

**Numerical Analysis Project  
Manuscript NA-86-36**

**October 1986**

# **The truncated SVD as a method for regularization**

**by**

**Pet Christian Hansen**

**Numerical Analysis Project  
Computer Science Department,  
Stanford University  
Stanford, California 94305**



# The 'truncated SVD as a method for regularization

Per Christian Hansen\*

## Abstract

The truncated singular value decomposition (TSVD) is considered as a method for regularization of ill-posed linear least squares problems. In particular, the truncated SVD solution is compared with the usual regularized solution. Necessary conditions are defined in which the two methods will yield similar results. This investigation suggests the truncated SVD as a favorable alternative to standard-form regularization in case of ill-conditioned matrices with a well-determined rank.

Key words: truncated singular value decomposition, regularization in standard form, perturbation theory for truncated SVD, numerical rank.

## 1. Introduction

This paper deals with methods for solving the unconstrained linear least squares problem

$$\min \|b - Ax\|, \quad A \in R^{m \times n}, \quad m \geq n. \quad (1)$$

Here, and throughout the paper,  $\|\cdot\| = \|\cdot\|_2$ . When the matrix  $A$  is ill-conditioned, the problem (1) is ill-posed in the sense that a small perturbation of  $b$  may lead to a large perturbation of the solution. The same is true for perturbations of  $A$ . A well-known and highly regarded method for dealing with such ill-posed problems is the method of regularization by Tikhonov [18] and Phillips [17]. In particular, regularization in standard form corresponds to defining a regularized solution  $x_\lambda$  as a function of the regularization parameter  $\lambda$ , by

$$x_\lambda \equiv \operatorname{argmin} \{ \|b - Ax\|^2 + \lambda^2 \|x\|^2 \}. \quad (2)$$

It is easy to show that  $x_\lambda$  is the least squares solution to the problems

$$\min \left\| \begin{pmatrix} b \\ 0 \end{pmatrix} - \begin{pmatrix} A \\ \lambda I_n \end{pmatrix} x \right\| \quad (3)$$

where  $I_n$  denotes the identity matrix of order  $n$ , and  $x_\lambda$  is unique since the augmented matrix in (3) has full rank.

Another well-known method for dealing with ill-conditioned matrices in problem (1) is the truncated singular value decomposition (TSVD), cf. Hanson [12] and Varah [21]. The use of the TSVD has certain similarities with the use of regularization in standard form, and it is generally known that the two methods often produce very similar results [22]. The purpose of this paper is to investigate the connection between the two methods and define necessary conditions in which the two methods will yield similar results.

---

\* On leave from Copenhagen University Observatory, Øster Voldgade 3, DK-1350 Copenhagen K, Denmark. This work was in part supported by National Science Foundation Grant Number DCR 8412314 and by a Fulbright Supplementary Grant.

The author is aware that it is often necessary to substitute  $\|Lx\|$  for  $\|x\|$  in (2), and that this may have a considerable effect on the solution [25]. However, it is stressed that analysis of the standard-form problem (2) will shed light on some general relations between TSVD and regularization, and that a problem not in standard form can be transformed into a problem in standard form as shown by Eldén [6].

The organization of the paper is as follows. In Section 2 the TSVD and standard-form regularization are stated in terms of the SVD of the matrix  $A$ , and in Section 3 a new perturbation theory for the TSVD is given. On basis of this, it is natural to divide ill-conditioned matrices into two classes of matrices with well-determined and ill-determined numerical rank, respectively, as discussed in Section 4. Section 5 treats the case of matrices with well-determined rank, and it is verified that TSVD and regularization can produce similar solutions. Finally, the case of matrices with ill-determined rank is treated in Section 6 where it is shown that the similarity of the two solutions now depends on the projection of the right-hand side  $b$  onto the left singular vectors of  $A$ .

The paper is similar in spirit to the papers of Varah [22] and Wedin [23,24], and extensions of the results of Wedin are given.

## 2. Truncated-SVD and standard-form regularization

Throughout the paper, the singular value decomposition (SVD) of the matrix  $A$  in (1) will be extensively used. To summarize the SVD briefly, let  $A$  be decomposed into the three matrices  $U$ ,  $\Sigma$ , and  $V$ :

$$A = U \Sigma V^T \quad (4)$$

where the left and right singular matrices  $U \in R^{m \times m}$  and  $V \in R^{n \times n}$  are orthogonal, and where the matrix  $\Sigma \in R^{m \times n}$  has diagonal form:

$$\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n) \quad (5)$$

The diagonal elements  $\{\sigma_i\}$  of  $\Sigma$  are the singular values of  $A$ , and they are ordered such that:

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > \sigma_{r+1} = \dots = \sigma_n = 0 \quad (6)$$

where  $r = \text{rank}(A)$ . In particular,  $\|A\| = \sigma_1$ . For a rigorous treatment of the SVD, see e.g. [10]. It is stressed that the SVD is mainly used here as a powerful analysis tool. The TSVD and the regularized solutions as defined below can be computed with much less computational effort by means of other methods [4,6].

The basic idea of TSVD as well as standard-form regularization is to impose the additional requirement on the solution that its norm be small, thus hopefully damping the contributions from the errors of the right-hand side. In the case of TSVD, this is achieved by neglectation of the components of the solution corresponding to the smallest singular values, since these contributions to the solution are most likely to be large. Thus, the TSVD of  $A$  is defined as the rank- $k$  matrix

$$A_k \equiv U \Sigma_k V^T = \sum_{i=1}^k u_i \sigma_i v_i^T, \quad \Sigma_k = \text{diag}(\sigma_1, \dots, \sigma_k, 0, \dots, 0) \in R^{m \times n} \quad (7)$$

where  $\Sigma_k$  equals  $\Sigma$  with the smallest  $n-k$  singular values replaced by zeroes, and  $k \leq r$ .  $u_i$  and  $v_i$  are the columns of the matrices  $U$  and  $V$ , respectively. When the number  $k$  is chosen properly, then the condition number  $\sigma_1/\sigma_k$  of the TSVD  $A_k$  will be small. The TSVD solution to (1), defined by:

$$x_k \equiv A_k^+ b, \quad (8)$$

is therefore not very sensitive to errors in  $b$  and  $A$ . The matrix  $A_k^+$  is the pseudoinverse of  $A_k$ :

$$A_k^+ = V \Sigma_k^+ U^T, \quad \Sigma_k^+ = \text{diag}(\sigma_1^{-1}, \dots, \sigma_k^{-1}, \mathbf{0}, \dots, \mathbf{0}) \in R^{n \times m} \quad (9)$$

and  $A_k^+$  is actually a {2,3,4}-inverse, or outer invcrsc, of A, cf. [2]. The TSVD solution can usually be computed from a Q-R factorization of A as described in [4].

Consider now regularization in standard form. As can be seen from Eq. (2), the additional requirement on the norm of the solution enters directly into the definition of the regularized solution  $x_\lambda$ . For theoretical investigations, this solution can also be expressed in terms of the SVD of A. To do this, it is convenient to write  $x_\lambda$  as

$$x_\lambda = A_\lambda^I b \quad (10)$$

where the matrix  $A_\lambda^I \in R^{n \times m}$  is a "regularized inverse", defined by:

$$A_\lambda^I \equiv (A^T A + \lambda^2 I)^{-1} A^T \quad (11)$$

( $A_\lambda^I$  is only a {3,4}-inverse of A [2] and therefore not really an invcrsc.) This matrix turns out to be closely related to the matrix  $A_k^+$  above. To see this, introduce the matrix

$$\Sigma_\lambda^+ \equiv \text{diag} \left[ \frac{\sigma_1}{\sigma_1^2 + \lambda^2}, \dots, \frac{\sigma_n}{\sigma_n^2 + \lambda^2} \right] \in R^{m \times n} \quad (12)$$

When (4) is inserted into (11), it is seen that  $A_\lambda^I$  can be written in terms of the SVD of A as

$$A_\lambda^I = V \Sigma_\lambda^+ U^T \quad (13)$$

This establishes the nice similarities between Eqs. (8), (9) and (10), (13), respectively. The matrix  $A_\lambda^I$  should not be computed in any of the forms (11) or (13); instead,  $x_\lambda$  can be computed efficiently directly from (3) as described in [6].

