Tikhonov Regularization for Large Scale Problems

Gene H. Golub[1] and Urs von Matt[2]

[1] Department of Computer Science, Stanford University, Stanford, CA 94305, golub@na-net.ornl.gov

[2] ISE Integrated Systems Engineering AG, Technopark Zürich, Technoparkstr. 1, CH-8005 Zürich, Switzerland, vonmatt@ise.ch

Abstract. Tikhonov regularization is a powerful tool for the solution of ill-posed linear systems and linear least squares problems. The choice of the regularization parameter is a crucial step, and many methods have been proposed for this purpose. However, efficient and reliable methods for large scale problems are still missing.

In this paper approximation techniques based on the Lanczos algorithm and the theory of Gauss quadrature are proposed to reduce the computational complexity for large scale problems. The new approach is applied to 5 different heuristics: Morozov's discrepancy principle, the Gfrerer/Raus-method, the quasi-optimality criterion, generalized cross-validation, and the L-curve criterion. Numerical experiments are used to determine the efficiency and robustness of the various methods.

1 Introduction

The solution of ill-posed linear systems and linear least squares problems is a frequent task in numerical analysis. We refer the reader to [5,17,30] for an overview of applications. In this paper we consider the (overdetermined) linear system

$$\boldsymbol{b} = A\boldsymbol{x} + \boldsymbol{e},$$

where A is an m-by-n matrix with $m \geq n$, $\boldsymbol{b}$ and $\boldsymbol{e}$ are vectors of size m, and $\boldsymbol{x}$ is an n-vector. The matrix A and the vector $\boldsymbol{b}$ are given, and $\boldsymbol{e}$ is assumed to be a random noise vector.

The direct solution of the least squares problem

$$\|A\boldsymbol{x} - \boldsymbol{b}\|_2 = \min \tag{1}$$

may lead to a vector $\boldsymbol{x}$ that is severely contaminated with noise. Therefore regularization is employed to get a more meaningful answer. In this paper we focus on Tikhonov regularization. This approach improves the condition of the problem by solving the linear least squares problem

$$\|A\boldsymbol{x} - \boldsymbol{b}\|_2^2 + \alpha\|\boldsymbol{x}\|_2^2 = \min \tag{2}$$

[1] The work of this author was in part supported by NSF under grant CCR-9505393.

instead of (1). The solution $\boldsymbol{x}_\alpha$ satisfies the equation

$$(A^{\mathrm{T}}A + \alpha I)\boldsymbol{x}_\alpha = A^{\mathrm{T}}\boldsymbol{b}.$$

Equivalently $\boldsymbol{x}_\alpha$ can be computed as the solution to the damped linear least squares problem

$$\left\| \begin{bmatrix} A \\ \sqrt{\alpha} I \end{bmatrix} \boldsymbol{x} - \begin{bmatrix} \boldsymbol{b} \\ \boldsymbol{0} \end{bmatrix} \right\|_2 = \min . \tag{3}$$

In [21] P. C. Hansen and D. P. O'Leary show that, for a given $\alpha > 0$, $\boldsymbol{x}_\alpha$ also solves the regularized total least squares problem

$$\begin{aligned} &\min_{A_0, \boldsymbol{b}_0, \boldsymbol{x}} \| [A, \boldsymbol{b}] - [A_0, \boldsymbol{b}_0] \|_F, \\ \text{s.t.} \qquad &\boldsymbol{b}_0 = A_0 \boldsymbol{x}, \qquad \|\boldsymbol{x}\|_2 \leq \delta := \|(A^{\mathrm{T}}A + \alpha I)^{-1} A^{\mathrm{T}} \boldsymbol{b}\|_2. \end{aligned}$$

The choice of an appropriate regularization parameter α is crucial, and many methods have been proposed to compute α. In Section 2 we give an overview of some frequently used criteria. The computational core of these methods can be expressed in terms of certain matrix moments. Since the main focus of this paper is the calculation of regularization parameters for large scale problems we will present an iterative method to approximate matrix moments. The required theory of Gauss quadrature is reviewed in Section 3, and stable methods based on the Lanczos algorithm are presented in Sections 4 and 5. As our methods of computing the regularization parameter α are based on an iterative algorithm, the Lanczos algorithm, we must provide an appropriate termination criterion. This turns out to be a nontrivial subject, and it is discussed in Section 6. Finally we present numerical results in Section 7.

2 Regularization Criteria

The choice of an appropriate regularization parameter α is crucial, and many methods have been proposed for this purpose. We will present a number of popular methods which may also be applied to large scale problems by means of the approximation techniques of Sections 3–5. The first two methods assume that the norm of the noise vector $\boldsymbol{e}$ is known, whereas the last three methods make no such assumption.

2.1 Morozov's Discrepancy Principle

The value of α is chosen such that the norm of the residual $\boldsymbol{b} - A\boldsymbol{x}_\alpha$ equals the norm of the error term [25,26]:

$$\|\boldsymbol{b} - A(A^{\mathrm{T}}A + \alpha I)^{-1} A^{\mathrm{T}} \boldsymbol{b}\|_2 = \|\boldsymbol{e}\|_2.$$

For $\alpha > 0$ the identity

$$I - A(A^{\mathrm{T}}A + \alpha I)^{-1}A^{\mathrm{T}} = \alpha(AA^{\mathrm{T}} + \alpha I)^{-1}$$

holds. Consequently the equation for α can also be written as

$$\phi_{\mathrm{M}}(\alpha) := \alpha^2 \boldsymbol{b}^{\mathrm{T}}(AA^{\mathrm{T}} + \alpha I)^{-2}\boldsymbol{b} = \|\boldsymbol{e}\|_2^2. \tag{4}$$

Because

$$\phi'_{\mathrm{M}}(\alpha) = 2\alpha \boldsymbol{b}^{\mathrm{T}}A(A^{\mathrm{T}}A + \alpha I)^{-3}A^{\mathrm{T}}\boldsymbol{b}$$

the function $\phi_{\mathrm{M}}(\alpha)$ is strictly increasing for $\alpha > 0$, and the equation (4) has a unique positive solution.

2.2 The Gfrerer/Raus-Method

The Gfrerer/Raus-method [8,17] may be seen as an improved variant of the discrepancy principle. It determines α such that

$$\phi_{\mathrm{GR}}(\alpha) := \alpha^3 \boldsymbol{b}^{\mathrm{T}}(AA^{\mathrm{T}} + \alpha I)^{-3}\boldsymbol{b} = \|\boldsymbol{e}\|_2^2. \tag{5}$$

The first derivative of $\phi_{\mathrm{GR}}(\alpha)$ is given by

$$\phi'_{\mathrm{GR}}(\alpha) = 3\alpha^2 \boldsymbol{b}^{\mathrm{T}}A(A^{\mathrm{T}}A + \alpha I)^{-4}A^{\mathrm{T}}\boldsymbol{b}.$$

Consequently $\phi_{\mathrm{GR}}(\alpha)$ is a strictly increasing function for $\alpha > 0$, and the solution of (5) is unique.

2.3 The Quasi-Optimality Criterion

The quasi-optimality criterion (cf. [1,24], [26, Sect. 27], [31, Sect. II.6], and [32]) determines $\alpha > 0$ such that

$$\phi_{\mathrm{Q}}(\alpha) := \alpha^2 \boldsymbol{b}^{\mathrm{T}}A(A^{\mathrm{T}}A + \alpha I)^{-4}A^{\mathrm{T}}\boldsymbol{b} = \min.$$

Note that we have $\phi_{\mathrm{Q}}(0) = 0$ such that the global minimizer of $\phi_{\mathrm{Q}}(\alpha)$ is attained at $\alpha = 0$, provided that the matrix A has full rank. As M. Hanke and P. C. Hansen point out in [17, p. 283] this is an artificial feature of a finite-dimensional regularization problem. The function $\phi_{\mathrm{Q}}(\alpha)$ usually has a large maximum in close vicinity of $\alpha = 0$, and the desired minimizer of $\phi_{\mathrm{Q}}(\alpha)$ lies to the right of this maximum.

2.4 Generalized Cross-Validation

Generalized cross-validation (GCV) [12,35,36] determines the regularization parameter α as the global minimizer of

$$\phi_{\mathrm{GCV}}(\alpha) := \frac{\|(AA^{\mathrm{T}} + \alpha I)^{-1}\boldsymbol{b}\|_2}{\mathrm{trace}\big((AA^{\mathrm{T}} + \alpha I)^{-1}\big)}.$$

The trace term in the denominator of $\phi_{\rm GCV}(\alpha)$ may be infeasible to evaluate for large matrices A. In [2,9,10,22] stochastic trace estimators are introduced to approximate the trace. We will use the approach proposed by M. F. Hutchinson [22]. Let U be the discrete random variable which takes the values $+1$ and -1 each with probability $1/2$, and let $\boldsymbol{u}$ be a vector of size m whose entries are independent samples from U. Then

$$\tilde{t}(\alpha) = \boldsymbol{u}^{\rm T}(AA^{\rm T} + \alpha I)^{-1}\boldsymbol{u} \tag{6}$$

is an unbiased estimator of

$$t(\alpha) = \operatorname{trace}\bigl((AA^{\rm T} + \alpha I)^{-1}\bigr).$$

Therefore we will only consider the minimization of the stochastic GCV function

$$\tilde{\phi}_{\rm GCV}(\alpha) := \frac{\sqrt{\boldsymbol{b}^{\rm T}(AA^{\rm T} + \alpha I)^{-2}\boldsymbol{b}}}{\boldsymbol{u}^{\rm T}(AA^{\rm T} + \alpha I)^{-1}\boldsymbol{u}}$$

from now on. In [15] it is shown that the minimizers of $\phi_{\rm GCV}$ and $\tilde{\phi}_{\rm GCV}$ are equally well suited for the Tikhonov regularization in (2).

