 7, 2001. This work was completed at the Departement Informatik, ETH Zürich, Switzerland. This work was supported in part by the U.S. National Science Foundation under grant CCR-97-32022 and by the National Institute of Standards and Technology.

http://www.siam.org/journals/sisc/23-4/35414.html

†Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742 (oleary@cs.umd.edu).

where L is a regularization operator chosen to obtain a solution with desirable properties, such as a small norm ($L = I$) or good smoothness (L a discrete approximation to a derivative operator), and $\lambda > 0$ is a scalar parameter. Throughout this paper we will use the 2-norm and denote it by $\| * \|$.

The central question in Tikhonov regularization is how to choose the parameter λ in order to produce a solution x close to the true noise-free solution x_{true}. Hoerl and Kennard [11] showed that on average a smaller error is produced using a nonzero λ, and numerous heuristics have been proposed for the choice of this parameter. Some of these (e.g., the discrepancy principle [14]) assume that the standard deviation of the noise is known. Others (e.g., generalized cross-validation [6] and the L-curve [8]) work with less knowledge of the noise properties. An interesting recent approach of Rust [17] uses visualization of residual and singular component plots to choose reasonable parameters. Pierce and Rust [15] minimize the lengths of confidence intervals using appropriate parameter choices, and Kilmer and O'Leary [13] discuss the choice of parameters when iterative solution methods are used.

In this work, we propose another rule for parameter choice. We go back to first principles: among all solutions in a given family such as Tikhonov, we want the solution that is a minimal distance from the true solution. Kay [12] has developed asymptotic expressions for this distance as the size of the problem grows large. Others have determined a Tikhonov parameter by minimizing a bound on this distance; Raus [16], Gfrerer [5], and Engl and Gfrerer [3] propose minimizing one such bound, while Hanke and Raus [7] propose an alternative. Rather than using asymptotic results or minimizing a bound, we compute in section 2 a parameter that approximately minimizes the distance to the true solution to the discretized problem and accomplishes this goal without a priori knowledge of the standard deviation or distribution of the noise in the observations. We discuss convergence of this choice in section 3. Section 4 contains a similar development for an alternative to Tikhonov regularization. Section 5 discusses some algorithmic issues, and in section 6 we show the effectiveness of these methods on numerical examples.

2. Choosing the Tikhonov regularization parameter. In order to analyze the problem, we convert to the coordinate system of the singular value decomposition (SVD) of A. For simplicity of exposition, we assume that the regularization operator L is the identity matrix. A similar development, using the generalized SVD, could be done for general L (see, for example, [10, sect. 2.1.2]), but the resulting function is considerably more complicated to compute and minimize.

Suppose $A = U\Sigma V^*$, where U and V have orthonormal columns and Σ is a matrix of zeros except for diagonal entries $\sigma_1 \geq \cdots \geq \sigma_n > 0$. Exploiting the property that $\|Uz\| = \|z\|$ and $\|V^* z\| = \|z\|$, the problem (1.2) takes the form

$$\min_z \|\Sigma z - \beta\|^2 + \lambda \|z\|^2 ,$$

where $\beta_i \equiv u_i^* b$ and $z = V^* x$. Setting the derivative equal to zero, we find that for a fixed value of λ, we need to solve the equation

$$(\Sigma^T \Sigma + \lambda I) z = \Sigma^T \beta.$$

Thus, the Tikhonov solution is

$$x_{tik} = \sum_{i=1}^{n} \frac{\beta_i \sigma_i}{\sigma_i^2 + \lambda} v_i, \tag{2.1}$$

where v_i is the ith column of V.

The true solution to the discrete (noise-free) problem is

$$x_{true} = \sum_{i=1}^{n} \frac{\beta_i - \epsilon_i}{\sigma_i} v_i,$$

where $\epsilon_i \equiv u_i^* e$ represents the unknown noise component.