The most important observation from Eqs. (9) and (13) is that regularization, like the TSVD, tends to filter out the contributions to the solution corresponding to the smallest singular values [22]. To elaborate on this, the  $i$ 'th diagonal element of  $\Sigma_k^+$  as well as  $\Sigma_\lambda^+$  can be written as the  $i$ 'th diagonal element of  $\Sigma_r^+$  times a *filter factor*  $f_i$ . For the TSVD, this filter factor has the form

$$f_i = \begin{cases} 1 & \text{for } \sigma_i \geq \sigma_k \\ 0 & \text{for } \sigma_i < \sigma_k \end{cases} \quad (14)$$

corresponding to a sharp filter that simply cuts off the last  $n - k$  components. For regularization, the filter factor takes the form

$$f_i = \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2}, \quad i = 1, 2, \dots, n \quad (15)$$

corresponding to a smooth filter that damps the components corresponding to  $\sigma_i < \lambda$ . When  $k$  is chosen such that  $\sigma_k = \lambda$ , the sharp filter of the TSVD can in fact be seen as an approximation to the smooth filter of the regularization method, cf. Fig. 1. This can be taken as a hint that  $x_k$  and  $x_\lambda$  may be similar, and in Sections 5 and 6 this will be investigated further.

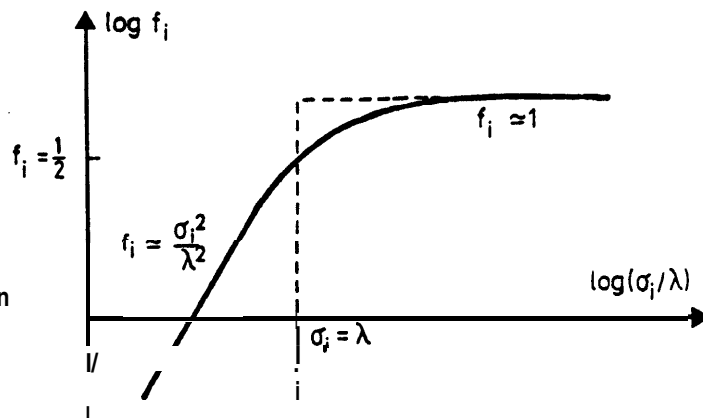


Figure 1.  
 --- Filter factor for regularization in standard form.  
 - - - Filter corresponding to truncated SVD with  $\sigma_k = \lambda$ .

### 3. Perturbation theory for the TSVD

To investigate the circumstances in which the TSVD is applicable as a regularization method, it is necessary to have a perturbation theory for the TSVD. Such a perturbation analysis is carried out in this section. The results here are strongly connected with those of Wedin [23,24].

$$\tilde{A} = A + E = \tilde{U} \tilde{\Sigma} \tilde{V}^T, \quad \tilde{b} = b + e \quad (16)$$

and let  $\tilde{x}_k$  denote the perturbed TSVD solution

$$\tilde{x}_k = \tilde{A}_k^+ \tilde{b} \quad (17)$$

with  $\tilde{A}_k$  and  $\tilde{A}_k^+$  defined as in (7) and (9). Also, define the following three useful quantities:

$$\kappa_k \equiv \|A\| \|A_k^+\| = \sigma_1 / \sigma_k \quad (18a)$$

$$\eta_k \equiv \|E\| \|A_k^+\| = \|E\| / \sigma_k = \kappa_k \frac{\|E\|}{\|A\|} \quad (18b)$$

$$\omega_k \equiv \|A - A_k\| \|A_k^+\| = \sigma_{k+1} / \sigma_k \quad (18c)$$

$\kappa_k$  is generally known as *the condition number* of  $A_k$ ,  $\eta_k$  is equal to  $\kappa_k$  times the relative error level  $\|E\|/\|A\|$ , and  $\omega_k$  is the size of the *relative gap* in the singular value spectrum between singular values  $\sigma_k$  and  $\sigma_{k+1}$ .

**Theorem 3.1.** Assume that  $\|E\| < \sigma_k$ . Then:

$$\|\tilde{A}_k^+\| = \frac{1}{\tilde{\sigma}_k} \leq \frac{1}{\sigma_k - \|E\|} = \frac{\|A_k^+\|}{1 - \eta_k} \quad (19)$$

Theorem 3.1 can also be found in e.g. [24, Lemma 3.1], but is included here for completeness. It states that  $\|\tilde{A}_k^+\|$  increases monotonically when the norm  $\|E\|$  approaches  $\sigma_k$ . Hence, for the TSVD to be useful,  $\|E\|$  must be small compared to the  $k$ 'th singular value of  $A$ , otherwise  $\tilde{A}_k^+$  may differ considerably from  $A_k^+$ . This point is elaborated in the following theorem:

**Theorem 3.2.** Assume that  $\|E\| < \sigma_k - \sigma_{k+1}$ . Then the relative error  $\|A_k^+ - \tilde{A}_k^+\|$  is bounded by:

$$\frac{\|A_k^+ - \tilde{A}_k^+\|}{\|A_k^+\|} \leq 3 \frac{\kappa_k}{(1 - \eta_k)(1 - \eta_k - \omega_k)} \frac{\|E\|}{\|A\|} \quad (20)$$

As a special case, for  $k = r = \text{rank}(A)$  and  $\|E\| < \sigma_r$ :

$$\frac{\|A^+ - \tilde{A}^+\|}{\|A^+\|} \leq 3 \frac{\kappa_r}{1 - \eta_r} \frac{\|E\|}{\|A\|} \quad (21)$$

*Proof.* The proof follows from [23, Eq. (4.6)] and Theorem 3.1 above:

$$\begin{aligned} \|A_k^+ - \tilde{A}_k^+\| &\leq \|E\| \left[ \|A_k^+\| \|\tilde{A}_k^+\| + \frac{\|A_k^+\| + \|\tilde{A}_k^+\|}{\sigma_k - \tilde{\sigma}_{k+1}} \right] \\ &\leq \|E\| \left[ \frac{\|A_k^+\|^2}{1 - \eta_k} + \frac{\|A_k^+\| + \|A_k^+\|/(1 - \eta_k)}{\sigma_k - \tilde{\sigma}_{k+1}} \right] \\ &= \|A_k^+\| \eta_k \left[ \frac{1}{1 - \eta_k} + \frac{2 - \eta_k}{(1 - \eta_k)(1 - \tilde{\sigma}_{k+1}/\sigma_k)} \right] \\ &= \|A_k^+\| \eta_k \frac{3 - \eta_k - \tilde{\sigma}_{k+1}/\sigma_k}{(1 - \eta_k)(1 - \tilde{\sigma}_{k+1}/\sigma_k)} \leq \frac{3\eta_k \|A_k^+\|}{(1 - \eta_k)(1 - \eta_k - \omega_k)}. \quad (22) \end{aligned}$$

Insertion of (18b) in (22) then yields (20). (21) follows from the fact that if  $\|E\| < \sigma_r$ , then

$\text{rank}(\tilde{A}) = \text{rank}(A)$  and  $\tilde{\sigma}_{r+1} = 0$ .  $\square$

Eq. (21) is a well-known result and states that for  $\tilde{A}^+$  to be close to  $A^+$ , both the condition number  $\kappa_k$  and the quantity  $\eta_k$  must be small. Eq. (20) is an extension of this result to the TSVD, and it is seen that for the perturbed pseudoinverse of the TSVD,  $\tilde{A}_k^+$ , to be close to  $A_k^+$  it is also necessary to require that the relative gap  $\omega_k$  be small. In other words, if the SVD is to be successfully truncated at  $k$ , then there must be a well-determined gap between the singular values  $\sigma_k$  and  $\sigma_{k+1}$ . This is also the essence of the following extension Of [23, Eq. (3.1)].

