Rounding error and perturbation bounds for the symplectic QR factorization

Sanja Singer a,*, Saša Singer b

aFaculty of Mechanical Engineering and Naval Architecture, University of Zagreb, Ivana Lučića 5, 10000 Zagreb, Croatia
bDepartment of Mathematics, University of Zagreb, P.O. Box 335, 10002 Zagreb, Croatia

Received 28 March 2001; accepted 11 December 2001
Submitted by J.L. Barlow

Abstract

To compute the eigenvalues of a skew-symmetric matrix $A$, we can use a one-sided Jacobi-like algorithm to enhance accuracy. This algorithm begins by a suitable Cholesky-like factorization of $A$, $A = G^T J G$. In some applications, $A$ is given implicitly in that form and its natural Cholesky-like factor $G$ is immediately available, but “tall”, i.e., not of full row rank. This factor $G$ is unsuitable for the Jacobi-like process. To avoid explicit computation of $A$, and possible loss of accuracy, the factor has to be preprocessed by a QR-like factorization. In this paper we present the symplectic QR algorithm to achieve such a factorization, together with the corresponding rounding error and perturbation bounds. These bounds fit well into the relative perturbation theory for skew-symmetric matrices given in factorized form.

Keywords: Symplectic QR factorization; Rounding error bounds; Perturbation bounds; Skew-symmetric eigenproblem

1. Introduction

Accurate floating point computation of eigensystems of matrices becomes an increasingly important subject in numerical linear algebra and its applications. Most of the effort has been concentrated on symmetric or Hermitian matrices (both positive
definite and indefinite). Skew-symmetric or skew-Hermitian matrices are in many ways analogous, except that their eigenvalues lie on the imaginary, instead of the real axis. With a little more effort, the known results can be extended to cover the “skewed” case.

The one-sided Jacobi-like algorithm for accurate computation of eigenvalues of nonsingular skew-symmetric matrix $A$ has been constructed by Pietzsch [6] as an analogue of the symmetric algorithm by Slapničar [9]. The skew-symmetric algorithm also begins with a factorization of $A$, but here

$$A = G^T J G, \quad (1.1)$$

where $G$ has full row rank, and $J$ is block diagonal elementary skew-symmetric matrix of order $2m$

$$J = \text{diag}(J_0, J_0, \ldots, J_0), \quad (1.2)$$

where $J_0$ are elementary (or orthogonal) skew-symmetric matrices of order 2

$$J_0 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \quad (1.3)$$

Note that nonsingularity of $A$ implies that it is of even order, say $2n$. Once we have this factorization of $A$, the Pietzsch algorithm operates on the full column rank matrix $G^T$.

In some problems, we are given the matrices $G$ and $J$ in (1.1), rather than $A$. For example, $A = B^T C - C^T B$ is such a matrix with implicitly given factors $\hat{J}$ and $\hat{G}$ in terms of $B$ and $C$, i.e.,

$$A = [B^T \ C^T] \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix} \begin{bmatrix} B \\ C \end{bmatrix} := \hat{G}^T \hat{J} \hat{G}. \quad (1.4)$$

It is obvious that $\hat{G}^T$ is not of full column rank and, therefore, cannot be used as a starting matrix for the Pietzsch algorithm. Our goal is to construct a new factorization $A = \hat{G}^T J \hat{G}$ with full column rank $G^T$, without explicit computation of $A$. This can be done by the so-called symplectic QR factorization of the given pair $\hat{G}, \hat{J}$, where we use symplectic (or $J$-orthogonal) transformations to obtain a new pair $G, J$ with triangular $G$.

To justify the name and as a reminder, a matrix $U_S \in \mathbb{C}^{2m \times 2m}$ is symplectic if $U_S^* J U_S = J$, where $J$ is given by (1.2) and (1.3), i.e., $U_S$ is a unitary matrix with respect to the “symplectic scalar product” matrix $J$. To be precise, the usual definition of a symplectic matrix (see, e.g., [3]) uses $\hat{J}$ from (1.4), instead of $J$. Note that $J$ and $\hat{J}$ are permutationally equivalent, which is sufficient for our purpose.

If one starts from $A$ as input data, the factorization (1.1) can be obtained by using the Bunch algorithm [2], which is, in fact, a Cholesky-like factorization for skew-symmetric matrices, and also yields a triangular factor $G$.

