UMIACS-TR-90-124 September 1990
CS-TR 2539

Perturbation Theory for the Singular Value Decomposition

G. W. Stewart*

ABSTRACT

The singular value decomposition has a number of applications in digital signal processing. However, the the decomposition must be computed from a matrix consisting of both signal and noise. It is therefore important to be able to assess the effects of the noise on the singular values and singular vectors — a problem in classical perturbation theory. In this paper we survey the perturbation theory of the singular value decomposition.

*Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742. This work was supported in part by the Air Force Office of Sponsored Research under Contract AFOSR-87-0188.

Perturbation Theory for the Singular Value Decomposition

G. W. Stewart

Department of Computer Science and Institute for Advanced Computer Studies
University of Maryland, College Park, MD 20742

Abstract

The singular value decomposition has a number of applications in digital signal processing. However, the the decomposition must be computed from a matrix consisting of both signal and noise. It is therefore important to be able to assess the effects of the noise on the singular values and singular vectors — a problem in classical perturbation theory. In this paper we survey the perturbation theory of the singular value decomposition.

1. INTRODUCTION

This paper is concerned with the effects of errors on the singular value decomposition of a matrix. The errors arise from two sources: rounding-errors made in computing the singular value decomposition and errors initially present in the matrix. The former are generally unimportant; for if a stable algorithm is used to compute the decomposition, their effect is as if the original matrix had been very slightly perturbed. The second kind of error can be large in comparison to rounding error, and it is important to know its effect on the decomposition.

To fix our notation, let A be an $m \times n$ matrix with, say, $m \geq n$. Then there are unitary matrices U and V such that[1]

$$U^{\mathrm{T}} AV = \begin{pmatrix} \Sigma \\ 0 \end{pmatrix}, \tag{1}$$

where

$$\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_n)$$

with

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n.$$

We will call (1) the SINGULAR VALUE DECOMPOSITION of A. The numbers σ_i are called the singular values of A.[2] The columns of U are LEFT SINGULAR VECTORS and the columns of V are RIGHT SINGULAR VECTORS.

Let $\tilde{A} = A + E$ be a perturbation of A, and let

$$\tilde{U}^{\mathrm{T}} \tilde{A}\tilde{V} = \begin{pmatrix} \tilde{\Sigma} \\ 0 \end{pmatrix}$$

be the singular value decomposition of $\tilde{A}$. The question we will be concerned with is how do Σ and $\tilde{\Sigma}$ (or U and $\tilde{U}$, or V and $\tilde{V}$) compare?

There are two ways in which we can answer this question: by exhibiting either a perturbation bound or a perturbation expansion. A perturbation bound gives an upper bound on the difference between the perturbed quantity and its original—say between σ_i and $\tilde{\sigma}_i$—in terms of a norm of E.

A perturbation expansion seeks to approximate, say, $\tilde{\sigma}_i$ as a function of E. In particular, a FIRST ORDER perturbation expansion expresses $\tilde{\sigma}_i$ in the form

$$\tilde{\sigma}_i = \sigma_i + \varphi(E) O(\|E\|^2),$$

where φ is a linear function.

The two approaches tend to be complementary and work best in different circumstances. Perturbation bounds are ideal when one has a crude bound on the error, but little specific information about its structure. Perturbations expansions are

[1] The singular value decomposition was discovered independently Beltrami [1, 1873] and Jordan [9, 1874]. Schmidt [12, 1907] used the infinite dimensional analogue of the decomposition in his work on integral equations. For elementary treatments of the singular value decomposition, see [6, 13].

[2] This unfortunate notation creates no end of confusion, since in probability and statistics the letter σ is traditionally reserved for a standard deviation.

most useful when the error is known, since it provides an approximation to the perturbed object. A survey of perturbation bounds may be found in [19]. For perturbation expansions see [10].

In this paper we will use two matrix norms, both of which reduce to the Euclidean vector norm $\|\cdot\|_2$. The first is the spectral norm, also written $\|\cdot\|_2$, which is defined by

$$\|E\|_2 \stackrel{\text{def}}{=} \max_{\|x\|_2=1} \|Ex\|_2.$$

The other is the Frobenius norm, defined by

$$\|E\|_{\mathrm{F}} \stackrel{\text{def}}{=} \sqrt{\sum_{i,j} \epsilon_{ij}^2}.$$

Whenever the particular choice of norm is unimportant, we will drop the subscript. Note that both norms are UNITARILY INVARIANT, in the sense that $\|U^{\mathrm{H}}EU\| = \|E\|$ for all unitary matrices U and V. From this it follows that the spectral norm of E is the largest singular value of E, while the square of the Frobenius norm is the sum of squares of the singular values of E.

