SIAM J. MATRIX ANAL. APPL.
Vol. 34, No. 2, pp. 417–444

DIVIDE AND CONQUER THE CS DECOMPOSITION*

BRIAN D. SUTTON†

Abstract. We develop a divide-and-conquer algorithm for the bidiagonal CS decomposition (CSD). This complements an earlier algorithm based on simultaneous QR iteration. The new algorithm is designed to provide the efficiency gains of familiar divide-and-conquer algorithms on both serial and parallel architectures. The solution uses many components of existing algorithms, particularly the bidiagonal SVD algorithm of Gu and Eisenstat, but extra steps and reparameterizations are required to maintain orthogonality and consistent singular vectors, especially when the singular vectors are ill conditioned. The algorithm supports the stable computation of the generalized singular value decomposition (GSVD) in addition to the CSD.

Key words. CS decomposition, generalized singular value decomposition, divide-and-conquer, bidiagonal

AMS subject classifications. primary 65F15, 15A18, 15A23, 15B10; secondary 65F25

DOI. 10.1137/120870785

1. Introduction.

To compute the *CS decomposition* (CSD) of a partitioned orthogonal matrix is to simultaneously diagonalize its blocks. An effective technique is to proceed in two steps: *simultaneous bidiagonalization* followed by a *bidiagonal CSD*. The bidiagonal CSD is defined as follows:

$$\underset{\substack{\text{orthonormal columns,}\\ \text{bidiagonal blocks}}}{\begin{bmatrix} B_{11} \\ B_{21} \end{bmatrix}} = \underset{\text{orthogonal}}{\begin{bmatrix} U_1 & \\ & U_2 \end{bmatrix}} \underset{\substack{\text{orthonormal columns,}\\ C,\, S \text{ diagonal}\\ \text{and nonnegative}}}{\begin{bmatrix} C \\ S \end{bmatrix}} \underset{\text{orthogonal}}{V_1^T} . \tag{1}$$

Simultaneous bidiagonalization has been proved backward stable [33], and the bidiagonal CSD has been computed previously by simultaneous QR iteration [32]. The present paper develops a new divide-and-conquer algorithm, motivated by the efficiency gains of other divide-and-conquer algorithms on both serial and parallel architectures.

The CSD is difficult to compute because it requires consistency in the face of ill conditioning. The columns of V_1 are simultaneously the right singular vectors of the top and bottom blocks, and they are sensitive to perturbations when associated with clustered singular values. A CSD algorithm must make a consistent choice of singular vectors even when that choice is arbitrary. Plus, the CSD is only defined when columns are orthonormal, a near impossibility in finite precision.

Our new divide-and-conquer algorithm navigates around these dangers. It is proved to be backward stable and has a number of advantageous numerical properties. In addition to absorbing roundoff errors, a perturbation of the input enforces orthogonality so that the CSD exists. The algorithm's handling of the top and bottom halves of the input matrix is completely symmetric. Rather than computing the diagonal entries of C or S directly, the algorithm computes the underlying angles $\theta_1, \ldots, \theta_n$, which are especially well conditioned. The columns of V_1 are computed

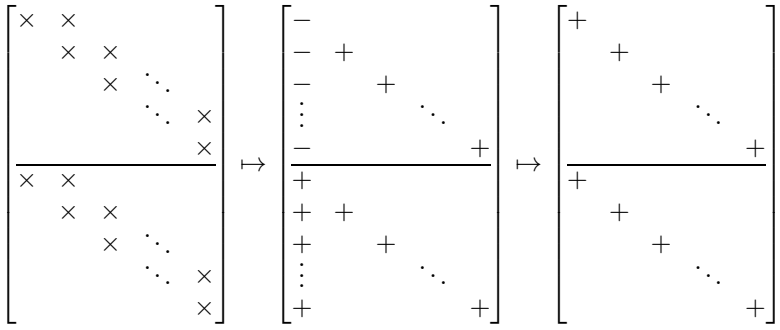
*Received by the editors March 21, 2012; accepted for publication (in revised form) by F. M. Dopico February 1, 2013; published electronically April 25, 2013. This material is based upon work supported by the National Science Foundation under grant DMS-0914559.

http://www.siam.org/journals/simax/34-2/87078.html

†Department of Mathematics, Randolph-Macon College, Ashland, VA 23005 (bsutton@rmc.edu).

directly without any postprocessing cleanup procedure, even when they are ill conditioned.

One of the key tools is an apparently new parameterization for matrices with orthonormal columns. It applies to matrices with "broken-arrow" blocks, illustrated by the intermediate structure in the following reduction:



Such a matrix must have the following form for its columns to be orthonormal:

$$\left[\begin{array}{ccccc} -r_n & & & & \\ -r_1\sin\phi_1 & \cos\phi_1 & & & \\ -r_2\sin\phi_2 & & \cos\phi_2 & & \\ \vdots & & & \ddots & \\ -r_{n-1}\sin\phi_{n-1} & & & & \cos\phi_{n-1} \\ \hline r_0 & & & & \\ r_1\cos\phi_1 & \sin\phi_1 & & & \\ r_2\cos\phi_2 & & \sin\phi_2 & & \\ \vdots & & & \ddots & \\ r_{n-1}\cos\phi_{n-1} & & & & \sin\phi_{n-1} \end{array}\right]. \tag{2}$$

Even if $r_0, \ldots, r_n$ and $\phi_1, \ldots, \phi_{n-1}$ are stored in finite precision, the columns are exactly orthogonal. Of course, if an entry $\cos\phi_i$, $\sin\phi_i$, $-r_i\sin\phi_i$, or $r_i\cos\phi_i$ were computed in finite precision, then the ensuing numerical roundoff error would likely destroy orthogonality, so we should refer to the underlying parameters $r_0, \ldots, r_n$, $\phi_1, \ldots, \phi_{n-1}$ whenever possible.

The bidiagonal CSD as defined above is an instance of the *two-by-one CSD* because the input matrix is tall and skinny and partitioned into a two-by-one block structure. Related problems include the *two-by-two CSD*, which applies to a square orthogonal matrix, and the *generalized* or *quotient SVD* (GSVD/QSVD), which places no orthogonality constraint on the input matrix. These decompositions are considered at the end of section 2 and in section 4.11.

The CSD, first formulated by Stewart [29], pulls together earlier ideas by Jordan [20], Davis and Kahan [8, 9], and Björck and Golub [5]. Van Loan [37] and Paige and Saunders [25] were instrumental in developing the GSVD and recognizing its connection to the CSD. The decompositions have many applications and connections to other areas of mathematics and engineering. They determine geodesics on the Grassman manifold [5, 8, 9, 13, 20, 29], reveal canonical correlations [14, 18], uncover the fast Fourier transform [36], compile quantum computer programs [23, 27, 34], and model random eigenvalues [14, 21, 31]. Additional overview is provided by a

pair of surveys [3, 26]. Other algorithms for the CSD include [4, 11, 12, 24, 30, 35], and relevant references on divide-and-conquer algorithms include [2, 7, 10, 19]. Simultaneous bidiagonalization is implemented in LAPACK [1] and has connections to orthogonal polynomial recurrences through CMV matrices [6, 28, 39].

2. The divide-and-conquer algorithm. The algorithm for computing the bidiagonal CSD (1) is specified next. The steps are broken into three major phases: dissect, descend, and merge. MATLAB notation is used for indexing and matrix construction. Although steps for ameliorating nonorthogonality and the effects of numerical roundoff error are included, the notation does not distinguish between exact and approximate quantities, and error terms are omitted. See section 4 for a numerical stability proof.

ALGORITHM 1 (Divide-and-Conquer CSD).

Input: B_{11} and B_{21}.

Output: U_1, U_2, V_1, and $\hat{\theta}_1, \ldots, \hat{\theta}_n$ for which $C = \operatorname{diag}(\cos(\hat{\theta}_i))$, $S = \operatorname{diag}(\sin(\hat{\theta}_i))$.

Procedure:

Dissect

1. Permute column $\lfloor n/2 \rfloor + 1$ to the front:
   $m := \lfloor n/2 \rfloor$
   $V_1 := I_n$; $V_1 = V_1(:, [m+1, 1:m, m+2:n])$
   $\begin{bmatrix} E \\ F \end{bmatrix} := \begin{bmatrix} B_{11} \\ B_{21} \end{bmatrix} V_1$
   Then

$$\begin{bmatrix} B_{11} \\ B_{21} \end{bmatrix} V_1 = \begin{bmatrix} E \\ F \end{bmatrix} \in \left[\begin{array}{cccccc} \circ & \bullet & \bullet & \circ & \circ & \circ \\ \circ & \circ & \bullet & \bullet & \circ & \circ \\ \bullet & \circ & \circ & \bullet & \circ & \circ \\ \bullet & \circ & \circ & \circ & \bullet & \circ \\ \circ & \circ & \circ & \circ & \bullet & \bullet \\ \circ & \circ & \circ & \circ & \circ & \bullet \\ \hline \circ & \bullet & \bullet & \circ & \circ & \circ \\ \circ & \circ & \bullet & \bullet & \circ & \circ \\ \bullet & \circ & \circ & \bullet & \circ & \circ \\ \bullet & \circ & \circ & \circ & \bullet & \circ \\ \circ & \circ & \circ & \circ & \bullet & \bullet \\ \circ & \circ & \circ & \circ & \circ & \bullet \end{array}\right].$$

2. Eliminate the subdiagonal entries of the lower-bidiagonal blocks:
   $\begin{bmatrix} P \\ Q \end{bmatrix} := \begin{bmatrix} E \\ F \end{bmatrix}$
   $U_1 := I_n$; $U_2 := I_n$
   for $j = 1 : n - m - 1$
   $G := \text{givens}(n, m+j, m+j+1, P([m+j, m+j+1], m+j+1))$
   $U_1 := U_1 G$; $P := G^T P$
   $G := \text{givens}(n, m+j, m+j+1, Q([m+j, m+j+1], m+j+1))$
   $U_2 := U_2 G$; $Q := G^T Q$
   end for

Then

$$\begin{bmatrix} U_1 & \\ & U_2 \end{bmatrix}^T \begin{bmatrix} B_{11} \\ B_{21} \end{bmatrix} V_1 = \begin{bmatrix} P \\ Q \end{bmatrix} \in \left[\begin{array}{cccccc} \circ & \bullet & \bullet & \circ & \circ & \circ \\ \circ & \circ & \bullet & \bullet & \circ & \circ \\ \bullet & \circ & \circ & \bullet & \circ & \circ \\ \bullet & \circ & \circ & \circ & \bullet & \bullet \\ \bullet & \circ & \circ & \circ & \circ & \bullet \\ \bullet & \circ & \circ & \circ & \circ & \circ \\ \hline \circ & \bullet & \bullet & \circ & \circ & \circ \\ \circ & \circ & \bullet & \bullet & \circ & \circ \\ \bullet & \circ & \circ & \bullet & \circ & \circ \\ \bullet & \circ & \circ & \circ & \bullet & \bullet \\ \bullet & \circ & \circ & \circ & \circ & \bullet \\ \bullet & \circ & \circ & \circ & \circ & \circ \end{array}\right]. \tag{3}$$

The function call givens(n, k, l, x) constructs an n-by-n Givens rotation whose projection onto the kth and lth standard basis vectors maps x to $(\|x\|, 0)$. Upon completion, the submatrices in columns $2, \ldots, m+1$ and $m+2, \ldots, n$ are orthonormal and have upper-bidiagonal blocks.

Descend

3. Recursively compute bidiagonal CSDs and update the larger matrix:
   $[U_1^{(a)}, U_2^{(a)}, \psi^{(a)}, V_1^{(a)}] := \text{csd}\left([P(1:m, 2:m+1); Q(1:m, 2:m+1)]\right)$
   $[U_1^{(b)}, U_2^{(b)}, \psi^{(b)}, V_1^{(b)}] := \text{csd}\left([P(m+1\!:\!n-1, m+2\!:\!n); Q(m+1\!:\!n-1, m+2\!:\!n)]\right)$
   $x := \text{blkdiag}(U_1^{(a)}, U_1^{(b)}, 1)^T\, P(:, 1)$
   $y := \text{blkdiag}(U_2^{(a)}, U_2^{(b)}, 1)^T\, Q(:, 1)$
   $U_1 := U_1\ \text{blkdiag}(U_1^{(a)}, U_1^{(b)}, 1)$
   $U_2 := U_2\ \text{blkdiag}(U_2^{(a)}, U_2^{(b)}, 1)$
   $V_1 := V_1\ \text{blkdiag}(1, V_1^{(a)}, V_1^{(b)})$
   Then

$$\begin{bmatrix} U_1 & \\ & U_2 \end{bmatrix}^T \begin{bmatrix} B_{11} \\ B_{21} \end{bmatrix} V_1 = \left[\begin{array}{ccc} x_{1:m} & \cos\Psi^{(a)} & 0 \\ x_{m+1:n-1} & 0 & \cos\Psi^{(b)} \\ x_n & 0 & 0 \\ \hline y_{1:m} & \sin\Psi^{(a)} & 0 \\ y_{m+1:n-1} & 0 & \sin\Psi^{(b)} \\ y_n & 0 & 0 \end{array}\right] \in \left[\begin{array}{cccccc} \bullet & \bullet & \circ & \circ & \circ & \circ \\ \bullet & \circ & \bullet & \circ & \circ & \circ \\ \bullet & \circ & \circ & \bullet & \circ & \circ \\ \bullet & \circ & \circ & \circ & \bullet & \circ \\ \bullet & \circ & \circ & \circ & \circ & \bullet \\ \bullet & \circ & \circ & \circ & \circ & \circ \\ \hline \bullet & \bullet & \circ & \circ & \circ & \circ \\ \bullet & \circ & \bullet & \circ & \circ & \circ \\ \bullet & \circ & \circ & \bullet & \circ & \circ \\ \bullet & \circ & \circ & \circ & \bullet & \circ \\ \bullet & \circ & \circ & \circ & \circ & \bullet \\ \bullet & \circ & \circ & \circ & \circ & \circ \end{array}\right], \tag{4}$$

   in which $\Psi^{(a)} = \text{diag}(\psi^{(a)})$ and $\Psi^{(b)} = \text{diag}(\psi^{(b)})$.
4. Permute rows and columns to achieve broken-arrow patterns and sorted diagonals. Then apply diagonal signature matrices to achieve the sign pattern of (2):
   $[\psi, \pi] := \text{sort}([\psi^{(a)}; \psi^{(b)}])$
   $z_n := x(n);\ z_{1:n-1} := x(\pi);\ w_0 := y(n);\ w_{1:n-1} := y(\pi)$
   $U_1 := U_1(:, [n; \pi]);\ U_2 := U_2(:, [n; \pi]);\ V_1 := V_1(:, [1; \pi + 1])$
   $D := \text{diag}(\text{sign}(-\sin(\psi) .* z_{1:n-1} + \cos(\psi) .* w_{1:n-1}))$
   $z_{1:n-1} := D\, z_{1:n-1};\ U_1 := U_1\ \text{blkdiag}(1, D)$
   $w_{1:n-1} := D\, w_{1:n-1};\ U_2 := U_2\ \text{blkdiag}(1, D)$
   $V_1 := V_1\ \text{blkdiag}(1, D)$

if $z_n > 0,\ z_n = -z_n;\ U_1(:,1) := -U_1(:,1);$ end if
if $w_0 < 0,\ w_0 = -w_0;\ U_2(:,1) := -U_2(:,1);$ end if
Then

$$\text{(5)}\qquad \begin{bmatrix} U_1 & \\ & U_2 \end{bmatrix}^T \begin{bmatrix} B_{11} \\ B_{21} \end{bmatrix} V_1 = \left[\begin{array}{cc} z_n & 0 \\ z_{1:n-1} & \cos\Psi \\ \hline w_0 & 0 \\ w_{1:n-1} & \sin\Psi \end{array}\right] \in \left[\begin{array}{ccccc} - & & & & \\ - & + & & & \\ - & & + & & \\ - & & & + & \\ - & & & & + \\ \hline + & & & & \\ + & + & & & \\ + & & + & & \\ + & & & + & \\ + & & & & + \end{array}\right],$$

in which $\Psi = \operatorname{diag}(\psi)$ and $0 \le \psi_1 \le \cdots \le \psi_{n-1} \le \pi/2$. For convenience below, also define $\psi_0 = 0$ and $\psi_n = \pi/2$.