The goal in regularization is to produce a solution as close as possible to the true solution, so let us (rather naively) try to minimize this distance:

$$\min_{\lambda} \|x_{tik} - x_{true}\|^2 \equiv \min_{\lambda} f(\lambda) \, .$$

Using the singular value representation, we see that

$$f(\lambda) = \sum_{i=1}^{n} \left[\frac{\beta_i \sigma_i}{\sigma_i^2 + \lambda} - \frac{\beta_i - \epsilon_i}{\sigma_i} \right]^2 \, .$$

Setting the derivative equal to zero yields

$$\begin{aligned} 0 = g(\lambda) \equiv \frac{1}{2} f'(\lambda) &= - \sum_{i=1}^{n} \left[\frac{\beta_i \sigma_i}{\sigma_i^2 + \lambda} - \frac{\beta_i - \epsilon_i}{\sigma_i} \right] \left[\frac{\beta_i \sigma_i}{(\sigma_i^2 + \lambda)^2} \right] \\ &= \sum_{i=1}^{n} \frac{\beta_i^2 \lambda}{(\sigma_i^2 + \lambda)^3} - \sum_{i=1}^{n} \frac{\beta_i \epsilon_i}{(\sigma_i^2 + \lambda)^2} \, . \end{aligned}$$

Now the first summation in this last expression is computable, but the second is not because the noise values ϵ_i are unknown. However, there are two interesting properties of the second summation.

- First, the terms for $i \approx n$ tend to be the largest because the denominators are the smallest.
- Second, the system satisfies the discrete Picard condition, meaning that for large enough values of the discretization parameter n, the sequence of true data values $\{\beta_i - \epsilon_i\}$ goes to zero faster than the sequence of singular values $\{\sigma_i\}$. Thus, for terms with i greater than or equal to some parameter k, $\epsilon_i \approx \beta_i$.

Therefore, although we cannot compute the function $g(\lambda)$, we can compute the following approximation to it:

$$\hat{g}(\lambda) \equiv \sum_{i=1}^{n} \frac{\beta_i^2 \lambda}{(\sigma_i^2 + \lambda)^3} - \sum_{i=k}^{n} \frac{\beta_i^2}{(\sigma_i^2 + \lambda)^2} - \mathcal{E} \left(\sum_{i=1}^{k-1} \frac{\beta_i \epsilon_i}{(\sigma_i^2 + \lambda)^2} \right)$$

for a suitable index k, depending on the standard deviation s. Finding the zero of this function yields an approximation to the optimal value of λ. The last term denotes the expected value of the quantity. Under assumption 3 of section 1, β_i is some true value plus noise ϵ_i, so $\mathcal{E}(\beta_i \epsilon_i) = \mathcal{E}(\epsilon_i^2) = s^2$, and

$$\hat{g}(\lambda) = \sum_{i=1}^{n} \frac{\beta_i^2 \lambda}{(\sigma_i^2 + \lambda)^3} - \sum_{i=k}^{n} \frac{\beta_i^2}{(\sigma_i^2 + \lambda)^2} - s^2 \sum_{i=1}^{k-1} \frac{1}{(\sigma_i^2 + \lambda)^2}. \tag{2.2}$$

As λ increases from zero, this function is monotonically increasing, and we denote the first zero by λ_{hat} and the corresponding solution vector x_{hat}.

3. Convergence for the Tikhonov parameter choice. We know that we cannot, in general, compute the optimal Tikhonov parameter. How far do we stray from the optimal vector when we use a nonoptimal parameter? The following theorem bounds the relative distance between the optimal solution and the computed one.

THEOREM 3.1. *Let λ_{opt} be the optimal parameter for the Tikhonov family (i.e., the (generally uncomputable) one that produces the solution closest to x_{true}). Then for any value of λ,*

$$\frac{\|x_{tik}(\lambda_{opt}) - x_{tik}(\lambda)\|}{\|x_{tik}(\lambda_{opt})\|} \leq \frac{|\lambda_{opt} - \lambda|}{\sigma_n^2 + \lambda}\,.$$