2. PERTURBATION BOUNDS FOR SINGULAR VALUES

The basic perturbation bounds for the singular values of a matrix are due to Weyl [21] and Mirsky [11].

Theorem 1 (Weyl).

$$|\tilde{\sigma}_i - \sigma_i| \le \|E\|_2, \qquad i = 1, \ldots, n.$$

Theorem 2 (Mirsky).

$$\sqrt{\sum_i (\tilde{\sigma}_i - \sigma_i)^2} \le \|E\|_{\mathrm{F}}. \tag{2}$$

In its original form, Mirsky's theorem holds for an arbitrary unitarily invariant norm and includes Weyl's theorem as a special case.

There are two remarkable facts about these theorems. First, there is no restriction on the size of the error: the theorems are true for *any* E. Second they show that ordering the singular values by magnitude provides a natural pairing: we know immediately which singular value is near which.[3]

In the language of numerical analysis, Weyl's theorem states that the singular values of a matrix are perfectly conditioned — no singular value can move more than the norm of the perturbations. If we divide both sides of (2) by $\sqrt{n}$, we see that Mirsky's theorem may be paraphrased as follows: the root mean square of the errors in the singular values is bounded by the root mean square of the singular values of the error. This theorem is less precise than Weyl's theorem; but it is often more useful, since the Frobenius norm is easy to calculate.

A word of caution. The fact that singular values are perfectly conditioned does not mean that they are determined to high *relative* accuracy. If a singular value is small compared with E, it may be entirely obliterated. We will return to the problem of small singular values in §5.

3. LOW RANK APPROXIMATIONS

Perhaps the most widespread application of the singular value decomposition is the detection of rank degeneracy. If A is of rank k, then

$$\sigma_k > 0 = \sigma_{k+1} = \cdots = \sigma_n.$$

Thus if A has small singular values, then A is near a matrix of defective rank. Specifically, set $\Sigma_k = \mathrm{diag}(\sigma_1, \ldots, \sigma_k, 0, \ldots, 0)$ and

$$A_k = U \begin{pmatrix} \Sigma_k \\ 0 \end{pmatrix} V^{\mathrm{T}}.$$

Then A_k has rank not greater than k and

$$\|A_k - A\|_{\mathrm{F}}^2 = \sigma_{k+1}^2 + \cdots + \sigma_n^2.$$

The above construction shows that small singular values are a sufficient condition for rank degeneracy. But are they necessary? Could we have a nearly degenerate matrix with no small singular values? The following argument shows that a nearly degenerate matrix *must* have small singular values.

[3] Mirsky's theorem in the Frobenius norm and specialized to eigenvalues of a symmetric matrix is sometimes said to be a a corollary of the Hoffman–Wielandt theorem, which bounds the sum of squares of the perturbations of the eigenvalues of a normal matrix [8]. However, the latter theorem does not state explicitly how the eigenvalues are to be paired.

Let B be any matrix of rank not greater than k, and let the singular values of B be denoted by $\psi_1 \geq \cdots \geq \psi_n$. Then

$$\psi_{k+1} = \cdots = \psi_n = 0.$$

By Mirsky's theorem

$$\|B - A\|_{\mathrm{F}}^2 \geq \sum_{i=1}^{n} |\psi_i - \sigma_i|^2 \geq \sigma_{k+1}^2 + \cdots + \sigma_n \geq \|A_k - A\|_{\mathrm{F}}^2.$$

Thus if A is near a matrix B of rank k, then the sum of squares of the of the k smallest singular values of A is not greater than $\|B - A\|_{\mathrm{F}}^2$. Moreover, we have proved the following theorem.

Theorem 3 (Schmidt). *The matrix A_k is a matrix of rank k that is nearest A in the Frobenius norm.*

There are three comments to be made about these results. First, the theorem about the minimality of A_k is usually attributed to Eckart and Young [5, 1939]; however, the theorem was first proved by Schmidt [12, 1907] for integral operators. It was later generalized by Mirsky [11, 1960] to all unitarily invariant norms.