Merge

5. Express $(w_i, -z_i)$, $i = 1, \ldots, n-1$, in polar coordinates:
   $r_0 := w_0$
   $r_{1:n-1} := -\sin(\psi) .* z_{1:n-1} + \cos(\psi) .* w_{1:n-1}$
   $r_n := -z_n$
   Then

$$\text{(6)}\qquad \begin{bmatrix} U_1 & \\ & U_2 \end{bmatrix}^T \begin{bmatrix} B_{11} \\ B_{21} \end{bmatrix} V_1 = \left[\begin{array}{ccc} -r_n & & \\ -r_1\sin\psi_1 & \cos\psi_1 & \\ \vdots & & \ddots \\ -r_{n-1}\sin\psi_{n-1} & & \cos\psi_{n-1} \\ \hline r_0 & & \\ r_1\cos\psi_1 & \sin\psi_1 & \\ \vdots & & \ddots \\ r_{n-1}\cos\psi_{n-1} & & \sin\psi_{n-1} \end{array}\right].$$

   This matrix has exactly orthogonal columns. Columns $2, \ldots, n$ are perfectly normalized, and the first column is almost normalized, in a sense to be made precise later. The singular values of the top and bottom blocks are the solutions of a secular equation specified below, which is expressed directly in terms of $r_i, \ldots, r_n$ rather than the entries $-r_i \sin\psi_i$, $r_i \cos\psi_i$.
6. Apply a deflation procedure if any r_i is tiny or if any ψ_i is exceptionally close to 0 or $\pi/2$ or another ψ_j. See section 4.5.
7. Compute angle gaps:
   for $i = 0 : n-1$
   $\Delta\phi_i := \psi_{i+1} - \psi_i$
   end for
   Then define, on paper, $\phi_i = \alpha \sum_{j=0}^{i-1} \Delta\phi_j$, in which $\alpha = \frac{\pi}{2} / \sum_{j=0}^{n-1} \Delta\phi_j$. We show in section 4.6 that, when $\Delta\phi_0, \ldots, \Delta\phi_{n-1}$ are computed in finite precision, ϕ_i is a small perturbation of ψ_i. The advantage of the angle gaps is that they allow various sums, differences, and trigonometric functions of angles to

be computed accurately. We have

(7)
$$\begin{bmatrix} U_1 & \\ & U_2 \end{bmatrix}^T \begin{bmatrix} B_{11} \\ B_{21} \end{bmatrix} V_1 = \left[\begin{array}{ccc} -r_n & & \\ -r_1 \sin\phi_1 & \cos\phi_1 & \\ \vdots & & \ddots \\ -r_{n-1}\sin\phi_{n-1} & & \cos\phi_{n-1} \\ \hline r_0 & & \\ r_1\cos\phi_1 & \sin\phi_1 & \\ \vdots & & \ddots \\ r_{n-1}\cos\phi_{n-1} & & \sin\phi_{n-1} \end{array}\right] =: \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix}.$$

8. Solve a secular equation:
   $[\sigma, \delta] :=$ solvesecular$(r_0, \ldots, r_n, \Delta\phi_0, \ldots, \Delta\phi_{n-1})$.
   This computes approximate solutions $\hat{\theta}_i = \phi_{\sigma(i)} + \hat{\delta}_i$, $i = 1, \ldots, n$, to the secular equation
   $$\sum_{k=0}^{n} \frac{r_k^2}{\sin(\phi_k + \theta)\sin(\phi_k - \theta)} = 0.$$
   Section 4.7 discusses accurate evaluation of the secular function. The cosines and sines of $\hat{\theta}_1, \ldots, \hat{\theta}_n$ are approximately the singular values of B_{11} and B_{21}, respectively.
9. Solve an inverse problem. Compute $t_0, \ldots, t_n$ as follows:
   for $i = 0 : n$
   $$t_i := \sqrt{\prod_{k=1}^{n} \sin(\phi_i + \hat{\theta}_k)\sin(\phi_i - \hat{\theta}_k) \Big/ \prod_{\substack{k=0 \\ k\neq i}}^{n} \sin(\phi_i + \phi_k)\sin(\phi_i - \phi_k)}$$
   end for
   For numerical stability, it is important to compute the intermediate quantities using the angle gaps $\Delta\phi_i$ and δ_i. The singular values of $\tilde{A}_{11}$ and $\tilde{A}_{21}$ in
   (8)
   $$\begin{bmatrix} \tilde{A}_{11} \\ \tilde{A}_{21} \end{bmatrix} := \left[\begin{array}{ccc} -t_n & & \\ -t_1 \sin\phi_1 & \cos\phi_1 & \\ \vdots & & \ddots \\ -t_{n-1}\sin\phi_{n-1} & & \cos\phi_{n-1} \\ \hline t_0 & & \\ t_1\cos\phi_1 & \sin\phi_1 & \\ \vdots & & \ddots \\ t_{n-1}\cos\phi_{n-1} & & \sin\phi_{n-1} \end{array}\right]$$
   are $\cos\hat{\theta}_i$ and $\sin\hat{\theta}_i$, respectively. That is, the singular values are determined by the computed solutions to the secular equation rather than the true solutions.
10. Compute singular vectors of (8) as follows:
    for $j = 1 : n$
    $\tilde{U}_1(1, j) := t_n/(\sin(\phi_n + \hat{\theta}_j)\sin(\phi_n - \hat{\theta}_j))$
    $\tilde{U}_1(2 : n, j) := (t_{1:n-1}\sin\phi_{1:n-1})/(\sin(\phi_{1:n-1} + \hat{\theta}_j)\sin(\phi_{1:n-1} - \hat{\theta}_j))$
    $\tilde{U}_2(1, j) := (t_0\cos\phi_0)/(\sin(\phi_0 + \hat{\theta}_j)\sin(\phi_0 - \hat{\theta}_j))$

$\tilde{U}_2(2:n,j) := (t_{1:n-1}\cos\phi_{1:n-1})/(\sin(\phi_{1:n-1}+\hat{\theta}_j)\sin(\phi_{1:n-1}-\hat{\theta}_j))$
$\tilde{V}_1(1,j) := -1$
$\tilde{V}_1(2:n,j) := t_{1:n-1}(\cos\phi_{1:n-1})(\sin\phi_{1:n-1}) \ \ldots$
$\qquad /(\sin(\phi_{1:n-1}+\hat{\theta}_j)\sin(\phi_{1:n-1}-\hat{\theta}_j))$
end for
Normalize columns of $\tilde{U}_1, \tilde{U}_2, \tilde{V}_1$.

Again, for numerical stability, it is important to compute the intermediate quantities using the angle gaps $\Delta\phi_i$ and δ_i. Now, the CSD of (8) has been computed:

$$\begin{bmatrix}\tilde{A}_{11}\\ \tilde{A}_{21}\end{bmatrix} = \begin{bmatrix}\tilde{U}_1 & \\ & \tilde{U}_2\end{bmatrix}\begin{bmatrix}C\\ S\end{bmatrix}\tilde{V}_1^T,$$

with $C = \operatorname{diag}(\cos\hat{\theta}_1,\ldots,\cos\hat{\theta}_n)$ and $S = \operatorname{diag}(\sin\hat{\theta}_1,\ldots,\sin\hat{\theta}_n)$.

11. Apply the singular vectors from the previous step to the original matrix:

$$U_1 := U_1\tilde{U}_1; \quad U_2 := U_2\tilde{U}_2; \quad V_1 := V_1\tilde{V}_1.$$

The bidiagonal CSD is finished:

$$\begin{bmatrix}U_1 & \\ & U_2\end{bmatrix}^T\begin{bmatrix}B_{11}\\ B_{21}\end{bmatrix}V_1 = \begin{bmatrix}\tilde{U}_1 & \\ & \tilde{U}_2\end{bmatrix}^T\begin{bmatrix}\tilde{A}_{11}\\ \tilde{A}_{21}\end{bmatrix}\tilde{V}_1 = \begin{bmatrix}C\\ S\end{bmatrix}.$$

If, instead of the two-by-one CSD, the desired decomposition is the two-by-two CSD

$$\begin{bmatrix}B_{11} & B_{12}\\ B_{21} & B_{22}\end{bmatrix} = \begin{bmatrix}U_1 & \\ & U_2\end{bmatrix}\begin{bmatrix}C & -S\\ S & C\end{bmatrix}\begin{bmatrix}V_1 & \\ & V_2\end{bmatrix}^T,$$

then the remaining right singular vectors can be found by

$$V_2 = -B_{12}^T U_1 S + B_{22}^T U_2 C.$$

This is proved stable in section 4.11.

The GSVD and the symmetric positive-definite generalized eigenvalue problem can also be reduced to the CSD [38, 35]. For the GSVD of matrices A and B, form the stacked matrix $\left[\begin{smallmatrix}A\\ B\end{smallmatrix}\right]$, orthonormalize the columns, simultaneously bidiagonalize the blocks, and then run the divide-and-conquer algorithm. In some cases, it may be advantageous to rescale and consider $\alpha^{-1}A$ and $\beta^{-1}B$, e.g., with $\alpha = \|A\|_2$ and $\beta = \|B\|_2$, in order to control the error.

The rest of this paper is organized as follows. Section 3 investigates the rank-one modification problem introduced by (7). Section 4 develops a numerical stability proof. Section 5 presents numerical experiments.

3. Rank-one modification problem. The matrix

$$\begin{bmatrix}A_{11}\\ A_{21}\end{bmatrix} = \left[\begin{array}{ccccc} & -r_n & & & \\ & -r_1\sin\phi_1 & \cos\phi_1 & & \\ & \vdots & & \ddots & \\ & -r_{n-1}\sin\phi_{n-1} & & & \cos\phi_{n-1}\\ \hline & r_0 & & & \\ & r_1\cos\phi_1 & \sin\phi_1 & & \\ & \vdots & & \ddots & \\ & r_{n-1}\cos\phi_{n-1} & & & \sin\phi_{n-1}\end{array}\right], \tag{9}$$

assuming $r_0^2 + \cdots + r_n^2 = 1$, has orthonormal columns and is a rank-one modification from having diagonal blocks. The CSD of this matrix is our present concern. We derive algebraic facts and perform a perturbation analysis. A study of numerical roundoff error is postponed to the next section.

3.1. Secular equation. The singular values $c_1, \ldots, c_n$ and $s_1, \ldots, s_n$ of A_{11} and A_{21}, respectively, are the solutions of the equations

$$1 + \sum_{k=1}^{n} \frac{r_k^2 \sin^2 \phi_k}{\cos^2 \phi_k - c^2} = 0 \quad \text{and} \quad 1 + \sum_{k=0}^{n-1} \frac{r_k^2 \cos^2 \phi_k}{\sin^2 \phi_k - s^2} = 0, \tag{10}$$

in which $\phi_0 = 0$ and $\phi_n = \pi/2$ [7, 15, 19]. We develop a single secular equation to replace this pair. Its solutions are $\theta_1, \ldots, \theta_n$, from which the singular values $c_i = \cos(\theta_i)$ and $s_i = \sin(\theta_i)$ are derived. This ensures that the singular values are perfectly consistent across the blocks and enables the stable and consistent computation of singular vectors.

THEOREM 2. *Given $0 = \phi_0 < \phi_1 < \cdots < \phi_{n-1} < \phi_n = \pi/2$ and positive $r_0, \ldots, r_n$, the secular equation*

$$f(\theta) = \sum_{k=0}^{n} \frac{r_k^2}{\sin(\phi_k + \theta) \sin(\phi_k - \theta)} = 0 \tag{11}$$

has exactly n solutions $\theta_1 < \cdots < \theta_n$ in $(0, \pi/2)$. The cosines and sines of these angles are the diagonal entries of C and S in the CSD

$$\begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix} = \begin{bmatrix} U_1 & \\ & U_2 \end{bmatrix} \begin{bmatrix} C \\ S \end{bmatrix} V_1^T.$$

Proof. The equation $f(\theta) = 0$ has n solutions interlacing $\phi_0, \ldots, \phi_n$ because there are poles at $\phi_0, \ldots, \phi_n$ and the function is monotonically increasing between each pair of adjacent poles.

By the existence of the CSD, the solutions $c_1, \ldots, c_n$ and $s_1, \ldots, s_n$ of (10) satisfy $c_i^2 + s_i^2 = 1$. Let $\theta_1, \ldots, \theta_n$ be the angles in $[0, \pi/2]$ for which $\cos(\theta_i) = c_i$ and $\sin(\theta_i) = s_i$. Applying the trigonometric identities $\cos^2 a - \cos^2 b = -\sin(a+b)\sin(a-b)$ and $\sin^2 a - \sin^2 b = \sin(a+b)\sin(a-b)$ to (10) and negating both sides of the left equation, we find

$$-1 + \sum_{k=1}^{n} \frac{r_k^2 \sin^2 \phi_k}{\sin(\phi_k + \theta) \sin(\phi_k - \theta)} = 0 \quad \text{and} \quad 1 + \sum_{k=0}^{n-1} \frac{r_k^2 \cos^2 \phi_k}{\sin(\phi_k + \theta) \sin(\phi_k - \theta)} = 0. \tag{12}$$

The sum of these equations is particularly attractive:

$$\frac{r_0^2}{\sin(\phi_0 + \theta) \sin(\phi_0 - \theta)} + \sum_{k=1}^{n-1} \frac{r_k^2 (\sin^2 \phi_k + \cos^2 \phi_k)}{\sin(\phi_k + \theta) \sin(\phi_k - \theta)} + \frac{r_n^2}{\sin(\phi_n + \theta) \sin(\phi_n - \theta)} = 0.$$

This is equivalent to (11). ∎

3.2. Inverse problem. Small perturbations to the singular values of A_{11} and A_{21} can cause large perturbations of the singular vectors. However, they cause only small perturbations to the entries, as shown in the following two theorems.