2.5 The L-Curve Criterion

The L-curve criterion is based on a plot of the solution norm $\|\boldsymbol{x}_\alpha\|_2$ versus the residual norm $\|\boldsymbol{b} - A\boldsymbol{x}_\alpha\|_2$ in a log-log scale [18,20]. The optimal regularization parameter α is characterized by a corner of this graph, i.e., the point on the L-curve with maximum curvature. Therefore the L-curve criterion maximizes the curvature

$$\phi_{\rm L}(\alpha) := \frac{\rho'\eta'' - \rho''\eta'}{\bigl((\rho')^2 + (\eta')^2\bigr)^{3/2}} = \max,$$

where

$$\begin{aligned}\rho(\alpha) &= \log\|\boldsymbol{b} - A(A^{\rm T}A + \alpha I)^{-1}A^{\rm T}\boldsymbol{b}\|_2 = \log\|\alpha(AA^{\rm T} + \alpha I)^{-1}\boldsymbol{b}\|_2,\\ \eta(\alpha) &= \log\|(A^{\rm T}A + \alpha I)^{-1}A^{\rm T}\boldsymbol{b}\|_2.\end{aligned}$$

The prime denotes differentiation with respect to α.

Note that

$$\begin{aligned}\rho'(\alpha) &= \frac{\boldsymbol{b}^{\rm T}A(A^{\rm T}A + \alpha I)^{-3}A^{\rm T}\boldsymbol{b}}{\alpha\boldsymbol{b}^{\rm T}(AA^{\rm T} + \alpha I)^{-2}\boldsymbol{b}},\\ \eta'(\alpha) &= -\frac{\boldsymbol{b}^{\rm T}A(A^{\rm T}A + \alpha I)^{-3}A^{\rm T}\boldsymbol{b}}{\boldsymbol{b}^{\rm T}A(A^{\rm T}A + \alpha I)^{-2}A^{\rm T}\boldsymbol{b}}.\end{aligned}$$

The numerator of $\phi_{\rm L}(\alpha)$ can be written as

$$\rho'\eta'' - \rho''\eta' = \rho'^2\left(\frac{\eta'}{\rho'}\right)' = -\eta'^2\left(\frac{\rho'}{\eta'}\right)'.$$

Therefore we have

$$\begin{aligned}\rho'\eta'' - \rho''\eta' = \Big(&\frac{\boldsymbol{b}^{\mathrm{T}} A(A^{\mathrm{T}}A+\alpha I)^{-3}A^{\mathrm{T}}\boldsymbol{b}}{\alpha \boldsymbol{b}^{\mathrm{T}}(AA^{\mathrm{T}}+\alpha I)^{-2}\boldsymbol{b}\cdot \boldsymbol{b}^{\mathrm{T}}A(A^{\mathrm{T}}A+\alpha I)^{-2}A^{\mathrm{T}}\boldsymbol{b}}\Big)^2 \cdot \\ &\big(- \boldsymbol{b}^{\mathrm{T}}(AA^{\mathrm{T}}+\alpha I)^{-2}\boldsymbol{b}\cdot \boldsymbol{b}^{\mathrm{T}}A(A^{\mathrm{T}}A+\alpha I)^{-2}A^{\mathrm{T}}\boldsymbol{b} \\ &+ 2\alpha \boldsymbol{b}^{\mathrm{T}}(AA^{\mathrm{T}}+\alpha I)^{-3}\boldsymbol{b}\cdot \boldsymbol{b}^{\mathrm{T}}A(A^{\mathrm{T}}A+\alpha I)^{-2}A^{\mathrm{T}}\boldsymbol{b} \\ &- 2\alpha \boldsymbol{b}^{\mathrm{T}}(AA^{\mathrm{T}}+\alpha I)^{-2}\boldsymbol{b}\cdot \boldsymbol{b}^{\mathrm{T}}A(A^{\mathrm{T}}A+\alpha I)^{-3}A^{\mathrm{T}}\boldsymbol{b}\big).\end{aligned}$$

2.6 Large Scale Problems

All of the regularization criteria discussed so far require the evaluation of matrix moments of the form

$$\mu_p(\alpha) = \boldsymbol{b}^{\mathrm{T}}(AA^{\mathrm{T}}+\alpha I)^p \boldsymbol{b}, \tag{7}$$

$$\nu_p(\alpha) = \boldsymbol{c}^{\mathrm{T}}(A^{\mathrm{T}}A+\alpha I)^p \boldsymbol{c}. \tag{8}$$

The exponent p is always a negative integer. If the matrix A is not too large, say $n \leq 1000$, then a direct method can be used to compute a regularization parameter α. For instance one can employ the singular value decomposition of A to evaluate the functions $\phi(\alpha)$ in the five heuristics above. This is the approach taken in the regularization tool box by P. C. Hansen [19].

As the size of the matrix A increases the direct evaluation of these moments becomes less and less feasible. We will therefore present an iterative method which will enable us to compute lower and upper bounds on these moments. The bounds will become tighter as the iteration index k increases. In the next section we will review the necessary theory from Gauss quadrature.

3 Gauss Quadrature

Let us consider the problem of approximating the quadratic form

$$s := \boldsymbol{g}^{\mathrm{T}}\varphi(M)\boldsymbol{g}, \tag{9}$$

where φ is an analytic function and M is a symmetric n-by-n matrix. Let

$$M = U\Lambda U^{\mathrm{T}}$$

be the eigenvalue decomposition of M, with $\lambda_1 \leq \cdots \leq \lambda_n$ as the eigenvalues. We assume that a and b denote lower and upper bounds on the spectrum of M, i.e., $a \leq \lambda_1$ and $\lambda_n \leq b$. The expression (9) can be rewritten as

$$s = \boldsymbol{h}^{\mathrm{T}}\varphi(\Lambda)\boldsymbol{h} = \sum_{i=1}^{n}\varphi(\lambda_i)h_i^2 = \int_a^b \varphi(\lambda)\,d\omega(\lambda), \tag{10}$$

where the vector $\boldsymbol{h}$ is given by

$$\boldsymbol{h} := U^{\mathrm{T}} \boldsymbol{g}.$$

In (10) we also expressed the finite sum as a Stieltjes integral where the measure $\omega(x)$ is a staircase function with steps of size h_i^2 at the eigenvalues λ_i. This representation of s as an integral suggests to use numerical integration to approximate it. Because the moments

$$\mu_k = \int_a^b \lambda^k \, d\omega(\lambda) = \boldsymbol{g}^{\mathrm{T}} M^k \boldsymbol{g}$$

are easy to compute for $k \geq 0$ we will focus on Gauss quadrature:

$$s = \int_a^b \varphi(\lambda) \, d\omega(\lambda) \approx \sum_{i=1}^{k} \varphi(x_i) w_i. \tag{11}$$

The quantities $x_1 < \cdots < x_k$ denote the abscissas or nodes of the quadrature rule, and the w_i's are the corresponding weights. The degree of the quadrature rule is given by the index k.

It is sometimes useful to prescribe an abscissa $x_1 = a$ or $x_k = b$. One obtains a Gauss-Radau quadrature rule if $m = 1$ abscissa is prescribed. If both endpoints a and b are prescribed, i.e., $m = 2$, one obtains a Gauss-Lobatto quadrature rule.

There is an extensive literature on Gauss quadrature (cf. [3,7,23,28,34]). In [16] it is shown how the nodes x_i and the weights w_i can be computed by means of an eigenvalue decomposition. In [11] this approach is extended to Gauss-Radau and Gauss-Lobatto quadrature rules. In this section we give a short overview of the relevant theory.

3.1 Orthogonal Polynomials

A sequence of orthogonal polynomials can be associated with a weight function $\omega(\lambda)$. For the sake of simplicity we may assume that all the eigenvalues λ_i of M are distinct and that all the entries in the vector $\boldsymbol{h}$ are nonzero. Then there exists the sequence $\{p_i\}_{i=0}^{n-1}$, where p_i is a polynomial of degree i, such that

$$\int_a^b p_i(\lambda) p_j(\lambda) \, d\omega(\lambda) = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$$

These polynomials satisfy the three-term recurrence relationship

$$\lambda \begin{bmatrix} p_0(\lambda) \\ \vdots \\ \vdots \\ p_{k-1}(\lambda) \end{bmatrix} = \begin{bmatrix} \alpha_1 & \beta_1 & & \\ \beta_1 & \ddots & \ddots & \\ & \ddots & \alpha_{k-1} & \beta_{k-1} \\ & & \beta_{k-1} & \alpha_k \end{bmatrix} \begin{bmatrix} p_0(\lambda) \\ \vdots \\ \vdots \\ p_{k-1}(\lambda) \end{bmatrix} + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \beta_k p_k(\lambda) \end{bmatrix}, \tag{12}$$

Algorithm 1. Lanczos Algorithm.

$\boldsymbol{q}_{-1} := \mathbf{0}$
$\beta_0 := 0$
$\boldsymbol{q}_0 := \boldsymbol{g}/\|\boldsymbol{g}\|_2$
$\alpha_1 := \boldsymbol{q}_0^{\mathrm{T}} M \boldsymbol{q}_0$
for $k := 2$ **to** n **do**
  $\boldsymbol{r}_{k-1} := (M - \alpha_{k-1} I)\boldsymbol{q}_{k-2} - \beta_{k-2}\boldsymbol{q}_{k-3}$
  $\beta_{k-1} := \|\boldsymbol{r}_{k-1}\|_2$
  $\boldsymbol{q}_{k-1} := \boldsymbol{r}_{k-1}/\beta_{k-1}$
  $\alpha_k := \boldsymbol{q}_{k-1}^{\mathrm{T}} M \boldsymbol{q}_{k-1}$
end

or,

$$\lambda \boldsymbol{p}_k(\lambda) = T_k \boldsymbol{p}_k(\lambda) + \beta_k p_k(\lambda) \boldsymbol{e}_k$$

for $k = 1, \ldots, n$. The coefficients α_i and β_i can be computed by means of the Lanczos algorithm 1 (cf. [14, Chapt. 9]). An arbitrary nonzero value can be chosen for β_n. The zeros of $p_k(\lambda)$ are also the eigenvalues of T_k.