Proof. The result follows from the computation

$$\begin{aligned}
\|x_{tik}(\lambda_{opt}) - x_{tik}(\lambda)\|^2 &= \sum_{i=1}^{n} \left(\frac{\beta_i \sigma_i}{\sigma_i^2 + \lambda_{opt}} - \frac{\beta_i \sigma_i}{\sigma_i^2 + \lambda} \right)^2 \\
&= \sum_{i=1}^{n} \beta_i^2 \sigma_i^2 \left(\frac{\lambda - \lambda_{opt}}{(\sigma_i^2 + \lambda_{opt})(\sigma_i^2 + \lambda)} \right)^2 \\
&\leq \frac{(\lambda - \lambda_{opt})^2}{(\sigma_n^2 + \lambda)^2} \sum_{i=1}^{n} \left(\frac{\beta_i \sigma_i}{\sigma_i^2 + \lambda_{opt}} \right)^2 \\
&= \frac{(\lambda - \lambda_{opt})^2}{(\sigma_n^2 + \lambda)^2} \|x_{tik}(\lambda_{opt})\|^2\,. \qquad \square
\end{aligned}$$

Our algorithm for choosing the regularization parameter also behaves well as the size of the observation noise is decreased.

THEOREM 3.2. *If we choose the parameter k so that $\epsilon_i \approx \beta_i$ for $i \geq k$, then in the limit as the standard deviation s of the noise converges to zero, the solution x_{hat} produced by our algorithm converges to the correct discrete solution x_{true}.*

Proof. As the standard deviation of the noise goes to zero, the value k increases to $n+1$, and the solution to $\hat{g}(\lambda) = 0$ becomes $\lambda = 0$, as desired. Thus, as the noise goes to zero, our solution converges to the noise-free solution. $\square$

4. An alternate family of solutions. We have studied how the regularization parameter might be chosen for one family of solutions, the Tikhonov solutions, which take the form

$$x_{tik} = \sum_{i=1}^{n} \frac{\beta_i \sigma_i}{\sigma_i^2 + \lambda} v_i\,.$$

A similar algorithm can be found for other solution families, and in this section we consider the family

$$x_{alt} = \sum_{i=1}^{n} \frac{\beta_i}{\sigma_i + \lambda} v_i\,.$$

This family was proposed by Franklin [4] for Hermitian positive definite A and is also associated with Lavrentiev [10, p.107]. Ekstrom and Rhoads [2] discussed the use of the algorithm for convolution problems symmetrized by reordering, and this method was also considered by Cullum [1].

In his Regularization Tool Package for Matlab [9], Hansen includes a function `dsvd` that can be used to apply the method to general problems. In this more general context, there is more than one interpretation. The solution x_{alt} satisfies the regularized equation

$$(A + \lambda UV^*)x = b\,.$$

However, it may be more intuitive to interpret the family as a set of filter factors [10, sect. 4.2]

$$\frac{\sigma_i}{\sigma_i + \lambda},$$

multiplying the corresponding terms in the least squares solution

$$\sum_{i=1}^{n} \frac{\beta_i}{\sigma_i} v_i\,.$$

To choose the parameter λ, we mimic the procedure in section 2: we naively try to minimize the distance between our solution and the true one:

$$\min_{\lambda} \|x_{alt} - x_{true}\|^2 \equiv \min_{\lambda} f(\lambda)\,.$$

Using the singular value representation, we see that

$$f(\lambda) = \sum_{i=1}^{n} \left[\frac{\beta_i}{\sigma_i + \lambda} - \frac{\beta_i - \epsilon_i}{\sigma_i} \right]^2 .$$

Setting the derivative equal to zero yields

$$\begin{aligned} 0 = g(\lambda) \equiv \frac{1}{2} f'(\lambda) &= -\sum_{i=1}^{n} \left[\frac{\beta_i}{\sigma_i + \lambda} - \frac{\beta_i - \epsilon_i}{\sigma_i} \right] \left[\frac{\beta_i}{(\sigma_i + \lambda)^2} \right] \\ &= \sum_{i=1}^{n} \frac{\beta_i^2 \lambda}{\sigma_i(\sigma_i + \lambda)^3} - \sum_{i=1}^{n} \frac{\beta_i \epsilon_i}{\sigma_i(\sigma_i + \lambda)^2}\,. \end{aligned}$$

