BIT 29 (1989), 491–504

REGULARIZATION, *GSVD* AND TRUNCATED *GSVD*

PER CHRISTIAN HANSEN

UNI•C, Building 305, Technical University of Denmark,
DK-2800 Lyngby, Denmark

Abstract.

The purpose of this paper is to analyze Tikhonov regularization in general form by means of generalized SVD (GSVD) in the same spirit as SVD is used to analyze standard-form regularization. We also define a truncated GSVD solution which is of interest in its own right and which sheds light on regularization as well. In addition, our analysis gives insight into a particular numerical method for solving the general-form problem via a transformation to standard form.

AMS subject classification: 65F20, 65F30.

Keywords: Ill-posed problems, regularization, generalized SVD, perturbation theory, generalized cross-validation, transformation to standard form.

1. Introduction.

This paper is concerned with the numerical solution of linear least-squares problems, min $\|A\,x - b\|_2$, where the matrix A is very ill-conditioned. Such problems typically arise in connection with the numerical treatment of linear ill-posed problems with a compact operator. Then it can be shown that the condition number of A must grow with A's dimension, and that the singular values of A must decay to zero without any particular gap in the spectrum [6,19]. Linear data-fitting and parameter estimation are examples of other problems that may lead to least squares problems with a very ill-conditioned matrix. For all these problems, the standard least-squares solution is useless: it may be dominated by rapid oscillations due to errors, or it may simply not fit with the underlying mathematical/physical model.

Let us first establish the notation: $\|\cdot\|$ denotes the 2-norm $\|\cdot\|_2$, $R(A)$ and $N(A)$ denote range and null space, respectively, of the matrix A, and A^+ is the pseudoinverse of A. The ith singular value of A is denoted by $\psi_i(A)$, and P_S is the orthogonal projection matrix onto the subspace S.

Part of this work was carried out while visiting the Mathematical Sciences Section, Oak Ridge National Laboratory, Tennessee, during the Numerical Linear Algebra Year 1987–88, and was supported by the Danish Natural Science Foundation.

Received May 1988. Revised May 1989.

A well-known and highly regarded method for solving discrete ill-posed problems is the method of *regularization* due to Tikhonov [14]. In its general form, the *regularized solution* x_λ is defined as the solution to the following least squares problem:

$$\min \{ \|Ax - b\|^2 + \lambda^2 \|Lx\|^2 \}. \tag{1.1}$$

Usually, the matrix L is a discrete approximation to some derivative operator. Typical examples of such L are $L_1 \in \mathbb{R}^{(n-1)\times n}$ and $L_2 \in \mathbb{R}^{(n-2)\times n}$ given by

$$L_1 = \begin{bmatrix} 1 & -1 & & & & \\ & 1 & -1 & & 0 & \\ & & 1 & -1 & & \\ & & & \ddots & \ddots & \\ & & & & \ddots & \ddots \\ & 0 & & & & 1 \quad -1 \end{bmatrix}, \quad L_2 = \begin{bmatrix} -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & 0 \\ & & \ddots & \ddots & \ddots & \\ & & & \ddots & \ddots & \ddots \\ & 0 & & & \ddots & \ddots \quad \ddots \\ & & & & -1 & 2 \quad -1 \end{bmatrix} \tag{1.2}$$

Except for a scaling factor, these matrices are approximations to the first and second derivative operators on a uniform net. The idea of including the term $\lambda^2 \|Lx\|^2$ in (1.1) is to control the smoothness or shape of the solution x. The *regularization parameter* λ controls the weight given to minimization of $\|Lx\|^2$ relative to minimization of $\|Ax - b\|^2$.

We make the following assumptions about the problem (1.1):

$$A \in \mathbb{R}^{m\times n}, \quad L \in \mathbb{R}^{p\times n}, \quad m \geq n \geq p, \quad \operatorname{rank}(L) = p \tag{1.3a}$$

$$N(A) \cap N(L) = \{0\} \Leftrightarrow \operatorname{rank}\left(\begin{bmatrix} A \\ L \end{bmatrix}\right) = n. \tag{1.3b}$$

For numerical reasons, it is also convenient to assume that L is scaled such that $\|L\| \approx \|A\|$. Notice that we make no assumption about the rank of A. Assumption (1.3b) ensures that there is a *unique* solution x_λ to (1.1) for all $\lambda > 0$. Once (1.3b) is satisfied, it is easy to show that the regularized solution x_λ is given by

$$x_\lambda = A_\lambda^I b, \quad A_\lambda^I \equiv (A^T A + \lambda^2 L^T L)^{-1} A^T. \tag{1.4}$$

The matrix A_λ^I satisfies only Penrose conditions 3 and 4 [Remark 4.1] and it is therefore not a generalized inverse of A. Eq. (1.4) should not be used for practical computations.

If $L = I_n$, the identity matrix of order n, then the regularization problem (1.1) is said to be in standard form. This is not always the optimal choice of L. An $L \neq I_n$ may provide a better filtering of the errors and/or provide a better means for singling out a particularly attractive solution when A is rank deficient. An excellent example of the latter case is given by Cox [2] who considers data approximation by bivariate splines. Here, $L = I_n$ gives a spline that is not geometrically invariant (which is unsatisfactory). But if L is taken as a discrete approximation to the Laplace operator

then the computed spline becomes geometrically invariant and thus independent of the origin of the axis system.

The case $L = I_n$ has been studied by means of the *singular value decomposition* (SVD), see e.g. [16]. Further, it is shown in [5, 7, 8] that the *truncated SVD* (TSVD), an alternative method to regularization, is very similar to regularization in standard form. Along this line, we shall in this paper use the *generalized SVD* (GSVD) to analyze regularization in general form, and our analysis sheds new light on this method. Also, we shall introduce a new method, *truncated GSVD* (TGSVD), which generalizes the TSVD method. Finally, we shall demonstrate how existing methods for standard-form regularization and TSVD can be used to solve the general problem (1.1) effectively.

In Section 2 we introduce the GSVD as a convenient tool for analysis of (1.1), and we define the TGSVD solution. In Section 3 we explain why Tikhonov regularization and TGSVD produce 'smooth' solutions, and we demonstrate why it is important that the matrix L be well-conditioned. Section 4 gives the perturbation theory for Tikhonov regularization and TGSVD, and in Section 5 we discuss generalized cross-validation as a method for choosing the regularization and truncation parameters. Finally, in Section 6 we show how (1.1) can be solved efficiently via a transformation to standard form as suggested by Eldén [3], thus avoiding explicit computation of the GSVD.