In the algorithm, $\theta_1, \ldots, \theta_n$ are approximated by $\hat{\theta}_1, \ldots, \hat{\theta}_n$, and then the entries of A_{11} and A_{21} are adjusted to compensate for the error. This requires the solution of

an inverse problem. The approach is due to Gu and Eisenstat [15], and the solution to the inverse problem is an extension of Löwner's work [22].

THEOREM 3. *Given interlaced angles* $0 = \phi_0 < \hat{\theta}_1 < \phi_1 < \hat{\theta}_2 < \cdots < \hat{\theta}_{n-1} < \phi_{n-1} < \hat{\theta}_n < \phi_n = \pi/2$, *define*

$$t_i = \sqrt{\prod_{k=1}^{n} \sin(\phi_i + \hat{\theta}_k)\sin(\phi_i - \hat{\theta}_k) \Big/ \prod_{\substack{k=0 \\ k\neq i}}^{n} \sin(\phi_i + \phi_k)\sin(\phi_i - \phi_k)} \tag{13}$$

for $i = 0, \ldots, n$. *Then the top and bottom blocks of* $\left[\begin{smallmatrix} \tilde{A}_{11} \\ \tilde{A}_{21} \end{smallmatrix}\right]$ *in* (8) *have singular values* $\cos\hat{\theta}_1, \ldots, \cos\hat{\theta}_n$ *and* $\sin\hat{\theta}_1, \ldots, \sin\hat{\theta}_n$, *respectively.*

Proof. According to [22], the singular values of

$$\begin{bmatrix} \tilde{z}_n & 0 \\ \tilde{z}_{1:n-1} & \cos\Phi \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \tilde{w}_0 & 0 \\ \tilde{w}_{1:n-1} & \sin\Phi \end{bmatrix}$$

equal $\cos\hat{\theta}_i$, $i = 1, \ldots, n$, and $\sin\hat{\theta}_i$, $i = 1, \ldots, n$, respectively, if

$$\tilde{z}_i^2 = \frac{\prod_{k=0}^{n-1}(\cos^2\hat{\theta}_{k+1} - \cos^2\phi_i)}{\prod_{\substack{k=1 \\ k\neq i}}^{n}(\cos^2\phi_k - \cos^2\phi_i)} = \frac{\prod_{k=0}^{n-1}\sin(\phi_i + \hat{\theta}_{k+1})\sin(\phi_i - \hat{\theta}_{k+1})}{\prod_{\substack{k=1 \\ k\neq i}}^{n}\sin(\phi_i + \phi_k)\sin(\phi_i - \phi_k)}, \quad i = 1, \ldots, n,$$

and

$$\tilde{w}_i^2 = \frac{\prod_{k=0}^{n-1}(\sin^2\hat{\theta}_{k+1} - \sin^2\phi_i)}{\prod_{\substack{k=0 \\ k\neq i}}^{n-1}(\sin^2\phi_k - \sin^2\phi_i)} = -\frac{\prod_{k=0}^{n-1}\sin(\phi_i + \hat{\theta}_{k+1})\sin(\phi_i - \hat{\theta}_{k+1})}{\prod_{\substack{k=0 \\ k\neq i}}^{n-1}\sin(\phi_i + \phi_k)\sin(\phi_i - \phi_k)}, \quad i = 0, \ldots, n-1.$$

We are free to choose $\tilde{z}_1, \ldots, \tilde{z}_n$ to be negative and $\tilde{w}_0, \ldots, \tilde{w}_{n-1}$ to be positive. Note that $-(\cos\phi_i)\tilde{z}_i = (\sin\phi_i)\tilde{w}_i$ for $i = 1, \ldots, n-1$. Let $t_i = -\tilde{z}_i/(\sin\phi_i) = \tilde{w}_i/(\cos\phi_i)$, $i = 1, \ldots, n-1$, and $t_0 = \tilde{w}_0$ and $t_n = -\tilde{z}_n$. Then $\tilde{z}_i = -t_i \sin\phi_i$, $\tilde{w}_i = t_i \cos\phi_i$, and the expression for t_i in the statement of the lemma is satisfied. ∎

The next theorem will show that the inverse problem is stable. Basically, if computed angles $\hat{\theta}_1, \ldots, \hat{\theta}_n$ produce small residuals in the secular equation (11), then $t_0, \ldots, t_n$ are small perturbations of $r_0, \ldots, r_n$. Some technical conditions are required in the proof, which are achieved by the deflation procedure in the code.

The perturbation result follows the path laid by Gu and Eisenstat [15], but it is not a direct corollary of their Theorem 3.3. Condition (6) of their paper requires that the entries in the first column of A_{11} and A_{21} be bounded away from zero by $\Omega(\mathbf{u})$. This we do not guarantee. Our deflation procedure ensures $r_i = \Omega(\mathbf{u})$, $i = 0, \ldots, n$, and also $\phi_i = \Omega(\mathbf{u})$ and $\pi/2 - \phi_i = \Omega(\mathbf{u})$, $i = 1, \ldots, n-1$, but this permits $r_i \sin(\theta_i) = O(\mathbf{u}^2)$ or $r_i \cos(\theta_i) = O(\mathbf{u}^2)$. See section 4.5 for details.

THEOREM 4. *Suppose* $\hat{\theta}_1, \ldots, \hat{\theta}_n$ *satisfy the following:*

1. *The interlacing inequalities* $0 = \phi_0 < \hat{\theta}_1 < \phi_1 < \hat{\theta}_2 < \cdots < \hat{\theta}_{n-1} < \phi_{n-1} < \hat{\theta}_n < \phi_n = \pi/2$.

2. *The residual bounds*

$$\left|f(\hat{\theta}_i)\right| \le \varepsilon \sum_{k=0}^{n} \frac{r_k^2}{|\sin(\phi_k + \hat{\theta}_i)\sin(\phi_k - \hat{\theta}_i)|}, \quad i = 1, \ldots, n,$$

for some $\varepsilon < 1/100$.

3. *The deflation-ensured criterion* $r_i \ge \tau$, $i = 0, \ldots, n$, *for some* $\tau \ge 2n\varepsilon$.

Then $|r_i - t_i| \le 4n\varepsilon$, $i = 0, \ldots, n$.

Proof. The argument follows the outline of section 3.3 of [16]. Some details are omitted here.

Let $\theta_1, \ldots, \theta_n$ be the true solutions to the secular equation. Because

$$\begin{aligned} f(\hat{\theta}_i) = f(\hat{\theta}_i) - f(\theta_i) &= \sum_{k=0}^{n} \frac{r_k^2}{\sin^2\phi_k - \sin^2\hat{\theta}_i} - \sum_{k=0}^{n} \frac{r_k^2}{\sin^2\phi_k - \sin^2\theta_i} \\ &= \sum_{k=0}^{n} \frac{-r_k^2\sin^2\theta_i + r_k^2\sin^2\hat{\theta}_i}{(\sin^2\phi_k - \sin^2\hat{\theta}_i)(\sin^2\phi_k - \sin^2\theta_i)} \\ &= -(\sin^2\theta_i - \sin^2\hat{\theta}_i)\sum_{k=0}^{n} \frac{r_k^2}{(\sin^2\phi_k - \sin^2\theta_i)(\sin^2\phi_k - \sin^2\hat{\theta}_i)}, \end{aligned}$$

we have

$$\begin{aligned} \left|\sin^2\theta_i - \sin^2\hat{\theta}_i\right| &\textstyle\sum_{k=0}^{n} \dfrac{r_k^2}{(\sin^2\phi_k - \sin^2\theta_i)(\sin^2\phi_k - \sin^2\hat{\theta}_i)} \\ &= \left|f(\hat{\theta}_i)\right| \le \varepsilon \sum_{k=0}^{n} \frac{r_k^2}{\left|\sin^2\phi_k - \sin^2\hat{\theta}_i\right|} \\ &\le \varepsilon \left(\sum_{k=0}^{n} \frac{r_k^2}{\left|\sin^2\phi_k - \sin^2\theta_i\right|} + \sum_{k=0}^{n} \frac{r_k^2}{\left|\sin^2\phi_k - \sin^2\hat{\theta}_i\right|} \right). \end{aligned}$$

The above bound enables us to show that $\left|\sin^2\theta_i - \sin^2\hat{\theta}_i\right|$ is small relative to $\left|\sin^2\phi_j - \sin^2\theta_i\right|$ for $j = 0, \ldots, n$, specifically

$$\left|\sin^2\theta_i - \sin^2\hat{\theta}_i\right| \le \frac{\beta_j}{1 - \beta_j/2}\left|\sin^2\phi_j - \sin^2\theta_i\right|,$$

in which $\beta_j = 2\varepsilon/((1-\varepsilon)r_j)$. Then it is possible to show

$$t_i = r_i\sqrt{\prod_{k=1}^{n}\left(1 + \frac{\lambda_{ki}}{r_i}\right)}, \qquad \lambda_{ki} = r_i\frac{\sin^2\theta_k - \sin^2\hat{\theta}_k}{\sin^2\phi_i - \sin^2\theta_k}.$$

The ratios λ_{ki} are uniformly bounded in absolute value by $\lambda \equiv \frac{2}{1-\varepsilon_0-(1/\nu)}\varepsilon$, in which $\varepsilon_0 = 1/100$ and $\nu = \tau/\varepsilon \ge 2n$. This leads to the absolute error bound

$$|t_i - r_i| = r_i\left|\sqrt{\prod_{k=1}^{n}\left(1 + \frac{\lambda_{ki}}{r_i}\right)} - 1\right| \le r_i\left|\left(1 + \frac{\lambda}{r_i}\right)^{n/2} - 1\right| \le r_i(e^{\frac{\lambda n}{2r_i}} - 1),$$

and the assumptions on ε, n, and r_i guarantee that $\lambda n/(2r_i) < 1$, and so

$$|t_i - r_i| \le r_i(e-1)\frac{\lambda n}{2r_i} = (e-1)\frac{\lambda n}{2} \le 4n\varepsilon. \qquad \square$$

As a corollary, the matrix entries undergo small perturbations as well: $|(-r_i \sin\phi_i) - (-t_i \sin\phi_i)| \le 4n\varepsilon$ and $|r_i \cos\phi_i - t_i \cos\phi_i| \le 4n\varepsilon$. Notice how the focus on radial factors $r_0, \ldots, r_n$ in the secular equation and the deflation procedure allow the perturbation theory to treat A_{11} and A_{21} equally.

3.3. Singular vectors. Upon finding singular vectors, the CSD will be complete. In the following theorem, the entries of an orthogonal matrix $X = [x_{ij}]$ are specified by the notation $x^{ij} \propto f(i,j)$, which translates to $x_{ij} = f(i,j)/\sqrt{\sum_i f(i,j)^2}$.

THEOREM 5. *The matrix* $\left[\begin{smallmatrix}\tilde{A}_{11}\\ \tilde{A}_{21}\end{smallmatrix}\right]$ *of* (8) *has orthonormal columns. In the CSD*

$$\begin{bmatrix}\tilde{A}_{11}\\ \tilde{A}_{21}\end{bmatrix} = \begin{bmatrix}\tilde{U}_1 & \\ & \tilde{U}_2\end{bmatrix}\begin{bmatrix}C\\ S\end{bmatrix}\tilde{V}_1^T,$$

the singular vectors are given by

$$\tilde{u}_1^{ij} \propto \begin{cases} \frac{t_n}{\sin(\phi_n+\hat{\theta}_j)\sin(\phi_n-\hat{\theta}_j)}, & i=1,\\ \frac{t_{i-1}\sin\phi_{i-1}}{\sin(\phi_{i-1}+\hat{\theta}_j)\sin(\phi_{i-1}-\hat{\theta}_j)}, & i=2,\ldots,n,\end{cases}$$

$$\tilde{u}_2^{ij} \propto \frac{t_{i-1}\cos\phi_{i-1}}{\sin(\phi_{i-1}+\hat{\theta}_j)\sin(\phi_{i-1}-\hat{\theta}_j)},$$

$$\tilde{v}_1^{ij} \propto \begin{cases} -1, & i=1,\\ \frac{t_{i-1}(\cos\phi_{i-1})(\sin\phi_{i-1})}{\sin(\phi_{i-1}+\hat{\theta}_j)\sin(\phi_{i-1}-\hat{\theta}_j)}, & i=2,\ldots,n.\end{cases}$$

Proof. According to the standard SVD theory, the singular vectors of the top block are given by

$$\tilde{u}_1^{ij} \propto \begin{cases} \frac{-t_n}{\cos^2\phi_n-\cos^2\hat{\theta}_j}, & i=1,\\ \frac{-t_{i-1}\sin\phi_{i-1}}{\cos^2\phi_{i-1}-\cos^2\hat{\theta}_j}, & i=2,\ldots,n,\end{cases}$$

$$\tilde{v}_1^{ij} \propto \begin{cases} -1, & i=1,\\ \frac{-t_{i-1}(\cos\phi_{i-1})(\sin\phi_{i-1})}{\cos^2\phi_{i-1}-\cos^2\hat{\theta}_j}, & i=2,\ldots,n,\end{cases}$$

and the singular vectors of the bottom block are given by

$$\tilde{u}_2^{ij} \propto \begin{cases} \frac{t_0}{\sin^2\phi_0-\sin^2\hat{\theta}_j}, & i=1,\\ \frac{t_{i-1}\cos\phi_{i-1}}{\sin^2\phi_{i-1}-\sin^2\hat{\theta}_j}, & i=2,\ldots,n,\end{cases}$$

$$\tilde{v}_1^{ij} \propto \begin{cases} -1, & i=1,\\ \frac{t_{i-1}(\cos\phi_{i-1})(\sin\phi_{i-1})}{\sin^2\phi_{i-1}-\sin^2\hat{\theta}_j}, & i=2,\ldots,n.\end{cases}$$

An application of $\cos^2 a = 1-\sin^2 a$ shows that the expressions for $\tilde{v}_1^{ij}$ are equal. The formulas in the statement of the lemma are obtained by applying the trigonometric identities $\sin^2 a-\sin^2 b = \sin(a+b)\sin(a-b)$ and $\cos^2 a-\cos^2 b = -\sin(a+b)\sin(a-b)$. The CSD has been constructed, and this proves that the matrix has orthonormal columns. ∎

4. Numerical stability. The theorem below is the main stability result. Its proof extends over the rest of the section. Our computational model is the usual

one, consisting of a set of floating-point numbers $\mathbf{F}$ and operations $+$, $-$, $\times$, and $\div$ satisfying $|(a \circ b) - \mathrm{fl}(a \circ b)| \leq |a \circ b|\,\mathbf{u}$ for all $a, b \in \mathbf{F}$, in which $\mathbf{u}$ is unit roundoff. Notation from Higham's book [17] is used, particularly $\gamma_k = \frac{k\mathbf{u}}{1-k\mathbf{u}}$ and $\tilde{\gamma}_k = \frac{Ck\mathbf{u}}{1-Ck\mathbf{u}}$ for an implicit constant C that depends on neither the problem size n nor the input matrix. We also need routines for $\sqrt{a}$, $a \geq 0$, and $\sin a$, $-\pi/2 \leq a \leq \pi/2$, that achieve low relative error, each bounded by some γ_k.