Again, the first summation in this last expression is computable. The second is not, because the observation noise values ϵ_i are unknown, but the terms for $i \approx n$ dominate, and for these $\epsilon_i \approx \beta_i$, so our approximate function becomes

$$\hat{g}(\lambda) \equiv \sum_{i=1}^{n} \frac{\beta_i^2 \lambda}{\sigma_i(\sigma_i + \lambda)^3} - \sum_{i=k}^{n} \frac{\beta_i^2}{\sigma_i(\sigma_i + \lambda)^2} - \mathcal{E} \left(\sum_{i=1}^{k-1} \frac{\beta_i \epsilon_i}{\sigma_i(\sigma_i + \lambda)^2} \right)$$

for a suitable index k that depends on the standard deviation of the noise. Finding the zero of the function

$$\hat{g}(\lambda) \equiv \sum_{i=1}^{n} \frac{\beta_i^2 \lambda}{\sigma_i(\sigma_i + \lambda)^3} - \sum_{i=k}^{n} \frac{\beta_i^2}{\sigma_i(\sigma_i + \lambda)^2} - s^2 \sum_{i=1}^{k-1} \frac{1}{\sigma_i(\sigma_i + \lambda)^2} \tag{4.1}$$

yields an approximation to the optimal value of λ.

We have a bound for the relative distance between the optimal solution and the computed one similar to the Tikhonov case.

THEOREM 4.1. *Let λ_{alt} be the optimal parameter for the alternate family (i.e., the one that produces the solution closest to x_{true}). Then for any value of λ,*

$$\frac{\|x_{alt}(\lambda_{alt}) - x_{alt}(\lambda)\|}{\|x_{alt}(\lambda_{alt})\|} \leq \frac{|\lambda_{alt} - \lambda|}{\sigma_n^2 + \lambda} .$$

Proof. The result follows from a computation similar to that in the proof of Theorem 3.1. ∎

Again, we can show that the solution converges to the true solution as the observation noise goes to zero.

THEOREM 4.2. *If we choose the parameter k so that $\epsilon_i \approx \beta_i$ for $i \geq k$, then in the limit as the standard deviation s of the noise converges to zero, the solution x_{hat} produced by our algorithm converges to x_{true}.*

Proof. The proof is the same as above. ∎

5. Algorithmic notes.

The standard deviation s of the noise is not assumed to be known, so we estimate it using the last $\max(m-n, 10)$ components of the right-hand side. If $|b_n| > 3.5s$, then we choose $k = n$. Otherwise we use a T-test to determine the index k. We choose k as the smallest index, among the values $n - 9, n - 14, \ldots$, for which a T-test with 0.05 significance level indicates that the sequence $\beta_k, \ldots, \beta_n$ has zero mean. If the mean of the noise-free sequence is likely to be near zero, then this test would not be appropriate, but many alternatives are available. One would be to use the Mann–Whitney Test, a nonparametric test to determine whether two independent groups of sampled data are taken from the same underlying distribution without making assumptions on the distribution.

A root of either function (2.2) or (4.1) can be found using standard algorithms (e.g., `fzero` in Matlab). Since $\hat{g}(0) < 0$ for both functions, we can find a lower bound on the root by searching $s, s/10, s/100, \ldots$ for a negative function value. The simple strategy of searching $100s, 1000s, \ldots$ has proved effective in finding a value for which $\hat{g}$ is positive, thus providing the root finder with an initial interval containing the root.

6. Performance of the algorithms.

The ideas of the previous sections were tested using two sets of test problems. In the first, the 200×200 matrix was diagonal, with entries ranging between 1 and 10^{-5}, evenly spaced on a log scale. The true solution was assumed to be the vector with elements evenly spaced between 1.0 and 0.9, and 100 sets of random noise were generated for the right-hand side. The value of the standard deviation of the noise was not made available to the algorithms; instead, we estimated it as in section 5. We generated solutions using the Tikhonov and the alternate method and calculated the distance between these computed solutions and the exact noise-free solution, tabulating the relative x-error $\|x - x_{true}\|/\|x_{true}\|$. Then we calculated the *optimal* Tikhonov and alternate solutions, the ones corresponding to the parameter values that minimize the distance to the noise-free solution. These optimal solutions, of course, cannot be computed in practical situations since the noise-free solution is unknown, but the results tell us how far we are from optimal. We also compared our results with three other methods:

1. We computed the the Tikhonov parameter by minimizing the generalized cross-validation (GCV) function using Matlab's `fmin` with tolerance 1.0e-07. In some sense this is an unfair comparison, since GCV aims to minimize the residual norm, not the x-error.