2. The GSVD of (A, L)

In this section we introduce the GSVD of the pair of matrices (A, L), write x_λ in terms of the GSVD and define the truncated GSVD solution. We also prove several important relations associated with the GSVD.

THEOREM 2.1: *Let the matrix pair (A, L) satisfy (1.3a-b). Then there exist matrices $U \in \mathbb{R}^{m \times n}$, $V \in \mathbb{R}^{p \times p}$ with $U^T U = I_n$, $V^T V = I_p$ and a nonsingular $X \in \mathbb{R}^{n \times n}$ such that*

$$A = U \Sigma X^{-1} \ , \quad L = V [M, 0] X^{-1}, \tag{2.1}$$

where

$$\Sigma = diag\,(\sigma_1, \ldots, \sigma_p, 1, \ldots, 1) \in \mathbb{R}^{n \times n} \ , \quad M = diag\,(\mu_1, \ldots, \mu_p) \in \mathbb{R}^{p \times p}, \tag{2.2a}$$

and such that

$$0 \le \sigma_1 \le \ldots \le \sigma_p \le 1 \ , \quad 1 \ge \mu_1 \ge \ldots \ge \mu_p > 0 \tag{2.2b}$$

$$\sigma_i^2 + \mu_i^2 = 1 \ , \quad i = 1, \ldots, p. \tag{2.2c}$$

PROOF: The existence of the GSVD was originally proved in [15] with a slightly different notation. The present notation is based on [1, Theorem 22.2]. ■

REMARK: It is convenient to partition the matrices U, Σ and X as follows:

$$U = [U_p, U_o], \quad \Sigma = \begin{bmatrix} \Sigma_p & 0 \\ 0 & I_o \end{bmatrix}, \quad X = [X_p, X_o] \tag{2.3}$$

where U_p and X_p have p columns, and where Σ_p is $p \times p$. The subscript 'o' is a short-hand notation for $n - p$. The quantities $\gamma_i \equiv \sigma_i/\mu_i$, $i = 1, \ldots, p$ are termed the *generalized singular values* of (A, L). Due to (2.2c), both σ_i and μ_i can be computed from the ratio γ_i:

$$\sigma_i = \gamma_i(\gamma_i^2 + 1)^{-\frac{1}{2}}, \quad \mu_i = (\gamma_i^2 + 1)^{-\frac{1}{2}}, \quad i = 1, \ldots, p. \tag{2.4}$$

If $L = I_n$, then $X^{-1} = M^{-1} V^T$ and $A = U \Sigma M^{-1} V^T$, and – except for the ordering – the generalized singular values γ_i are equal to the usual singular values of A, while $\mu_i \neq 1$ due to (2.2c).

Using the GSVD of (A, L) it is easy to derive the following expression for A_λ^I in (1.4):

$$A_\lambda^I = X \begin{bmatrix} F & 0 \\ 0 & I_o \end{bmatrix} \Sigma^+ U^T = X_p F \Sigma_p^+ U_p^T + X_o U_o^T \tag{2.5}$$

where the matrix $F = \operatorname{diag}(f_i) \in \mathbb{R}^{p \times p}$ has diagonal elements given by

$$f_i = \frac{(\sigma_i/\mu_i)^2}{(\sigma_i/\mu_i)^2 + \lambda^2}, \quad i = 1, \ldots, p. \tag{2.6}$$

A straightforward extension of truncated SVD (TSVD) [5, 7, 8, 16] to general-form regularization is easily derived from Eqs. (2.5) and (2.6). We introduce the *truncated GSVD solution* x_k by substituting for F in (2.5) a diagonal matrix with k unit elements corresponding to the k largest σ_i in Σ_p and otherwise 0, thus simply neglecting the contributions corresponding to the $p - k$ smallest σ_i.

DEFINITION 2.2: Define the matrix $\hat{\Sigma}_k^+$ by

$$\hat{\Sigma}_k^+ \equiv \operatorname{diag}(0, \ldots, 0, \sigma_{p-k+1}^{-1}, \ldots, \sigma_p^{-1}). \tag{2.7}$$

Then the truncated GSVD (TGSVD) solution x_k to (1.1), defined by neglecting the components of $X_p \Sigma_p^+ U^T$ corresponding to the $p - k$ smallest σ_i, is given by

$$x_k = A_k^I, \quad A_k^I \equiv X \begin{bmatrix} \hat{\Sigma}_k^+ & 0 \\ 0 & I_o \end{bmatrix} U^T = X_p \hat{\Sigma}_k^+ U_p^T + X_o U_o^T. \tag{2.8}$$

The matrix A_k^I satisfies Penrose conditions 2 and 3 [1, Remark 4.1].

The following results are important tools for the analysis and understanding of TGSVD and Tikhonov regularization is general form.

THEOREM 2.3: *If* $Z = \begin{bmatrix} A \\ L \end{bmatrix}$, *then*

$$\|X^{-1}\| = \|Z\| \leq \|A\| + \|L\|, \quad \|X\| = \|Z^+\| \leq \Pi_p^{-1} \tag{2.9}$$

where Π_p is defined by

$$\Pi_p \equiv \begin{cases} \min\{\|L^+\|^{-1}, \inf(A P_{N(L)})\} & , \quad p < n \\ \|L^+\|^{-1} & , \quad p = n \end{cases} . \tag{2.10}$$

Here, $\inf(A P_{N(L)})$ denotes the smallest nonzero singular value of $A P_{N(L)}$.