Second, although the singular value decomposition is widely recommended as a way of detecting rank degeneracy, it is expensive to compute. There are other techniques, based on the QR factorization, that in practice are equally reliable and require far less work [4, 15, 16, 3, 7, 2].

Finally, the singular values of a matrix change when the columns of the matrix are scaled. For example, when a column is forced to zero, one of the singular values must also approach zero. This lack of scale invariance on the part of the singular value decomposition makes its use in rank detection problematical. In particular, one usually has to know something about the structure of the errors in the elements of the matrix to make a meaningful statement about rank. For more on this problem see [16].

4. PERTURBATION EXPANSIONS

In order to obtain a perturbation expansion for a singular value we must place restrictions on the singular value and the error matrix E. Specifically, we must assume that the singular value is SIMPLE; i.e., it is not repeated. We must also assume that the error is sufficiently small.

Let $\sigma \neq 0$ be a simple singular value of A with left singular vector u and right singular vector v. Then as E approaches zero, there is a unique singular value $\tilde{\sigma}$ of $\tilde{A}$ such that

$$\tilde{\sigma} = \sigma + u^{\mathrm{T}} E v + O(\|E\|^2). \tag{3}$$

The hypothesis that $\sigma \neq 0$ is really included in the hypothesis that σ is simple — at least when $m > n$. For in this case, the last $m - n$ columns of U are null vectors of A^{T}, so that zero can be regarded as a ghost singular value with multiplicity $m - n$. In the next section we will see how these ghosts can haunt the smaller singular values of a matrix.

The perturbation expansion is quite accurate provided E is sufficiently small. How small depends on the separation of σ from its neighbors (including the ghosts). If this separation is denoted by δ, then the second order term is approximately bounded by $\|E\|^2/\delta$. In particular, $\|E\|/\delta$ should be considerably less than one — say one tenth — before one trusts the approximation.

The perturbation expansion can be used to give some insight into how much Mirsky's theorem overestimates the variation of the singular values. Let us suppose that the elements of E are independent random variables with mean zero and standard deviation ϵ. Then if the singular values of A are simple and second order terms in (3) are ignored, the perturbation in the ith singular value is $u_i^{\mathrm{T}} E v_i$, where u_i and v_i are the corresponding left and right singular vectors. Thus the expected value of the sum of squares of the errors in the singular values is

$$\mathbf{E}\left[\sum_{i=1}^{n} (u_i^{\mathrm{T}} E v_i)^2\right] = n\epsilon^2$$

while the expected value of the square of the Frobenius norm of E is

$$\mathbf{E}\left(\|E\|_{\mathrm{F}}^2\right) = mn\epsilon^2.$$

Thus Mirsky's theorem tends to overestimate the root mean square error in the singular values by a factor of $\sqrt{m}$. For more on the subject of stochastic perturbation estimates, see [18].

However, when there are multiple singular values, Mirsky's bound can be sharp. For example, let

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \tilde{A} = \begin{pmatrix} 1+\epsilon & \epsilon \\ \epsilon & 1+\epsilon \end{pmatrix}$$

The singular values of A are 1 and 1. The singular values of $\tilde{A}$ are 1 and $1+2\epsilon$. Hence the sum of squares of the differences is

$$4\epsilon^2 = \|E\|_{\mathrm{F}}^2.$$

5. SMALL SINGULAR VALUES

We have already mentioned that a rectangular matrix is haunted by $m-n$ ghost singular values of zero. One seldom sees them in applications, since they remain invariant under perturbations of the matrix; but they make their presence know through their effect on small singular values. To see what is going on let us consider another expression for the perturbed singular values [14, 17].

Let P be the orthogonal projection onto the column space of A. Let $P_\perp = I - P$. Then

$$\tilde{\sigma}_i^2 = (\sigma_i + \gamma_i)^2 + \eta_i^2, \qquad i = 1, \ldots, p, \tag{4}$$

where

$$|\gamma_i| \le \|PE\|_2$$

and

$$\inf_2(P_\perp E) \le \eta_i \le \|P_\perp E\|_2$$

Here $\inf_2(X)$ is the smallest singular value of X.

