SIAM J. SCI. STAT. COMPUT.
Vol. 4, No. 4, December 1983

0196-5204/83/0404-0011 $01.25/0

IMPROVING THE ACCURACY OF COMPUTED SINGULAR VALUES*

J. J. DONGARRA†

Abstract. This paper describes a computational method for improving the accuracy of a given singular value and its associated left and right singular vectors. The method is analogous to iterative improvement for the solution of linear systems. That is, by means of a low-precision computation, an iterative algorithm is applied to increase the accuracy of the singular value and vectors; extended precision computations are used in the residual calculation. The method is related to Newton's method applied to the singular value problem and inverse iteration for the eigenvalue problem.

Key words. singular values improvement, iterative method, singular values

1. The basic algorithm. In a recent paper, Dongarra, Moler and Wilkinson [1] described an algorithm for improving an approximation to a simple eigenvalue and the corresponding eigenvector. In this paper we extend and modify the algorithm to cover the singular value problem. We know that the singular values of a matrix are well conditioned in the sense that small changes in the matrix result in small changes in the singular values. The singular vectors may not be well determined and may vary drastically with small changes in the matrix. In [3], Stewart describes a somewhat analogous procedure for determining error bounds and obtaining corrections to the singular values and vectors associated with invariant subspaces. Here we describe a procedure for improving a single or arbitrary singular value and singular vectors using the previously computed factorization.

We begin with a brief description of the basic algorithm.

Given an $m \times n$ rectangular matrix A, we are interested in the decomposition

$$A = U \Sigma V^T, \tag{1.1}$$

where U and V are unitary matrices and Σ is a rectangular diagonal matrix of the same dimension as A with real nonnegative diagonal entries. The equations can also be written as

$$Av_i = \sigma_i u_i \tag{1.2}$$

and

$$A^T u_i = \sigma_i v_i \quad \text{for each singular value } \sigma_i. \tag{1.3}$$

If σ, u, and v have been derived from some computation on a computer with finite precision or by some insight into the problem, they are generally not the true singular value and vectors, but approximations. We know, however, that there exist μ_1, μ_2, y, and z such that

$$A(v+y) = (\sigma + \mu_1)(u+z) \tag{1.4}$$

and

$$A^T(u+z) = (\sigma + \mu_2)(v+y), \tag{1.5}$$

where μ_1, μ_2, y, and z, when added to computed σ, u, and v, give the exact left and

* Received by the editors February 5, 1982. This research was supported in part by the Applied Mathematical Sciences Research Program (KC-04-02) of the Office of Energy Research of the U.S. Department of Energy under contract W-31-109-Eng-38.

† Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Illinois 60439.

right singular vectors and the exact singular value. The corrections μ_1 and μ_2 come about by the separate nature of (1.2) and (1.3). We compute the correction to σ as $\mu = (\mu_1 + \mu_2)/2$.

The above equations can be expanded to obtain

$$Ay - \sigma z - \mu_1 u = \sigma u - Av + \mu_1 z$$

and

$$A^T z - \sigma y - \mu_2 v = \sigma v - A^T u + \mu_2 y. \tag{1.6}$$

If the orthogonality conditions

$$(v+y)^T(v+y) = 1 \tag{1.7}$$

and

$$(u+z)^T(u+z) = 1$$

are included, we then have $m+n+2$ equations in $m+n+2$ unknowns. We can now rewrite the equations in matrix notation to obtain

$$\begin{pmatrix} -\sigma I & A & -u & 0 \\ A^T & -\sigma I & 0 & -v \\ 2u^T & 0 & 0 & 0 \\ 0 & 2v^T & 0 & 0 \end{pmatrix} \begin{pmatrix} z \\ y \\ \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} \sigma u - Av + \mu_1 z \\ \sigma v - A^T u + \mu_2 y \\ 1 - u^T u - z^T z \\ 1 - v^T v - y^T y \end{pmatrix}. \tag{1.8}$$

Note that this is a mildly nonlinear matrix equation. We can determine the unknowns $(z, y, \mu_1, \mu_2)^T$ iteratively by solving

$$\begin{pmatrix} -\sigma^{(p)} I & A & -u^{(p)} & 0 \\ A^T & -\sigma^{(p)} I & 0 & -v^{(p)} \\ 2u^{(p)T} & 0 & 0 & 0 \\ 0 & 2v^{(p)T} & 0 & 0 \end{pmatrix} \begin{pmatrix} z^{(p+1)} \\ y^{(p+1)} \\ \mu_1^{(p+1)} \\ \mu_2^{(p+1)} \end{pmatrix} = \begin{pmatrix} \sigma^{(p)} u^{(p)} - Av^{(p)} + \mu_1^{(p)} z^{(p)} \\ \sigma^{(p)} v^{(p)} - A^T u^{(p)} + \mu_2^{(p)} y^{(p)} \\ 1 - u^{(p)T} u^{(p)} - z^{(p)T} z^{(p)} \\ 1 - v^{(p)T} v^{(p)} - y^{(p)T} y^{(p)} \end{pmatrix}, \tag{1.9}$$