PROOF: It follows immediately from [1, Theorem 22.2] that $\psi_i(X^{-1}) = \psi_i(Z)$. The bound for $\|X^{-1}\|$ then follows trivially from the definition of Z. Concerning the bound for $\|X\|$, the interlacing inequalities for singular values [1, Theorem 3.5] immediately lead to:

$$\psi_i(Z) \geq \psi_i\left(\begin{bmatrix} 0 \\ L \end{bmatrix}\right) = \begin{cases} \psi_i(L) & , \quad i = 1, \ldots, p \\ 0 & , \quad i = p+1, \ldots, n \end{cases} .$$

To obtain nonzero bounds for $i = p+1, \ldots, n$, we can consider $\begin{bmatrix} A \\ 0 \end{bmatrix}$ a perturbation of $\begin{bmatrix} 0 \\ L \end{bmatrix}$. Then it follows from [13, Eq. (4)] that

$$\psi_i(Z) \geq \inf\left(\begin{bmatrix} A \\ 0 \end{bmatrix} P_{N\left(\left[\begin{smallmatrix} 0 \\ L \end{smallmatrix}\right]\right)}\right) = \inf(A P_{N(L)}), \quad i = 1, \ldots, n$$

where $\inf(\cdot)$ is defined above. Hence, a lower bound for $\|X\|^{-1} = \psi_n(Z)$ is $\psi_n(L)$ for $p = n$, and $\min\{\psi_p(L), \inf(A P_{N(L)})\}$ for $p < n$. ■

This means that if Π_p is sufficiently large and if $\|A\| \approx \|L\|$, then X is guaranteed to be wellconditioned. A large Π_p also ensures that there is no unit vector z such that $\|A x\|$ and $\|L z\|$ are small simultaneously, and we can say that Eq. (1.3b) is satisfied *numerically*. We return to the magnitude of Π_p in the next section.

We can now relate σ_i and μ_i in (2.1)-(2.2) to the usual singular values of A and L:

THEOREM 2.4: *Let $\psi_i(A)$ and $\psi_i(L)$ denote the singular values of A and L, respectively, ordered such that $\psi_1(A) \geq \psi_2(A) \geq \ldots \geq \psi_n(A) \geq 0$ and $\psi_1(L) \geq \psi_2(L) \geq \ldots \geq \psi_p(L) > 0$. Then for all $\sigma_i \neq 0$ and for all μ_i:*

$$\Pi_p \leq \frac{1}{\|Z^+\|} \leq \frac{\psi_{n-i+1}(A)}{\sigma_i} \leq \|Z\| \leq \|A\| + \|L\| \tag{2.11a}$$

$$\Pi_p \leq \frac{1}{\|Z^+\|} \leq \frac{\psi_i(L)}{\mu_i} \leq \|Z\| \leq \|A\| + \|L\| \tag{2.11b}$$

where $Z = \begin{bmatrix} A \\ L \end{bmatrix}$ and Π_p is defined in (2.10).

PROOF: For a product of two matrices, $\psi_i(A B) \leq \psi_i(A)\|B\|$ [12, p. 89]. This relation and Eq. (2.1) lead to $\psi_i(A) = \psi_i(U \Sigma X^{-1}) \leq \psi_i(\Sigma)\|X^{-1}\| = \sigma_{n-i+1}\|X^{-1}\|$ and $\sigma_{n-i+1} = \psi_i(\Sigma) = \psi_i(U^T A X) \leq \psi_i(A)\|X\|$ (where the identity $\sigma_{n-i+1} = \psi_i(\Sigma)$ is

due to the ordering in (2.2b)), and thus $\|X\|^{-1} \leq \psi_i(A)/\sigma_{n-i+1} \leq \|X^{-1}\|$. The same bounds for $\psi_i(L)/\mu_i$ are derived analogously. The upper and lower bounds in (2.11a-b) then follow immediately from Theorem 2.3. ■

3. Understanding regularization and truncation

In this section we use the GSVD of (A, L) to analyze the regularized solution $\boldsymbol{x}_\lambda$ and the TGSVD solution $\boldsymbol{x}_k$. First of all, it follows immediately from Eq. (2.5) that $\boldsymbol{x}_\lambda$ is given by

$$\boldsymbol{x}_\lambda = \boldsymbol{x}_\lambda^{(1)} + \boldsymbol{x}_\lambda^{(2)} \ , \quad \boldsymbol{x}_\lambda^{(1)} = X_p F \Sigma_p^+ U_p^T \boldsymbol{b} \ , \quad \boldsymbol{x}_\lambda^{(2)} = X_o U_o^T \boldsymbol{b}. \tag{3.1}$$

The latter component, $\boldsymbol{x}_\lambda^{(2)}$, lies in $N(L)$, and it vanishes for $p = n$. It is an intrinsic property of a discrete derivative operator L that its null space $N(L) = \text{span}\{\boldsymbol{x}_{p+1}, \ldots, \boldsymbol{x}_n\}$ is spanned by vectors which are very 'smooth' (in the sense: few sign changes). For example, for L_1 and L_2 given by (1.2) we have

$$N(L_1) = \text{span}\{[1, 1, \ldots, 1]^T\} \ , \quad N(L_2) = \text{span}\{[1, 1, \ldots, 1]^T, [1, 2, \ldots,]^T\}. \tag{3.2}$$

Thus, we are guaranteed that the component $\boldsymbol{x}_\lambda^{(2)}$ is actually 'smooth'.

Consider now the component $\boldsymbol{x}_\lambda^{(1)}$. Regularization obviously corresponds to damping the terms $(\boldsymbol{u}_i^T \boldsymbol{b}/\sigma_i)\boldsymbol{x}_i$ in $\boldsymbol{x}_\lambda^{(1)}$ corresponding to small σ_i, via the matrix F (2.6). It is easy to see that this prevents any problem associated with division by the very small σ_i. The reason why this damping, in general, also gives a 'smooth' $\boldsymbol{x}_\lambda^{(1)}$ is because the number of oscillations (or sign changes) in the generalized singular vectors $\boldsymbol{x}_i$ tend to increase with decreasing σ_i, such that the matrix F filters out the most oscillating contributions to $\boldsymbol{x}_\lambda^{(1)}$. The success of this approach to producing a 'smooth' solution relies on the following heuristic:

HEURISTIC 3.1: *The number of oscillations in the left and right singular vectors of A tend to increase with increasing i.*

Although this oscillation property of the singular vectors has only been proved for totally positive matrices A, experience with discrete ill-posed problems suggests that it is in fact satisfied for a much broader class of matrices. To see why the oscillation property carries over to the vectors $\boldsymbol{x}_i$ – in reverse order – we need the following theorem.