TABLE 6.1
Relative errors in experiments on a diagonal matrix of size 200.

Standard dev. of noise	Optimal Tikhonov	Computed Tikhonov	Optimal alternate	Computed alternate	GCV Tikhonov	Hanke–Raus Tikhonov	Discrep. Tikhonov
Mean							
1.0e-03	6.17e-01	7.05e-01	6.59e-01	1.12e+00	8.99e-01	8.20e-01	6.80e-01
1.0e-04	4.34e-01	4.61e-01	4.63e-01	4.94e-01	8.37e-01	6.82e-01	5.16e-01
1.0e-05	1.75e-01	2.11e-01	1.79e-01	1.97e-01	7.70e-01	4.97e-01	3.09e-01
1.0e-06	2.18e-02	6.52e-02	2.19e-02	5.34e-02	7.32e-01	4.58e-01	1.38e-01
Median							
1.0e-03	6.17e-01	6.43e-01	6.62e-01	6.71e-01	9.01e-01	8.21e-01	6.71e-01
1.0e-04	4.37e-01	4.57e-01	4.66e-01	4.74e-01	8.38e-01	6.83e-01	5.16e-01
1.0e-05	1.73e-01	2.10e-01	1.76e-01	1.95e-01	7.72e-01	4.83e-01	3.07e-01
1.0e-06	2.13e-02	4.10e-02	2.13e-02	3.05e-02	7.35e-01	4.63e-01	1.37e-01
Maximum							
1.0e-03	6.46e-01	4.95e+00	6.88e-01	7.34e+00	9.21e-01	8.49e-01	1.62e+00
1.0e-04	4.78e-01	9.86e-01	5.11e-01	9.93e-01	8.68e-01	7.52e-01	5.75e-01
1.0e-05	2.42e-01	2.82e-01	2.60e-01	2.68e-01	7.99e-01	5.89e-01	3.85e-01
1.0e-06	3.41e-02	1.61e-01	3.41e-02	1.40e-01	7.53e-01	4.71e-01	1.83e-01

2. We also compared our results with those of the Tikhonov algorithm of Hanke and Raus [7], which chooses the parameter by minimizing

$$f(\lambda) = \sqrt{1 + 1/\lambda}\sqrt{r_1^T(\lambda) r_0(\lambda)}\,,$$

where

$$\begin{aligned} x_0 &= (A^*A + \lambda I)^{-1} A^* b\,, \\ r_0(\lambda) &= b - Ax_0\,, \\ x_1 &= (A^*A + \lambda I)^{-1} A^* r_0 + x_0\,, \\ r_1(\lambda) &= b - Ax_1\,. \end{aligned}$$

3. We used our estimate of the standard deviation of the noise to apply the discrepancy principle [14].

The results are summarized in Table 6.1. Several trends are apparent. First, the average relative x-errors in the solutions computed by our algorithm are within a factor of 2 of the average relative x-errors for the optimal parameter values. Second, for large noise in the observations, the Tikhonov solution is on average closer to the true solution, but for small noise the alternate algorithm does somewhat better than Tikhonov. Third, the Tikhonov solutions computed by our algorithm are on average better than the GCV Tikhonov solutions and the Hanke–Raus solutions over the full range of noise values, and for small noise, the alternate solutions are better too. Our Tikhonov solutions are better than the discrepancy principle solutions except for a noise level of 10^{-3}.

The trends in the medians are similar to those of the averages, except that our values are always better than the discrepancy principle. The maximum relative errors show that only in the small number of cases in which the standard deviation of the error fails to be computed accurately are the GCV and Hanke–Raus solutions much better than our solutions.