In exact arithmetic Algorithm 1 computes the orthogonal n-by-n matrix

$$Q_n = \left[\boldsymbol{q}_0 \, \cdots \, \boldsymbol{q}_{n-1}\right]$$

and the tridiagonal matrix T_n such that

$$M = Q_n T_n Q_n^{\mathrm{T}}.$$

If we execute only $k < n$ iterations we get Q_k, the first k columns of Q_n, and T_k, the k-by-k leading principal submatrix of T_n. It is easy to see that these quantities satisfy the equation

$$M Q_k = Q_k T_k + \beta_k \boldsymbol{q}_k \boldsymbol{e}_k^{\mathrm{T}}. \tag{13}$$

3.2 Construction of Gauss Quadrature Rules

There are $2k - m$ degrees of freedom when determining a Gauss quadrature rule (11) of degree k with m prescribed abscissas. Thus we can determine the nodes x_i and the corresponding weights w_i such that all polynomials up to degree $2k - m - 1$ are integrated exactly. As V. I. Krylov shows in [23, p. 161] an equivalent requirement is to find a polynomial $\tilde{p}_k(\lambda)$ of degree k, whose zeros are the nodes $x_1 < \cdots < x_k$, and the weights $w_1, \ldots, w_k$, such that

1. the equation

$$\int_a^b \tilde{p}_k(\lambda) p(\lambda) \, d\omega(\lambda) = 0 \tag{14}$$

holds for all polynomials p of degree $d < k - m$, and,

2. the equation

$$\int_a^b p(\lambda)\, d\omega(\lambda) = \sum_{i=1}^{k} p(x_i) w_i \tag{15}$$

holds for all polynomials p of degree $d < k$.

3.3 Abscissas

First we show how the polynomial $\tilde{p}_k(\lambda)$ can be determined such that condition (14) holds. In the case of a Gauss quadrature rule, where no abscissas are prescribed, we can choose $\tilde{p}_k = p_k$, the kth orthogonal polynomial with respect to the weight function $\omega(\lambda)$. This ensures that $\tilde{p}_k$ is orthogonal to all polynomials of degree $d < k$.

Next we consider a Gauss-Radau quadrature rule where we prescribe $x_1 = a$. We choose the following representation for $\tilde{p}_k$:

$$\tilde{p}_k = \beta_k p_k + (\alpha_k - \tilde{\alpha}_k) p_{k-1}.$$

This ensures that $\tilde{p}_k$ is orthogonal to all polynomials of degree $d < k-1$. It is not hard to see that $\tilde{p}_k$ also satisfies the equation

$$\lambda \begin{bmatrix} p_0(\lambda) \\ \vdots \\ \vdots \\ p_{k-1}(\lambda) \end{bmatrix} = \begin{bmatrix} \alpha_1 & \beta_1 & & \\ \beta_1 & \ddots & \ddots & \\ & \ddots & \alpha_{k-1} & \beta_{k-1} \\ & & \beta_{k-1} & \tilde{\alpha}_k \end{bmatrix} \begin{bmatrix} p_0(\lambda) \\ \vdots \\ \vdots \\ p_{k-1}(\lambda) \end{bmatrix} + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \tilde{p}_k(\lambda) \end{bmatrix},$$

or,

$$\lambda \boldsymbol{p}_k(\lambda) = \tilde{T}_k \boldsymbol{p}_k(\lambda) + \tilde{p}_k(\lambda) \boldsymbol{e}_k.$$

Thus the zeros of $\tilde{p}_k$ are just the eigenvalues of $\tilde{T}_k$, and if we set

$$\tilde{\alpha}_k = a + \beta_{k-1}^2 \boldsymbol{e}_{k-1}^{\mathrm{T}} (T_{k-1} - aI)^{-1} \boldsymbol{e}_{k-1},$$

we ensure that $\tilde{T}_k$ has the eigenvalue $x_1 = a$.

If we want to compute a Gauss-Radau quadrature rule with the prescribed abscissa $x_k = b$ we would just set

$$\tilde{\alpha}_k = b + \beta_{k-1}^2 \boldsymbol{e}_{k-1}^{\mathrm{T}} (T_{k-1} - bI)^{-1} \boldsymbol{e}_{k-1}.$$

Finally we want to determine the polynomial $\tilde{p}_k$ for a Gauss-Lobatto quadrature rule, where $x_1 = a$ and $x_k = b$ are prescribed. We may write $\tilde{p}_k$ as

$$\tilde{p}_k = \frac{\beta_k \beta_{k-1}}{\tilde{\beta}_{k-1}} p_k + (\alpha_k - \tilde{\alpha}_k) \frac{\beta_{k-1}}{\tilde{\beta}_{k-1}} p_{k-1} + \frac{\beta_{k-1}^2 - \tilde{\beta}_{k-1}^2}{\tilde{\beta}_{k-1}} p_{k-2},$$

where the parameters $\tilde{\alpha}_k$ and $\tilde{\beta}_{k-1}$ are to be determined such that $\tilde{p}_k(a) = \tilde{p}_k(b) = 0$. Obviously $\tilde{p}_k$ is orthogonal to all polynomials of degree $d < k-2$. It can also be verified that $\tilde{p}_k$ satisfies the equation

$$\lambda \begin{bmatrix} p_0(\lambda) \\ \vdots \\ \vdots \\ \tilde{p}_{k-1}(\lambda) \end{bmatrix} = \begin{bmatrix} \alpha_1 & \beta_1 & & \\ \beta_1 & \ddots & \ddots & \\ & \ddots & \alpha_{k-1} & \tilde{\beta}_{k-1} \\ & & \tilde{\beta}_{k-1} & \tilde{\alpha}_k \end{bmatrix} \begin{bmatrix} p_0(\lambda) \\ \vdots \\ \vdots \\ \tilde{p}_{k-1}(\lambda) \end{bmatrix} + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \tilde{p}_k(\lambda) \end{bmatrix},$$

or,

$$\lambda \tilde{\boldsymbol{p}}_k(\lambda) = \tilde{T}_k \tilde{\boldsymbol{p}}_k(\lambda) + \tilde{p}_k(\lambda)\boldsymbol{e}_k.$$

By setting

$$\tilde{\alpha}_k = \frac{b y_{k-1} - a z_{k-1}}{y_{k-1} - z_{k-1}},$$
$$\tilde{\beta}_{k-1} = \sqrt{\frac{b-a}{y_{k-1} - z_{k-1}}},$$

where

$$y_{k-1} = \boldsymbol{e}_{k-1}^{\mathrm{T}} (T_{k-1} - aI)^{-1} \boldsymbol{e}_{k-1},$$
$$z_{k-1} = \boldsymbol{e}_{k-1}^{\mathrm{T}} (T_{k-1} - bI)^{-1} \boldsymbol{e}_{k-1},$$

we can assign the two eigenvalues a and b to the matrix $\tilde{T}_k$. Since the eigenvalues of $\tilde{T}_k$ are the zeros of $\tilde{p}_k$ we have the desired property $\tilde{p}_k(a) = \tilde{p}_k(b) = 0$.

3.4 Weights

The weights $\{w_i\}_{i=1}^k$ must be determined in such a way that condition (15) is satisfied. We will only discuss the case of a Gauss quadrature rule. The cases of a Gauss-Radau or a Gauss-Lobatto quadrature rule are handled similarly: just replace the matrix T_k by $\tilde{T}_k$.

Condition (15) is equivalent to the requirement that the orthogonal polynomials $\{p_i\}_{i=0}^{k-1}$ are integrated exactly:

$$\int_a^b p_0 \, d\omega(\lambda) = \sum_{j=1}^k p_0 w_j = \frac{1}{p_0} = \sqrt{\int_a^b d\omega(\lambda)} = \sqrt{\mu_0},$$
$$\int_a^b p_i(\lambda) \, d\omega(\lambda) = \sum_{j=1}^k p_i(x_j) w_j = 0, \qquad i = 1, \ldots, k-1,$$

or, in matrix terms:

$$P\boldsymbol{w} = \sqrt{\mu_0}\boldsymbol{e}_1. \tag{16}$$

Note that the abscissas $x_1 < \cdots < x_k$ are the zeros of p_k, which are also the eigenvalues of T_k. This is obvious from the three-term recurrence relationship (12). Therefore the columns of P are the eigenvectors of T_k, and they are orthogonal to each other:

$$PX = T_k P.$$

Here the quantity X denotes the diagonal eigenvalue matrix $\operatorname{diag}(x_i)$. However the columns of P are not normalized. Let

$$T_k = QXQ^{\mathrm{T}}$$

denote the eigenvalue decomposition of T_k with $Q^{\mathrm{T}}Q = I$. Then we have

$$Q = PD,$$

where $D = \operatorname{diag}(q_{1i}/p_0)$. Consequently the solution of the linear system (16) is given by

$$\boldsymbol{w} = \sqrt{\mu_0} D Q^{\mathrm{T}} \boldsymbol{e}_1,$$

or,

$$w_i = \mu_0 q_{1i}^2.$$

3.5 Evaluation of Gauss Quadrature Rules

In order to approximate the quadratic form (9) by a quadrature rule of degree k we must execute k iterations of the Lanczos Algorithm 1. The value of the Gauss quadrature rule can then be evaluated as follows:

$$\begin{aligned} G &= \sum_{i=1}^{k} \varphi(x_i) w_i = \mu_0 \sum_{i=1}^{k} \varphi(x_i) q_{1i}^2 \\ &= \|\boldsymbol{g}\|_2^2 \boldsymbol{e}_1^{\mathrm{T}} Q \varphi(X) Q^{\mathrm{T}} \boldsymbol{e}_1 = \|\boldsymbol{g}\|_2^2 \boldsymbol{e}_1^{\mathrm{T}} \varphi(T_k) \boldsymbol{e}_1. \end{aligned}$$

A Gauss-Radau quadrature rule R and a Gauss-Lobatto quadrature rule L can be expressed in the same way. We only need to replace the matrix T_k by $\tilde{T}_k$ as discussed in Section 3.3.