THEOREM 3.2: *Let $\boldsymbol{v}_k(A)$, $k = 1, \ldots, n$ denote the right singular vectors corresponding to the singular values $\psi_k(A)$ in the SVD of A, and express x_i in terms of these vectors as*

$$\boldsymbol{x}_i = \sum_{k=1}^{n} \xi_{ki} \boldsymbol{v}_k(A) \ , \quad i = 1, \ldots, p. \tag{3.3}$$

If $\sigma_i \neq 0$ then the coefficients ξ_{ki} are bounded by

$$|\xi_{ki}| \leq \begin{cases} \min\{\sigma_i/\psi_k(A), \Pi_p^{-1}\}, & \psi_k(A) \neq 0 \\ \Pi_p^{-1}, & \psi_k(A) = 0. \end{cases} \tag{3.4}$$

PROOF: It is trivial that $|\xi_{ki}| \leq \|\boldsymbol{x}_i\| \leq \|X\| \leq \Pi_p^{-1}$. The other bound in (3.4) follows from the relation

$$\sigma_i^2 = \|A\,\boldsymbol{x}_i\|^2 = \sum_{j=1}^{n} \psi_j(A)^2\,\xi_{ji}^2 \;\Leftrightarrow\; \xi_{ki}^2 = \frac{\sigma_i^2 - \sum_{j \neq k} \psi_j(A)^2\,\xi_{ji}^2}{\psi_k(A)^2} \leq \left(\frac{\sigma_i}{\psi_k(A)}\right)^2. \qquad \blacksquare$$

This theorem shows that $\boldsymbol{x}_i$ is dominated by those vectors $\boldsymbol{v}_k(A)$ for which $\psi_k(A) < \sigma_i$. Roughly speaking, $\boldsymbol{x}_1$ is dominated by $\boldsymbol{v}_n(A)$, $\boldsymbol{x}_2$ is dominated by $\boldsymbol{v}_{n-1}(A)$ and $\boldsymbol{v}_n(A)$, etc. If X is well-conditioned, then all its columns must be significantly linearly independent, such that the highly oscillating $\boldsymbol{v}_k(A)$ cannot dominate $\boldsymbol{x}_i$ for $i \approx n - p$. This explains why the oscillations in $\boldsymbol{x}_i$ tend to decrease with i, provided that A satisfies Heuristic 3.1.

Heuristic 3.1 also lets us elaborate on the quantity Π_p defined in Eq. (2.10). Remember that $N(L)$ is spanned by $n - p$ very 'smooth' vectors. Thus, the matrix $A\,P_{N(L)}$ will be dominated by the $n - p$ 'smoothest' singular vectors $\boldsymbol{u}_i(A)$ and $\boldsymbol{v}_i(A)$ of A, and $A\,P_{N(L)}$ will resemble a truncated SVD of A of rank $n - p$,

$$A\,P_{P(L)} \approx \sum_{i=1}^{n-p} \boldsymbol{u}_i(A)\psi_i(A)\boldsymbol{v}_i(A)^T,$$

and therefore $\inf(A\,P_{N(L)}) \approx \psi_{n-p}(A)$. Since $n - p$ is always a very small integer, it readily follows that $\psi_{n-p}(A) \approx \psi_1(A) = \|A\|$, such that $\inf(A\,P_{N(L)}) \approx \|A\|$. Thus, if L is reasonably scaled such that $\|L\| \approx \|A\|$, then $\|L^+\|^{-1} < \inf(A\,P_{N(L)})$, and it follows that Π_p is equal to $\|L^+\|^{-1}$. Due to Theorem 2.3 we can therefore say that $\|X\|$ is approximately bounded by $\|L^+\|$. The conclusion is that L should be a well-conditioned matrix in order to ensure that Eq. (1.3b) is satisfied numerically and also – as we shall see in the next section – to ensure a small condition number.

Returning to the TGSVD solution $\boldsymbol{x}_k$ (2.8), we now see that the neglection of the contributions $(\boldsymbol{u}_i^T\boldsymbol{b}/\sigma_i)\boldsymbol{x}_i$, corresponding to the smallest σ_i, is simply another way of achieving a regularized 'smooth' solution having basically the same properties as $\boldsymbol{x}_\lambda$. For more details about this, we refer to [10] where a careful analysis of the exact conditions in which $\boldsymbol{x}_k$ and $\boldsymbol{x}_\lambda$ are similar is carried out.

4. Perturbation analysis

In this section we give the perturbation bounds for the TGSVD solution $\boldsymbol{x}_k$ (2.8) and for the regularized solution $\boldsymbol{x}_\lambda$ (1.4). For simplicity, we restrict ourselves to the case when only the right-hand side is perturbed. This is not an unusual situation in applications such as data fitting and numerical solution of integral equations.

THEOREM 4.1: *Let $\boldsymbol{e}$ denote the perturbation of the right-hand side $\tilde{\boldsymbol{b}} = \boldsymbol{b} + \boldsymbol{e}$, and let $\tilde{\boldsymbol{x}}_k$ and $\tilde{\boldsymbol{x}}_\lambda$ denote the perturbed TGSVD and regularized solutions. Then the relative perturbations of $\boldsymbol{x}_k$ and $\boldsymbol{x}_\lambda$ are bounded as*

$$(4.1)\qquad \frac{\|\tilde{\boldsymbol{x}}_k - \boldsymbol{x}_k\|}{\|\boldsymbol{x}_k\|} \le \frac{\|A\|\,\|X\|}{\sigma_{p-k+1}}\,\kappa(X)\,\frac{\|\boldsymbol{e}\|}{\|\boldsymbol{b}_k\|}\,, \qquad \frac{\|\tilde{\boldsymbol{x}}_\lambda - \boldsymbol{x}_\lambda\|}{\|\boldsymbol{x}_\lambda\|} \le \phi_\lambda\,\|A\|\,\|X\|\,\frac{\|\boldsymbol{e}\|}{\|\boldsymbol{b}_\lambda\|}$$

where $\boldsymbol{b}_k = A\,\boldsymbol{x}_k$, $\boldsymbol{b}_\lambda = A\,\boldsymbol{x}_\lambda$, and ϕ_λ is defined by

$$(4.2)\qquad \phi_\lambda \equiv \begin{cases} 1 & \text{for} \quad \lambda > 1/\sqrt{2} \\ \dfrac{1}{2\lambda}(1-\lambda^2)^{-\frac{1}{2}} & \text{for} \quad \lambda \le 1/\sqrt{2} \end{cases}.$$

In particular, if $p = n$ the bounds simplify to

$$(4.3)\qquad \frac{\|\tilde{\boldsymbol{x}}_k - \boldsymbol{x}_k\|}{\|\boldsymbol{x}_k\|} \le \frac{\|A\|\,\|L^{-1}\|}{\gamma_{p-k+1}}\,\frac{\|\boldsymbol{e}\|}{\|\boldsymbol{b}_k\|}\,, \qquad \frac{\|\tilde{\boldsymbol{x}}_\lambda - \boldsymbol{x}_\lambda\|}{\|\boldsymbol{x}_\lambda\|} \le \frac{\|A\|\,\|L^{-1}\|}{2\lambda}\,\frac{\|\boldsymbol{e}\|}{\|\boldsymbol{b}_\lambda\|}\,.$$