THEOREM 6. *The divide-and-conquer algorithm for the bidiagonal CSD is backward stable. Given upper-bidiagonal B_{11} and B_{21}, it computes*

$$\begin{bmatrix} U_1 & \\ & U_2 \end{bmatrix}^T \begin{bmatrix} B_{11} + \Delta B_{11} \\ B_{21} + \Delta B_{21} \end{bmatrix} V_1 = \begin{bmatrix} C \\ S \end{bmatrix} \tag{14}$$

with the following properties:

1. *The right-hand side has diagonal blocks and exactly orthonormal columns. The matrices $C = \mathrm{diag}(\cos\hat{\theta}_1, \ldots, \cos\hat{\theta}_n)$ and $S = \mathrm{diag}(\sin\hat{\theta}_1, \ldots, \sin\hat{\theta}_n)$ are defined implicitly by angles $\hat{\theta}_1, \ldots, \hat{\theta}_n$ computed in finite precision.*
2. *The backward error is small: $\left\|\Delta B_{11}\right\|_F \leq C_1\tilde{\gamma}_{n^{5/2}}$ and $\left\|\Delta B_{21}\right\|_F \leq C_1\tilde{\gamma}_{n^{5/2}}$, in which*
$$C_1 = \max\left(1, \left\|I - \begin{bmatrix} B_{11} \\ B_{21} \end{bmatrix}^T \begin{bmatrix} B_{11} \\ B_{21} \end{bmatrix}\right\|_2 \Big/ (n^{5/2}\mathbf{u})\right).$$
3. *U_1, U_2, and V_1 refer to exactly orthogonal matrices that are approximated by $\hat{U}_1$, $\hat{U}_2$, and $\hat{V}_1$ satisfying $\|U_1 - \hat{U}_1\|_F \leq \tilde{\gamma}_{n^2}$, $\|U_2 - \hat{U}_2\|_F \leq \tilde{\gamma}_{n^2}$, and $\|V_1 - \hat{V}_1\|_F \leq \tilde{\gamma}_{n^2}$.*

Note that simultaneous bidiagonalization produces matrices for which C_1 is small, independent of the original dense matrix's deviation from orthogonality [33]. (If the input matrix is far from orthogonal, then the backward error is large, but $\left[\begin{smallmatrix} B_{11} \\ B_{21} \end{smallmatrix}\right]$ has orthonormal columns regardless.)

Gu and Eisenstat comment that their error bounds are conservative and may include a factor of n that is not seen in practice [15, 16]. If that were proved, then $n^{5/2}$ could be replaced by n^2 in the bounds on $\left\|\Delta B_{11}\right\|_F$ and $\left\|\Delta B_{21}\right\|_F$ and in the definition of C_1.

The proof assumes the existence of a bidiagonal CSD routine—perhaps simultaneous QR iteration—that satisfies the conclusions of Theorem 6 for matrices with no more than $n/2$ columns. The divide-and-conquer algorithm extends this stability to a problem of size n. Of course, an even larger problem could be solved through recursion. The proof also assumes that $n \geq 3$, so that the subproblem sizes m and $n-m-1$ are greater than zero, and that C_1 and n are small enough, and the arithmetic precise enough, that $C_1\gamma_k \leq 1/2$ for every γ_k encountered.

4.1. Numerical lemmas. The following numerical lemmas are needed. Their proofs are omitted.

LEMMA 7. *Suppose $\left\|I - Q^TQ\right\|_2 \leq c_1\gamma_j$ and $\|E\|_2 \leq c_2\gamma_k$. If c_2 is small enough that $c_2\gamma_k \leq 1$, then $\left\|I - (Q+E)^T(Q+E)\right\|_2 \leq \max(c_1, c_2)\,(\gamma_{3j+3k})$.*

LEMMA 8. *If Q is an n-by-n orthogonal matrix and $\hat{Q}$ is an approximation satisfying $\|Q - \hat{Q}\|_F \leq \gamma_k$, then $\mathrm{fl}(\hat{Q}x) = Q(x + \Delta x)$ for a backward error Δx of size $\|\Delta x\|_2 \leq \gamma_{k+n^{3/2}}\|x\|_2$.*

LEMMA 9. *Suppose A and B are n-by-n matrices whose 2-norms are at most 1, and let $\hat{A}$ and $\hat{B}$ be approximations satisfying $\|A - \hat{A}\|_F \leq \gamma_k$ and $\|B - \hat{B}\|_F \leq \gamma_l$ with $\gamma_k, \gamma_l \leq 1$. Then $\|AB - \mathrm{fl}(\hat{A}\hat{B})\|_F \leq \tilde{\gamma}_{n^2+k+l}$.*

4.2. Dissection. The following lemma revisits (3).

LEMMA 10. *The algorithm computes*

$$\begin{bmatrix} U_1 & \\ & U_2 \end{bmatrix}^T \begin{bmatrix} B_{11}+\Delta B_{11} \\ B_{21}+\Delta B_{21} \end{bmatrix} V_1 = \begin{bmatrix} \hat{P} \\ \hat{Q} \end{bmatrix} = \left[\begin{array}{ccc} \hat{p}_{11} & \hat{P}_{12} & 0 \\ \hat{p}_{21} & 0 & \hat{P}_{23} \\ \hat{p}_{31} & 0 & 0 \\ \hline \hat{q}_{11} & \hat{Q}_{12} & 0 \\ \hat{q}_{21} & 0 & \hat{Q}_{23} \\ \hat{q}_{31} & 0 & 0 \end{array}\right]$$

with the following properties:

1. $\begin{bmatrix} \hat{P}_{12} \\ \hat{Q}_{12} \end{bmatrix}$ *and* $\begin{bmatrix} \hat{P}_{23} \\ \hat{Q}_{23} \end{bmatrix}$ *have upper-bidiagonal blocks and satisfy*

$$\left\| I - \begin{bmatrix} \hat{P}_{12} \\ \hat{Q}_{12} \end{bmatrix}^T \begin{bmatrix} \hat{P}_{12} \\ \hat{Q}_{12} \end{bmatrix} \right\|_2 \leq C_1 \tilde{\gamma}_{m^{5/2}}, \tag{15}$$

$$\left\| I - \begin{bmatrix} \hat{P}_{23} \\ \hat{Q}_{23} \end{bmatrix}^T \begin{bmatrix} \hat{P}_{23} \\ \hat{Q}_{23} \end{bmatrix} \right\|_2 \leq C_1 \tilde{\gamma}_{(n-m-1)^{5/2}}. \tag{16}$$

2. $\left\|\Delta B_{11}\right\|_F \leq \tilde{\gamma}_{n^{3/2}}$ *and* $\left\|\Delta B_{21}\right\|_F \leq \tilde{\gamma}_{n^{3/2}}$.
3. U_1 *and* U_2 *are products of Givens rotations and are approximated by* $\hat{U}_1$, $\hat{U}_2$ *satisfying* $\|U_1 - \hat{U}_1\|_F \leq \tilde{\gamma}_{n^{3/2}}$ *and* $\|U_2 - \hat{U}_2\|_F \leq \tilde{\gamma}_{n^{3/2}}$. V_1 *is a permutation matrix.*

The proof is a straightforward application of Givens rotations and permutations. It is omitted for brevity.

4.3. Recursion. The next lemma considers the recursive CSD computation.

LEMMA 11. *The algorithm computes*

$$\begin{bmatrix} U_1 & \\ & U_2 \end{bmatrix}^T \begin{bmatrix} B_{11}+\Delta B_{11} \\ B_{21}+\Delta B_{21} \end{bmatrix} V_1 = \left[\begin{array}{cc} \hat{z}_n & 0 \\ \hat{z}_{1:n-1} & \cos\hat{\Psi} \\ \hline \hat{w}_0 & 0 \\ \hat{w}_{1:n-1} & \sin\hat{\Psi} \end{array}\right]$$

with the following properties:

1. $\hat{\Psi}$ *is a diagonal matrix with finite-precision diagonal entries* $\hat{\psi}_1, \ldots, \hat{\psi}_{n-1}$ *satisfying* $0 \leq \hat{\psi}_1 \leq \cdots \leq \hat{\psi}_{n-1} \leq \pi/2$.
2. $\left\|\Delta B_{11}\right\|_F \leq C_1 \tilde{\gamma}_{n^{5/2}}$ *and* $\left\|\Delta B_{21}\right\|_F \leq C_1 \tilde{\gamma}_{n^{5/2}}$.
3. U_1, U_2, *and* V_1 *refer to exactly orthogonal matrices that are approximated by* $\hat{U}_1$, $\hat{U}_2$, *and* $\hat{V}_1$ *satisfying* $\|U_1 - \hat{U}_1\|_F \leq \tilde{\gamma}_{n^2}$, $\|U_2 - \hat{U}_2\|_F \leq \tilde{\gamma}_{n^2}$, *and* $\|V_1 - \hat{V}_1\|_F \leq \tilde{\gamma}_{n^2}$.

Proof. The proof shows that (4) is computed stably. The subsequent step to (5) consists of a permutation and sign changes and introduces no additional error.

By assumption, we can stably compute CSDs of matrices with fewer than n columns, specifically,

$$\begin{bmatrix} \hat{P}_{12}+\Delta P_{12} \\ \hat{Q}_{12}+\Delta Q_{12} \end{bmatrix} = \begin{bmatrix} U_1^{(a)} & \\ & U_2^{(a)} \end{bmatrix} \begin{bmatrix} \cos\hat{\Psi}^{(a)} \\ \sin\hat{\Psi}^{(a)} \end{bmatrix} (V_1^{(a)})^T,$$

$$\begin{bmatrix} \hat{P}_{23}+\Delta P_{23} \\ \hat{Q}_{23}+\Delta Q_{23} \end{bmatrix} = \begin{bmatrix} U_1^{(b)} & \\ & U_2^{(b)} \end{bmatrix} \begin{bmatrix} \cos\hat{\Psi}^{(b)} \\ \sin\hat{\Psi}^{(b)} \end{bmatrix} (V_1^{(b)})^T$$

for some backward errors ΔP_{12}, ΔQ_{12}, ΔP_{23}, and ΔQ_{23}. The orthogonality bounds of (15)–(16) can be written $C_1 C_2 m^{5/2}\mathbf{u}$ and $C_1 C_3 (n-m-1)^{5/2}\mathbf{u}$, respectively, for constants C_2 and C_3 independent of n, m, and C_1. Then the backward errors are

$$\begin{aligned} \|\Delta P_{12}\|_F &\le C_1 \tilde{\gamma}_{m^{5/2}}, & \|\Delta P_{23}\|_F &\le C_1 \tilde{\gamma}_{(n-m-1)^{5/2}}, \\ \|\Delta Q_{12}\|_F &\le C_1 \tilde{\gamma}_{m^{5/2}}, & \|\Delta Q_{23}\|_F &\le C_1 \tilde{\gamma}_{(n-m-1)^{5/2}}, \end{aligned}$$

having reabsorbed the constants C_2, C_3 into the $\tilde{\gamma}_\cdot$ factors. The orthogonal transformations are approximated as follows:

$$\begin{aligned} \left\|U_1^{(a)} - \hat{U}_1^{(a)}\right\|_F &\le \tilde{\gamma}_{m^2}, & \left\|U_1^{(b)} - \hat{U}_1^{(b)}\right\|_F &\le \tilde{\gamma}_{(n-m-1)^2}, \\ \left\|U_2^{(a)} - \hat{U}_2^{(a)}\right\|_F &\le \tilde{\gamma}_{m^2}, & \left\|U_2^{(b)} - \hat{U}_2^{(b)}\right\|_F &\le \tilde{\gamma}_{(n-m-1)^2}, \\ \left\|V_1^{(a)} - \hat{V}_1^{(a)}\right\|_F &\le \tilde{\gamma}_{m^2}, & \left\|V_1^{(b)} - \hat{V}_1^{(b)}\right\|_F &\le \tilde{\gamma}_{(n-m-1)^2}. \end{aligned}$$

To update the first column of the larger $2n$-by-n matrix, define x_{11}, x_{21}, y_{11}, y_{21} and compute $\hat{x}_{11}$, $\hat{x}_{21}$, $\hat{y}_{11}$, $\hat{y}_{21}$, as follows:

$$\begin{aligned} x_{1:m} &= (U_1^{(a)})^T \hat{p}_{11}, & \hat{x}_{1:m} &= (\hat{U}_1^{(a)})^T (\hat{p}_{11} + \Delta p_{11}), & \|\Delta p_{11}\|_2 &\le \tilde{\gamma}_{m^2}, \\ x_{m+1:n-1} &= (U_1^{(b)})^T \hat{p}_{21}, & \hat{x}_{m+1:n-1} &= (\hat{U}_1^{(b)})^T (\hat{p}_{21} + \Delta p_{21}), & \|\Delta p_{21}\|_2 &\le \tilde{\gamma}_{(n-m-1)^2}, \\ x_n &= \hat{p}_{31}, & \hat{x}_n &= \hat{p}_{31}, & & \\ y_{1:m} &= (U_2^{(a)})^T \hat{q}_{11}, & \hat{y}_{1:m} &= (\hat{U}_2^{(a)})^T (\hat{q}_{11} + \Delta q_{11}), & \|\Delta q_{11}\|_2 &\le \tilde{\gamma}_{m^2}, \\ y_{m+1:n-1} &= (U_2^{(b)})^T \hat{q}_{21}, & \hat{y}_{m+1:n-1} &= (\hat{U}_2^{(b)})^T (\hat{q}_{21} + \Delta q_{21}), & \|\Delta q_{21}\|_2 &\le \tilde{\gamma}_{(n-m-1)^2}, \\ y_n &= \hat{q}_{31}, & \hat{y}_n &= \hat{q}_{31}. & & \end{aligned}$$

(The error bounds follow from Lemma 8.)