3.6 Quadrature Error

In [23, pp. 162–163] and [27, p. 134] explicit expressions for the quadrature error

$$E[\varphi] := \int_a^b \varphi(\lambda)\, d\omega(\lambda) - \sum_{i=1}^{k} \varphi(x_i) w_i$$

are derived. In the case of a Gauss quadrature rule, there exists a ξ_G such that

$$E[\varphi] = \frac{\varphi^{(2k)}(\xi_G)}{(2k)!} \int_a^b \prod_{i=1}^{k} (x - x_i)^2 \, d\omega(x), \qquad a \leq \xi_G \leq b,$$

and for a Gauss-Radau quadrature rule with $x_1 = a$ we have

$$E[\varphi] = \frac{\varphi^{(2k-1)}(\xi_R)}{(2k-1)!} \int_a^b (x - a) \prod_{i=2}^{k} (x - x_i)^2 \, d\omega(x), \qquad a \leq \xi_R \leq b.$$

Similarly the error for a Gauss-Lobatto quadrature rule is given by

$$E[\varphi] = \frac{\varphi^{(2k-2)}(\xi_L)}{(2k-2)!} \int_a^b (x - a)(x - b) \prod_{i=2}^{k-1} (x - x_i)^2 \, d\omega(x), \qquad a \leq \xi_L \leq b.$$

3.7 Matrix Moments

As we have seen in Section 2.6 we are particularly interested in approximating the matrix moments $\mu_p(\alpha)$ and $\nu_p(\alpha)$. These moments can be expressed as a quadratic form (9) with

$$\varphi(\lambda) = (\lambda + \alpha)^p.$$

The key derivatives of φ are given as follows:

$$\begin{aligned} \varphi^{(2k)}(\lambda) &= p(p-1)\cdots(p-2k+1)(\lambda+\alpha)^{p-2k}, \\ \varphi^{(2k-1)}(\lambda) &= p(p-1)\cdots(p-2k+2)(\lambda+\alpha)^{p-2k+1}, \\ \varphi^{(2k-2)}(\lambda) &= p(p-1)\cdots(p-2k+3)(\lambda+\alpha)^{p-2k+2}. \end{aligned}$$

The exponent p is always negative. We may also assume that $\alpha > 0$ and $\lambda \geq 0$ since both AA^{T} and $A^{\mathrm{T}}A$ are symmetric positive semidefinite matrices. Consequently we know the sign of the derivatives of φ:

$$\begin{aligned} \varphi^{(2k)}(\lambda) &\geq 0, \\ \varphi^{(2k-1)}(\lambda) &\leq 0, \\ \varphi^{(2k-2)}(\lambda) &\geq 0. \end{aligned}$$

Therefore we can use a Gauss or a Gauss-Radau ($x_k = b$) quadrature rule to compute a lower bound on the matrix moments of Section 2.6. Conversely we get upper bounds from a Gauss-Radau ($x_1 = a$) or a Gauss-Lobatto quadrature rule. In the remainder of this paper we will only use a Gauss quadrature rule to obtain a lower bound, and a Gauss-Radau quadrature rule with $x_1 = a = 0$ for the upper bound.

Algorithm 2. Lanczos Bidiagonalization I.

$\boldsymbol{q}_0 := \boldsymbol{b}/\|\boldsymbol{b}\|_2$
$\boldsymbol{s}_0 := A^{\mathrm{T}}\boldsymbol{q}_0$
$\gamma_1 := \|\boldsymbol{s}_0\|_2$
$\boldsymbol{w}_0 := \boldsymbol{s}_0/\gamma_1$
for $k := 2$ **to** n **do**
  $\boldsymbol{r}_{k-1} := A\boldsymbol{w}_{k-2} - \gamma_{k-1}\boldsymbol{q}_{k-2}$
  $\delta_{k-1} := \|\boldsymbol{r}_{k-1}\|_2$
  $\boldsymbol{q}_{k-1} := \boldsymbol{r}_{k-1}/\delta_{k-1}$
  $\boldsymbol{s}_{k-1} := A^{\mathrm{T}}\boldsymbol{q}_{k-1} - \delta_{k-1}\boldsymbol{w}_{k-2}$
  $\gamma_k := \|\boldsymbol{s}_{k-1}\|_2$
  $\boldsymbol{w}_{k-1} := \boldsymbol{s}_{k-1}/\gamma_k$
end

4 Lanczos Bidiagonalization I

In this section we consider the numerical aspects of approximating the matrix moment $\mu_p(\alpha)$ in (7). The discussion in Section 3 seems to suggest that we apply Algorithm 1 to the matrix $M = AA^{\mathrm{T}}$. However, we may lose accuracy if we compute AA^{T} explicitly. This problem can be avoided by using the Lanczos bidiagonalization of Algorithm 2 (cf. [13]). If we execute k iterations of this algorithm in exact arithmetic we get the orthogonal m-by-k matrix

$$Q_k = \left[\boldsymbol{q}_0 \; \ldots \; \boldsymbol{q}_{k-1}\right]$$

and the orthogonal n-by-k matrix

$$W_k = \left[\boldsymbol{w}_0 \; \ldots \; \boldsymbol{w}_{k-1}\right].$$

The matrices Q_k and W_k are connected by the two equations

$$AW_k = Q_k B_k + \delta_k \boldsymbol{q}_k \boldsymbol{e}_k^{\mathrm{T}}, \tag{17}$$
$$A^{\mathrm{T}} Q_k = W_k B_k^{\mathrm{T}}, \tag{18}$$

where B_k denotes the k-by-k lower bidiagonal matrix

$$B_k = \begin{bmatrix} \gamma_1 & & & \\ \delta_1 & \ddots & & \\ & \ddots & \ddots & \\ & & \delta_{k-1} & \gamma_k \end{bmatrix}.$$

By combining (17) and (18) we get

$$AA^{\mathrm{T}} Q_k = Q_k B_k B_k^{\mathrm{T}} + \gamma_k \delta_k \boldsymbol{q}_k \boldsymbol{e}_k^{\mathrm{T}}.$$

It is now straightforward to identify $B_k B_k^{\mathrm{T}}$ with the matrix T_k in (13), and $\gamma_k \delta_k$ is equal to β_k in (13). Thus the Gauss quadrature rule for the approximation of (7) can be written as

$$G_p(\alpha) = \|\boldsymbol{b}\|_2^2 \boldsymbol{e}_1^{\mathrm{T}} (B_k B_k^{\mathrm{T}} + \alpha I)^p \boldsymbol{e}_1 .$$

To compute a Gauss-Radau quadrature rule we prescribe the abscissa $x_1 = a = 0$. This is an obvious choice since AA^{T} is always positive semidefinite. Thus we have to compute the entry $\tilde{\alpha}_k$ in $\tilde{T}_k$ such that $\tilde{T}_k$ becomes singular. It is not hard to see that we can express $\tilde{T}_k$ by

$$\tilde{T}_k = \tilde{B}_k \tilde{B}_k^{\mathrm{T}} ,$$

where we obtain $\tilde{B}_k$ from B_k by setting γ_k to zero. Consequently, the Gauss-Radau quadrature rule is given by

$$R_p(\alpha) = \|\boldsymbol{b}\|_2^2 \boldsymbol{e}_1^{\mathrm{T}} (\tilde{B}_k \tilde{B}_k^{\mathrm{T}} + \alpha I)^p \boldsymbol{e}_1 .$$

As discussed in Section 3.7 the Gauss and Gauss-Radau quadrature rules are actually lower and upper bounds on the matrix moment $\mu_p(\alpha)$:

$$G_p(\alpha) \le \mu_p(\alpha) \le R_p(\alpha).$$

Note that

$$\lim_{\alpha \downarrow 0} R_p(\alpha) = \infty$$

in general for $p < 0$. On the other hand $G_p(\alpha)$ remains bounded for $\alpha \downarrow 0$ if B_k is nonsingular.

We do not want to clutter the notation more than necessary, and we omit the degree k of the quadrature rule from the functions G_p and R_p. The reader should bear in mind, however, that these functions will approximate $\mu_p(\alpha)$ more and more accurately as k increases.

In order to evaluate $G_p(\alpha)$ and $R_p(\alpha)$ we need to solve linear systems of the form

$$(B_k B_k^{\mathrm{T}} + \alpha I)\boldsymbol{x} = \boldsymbol{y}.$$

It is important to note that for $\alpha > 0$ the solution $\boldsymbol{x}$ also satisfies the equivalent linear least squares problem

$$\left\| \begin{bmatrix} B_k^{\mathrm{T}} \\ \sqrt{\alpha} I \end{bmatrix} \boldsymbol{x} - \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{y}/\sqrt{\alpha} \end{bmatrix} \right\|_2 = \min .$$

This least squares problem can be solved efficiently by a sequence of Givens transformations. The details of this approach are discussed in [4, pp. 140–141], [6, pp. 45–52], and [34, pp. 103–108].

Algorithm 3. Lanczos Bidiagonalization II.