PROOF: The norms of $\boldsymbol{b}_k$ and $\boldsymbol{b}_\lambda$ satisfy $\|\boldsymbol{b}_k\| \le \|A\|\,\|\boldsymbol{x}_k\|$ and $\|\boldsymbol{b}_\lambda\| \le \|A\|\,\|\boldsymbol{x}_\lambda\|$. From (2.8) we have $\tilde{\boldsymbol{x}}_k - \boldsymbol{x}_k = A_k^I \boldsymbol{e} \Rightarrow \|\tilde{\boldsymbol{x}}_k - \boldsymbol{x}_k\| \le \|X\|\,\max\{\|\hat{\Sigma}_k^+\|, 1\}\,\|U^T\|\,\|\boldsymbol{e}\|$ $= \|X\|\,\dfrac{1}{\sigma_{p-k+1}}\,\|\boldsymbol{e}\|$. This immediately establishes the left part of Eq. (4.1). For the regularized solution, (2.5) yields $\tilde{\boldsymbol{x}}_\lambda - \boldsymbol{x}_\lambda = A_\lambda^I \boldsymbol{e} \Rightarrow \|\boldsymbol{x}_\lambda - \boldsymbol{x}_\lambda\| \le$ $\|X\|\,\max\{\|F\,\Sigma_p^+\|, 1\}\,\|\boldsymbol{e}\|$ in which

$$F\,\Sigma_p^+ = \operatorname{diag}\left(\frac{(\sigma_i/\mu_i)^2}{(\sigma_i/\mu_i)^2 + \lambda^2}\cdot\frac{1}{\sigma_i}\right) = \operatorname{diag}\left(\frac{\gamma_i}{\gamma_i^2 + \lambda^2}(\gamma_i^2+1)^{\frac{1}{2}}\right)$$

where $\gamma_i = \sigma_i/\mu_i$ and we have used (2.4). If we define the function $\phi(\gamma, \lambda) = \gamma(\gamma^2+1)^{\frac{1}{2}}$ $(\gamma^2 + \lambda^2)^{-1}$, then $\|F\,\Sigma_p^+\| \le \max \phi(\gamma, \lambda)$ for all $\lambda \ge 0$. It is straightforward to show that $\max \phi(\gamma, \lambda) = \phi_\lambda$ given in (4.2), and since $\phi_\lambda \ge 1$ we have proved the right part of (4.1). To prove (4.3), we use $L = V\,M\,X^{-1} \Leftrightarrow X = L^{-1}\,V\,M$ to obtain $\|A_k^I\| \le \|L^{-1}\|\,\|M\,\hat{\Sigma}_k^+\| = \|L^{-1}\|/\gamma_{p-k+1}$ and $\|A_\lambda^I\| \le \|L^{-1}\|\,\|F\,\Sigma_n^+\| \le \|L^{-1}\|/(2\lambda)$.

REMARK: The bounds in (4.1) become much more complicated if also A is perturbed, cf. [5, Theorem 3.4] for the case $L = I_n$.

COROLLARY 4.2: *The condition numbers associated with TGSVD and regularization satisfy*

$$(4.4a)\qquad \kappa_k \equiv \lim_{\|\boldsymbol{e}\| \to 0} \sup \frac{\|\boldsymbol{x}_k - \tilde{\boldsymbol{x}}_k\|}{\|\boldsymbol{x}_k\|} \le \frac{\kappa(X)}{\sigma_{p-k+1}}\,,$$

$$(4.4b)\qquad \kappa_\lambda \equiv \lim_{\|\boldsymbol{e}\| \to 0} \sup \frac{\|\boldsymbol{x}_\lambda - \tilde{\boldsymbol{x}}_\lambda\|}{\|\boldsymbol{x}_\lambda\|} \le \phi_\lambda\,\kappa(X).$$

REMARK: For $L = I_n$, these condition numbers simplify to $\kappa_k = \psi_1(A)/\psi_k(A)$ and $\kappa_\lambda = \psi_1(A)/(2\lambda)$ as obtained in [7, Corollary 4.3].

Since $\|X\|$ is approximately bounded by $\|L^+\|$ (cf. Theorem 2.3 and the remarks about Π_p in Section 3), we see that it is always possible to choose parameters k and λ that improve the condition of the problem, i.e. that improve κ_k and κ_λ compared to the condition number $\psi_1(A)/\psi_n(A)$ of the original least-squares problem: min $\|A\,x - b\|$. It is interesting to notice that if we choose $\lambda \approx \sigma_{p-k+1}$ such that $x_\lambda \approx x_k$, cf. [10], then Theorem 4.1 shows that both solutions x_λ and x_k are approximately equally sensitive to perturbations in b.

5. The choice of λ and k

This section takes its basis in a brief discussion of the behavior of the solutions x_k and x_λ under the influence of errors appearing linearly in the right-hand side. Thus, we are given a perturbed right-hand side $\tilde{b} = b + e$, where b represents underlying exact data while e denotes the errors. We assume that e is a random vector that appears as 'white noise' in the sense that the expected value of all the quantities $|u_i^T b|$, $i = 1, \ldots, n$ is a constant independent of i. We also assume that the coefficients $|u_i^T b|$ satisfy the Discrete Picard Condition as defined in [7,10]; i.e., on the average they decay to zero faster than σ_i. This means that, for increasing values of k, the TGSVD solution x_k will contain increasing amounts $(u_i^T \tilde{b}/\sigma_i)x_i$ of both the 'signal' from b and the 'noise' from e. For small k, the 'signal' dominates. When k is increased, the contributions $(u_i^T e/\sigma_i)x_i$ from the noise eventually tend to dominate, and then x_k starts to oscillate and $\|L\,x_k\|$ grows rapidly. The proper choice of k is obviously the one for which $|u_k^T b|$ is approximately equal to the average noise-level set by all the $|u_i^T e|$. Then the signal-to-noise ratio in the solution is optimal.

Consider now the regularized solution x_λ. As long as $\tilde{b}$ satisfies the above assumptions, for any k there always exists a $\lambda \approx \sigma_{p-k+1}$ such that $x_\lambda \approx x_i$ and such that the behavior of x_λ under influence of errors e is similar to that of x_k – this is a consequence of the fact that the filter properties of regularization and TGSVD are roughly the same for $\lambda \approx \sigma_{p-k+1}$; see [10] for more details. Thus, when λ is decreased, x_λ behaves similarly to x_k when k is increased as described immediately above.