To see what this bound means let us return to the probabilistic model of the last section. First let us suppose that $\sigma_n = 0$. Then $\tilde{\sigma}_i = \gamma_i^2 + \eta_i^2$. As m grows, γ_i^2 will on the average be of order unity, while η_i^2 will be of order m. Thus instead of a zero singular value, we will find a nonzero singular value that tends to grow as $\sqrt{m}$. Thus,

> *small singular values tend to increase under perturbation, and the increment is proportional to* $\sqrt{m}$.

This result should be a caution to people who attempt to detect rank by looking for small singular values. In the presence of noise, the singular values corresponding to zero singular values in the unperturbed matrix will be larger than the noise by a factor proportional to the square root of the sample size. In particular if the signal to noise ratio is near $\sqrt{m}$, there is a real possibility of getting the rank wrong.

As σ_n grows, the term η_n becomes negligible, and the expression (4) becomes

$$\tilde{\sigma}_n \cong \sigma_n + \gamma_n.$$

In this case there is no upward bias in the perturbation of σ_n.

6. SINGULAR VECTORS AND SINGULAR SUBSPACES

The perturbation theory for singular vectors is complicated by the fact that singular vectors corresponding to close singular values are extremely sensitive. For example, consider the matrix

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1+\epsilon \end{pmatrix},$$

whose right singular vectors are given by

$$V = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

If A is perturbed to give

$$\tilde{A} = \begin{pmatrix} 1 & \epsilon \\ \epsilon & 1 \end{pmatrix},$$

then the perturbed right singular vectors are

$$\tilde{V} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Since ϵ can be as small as we like, this example shows that there are matrices for which arbitrarily small perturbations completely change its singular vectors.

At first glance this observation appears quite discouraging. There are computations in which the singular vectors are used quite freely, and the complete inaccuracy of one of them would seem to invalidate the entire computation. However, in most applications the singular vectors are used only as transformations to and from the coordinate systems in which A assumes the diagonal form

$$\begin{pmatrix} \Sigma \\ 0 \end{pmatrix}.$$

Now if the singular value decomposition is computed by a stable algorithm, the *computed* U, V and Σ satisfy

$$A + G = U \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} V^{\mathrm{T}},$$

where G represents a perturbation that is on the order of rounding the matrix A. In this case, the use of the perturbed singular vectors amounts to a negligible change in the original problem.

However, in some cases it is necessary to compute perturbation bounds on singular vectors. Since the individual singular vectors corresponding to a cluster of singular values is unstable, we must compute bounds for the subspace spanned by the singular vectors. Such a subspace is called a SINGULAR SUBSPACE of the matrix.

In order to compare singular subspaces we need a notion of distance between subspaces. For two one dimensional subspaces $\mathcal{X}$ and $\mathcal{Y}$, a natural distance is the angle between them. From elementary geometry, we know that this angle is given by

$$\angle(\mathcal{X}, \mathcal{Y}) = \cos^{-1} |x^{\mathrm{T}} y|,$$

where x and y are vectors of norm one spanning $\mathcal{X}$ and $\mathcal{Y}$.

This construction can be generalized. Let $\mathcal{X}$ and $\mathcal{Y}$ be subspaces of dimension k. Let X and Y be orthonormal bases for $\mathcal{X}$ and $\mathcal{Y}$. Finally, let $\gamma_1 \geq \cdots \geq \gamma_k$ be the singular values of $X^{\mathrm{T}} Y$. Then the numbers

$$\theta_i = \cos^{-1} \gamma_i$$

are called the CANONICAL angles between $\mathcal{X}$ and $\mathcal{Y}$. We will say that $\mathcal{X}$ and $\mathcal{Y}$ are near if the largest canonical angle, θ_1, is small.

It would take us too far afield to give a complete justification of the use of canonical angles to measure the distance between subspaces. However, it is worthwhile to note the connection between canonical angles and projections.

Let $P_{\mathcal{X}}$ and $P_{\mathcal{Y}}$ be the orthogonal projections onto $\mathcal{X}$ and $\mathcal{Y}$. If $\mathcal{X} = \mathcal{Y}$, then $P_{\mathcal{X}} = P_{\mathcal{Y}}$, and it is natural to take $\|P_{\mathcal{X}} - P_{\mathcal{Y}}\|$ as another measure of the distance between $\mathcal{X}$ and $\mathcal{Y}$. In fact, the two measures — one from canonical angles and one from projections — are essentially the same. For it can be shown that

$$\|P_{\mathcal{X}} - P_{\mathcal{Y}}\|_{\mathrm{F}} = \sqrt{2} \|\sin \Theta\|_{\mathrm{F}}.$$

Thus the two measures go to zero at the same rate. For a detailed treatment of metrics between subspaces, see [19, Ch.2].