Our algorithm for choosing the parameter k tested values in increments of 5. (See section 5.) Results are relatively insensitive to this increment: for experiments with noise level 1.0e-03 and increments of 1, 5, or 10, for example, the mean for

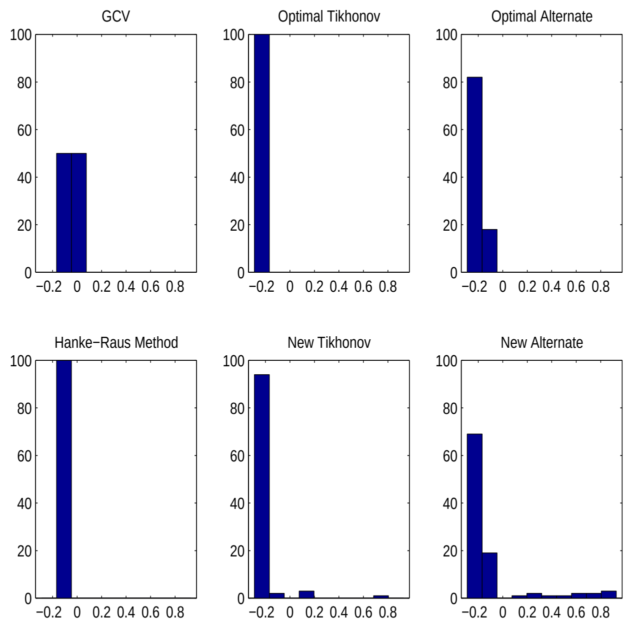

Fig. 6.1. *Histograms of the relative errors in the solutions computed for the diagonal matrix problem with standard deviation of the observation noise equal to* 1.0e-03. *The horizontal axis indicates the log of the relative error. The bars have height equal to the number of test problems yielding errors in that range.*

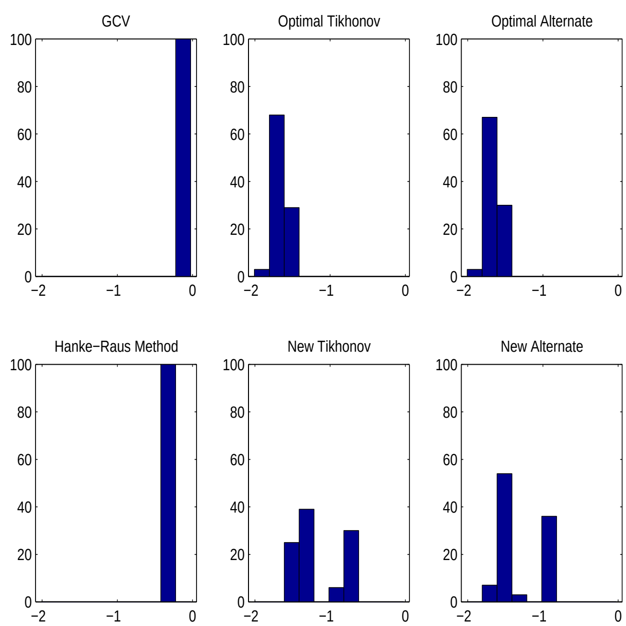

Fig. 6.2. *Histograms of the relative errors in the solutions computed for the diagonal matrix problem with standard deviation of the observation noise equal to* 1.0e-06. *The horizontal axis indicates the log of the relative error. The bars have height equal to the number of test problems yielding errors in that range.*

TABLE 6.2
Relative errors in experiments on a helioseismatic matrix of size 212×100.

Standard dev. of noise	Optimal Tikhonov	Computed Tikhonov	Optimal alternate	Computed alternate	GCV Tikhonov	Hanke–Raus
Mean values:						
1.0e-02	5.83e-01	5.75e+04	6.46e-01	5.08e+04	8.87e-01	8.63e-01
1.0e-04	4.58e-01	5.75e+02	4.91e-01	5.08e+02	6.60e-01	5.88e-01
1.0e-06	3.54e-01	3.49e+00	3.71e-01	3.35e+00	5.71e-01	5.71e-01
Median values:						
1.0e-02	5.92e-01	6.02e-01	6.49e-01	6.52e-01	8.87e-01	8.63e-01
1.0e-04	4.58e-01	4.90e-01	4.93e-01	4.97e-01	6.61e-01	5.88e-01
1.0e-06	3.54e-01	3.62e-01	3.73e-01	3.75e-01	5.71e-01	5.71e-01
Max values:						
1.0e-02	6.11e-01	4.55e+06	6.88e-01	4.00e+06	8.92e-01	8.66e-01
1.0e-04	4.95e-01	4.55e+04	5.18e-01	4.00e+04	6.68e-01	5.88e-01
1.0e-06	3.70e-01	3.10e+02	3.95e-01	2.96e+02	5.74e-01	5.74e-01