Concerning the behavior of the residuals $\tilde{r}_k = \tilde{b} - A\,\tilde{x}_k$ and $\tilde{r}_\lambda = \tilde{b} - A\,\tilde{x}_\lambda$, the relevant quantities to consider are the functions:

$$V(k) = \frac{\|\tilde{r}_k\|^2}{\operatorname{trace}(I_m - A\,A_k^I)} = \frac{\|\tilde{r}_k\|^2}{m - k - (n-p)} \tag{5.1}$$

$$V(\lambda) = \frac{\|\tilde{r}_\lambda\|^2}{\operatorname{trace}(I_m - A\,A_\lambda^I)} = \|r_\lambda\|^2 \left(m - n + \sum_{i=1}^{p} \frac{\lambda^2}{(\sigma_i/\mu_i)^2 + \lambda^2} \right)^{-1}. \tag{5.2}$$

In (5.1), $k + (n-p)$ is the number of nonzero terms in the expression (2.8) for $\tilde{x}_k$, and $m - k - (n-p)$ is therefore the number of degrees of freedom in the residual $\tilde{r}_k$.

Thus if $\tilde{b}$ satisfies the assumptions made above, then $V(k)$ will be an overall decreasing function of k until it settles at a level where it stays almost constant. Then, most of the 'signal' is extracted, and $V(k)$ becomes an estimate for the variance $\|e\|^2/m$ of the noise. The function $V(\lambda)$ behaves similarly when λ is decreased. The optimum parameters k and λ are exactly those for which $V(k)$ and $V(\lambda)$ start to level off. For more details on these aspects, cf. [8].

A computationally attractive method for determining these optimum values of k and λ is to compute the minimizers of the *generalized cross-validation* (GCV) functions [4] which, except for a constant factor, are defined as:

$$G(k) \equiv \frac{V(k)}{m - k - (n-p)}, \quad G(\lambda) \equiv \frac{V(\lambda)}{\operatorname{trace}(I_m - A A_\lambda^I)}. \tag{5.3}$$

Since the denominators of these functions are increasing functions of k and λ, $G(k)$ and $G(\lambda)$ will have *minima* at the optimal k and λ where $V(k)$ and $V(\lambda)$ start to level off. Notice that there are several other statistical reasons for choosing k and λ as the minimizers of the GCV functions, cf. e.g. [17, 18].

6. Solution via transformation to standard form

We conclude this theoretical investigation by analyzing a particular numerical method for solving (1.1) via a transformation to a standard-form problem:

$$\min\{\|\bar{A}\bar{x} - \bar{b}\|^2 + \lambda^2 \|\bar{x}\|^2\}. \tag{6.1}$$

The transformation is due to Hilgers [11] and Eldén [3], and it is performed by the following numerically stable algorithm [3]:

1) Compute the QR factorization of L^T:

$$L^T = [K_p, K_o]\begin{bmatrix} R_p \\ 0 \end{bmatrix}, \quad K_p \in \mathbb{R}^{n\times p}, \quad K_o \in \mathbb{R}^{n\times(n-p)}. \tag{6.2}$$

2) Compute the QR factorization of AK_o:

$$AK_o = [H_o, H_1]\begin{bmatrix} T_o \\ 0 \end{bmatrix}, \quad H_o \in \mathbb{R}^{m\times(n-p)}, \quad H_1 \in \mathbb{R}^{m\times(m-(n-p))}. \tag{6.3}$$

3) Solve the standard-form problem (6.1) with

$$\bar{A} = H_1^T A L^+, \quad L^+ = K_p R_p^{-T}, \quad \bar{b} = H_1^T b. \tag{6.4}$$

4) Compute the solution to (1.1) as

$$x = L^+\bar{x} + K_o T_o^{-1} H_o^T (b - A L^+ \bar{x}). \tag{6.5}$$

One might suppose that there is a connection between the standard-form regularized solution $\bar{x}_\lambda$ to (6.1) and the regularized solution x_λ to (1.1). One may also expect

a connection between the TSVD solution $\bar{x}_k$ to (6.1) and the TGSVD solution x_k to (1.1). That this is indeed the case is demonstrated by the following theorem.

THEOREM 6.1: *Let the SVD of* $\bar{A}$ (6.4) *be given by*

$$\bar{A} = \bar{U}\,\psi\,\bar{V}^T \tag{6.6}$$

and let the GSVD of (A, L) *be given by* (2.1)-(2.3). *Then*

$$\bar{U} = H_1^T U_p E \ , \quad \psi = E\,\Sigma_p M^{-1} E \ , \quad \bar{V} = V E \tag{6.7}$$

where $E = \text{antidiag}(1, \ldots, 1)$ *is the* $p \times p$ *exchange matrix. Further, let* $\bar{x}_k$ *and* $\bar{x}_\lambda$ *denote the TSVD and regularized solutions to* (6.1),

$$x_k = \bar{A}_k^+ b \ , \quad \bar{A}_k^+ \equiv \bar{V}\,\text{diag}(\psi_1^{-1}, \ldots, \psi_k^{-1}, 0, \ldots, 0)\,\bar{U}^T \tag{6.8a}$$

$$\bar{x}_\lambda = \bar{A}_\lambda^I b \ , \quad \bar{A}_\lambda^I \equiv \bar{V}\,\text{diag}\left(\frac{\psi_i^2}{\psi_i^2 + \lambda^2}\right)\Psi^+\,\bar{U}^T. \tag{6.8b}$$

Then the solutions obtained by inserting these x_k *and* x_λ *into* (6.5) *are exactly the TGSVD and regularized solutions* x_k (2.8) *and* x_λ (1.4).

PROOF: Eq. (6.7) is proved in [9, Theorem 2]. The relations between $\bar{x}_k$ and x_k, and between $\bar{x}_\lambda$ and x_λ, follow immediately by insertion of (6.7) into (6.8a-b) and noting that $\psi_i = \sigma_{p-i+1}/\mu_{p-i+1}$ such that $E\,\text{diag}\,(\psi_i^2/(\psi_i^2 + \lambda^2))E = F$. ■

Theorem 6.1 shows that when (6.1) is solved by means of standard-form regularization, then the solution obtained via insertion of $\bar{x}_\lambda$ into (6.5) is *exactly* the regularized solution x_λ to (1.1) – and not just some 'similar' solution. Likewise, when the TSVD solution $\bar{x}_k$ is inserted into (6.5), one is *guaranteed* to obtain the TGSVD solution x_k as defined in (2.8). We remind that in both $\bar{x}_k$ and x_k, $p - k$ is the number of neglected (generalized) singular values.