$\boldsymbol{q}_0 := \boldsymbol{c}/\|\boldsymbol{c}\|_2$
$\boldsymbol{r}_0 := A\boldsymbol{q}_0$
$\gamma_1 := \|\boldsymbol{r}_0\|_2$
$\boldsymbol{p}_0 := \boldsymbol{r}_0/\gamma_1$
for $k := 2$ **to** n **do**
  $\boldsymbol{s}_{k-1} := A^{\mathrm{T}}\boldsymbol{p}_{k-2} - \gamma_{k-1}\boldsymbol{q}_{k-2}$
  $\delta_{k-1} := \|\boldsymbol{s}_{k-1}\|_2$
  $\boldsymbol{q}_{k-1} := \boldsymbol{s}_{k-1}/\delta_{k-1}$
  $\boldsymbol{r}_{k-1} := A\boldsymbol{q}_{k-1} - \delta_{k-1}\boldsymbol{p}_{k-2}$
  $\gamma_k := \|\boldsymbol{r}_{k-1}\|_2$
  $\boldsymbol{p}_{k-1} := \boldsymbol{r}_{k-1}/\gamma_k$
end

5 Lanczos Bidiagonalization II

We will now address the approximation of the matrix moment $\nu_p(\alpha)$ in (8). The Lanczos bidiagonalization of Algorithm 3 operates on the matrix A instead of $A^{\mathrm{T}}A$ (cf. [13]). After k iterations in exact arithmetic we obtain the orthogonal m-by-k matrix

$$P_k = \begin{bmatrix}\boldsymbol{p}_0 & \cdots & \boldsymbol{p}_{k-1}\end{bmatrix}$$

and the orthogonal n-by-k matrix

$$Q_k = \begin{bmatrix}\boldsymbol{q}_0 & \cdots & \boldsymbol{q}_{k-1}\end{bmatrix}.$$

If we define the k-by-k upper bidiagonal matrix

$$B_k = \begin{bmatrix} \gamma_1 & \delta_1 & & \\ & \ddots & \ddots & \\ & & \ddots & \delta_{k-1} \\ & & & \gamma_k \end{bmatrix},$$

then P_k and Q_k satisfy the equations

$$AQ_k = P_k B_k, \tag{19}$$

$$A^{\mathrm{T}}P_k = Q_k B_k^{\mathrm{T}} + \delta_k \boldsymbol{q}_k \boldsymbol{e}_k^{\mathrm{T}}. \tag{20}$$

From (19) and (20) we conclude that

$$A^{\mathrm{T}}AQ_k = Q_k B_k^{\mathrm{T}} B_k + \gamma_k \delta_k \boldsymbol{q}_k \boldsymbol{e}_k^{\mathrm{T}}.$$

Obviously the matrix $B_k^{\mathrm{T}}B_k$ is the same as T_k in (13), and $\gamma_k\delta_k$ corresponds to β_k in (13). We obtain the following Gauss quadrature rule for the approximation of (8):

$$G_p(\alpha) = \|\boldsymbol{c}\|_2^2 \boldsymbol{e}_1^{\mathrm{T}} (B_k^{\mathrm{T}} B_k + \alpha I)^p \boldsymbol{e}_1.$$

Again, we can also use a Gauss-Radau quadrature rule to approximate $\nu_p(\alpha)$, and we will prescribe the abscissa $x_1 = a = 0$. Similarly as in Section 4 we can write the matrix $\tilde{T}_k$ as

$$\tilde{T}_k = \tilde{B}_k^{\mathrm{T}} \tilde{B}_k,$$

where we obtain $\tilde{B}_k$ from B_k by setting γ_k to zero. Consequently, the Gauss-Radau quadrature rule is given by

$$R_p(\alpha) = \|\boldsymbol{c}\|_2^2 \boldsymbol{e}_1^{\mathrm{T}} (\tilde{B}_k^{\mathrm{T}} \tilde{B}_k + \alpha I)^p \boldsymbol{e}_1.$$

We are now able to bound $\nu_p(\alpha)$ from below and above:

$$G_p(\alpha) \le \nu_p(\alpha) \le R_p(\alpha).$$

In general the function $R_p(\alpha)$ is unbounded for $p < 0$ as α goes to zero:

$$\lim_{\alpha \downarrow 0} R_p(\alpha) = \infty.$$

If B_k is nonsingular, however, $G_p(\alpha)$ remains bounded for $\alpha \downarrow 0$. Both bounds become tighter as we execute more and more iterations in Algorithm 3.

The evaluation of $G_p(\alpha)$ and $R_p(\alpha)$ involves the solution of linear systems of the form

$$(B_k^{\mathrm{T}} B_k + \alpha I)\boldsymbol{x} = \boldsymbol{y}.$$

For $\alpha > 0$ we prefer to compute $\boldsymbol{x}$ by means of a linear least squares problem as discussed in the previous section.

6 Termination Criteria

The theory of Sections 3–5 allows us to compute lower and upper bounds on the functions $\phi(\alpha)$ of Section 2. If we execute k steps of Algorithm 2 (Lanczos bidiagonalization I) with the initial vector $\boldsymbol{b}$ we get the following bounds on the moment $\mu_p(\alpha)$:

$$\|\boldsymbol{b}\|_2^2 \boldsymbol{e}_1^{\mathrm{T}} (B_k B_k^{\mathrm{T}} + \alpha I)^p \boldsymbol{e}_1 \le \boldsymbol{b}^{\mathrm{T}} (AA^{\mathrm{T}} + \alpha I)^p \boldsymbol{b} \le \|\boldsymbol{b}\|_2^2 \boldsymbol{e}_1^{\mathrm{T}} (\tilde{B}_k \tilde{B}_k^{\mathrm{T}} + \alpha I)^p \boldsymbol{e}_1.$$

Note that the same Lanczos algorithm can also be used to compute bounds on $\tilde{t}(\alpha)$ defined in (6). We just have to use the random vector $\boldsymbol{u}$ as the initial vector.

Similarly we run Algorithm 3 (Lanczos bidiagonalization II) with the initial vector $\boldsymbol{c} = A^{\mathrm{T}}\boldsymbol{b}$. Thus we get the following bounds on $\nu_p(\alpha)$:

$$\|\boldsymbol{c}\|_2^2 \boldsymbol{e}_1^{\mathrm{T}} (B_k^{\mathrm{T}} B_k + \alpha I)^p \boldsymbol{e}_1 \le \boldsymbol{c}^{\mathrm{T}} (A^{\mathrm{T}} A + \alpha I)^p \boldsymbol{c} \le \|\boldsymbol{c}\|_2^2 \boldsymbol{e}_1^{\mathrm{T}} (\tilde{B}_k^{\mathrm{T}} \tilde{B}_k + \alpha I)^p \boldsymbol{e}_1.$$

It is straightforward to compute lower and upper bounds on the various functions $\phi(\alpha)$ defined in Section 2:

$$\mathfrak{L}_k(\alpha) \leq \phi(\alpha) \leq \mathfrak{U}_k(\alpha).$$

These bounds get tighter as the number k of Lanczos iterations increases. We will now discuss the choice of appropriate stopping criteria.

Only certain values of the regularization parameter α are meaningful in the least squares problem (3). We will only consider α's in the range

$$\alpha_{\min} \leq \alpha \leq \alpha_{\max},$$

where

$$\begin{aligned} \alpha_{\min} &= \|A\|_2^2 \, \varepsilon^2, \\ \alpha_{\max} &= \|A\|_2^2, \end{aligned}$$

where ε denotes the unit roundoff of the computer. If $\alpha < \alpha_{\min}$ the matrix $\sqrt{\alpha} I$ in (3) is numerically zero compared to A. On the other hand, if $\alpha > \alpha_{\max}$, the damping matrix $\sqrt{\alpha} I$ becomes larger than the data matrix A.

As a simple guard against premature termination we require that at least

$$k_{\min} := \lceil 3 \log \min(m, n) \rceil$$

Lanczos iterations are executed. Thus the following stopping criteria are only checked for $k \geq k_{\min}$.

In the case of Morozov's discrepancy principle and the Gfrerer/Raus-method, $\phi(\alpha)$ is a strictly increasing function for $\alpha > 0$. Therefore, there is a unique regularization parameter α_* such that $\phi(\alpha) = \|\boldsymbol{e}\|_2^2$. The solution of the two equations

$$\begin{aligned} \mathfrak{U}_k(\alpha) &= \|\boldsymbol{e}\|_2^2, \\ \mathfrak{L}_k(\alpha) &= \|\boldsymbol{e}\|_2^2, \end{aligned}$$

yields the bounds α_l and α_u, respectively, such that

$$\alpha_l \leq \alpha_* \leq \alpha_u.$$

Consequently we may stop the iteration as soon as the bounds on α_* are tight enough. We use the criterion

$$0.99 \, \alpha_u \leq \alpha_l, \tag{21}$$

which allows us to determine α_* to two digits of accuracy.

The criterion (21) works well if $\phi'(\alpha_*)$ is sufficiently large, say, $\phi'(\alpha_*) \geq 1$. In the case of $\phi'(\alpha_*) \approx 0$ we need to relax the criterion (21). We only require that $\phi(\alpha)$ is approximated to two digits of accuracy in the range $\alpha_l \leq \alpha \leq \alpha_u$. In our implementation we use the condition

$$0.99 \, \mathfrak{U}_k(\alpha_l) \leq \|\boldsymbol{e}\|_2^2 \leq 1.01 \, \mathfrak{L}_k(\alpha_l). \tag{22}$$

The Lanczos iteration is stopped as soon as either (21) or (22) is satisfied.

In the case of the quasi-optimality criterion, GCV, and the L-curve criterion we need to compute a global minimizer or maximizer of $\phi(\alpha)$. Unfortunately, $\phi(\alpha)$ is usually neither convex nor concave, and thus no simple bounds on the global minimizer/maximizer α_* are available.

But the bounds on $\phi(\alpha)$ can still be used to compute approximations to α_*. We observe that the bounds $\mathfrak{L}_k(\alpha)$ and $\mathfrak{U}_k(\alpha)$ are very tight for large α's. As $\alpha \downarrow 0$, these bounds become increasingly loose. It is usually infeasible to execute the Lanczos algorithm until $\phi(\alpha)$ is approximated well over the whole interval $[\alpha_{\min}, \alpha_{\max}]$. In our implementation we only iterate until the largest local minimizer or maximizer of $\phi(\alpha)$ has been identified. Obviously this strategy fails if the global solution α_* is smaller than the largest local minimizer/maximizer. In Section 7 we will discuss the robustness of the three regularization criteria with respect to this type of premature termination.

We will now discuss the approximation of the largest local minimizer of $\phi(\alpha)$ for the quasi-optimality criterion and GCV. The largest local maximizer of $\phi_{\mathrm{L}}(\alpha)$ can be computed analogously.