Cholesky-like factorization of skew-symmetric matrices has recently been rediscovered by Benner et al. [1], with motivation from eigenvalue problems with Hamiltonian structure.
In the symmetric case, if \( A = G^T G \) is positive definite, then the QR factorization of \( G \) gives the triangular Cholesky factor of \( A \). For indefinite matrices, the situation is more complicated, but a similar relationship holds between the indefinite QR and the symmetric indefinite factorization of \( A \) (see [7]), although equality is not fully preserved. As we expect, the same is true in the skew-symmetric case: the symplectic QR factorization of \( G \) can produce the Bunch triangular factor of \( A \).

Bunch constructed his algorithm in order to factor skew-symmetric matrices in real arithmetic. The symplectic QR does the same. In view of this, the skew-Hermitian case can almost be regarded as an uninteresting one. Namely, if \( A \) is skew-Hermitian, then \( B = iA \) is Hermitian, and \( B = iG^T JG \) can be found by Hermitian indefinite factorization or implicitly solved by a simple modification of the indefinite QR factorization [7]. In one way or the other, we cannot avoid complex arithmetic without doubling the dimension of the problem.

In this paper, we present and analyze the Givens-like algorithm for the symplectic QR factorization. This algorithm requires plane trigonometric and block symplectic rotations. So, the paper is, naturally, in many ways analogous to the previous one on the indefinite QR [8].

The rest of the paper is organized as follows. In Section 2 we construct the required elementary symplectic matrices and give a description of the algorithm. Section 3 contains the floating point error analysis of the algorithm. In Section 4 we derive the perturbation bounds for the symplectic QR factorization. A combination of these results gives the relative perturbation bounds for the eigenvalues of \( A \). Finally, we give some examples illustrating the algorithm and its accuracy.

2. Symplectic QR factorization

Let \( G \in \mathbb{R}^{2m \times 2n} \), \( m \geq n \), be a given "tall" matrix, and let \( J \in \mathbb{R}^{2m \times 2m} \) be given by (1.2) and (1.3). A symplectic QR factorization of \( G \) is a factorization

\[
G = P_1 Q R P_2^T = P_1 Q \begin{bmatrix} R_1 & 0 \\ 0 & R_2 \end{bmatrix} P_2^T, \quad Q^T J Q = J, \quad J = P_1^T J P_1,
\]

where \( Q \) is symplectic of order \( 2m \), \( R_1 \) is upper triangular of order \( 2n \), and \( P_1 \) and \( P_2 \) are permutation matrices corresponding to some row and column pivoting strategies, respectively.

The first goal is to construct an algorithm for this factorization. This will be a constructive proof of its existence (under some mild conditions). As will be seen, the row pivoting is not essential for this construction, so we can take \( P_1 = I \).

Symplectic matrices of order \( 2m \) form a multiplicative group. Therefore, we can use a sequence of elementary symplectic transformations to transform \( G \) into a triangular form. In the Givens-like algorithm, we use elementary matrices that resemble rotations, i.e., they are equal to identity matrix except for a block of small
order. To construct the algorithm, we need several types of these blocks of order 2 and 4.

2.1. Elementary symplectic and block symplectic matrices

Symplectic matrices of order 2 are characterized by the following lemma.

**Lemma 2.1.** A matrix $S$ of order 2 is symplectic (with $J = J_0$) if and only if $\det S = 1$.

This means that ordinary plane rotations are symplectic matrices in $\mathbb{R}^{2 \times 2}$. Elementary rotations $U_G(i, \ell)$ in the $(i, \ell)$ plane, which are equal to identity matrix $I_{2m}$, except that

$$ U_G(i, \ell)([i, \ell], [i, \ell]) = \begin{bmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{bmatrix} $$

are symplectic matrices (with respect to $J$) in $\mathbb{R}^{2m \times 2m}$, if and only if indices $i$ and $\ell$ refer to the same $J_0$ block in $J$, i.e., $i = 2k - 1$ and $\ell = 2k$, for $k = 1, \ldots, m$. Symplectic elementary rotations will be denoted by

$$ U_G(k) = U_G(2k - 1, 2k). $$

They can be used to annihilate elements, but only in (odd, even) planes corresponding to a single $J_0$ block in $J$.

To work across two different $J_0$ blocks of $J$, we need two types of symplectic matrices of order 4, with $J = \text{diag}(J_0, J_0)$.