Another important implication of Theorem 6.1 is that it lets one compute the GSVD of (A, L) *stably* from the SVD of $\bar{A}$, without actually performing the complicated GSVD computation, when L is well-conditioned:

THEOREM 6.2: *Let* $H = [H_o, H_1]$ *denote the matrix in* (6.3), *and let* $\bar{U}\,\psi\,\bar{V}T$ *be the SVD of* $\bar{A}$. *Then the matrices* Σ_p *and* M *in the GSVD* (2.1) *of* (A, L) *can be computed from* $\psi = \text{diag}(\sigma_{p-i+1}/\mu_{p-1+1})$ *via Eq.* (2.4), *while the remaining matrices in the GSVD are given by*

$$U = [U_p, U_o] = [H_1\,\bar{U}\,E, H_o] \ , \quad V = \bar{V}E \ , \quad X = \begin{bmatrix} M^{-1}\,V^T L \\ H_o^T A \end{bmatrix}^{-1}. \tag{6.9}$$

PROOF: The expression for U is proved in [9, Theorem 2], and the expression for V follows trivially from (6.7). Concerning X, we introduce $X^{-1} = W^T = [W_p, W_o]^T$, such that

$$U^T A = \begin{bmatrix} U_p^T A \\ U_o^T A \end{bmatrix} = \Sigma W^T = \begin{bmatrix} \Sigma_p & 0 \\ 0 & I_o \end{bmatrix} \begin{bmatrix} W_p^T \\ W_o^T \end{bmatrix} = \begin{bmatrix} \Sigma_p W_p^T \\ W_o^T \end{bmatrix} \Rightarrow W_o^T = U_o^T A = H_o^T A$$

where we have ued (6.9a). Also,

$$V^T L = [M, 0] W^T = [M, 0] \begin{bmatrix} W_p^T \\ W_o^T \end{bmatrix} = M W_p^T \Rightarrow W_p^T = M^{-1} V^T L.$$

Inserting these results into $X^{-1} = W^T$, we obtain the third identity in (6.9). ∎

The crucial computations involved here are the QR factorization in (6.2) and the inversion in (6.9b), and if L is well-conditioned then both of these computations are numerically stable. The outlined method is not applicable in general when both A and L are ill-conditioned [9].

In order to determine the optimum parameters k and λ, one may of course compute the GCV-functions $G(k)$ and $G(\lambda)$ (5.3) via insertion of $\bar{\boldsymbol{x}}_k$ or $\bar{\boldsymbol{x}}_\lambda$ into (6.5). However, the optimum parameters can be determined *directly* from the problem (6.1) and the corresponding GCV function $G(\cdot)$ without reference to the original problem. This is an attractive feature that substantially simplifies the numerical procedure for solving (1.1) via (6.1). The result relies on the following theorem:

THEOREM 6.3: *Let $\boldsymbol{r}_k$ and $\boldsymbol{r}_\lambda$ denote the residuals corresponding to $\boldsymbol{x}_k$ and $\boldsymbol{x}_\lambda$, respectively. Similarly, let $\bar{\boldsymbol{x}}_k = \bar{\boldsymbol{b}} - \bar{A}\bar{\boldsymbol{x}}_k$ and $\bar{\boldsymbol{r}}_\lambda = \bar{\boldsymbol{b}} - \bar{A}\bar{\boldsymbol{x}}_\lambda$ denote the residuals corresponding to $\bar{\boldsymbol{x}}_k$ (6.8a) and $\bar{\boldsymbol{x}}_\lambda$ (6.8b). Then*

$$\|\bar{\boldsymbol{r}}_k\| = \|\boldsymbol{r}_k\| \ , \quad \|\bar{\boldsymbol{r}}_\lambda\| = \|\boldsymbol{r}_\lambda\|. \tag{6.10}$$

The corresponding GCV functions $\bar{G}(k)$ and $\bar{G}(\lambda)$ satisfy

$$\bar{G}(k) \equiv \frac{\|\bar{\boldsymbol{r}}_k\|^2}{[\operatorname{trace}(I_{m-(n-p)} - \bar{A}\bar{A}_k^+)]^2} = G(k) \tag{6.11a}$$

$$\bar{G}(\lambda) \equiv \frac{\|\bar{\boldsymbol{r}}_\lambda\|^2}{[\operatorname{trace}(I_{m-(n-p)} - \bar{A}\bar{A}_\lambda^I)]^2} = G(\lambda). \tag{6.11b}$$

PROOF: First, we extend the matrix $U = [U_p, U_o]$ in (2.3) with U_q to an orthogonal matrix: $\hat{U} \equiv [U_p, U_o, U_q] \in \mathbb{R}^{m \times m}$. Eq. (6.9a) shows that $H_o = U_o$, and therefore $H_1^T U_o = 0$. Since $H^T \hat{U}$ is orthogonal, the matrix $H_1^T [U_p, U_q]$ must also be orthogonal. If we define $\boldsymbol{b} = [\beta_1, \ldots, \beta_m]^T$, then insertion of (2.1) and (2.8) into $\boldsymbol{r}_k$ yields

$$\boldsymbol{r}_k = \hat{U}[\beta_1, \ldots, \beta_{p-k}, 0, \ldots, 0, \beta_{n+1}, \ldots, \beta_m]^T \Rightarrow$$

$$\|\boldsymbol{r}_k\|^2 = \beta_1^2 + \ldots + \beta_{p-k}^2 + \beta_{n+1}^2 + \ldots + \beta_m^2.$$