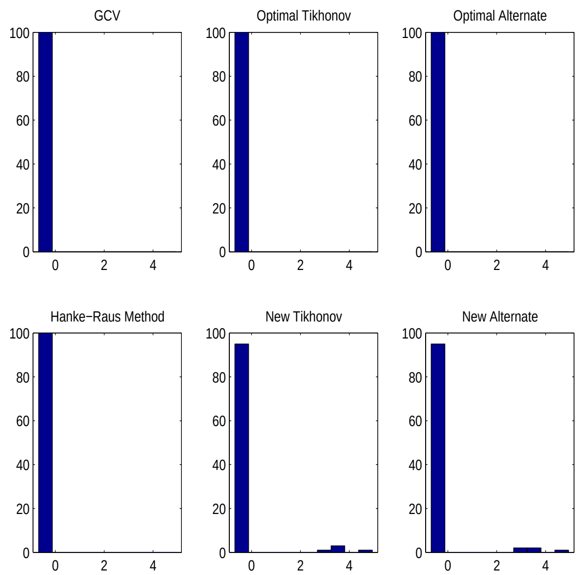

FIG. 6.3. *Histograms of the relative errors in the solutions computed for the helioseismatic matrix problem with standard deviation of the observation noise equal to* 1.0e-04. *The horizontal axis indicates the log of the relative error. The bars have height equal to the number of test problems yielding errors in that range.*

the computed Tikhonov values was between 7.01e-01 and 7.56e-01, while the median changed by at most 3 in the third significant digit. Similarly, the mean for the computed alternate values was between 1.06 and 1.34, while the median changed by at most 1 in the third significant digit.

Histograms of the relative errors are presented in Figures 6.1 and 6.2.

The second experiment used the inverse helioseismatic data of Hansen (`helio.mat`, taken from the Regularization Tool Package homepage [9]). The problem is an integral equation of the first kind modeling internal rotation of the sun as a function of radius. The matrix A of size 212×100 and the true solution x were obtained from there, and random observation noise was added as before. The right-hand side values had a mean close to zero, so a rather primitive scheme was used to determine k; it was determined so that the values b_j for $j > k$ were not larger than 3.5 times the estimated standard deviation. The results (Table 6.2) show that the median relative x-errors

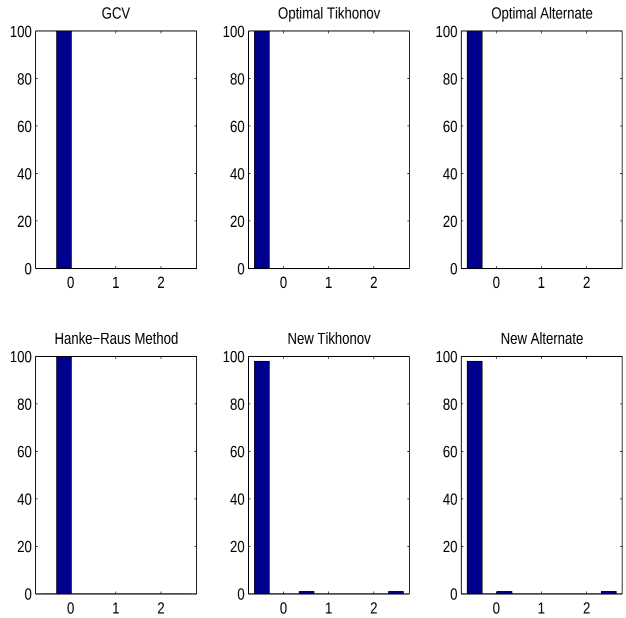

Fig. 6.4. *Histograms of the relative errors in the solutions computed for the helioseismatic matrix problem with standard deviation of the observation noise equal to* 1.0e-06. *The horizontal axis indicates the log of the relative error. The bars have height equal to the number of test problems yielding errors in that range.*

are at most 1.1 times as large as the optimal and at most 0.8 times the GCV values or the Hanke–Raus values. The trends are similar to the diagonal matrix problem: when the value of k is estimated well, the new algorithms perform much better than GCV and Hanke–Raus. However, since the k estimation problem is more difficult with this right-hand side, the mean and maximum values of the relative errors are not well behaved.