Let $\bar{U}_1$, $\bar{U}_2$, and $\bar{V}_1$ be the orthogonal transformations from the previous lemma, and define $U_1 = \bar{U}_1(U_1^{(a)} \oplus U_1^{(b)} \oplus 1)$, $U_2 = \bar{U}_2(U_2^{(a)} \oplus U_2^{(b)} \oplus 1)$, and $V_1 = \bar{V}_1(1 \oplus V_1^{(a)} \oplus V_1^{(b)})$, in which $A \oplus B$ is shorthand for the block-diagonal matrix with blocks A and B. In finite precision, Lemma 9 gives

$$\begin{aligned} \hat{U}_1 &= \mathrm{fl}(\mathrm{fl}(\bar{U}_1)(\hat{U}_1^{(a)} \oplus \hat{U}_1^{(b)} \oplus 1)), & \left\|U_1 - \hat{U}_1\right\|_F &\le \tilde{\gamma}_{n^2}, \\ \hat{U}_2 &= \mathrm{fl}(\mathrm{fl}(\bar{U}_2)(\hat{U}_2^{(a)} \oplus \hat{U}_2^{(b)} \oplus 1)), & \left\|U_2 - \hat{U}_2\right\|_F &\le \tilde{\gamma}_{n^2}, \end{aligned}$$

and since $\bar{V}_1$ is just a permutation,

$$\hat{V}_1 = \bar{V}_1(1 \oplus \hat{V}_1^{(a)} \oplus \hat{V}_1^{(b)}), \qquad \left\|V_1 - \hat{V}_1\right\|_F \le \tilde{\gamma}_{n^2}.$$

For the backward error analysis, let $\Delta\bar{B}_{11}$ and $\Delta\bar{B}_{21}$ be the backward errors from the previous lemma. Update these by defining

$$\begin{aligned} \Delta B_{11} &= \Delta\bar{B}_{11} + \bar{U}_1 \begin{bmatrix} \Delta p_{11} & \Delta P_{12} & 0 \\ \Delta p_{21} & 0 & \Delta P_{23} \\ 0 & 0 & 0 \end{bmatrix} \bar{V}_1^T, \\ \Delta B_{21} &= \Delta\bar{B}_{21} + \bar{U}_2 \begin{bmatrix} \Delta q_{11} & \Delta Q_{12} & 0 \\ \Delta q_{21} & 0 & \Delta Q_{23} \\ 0 & 0 & 0 \end{bmatrix} \bar{V}_1^T. \end{aligned}$$

Then

$$\begin{bmatrix} U_1 & \\ & U_2 \end{bmatrix}^T \begin{bmatrix} B_{11}+\Delta B_{11} \\ B_{21}+\Delta B_{21} \end{bmatrix} V_1 = \left[\begin{array}{ccc} \hat{x}_{1:m} & \cos\hat{\Psi}^{(a)} & 0 \\ \hat{x}_{m+1:n-1} & 0 & \cos\hat{\Psi}^{(b)} \\ \hat{x}_n & 0 & 0 \\ \hline \hat{y}_{1:m} & \sin\hat{\Psi}^{(a)} & 0 \\ \hat{y}_{m+1:n-1} & 0 & \sin\hat{\Psi}^{(b)} \\ \hat{y}_n & 0 & 0 \end{array}\right].$$

The backward errors satisfy

$$\begin{aligned} \|\Delta B_{11}\|_F &\le \left\|\Delta\bar{B}_{11}\right\|_F + \sqrt{\|\Delta p_{11}\|_2^2 + \|\Delta p_{21}\|_2^2 + \|\Delta P_{12}\|_F^2 + \|\Delta P_{22}\|_F^2} \\ &\le \tilde{\gamma}_{n^{3/2}} + \sqrt{\tilde{\gamma}_{m^2}^2 + \tilde{\gamma}_{(n-m-1)^2}^2 + C_1^2\tilde{\gamma}_{m^{5/2}}^2 + C_1^2\tilde{\gamma}_{(n-m-1)^{5/2}}^2} \le C_1\tilde{\gamma}_{n^{5/2}}, \end{aligned}$$

and similarly $\|\Delta B_{21}\|_F \le C_1\tilde{\gamma}_{n^{5/2}}$. ∎

4.4. Reparameterization to enforce orthogonality. At this point, little has been done to the first column. By converting from Cartesian to polar coordinates, the next step in the algorithm orthonormalizes the first column against the rest. The CSD is then well defined, and many later steps are expressed in terms of the polar coordinates instead of the actual matrix entries.

LEMMA 12. *The algorithm computes*

$$\begin{bmatrix} U_1 & \\ & U_2 \end{bmatrix}^T \begin{bmatrix} B_{11}+\Delta B_{11} \\ B_{21}+\Delta B_{21} \end{bmatrix} V_1 = \left[\begin{array}{ccc} -\beta\hat{r}_n & & \\ -\beta\hat{r}_1\sin\hat{\psi}_1 & \cos\hat{\psi}_1 & \\ \vdots & & \ddots \\ -\beta\hat{r}_{n-1}\sin\hat{\psi}_{n-1} & & \cos\hat{\psi}_{n-1} \\ \hline \beta\hat{r}_0 & & \\ \beta\hat{r}_1\cos\hat{\psi}_1 & \sin\hat{\psi}_1 & \\ \vdots & & \ddots \\ \beta\hat{r}_{n-1}\cos\hat{\psi}_{n-1} & & \sin\hat{\psi}_{n-1} \end{array}\right], \tag{17}$$

with the following properties:

1. *The matrix on the right-hand side has exactly orthonormal columns.*
2. *$\hat{r}_0, \ldots, \hat{r}_n$ and $\hat{\psi}_1, \ldots, \hat{\psi}_{n-1}$ are stored in finite precision.*
3. *The normalizing coefficient β is not computed in finite precision but satisfies $|1-\beta| \le C_1\tilde{\gamma}_{n^{5/2}}$.*
4. *$\left\|\Delta B_{11}\right\|_F \le C_1\tilde{\gamma}_{n^{5/2}}$ and $\left\|\Delta B_{21}\right\|_F \le C_1\tilde{\gamma}_{n^{5/2}}$.*
5. *U_1, U_2, and V_1 are unchanged from the previous lemma.*

Proof. Let $\Delta\bar{B}_{11}$ and $\Delta\bar{B}_{21}$ be the backward errors from the previous lemma. Define

$$G = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & \cos\hat{\Psi} & 0 & -\sin\hat{\Psi} \\ 0 & 0 & 1 & 0 \\ 0 & \sin\hat{\Psi} & 0 & \cos\hat{\Psi} \end{bmatrix}$$

and note that

$$G^T \begin{bmatrix} U_1 & \\ & U_2 \end{bmatrix}^T \begin{bmatrix} B_{11} + \Delta\bar{B}_{11} \\ B_{21} + \Delta\bar{B}_{21} \end{bmatrix} V_1 = \left[\begin{array}{cc} -\hat{z}_n & 0 \\ (\cos\hat{\Psi})\hat{z}_{1:n-1} + (\sin\hat{\Psi})\hat{w}_{1:n-1} & I \\ \hline \hat{w}_0 & 0 \\ -(\sin\hat{\Psi})\hat{z}_{1:n-1} + (\cos\hat{\Psi})\hat{w}_{1:n-1} & 0 \end{array}\right] =: M.$$

M has nearly orthonormal columns. Specifically, Lemma 7 shows that

$$\left\|I - M^T M\right\|_2 \leq C_1 \tilde{\gamma}_{n^{5/2}}.$$

Hence, $m_{21} := (\cos\hat{\Psi})\hat{z}_{1:n-1} + (\sin\hat{\Psi})\hat{w}_{1:n-1}$ must already be approximately zero. In fact, this vector lies in the first column of $M^T M - I$ and, therefore, satisfies

$$\|m_{21}\|_2 \leq C_1 \tilde{\gamma}_{n^{5/2}}.$$

We shall replace this vector by zeros, and then the columns will become exactly orthogonal. Let

$$(r_1, \ldots, r_{n-1}) = -(\sin\hat{\Psi})\hat{z}_{1:n-1} + (\cos\hat{\Psi})\hat{w}_{1:n-1},$$

$r_0 = \hat{w}_0$, and $r_n = -\hat{z}_0$, and compute

$$(\hat{r}_1, \ldots, \hat{r}_{n-1}) = \mathrm{fl}\left(-(\sin\hat{\Psi})\hat{z}_{1:n-1} + (\cos\hat{\Psi})\hat{w}_{1:n-1}\right),$$

$\hat{r}_0 = \hat{w}_0$, and $\hat{r}_n = -\hat{z}_n$. The forward errors are small: $|r_i - \hat{r}_i| \leq \tilde{\gamma}_1$. We have

$$G^T \begin{bmatrix} U_1 & \\ & U_2 \end{bmatrix}^T \begin{bmatrix} B_{11} + \Delta\bar{B}_{11} \\ B_{21} + \Delta\bar{B}_{21} \end{bmatrix} V_1 = \left[\begin{array}{cc} -\hat{r}_n & 0 \\ 0 & I \\ \hline \hat{r}_0 & 0 \\ \hat{r}_{1:n-1} & 0 \end{array}\right] + \left[\begin{array}{cc} 0 & 0 \\ m_{21} & 0 \\ \hline 0 & 0 \\ r_{1:n-1} - \hat{r}_{1:n-1} & 0 \end{array}\right] =: R + E.$$

R will have exactly orthonormal columns once its first column is normalized. To do that, let $\beta = 1/\sqrt{\hat{r}_0^2 + \cdots + \hat{r}_n^2}$. By Lemma 7, $\left\|I - R^T R\right\|_2 \leq C_1 \tilde{\gamma}_{n^{5/2}}$. The top-left entry of $I - R^T R$ equals $1 - 1/\beta^2$, and, therefore, $\left|1 - 1/\beta^2\right| \leq C_1 \tilde{\gamma}_{n^{5/2}}$, and so $|1 - \beta| \leq C_1 \tilde{\gamma}_{n^{5/2}}$ as well. Then

$$G^T \begin{bmatrix} U_1 & \\ & U_2 \end{bmatrix}^T \begin{bmatrix} B_{11} + \Delta\bar{B}_{11} \\ B_{21} + \Delta\bar{B}_{21} \end{bmatrix} V_1 = \left[\begin{array}{cc} -\beta\hat{r}_n & 0 \\ 0 & I \\ \hline \beta\hat{r}_0 & 0 \\ \beta\hat{r}_{1:n-1} & 0 \end{array}\right] + \left[\begin{array}{cc} -(1-\beta)\hat{r}_n & 0 \\ m_{21} & 0 \\ \hline (1-\beta)\hat{r}_0 & 0 \\ -\beta\hat{r}_{1:n-1} + r_{1:n-1} & 0 \end{array}\right].$$

Let

$$\begin{bmatrix} \Delta B_{11} \\ \Delta B_{21} \end{bmatrix} = \begin{bmatrix} \Delta\bar{B}_{11} \\ \Delta\bar{B}_{21} \end{bmatrix} - \begin{bmatrix} U_1 & \\ & U_2 \end{bmatrix} G \left[\begin{array}{cc} -(1-\beta)\hat{r}_n & 0 \\ m_{21} & 0 \\ \hline (1-\beta)\hat{r}_0 & 0 \\ -\beta\hat{r}_{1:n-1} + r_{1:n-1} & 0 \end{array}\right] V_1^T.$$

Then

$$\begin{bmatrix} U_1 & \\ & U_2 \end{bmatrix}^T \begin{bmatrix} B_{11} + \Delta B_{11} \\ B_{21} + \Delta B_{21} \end{bmatrix} V_1 = G \left[\begin{array}{cc} -\beta\hat{r}_n & 0 \\ 0 & I \\ \hline \beta\hat{r}_0 & 0 \\ \beta\hat{r}_{1:n-1} & 0 \end{array}\right].$$

This is the decomposition in the statement of the theorem.

The backward errors satisfy

$$\begin{aligned}\|\Delta B_{11}\|_F &\le \left\|\Delta \bar{B}_{11}\right\|_F + \sqrt{(1-\beta)^2 \hat{r}_n^2 + \|m_{21}\|_2^2} \\ &\le C_1 \tilde{\gamma}_{n^{5/2}} + \sqrt{C_1^2 \tilde{\gamma}_{n^{5/2}}^2 + C_1^2 \tilde{\gamma}_{n^{5/2}}^2} \le C_1 \tilde{\gamma}_{n^{5/2}}\end{aligned}$$

and

$$\begin{aligned}\|\Delta B_{21}\|_F &\le \left\|\Delta \bar{B}_{21}\right\|_F + \sqrt{(1-\beta)^2 \left\|\hat{r}_{0:n-1}\right\|_2^2 + \left\|r_{1:n-1} - \hat{r}_{1:n-1}\right\|_2^2} \\ &\le C_1 \tilde{\gamma}_{n^{5/2}} + \sqrt{C_1^2 \tilde{\gamma}_{n^{5/2}}^2 + n\tilde{\gamma}_1^2} \le C_1 \tilde{\gamma}_{n^{5/2}}. \qquad \square\end{aligned}$$

4.5. Deflation. Soon, the secular equation must be solved. The tolerance τ of Theorem 4 balances speed against accuracy in the zero finder. When a feature of the secular equation is too fine to be resolved at the given tolerance, one of four deflation procedures reduces the problem. Together, the deflation procedures guarantee the following:

1. $|r_0| \ge \tau$ and $|r_n| \ge \tau$,
2. $|r_i| \ge \tau$, $i = 1, \ldots, n-1$,
3. $\tau \le \psi_1 \le \cdots \le \psi_{n-1} \le \pi/2 - \tau$,
4. $\left|\psi_{i+1} - \psi_i\right| \ge \tau$, $i = 1, \ldots, n-2$.

The deflation procedures closely parallel those for the bidiagonal SVD problem [16].

First, if $|r_0| < \tau$ or $|r_n| < \tau$, then the offending parameter can be replaced by τ, incurring only a small backward error.

Second, if $|r_i| < \tau$ is small for some $1 \le i \le n-1$, then entries $i+1$ and $n+i+1$ of the first column of (17) are negligible. In this case, they can be truncated to zero and rows and columns can be permuted to reduce the problem size.

Third, if $\psi_1 < \tau$, then the $(2,1)$- and $(n+2,2)$-entries of (17) are negligible and can be truncated to zero. Then a Givens rotation on rows $n+1$ and $n+2$ can zero the $(n+2,1)$-entry, and a permutation of rows and columns produces a deflated problem with $n-1$ columns. Similarly, if $\psi_{n-1} > \pi/2 - \tau$, then two entries are negligible, and a truncation, rotation, and permutation deflate the problem.