7. WEDIN'S THEOREM

In this section we will present a perturbation bound, due to Wedin [20], for singular subspaces. It is an unusual result in that it provides a single bound for both the right and left singular subspaces corresponding to a set of singular spaces.

We will need a basis for the singular subspaces we wish to bound. Let

$$(U_1\ U_2\ U_3)^{\mathrm{H}} A (V_1\ V_2) = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \\ 0 & 0 \end{pmatrix},$$

be a singular value decomposition of A, in which the singular values are not necessarily in descending order. The singular subspaces we will bound are the column spaces of U_1 and V_1. The perturbed subspaces will be the columns spaces of $\tilde{U}_1$ and $\tilde{V}_1$ in the decomposition

$$(\tilde{U}_1\ \tilde{U}_2\ \tilde{U}_3)^{\mathrm{H}} \tilde{A} (\tilde{V}_1\ \tilde{V}_2) = \begin{pmatrix} \tilde{\Sigma}_1 & 0 \\ 0 & \tilde{\Sigma}_2 \\ 0 & 0 \end{pmatrix}.$$

Let Φ be the matrix of canonical angles between $\mathcal{R}(U_1)$ and $\mathcal{R}(\tilde{U}_1)$, and let Θ be the matrix of canonical angles between $\mathcal{R}(V_1)$ and $\mathcal{R}(\tilde{V}_1)$. Our problem is to derive bounds on Φ and Θ.

The bounds will not be cast in terms of E but in terms of the residuals

$$R = A\tilde{V}_1 - \tilde{U}_1\tilde{\Sigma}_1$$

and

$$S = A^{\mathrm{H}}\tilde{U}_1 - \tilde{V}_1\tilde{\Sigma}_1.$$

Note that if E is zero, then R and S are zero. More generally

$$\|R\| \le \|(\tilde{A} - E)\tilde{V}_1 - \tilde{U}_1\tilde{\Sigma}_1\| \le \|E\tilde{V}_1\| \le \|E\|,$$

with a similar bound for S.

We are now in a position to state Wedin's theorem.

Theorem 4 (Wedin). *If there is a $\delta > 0$ such that*

$$\min|\sigma(\tilde{\Sigma}_1) - \sigma(\Sigma_2)| \geq \delta \tag{5}$$

and

$$\min \sigma(\tilde{\Sigma}_1) \geq \delta, \tag{6}$$

then

$$\sqrt{\|\sin\Phi\|_{\mathrm{F}}^2 + \|\sin\Theta\|_{\mathrm{F}}^2} \leq \frac{\sqrt{\|R\|_{\mathrm{F}}^2 + \|S\|_{\mathrm{F}}^2}}{\delta}. \tag{7}$$

Let us make some observations about this theorem.

As we pointed out previously, the bound (7) is a combined bound. The left-hand side combines the matrices of canonical angles for the left and right singular subspaces: the right-hand side combines what might be called left and right residuals.

The conditions (5) and (6) are separation conditions. The first says that the singular values in Σ_1 are separated from those in Σ_2.[4] The second condition says that the singular values in Σ_1 are separated from the ghost singular values.

The second condition is unfortunately necessary, as the following example shows. Let

$$A = \begin{pmatrix} \epsilon & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$$

and

$$\tilde{A} = \begin{pmatrix} \epsilon & 0 \\ 0 & 1 \\ \epsilon & 0 \end{pmatrix}.$$

Then $u_1 = (1\ 0\ 0)^{\mathrm{T}}$ while $\sqrt{2}\tilde{u}_1 = (1\ 0\ 1)$, so that u_1 and $\tilde{u}_1$ subtend an angle of 45 degrees.

This example highlights an unsavory aspect of Wedin's theorem. Although the left singular vector u_1 is unstable, the corresponding right singular vector v_1 is quite stable—i.e., only u_1 is haunted by the ghost singular values. Since the bound (7) must bound the error in both vectors, it belies the relative stability of

[4] Strictly speaking, the separation is between $\tilde{\Sigma}_1$ and Σ_2. However, if E is small compared with δ, then Weyl's theorem guarantees that the two conditions are essentially equivalent.

v_1. This problem is especially acute in signal processing applications, where the interest is typically in the right singular subspaces.