**Lemma 2.2.** Matrices $S_1$ and $S_2$ of order 4 defined by

$$ S_1 = \begin{bmatrix} \cos \phi I_2 & \sin \phi I_2 \\ -\sin \phi I_2 & \cos \phi I_2 \end{bmatrix}, \quad S_2 = \begin{bmatrix} I_2 & -X^* \\ -X & I_2 \end{bmatrix}, \quad X = \begin{bmatrix} 0 & a \\ 0 & 0 \end{bmatrix} $$

are symplectic with $J = \text{diag}(J_0, J_0)$ for any choice of $\phi$ and $a$, respectively.

Their inverses, which will be used for annihilation, are given by

$$ S_1^{-1} = \begin{bmatrix} \cos \phi I_2 & -\sin \phi I_2 \\ \sin \phi I_2 & \cos \phi I_2 \end{bmatrix}, \quad S_2^{-1} = \begin{bmatrix} I_2 & X \\ X & I_2 \end{bmatrix}. $$

The corresponding elementary block symplectic matrices $U_{S1}(i, k)$ and $U_{S2}(k)$ of order $2m$ are equal to identity matrix $I_{2m}$, except that

$$ U_{S1}(i, k)([2i - 1, 2i, 2k - 1, 2k], [2i - 1, 2i, 2k - 1, 2k]) = S_1, $$
$$ U_{S2}(k)(2k - 1 : 2k + 2, 2k - 1 : 2k + 2) = S_2. $$
With a suitable choice of $\varphi$ and $\alpha$, premultiplication by $U_{S_1}^{-1}(i, k)$ or $U_{S_2}^{-1}(k)$ can also be used to annihilate elements in the working matrix.

Note that $U_{S_1}(i, k)$ can be viewed as a product of two independent interlaced plane rotations (in any order)

$$U_{S_1}(i, k) = U_G(2i - 1, 2k - 1)U_G(2i, 2k) = U_G(2i, 2k)U_G(2i - 1, 2k - 1),$$

with the same angle $\varphi$. This fact will be used in rounding error analysis.

### 2.2. Algorithm for the symplectic QR factorization

Now, we can prove the main factorization theorem.

**Theorem 2.1.** Let $G \in \mathbb{R}^{2m \times 2n}$, $m \geq n$, be a given matrix, and let $J \in \mathbb{R}^{2m \times 2m}$ be given by (1.2) and (1.3). If $A = G^T J G$ is nonsingular, then $G$ can be factorized as

$$G = Q R P_2^T = Q \begin{bmatrix} R_1 & 0 \\ 0 & \end{bmatrix} P_2^T, \quad Q^T J Q = J,$$

where $Q$ is symplectic of order $2m$, $R_1$ is upper triangular of order $2n$, and $P_2$ is a permutation matrix.

**Proof.** The proof is algorithmic. We will construct a sequence of elementary symplectic transformations that transforms $G$ into a triangular form.

The annihilation of elements in $G$ is performed in a sequence of $n$ stages. Initially, let $G^{(0)} = G$. In stage $k$, we annihilate the elements below the main diagonal in pair of columns $2k - 1$ and $2k$ of the working matrix $G^{(k-1)}$ to obtain $G^{(k)}$. We will describe only the first stage $G \rightarrow G^{(1)}$ of the reduction. The rest of the proof follows by induction.

To simplify the notation, suppose that the working matrix $G$ is stored in array $G$. Column $\ell$ of the working array $G$ is denoted by $g_\ell$. In each of the following steps, $G$ refers to the state of the working array before the corresponding transformation, while $G'$ refers to the state after the transformation.

**Step 1.** We choose the first pair $(i, \ell)$, $i \neq \ell$, of “pivot” columns $g_i$, $g_\ell$ such that $g_i^T J g_\ell \neq 0$. Nonsingularity of $A = G^T J G$ guarantees the existence of such a pair $(i, \ell)$. By permuting columns of $G$, this pair can be brought to the first two places in $G' = G P_2$. Additional column pivoting requirements can also be included in $P_2$.

Note that we have a freedom of choice which of these two columns will be the first one in the new $G$. Therefore, we can further achieve $g_1^T J g_2 > 0$, since permuting the first two columns of $G$ changes the sign of $g_1^T J g_2$. This step is not essential for the proof, but will be used later to obtain a Bunch-like factor $R_B$ from $R$. 