Finally, if $\psi_{i+1} - \psi_i < \tau$ for some $1 \le i < n-1$, then ψ_{i+1} may be set equal to ψ_i at the cost of a small backward error, producing multiples of the 2-by-2 identity matrix in columns $i+1$ and $i+2$. Givens rotations from the left and right can then zero entries $i+2$ and $n+i+2$ of the first column while leaving columns $i+1$ and $i+2$ fixed, and the newly null entries can be swept away. This leaves a deflated problem with $n-1$ columns.

To keep the notation clean in the rest of the paper, we simply assume that no deflation is necessary.

4.6. Focus on angle gaps. In the upcoming secular equation, differences of the form $\hat{\psi}_i - \theta$ play an important role. Unfortunately, these cannot be computed to high relative accuracy given the current representation. The solution is to store angle gaps $\Delta\hat{\phi}_i = \hat{\psi}_{i+1} - \hat{\psi}_i$, $i = 0, \ldots, n-1$, in finite precision. Then differences and trigonometric functions can be computed more accurately. It may appear that accuracy is conjured from thin air, but the error does not disappear; it is thrown backward and absorbed into ΔB_{11} and ΔB_{21}. The following lemma executes this idea and introduces our final rank-one modification problem.

LEMMA 13. *The algorithm computes*

(18)

$$\begin{bmatrix} U_1 & \\ & U_2 \end{bmatrix}^T \begin{bmatrix} B_{11}+\Delta B_{11} \\ B_{21}+\Delta B_{21} \end{bmatrix} V_1 = \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix} = \left[\begin{array}{cccc} -\beta\hat{r}_n & & & \\ -\beta\hat{r}_1\sin\phi_1 & \cos\phi_1 & & \\ \vdots & & \ddots & \\ -\beta\hat{r}_{n-1}\sin\phi_{n-1} & & & \cos\phi_{n-1} \\ \hline \beta\hat{r}_0 & & & \\ \beta\hat{r}_1\cos\phi_1 & \sin\phi_1 & & \\ \vdots & & \ddots & \\ \beta\hat{r}_{n-1}\cos\phi_{n-1} & & & \sin\phi_{n-1} \end{array}\right]$$

with the following properties:

1. *The matrix on the right-hand side has exactly orthonormal columns.*
2. *The angles are defined by*

$$\phi_i = \alpha\sum_{j=0}^{i-1}\Delta\hat{\phi}_j,\ i=1,\dots,n-1, \qquad \alpha^{-1} = \frac{2}{\pi}\sum_{j=0}^{n-1}\Delta\hat{\phi}_j,$$

   with the angle gaps $\Delta\hat{\phi}_0,\dots,\Delta\hat{\phi}_{n-1}$ stored in finite precision.
3. *The normalizing coefficient α is not stored in finite precision but satisfies $|1-\alpha|\le\tilde{\gamma}_n$.*
4. *Deflation conditions* 1–4 *of section* 4.5 *hold.*
5. *$\|\Delta B_{11}\|_F \le C_1\tilde{\gamma}_{n^{5/2}}$ and $\|\Delta B_{21}\|_F \le C_1\tilde{\gamma}_{n^{5/2}}$.*
6. *U_1, U_2, V_1, β, and $\hat{r}_0,\dots,\hat{r}_n$ are unchanged from the previous lemma.*

Proof. Compute $\Delta\hat{\phi}_i = \mathrm{fl}(\hat{\psi}_{i+1}-\hat{\psi}_i)$, $i=0,\dots,n-1$, using the shorthand $\hat{\psi}_0=0$, $\hat{\psi}_n=\pi/2$. Then

$$\left|1-\alpha^{-1}\right| = \left|1-\frac{2}{\pi}\sum_{j=0}^{n-1}\Delta\hat{\phi}_j\right| \le \left|1-\frac{2}{\pi}\sum_{j=0}^{n-1}(\hat{\psi}_{j+1}-\hat{\psi}_j)\right| + \tilde{\gamma}_n = \tilde{\gamma}_n,$$

so $|1-\alpha|\le\tilde{\gamma}_n$ for some implicit constant as well. Also,

$$\left|\hat{\psi}_i-\phi_i\right| = \left|\hat{\psi}_i-\alpha\sum_{j=0}^{i-1}\Delta\hat{\phi}_j\right| \le \left|\hat{\psi}_i-\alpha\sum_{j=0}^{i-1}(\hat{\psi}_{j+1}-\hat{\psi}_j)\right| + \tilde{\gamma}_n \le |1-\alpha|\,\hat{\psi}_i + \tilde{\gamma}_n \le \tilde{\gamma}_n.$$

Replacing every $\hat{\psi}_i$ by ϕ_i in the right-hand side of (17) perturbs each entry by at most $\tilde{\gamma}_n$. There are $4n-4$ perturbed entries, so the Frobenius norm of the perturbation is at most $\sqrt{4n-4}\,\tilde{\gamma}_n \le \tilde{\gamma}_{n^{3/2}}$. This perturbation is moved to the left-hand side and absorbed into ΔB_{11} and ΔB_{21}. ∎

Technically, the perturbations to $\hat{\psi}_1,\dots,\hat{\psi}_{n-1}$ could cause one of the deflation conditions to fail, but this can be remedied by slightly inflating τ when deciding whether deflation is necessary in the first place.

Many quantities involving $\phi_1,\dots,\phi_{n-1}$ can be computed to high relative accuracy using the angle-gap representation, as follows.

LEMMA 14. *For any $0\le i,k\le n$ and $\theta=\phi_i+\delta$ with $0\le\theta\le\pi/2$ and $-\Delta\hat{\phi}_{i-1}/2\le\delta\le\Delta\hat{\phi}_i/2$, the computations below achieve high relative accuracy. Each*

$\tilde{\varepsilon}_n$ represents a different error term bounded by $\tilde{\gamma}_n$.

$$\begin{aligned}
\phi_k &= \alpha^{-1}\textstyle\sum_{j=0}^{k-1}\Delta\hat{\phi}_j = \mathrm{fl}(\sum_{j=0}^{k-1}\Delta\hat{\phi}_j)(1+\tilde{\varepsilon}_n),\\
\pi/2-\phi_k &= \alpha^{-1}\textstyle\sum_{j=k}^{n-1}\Delta\hat{\phi}_j = \mathrm{fl}(\sum_{j=k}^{n-1}\Delta\hat{\phi}_j)(1+\tilde{\varepsilon}_n),\\
\phi_k+\theta &= \phi_k+\phi_i+\delta = \mathrm{fl}\left(\mathrm{fl}(\phi_k)+\mathrm{fl}(\phi_i)+\delta\right)(1+\tilde{\varepsilon}_n),\\
\pi-(\phi_k+\theta) &= (\tfrac{\pi}{2}-\phi_k)+(\tfrac{\pi}{2}-\phi_i)-\delta = \mathrm{fl}\left(\mathrm{fl}(\tfrac{\pi}{2}-\phi_k)+\mathrm{fl}(\tfrac{\pi}{2}-\phi_i)-\delta\right)(1+\tilde{\varepsilon}_n),\\
\phi_k-\theta &= \begin{cases}\alpha\sum_{j=i}^{k-1}\Delta\hat{\phi}_j-\delta, & k>i,\\ -\delta, & k=i,\\ -\alpha\sum_{j=k}^{i-1}\Delta\hat{\phi}_j-\delta, & k<i\end{cases}\\
&= \begin{cases}\mathrm{fl}\left(\sum_{j=i}^{k-1}\Delta\hat{\phi}_j-\delta\right)(1+\tilde{\varepsilon}_n), & k>i,\\ \mathrm{fl}(-\delta)(1+\tilde{\varepsilon}_1), & k=i,\\ \mathrm{fl}\left(-\sum_{j=k}^{i-1}\Delta\hat{\phi}_j-\delta\right)(1+\tilde{\varepsilon}_n), & k<i.\end{cases}\\
\sin(\phi_k+\theta) &= \begin{cases}\mathrm{fl}(\sin(\mathrm{fl}(\phi_k+\theta)))(1+\tilde{\varepsilon}_n), & \phi_k+\theta\le\frac{\pi}{2},\\ \mathrm{fl}(\sin(\mathrm{fl}(\pi-(\phi_k+\theta))))(1+\tilde{\varepsilon}_n), & \phi_k+\theta\ge\frac{\pi}{2},\end{cases}\\
\sin(\phi_k-\theta) &= \mathrm{fl}(\sin(\mathrm{fl}(\phi_k-\theta)))(1+\tilde{\varepsilon}_n),\\
\cos\phi_k &= \mathrm{fl}(\sin(\mathrm{fl}(\tfrac{\pi}{2}-\phi_k)))(1+\tilde{\varepsilon}_n).
\end{aligned}$$

Proof. In the first two computations, the arithmetic error is relatively small because all terms are positive. Also, the coefficient α was already shown to satisfy $\alpha = 1+\tilde{\varepsilon}_n$, so it can safely be ignored.

In the next three computations, loss of significance is avoided by restricting δ to $[-\Delta\hat{\phi}_{i-1}/2, \Delta\hat{\phi}_i/2]$; in every computation, either δ has the same sign as, or its magnitude is no more than half as much as, the other operand.

For the trigonometric computations, we note that sin can be computed to high relative accuracy on $[-\pi/2, \pi/2]$, and we compute the argument to high relative accuracy using the previous five recipes. ∎

4.7. Secular equation. The secular function for matrix (18) is

$$f(\theta) = \sum_{k=0}^{n}\frac{\hat{r}_k^2}{\sin(\phi_k+\theta)\sin(\phi_k-\theta)},$$

with ϕ_k defined implicitly by $\Delta\hat{\phi}_0, \ldots, \Delta\hat{\phi}_{n-1}$. It is computed accurately by translating the region of interest to the origin.

LEMMA 15. *Let $f_i(\delta) = f(\phi_i+\delta)$. Then $f_i(\delta)$ can be computed accurately for any δ satisfying $-\Delta\hat{\phi}_{i-1}/2 \le \delta \le \Delta\hat{\phi}_i/2$ and $0 \le \phi_i+\delta \le \pi/2$, specifically*

$$|f_i(\delta) - \mathrm{fl}(f_i(\delta))| \le C_4 n\mathbf{u}\sum_{k=0}^{n}\frac{\hat{r}_k^2}{|\sin(\phi_k+\phi_i+\delta)\sin(\phi_k-\phi_i-\delta)|}$$

for some constant C_4.

Proof. By Lemma 14, the kth term in $f_i(\delta)$ can be computed with error

$$\varepsilon_k\frac{\hat{r}_k^2}{\sin(\phi_k+\phi_i+\delta)\sin(\phi_k-\phi_i-\delta)},$$

in which $|\varepsilon_k| \le \tilde{\gamma}_n$ and the implicit constant in $\tilde{\gamma}_n$ is independent of i, k, and n. Because each of its terms is computed to high relative accuracy, the sum is computed to high accuracy relative to the sum of the terms' absolute values. ∎

The lemma shows that $f(\theta)$ can be computed accurately though indirectly over $[\frac{\phi_{i-1}+\phi_i}{2}, \frac{\phi_i+\phi_{i+1}}{2}]$. As i ranges from 0 to n, the entire interval $[0, \frac{\pi}{2}]$ is covered. The bound implies relative accuracy near the poles and absolute accuracy near the zeros.

Gu and Eisenstat suggest a stopping criterion for the zero finder [15]. The analogue for the CSD problem is

$$|\mathrm{fl}(f_i(\delta))| \leq C_4 n \mathbf{u} \sum_{k=0}^{n} \frac{\hat{r}_k^2}{|\sin(\phi_k + \phi_i + \delta)\sin(\phi_k - \phi_i - \delta)|}. \tag{19}$$

If f_i has a zero $\delta \in [-\Delta\hat{\phi}_{i-1}/2, \Delta\hat{\phi}_i/2]$, then $\mathrm{fl}(\delta)$ will satisfy the stopping criterion. When an approximate zero is found, $f_i(\hat{\delta}) \approx 0$, we record $\sigma(j) = i$ and $\hat{\delta}_j = \hat{\delta}$ to indicate that the original secular equation has a solution near $\hat{\theta}_j := \phi_{\sigma(j)} + \hat{\delta}_j$.

4.8. Inverse problem. The computed angles $\hat{\theta}_1, \ldots, \hat{\theta}_n$ are treated as exact by solving the inverse problem of section 3.2. This is done by computing $t_0, \ldots, t_n$, defined by (13).

LEMMA 16. *If the computed solutions $\hat{\theta}_1, \ldots, \hat{\theta}_n$ of the secular equation satisfy the stopping criterion* (19) *and $\tau = 2C_4 n^2 \mathbf{u}$ in the deflation criteria, then the algorithm finds*

$$\begin{bmatrix} U_1 & \\ & U_2 \end{bmatrix}^T \begin{bmatrix} B_{11} + \Delta B_{11} \\ B_{21} + \Delta B_{21} \end{bmatrix} V_1 = \begin{bmatrix} \tilde{A}_{11} \\ \tilde{A}_{21} \end{bmatrix} = \left[\begin{array}{ccc} -t_n & & \\ -t_1 \sin\phi_1 & \cos\phi_1 & \\ \vdots & & \ddots \\ -t_{n-1}\sin\phi_{n-1} & & \cos\phi_{n-1} \\ \hline t_0 & & \\ t_1\cos\phi_1 & \sin\phi_1 & \\ \vdots & & \ddots \\ t_{n-1}\cos\phi_{n-1} & & \sin\phi_{n-1} \end{array}\right]$$

with the following properties:

1. *The matrix on the right-hand side has exactly orthonormal columns, and the singular values of its blocks are* $\cos(\hat{\theta}_i)$ *and* $\sin(\hat{\theta}_i)$*, respectively, for* $i = 1, \ldots, n$.
2. t_i *is approximated by* $\hat{t}_i$ *satisfying* $|t_i - \hat{t}_i| \leq |t_i|\,\tilde{\gamma}_{n^2}$.
3. $\|\Delta B_{11}\|_F \leq C_1 \tilde{\gamma}_{n^{5/2}}$ *and* $\|\Delta B_{21}\|_F \leq C_1 \tilde{\gamma}_{n^{5/2}}$.
4. U_1, U_2, *and* V_1 *are unchanged from the previous lemma.*

Proof. Define t_i by (13), with $\phi_i = \alpha \sum_{j=0}^{i-1} \Delta\hat{\phi}_j$ as above.

The first conclusion was already proved in Theorem 3.

In approximating t_i numerically, every factor of the form $\sin(\cdot)$ can be computed with relative error below $\tilde{\gamma}_n$ by Lemma 14. There are $O(n)$ arithmetic operations on these factors, so the combined error is bounded by $\tilde{\gamma}_{n^2}$.