to obtain corrections to $u^{(p)}$, $v^{(p)}$, and $\mu^{(p)}$ by the updates

$$u^{(p+1)} = u^{(p)} + z^{(p+1)},$$
$$v^{(p+1)} = v^{(p)} + y^{(p+1)},$$
$$\sigma^{(p+1)} = \sigma^{(p)} + (\mu_1^{(p+1)} + \mu_2^{(p+1)})/2.$$

If A is $m \times n$, then this is an $(m+n+2) \times (m+n+2)$ system to be solved. If this system is solved, we can compute corrections μ, y, and z to the singular value and the singular vectors, thereby obtaining a more accurate value for the singular value and singular vectors.

If we handle this as we do in the eigenvalue case [1], we will improve the accuracy of σ, u, and v. The accuracy obtained by the algorithm will be full working precision, with only the residual calculations (the right-hand side of (1.9)) done in extended precision.

2. Relationship to Newton's method. The algorithm as described above can be derived by the use of Newton's method applied to (1.2) and (1.3). We define functions f_i and f as follows:

$$\begin{aligned} f_1(u, v, \sigma_1, \sigma_2) &= Av - \sigma_1 u, & f_3(u, v, \sigma_1, \sigma_2) &= u^T u - 1, \\ f_2(u, v, \sigma_1, \sigma_2) &= A^T u - \sigma_2 v, & f_4(u, v, \sigma_1, \sigma_2) &= v^T v - 1, \end{aligned} \tag{2.1}$$

and

$$f(x) = (f_1(x), f_2(x), f_3(x), f_4(x)),$$

where

$$x = \begin{pmatrix} u \\ v \\ \sigma_1 \\ \sigma_2 \end{pmatrix}.$$

The approach is to find the zeros of $f(x)$. Newton's method applied to this problem is

$$f'(x_i)(x_{i+1} - x_i) = -f(x_i), \tag{2.2}$$

where

$$x_i = \begin{pmatrix} u^{(i)} \\ v^{(i)} \\ \sigma_1^{(i)} \\ \sigma_2^{(i)} \end{pmatrix}.$$

The derivative of $f(x)$ is

$$f'(x) = \begin{pmatrix} -\sigma_1 I & A & -u & 0 \\ A^T & -\sigma_2 I & 0 & -v \\ 2u^T & 0 & 0 & 0 \\ 0 & 2v^T & 0 & 0 \end{pmatrix}. \tag{2.3}$$

The above method expressed in matrix notation is then just a restatement of (1.8), ignoring the second order terms in the right-hand side.

Notice that since the method is equivalent to Newton's method, we could compute the left and right singular vectors, given a close approximation to the singular value.

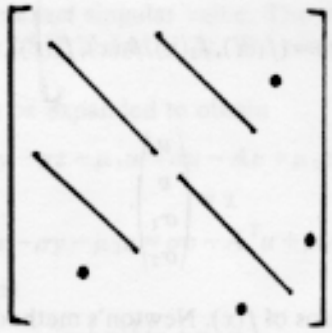
3. Effects of various factorizations. If we have computed the singular value decomposition and retained the matrices produced during the factorization, each singular value and the corresponding singular vectors can be improved in $O(mn)$ operations. We will assume that the matrices U, Σ, and V are available such that $A \approx U\Sigma V^T$. Then the coefficient matrix in (1.8) can be decomposed into the form

$$\begin{pmatrix} U & 0 & 0 & 0 \\ 0 & V & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -\sigma_s I & \Sigma & -e_s & 0 \\ \Sigma & -\sigma_s I & 0 & -e_s \\ e_s^T & 0 & 0 & 0 \\ 0 & e_s^T & 0 & 0 \end{pmatrix} \begin{pmatrix} U^T & 0 & 0 & 0 \\ 0 & V^T & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{3.1}$$

where e_s is the sth column of the identity matrix and σ_s is the approximation being improved.