Step 2. Now, we can reduce the first column of $G$ to at most $m$ nonzero elements by using $m$ ordinary trigonometric rotations inside each of the $J_0$ blocks in $J$. This transformation can be written as

$$G' = \left( \prod_{k=1}^{m} U_{G}^{-1}(k) \right) G,$$

where the angle $\varphi_k$ for $U_{G}^{-1}(k)$ is chosen to annihilate the element $g_{2k,1}$. It is obvious that these $m$ rotations can be implemented independently (in any order), which is crucial for the error analysis.

Step 3. After the sequence of ordinary rotations, the first column of $G$ has the following form:

$$g_1 = \begin{bmatrix} g_{11} \\ 0 \\ g_{31} \\ 0 \\ \vdots \\ g_{2m-1,1} \\ 0 \end{bmatrix}.$$

Potential nonzero elements $g_{2k-1,1}$ correspond to different $J_0$ blocks of $J$. To annihilate all of them but one, say $g_{11}$, we have to use elementary block transformations $U_{S_1}^{-1}(i,k)$ (interlaced rotations). An obvious way to implement this step would be

$$G' = \left( \prod_{k=2}^{m} U_{S_1}^{-1}(1,k) \right) G,$$

as a sequence of $m - 1$ transformations. The same can be achieved in a sequence of only $\lceil \log_2 m \rceil$ steps, by applying $m/2$ independent transformations $U_{S_1}^{-1}(i,k)$ (with disjoint indices $i,k$) in the first step, then $m/4$ in the second step, etc. This completes the reduction of the first column.

The same result can be obtained in many different ways. Any sequence of $U_{G}^{-1}(k)$ and $U_{S_1}^{-1}(i,k)$ transformations producing the same final form would do. But the number of transformations acting on each element of $G$ may be different, resulting in a different floating point error bound. The “recursive halving” annihilation strategy gives the smallest overall error bound, and is just as easy to implement as the sequential one.

Step 4. The second column below the diagonal $g_2(3 : m)$ can be reduced in the similar way, by applying steps 2 and 3 on the matrix $G(3 : m, 2 : n)$.

Step 5. The first two columns of $G$ now have the following form:
\[ [g_1, g_2] = \begin{bmatrix} g_{11} & g_{12} \\ 0 & g_{22} \\ 0 & g_{32} \\ \vdots & \vdots \\ 0 & 0 \end{bmatrix}. \]

It remains to annihilate \( g_{32} \) without destroying the zero pattern of the first column. Let us denote
\[ G_1 = \begin{bmatrix} g_{11} & g_{12} \\ 0 & g_{22} \end{bmatrix}, \quad G_2 = \begin{bmatrix} 0 & g_{32} \\ 0 & 0 \end{bmatrix}. \]

First, note that both \( g_{11} \) and \( g_{22} \) are nonzero (otherwise, the corresponding 2 \( \times \) 2 block in \( A \) is singular). Therefore, \( G_1 \) is nonsingular.

By a suitable choice of \( a \), we can use the matrix \( U_{S_2}^{-1}(1) \) to annihilate \( G_2 U_{S_2}^{-1}(1) G = \begin{bmatrix} G'_1 \\ 0 \end{bmatrix} \).

To see that this can be done, the second row of this matrix equation yields \( XG_1 + G_2 = 0 \), or
\[ X = -G_2 G_1^{-1} \begin{bmatrix} 0 & a \\ 0 & 0 \end{bmatrix}, \]
with \( a = -g_{32}/g_{22} \).

The transformation \( G' = U_{S_2}^{-1}(1)G \) changes only the first and the third row of \( G \), and we have
\[ \begin{align*}
&g'_{1\ell} = g_{1\ell} + ag_{4\ell}, \\
&g'_{3\ell} = ag_{2\ell} + g_{3\ell}
\end{align*} \quad (2.2) \]

for \( \ell = 1, \ldots, 2n \). Especially, the zero pattern in the first column is not disturbed and we have
\[ \begin{bmatrix} g'_{11} & g'_{12} \\ g'_{21} & g'_{22} \\ g'_{31} & g'_{32} \\ g'_{41} & g'_{42} \end{bmatrix} = \begin{bmatrix} g_{11} & g_{12} \\ 0 & g_{22} \\ 0 & ag_{22} + g_{32} = 0 \\ 0 & 0 \end{bmatrix} \]

or \( G'_1 = G_1 \) and \( G'_2 = 0 \). The first two columns of \( G' \) are in the upper triangular form, which completes the first stage of the reduction, \( G^{(1)} = G' \).