**Theorem 3.3.** Assume that  $\|E\| < \sigma_k - \sigma_{k+1}$  and let  $\theta_k$  denote the subspace angle

$$\theta_k = \theta \{S(A_k), S(\tilde{A}_k)\} \quad (23)$$

where  $S$  is any of the four fundamental subspaces  $N$ ,  $N^-$ ,  $R$ , and  $R^-$ . Then:

$$\begin{aligned} \sin \theta_k &\leq \frac{\|E\|}{\sigma_k - \tilde{\sigma}_{k+1}} = \frac{\eta_k}{1 - \tilde{\sigma}_{k+1}/\sigma_k} \\ &\leq \frac{\eta_k}{1 - \eta_k - \omega_k} = \frac{\kappa_k}{1 - \eta_k - \omega_k} \frac{\|E\|}{\|A\|} \end{aligned} \quad (24)$$

As a special case, for  $k = r = \text{rank}(A)$  and  $\|E\| < \sigma_r$ :

$$\sin \theta_r \leq \eta_r = \|E\| \|A^+\| = \kappa_r \frac{\|E\|}{\|A\|} \quad (25)$$

Finally, consider the perhaps most important result: the relative perturbation of the TSVD solution (8). The following theorem is an extension of [24, Theorem 5.1].

**Theorem 3.4.** Assume that  $\|E\| < \sigma_k - \sigma_{k+1}$ , and, let  $r_k = b - A x_k$  denote the residual corresponding to the TSVD solution  $x_k$ . Then:

$$\frac{\|x_k - \tilde{x}_k\|}{\|x_k\|} \leq \frac{\kappa_k}{1 - \eta_k} \left[ \frac{\|E\|}{\|A\|} + \frac{\|e\|}{\|b\|} + \frac{\eta_k}{1 - \eta_k - \omega_k} \frac{\|r_k\|}{\|b\|} \right] + \frac{\eta_k}{1 - \eta_k - \omega_k} \quad (26)$$

As a special case, for  $k = r = \text{rank}(A)$  and  $\|E\| < \sigma_r$ :

$$\frac{\|x_r - \tilde{x}_r\|}{\|x_r\|} \leq \frac{\kappa_r}{1 - \eta_r} \left[ \frac{\|E\|}{\|A\|} + \frac{\|e\|}{\|b\|} + \eta_r \frac{\|r_r\|}{\|b\|} \right] + \eta_r \quad (27)$$

In both equations, the denominator  $\|b\|$  can be replaced by  $\|A x_k\|$  and  $\|A\| \|x_k\|$ , thus tightening the bounds.

**Proof:** This proof follows a different line than that of [24]. The error of the TSVD solution is:

$$\begin{aligned} \tilde{x}_k - x_k &= \tilde{A}_k^+ \tilde{b} - x_k = \tilde{A}_k^+ (b + e) - x_k = \tilde{A}_k^+ (A x_k + r_k + e) - x_k \\ &= \tilde{A}_k^+ ((\tilde{A} - E) x_k + r_k + e) - x_k = \tilde{A}_k^+ (\tilde{A} x_k - E x_k + r_k + e) - x_k \\ &= \tilde{A}_k^+ (-E x_k + e + r_k) - (I_n - \tilde{A}_k^+ \tilde{A}_k) x_k \end{aligned}$$

Taking norms on both sides yields

$$\|\tilde{x}_k - x_k\| \leq \|\tilde{A}_k^+\| \left[ \|E\| \|x_k\| + \|e\| + \|r_k\| \right] + \|(I_n - \tilde{A}_k^+ \tilde{A}_k) x_k\| \quad (28)$$

The contribution to  $\|\tilde{x}_k - x_k\|$  from the vector  $-E x_k + e + r_k$  comes from its component in  $R(\tilde{A}_k)$ , the range of  $\tilde{A}_k$ , and in the worst case both  $E x_k$  and  $e$  belong to  $R(\tilde{A}_k)$ . The contribution from  $r_k$  is, however, bounded by  $\|r_k\| \sin \theta_k$ , where  $\theta_k = \theta \{R(A_k), R(\tilde{A}_k)\}$ , cf. Fig. 2(a). Concerning the second term in (28), the matrix  $I_n - \tilde{A}_k^+ \tilde{A}_k$  is the projection matrix for orthogonal projection onto  $N(A_k)$ , the null space of  $A_k$ , and from Fig. 2(b) it follows that  $\|(I_n - \tilde{A}_k^+ \tilde{A}_k) x_k\| = \|x_k\| \sin \varphi_k$ , where  $\varphi_k = \varphi \{N-(A_k), N-(\tilde{A}_k)\}$ . Upper bounds for both

angles  $\theta_k$  and  $\varphi_k$  are given in Theorem 3.3, and an upper bound for  $\|\tilde{A}_k^+\|$  is given in Theorem 3.1. This gives Eqs. (26) and (27).  $\square$

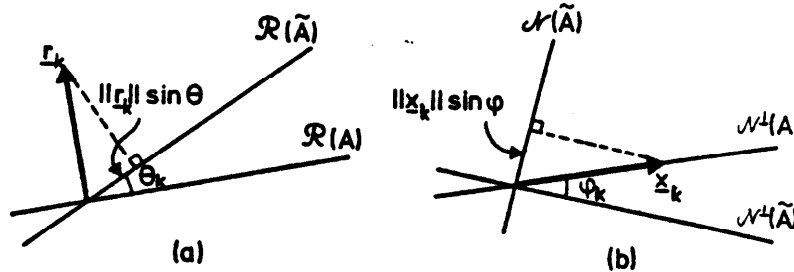


Figure 2. The contributions to (28) from (a)  $r_k$  and (b)  $x_k$ .

Again, Eq. (27) is a well-known result. The new main result (26) supplements Theorems 3.2-3.3 above and states again that the perturbed TSVD solution  $x_k$  can only be guaranteed to be close to the true solution when there is a well-determined gap between singular values  $\sigma_k$  and  $\sigma_{k+1}$ .

#### 4. Matrices with well-determined and ill-determined numerical rank

Although the concept of matrix rank is not necessary for the use of TSVD as a method for regularization, it is appropriate to discuss this concept here since it is so strongly connected with the above perturbation theory. This leads to a natural division of ill-conditioned matrices into two classes: those with well-determined numerical rank and those with ill-determined numerical rank. Such a characterization was also discussed by Golub, Kicma & Stewart [9].

It is well-known that, due to approximations, rounding errors, and other sources of errors, it is very unlikely that true zero singular values occur in practical numerical applications. It is therefore common to neglect the singular values smaller than a certain threshold, which obviously corresponds to the use of the TSVD. The choice of a suitable threshold takes its basis in the following classical perturbation bound for the singular values [LO, p. 286]:

$$|\sigma_i - \tilde{\sigma}_i| \leq \|E\|, \quad i = 1, 2, \dots, n \quad (29)$$

This implies that singular values  $\tilde{\sigma}_i$  of  $\tilde{A}$  larger than  $\|E\|$  are guaranteed to represent nonzero singular values  $\sigma_i$  of  $A$ . However, one cannot distinguish the singular values  $\tilde{\sigma}_i$  below  $\|E\|$  from exact zeroes. As a consequence, when

$$\tilde{\sigma}_k > \|E\| \geq \tilde{\sigma}_{k+1} \quad (30)$$

for some  $k$ , one can only guarantee that the rank of  $A$  is at least  $k$ .