First we compute a global minimizer α_u of the upper bound $\mathfrak{U}_k(\alpha)$. As a next step we would like to determine $\alpha_1 < \alpha_u < \alpha_2$ such that

$$\mathfrak{L}_k(\alpha_1) > \mathfrak{U}_k(\alpha_u) < \mathfrak{L}_k(\alpha_2). \tag{23}$$

This would mean that the interval $[\alpha_1, \alpha_2]$ contains a local minimizer of $\phi(\alpha)$. Unfortunately, condition (23) may need a lot of iterations to be satisfied in practice. Therefore we only try to find an $\alpha_1 < \alpha_u$ such that

$$\mathfrak{L}_k(\alpha_1) > \mathfrak{L}_k(\alpha_u).$$

To account for rounding errors in the calculation of the bounds we require that

$$\begin{aligned} \alpha_1 &\leq (1 - \sqrt{\varepsilon})\, \alpha_u, \\ \mathfrak{L}_k(\alpha_1) &\geq (1 + \sqrt{\varepsilon})\, \mathfrak{L}_k(\alpha_u), \end{aligned} \tag{24}$$

where ε denotes the unit roundoff of the computer.

7 Numerical Results

We implemented our methods to compute regularization parameters for large scale problems in MATLAB [29]. Our experiments were conducted on a Sun Ultra workstation with a unit roundoff of $\varepsilon = 2^{-52} \approx 2.2204 \cdot 10^{-16}$. The software is available from the second author upon request.

We used the functions "baart," "phillips," and "shaw" in Per Christian Hansen's regularization tool box [19] to generate test matrices. These matrices represent discretizations of Fredholm integral equations, a frequent source of ill-posed linear systems. However, they are dense, and we will only solve them for small values of n.

Table 1. Test Cases.

	Test Matrix	m	n	σ_1	σ_n	$\mathrm{relerr}_{\boldsymbol{e}}$
1	large	2000	1000	1.000	$1.690 \cdot 10^{-87}$	10^{-1}
2	large	2000	1000	1.000	$1.690 \cdot 10^{-87}$	10^{-2}
3	large	2000	1000	1.000	$1.690 \cdot 10^{-87}$	10^{-3}
4	large	20000	10000	1.000	$3.147 \cdot 10^{-869}$	10^{-1}
5	large	20000	10000	1.000	$3.147 \cdot 10^{-869}$	10^{-2}
6	large	20000	10000	1.000	$3.147 \cdot 10^{-869}$	10^{-3}
7	baart	100	100	3.229	$1.581 \cdot 10^{-18}$	10^{-1}
8	baart	100	100	3.229	$1.581 \cdot 10^{-18}$	10^{-3}
9	baart	100	100	3.229	$1.581 \cdot 10^{-18}$	10^{-5}
10	baart	100	100	3.229	$1.581 \cdot 10^{-18}$	10^{-7}
11	phillips	200	200	5.803	$1.372 \cdot 10^{-7}$	10^{-1}
12	phillips	200	200	5.803	$1.372 \cdot 10^{-7}$	10^{-2}
13	phillips	200	200	5.803	$1.372 \cdot 10^{-7}$	10^{-3}
14	phillips	200	200	5.803	$1.372 \cdot 10^{-7}$	10^{-4}
15	shaw	100	100	2.993	$5.486 \cdot 10^{-19}$	10^{-1}
16	shaw	100	100	2.993	$5.486 \cdot 10^{-19}$	10^{-3}
17	shaw	100	100	2.993	$5.486 \cdot 10^{-19}$	10^{-5}
18	shaw	100	100	2.993	$5.486 \cdot 10^{-19}$	10^{-7}

In order to obtain large scale test problems, which can also be solved explicitly, we use the m-by-n matrix A defined by its singular value decomposition

$$A = U \Sigma V^{\mathrm{T}},$$

where

$$U = I - 2\frac{\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}}{\|\boldsymbol{u}\|_2^2},$$
$$V = I - 2\frac{\boldsymbol{v}\boldsymbol{v}^{\mathrm{T}}}{\|\boldsymbol{v}\|_2^2},$$

and the entries of $\boldsymbol{u}$ and $\boldsymbol{v}$ are random numbers. Σ is a diagonal matrix with the singular values $\sigma_i = e^{-0.2(i-1)}$. We will call this the “large” test problem.

We use a random vector $\boldsymbol{x}_0$ as the true solution of (1) and compute the right-hand side $\boldsymbol{b}$ according to

$$\boldsymbol{b} = A\boldsymbol{x}_0 + \boldsymbol{e},$$

where $\boldsymbol{e}$ is a random vector with

$$\|\boldsymbol{e}\|_2 = \mathrm{relerr}_{\boldsymbol{e}} \, \|A\boldsymbol{x}_0\|_2.$$

We used the 18 problems from Table 1 to test our regularization algorithms. The columns σ_1 and σ_n show the values of the largest and the smallest singular values of A, respectively. Note that σ_1 and σ_n were computed by using

Table 2. Average Value of the Exact Regularization Parameter α_*.

	Morozov	Gfrerer/Raus	Quasi-Optimality	GCV	L-Curve
1	$2.998 \cdot 10^{-4}$	$5.420 \cdot 10^{-4}$	$9.538 \cdot 10^{-4}$	$3.319 \cdot 10^{-5}$	$9.380 \cdot 10^{-4}$
2	$2.303 \cdot 10^{-6}$	$4.737 \cdot 10^{-6}$	$2.452 \cdot 10^{-4}$	$1.800 \cdot 10^{-7}$	$7.384 \cdot 10^{-6}$
3	$6.278 \cdot 10^{-9}$	$1.409 \cdot 10^{-8}$	$1.630 \cdot 10^{-5}$	$4.630 \cdot 10^{-10}$	$2.136 \cdot 10^{-8}$
4	$8.784 \cdot 10^{-6}$	$1.674 \cdot 10^{-5}$	$5.333 \cdot 10^{-5}$	$8.359 \cdot 10^{-7}$	$1.326 \cdot 10^{-4}$
5	$4.553 \cdot 10^{-7}$	$1.036 \cdot 10^{-6}$	$7.441 \cdot 10^{-7}$	$3.786 \cdot 10^{-8}$	$8.779 \cdot 10^{-6}$
6	$3.885 \cdot 10^{-9}$	$8.053 \cdot 10^{-9}$	$8.719 \cdot 10^{-6}$	$1.809 \cdot 10^{-10}$	$1.756 \cdot 10^{-8}$
7	$5.929 \cdot 10^{-4}$	$1.367 \cdot 10^{-3}$	$4.394 \cdot 10^{-4}$	$1.362 \cdot 10^{-4}$	$5.894 \cdot 10^{-4}$
8	$3.869 \cdot 10^{-6}$	$7.385 \cdot 10^{-6}$	$1.191 \cdot 10^{-6}$	$3.138 \cdot 10^{-7}$	$4.887 \cdot 10^{-7}$
9	$3.062 \cdot 10^{-9}$	$9.005 \cdot 10^{-9}$	$2.037 \cdot 10^{-9}$	$2.809 \cdot 10^{-10}$	$2.614 \cdot 10^{-10}$
10	$1.227 \cdot 10^{-14}$	$2.979 \cdot 10^{-14}$	$3.921 \cdot 10^{-12}$	$9.020 \cdot 10^{-16}$	$4.760 \cdot 10^{-15}$
11	$1.398 \cdot 10^{-2}$	$2.064 \cdot 10^{-2}$	$3.593 \cdot 10^{-33}$	$1.224 \cdot 10^{-3}$	$7.050 \cdot 10^{-2}$
12	$1.871 \cdot 10^{-4}$	$3.756 \cdot 10^{-4}$	$1.181 \cdot 10^{-32}$	$3.103 \cdot 10^{-5}$	$7.388 \cdot 10^{-4}$
13	$3.131 \cdot 10^{-6}$	$6.626 \cdot 10^{-6}$	$2.562 \cdot 10^{-32}$	$3.663 \cdot 10^{-7}$	$5.803 \cdot 10^{-6}$
14	$4.941 \cdot 10^{-8}$	$8.631 \cdot 10^{-8}$	$2.562 \cdot 10^{-32}$	$8.465 \cdot 10^{-9}$	$9.112 \cdot 10^{-9}$
15	$7.259 \cdot 10^{-3}$	$9.898 \cdot 10^{-3}$	$1.885 \cdot 10^{-2}$	$3.049 \cdot 10^{-4}$	$1.307 \cdot 10^{-2}$
16	$2.077 \cdot 10^{-6}$	$3.283 \cdot 10^{-6}$	$1.660 \cdot 10^{-2}$	$8.562 \cdot 10^{-8}$	$4.283 \cdot 10^{-7}$
17	$5.511 \cdot 10^{-9}$	$2.597 \cdot 10^{-8}$	$1.876 \cdot 10^{-8}$	$2.604 \cdot 10^{-10}$	$1.503 \cdot 10^{-10}$
18	$7.917 \cdot 10^{-14}$	$5.636 \cdot 10^{-14}$	$2.008 \cdot 10^{-8}$	$4.422 \cdot 10^{-16}$	$1.927 \cdot 10^{-14}$

MATLAB's `svd` command. If $\sigma_n \approx \varepsilon\sigma_1$, the matrix A is numerically singular and the value of σ_n may carry no significant digits. However, for the large test matrix, the extreme singular values are exact. Note that the test cases 11–14 are only moderately ill-conditioned, whereas all the other test cases are severely ill-conditioned.

We solved each test problem 10 times with a given regularization technique. Each time we used a different random perturbation $\boldsymbol{e}$ of the right-hand side. This gives us a better way to assess the average properties of each regularization method.

Table 2 shows the average values for the exact regularization parameters α_* produced by each regularization technique. The value of α_* tends to 0 as the norm of the perturbation becomes smaller and smaller. For the test cases 11–14 the quasi-optimality criterion fails to compute a regularization parameter. As already pointed out in Section 2.3 this anomaly is caused by the properties of the finite-dimensional regularization problem. The function $\phi_{\mathrm{Q}}(\alpha)$ has a large maximum around $\alpha \approx 10^{-14}$, and the desired minimizer α_* lies to the right of this maximum. Unfortunately, $\phi_{\mathrm{Q}}(\alpha)$ tends to 0 as $\alpha \downarrow 0$, and our black box minimizer determined $\alpha = \alpha_{\min}$ as the global minimizer of $\phi_{\mathrm{Q}}(\alpha)$. Because of this drawback the quasi-optimality criterion is not as robust as the other regularization methods.