The rest of the proof follows by induction. We repeat the reduction process on the \( (2m-2) \times (2n-2) \) matrix \( G(3 : 2m, 3 : 2n) \), until all columns are properly reduced. The final matrix is \( R = G^{(n)} \)
\[ R = \begin{bmatrix} R_1 \\ 0 \end{bmatrix}, \]
where \( R_1 \) is nonsingular upper triangular of order \( 2n \). \( \qed \)
If the matrix $A$ is given in form (1.4), the first step is to reorganize $\hat{J}$ and $\hat{G}$ by using the so-called perfect shuffle permutation $P$ to transform $\hat{J}$ into $J$

$$J = P \hat{J} P^T = P \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix} P^T, \quad G = P \hat{G}.$$  

2.3. Implicit Bunch factorization

A nonsingular skew-symmetric matrix $A$ of order $2n$ can be factorized as [2]

$$A = P_2 M^T D M P_2^T, \quad (2.3)$$

where $M$ is block unit upper triangular, and $D$ is skew-symmetric, block diagonal with $2 \times 2$ diagonal blocks $D_k$, $k = 1, \ldots, n$,

$$D_k = \begin{bmatrix} 0 & d_k \\ -d_k & 0 \end{bmatrix}.$$

This factorization is called the Bunch (explicit) factorization of $A$. Moreover, we may assume that $d_k > 0$. Then $D_k$ can be written as

$$D_k = \left(\sqrt{d_k} I_2 \right)^T J_0 \left(\sqrt{d_k} I_2 \right).$$

Let

$$R'_B := \text{diag} \left(\sqrt{d_1} I_2, \ldots, \sqrt{d_n} I_2\right) M.$$

From (2.3) we obtain the following factorization of $A$:

$$A = P_2 R'_B J R'_B P_2^T, \quad (2.4)$$

where $R'_B$ is upper triangular with scalar diagonal blocks $\sqrt{d_k} I_2$ and $J$ of order $2n$. The matrix $R'_B$ will be called the Bunch factor of $A$.

If $A$ is given implicitly by (1.1), a factor with the same structure as the Bunch factor can be obtained from the symplectic QR factorization of $G$.

**Theorem 2.2.** Let $R_1 = (r_{ij})$ be the triangular matrix from the symplectic QR factorization (2.1) of $G$. Matrices $Y_k$ which are equal to identity matrix $I_{2n}$, except the $2 \times 2$ block

$$Y_k([2k-1, 2k], [2k-1, 2k]) = \begin{bmatrix} s_k \sqrt{\frac{r_{2k, 2k}}{r_{2k-1, 2k-1}}} & \frac{r_{2k-1, 2k}}{\sqrt{r_{2k-1, 2k-1} r_{2k, 2k}}} \\ \frac{r_{2k-1, 2k}}{\sqrt{r_{2k-1, 2k-1} r_{2k, 2k}}} & s_k \sqrt{\frac{r_{2k-1, 2k-1}}{r_{2k, 2k}}} \end{bmatrix},$$

with $s_k = \text{sign}(r_{2k-1, 2k-1}) = \text{sign}(r_{2k, 2k})$, are symplectic (with $J$ of order $2n$) for $k = 1, \ldots, n$. The matrix
\[ R_B = Y R_1, \quad Y = \prod_{k=1}^{n} Y_k, \]  \hspace{1cm} (2.6)

where \( Y_k \) can be applied in any order, has the same structure as the Bunch factor of \( A = G^T J G \), and will be called the implicit Bunch factor of \( G \) (or \( A \)).

**Proof.** In the proof of Theorem 2.1, the order of the pivoting columns is chosen so that \( g_1^T J g_2 > 0 \) holds after step 1. The same is true for the working parts of all other pairs of pivoting columns

\[ (g_{2k-1}^T (2k - 1 : 2m))^T J_k g_{2k} (2k - 1 : 2m) > 0, \]

where \( J_k = \text{diag}(J_0, \ldots, J_0) \) is of order \( 2(m - k + 1), k = 1, \ldots, n \). Since we perform symplectic transformations of \( G \), the same remains valid in the final \( R \) or \( R_1 \), and we conclude that \( r_{2k-1,2k-1} r_{2k,2k} > 0 \). This proves that \( Y_k \) are correctly defined.