This leads to the definition of the *numerical rank*  $r_\tau$  of  $A$ , with respect to the error level  $\tau > 0$ , as the number of singular values strictly greater than  $\tau$ :

$$\sigma_1 \geq \dots \geq \sigma_{r_\tau} > \tau \geq \sigma_{r_\tau+1} \quad (31)$$

Equivalent names for 'numerical rank' are 'effective rank' [8] and 'pseudorank' [12]. The so defined TSVD  $A_{r_\tau}$  (7) consists only of those contributions  $u_i \sigma_i v_i^T$  to  $A$  with a significant magnitude as measured by the error level  $\tau$ , while the uncertain contributions (corresponding to  $i > r_\tau$ ) are discarded.



The above definition of the numerical rank is independent of the particular distribution of the singular values of  $A$ . The numerical rank  $r_\tau$  is, however, only useful when  $r_\tau$  is well-determined with respect to  $\tau$ ; i.e.,  $r_\tau$  must be insensitive to small variations in  $\tau$ . As seen from Fig. 3, this is only the case when there is a well-defined gap between the singular values  $\sigma_{r_\tau}$  and  $\sigma_{r_\tau+1}$ . Exactly the same conclusion can be drawn from Theorems 3.2-3.4 of the perturbation theory in the above section, with  $k = r_\tau$ . Depending on the behavior of the singular value spectrum, it is therefore natural to characterize an ill-conditioned matrix as either a matrix with a *well-determined numerical rank* or an *ill-determined numerical rank*. One should look for a numerical rank only if it actually can be expected to be there, as is the case for ill-conditioned matrices with a well-determined numerical rank.

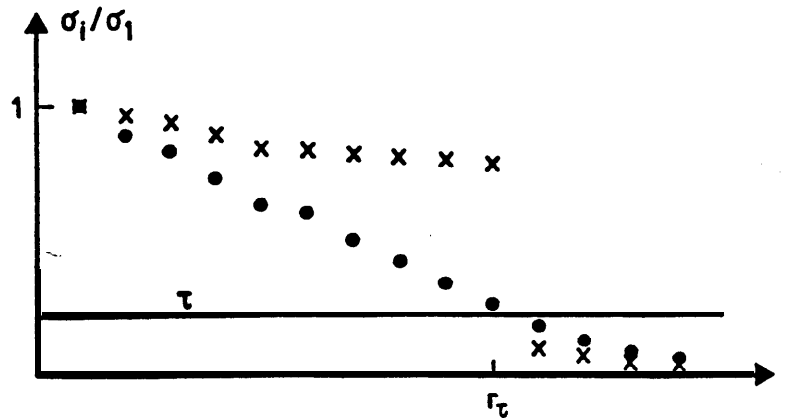


Figure 3. Singular value spectra corresponding to an ill-conditioned matrix with X well-determined and • ill-determined numerical rank.

It should be noted that the scaling of  $A$  has a considerable effect on its singular value spectrum. Inherent in the above discussion is therefore that the matrix  $A$  has been properly scaled. Good scaling strategies seem to be to scale so that, as far as possible, the uncertainties in all the elements of  $A$  are of the same order of magnitude, or so that all columns of  $A$  have approximately the same norm  $\| \cdot \|$ .

Any distribution of singular values in between the two extremes of Fig. 3 may of course be expected in practical applications. However, there are certain categories of problems that clearly lead to ill-conditioned matrices with either well-determined or ill-determined numerical rank. Matrices with well-determined numerical rank are most likely to occur when the algebraic least squares problem (1) is obtained from some underlying problem for which the concept of rank makes sense. Examples of such problems are:

- observation of signal components in noisy data [19],
- solution of some Fredholm integral equations of the first kind [7,11],
- determination of  $(A, B)$ -invariant and controllability subspaces [15,20].

Matrices with ill-determined numerical rank, on the other hand, are obtained from underlying ill-posed problems where the concept of rank has no intuitive interpretation. Examples of such problems are:

- digital image restoration [1],
- solution of integral equations in solid state physics [5],
- inverse Radon and Laplace transformation [14,16,21].

## 5. TSVD and matrices with well-determined numerical rank

When the matrix  $A$  in problem (1) has a well-determined numerical rank as discussed in Section 4, it seems natural to apply the TSVD as a method for regularization. The question then arises: under which circumstances is the TSVD solution  $x_k$  close to the regularized solution  $x_\lambda$ ? This obviously depends on the choice of the regularization parameter  $\lambda$ , and the filter factors (14) and (15) in Section 2 suggest that  $\lambda$  should be chosen somewhere between the singular values  $\sigma_k$  and  $\sigma_{k+1}$ , where  $k$  is the numerical rank. The choice of the regularization parameter  $\lambda$  is not a trivial problem as can be seen from the following theorem.

**Theorem 5.1.** *Let  $\lambda$  be chosen such that  $\lambda \in [\sigma_{k+1}, \sigma_k]$ . Then:*

$$\frac{\|A_k^+\|}{2} \leq \|A_\lambda^I\| \leq \frac{\|A_k^+\|}{2\omega_k} \quad (32)$$

where  $\omega_k$  is the relative gap (18c).

*Proof.* From Eq. (13) it follows that

$$\|A_\lambda^I\| \equiv \max_i \left\{ \frac{\sigma_i}{\sigma_i^2 + \lambda^2} \right\} = \max \left\{ \frac{\sigma_k}{\sigma_k^2 + \lambda^2}, \frac{\sigma_{k+1}}{\sigma_{k+1}^2 + \lambda^2} \right\} .$$

Both of these quantities are monotonically decreasing functions of  $\lambda$ , and they intersect at  $\lambda = (\sigma_k \sigma_{k+1})^{1/2}$ . Thus, the minimum and maximum values of  $\|A_\lambda^I\|$  are attained at  $\lambda = \sigma_k$  and  $\lambda = \sigma_{k+1}$ , respectively:

$$\begin{aligned} \lambda = \sigma_k &\Rightarrow \|A_\lambda^I\| = \frac{1}{2} \sigma_k^{-1} = \frac{1}{2} \|A_k^+\| , \\ \lambda = \sigma_{k+1} &\Rightarrow \|A_\lambda^I\| = \frac{1}{2} \sigma_{k+1}^{-1} = \frac{1}{2} \sigma_k^{-1} \omega_k = \frac{1}{2} \|A_k^+\| \omega_k^{-1} . \quad \square \end{aligned}$$

Theorem 5.1 states that if  $\omega_k$  is small, i.e. if there is a large gap in the singular value spectrum of  $A$  at  $k$ , then  $A_\lambda^I$  may differ considerably from  $A_k^+$  if  $\lambda$  is chosen close to  $\sigma_{k+1}$ . On the other hand, if  $\lambda$  is chosen close to  $\sigma_k$ , then  $A_\lambda^I$  and  $A_k^+$  might be similar. The closeness of these two matrices is investigated in the following theorem.