Table 3. Average Relative Error relerr_α of the Regularization Parameter α.

	Morozov	Gfrerer/Raus	Quasi-Optimality	GCV	L-Curve
1	$1.274 \cdot 10^{-1}$	$1.104 \cdot 10^{-1}$	$7.718 \cdot 10^{-2}$	$3.744 \cdot 10^{-2}$	$4.186 \cdot 10^{-2}$
2	$1.765 \cdot 10^{-1}$	$1.337 \cdot 10^{-1}$	$1.818 \cdot 10^{-1}$	$8.277 \cdot 10^{-2}$	$4.069 \cdot 10^{-2}$
3	$1.542 \cdot 10^{-1}$	$1.411 \cdot 10^{-1}$	$6.483 \cdot 10^{+1}$	$9.226 \cdot 10^{-2}$	$5.027 \cdot 10^{+4}$
4	$5.045 \cdot 10^{-1}$	$4.552 \cdot 10^{-1}$	$8.546 \cdot 10^{+0}$	$8.670 \cdot 10^{-2}$	$1.349 \cdot 10^{-15}$
5	$5.972 \cdot 10^{-1}$	$3.591 \cdot 10^{-1}$	$4.657 \cdot 10^{+2}$	$7.426 \cdot 10^{-2}$	$4.728 \cdot 10^{-2}$
6	$5.047 \cdot 10^{-1}$	$5.455 \cdot 10^{-1}$	$2.563 \cdot 10^{+4}$	$5.223 \cdot 10^{-2}$	$9.272 \cdot 10^{+3}$
7	$3.909 \cdot 10^{-13}$	$7.513 \cdot 10^{-14}$	$1.309 \cdot 10^{-8}$	$4.186 \cdot 10^{+5}$	$3.995 \cdot 10^{-8}$
8	$1.759 \cdot 10^{-12}$	$2.482 \cdot 10^{-12}$	$4.026 \cdot 10^{-7}$	$2.567 \cdot 10^{+19}$	$2.047 \cdot 10^{-8}$
9	$2.226 \cdot 10^{-9}$	$2.629 \cdot 10^{-9}$	$2.052 \cdot 10^{-5}$	$1.278 \cdot 10^{+10}$	$9.772 \cdot 10^{-3}$
10	$7.754 \cdot 10^{-3}$	$1.084 \cdot 10^{-2}$	$3.735 \cdot 10^{-2}$	$1.691 \cdot 10^{+13}$	$3.987 \cdot 10^{+12}$
11	$2.564 \cdot 10^{-2}$	$3.739 \cdot 10^{-5}$	$2.028 \cdot 10^{+31}$	$8.344 \cdot 10^{-1}$	$1.713 \cdot 10^{-8}$
12	$3.352 \cdot 10^{-2}$	$1.962 \cdot 10^{-2}$	$2.028 \cdot 10^{+31}$	$7.778 \cdot 10^{-1}$	$2.699 \cdot 10^{-2}$
13	$2.398 \cdot 10^{-2}$	$2.479 \cdot 10^{-2}$	$2.316 \cdot 10^{+30}$	$1.228 \cdot 10^{+0}$	$6.415 \cdot 10^{+1}$
14	$1.185 \cdot 10^{-2}$	$1.405 \cdot 10^{-2}$	$1.717 \cdot 10^{+30}$	$5.042 \cdot 10^{-1}$	$7.539 \cdot 10^{+5}$
15	$7.889 \cdot 10^{-14}$	$7.845 \cdot 10^{-14}$	$2.024 \cdot 10^{-16}$	$1.145 \cdot 10^{+25}$	$1.510 \cdot 10^{-8}$
16	$1.180 \cdot 10^{-2}$	$1.587 \cdot 10^{-2}$	$2.344 \cdot 10^{+5}$	$8.841 \cdot 10^{+3}$	$2.890 \cdot 10^{-2}$
17	$2.500 \cdot 10^{-2}$	$2.762 \cdot 10^{-2}$	$7.222 \cdot 10^{+3}$	$5.539 \cdot 10^{+9}$	$7.629 \cdot 10^{+7}$
18	$1.613 \cdot 10^{-2}$	$6.770 \cdot 10^{-2}$	$4.124 \cdot 10^{+6}$	$3.262 \cdot 10^{+15}$	$9.405 \cdot 10^{+11}$

Let α be an approximation to α_* computed by one of our approximation methods. Then we define the relative error

$$\mathrm{relerr}_\alpha := \frac{|\alpha - \alpha_*|}{\min(\alpha, \alpha_*)}.$$

Table 3 presents the average relative errors for all the test cases.

Our approximation techniques work extremely well for Morozov's discrepancy principle and the Gfrerer/Raus-method. Because $\phi_{\mathrm{M}}(\alpha)$ and $\phi_{\mathrm{GR}}(\alpha)$ are strictly increasing for $\alpha > 0$ it is easy to give lower and upper bounds on α_*.

In the cases of the quasi-optimality criterion, GCV, and the L-curve criterion, however, a global minimizer or maximizer of $\phi(\alpha)$ must be determined. As discussed in Section 6 we terminate as soon as the largest local minimizer or maximizer has been identified. This explains the large value of relerr_α for some of the test cases.

On the other hand it should be noted that even regularization parameters α of vastly different sizes can lead to acceptable solutions $\boldsymbol{x}_\alpha$. This was also observed by J. M. Varah [33, p. 175]. We call a solution $\boldsymbol{x}_\alpha$ acceptable if the norm of the corresponding residual $\|A\boldsymbol{x}_\alpha - \boldsymbol{b}\|_2$ is a good estimate for the norm of the perturbation $\|\boldsymbol{e}\|_2$ of the right-hand side. Therefore, we also compute the relative error of the residual, i.e.,

$$\mathrm{relerr}_{\mathrm{res}} := \left| \frac{\|A\boldsymbol{x}_\alpha - \boldsymbol{b}\|_2 - \|\boldsymbol{e}\|_2}{\|\boldsymbol{e}\|_2} \right|.$$

Table 4. Average Relative Error $\mathrm{relerr}_{\mathrm{res}}$ of the Residual Norm $\|\alpha(AA^{\mathrm{T}}+\alpha I)^{-1}\boldsymbol{b}\|_2$.

	Morozov	Gfrerer/Raus	Quasi-Optimality	GCV	L-Curve
1	$5.447 \cdot 10^{-4}$	$6.839 \cdot 10^{-3}$	$1.782 \cdot 10^{-2}$	$6.650 \cdot 10^{-3}$	$1.493 \cdot 10^{-2}$
2	$1.558 \cdot 10^{-3}$	$1.207 \cdot 10^{-2}$	$5.944 \cdot 10^{-1}$	$1.003 \cdot 10^{-2}$	$2.042 \cdot 10^{-2}$
3	$1.735 \cdot 10^{-3}$	$1.297 \cdot 10^{-2}$	$5.350 \cdot 10^{+1}$	$1.385 \cdot 10^{-2}$	$5.350 \cdot 10^{+1}$
4	$4.794 \cdot 10^{-4}$	$1.672 \cdot 10^{-3}$	$9.107 \cdot 10^{-3}$	$1.154 \cdot 10^{-3}$	$1.253 \cdot 10^{-2}$
5	$5.789 \cdot 10^{-4}$	$2.054 \cdot 10^{-3}$	$6.362 \cdot 10^{-1}$	$1.094 \cdot 10^{-3}$	$2.357 \cdot 10^{-2}$
6	$8.176 \cdot 10^{-4}$	$4.494 \cdot 10^{-3}$	$7.972 \cdot 10^{+0}$	$1.405 \cdot 10^{-3}$	$1.139 \cdot 10^{+1}$
7	$2.417 \cdot 10^{-15}$	$1.220 \cdot 10^{-2}$	$1.510 \cdot 10^{-2}$	$2.285 \cdot 10^{-2}$	$1.789 \cdot 10^{-2}$
8	$5.657 \cdot 10^{-14}$	$4.798 \cdot 10^{-2}$	$1.557 \cdot 10^{-2}$	$2.539 \cdot 10^{-2}$	$1.785 \cdot 10^{-2}$
9	$1.931 \cdot 10^{-11}$	$8.806 \cdot 10^{-2}$	$1.792 \cdot 10^{-2}$	$2.520 \cdot 10^{-2}$	$3.329 \cdot 10^{-2}$
10	$2.535 \cdot 10^{-4}$	$8.193 \cdot 10^{-2}$	$1.564 \cdot 10^{+0}$	$2.632 \cdot 10^{-2}$	$1.023 \cdot 10^{+6}$
11	$1.409 \cdot 10^{-3}$	$2.039 \cdot 10^{-2}$	$8.802 \cdot 10^{-2}$	$3.733 \cdot 10^{-2}$	$8.982 \cdot 10^{-2}$
12	$1.921 \cdot 10^{-3}$	$4.432 \cdot 10^{-2}$	$5.018 \cdot 10^{+0}$	$7.698 \cdot 10^{-2}$	$1.248 \cdot 10^{-1}$
13	$2.079 \cdot 10^{-3}$	$6.512 \cdot 10^{-2}$	$1.691 \cdot 10^{+1}$	$1.347 \cdot 10^{-1}$	$1.095 \cdot 10^{+0}$
14	$2.361 \cdot 10^{-3}$	$1.582 \cdot 10^{-1}$	$1.559 \cdot 10^{+2}$	$2.138 \cdot 10^{-1}$	$9.481 \cdot 10^{+1}$
15	$4.828 \cdot 10^{-4}$	$9.240 \cdot 10^{-3}$	$5.335 \cdot 10^{-2}$	$3.824 \cdot 10^{-2}$	$4.141 \cdot 10^{-2}$
16	$4.672 \cdot 10^{-4}$	$3.422 \cdot 10^{-2}$	$2.473 \cdot 10^{+1}$	$4.660 \cdot 10^{-2}$	$4.511 \cdot 10^{-2}$
17	$1.089 \cdot 10^{-3}$	$3.127 \cdot 10^{-1}$	$7.315 \cdot 10^{+1}$	$6.580 \cdot 10^{-2}$	$1.990 \cdot 10^{+3}$
18	$8.937 \cdot 10^{-4}$	$4.159 \cdot 10^{-2}$	$4.121 \cdot 10^{+5}$	$7.254 \cdot 10^{-2}$	$2.987 \cdot 10^{+5}$

Table 4 presents the average relative error of the residual norm for each test case.