Still, the histograms of the relative errors presented in Figures 6.3 and 6.4 show that the new algorithms can be expected to produce much better results than GCV or Hanke–Raus when the errors are small enough that k is easily estimated.

7. Conclusions. We have proposed a method for choosing a regularization parameter that approximately minimizes the Euclidean distance between the computed solution and the noise-free solution, and we have demonstrated by numerical experiments that it produces solutions quite close to optimal.

We have demonstrated the use of these methods of parameter choice when the SVD of the matrix A can be explicitly computed. If the problem is too large for this to be practical, the ideas could be used in conjunction with iterative methods by applying them in the subspace generated by the iteration. For example, the SVD of the reduced matrix produced by a GMRES iteration could be substituted for the SVD of the full matrix. The effectiveness of this general methodology is discussed in [13].

Acknowledgments. I am grateful for the hospitality provided by Professor Walter Gander and the Departement Informatik, ETH Zürich, Switzerland, which enabled

this work to be completed. The manuscript benefited greatly from helpful comments by Misha E. Kilmer, Per Christian Hansen, Bert W. Rust, and the referees.

REFERENCES

[1] J. Cullum, *Ill-posed deconvolutions: Regularization and singular value decompositions*, in Proceedings of the 19th IEEE Conference on Decision and Control, Albuquerque, NM, 1980, pp. 29–35.

[2] M. P. Ekstrom and R. L. Rhoads, *On the application of eigenvector expansions to numerical deconvolution*, J. Comput. Phys., 14 (1974), pp. 319–340.

[3] H. W. Engl and H. Gfrerer, *A posteriori parameter choice for general regularization methods for solving linear ill-posed problems*, Appl. Numer. Math., 4 (1988), pp. 395–417.

[4] J. N. Franklin, *Minimum principles for ill-posed problems*, SIAM J. Math. Anal., 9 (1978), pp. 638–650.

[5] H. Gfrerer, *An a posteriori parameter choice for ordinary and iterated Tikhonov regularization of ill-posed problems leading to optimal convergence rates*, Math. Comp., 49 (1987), pp. 507–522.

[6] G. Golub, M. Heath, and G. Wahba, *Generalized cross-validation as a method for choosing a good ridge parameter*, Technometrics, 21 (1979), pp. 215–223.

[7] 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.

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

[9] P. C. Hansen, *Regularization tools: A Matlab package for analysis and solution of discrete ill-posed problems*, Numer. Algorithms, 6 (1994), pp. 1–35; also available online from http://www.imm.dtu.dk/documents/users/pch/Regutools/regutools.html.

[10] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*, SIAM, Philadelphia, 1997.

[11] A. E. Hoerl and R. W. Kennard, *Ridge regression: Biased estimation for nonorthogonal problems*, Technometrics, 12 (1970), pp. 55–67.

[12] J. Kay, *Asymptotic comparison factors for smoothing parameter choices in regression problems*, Statist. Probab. Lett., 15 (1992), pp. 329–335.

[13] M. E. Kilmer and D. P. O'Leary, *Choosing regularization parameters in iterative methods for ill-posed problems*, SIAM J. Matrix Anal. Appl., 22 (2001), pp. 1204–1221.

[14] V. A. Morozov, *On the solution of functional equations by the method of regularization*, Soviet Math. Dokl., 7 (1966), pp. 414–417. Cited in [10].

[15] J. E. Pierce and B. W. Rust, *Constrained least squares interval estimation*, SIAM J. Sci. Statist. Comput., 6 (1985), pp. 670–683.

[16] T. Raus, *The principle of the residual in the solution of ill-posed problems with nonselfadjoint operator*, Uchen. Zap. Tartu Gos. Univ., 715 (1985), pp. 12–20 (in Russian). Cited in [7].

[17] B. W. Rust, *Truncating the Singular Value Decomposition for Ill-Posed Problems*, Tech. Rep. NISTIR 6131, National Institute of Standards and Technology, U.S. Department of Commerce, Gaithersburg, MD, 1998.