**Theorem 5.2.** *Assume that  $\lambda \in [\sigma_{k+1}, \sigma_k]$ . Then:*

$$\frac{\omega_k^2}{1 + \omega_k^2} \leq \min_\lambda \frac{\|A_\lambda^I - A_k^+\|}{\|A_k^+\|} \leq \frac{\omega_k^2}{1 + \omega_k^{3/2}} \quad (33)$$

and the minimum is obtained when

$$\lambda \approx (\sigma_k^2 \sigma_{k+1})^{1/2} . \quad (33a)$$

Under the same assumption:

$$\min_\lambda \|A(A_\lambda^I - A_k^+)\| = \frac{\omega_k}{1 + \omega_k} \quad (34)$$

and this minimum is obtained for

$$\lambda = (\sigma_k \sigma_{k+1})^{1/2} . \quad (34a)$$

*Proof.* The proof follows directly from the appendix and the following relations:

$$\begin{aligned} \|A_\lambda^I - A_k^+\| &= \|\Sigma_\lambda^+ - \Sigma_k^+\| = \max_i |[\xi_\lambda - \xi_k]_i| \quad \text{for } p = 0 , \\ \|A(A_\lambda^I - A_k^+)\| &= \|\Sigma(\Sigma_\lambda^+ - \Sigma_k^+)\| = \max_i |[\xi_\lambda - \xi_k]_i| \quad \text{for } p = 1 . \quad \square \end{aligned}$$

The essence of Theorem 5.2 is that the difference between the matrices  $A_\lambda^l$  and  $A_k^+$  is small when  $\lambda$  is chosen according to (33a) and when  $\omega_k$  is small, in which case the minimum (33) is approximately equal to  $\omega_k^{\frac{1}{2}}$ . The difference between the solutions  $x_\lambda$  and  $x_k$  is therefore also small for this value of  $\lambda$ . Similarly, it is seen that the difference between the residuals corresponding to the two solutions:

$$(b - A x_k) - (b - A x_\lambda) = A (A_\lambda^l - A_k^+) b, \quad (35)$$

is small when  $\lambda$  is chosen according to (34a) and when  $\omega_k$  is small, in which case the minimum (34) is approximately equal to  $\omega_k$ . There is a trade-off between (33) and (34); but since (34) is a factor  $\omega_k^{\frac{1}{2}}$  smaller than (33) it seems appropriate to choose  $\lambda$  according to (33a).

To illustrate the impact of the above theorem it is convenient to compare the two methods graphically as in Fig. 4. The solid curve, which is associated with the regularization method, is given by

$$(\|b_R - A x_\lambda\|, \|x_\lambda\|), \quad \lambda \geq 0 \quad (36)$$

where  $b_R$  is the projection of the right-hand side on the range of  $A$ . Similarly, the points marked  $X$  are associated with the TSVD and represent the point set

$$(\|b_R - A x_k\|, \|x_k\|), \quad k = 0, 1, \dots, n. \quad (37)$$

The behavior of (36) and (37) is described in the following theorem.

**Theorem 5.3.** *In (36),  $\|x_\lambda\|$  is a decreasing function of  $\|b_R - A x_\lambda\|$ , and in (37),  $\|x_k\|$  is a decreasing function of  $\|b_R - A x_k\|$  on a finite set. The curve coincides with the point set at the endpoints ( $k = 0, \lambda = 0$ ) and ( $k = n, \lambda = \infty$ ) where they both touch the axes. The remaining points of (37) lie above the solid curve (36).*

*Proof* The fact that  $\|x_\lambda\|$  and  $\|x_k\|$  are decreasing functions follows from Eqs. (3) and (10) and the following expressions:

$$\|b_R - A x_\lambda\| = \sum_{i=1}^n \frac{\lambda^2}{\sigma_i^2 + \lambda^2} (u_i^T b)^2, \quad \|b_R - A x_k\| = \sum_{i=k+1}^n (u_i^T b)^2$$

in which  $u_i$  is the  $i$ 'th column of the matrix  $U$  in (4). Eldén [6] has shown that  $x_\lambda$  can also be characterized by

$$\min \|b - A x\| \text{ subject to } \|x\| \leq \gamma$$

where  $\gamma$  is a free parameter, and that normally the solution occurs when  $\|x\| = \gamma$ . Hence,  $\|x\| \geq \|x_\lambda\|$  for any  $x$  that satisfies  $\|b - A x\| = \|b - A x_\lambda\|$ .  $\square$

Fig. 4 is drawn for the case when  $A$  has a well-determined numerical rank, and the 'corner' of the solid curve is characteristic for such matrices. It is intuitively clear that in order to yield a fair trade-off between minimization of the residual norm and the solution norm (2),  $\lambda$  should be chosen such that  $x_\lambda$  is represented by a point near the 'corner' of the curve. The figure shows that this is actually the case when  $\lambda$  is chosen according to (33a) as well as (34a). The figure also shows that the TSVD solution  $x_k$  is in fact close to  $x_\lambda$  for this choice of  $\lambda$ .

The conclusion to be drawn from this discussion is that if the matrix  $A$  is ill-conditioned and has a well-determined numerical rank, and if  $\lambda$  is chosen near the intuitive optimum value, then the TSVD solution  $x_k$  is guaranteed to be similar to the regularized solution  $x_\lambda$ . This suggests that for this class of matrices, from a theoretical as well as a computational point of view, a suitable solution to (1) is the TSVD solution  $x_k$  which, in most cases, can be computed efficiently from a Q-R factorization of the matrix  $A$  [4].