Concerning $\boldsymbol{r}_k$, Eqs. (6.4), (6.7) and (6.8) lead to $\bar{A}\bar{\boldsymbol{x}}_k = H_1^T U_p E \psi \psi_k^+ \bar{U}^T \bar{\boldsymbol{b}}$, where $\psi_k^+ = \operatorname{diag}(\psi_1^{-1}, \ldots, \psi_k^{-1}, 0, \ldots, 0)$ and $\bar{U}^T \bar{\boldsymbol{b}} = (H_1^T U_p E)^T H_1^T \boldsymbol{b} = E(H_1 H_1^T U_p)^T \boldsymbol{b} = E U_p^T \boldsymbol{b}$, such that $\bar{A}\bar{\boldsymbol{x}}_k$ becomes $\bar{A}\bar{\boldsymbol{x}}_k = H_1^T \hat{U}[0, \ldots, 0, \beta_{p-k+1}, \ldots, \beta_p, 0, \ldots, 0]^T$. Inserting this relation and $\boldsymbol{b} = H_1^T \boldsymbol{b} = H_1^T \hat{U} \boldsymbol{\beta}$ into $\bar{\boldsymbol{r}}_k$, we get:

$$\bar{r}_k = H_1^T \hat{U}[\beta_1, \ldots, \beta_{p-k}, 0, \ldots, 0, \beta_{p+1}, \ldots, \beta_m]^T =$$

$$H_1^T[U_p, U_q][\beta_1, \ldots, \beta_{p-k}, 0, \ldots, 0, \beta_{n+1}, \ldots, \beta_m]^T$$

because $H_1^T U_o = 0$. Since $H_1^T[U_p, U_q]$ is orthogonal, we see that $\|\bar{r}_k\| = \|r_k\|$. The same argument and the expression (6.8) for $\bar{x}_\lambda$ lead to the second result $\|\bar{r}_\lambda\| = \|r_\lambda\|$. Eqs. (6.11a-b) follow immediately from insertion of (6.7), (6.8a-b) and (6.10) into the expressions for $\bar{G}(k)$ and $\bar{G}(\lambda)$. ■

Theorem 6.3 shows that the minimizer of $\bar{G}(\cdot)$ is *identical* to the minimizer of $G(\cdot)$. This means that the truncation parameter k or the regularization parameter λ, chosen by the GCV method applied directly to the standard-form problem (6.1), is identical to the parameter that would be obtained by application of GCV to the original problem (1.1). The same argument shows that the variance $\|e\|^2/m$ of the errors can be estimated directly from (6.1).

In this connection we would like to emphasize the use of the TSVD method for solving (6.1) because of its simplicity and because the SVD of the transformed problem sheds light on both $\bar{x}_k$ and the original problem, due to the strong connection between the SVD of $\bar{A}$ and the GSVD of (A, L). We stress that the success of TSVD relies only on satisfaction of the Discrete Picard Condition [7,10], and not on the existence of a particular gap in the singular value spectrum of A. See also [8] where several efficient methods for computing $\bar{x}_k$ are discussed.

7. Conclusion

The first part of our investigations deals with the properties and similarities of the regularized solution and the TGSVD solution to (1.1). The main conclusion is that if A satisfies Heuristic 3.1 and if the matrix L is well-conditioned, then both Tikhonov regularization and TGSVD lead to 'smooth' solutions which are approximately identical and approximately equally sensitive to perturbations.

In the second part of our investigations we show that if (1.1) is transformed into standard form (6.1), then it is not necessary to transform the solutions back to the form (1.1) in order to determine the appropriate regularization or truncation parameter, since these parameters can be determined directly from the standard-form problem. Hence, any numerical method for solving a standard-form problem can immediately be used to solve the problem in general form.

REFERENCES

[1] Å. Björck, *Least Squares Methods;* in P. G. Ciarlet & J. L. Lions (Eds), *Handbook of Numerical Analysis, Vol. III: Finite Difference Methods – Solution of Equations in* $\mathbb{R}^n$, Elsevier (1989), to appear.

[2] M. G. Cox, *Data approximation by splines in one and two independent variables;* in A. Iserles & M. J. D. Powell, *The State of the Art in Numerical Analysis*, Clarendon Press (1987), pp. 111-138.

[3] L. Eldén, *Algorithms for regularization of ill-conditioned least-squares problems*, BIT 17 (1977), 134-145.

[4] G. H. Golub, M. T. Heath & G. Wahba, *Generalized cross-validation as a method for choosing a good ridge parameter*, Technometrics 21 (1979), 215-223.
[5] P. C. Hansen, *The truncated SVD as a method for regularization*, BIT 27 (1987), 534-553.
[6] P. C. Hansen, *Computation of the singular value expansion*, Computing 40 (1988), 185-199.
[7] P. C. Hansen, *Truncated SVD solutions to discrete ill-posed problems with ill-determined numerical rank*, SIAM J. Sci. Stat. Comput., to appear.
[8] P. C. Hansen, *Solution of ill-posed problems by means of truncated SVD*, in: R. P. Agarwal, Y. M. Chow & S. J. Wilson (Eds.), *Numerical Mathematics, Singapore* 1986, ISNM 88, Birkhäuser (1988), pp. 179-192.
[9] P. C. Hansen, *Relations between SVD and GSVD of discrete regularization problems in standard and general form*, to appear in Lin. Alg. & Its Appl.
[10] P. C. Hansen, *The discrete Picard condition for discrete ill-posed problems*, submitted to BIT.
[11] J. W. Hilgers, *On the equivalence of regularization and certain reproducing kernel Hilbert space approaches for solving first kind problems*, SIAM J. Numer. Anal. 13 (1976), 172-184.
[12] A. S. Householder, *The Theory of Matrices in Numerical Analysis*, Dover Publications, New York (1974).
[13] G. W. Stewart, *A note on the perturbation of singular values*, Lin. Alg. & Its Appl. 28 (1979), 213-216.
[14] A. N. Tikhonov, *Solution of incorrectly formulated problems and the regularization method*, Doklady Akad. Nauk SSSR 151 (1963), 501-504 = Soviet Math. 4 (1963), 1035-1038.
[15] C. F. Van Loan, *Generalizing the singular value decomposition*, SIAM J. Numer. Anal. 13 (1976), 76-83.
[16] J. M. Varah, *A practical examination of some numerical methods for linear discrete ill-posed problems*, SIAM Review 21 (1979), 100-111.
[17] C. R. Vogel, *Optimal choice of a truncation level for the truncated SVD solution of linear first kind integral equations when data are noisy*, SIAM J. Numer. Anal. 23 (1986), 109-117.
[18] G. Wahba, *Practical approximate solutions to linear operator equations when the data are noisy*, SIAM J. Numer. Anal. 14 (1977), 651-667.
[19] G. M. Wing, *Condition numbers of matrices arising from the numerical treatment of linar integral equations of the first kind*, J. Integral Equations 9 (Suppl.) (1985), 191-204.