Let $\Delta\bar{B}_{11}$ and $\Delta\bar{B}_{21}$ be the backward errors from the previous lemma, and let

$$\begin{bmatrix}\Delta B_{11}\\ \Delta B_{21}\end{bmatrix} = \begin{bmatrix}\Delta\bar{B}_{11}\\ \Delta\bar{B}_{21}\end{bmatrix} - \begin{bmatrix}U_1 & \\ & U_2\end{bmatrix} \left[\begin{array}{cc} -\beta\hat{r}_n + t_n & 0\\ (-\beta\hat{r}_1 + t_1)\sin\phi_1 & 0\\ \vdots & 0\\ (-\beta\hat{r}_{n-1} + t_{n-1})\sin\phi_{n-1} & 0\\ \hline \beta\hat{r}_0 - t_0 & 0\\ (\beta\hat{r}_1 - t_1)\cos\phi_1 & 0\\ \vdots & 0\\ (\beta\hat{r}_{n-1} - t_{n-1})\cos\phi_{n-1} & 0\end{array}\right] V_1^T.$$

Then the matrix decomposition of the lemma is satisfied and, making use of Theorem 4,

$$\begin{aligned}\left\|\Delta B_{11}\right\|_F &\le \left\|\Delta\bar{B}_{11}\right\|_F + \sqrt{\sum_{k=1}^{n}(\beta\hat{r}_k - t_k)^2}\\ &\le \left\|\Delta\bar{B}_{11}\right\|_F + \sqrt{\sum_{k=1}^{n}(\beta\hat{r}_k - \hat{r}_k)^2} + \sqrt{\sum_{k=1}^{n}(\hat{r}_k - t_k)^2}\\ &\le \left\|\Delta\bar{B}_{11}\right\|_F + (\beta-1)\sqrt{\sum_{k=1}^{n}\hat{r}_k^2} + 4C_4n^2\mathbf{u}\sqrt{\sum_{k=1}^{n}1^2}\\ &\le C_1\tilde{\gamma}_{n^{5/2}} + C_1\tilde{\gamma}_{n^{5/2}} + 4C_4n^{5/2}\mathbf{u} \le C_1\tilde{\gamma}_{n^{5/2}}.\end{aligned}$$

The bound on $\left\|\Delta B_{21}\right\|_F$ is similar. ∎

4.9. Singular vectors. To complete the two-by-one CSD, singular vectors are computed.

LEMMA 17. *The singular vectors of $\tilde{A}_{11}$ and $\tilde{A}_{21}$ from the previous lemma are computed to high relative accuracy:*

$$\begin{aligned}\left|\tilde{U}_1(i,j) - \mathrm{fl}(\tilde{U}_1(i,j))\right| / \left|\tilde{U}_1(i,j)\right| &\le \tilde{\gamma}_n,\\ \left|\tilde{U}_2(i,j) - \mathrm{fl}(\tilde{U}_2(i,j))\right| / \left|\tilde{U}_2(i,j)\right| &\le \tilde{\gamma}_n,\\ \left|\tilde{V}_1(i,j) - \mathrm{fl}(\tilde{V}_1(i,j))\right| / \left|\tilde{V}_1(i,j)\right| &\le \tilde{\gamma}_n.\end{aligned}$$

As a consequence, $\|\tilde{U}_1 - \mathrm{fl}(\tilde{U}_1)\|_F$, $\|\tilde{U}_2 - \mathrm{fl}(\tilde{U}_2)\|_F$, and $\|\tilde{V}_1 - \mathrm{fl}(\tilde{V}_1)\|_F$ are bounded by $\tilde{\gamma}_{n^2}$.

Proof. Each formula in Theorem 5 involves $O(1)$ arithmetic operations. Every factor is known exactly or can be computed with relative error below $\tilde{\gamma}_n$ with the formulas in Lemma 14. Hence, the combined relative error is at most $\tilde{\gamma}_n$. ∎

4.10. Summary of the bidiagonal CSD. The main stability result can now be proved.

Proof of Theorem 6. Let $\bar{U}_1$, $\bar{U}_2$, and $\bar{V}_1$ be the orthogonal transformations from Lemma 11. By Lemma 16 there are backward errors ΔB_{11}, ΔB_{21} for which

$$\begin{bmatrix}\bar{U}_1 & \\ & \bar{U}_2\end{bmatrix}^T \begin{bmatrix}B_{11} + \Delta B_{11}\\ B_{21} + \Delta B_{21}\end{bmatrix} \bar{V}_1 = \begin{bmatrix}\tilde{A}_{11}\\ \tilde{A}_{21}\end{bmatrix} \tag{20}$$

and $\left\|\Delta B_{11}\right\|_F \leq C_1\tilde{\gamma}_{n^{5/2}}$, $\left\|\Delta B_{21}\right\|_F \leq C_1\tilde{\gamma}_{n^{5/2}}$. The blocks $\tilde{A}_{11}$ and $\tilde{A}_{21}$ are in broken-arrow form and are diagonalized by $\tilde{U}_1$, $\tilde{U}_2$, and $\tilde{V}_1$ of the previous lemma. Defining $U_1 = \bar{U}_1\tilde{U}_1$, $U_2 = \bar{U}_2\tilde{U}_2$, and $V_1 = \bar{V}_1\tilde{V}_1$, we have found

$$\begin{bmatrix} U_1 & \\ & U_2 \end{bmatrix}^T \begin{bmatrix} B_{11}+\Delta B_{11} \\ B_{21}+\Delta B_{21} \end{bmatrix} V_1 = \begin{bmatrix} C \\ S \end{bmatrix} = \begin{bmatrix} \operatorname{diag}(\cos\hat{\theta}_1,\ldots,\cos\hat{\theta}_n) \\ \operatorname{diag}(\sin\hat{\theta}_1,\ldots,\sin\hat{\theta}_n) \end{bmatrix}. \tag{21}$$

Considering Lemmas 9, 11, and 17, the orthogonal matrices U_1, U_2, and V_1 can be approximated numerically by $\hat{U}_1$, $\hat{U}_2$, and $\hat{V}_1$ for which $\|U_1 - \hat{U}_1\|_F$, $\|U_2 - \hat{U}_2\|_F$, and $\|V_1 - \hat{V}_1\|_F$ are bounded by $\tilde{\gamma}_{n^2}$.

So far, $\hat{\theta}_1,\ldots,\hat{\theta}_n$ are stored implicitly in terms of offsets: $\hat{\theta}_i = \phi_{\sigma(i)} + \hat{\delta}_i$. If the $\hat{\theta}_1,\ldots,\hat{\theta}_n$ are desired explicitly, they can be computed accurately using Lemma 14. The resulting matrix of errors has Frobenius norm bounded by $\tilde{\gamma}_{n^{3/2}}$ and can be absorbed into ΔB_{11}, ΔB_{21}. ∎

4.11. Two-by-two CSD. The two-by-two CSD requires one additional step.

THEOREM 18. *The two-by-two bidiagonal CSD is computed backward stably. Suppose B_{11} and B_{21} are n-by-n upper-bidiagonal and B_{12} and B_{22} are n-by-n lower-bidiagonal, and let*

$$C_5 = \max\left(1, \left\| I - \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}^T \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \right\|_2 \Big/ (n^{5/2}\mathbf{u})\right).$$

Run the divide-and-conquer algorithm on $\begin{bmatrix} B_{11} \\ B_{21} \end{bmatrix}$ and then compute $\hat{V}_2 = \mathrm{fl}(-B_{12}^T\hat{U}_1 S + B_{22}^T\hat{U}_2 C)$. In addition to the properties in Theorem 6, this gives

$$\begin{bmatrix} U_1 & \\ & U_2 \end{bmatrix}^T \begin{bmatrix} B_{11}+\Delta B_{11} & B_{12}+\Delta B_{12} \\ B_{21}+\Delta B_{21} & B_{22}+\Delta B_{22} \end{bmatrix} \begin{bmatrix} V_1 & \\ & V_2 \end{bmatrix} = \begin{bmatrix} C & -S \\ S & C \end{bmatrix} \tag{22}$$

with $\|\Delta B_{21}\|_2 \leq C_5\tilde{\gamma}_{n^{5/2}}$, $\|\Delta B_{22}\|_2 \leq C_5\tilde{\gamma}_{n^{5/2}}$, and $\left\|V_2 - \hat{V}_2\right\|_F \leq C_5\tilde{\gamma}_{n^3}$.

The proof assumes that C_5 and n are small enough, and the arithmetic precise enough, that $C_5\gamma_k \leq 1/2$ for every γ_k encountered.

Proof. Theorem 6 provides

$$\begin{bmatrix} U_1 & \\ & U_2 \end{bmatrix}^T \begin{bmatrix} B_{11}+\Delta B_{11} \\ B_{21}+\Delta B_{21} \end{bmatrix} V_1 = \begin{bmatrix} C \\ S \end{bmatrix}$$

with $\left\|\Delta B_{11}\right\|_F$ and $\left\|\Delta B_{21}\right\|_F$ bounded by $C_1\tilde{\gamma}_{n^{5/2}} \leq C_5\tilde{\gamma}_{n^{5/2}}$.

Let $-SU_1^TB_{12} + CU_2^TB_{22} = YV_2^T$ be a polar decomposition with Y symmetric positive semidefinite and V_2 orthogonal. We have

$$\begin{bmatrix} C & -S \\ S & C \end{bmatrix}^T \begin{bmatrix} U_1 & \\ & U_2 \end{bmatrix}^T \begin{bmatrix} B_{11}+\Delta B_{11} & B_{12} \\ B_{21}+\Delta B_{21} & B_{22} \end{bmatrix} \begin{bmatrix} V_1 & \\ & V_2 \end{bmatrix}$$
$$= \begin{bmatrix} I & C(U_1^TB_{12}V_2) + S(U_2^TB_{22}V_2) \\ 0 & -S(U_1^TB_{12}V_2) + C(U_2^TB_{22}V_2) \end{bmatrix} = \begin{bmatrix} I & E_{12} \\ 0 & Y \end{bmatrix}$$

for $E_{12} = CU_1^TB_{12}V_2 + SU_2^TB_{22}V_2$. By Lemma 7 and the definition of C_5,

$$\left\| I - \begin{bmatrix} I & E_{12} \\ 0 & Y \end{bmatrix}^T \begin{bmatrix} I & E_{12} \\ 0 & Y \end{bmatrix} \right\|_2 \leq C_5\tilde{\gamma}_{n^{5/2}},$$

so $\|E_{12}\|_2 \le C_5\tilde{\gamma}_{n^{5/2}}$. In addition,

$$\left\| I - \begin{bmatrix} I & E_{12} \\ 0 & Y \end{bmatrix} \begin{bmatrix} I & E_{12} \\ 0 & Y \end{bmatrix}^T \right\|_2 \le C_5\tilde{\gamma}_{n^{5/2}},$$

so $\|I - Y\|_2 \le \left\|I - Y^2\right\|_2 = \left\|I - YY^T\right\|_2 \le C_5\tilde{\gamma}_{n^{5/2}}$, since Y is symmetric positive semidefinite.

Define

$$\begin{bmatrix} \Delta B_{12} \\ \Delta B_{22} \end{bmatrix} = \begin{bmatrix} -B_{12} - U_1 S V_2^T \\ -B_{22} + U_2 C V_2^T \end{bmatrix}$$

so that

$$\begin{bmatrix} U_1 & \\ & U_2 \end{bmatrix}^T \begin{bmatrix} B_{11} + \Delta B_{11} & B_{12} + \Delta B_{12} \\ B_{21} + \Delta B_{21} & B_{22} + \Delta B_{22} \end{bmatrix} \begin{bmatrix} V_1 & \\ & V_2 \end{bmatrix} = \begin{bmatrix} C & -S \\ S & C \end{bmatrix}.$$

Alternatively, we can express the new backward errors as

$$\begin{aligned} \begin{bmatrix} \Delta B_{12} \\ \Delta B_{22} \end{bmatrix} &= \begin{bmatrix} U_1 & \\ & U_2 \end{bmatrix} \begin{bmatrix} C & -S \\ S & C \end{bmatrix} \begin{bmatrix} C & -S \\ S & C \end{bmatrix}^T \begin{bmatrix} -U_1^T B_{12} V_2 - S \\ -U_2^T B_{22} V_2 + C \end{bmatrix} V_2^T \\ &= \begin{bmatrix} U_1 & \\ & U_2 \end{bmatrix} \begin{bmatrix} C & -S \\ S & C \end{bmatrix} \begin{bmatrix} -E_{12} \\ I - Y \end{bmatrix} V_2^T, \end{aligned}$$

so

$$\left\| \begin{bmatrix} \Delta B_{12} \\ \Delta B_{22} \end{bmatrix} \right\|_2 \le \sqrt{\|E_{12}\|_2^2 + \|I - Y\|_2^2} \le \sqrt{C_5^2\tilde{\gamma}_{n^{5/2}}^2 + C_5^2\tilde{\gamma}_{n^{5/2}}^2} \le C_5\tilde{\gamma}_{n^{5/2}}.$$

Finally, consider $\|V_2 - \hat{V}_2\|_F$. From

$$\begin{aligned} \left\|(-B_{12}^T U_1 S + B_{22}^T U_2 C)Y - (-B_{12}^T U_1 S + B_{22}^T U_2 C)\right\|_F &= \left\|(-B_{12}^T U_1 S + B_{22}^T U_2 C)(I - Y)\right\|_F \\ &\le \left(\left\|B_{12}^T U_1 S\right\|_F + \left\|B_{22}^T U_2 C\right\|_F\right) \|I - Y\|_2 \le C_5\tilde{\gamma}_{n^3} \end{aligned}$$

and

$$\left\|(-B_{12}^T U_1 S + B_{22}^T U_2 C) - \mathrm{fl}(-B_{12}^T \hat{U}_1 S + B_{22}^T \hat{U}_2 C)\right\|_F \le \tilde{\gamma}_{n^2},$$

it follows that

$$\left\|V_2 - \hat{V}_2\right\|_F \le C_5\tilde{\gamma}_{n^3}. \qquad \square$$

5. Numerical experiments. The numerical experiments in this section compare the following algorithms:

- D&C: the divide-and-conquer algorithm of this article with simultaneous QR iteration as the base case for $n \le 25$.
- QR: simultaneous QR iteration, specified in [32] and implemented as LAPACK's DBBCSD.
- Kog: the Kogbetliantz algorithm developed in [4, 24] and implemented as LAPACK's DGGSVD. The upper-triangular factor from the GSVD is discarded to obtain a CSD.

There are four classes of test matrices:

- Haar: a random matrix from Haar measure that has been reduced to bidiagonal-block form.
- Haar + noise: a matrix from the Haar class with uncorrelated Gaussian noise added to its bidiagonal bands.
- Clustered: a random matrix designed to have clustered principal angles $\theta_1, \ldots, \theta_n$ generated by the following procedure:
  $x = \text{rand}(n+1, 1)$
  $\delta = 10^{-18x}$
  $\theta = \text{cumsum}(\delta)$
  $\theta = (\pi/2)\theta_{1:n}/\theta_{n+1}$
  Diagonal matrices $C = \text{diag}(\cos(\theta))$ and $S = \text{diag}(\sin(\theta))$ are constructed, pre- and postmultiplied by Haar-distributed random matrices, and then simultaneously bidiagonalized.
- Clustered + noise: a matrix from the previous test class with uncorrelated Gaussian noise added to its bidiagonal bands.