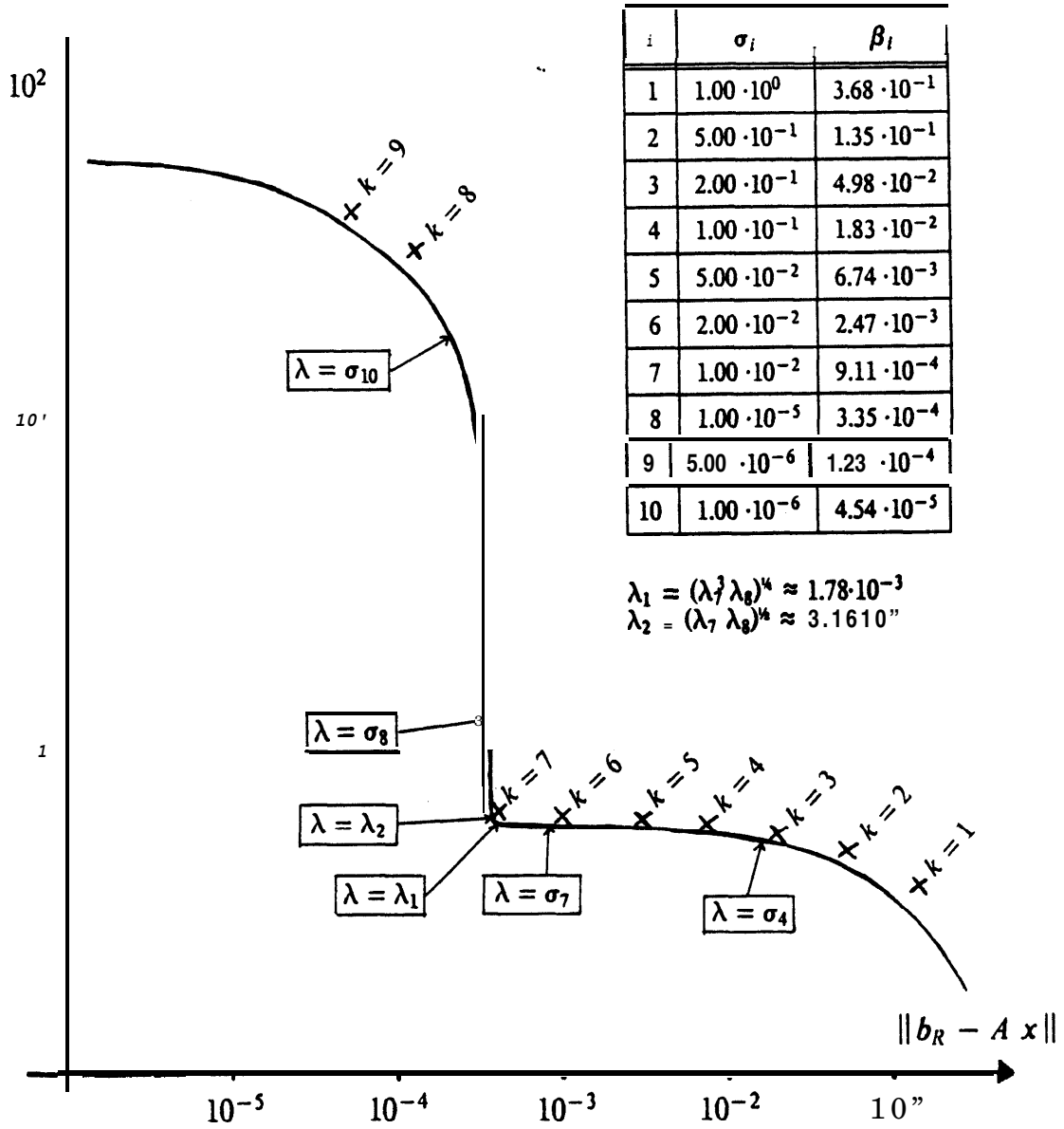


Figure 4. Comparison of the TSVID and regularization methods for an ill-conditioned matrix with well-determined numerical rank. The numerical rank is obviously equal to  $k = 7$ .

## 6. TSVD and matrices with ill-determined numerical rank

In spite of the conclusion from the previous section, the TSVD has also been reported [22] to yield results similar to those of regularization when the matrix  $A$  has an ill-determined numerical rank. For such matrices, Theorem 5.2 is not useful since it turns out that the similarity of  $x_\lambda$  and  $x_k$  now depends strongly on the right-hand side  $b$ . More precisely, it depends on the projection of  $b$  on the left singular vectors of  $A$ , as can be seen from the following theorem.

**Theorem 6.1.** Let  $\beta_i = u_i^T b$ ,  $i = 1, \dots, n$ , where  $\{u_i\}$  are the columns of the left singular matrix  $U$  in the SVD of  $A$ . Assume that the  $\{\beta_i\}$  decay as:

$$\beta_i = \sigma_i^p, \quad p = 0, 1, 2, 3, 4 \quad (38)$$

and assume that  $\lambda$  is chosen such that  $\lambda \in [\sigma_{k+1}, \sigma_k]$ . Then the difference between  $x_\lambda$  and  $x_k$  as measured by

$$d \equiv \min_{\lambda} \|x_\lambda - x_k\|_\infty, \quad \lambda \in [\sigma_{k+1}, \sigma_k] \quad (39)$$

is a function of  $\omega_k$  and  $\kappa_k$  as shown in Table 1 below. When  $\omega_k \approx 1$ , which corresponds to an  $A$  with ill-determined numerical rank, the relative measure  $d \|b\|_\infty^{-1}$  is a function of  $\kappa_k$  only as shown in Table 1.

$p$	$\beta_i$	$d$	$\frac{d}{\ b\ _\infty}, \omega_k \approx 1$
0	1	$\frac{1}{\sigma_k} \frac{\omega_k^{1/2}}{1 + \omega_k^{1/2}} \leq d \leq \frac{1}{\sigma_k} \frac{\omega_k^{1/2}}{1 + \omega_k^{3/2}}$	$\sigma_1^{-1} \frac{\kappa_k}{2}$
1	$\sigma_i$	$\frac{\omega_k}{1 + \omega_k}$	$\sigma_1^{-1} \frac{1}{2}$
2	$\sigma_i^2$	$\sigma_k \frac{\omega_k^{3/2}}{1 + \omega_k^{1/2}} \leq d \leq \sigma_k \frac{\omega_k^{3/2}}{1 + \omega_k^{3/2}}$	$\sigma_1^{-1} \frac{1}{2 \kappa_k}$
3	$\sigma_i^3$	$\sigma_1^3 \frac{\omega_k^2}{\kappa_k^2 + \omega_k^2}$	$\sigma_1^{-1} \frac{1}{\kappa_k^2}$
4	$\sigma_i^4$	$\sigma_1^4 \frac{\omega_k^2}{\kappa_k^2 + \omega_k^2}$	$\sigma_1^{-1} \frac{1}{\kappa_k^2}$

Table 1. The similarity of  $x_\lambda$  and  $x_k$  as a function the decay of  $\{\beta_i\}$ .

*Proof.* See the Appendix.

Although Theorem 6.1 covers only a very special case of right-hand sides  $b$  it gives a clear indication of the importance of the decay of the  $\{\beta_i\}$ . For the case of Fredholm integral equations of the first kind, the well-known Picard condition (cf. e.g. [22]) states that for a solution to exist, the corresponding  $\beta_i$ -coefficients must decay faster than the singular values  $\{\mu_i\}$  of the kernel such that  $\sum (\beta_i / \mu_i)^2 < \infty$ . Theorem 6.1 is a kind of 'discrete Picard condition' for the TSVD and states that the faster decay of the  $\{\beta_i\}$ , the closer  $x_k$  gets to  $x_\lambda$ .

The case  $p = 0$  is unrealistic for most practical right-hand sides  $b$  and is included here only for completeness. The case  $p = 1$  is slightly unrealistic; but it sometimes occurs in practical applications. Both cases do, however, apply to the perturbation  $e$  of  $b$  in many practical applications when the right-hand side  $b$  consists of measured quantities contaminated with measurement errors.

If, on the other hand, the coefficients  $\{u_i^T e\}$  of the perturbation  $e$  decay like the  $\beta_i$ -coefficients for  $p \geq 2$ , then Theorem 6.1 leads to a small perturbation bound on the TSVD solution. This result agrees with the perturbation bound given in terms of the 'effective condition number' as defined in [3].