Since \( \det Y_k ([2k - 1, 2k], [2k - 1, 2k]) = 1 \), by Lemma 2.1, this matrix is symplectic (with respect to \( J_0 \)), and from its position in \( Y_k \) it follows that \( Y_k \) is symplectic. It is obvious that \( Y_1, \ldots, Y_n \) are mutually independent (the small blocks are disjoint), i.e., they all commute. Finally, \( R_B \) is block upper triangular and its diagonal blocks are

\[ (Y_k R_1) ([2k - 1, 2k], [2k - 1, 2k]) = \sqrt{r_{2k-1,2k-1} r_{2k,2k}}, \]

which means that \( R_B \) is triangular with scalar diagonal blocks. \( \square \)

If we choose the same column permutation \( P_2 \) in both factorizations (2.1) and (2.4), then \( R_B' = R_B \). The proof is analogous to the proof of uniqueness of Cholesky factorization of a positive definite matrix.

Of course, in floating point arithmetic, the computed values of \( R_B \) and \( R_B' \) may differ. To avoid explicit computation of \( A \), only the implicit Bunch factor \( R_B \) from (2.6) will be considered.

### 3. Rounding error analysis

The usual technique for the error analysis of the ordinary QR factorization of a matrix is to examine norms of errors in computed columns of \( R \), i.e., errors in diagonal elements of \( A \). But, in the skew-symmetric case, diagonal elements of \( A \) are 0. Let \( \tilde{R}_1 \) be the computed \( R_1 \) from Theorem 2.1 in floating point arithmetic. Then

\[ \text{diag}(\tilde{A}) = \text{diag}(\text{fl}(P_2 R_1^T J \tilde{R}_1 P_2^T)) \approx \text{diag}(0, \ldots, 0) \]

regardless of errors in \( \tilde{R}_1 \). From this, it is clear that “J-norms” of errors in columns of \( R_1 \), that is, in diagonal elements of \( A \), are useless as an error measure. The best we can do is to measure errors in some other, more meaningful norms, for example, in the Euclidean norm of computed columns of \( R_1 \).
We use the IEEE standard model of floating point arithmetic
\[ \text{fl}(a \circ b) = (a \circ b)(1 + \varepsilon_a), \quad |\varepsilon_a| \leq \varepsilon, \]
where \( \circ \) is any of the four elementary arithmetic operations and \( \varepsilon \) is the unit roundoff error. Furthermore, we assume that square roots can be computed with the same accuracy
\[ \text{fl}(\sqrt{a}) = \sqrt{a}(1 + \varepsilon_{\sqrt{a}}), \quad |\varepsilon_{\sqrt{a}}| \leq \varepsilon. \]

To obtain a rounding error bound for the symplectic QR factorization, we have to follow all the steps of the algorithm. The first two subsections contain the error analysis for the main types of transformations used in the algorithm. The overall backward error bounds for the symplectic QR and the implicit Bunch factorizations are presented in the last two subsections.

### 3.1. Error analysis for trigonometric rotations

Let \( U_G^{-1}(k) \) be the real trigonometric rotation which annihilates the element \( g_{2k,1} \) in the first column of \( G \). The requirement
\[
\begin{bmatrix}
\cos \varphi & -\sin \varphi \\
\sin \varphi & \cos \varphi
\end{bmatrix}
\begin{bmatrix}
g_{2k-1,1} \\
g_{2k,1}
\end{bmatrix}
= \begin{bmatrix}
g'_{2k-1,1} \\
0
\end{bmatrix}
\]
leads to
\[ \tan \varphi = -\frac{g_{2k,1}}{g_{2k-1,1}}. \]
Premultiplication by \( U_G^{-1}(k) \) changes only the rows \( 2k - 1 \) and \( 2k \) in the working matrix \( G \). For computed elements \( g'_{2k-1,\ell}, g'_{2k,\ell} \) in any column \( \ell \), we have
\[
\sqrt{\text{fl}(g'_{2k-1,\ell}) - g'_{2k-1,\ell})^2 + (\text{fl}(g'_{2k,\ell}) - g'_{2k,\ell})^2} 
\leq e_G \sqrt{g_{2k-1,\ell}^2 + g_{2k,\ell}^2}, \quad (3.1)
\]
with \( e_G = (3 + 2\sqrt{2})\varepsilon \approx 5.83\varepsilon \) (see [4]).