This factored form can be used to simplify (1.8). Since U and V are orthogonal, systems of equations involving the left and the right matrices of (3.1) can be easily solved by simply multiplying by the transposes. Systems of equations involving the matrix in the center can be handled by solving 2×2 or 4×4 subsystems of equations as can be seen from the nonzero structure of the matrix:

(3.2)



If we have a bidiagonal factorization of A, say $A = UBV^T$, where B is bidiagonal, then we can improve the accuracy in $O(mn)$ operations. Let us assume we have the matrices V and B from the bidiagonalization procedure. We will concentrate only on the matrix

$$\begin{pmatrix} -\sigma I & A \\ A^T & -\sigma I \end{pmatrix}. \tag{3.3}$$

This matrix is the interesting part of the one in (1.8) and can be thought of as a rank 2 modification of that equation. The matrix can then be written as

$$\begin{pmatrix} I & 0 \\ -\frac{1}{\sigma}A^T & V^T \end{pmatrix} \begin{pmatrix} -\sigma I & 0 \\ 0 & -\sigma I + \frac{1}{\sigma}B^T B \end{pmatrix} \begin{pmatrix} I & -\frac{1}{\sigma}A \\ 0 & V \end{pmatrix}. \tag{3.4}$$

Note that solving systems based on this factored form is a simple task since $V^{-1} = V^T$. The only actual need for an equation solver comes from

$$-\sigma I + \frac{1}{\sigma}B^T B, \tag{3.5}$$

and this matrix is tridiagonal. Thus, given the bidiagonal matrix and the V matrix of the transformation, we can improve the accuracy of the singular values.

If we have instead the QR factorization of A, namely $A = QR$, where R is upper triangular and Q is orthogonal, then we can improve the accuracy of the singular value in $O(mn + n^3)$ operations, provided we have some approximation to it. We will concentrate on the matrix in (3.3). This matrix can be rewritten in factored form as

$$\begin{pmatrix} I & 0 \\ -\frac{1}{\sigma}A^T & Q^T \end{pmatrix} \begin{pmatrix} -\sigma I & 0 \\ 0 & -\sigma I + \frac{1}{\sigma}R^T R \end{pmatrix} \begin{pmatrix} I & -\frac{1}{\sigma}A \\ 0 & Q \end{pmatrix}. \tag{3.6}$$

As in (3.4) it becomes a matter of solving equations with a matrix of the form

$$-\sigma I + \frac{1}{\sigma}R^T R. \tag{3.7}$$

Unlike (3.5), this matrix is full and, unfortunately, the factor R cannot be used to simplify the process since the matrix $R^T R$ is being modified by a rank n matrix, σI. Equation (3.7) requires a further factorization to solve systems based upon it.

4. Convergence of the update process. The convergence results for this method are the same as for the eigenvalue case. We state the results here but omit the proof which can be found in [1].

In the presence of round-off error, if the initial error in the singular value is small enough in some sense and the singular value is an isolated one, the iterative process will converge.

If working precision is used in computing the approximate singular values and extended precision is used in the residual calculation, then when the method converges, it produces results that are accurate to at least full working precision.

The method is equivalent to Newton's method; therefore, the convergence is quadratic.

The method just described has a deficiency: When there are multiple singular values, the matrix in (1.8) becomes ill-conditioned. The degree of ill-conditioning is related to the separation between the singular value being improved and its closest neighbor. Existence of close or multiple singular values can be monitored by examining the condition number of the matrix in (1.8). If the matrix of (1.8) has a large condition number, then the iteration will converge with a less than quadratic rate. For identical singular values, the matrix involved is exactly singular.

This deficiency can be illustrated by an example. For a 2×2 system the matrix has the form

$$\begin{pmatrix} -\sigma & \sigma_i \\ \sigma_i & -\sigma \end{pmatrix},$$

where σ is an approximation to σ_s. If any σ_i is close to σ_s, then this system will be ill-conditioned, and the conditioning depends upon $1/(\sigma-\sigma_i)$. In this situation one cannot improve just one singular value but must work with a cluster of them, as well as the invariant subspace of singular vectors.

5. Results. The following numerical tests were run on a VAX 11/780. The initial reduction was performed in single precision; double precision was used only to compute the residuals and to add the correction to the previous result. In single precision, the working accuracy is 2^{-28}; in double precision, the accuracy is 2^{-56}.

The matrices used here come from the original paper by Golub and Reinsch [2]. The first matrix has the form

$$A=\begin{pmatrix} 22 & 10 & 2 & 3 & 7 \\ 14 & 7 & 10 & 0 & 8 \\ -1 & 13 & -1 & -11 & 3 \\ -3 & -2 & 13 & -2 & 4 \\ 9 & 8 & 1 & -2 & 4 \\ 9 & 1 & -7 & 5 & -1 \\ 2 & -6 & 6 & 5 & 1 \\ 4 & 5 & 0 & -2 & 2 \end{pmatrix},$$

with singular values

$$\sigma_1=\sqrt{1248},\quad \sigma_2=20,\quad \sigma_3=\sqrt{384},\quad \sigma_4=\sigma_5=0.$$

The results from the improvement algorithm on this problem are given in Table 1. All results were achieved using single precision computations except to accumulate the residuals. The method used was based on the factored form of (3.1).