All three algorithms perform stably, as demonstrated in Tables 1 and 2. There are forty test matrices, one per row. When the input matrix is orthogonal to machine precision, as in the first and third test classes, the residual is on the order of unit roundoff. When the input matrix is farther from orthogonality, it must be pushed onto the Stiefel manifold, and the residual reflects this—the residuals for the second and fourth test classes are small relative to

$$\varepsilon := \left\| I - \begin{bmatrix} B_{11} \\ B_{21} \end{bmatrix}^T \begin{bmatrix} B_{11} \\ B_{21} \end{bmatrix} \right\|_2,$$

which is about 10^{-10}. A perusal of the data suggests that divide-and-conquer and QR iteration provide somewhat higher accuracy than the Kogbetliantz algorithm, especially for larger matrices, with divide-and-conquer perhaps enjoying a small edge over QR iteration.

6. Conclusions. We have developed a new divide-and-conquer algorithm for the bidiagonal CSD and proved it numerically stable. The key steps in ensuring stability are the following:

- After reduction to broken-arrow form, pairs of matrix entries are expressed in polar coordinates to ameliorate nonorthogonality.
- The deflation procedure works directly with polar factors rather than the entries themselves.
- Before performing the rank-one update, the singular values are expressed in terms of angle gaps: $\cos(\sum \Delta\phi_j)$ and $\sin(\sum \Delta\phi_j)$. This supports higher-accuracy floating-point computations.
- In performing the rank-one update, a single secular equation reveals underlying angles θ_i, $i = 1, \ldots, n$. The singular values $\cos(\theta_i)$ and $\sin(\theta_i)$ are then computed easily.
- A new inverse eigenvalue problem is defined and solved. This allows the broken-arrow blocks of (2) to be reconstructed from their singular values. The reconstructed matrix is exactly orthogonal because of its representation, and its singular vectors are well conditioned.
- The right singular vectors are computed from underlying polar factors. The formulas for the top and bottom halves of the matrix are identical,

TABLE 1
Residuals.

Test class	n	Input: $\varepsilon := \left\|I - X^T X\right\|_2$, $X = \begin{bmatrix} B_{11} \\ B_{21} \end{bmatrix}$	Residual: $\left\| \begin{bmatrix} \hat{U}_1 & \\ & \hat{U}_2 \end{bmatrix}^T X \hat{V}_1 - \begin{bmatrix} \hat{C} \\ \hat{S} \end{bmatrix} \right\|_2 \Big/ \varepsilon$		
			D&C	QR	Kog
Haar	30	2.5×10^{-16}	21	22	23
	42	4.5×10^{-16}	15	24	15
	60	4.4×10^{-16}	10	14	27
	85	3.4×10^{-16}	16	33	41
	120	2.8×10^{-16}	32	45	96
	170	4.5×10^{-16}	25	27	57
	240	2.5×10^{-16}	41	54	154
	339	4.5×10^{-16}	16	59	107
	480	4.5×10^{-16}	24	55	191
	679	4.6×10^{-16}	26	76	239
Haar + noise	30	4.0×10^{-10}	0.65	0.65	0.83
	42	4.8×10^{-10}	0.56	0.56	0.67
	60	6.5×10^{-10}	0.51	0.51	0.70
	85	4.9×10^{-10}	0.67	0.67	0.73
	120	5.9×10^{-10}	0.54	0.52	0.70
	170	6.3×10^{-10}	0.55	0.55	0.73
	240	6.6×10^{-10}	0.53	0.53	0.70
	339	6.6×10^{-10}	0.54	0.54	0.69
	480	7.8×10^{-10}	0.52	0.52	0.71
	679	8.4×10^{-10}	0.51	0.51	0.62
Clustered	30	2.4×10^{-16}	88	43	20
	42	2.5×10^{-16}	37	33	27
	60	3.3×10^{-16}	37	24	36
	85	2.8×10^{-16}	51	37	73
	120	3.4×10^{-16}	42	32	108
	170	2.8×10^{-16}	65	53	86
	240	2.8×10^{-16}	79	54	338
	339	3.3×10^{-16}	83	37	416
	480	4.5×10^{-16}	76	56	334
	679	4.5×10^{-16}	94	76	575
Clustered + noise	30	5.0×10^{-10}	0.62	0.62	0.71
	42	4.4×10^{-10}	0.64	0.64	0.63
	60	5.2×10^{-10}	0.63	0.58	0.68
	85	6.4×10^{-10}	0.53	0.53	0.69
	120	5.1×10^{-10}	0.61	0.61	0.70
	170	6.8×10^{-10}	0.72	0.72	0.67
	240	5.9×10^{-10}	0.67	0.67	0.64
	339	6.0×10^{-10}	0.62	0.62	0.82
	480	6.5×10^{-10}	0.72	0.70	0.73
	679	6.7×10^{-10}	0.55	0.55	0.69

TABLE 2
Orthogonality.

Test class	n	Orthogonality								
		$\|I - U_1^T U_1\|_2 / \mathbf{u}$			$\|I - U_2^T U_2\|_2 / \mathbf{u}$			$\|I - V_1^T V_1\|_2 / \mathbf{u}$		
		D&C	QR	Kog	D&C	QR	Kog	D&C	QR	Kog
Haar	30	48	25	79	27	23	82	25	28	31
	42	38	31	105	34	29	119	38	34	45
	60	42	35	179	37	35	148	33	41	57
	85	42	52	241	36	38	218	42	63	61
	120	47	54	296	49	49	329	46	58	78
	170	55	71	423	52	59	436	48	64	94
	240	61	70	591	54	85	607	53	87	113
	339	63	89	855	62	119	905	70	89	137
	480	84	112	1469	84	100	1636	77	108	186
	679	89	146	1904	93	109	1944	86	130	232
Haar + noise	30	27	28	85	26	22	76	24	31	29
	42	40	32	101	38	35	109	29	41	42
	60	32	50	162	46	35	179	27	42	47
	85	42	45	217	39	38	208	38	42	60
	120	43	70	304	41	66	330	37	58	77
	170	52	62	434	57	66	467	57	62	101
	240	57	66	623	55	74	641	59	84	134
	339	62	80	870	63	97	942	64	109	160
	480	79	113	1286	84	99	1264	82	115	186
	679	92	117	1996	90	135	2051	101	167	296
Clustered	30	24	16	71	34	20	53	25	18	24
	42	30	30	108	47	17	103	35	22	35
	60	22	55	187	27	54	191	29	41	32
	85	37	43	360	36	46	369	45	55	66
	120	48	57	632	36	61	612	51	62	79
	170	50	67	428	43	63	431	75	67	59
	240	50	101	1774	46	68	1715	42	76	146
	339	71	86	2348	58	112	2383	81	101	228
	480	75	125	2659	63	100	2749	63	115	229
	679	103	154	4576	71	131	4636	77	134	336
Clustered + noise	30	24	23	51	21	18	57	18	25	18
	42	37	36	81	26	20	78	26	35	26
	60	26	29	149	30	31	160	31	32	35
	85	34	46	277	36	55	302	33	45	66
	120	39	56	424	42	76	477	41	67	69
	170	40	69	1012	43	66	984	40	81	108
	240	46	78	1330	41	119	1378	42	119	148
	339	60	83	2071	57	141	1962	51	95	177
	480	66	107	4486	61	107	4428	63	128	349
	679	67	141	6244	136	145	6284	73	167	448

enabling perfect consistency even when the singular vectors were originally ill-conditioned.

Fortunately, none of these steps require a significant number of floating-point operations or memory accesses that are not already present in the SVD algorithm. The computation of the $2n$-by-n CSD may indeed be faster than two separate n-by-n divide-and-conquer SVDs because the singular vectors in V_1 need be computed only once.

Acknowledgments. Thanks are extended to James Demmel for suggesting the problem, to Yuji Nakatsukasa for suggesting the rescaling of A and B at the end of section 2, and to the anonymous referees for their much appreciated comments.

REFERENCES

[1] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen, *LAPACK Users' Guide*, 3rd ed., SIAM, Philadelphia, 1999.

[2] P. Arbenz and G. H. Golub, *On the spectral decomposition of Hermitian matrices modified by low rank perturbations with applications*, SIAM J. Matrix Anal. Appl., 9 (1988), pp. 40–58.

[3] Z. Bai, *The CSD, GSVD, their Applications and Computations*, Technical report 958, IMA Preprint Series, Institute for Mathematics and Its Applications, University of Minnesota, Minneapolis, MN, 1992.

[4] Z. Bai and J. W. Demmel, *Computing the generalized singular value decomposition*, SIAM J. Sci. Comput., 14 (1993), pp. 1464–1486.

[5] A. Bjorck and G. H. Golub, *Numerical methods for computing angles between linear subspaces*, Math. Comp., 27 (1973), pp. 579–594.

[6] M. J. Cantero, L. Moral, and L. Velázquez, *Five-diagonal matrices and zeros of orthogonal polynomials on the unit circle*, Linear Algebra Appl., 362 (2003), pp. 29–56.

[7] J. J. M. Cuppen, *A divide and conquer method for the symmetric tridiagonal eigenproblem*, Numer. Math., 36 (1980), pp. 177–195.

[8] C. Davis and W. M. Kahan, *Some new bounds on perturbation of subspaces*, Bull. Amer. Math. Soc., 75 (1969), pp. 863–868.

[9] C. Davis and W. M. Kahan, *The rotation of eigenvectors by a perturbation.* III, SIAM J. Numer. Anal., 7 (1970), pp. 1–46.

[10] J. J. Dongarra and D. C. Sorensen, *A fully parallel algorithm for the symmetric eigenvalue problem*, SIAM J. Sci. Stat. Comput., 8 (1987), pp. S139–S154.

[11] Z. Drmač, *A tangent algorithm for computing the generalized singular value decomposition*, SIAM J. Numer. Anal., 35 (1998), pp. 1804–1832.

[12] Z. Drmač and E. R. Jessup, *On accurate quotient singular value computation in floating-point arithmetic*, SIAM J. Matrix Anal. Appl., 22 (2000), pp. 853–873.

[13] A. Edelman, T. A. Arias, and S. T. Smith, *The geometry of algorithms with orthogonality constraints*, SIAM J. Matrix Anal. Appl., 20 (1998), pp. 303–353.

[14] A. Edelman and B. D. Sutton, *The beta-Jacobi matrix model, the CS decomposition, and generalized singular value problems*, Found. Comput. Math., 8 (2008), pp. 259–285.

[15] M. Gu and S. C. Eisenstat, *A divide-and-conquer algorithm for the bidiagonal SVD*, SIAM J. Matrix Anal. Appl., 16 (1995), pp. 79–92.

[16] M. Gu and S. C. Eisenstat, *A divide-and-conquer algorithm for the symmetric tridiagonal eigenproblem*, SIAM J. Matrix Anal. Appl., 16 (1995), pp. 172–191.

[17] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*, 2nd ed., SIAM, Philadelphia, 2002.

[18] H. Hotelling, *Relations between two sets of variates*, Biometrika, 28 (1936), pp. 321–377.

[19] E. R. Jessup and D. C. Sorensen, *A divide and conquer algorithm for computing the singular value decomposition*, in Parallel Processing for Scientific Computing (Los Angeles, CA, 1987), SIAM, Philadelphia, 1989, pp. 61–66.

[20] C. Jordan, *Essai sur la géométrie á n dimensions*, Bull. Soc. Math. France, 3 (1875), pp. 103–174.

[21] R. Killip and I. Nenciu, *Matrix models for circular ensembles*, Int. Math. Res. Not. 50, (2004), pp. 2665–2701.

[22] K. Löwner, *Über monotone Matrixfunktionen*, Math. Z., 38 (1934), pp. 177–216.

[23] M. Möttönen, J. J. Vartiainen, V. Bergholm, and M. M. Salomaa, *Quantum circuits for general multiqubit gates*, Phys. Rev. Lett., 93 (2004), 130502.

[24] C. C. Paige, *Computing the generalized singular value decomposition*, SIAM J. Sci. Statist. Comput., 7 (1986), pp. 1126–1146.

[25] C. C. Paige and M. A. Saunders, *Towards a generalized singular value decomposition*, SIAM J. Numer. Anal., 18 (1981), pp. 398–405.

[26] C. C. Paige and M. Wei, *History and generality of the CS decomposition*, Linear Algebra Appl., 208/209 (1994), pp. 303–326.

[27] V. V. Shende, S. S. Bullock, and I. L. Markov, *Synthesis of quantum logic circuits*, in Proceedings of the 2005 Conference on Asia South Pacific Design Automation, Shanghai, China, 2005, ACM, New York, pp. 272–275.

[28] B. Simon, *CMV matrices: five years after*, J. Comput. Appl. Math., 208 (2007), pp. 120–154.
[29] G. W. Stewart, *On the perturbation of pseudo-inverses, projections and linear least squares problems*, SIAM Rev., 19 (1977), pp. 634–662.
[30] G. W. Stewart, *Computing the CS decomposition of a partitioned orthonormal matrix*, Numer. Math., 40 (1982), pp. 297–306.
[31] B. D. Sutton, *The Stochastic Operator Approach to Random Matrix Theory*, Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, 2005.
[32] B. D. Sutton, *Computing the complete CS decomposition*, Numer. Algorithms, 50 (2009), pp. 33–65.
[33] B. D. Sutton, *Stable computation of the CS decomposition: Simultaneous bidiagonalization*, SIAM J. Matrix Anal. Appl., 33 (2012), pp. 1–21.
[34] R. R. Tucci, *A rudimentary quantum compiler (2nd ed.)*, http://arxiv.org/abs/quant-ph/9902062 1999.
[35] C. F. Van Loan, *Computing the CS and the generalized singular value decompositions*, Numer. Math., 46 (1985), pp. 479–491.
[36] C. F. Van Loan and J. Speiser, *Computation of the C-S decomposition, with application to signal processing*, in SPIE Conference on Advanced Algorithms and Architectures for Signal Processing, 1986, San Diego, CA, SPIE, Bellingham, WA, 1987, pp. 71–78.
[37] C. F. Van Loan, *Generalizing the singular value decomposition*, SIAM J. Numer. Anal., 13 (1976), pp. 76–83.
[38] S. Wang and S. Zhao, *An algorithm for $Ax = \lambda Bx$ with symmetric and positive-definite A and B*, SIAM J. Matrix Anal. Appl., 12 (1991), pp. 654–660.
[39] D. S. Watkins, *Some perspectives on the eigenvalue problem*, SIAM Rev., 35 (1993), pp. 430–471.