Morozov's discrepancy principle and the Gfrerer/Raus-method give excellent estimates for $\|\boldsymbol{e}\|_2$. This could be expected because we can approximate the optimal regularization parameter α_* very well for these methods. Furthermore, GCV also provides amazingly accurate estimates for $\|\boldsymbol{e}\|_2$. This is particularly remarkable as the value of relerr_α can be quite large. On the other hand, the quasi-optimality and the L-curve criteria fail to compute acceptable solutions $\boldsymbol{x}_\alpha$ in many cases. These methods suffer from the fact that $\phi_{\mathrm{Q}}(\alpha)$ and $\phi_{\mathrm{L}}(\alpha)$ have many local minimizers/maximizers that are significantly larger than α_*. Therefore these methods are not very robust in connection with our approximation techniques.

The computational cost of our regularization techniques is proportional to the number k of Lanczos iterations. In Table 5 we give an overview of the average number of required iterations. Morozov's discrepancy principle and the Gfrerer/Raus-method require roughly the same number of iterations to converge. GCV needs more iterations but it is always able to provide good estimates for $\|\boldsymbol{e}\|_2$. On the other hand, the quasi-optimality and the L-curve criteria terminate prematurely in many cases.

Table 5. Average Number of Lanczos Iterations k.

	Morozov	Gfrerer/Raus	Quasi-Optimality	GCV	L-Curve
1	22.5	22.4	22.8	53.2	22.4
2	76.5	72.4	29.6	177.3	84.6
3	346.6	316.7	22.0	854.4	22.0
4	47.5	47.2	60.2	141.2	31.7
5	104.0	98.2	35.6	283.2	82.3
6	346.0	340.9	43.8	1450.1	29.0
7	15.0	15.0	15.0	16.2	15.0
8	15.0	15.0	15.0	15.0	15.0
9	15.0	15.0	15.0	15.0	15.1
10	22.1	21.7	17.0	28.2	15.0
11	17.0	17.0	17.0	21.3	17.0
12	27.2	25.5	17.0	52.8	25.2
13	84.7	74.4	17.0	176.7	70.1
14	350.9	329.8	17.0	726.0	17.0
15	15.0	15.0	15.0	15.1	15.0
16	15.7	15.5	15.0	23.7	16.6
17	24.8	22.4	15.0	46.4	15.0
18	70.7	80.2	15.0	151.8	15.0

8 Conclusions

We presented iterative methods to compute regularization parameters for ill-posed linear systems and linear least squares problems. Our approach is based on the Lanczos algorithm and the theory of Gauss quadrature. We are able to compute lower and upper bounds on the functions $\phi(\alpha)$ that arise in Morozov's discrepancy principle, the Gfrerer/Raus-method, the quasi-optimality criterion, GCV, and the L-curve criterion. These bounds are used to compute approximations to the optimal regularization parameter α_*.

When we apply our approximation techniques to the 5 regularization methods we obtain the following results: If the norm of the perturbation $\boldsymbol{e}$ of the right-hand side $\boldsymbol{b}$ is known, Morozov's discrepancy principle and the Gfrerer/Raus-method are reliable methods to compute a regularization parameter α. They converge equally fast and provide similar values for α. It can be argued that the Gfrerer/Raus-method is actually an improved variant of Morozov's discrepancy principle (cf [17, p. 280]).

If no accurate estimate of $\|\boldsymbol{e}\|_2$ is available then GCV works much more reliably than the quasi-optimality or the L-curve criterion. The quasi-optimality and the L-curve criteria are likely to terminate prematurely and return a regularization parameter α which is significantly too large. GCV may require more Lanczos iterations but it provides reliable answers in return.

References

1. A. B. Bakushinskii, *Remarks on choosing a regularization parameter using the quasi-optimality and ratio criterion*, U.S.S.R. Computational Mathematics and Mathematical Physics, 24/4 (1984), pp. 181–182.
2. S. Bernardson, P. McCarty and C. Thron, *Monte Carlo methods for estimating linear combinations of inverse matrix entries in lattice QCD*, Computer Physics Communications, 78 (1994), pp. 256–264.
3. P. J. Davis and P. Rabinowitz, *Methods of Numerical Integration*, Academic Press, Orlando, 1984.
4. L. Eldén, *Algorithms for the regularization of ill-conditioned least squares problems*, BIT, 17 (1977), pp. 134–145.
5. Heinz W. Engl, *Regularization methods for the stable solution of inverse problems*, Surveys on Mathematics for Industry, 3 (1993), pp. 71–143.
6. Walter Gander, *On the Linear Least Squares Problem with a Quadratic Constraint*, Habilitationsschrift ETH Zürich, STAN-CS-78-697, Stanford University, 1978.
7. Walter Gautschi, *Orthogonal polynomials: applications and computation*, Acta Numerica, to appear.
8. Helmut 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.
9. D. A. Girard, *A Fast 'Monte-Carlo Cross-Validation' Procedure for Large Least Squares Problems with Noisy Data*, Numer. Math., 56 (1989), pp. 1–23.
10. D. A. Girard, *Asymptotic optimality of the fast randomized versions of GCV and C_L in ridge regression and regularization*, The Annals of Statistics, 19 (1991), pp. 1950–1963.
11. G. H. Golub, *Some Modified Matrix Eigenvalue Problems*, SIAM Review, 15 (1973), pp. 318–334.
12. G. H. Golub, M. Heath and G. Wahba, *Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter*, Technometrics, 21 (1979), pp. 215–223.
13. G. H. Golub and W. Kahan, *Calculating the singular values and pseudo-inverse of a matrix*, SIAM J. Numer. Anal., 2 (1965), pp. 205–224.
14. G. H. Golub and C. F. Van Loan, *Matrix Computations*, Second Edition, The Johns Hopkins University Press, Baltimore, 1989.
15. Gene H. Golub and Urs von Matt, *Generalized Cross-Validation for Large Scale Problems*, Journal of Computational and Graphical Statistics, to appear, `ftp://ftp.cscs.ch/pub/CSCS/techreports/1996/TR-96-28.ps.gz`.
16. G. H. Golub and J. H. Welsch, *Calculation of Gauss Quadrature Rules*, Math. Comp., 23 (1969), pp. 221–230.
17. Martin Hanke and Per Christian Hansen, *Regularization methods for large-scale problems*, Surveys on Mathematics for Industry, 3 (1993), pp. 253–315.
18. Per Christian Hansen, *Analysis of Discrete Ill-Posed Problems by Means of the L-curve*, SIAM Review, 34 (1992), pp. 561–580.
19. Per Christian Hansen, *REGULARIZATION TOOLS: A Matlab package for analysis and solution of discrete ill-posed problems*, Numerical Algorithms, 6 (1994), pp. 1–35.
20. Per Christian Hansen and Dianne P. 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.

21. Per Christian Hansen and Dianne P. O'Leary, *Regularization Algorithms Based on Total Least Squares*, Department of Computer Science, University of Maryland, Technical Report CS-TR-3684, September 1996, `ftp://ftp.cs.umd.edu/pub/papers/papers/3684/3684.ps.Z`.
22. M. F. Hutchinson, *A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines*, Communications in Statistics, Simulation and Computation, 19 (1990), pp. 433–450.
23. V. I. Krylov, *Approximate Calculation of Integrals*, translated by A. H. Stroud, Macmillan, New York, 1962.
24. A. S. Leonov, *On the choice of regularization parameters by means of the quasi-optimality and ratio criteria*, Soviet Math. Dokl., 19 (1978), pp. 537–540.
25. V. A. Morozov, *On the solution of functional equations by the method of regularization*, Soviet Math. Dokl., 7 (1966), pp. 414–417.
26. V. A. Morozov, *Methods for Solving Incorrectly Posed Problems*, Springer, New York, 1984.
27. J. Stoer, *Einführung in die Numerische Mathematik I*, Springer-Verlag, 1983.
28. A. H. Stroud and D. Secrest, *Gaussian Quadrature Formulas*, Prentice-Hall, Englewood Cliffs, New Jersey, 1966.
29. The MathWorks Inc., *MATLAB, High-Performance Numeric Computation and Visualization Software*, Natick, Massachusetts, 1992.
30. Andrei N. Tikhonov, *Ill-Posed Problems in Natural Sciences*, Proceedings of the International Conference, TVP Science Publishers, Moscow, 1992.
31. Andrey N. Tikhonov and Vasiliy Y. Arsenin, *Solutions of Ill-Posed Problems*, John Wiley & Sons, New York, 1977.
32. A. N. Tikhonov and V. B. Glasko, *Use of the regularization method in non-linear problems*, U.S.S.R. Computational Mathematics and Mathematical Physics, 5/3 (1965), pp. 93–107.
33. J. M. Varah, *Pitfalls in the numerical solution of linear ill-posed problems*, SIAM J. Sci. Stat. Comput., 4 (1983), pp. 164–176.
34. U. von Matt, *Large Constrained Quadratic Problems*, Verlag der Fachvereine, Zürich, 1993, `http://vdf.ethz.ch/vdf/info/2023.html`.
35. Grace Wahba, *Practical approximate solutions to linear operator equations when the data are noisy*, SIAM J. Numer. Anal., 14 (1977), pp. 651–667.
36. Grace Wahba, *Spline Models for Observational Data*, CBMS-NSF, Regional Conference Series in Applied Mathematics, Vol. 59, SIAM, Philadelphia, 1990.