Fortunately, there is a way out of the difficulty. Returning to the general problem, suppose that the singular subspace spanned by U_1 is haunted by ghost singular values. Use Wedin's theorem to bound the perturbations in the subspaces spanned by U_2 and V_2. Since $(V_1\ V_2)$ and $(\tilde{V}_1\ \tilde{V}_2)$ are both orthogonal, it follows that the same bound holds for the space spanned by V_1.

ACKNOWLEDGEMENT

This work was supported in part by the Air Force Office of Sponsored Research under Contract AFOSR-87-0188.

References

[1] E. Beltrami (1873). "Sulle Funzioni Bilineari." *Giornale di Matematiche ud uso Degli Studenti Delle Universita*, **11**, 98–106.

[2] C. H. Bischof (1990). "Incremental Condition Estimation." *SIAM Journal on Matrix Analysis and Applications*, **11**, 312–322.

[3] T. F. Chan (1987). "Rank Revealing QR Factorizations." *Linear Algebra and Its Applications*, **88/89**, 67–82.

[4] J. J. Dongarra, J. R. Bunch, C. B. Moler, and G. W. Stewart (1979). *LINPACK User's Guide.* SIAM, Philadelphia.

[5] C. Eckart and G. Young (1939). "A Principal Axis Transformation for Non-Hermitian Matrices." *Bulletin of the American Mathematical Society*, **45**, 118–121.

[6] G. H. Golub and C. F. Van Loan (1989). *Matrix Computations* (2nd ed.). Johns Hopkins University Press, Baltimore, Maryland.

[7] N. J. Higham (1987). "A Survey of Condition Number Estimation for Triangular Matrices." *SIAM Review*, **29**, 575–596.

[8] A. J. Hoffman and H. W. Wielandt (1953). "The Variation of the Spectrum of a Normal Matrix." *Duke Mathematical Journal*, **20**, 37–39.

[9] C. Jordan (1874). "Mémoire sur les formes bilinéaires." *Journal de Mathématiques Pures et Appliquées, Deuxiéme Série*, **19**, 35–54.

[10] T. Kato (1966). *Perturbation Theory for Linear Operators.* Springer Verlag, New York.

[11] L. Mirsky (1960). "Symmetric Gage Functions and Unitarily Invariant Norms." *Quarterly Journal of Mathematics*, **11**, 50–59.

[12] E. Schmidt (1907). "Zur Theorie der linearen und nichtlinearen Integralgleichungen. I Tiel. Entwicklung willkürlichen Funktionen nach System vorgeschriebener." *Mathematische Annalen*, **63**, 433–476.

[13] G. W. Stewart (1974). *Introduction to Matrix Computations.* Academic Press, New York.

[14] G. W. Stewart (1979). "A Note on the Perturbation of Singular Values." *Linear Algebra and Its Applications*, **28**, 213–216.

[15] G. W. Stewart (1980). "The Efficient Generation of Random Orthogonal Matrices with an Application to Condition Estimators." *SIAM Journal on Numerical Analysis*, **17**, 403–404.

[16] G. W. Stewart (1984). "Rank Degeneracy." *SIAM Journal on Scientific and Statistical Computing*, **5**, 403–413.

[17] G. W. Stewart (1984). "A Second Order Perturbation Expansion for Small Singular Values." *Linear Algebra and Its Applications*, **56**, 231–235.

[18] G. W. Stewart (1988). "Stochastic Perturbation Theory." Technical Report CS-TR2129, Department of Computer Science, University of Maryland. To appear in *SIAM Review.*

[19] G. W. Stewart and Ji guang Sun (1990). *Matrix Perturbation Theory.* Academic Press, Boston.

[20] P.-Å. Wedin (1972). "Perturbation Bounds in Connection with Singular Value Decomposition." *BIT*, **12**, 99–111.

[21] H. Weyl (1912). "Das asymptotische Verteilungsgestez der Eigenwert linearer partieller Differentialgleichungen (mit einer Anwendung auf der Theorie der Hohlraumstrahlung)." *Mathematische Annalen*, **71**, 441–479.