TABLE 1

Iteration	σ	$u^T u$	$v^T v$
0	35.3270149	0.999999718	0.999999683
1	35.327043465315658	1.000000000000101	1.000000000000304
2	35.327043465311387	1.000000000000000	1.000000000000000
true	35.327043465311387419		
0	19.9999790	0.999999520	0.999999326
1	20.000000000006048	1.000000000003621	1.000000000003431
2	20.000000000000000	1.000000000000000	1.000000000000000
true	20.0		
0	19.5958881	0.999999043	0.999999379
1	19.595917942277176	1.000000000003258	1.000000000003183
2	19.595917942265425	1.000000000000000	1.000000000000000
true	19.595917942265424785		
0	0.00000718535284	0.999998454	0.999999228
1	−0.000000000004162	1.000000000000745	1.000000000000306
2	0.000000000000000	1.000000000533098	1.000000000281307
true	0.0		
0	0.00000120505399	0.999998900	0.999999509
1	−0.000000000000479	1.000000000000304	1.000000000000061
2	0.000000000000000	1.000000018476308	1.000000001164373
true	0.0		

The results in Table 1 show the iteration converging very rapidly. The singular values are initially correct to working precision, and two iterations have gained full extended precision.

For the next example we use a standard symmetric eigenvalue problem. The matrix, W_{2k+1}^{+} [4], is symmetric tridiagonal, and has some pathologically close eigenvalues and singular values. It is defined by the relations

$$\alpha_i = k+1-i, \qquad i = 1, \cdots, k+1,$$
$$\alpha_i = i-k-1, \qquad i = k+2, \cdots, 2k+1,$$
$$\beta_i = 1, \qquad i = 2, \cdots, 2k+1,$$

where $k = 5$, α_i is the ith diagonal element, and β_i is the ith subdiagonal element. See Table 2.

TABLE 2

Iteration	σ	$u^T u$	$v^T v$
0	5.7462210	0.999998079	0.999997771
1	5.746231847961203	1.001536038280316	1.001536023270078
2	5.746231833605774	1.000033093486984	1.000033093488725
3	5.746231833805267	1.000000000729813	1.000000000729813
4	5.746231833809865	1.000000000000009	1.000000000000002
5	5.746231833809865	1.000000000000000	1.000000000000000
0	5.7461471	0.999998012	0.999997719
1	5.746157555822260	1.000863731344646	1.000863745222875
2	5.746157545424549	1.000016916083231	1.000016916084818
3	5.746157545577390	1.000000000495525	1.000000000495525
4	5.746157545580572	1.000000000000011	1.000000000000011
5	5.746157545580572	1.000000000000000	1.000000000000000

The singular values displayed in Table 2 are the largest ones of W_{11}^{+} and happen to be the closest. The matrix has a condition number of 10^5, as a result, each iteration makes an improvement of approximately three digits. Note the contrast to the previous case where the matrix was well conditioned and each iteration gained a full seven digits.

6. Multiple singular values. We are interested in improving more than one singular value at a time. We are motivated to do so since the approach for improving one singular value breaks down when there are multiple singular values with close numerical values. For simplicity we will restrict the discussion initially to two singular values σ_1 and σ_2 and the corresponding vectors u_1, v_1 and u_2, v_2.

We know that the two-space in which u_1 and u_2 and v_1 and v_2 lie is numerical well determined. Hence we have

$$\begin{aligned} A(v_1+y_1) &= (\sigma_1+\mu_{11})(u_1+z_1)+\mu_{21}(u_2+z_2), \\ A(v_2+y_2) &= \mu_{12}(u_1+z_1)+(\sigma_2+\mu_{22})(u_2+z_2) \end{aligned} \tag{6.1}$$

and

$$\begin{aligned} A^T(u_1+z_1) &= (\sigma_1+\mu_{11})(v_1+y_1)+\mu_{21}(v_2+y_2), \\ A^T(u_2+z_2) &= \mu_{12}(v_1+y_1)+(\sigma_2+\mu_{22})(v_2+y_2), \end{aligned}$$

where the corrections y_i, z_i, and μ_{ij} are expected to be small.