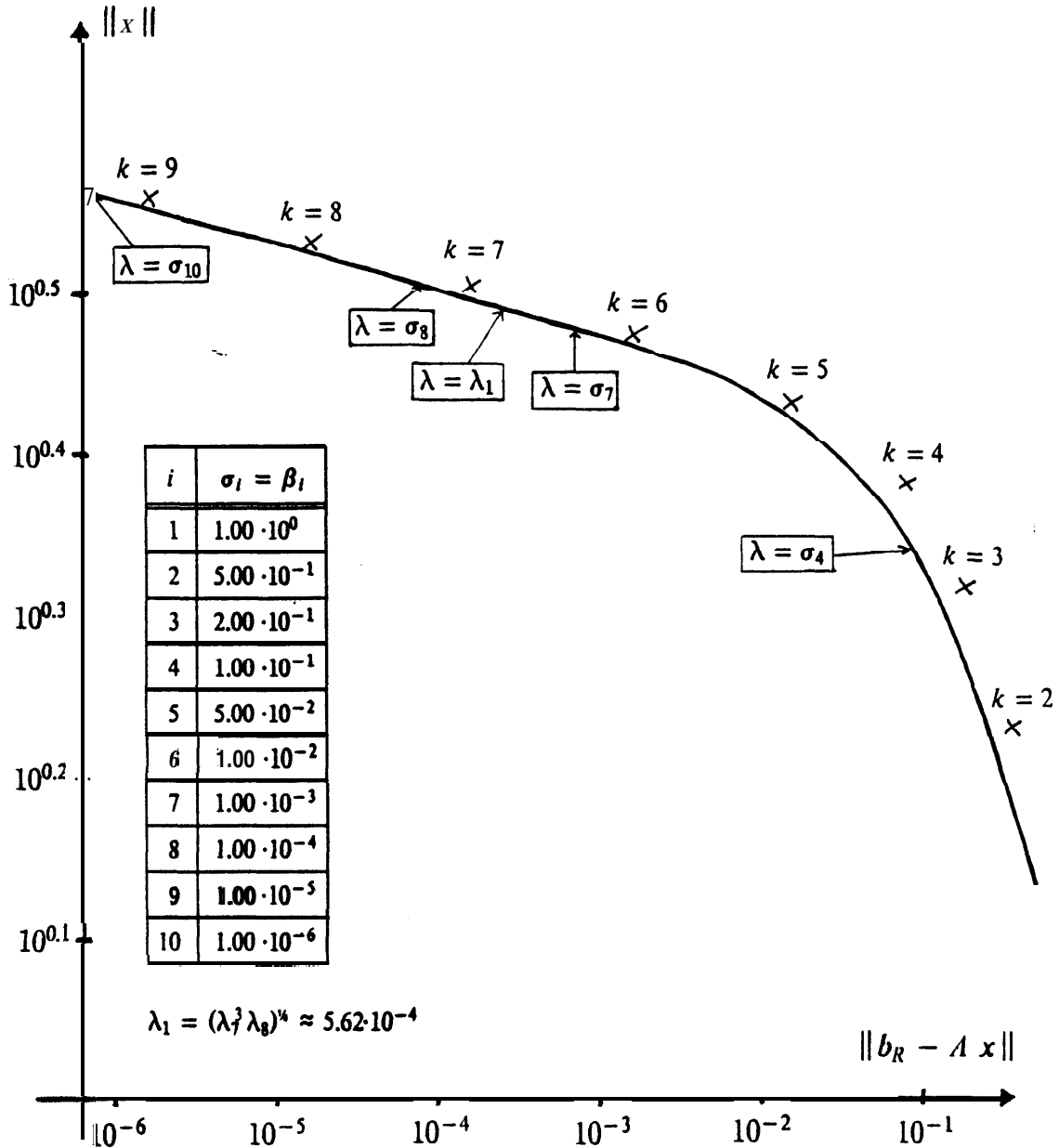


Figure 5. Comparison of the TSVD and regularization methods for an ill-conditioned matrix with ill-determined numerical rank. There is no intuitive way of choosing  $\lambda$  or  $k$ .

It is interesting to notice the different behavior of  $d$  and  $d \| b \|_{\infty}^{-1}$  for  $p \leq 2$  and for  $p > 2$ . The reason for this behavior is that for  $p \leq 2$  the maximum element of  $x_{\lambda} - x_k$  is associated with  $\sigma_k$ , while for  $p > 2$  it is associated with  $\sigma_1$ . It is also interesting to note that  $d \| b \|_{\infty}^{-1}$  actually decreases with increasing  $\kappa_k$ , thus suggesting that  $k$  be chosen as large as possible to maximize  $\kappa_k$ . The perturbation theory of Section 3 does, however, still apply and there is therefore a trade-off on  $\kappa_k$  between minimization of  $d \| b \|_{\infty}^{-1}$  and minimization of the condition number  $\kappa_k$ .

The general situation is illustrated in Fig. 5, showing the typical behavior of (36) and (37) when  $A$  has an ill-determined numerical rank for the case  $p = 1$  above. Here, there is obviously no intuitive way of choosing a suitable regularization parameter  $\lambda$ . Neither does the singular value spectrum of  $A$  suggest a suitable value of  $k$ .  $x_k$  will be close to  $x_{\lambda}$  for any  $k$  provided that the 'discrete Picard condition' is satisfied

The conclusion to be drawn in this section is that if  $A$  is ill-conditioned and has an ill-determined numerical rank then the TSVD solution  $x_k$  will be close to the regularized solution  $x_{\lambda}$  if the  $\beta_i$ -coefficients of the right-hand side  $b$  decay sufficiently fast. Hence, use of the TSVD as a regularization method might give good results. In general, one can not guarantee a small perturbation bound on the solution  $x_k$  for any value of  $k$ ; but if the perturbation  $e$  of the right-hand side also satisfies the 'discrete Picard condition' then the perturbation bound on  $x_k$  is small.

## 7. Conclusion

From a theoretical as well as a practical point of view, the truncated singular value decomposition (TSVD) is a suitable method for regularization of the ill-posed problem (1) when the coefficient matrix  $A$  is ill-conditioned with a well-determined numerical rank. If the parameter  $k$  of the TSVD  $A_k$  (7) is chosen equal to the numerical rank  $r_r$  (31) of  $A$ , then the TSVD solution is little sensitive to errors in the matrix  $A$  and right-hand side  $b$ , and the TSVD solution is close to the regularized solution with the regularization parameter chosen near its intuitive optimum value. When  $A$  has an ill-determined numerical rank, the TSVD and regularization methods may also produce similar results, provided that the  $\beta_i$ -coefficients (38) of  $b$  decay sufficiently fast, and if the corresponding coefficients of the perturbation  $e$  of  $b$  also decay sufficiently fast then the TSVD solution is little sensitive to these errors.

## Acknowledgements

I would like to thank Prof. Gene H. Golub for the invitation to visit Stanford University and for providing nice working conditions. Prof. Åke Björck and Prof. Hans Bruun Nielsen are acknowledged for valuable discussions and suggestions. I am indebted to Prof. Per-Åke Wedin for his great influence on this presentation.

## Appendix: Proofs of Theorems 5.2 and 6.1

In this appendix, the quantity  $d$  as defined in Eq. (39) is investigated for the special case when  $\beta_i = \sigma_i^p$  as in Eq. (38). Write  $x_{\lambda} = V \xi_{\lambda}$  and  $x_k = V \xi_k$ , where  $V$  is the right singular matrix in the SVD (4) of  $A$ . The elements of  $\xi_{\lambda}$  and  $\xi_k$  are then:

$$[\xi_{\lambda}]_i = \frac{\sigma_i^{p+1}}{\sigma_i^2 + \lambda^2}, \quad i = 1, \dots, n \quad \text{and} \quad [\xi_k]_i = \begin{cases} \sigma_i^{p-1}, & i = 1, \dots, k \\ 0, & i = k+1, \dots, n \end{cases} \quad (A1)$$

giving

$$[\xi_\lambda - \xi_k]_i = \begin{cases} -\sigma_i^{p-1} \frac{\lambda^2}{\sigma_i^2 + \lambda^2} & , i = 1, \dots, k \\ \sigma_i^{p-1} \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2} & , i = k+1, \dots, n \end{cases} \quad (A2)$$

and since  $\|x\| = \|V\xi\| = \|\xi\|$ , it follows that:

$$\begin{aligned} d &\equiv \min_\lambda \|x_\lambda - x_k\|_\infty = \min_\lambda \|\xi_\lambda - \xi_k\|_\infty = \min_\lambda \{ \max_i |[\xi_\lambda - \xi_k]_i| \} \\ &= \min_\lambda \left\{ \max_i \left[ \frac{\sigma_1^{p-1} \lambda^2}{\sigma_1^2 + \lambda^2}, \dots, \frac{\sigma_k^{p-1} \lambda^2}{\sigma_k^2 + \lambda^2}, \frac{\sigma_{k+1}^{p+1}}{\sigma_{k+1}^2 + \lambda^2}, \dots, \frac{\sigma_n^{p+1}}{\sigma_n^2 + \lambda^2} \right] \right\} \end{aligned} \quad (A3)$$

Consider first the situation when  $\lambda = \sigma_k$ . In this case, it is straightforward (but quite cumbersome) to show that:

$$d = f(\lambda, p) = \begin{cases} \frac{\sigma_k^{p-1} \lambda^2}{\sigma_k^2 + \lambda^2} = \frac{1}{2} \sigma_k^{p-1} & , p = 0, 1, 2 \\ \frac{\sigma_1^{p-1} \lambda^2}{\sigma_1^2 + \lambda^2} = \frac{\sigma_1^{p-1} \sigma_k^2}{\sigma_1^2 + \sigma_k^2} & , p = 3, 4 \end{cases} \quad (A4)$$

Similarly, when  $\lambda = \sigma_{k+1}$  one can show that:

$$d = g(\lambda, p) = \begin{cases} \frac{\sigma_{k+1}^{p+1}}{\sigma_{k+1}^2 + \lambda^2} = \frac{1}{2} \sigma_{k+1}^{p+1} & , p = 0, 1, 2 \\ \frac{\sigma_1^{p-1} \lambda^2}{\sigma_1^2 + \lambda^2} = \frac{\sigma_1^{p-1} \sigma_{k+1}^2}{\sigma_1^2 + \sigma_{k+1}^2} & , p = 3, 4 \end{cases} \quad (A5)$$

Hence, for  $p = 0, 1, 2$  the maximum element of  $\xi_\lambda - \xi_k$  is element no.  $k$  or  $k+1$ , while for  $p = 3, 4$  the maximum element is no. 1.

For  $p = 0, 1, 2$  the function  $f$  is an increasing function of  $\lambda$ , and  $g$  is a decreasing function of  $\lambda$ . Hence,  $d$  can be written as:

$$d = \min_\lambda \{ f(\lambda, p), g(\lambda, p) \} \quad (A6)$$

and the minimum occurs when  $f(\lambda, p) = g(\lambda, p)$ ; i.e., when elements no.  $k$  and  $k+1$  are equal. This leads to the equation

$$\frac{\sigma_k^{p-1} \lambda^2}{\sigma_k^2 + \lambda^2} = \frac{\sigma_{k+1}^{p+1}}{\sigma_{k+1}^2 + \lambda^2} \quad (A7)$$

which has the solution

$$\frac{\lambda^2}{\sigma_k^2} = \begin{cases} \omega_k^{1/2} \left[ -\frac{1}{2} \omega_k^{1/2} (\omega_k - 1) + \sqrt{1 + [\frac{1}{2} \omega_k^{1/2} (\omega_k - 1)]^2} \right] & p = 0 \\ \omega_k & p = 1 \\ \omega_k^{3/2} \left[ -\frac{1}{2} \omega_k^{1/2} (1 - \omega_k) + \sqrt{1 + [\frac{1}{2} \omega_k^{1/2} (1 - \omega_k)]^2} \right] & p = 2 \end{cases} \quad (A8)$$

Insertion of the solution for  $p = 1$  into  $f(\lambda, 1)$  gives  $\omega_k / (1 + \omega_k)$  as given in Table 1. For  $\omega_k < 1$  and  $p = 0, 2$  the above expressions simplify to



$$\frac{\lambda^2}{\sigma_k^2} \approx \begin{cases} \omega_k^4 & , \quad p = 0 \\ \omega_k^{3/2} & , \quad p = 2 \end{cases} \quad (\text{A9})$$

and insertion of this into  $f(\lambda, p)$  and  $g(\lambda, p)$  gives the upper and lower bounds in the third column of Table 1. For  $p = 3, 4$ ,  $d$  can be written as:

$$d = \min_{\lambda} f(\lambda, p) \quad (\text{A10})$$

and since  $f$  is an increasing function of  $\lambda$ , the minimum of (A10) is obtained for  $\lambda = \sigma_{k+1}$ . This leads to the remaining results in the third column of Table 1. The rightmost column of Table 1 follows from the simple fact that  $\|b\|_{\infty} = \sigma_1$  for all  $p$ .

## References

- 1 H.C. Andrews & B.R. Hunt, *Digital image restoration*, Prentice-Hall (1977).
- 2 A. Ben-Israel & T.N.E. Greville, *Generalized inverses: theory and applications*, Wiley-Interscience (1974).
- 3 T.F. Chan & D. Foulser, *Effective condition numbers for linear systems*, Tech Memo 86-05, Saxy Computer Corporation, Sunnyvale (1986).
- 4 T.F. Chan & P.C. Hansen, *Truncated SVD solutions by means of rank revealing QR-factorization* in preparation.
- 5 U. Eckhardt & K. Mika, *Numerical treatment of incorrectly posed problems - a case study*; in J. Albrecht & L. Collatz (Eds.), *Numerical treatment of integml equations Workshop on numerical treatment of integml equations Oberwolfach. November 18-24, 1979*, Birkhäuser Verlag (1980), pp. 92-101.
- 6 L. Eldén, *Algorithms for regularization of ill-conditioned least squares problems* BIT 17 (1977), 134-145.
- 7 L. Eldén, *The numerical solution of a non-characteristic cauchy problem for a pambolic equation*; in P. Duffhard & E. Hairer (Eds.), *Numerical treatment of inverse problems in differential and integral equations* Birkhäuser Verlag (1983); pp. 246-268.
- 8 G.E. Forsythe, M.A. Malcolm & C.B. Moler, *Computer methods for mathematical computations*, Prentice-Hall (1977).
- 9 G.H. Golub, V. Klema & G.W. Stewart, *Rank degeneracy and least squares problems*, Technical Report TK-456, Computer Science Department, University of Maryland (1976).
- 10 G.H. Golub & C.F. Van Loan, *Matrix Computations* North Oxford Academic (1983).
- 11 P.C. Hansen & S. Christiansen, *An SVD analysis of linear algebraic equations derived from first kind integral equations*, J. Comp. Appl. Math. 1 2&13 (1985), 341-357.
- 12 R.J. Hanson, *A numerical method for solving Fredholm integral equations of the first kind using singular values* SIAM J. Numer. Anal. 8 (1971), 616-622.
- 13 C.L. Lawson & R.J. Hanson, *Solving least squares problems* Prentice Hall (1974).
- 14 A.K. Louis & F. Natterer, *Mathematical problems of computerized tomography*, Proc. IEEE 71 (1983), 379-389.
- 15 B.C. Moore & A.J. Laub, *Computation of supremal (A, B)-invariant and controllability subspaces* IEEE Trans. Automat. Contr. AC-23 (1978), 783-792.
- 16 F. Natterer, *Numerical inversion of the Radon transform*, Numer. Math. 30 (1978), 81-91.
- 17 D.L. Phillips, *A technique for the numerical solution of certain integral equations of the first kind* J. ACM 9 (1962), 84-97.

- 18 A.N. Tikhonov, *Solution of incorrectly formulated problems and the regularization method*, Dokl. Akad. Nauk. SSSR **151** (1963), 501-504 = Soviet Math. Dokl. 4 (1963), 1035-1038.
- 19 D.W. Tufts & R. Kumaresan, *Singular value decomposition and improved frequency estimation using linear prediction*, IEEE Trans. Acoust., Speech, Signal Processing **ASSP-30** (1982), 671-675.
- 20 P.M. Van Dooren, *The generalized eigenstructure problem in linear system theory*, IEEE Trans. Automat. Contr. **AC-26** (1981) **111-129**.
- 21 J.M. Varah, *On the numerical solution of ill-conditioned linear systems with applications to ill-posed problems*, SIAM J. Numer. Anal. **10** (1973), 257-267.
- 22 J.M. Varah, *A practical examination of some numerical methods for linear discrete ill-posed problems* SIAM Review **21** (1979), 100-111.
- 23 P.-Å. Wedin, *Perturbation bounds in connection with the singular value decomposition*, BIT **12** (1972), 99-111.
- 24 P.-Å. Wedin, *Perturbation theory for pseudo-inverses*, BIT **13** (1973), 217-232.
- 25 P.-Å. Wedin, *On the almost rank deficient case of the least squares problem*, BIT **13** (1973), 344-354.