From above we form

$$\begin{aligned} A(v_1+y_1|v_2+y_2) &= (u_1+z_1|u_2+z_2)\begin{pmatrix}\sigma_1+\mu_{11} & \mu_{12}\\ \mu_{21} & \sigma_2+\mu_{22}\end{pmatrix}, \\ A^T(u_1+z_1|u_2+z_2) &= (v_1+y_1|v_2+y_2)\begin{pmatrix}\sigma_1+\mu_{11} & \mu_{12}\\ \mu_{21} & \sigma_2+\mu_{22}\end{pmatrix}. \end{aligned} \tag{6.2}$$

We will require in addition that the orthogonality conditions

$$(v_i+y_i)^T(v_j+y_j)=\begin{cases}0 & \text{if } i=j,\\ 1 & \text{if } i\neq j\end{cases} \tag{6.3}$$

and

$$(u_i+z_i)^T(u_j+z_j)=\begin{cases}0 & \text{if } i=j,\\ 1 & \text{if } i\neq j\end{cases}$$

be satisfied. For simplicity we will assume that the 2×2 matrix of (6.2) which contains the correction to the singular values is symmetric, therefore, $\mu_{12}=\mu_{21}$. In order to produce the improved singular values this 2×2 matrix must be diagonalized.

Equation (6.2) together with (6.3) give rise to $2n+2m+6$ equations in $2n+2m+6$ unknowns. This matrix equation has the form

$$\begin{pmatrix}
-\sigma_1 I & A & & & -u_2 & & -u_1 & & & \\
A^T & -\sigma_1 & & & & -v_2 & & -v_1 & & \\
 & & -\sigma_2 I & A & -u_1 & & & & -u_2 & \\
 & & A^T & -\sigma_2 I & & -v_1 & & & & -v_2 \\
u_2^T & & u_1^T & & & & & & & \\
 & v_2^T & & v_1^T & & & & & & \\
2u_1^T & & & & & & & & & \\
 & 2v_1^T & & & & & & & & \\
 & & 2u_2^T & & & & & & & \\
 & & & 2v_2^T & & & & & &
\end{pmatrix}
\begin{pmatrix} z_1\\ y_1\\ z_2\\ y_2\\ \mu_{12}^1\\ \mu_{12}^2\\ \mu_{11}^1\\ \mu_{11}^2\\ \mu_{22}^1\\ \mu_{22}^2 \end{pmatrix}
=
\begin{pmatrix}
\sigma_1u_1-Av_1+\mu_{11}^1z_1+\mu_{12}^1z_2\\
\sigma_2u_2-Av_2+\mu_{12}^1z_1+\mu_{22}^1z_2\\
\sigma_1v_1-A^Tu_1+\mu_{11}^2y_1+\mu_{12}^2y_2\\
\sigma_2v_2-A^Tu_2+\mu_{12}^2y_1+\mu_{22}^2y_2\\
-\mu_1^Tu_2-z_1^Tz_2\\
-v_1^Tv_2-y_1^Ty_2\\
1-u_2^Tu_2-z_2^Tz_2\\
1-v_1^Tv_1-y_1^Ty_1\\
1-u_1^Tu_1-z_1^Tz_1\\
1-v_2^Tv_2-y_2^Ty_2
\end{pmatrix}.$$

As in the case of a single singular value, if one has access to the matrix factorization then the matrix problem can easily be solved.

In general, if we extend the procedure to handle k close singular values we have

$$A[v_1+y_1, \cdots, v_k+y_k] = [u_1+z_1, \cdots, u_k+z_k][\operatorname{diag}(\sigma_i)+M]$$

and

$$A^T[u_1+z_1, \cdots, u_k+z_k] = [v_1+y_1, \cdots, v_k+y_k][\operatorname{diag}(\sigma_i)+M],$$

where $m_{ij} = \mu_{ij}$ and $M = M^T$ and it is expected that y_i, z_i, and μ_{ij} will be small. These equations together with (6.3) lead to a system of equations of order $k(m+n)+k(k+1)$ and an eigenvalue problem of order k.

REFERENCES

[1] J. J. DONGARRA, C. B. MOLER AND J. H. WILKINSON, *Improving the accuracy of computed eigenvalues and eigenvectors*, SIAM J. Numer. Anal., 20 (1983), pp. 23–45.

[2] G. H. GOLUB AND C. REINSCH, *Singular value decomposition and least squares solutions*, Numer. Math., 14 (1970), pp. 403–420.

[3] G. W. STEWART, *Error and perturbation bounds for subspaces associated with certain eigenvalue problems*, SIAM Rev., 15 (1973), pp. 727–764.

[4] J. H. WILKINSON, *The Algebraic Eigenvalue Problem*, Oxford Univ. Press, London, 1965.