Note that all transformations in step 2 of the reduction are mutually independent. The computed column \( g'_{\ell} \) satisfies
\[ \|\text{fl}(g'_{\ell}) - g'_{\ell}\|_2 \leq e_G \|g_{\ell}\|_2. \quad (3.2) \]
Let \( U_1 \) be the product of all independent rotations applied in step 2. The computed \( \ell \)th column \( \text{fl}(g'_{\ell}) = \text{fl}(U_1^{-1} g_\ell) \) after step 2 can be interpreted as an exact result with a slightly perturbed unitary matrix \( \tilde{U}_1^{-1} \)
\[
\text{fl}(g'_{\ell}) = \tilde{U}_1^{-1} g_\ell.
\]
The similar reasoning is valid in step 3, where we apply \([\log_2 m]\) sequences of independent interlaced rotations. Repeating the same argument for step 4 (the second column of \(G\)) this gives a total of

\[ p_1 = [\log_2 m] + [\log_2 (m - 1)] + 2 \]  

(3.3)

sequences of independent rotations and interlaced rotations for the first two columns. Similarly as in [8], from (3.2), we can conclude that

\[
\| \tilde{U}^{-1} \tilde{U}^{-1}_{p_1-1} \cdots \tilde{U}^{-1}_1 \tilde{g}_\ell - U^{-1}_{p_1} U^{-1}_{p_1-1} \cdots U^{-1}_1 g_\ell \|_2 \\
\leq ((1 + e_G)^{p_1} - 1) \| g_\ell \|_2
\]  

(3.4)

after the first four steps.

If we use a different sequence of rotations to reduce the first two columns, (3.4) remains valid with appropriate \(p_1\) for that sequence. For sequential annihilation, \(p_1\) can be as high as \(4m - 6\).

3.2. Error analysis for one symplectic transformation

To complete the reduction of the first two columns, in step 5 we apply a single block symplectic transformation \(U^{-1}_{S2}(1)\).

From (2.2) we can conclude that

\[
\begin{align*}
\tilde{f}(g'_1) &= (1 + \epsilon_1)(g_{1\ell} + (1 + \epsilon_2)(1 + \epsilon_a)ag_{4\ell}) \\
\tilde{f}(g'_{3\ell}) &= (1 + \epsilon_3)((1 + \epsilon_4)(1 + \epsilon_a)ag_{2\ell} + g_{3\ell}).
\end{align*}
\]

Neglecting the terms of order \(O(\epsilon^2)\), we have

\[
\begin{bmatrix}
\tilde{f}(g'_1) - g'_1 \\
\tilde{f}(g'_{3\ell}) - g'_{3\ell}
\end{bmatrix} = \begin{bmatrix}
\epsilon_1 g_{1\ell} + (\epsilon_1 + \epsilon_2 + \epsilon_a)ag_{4\ell} \\
(\epsilon_3 + \epsilon_4 + \epsilon_a)ag_{2\ell} + \epsilon_3 g_{3\ell}
\end{bmatrix}.
\]

Since \(U^{-1}_{S2}(1)\) changes only the first and the third row of \(G\), we have

\[
\| \tilde{f}(g'_1) - g'_1 \|_2 = \left\| \begin{bmatrix}
\tilde{f}(g'_1) - g'_1 \\
\tilde{f}(g'_{3\ell}) - g'_{3\ell}
\end{bmatrix} \right\|_2.
\]

By using \(|x| + |y| \leq \sqrt{2} \sqrt{x^2 + y^2}\), this immediately yields

\[
\begin{align*}
\| \tilde{f}(g_i) - g'_i \|_2 &= \sqrt{(\epsilon_1 g_{1\ell} + (\epsilon_1 + \epsilon_2 + \epsilon_a)ag_{4\ell})^2 + ((\epsilon_3 + \epsilon_4 + \epsilon_a)ag_{2\ell} + \epsilon_3 g_{3\ell})^2} \\
&\leq \epsilon \sqrt{(|g_{1\ell}| + 3|a||g_{4\ell}|)^2 + (3|a||g_{2\ell}| + |g_{3\ell}|)^2}
\end{align*}
\]
\begin{align*}
&\leq 3\varepsilon \max\{1, |a|\} \sqrt{(|g_{1\ell}| + |g_{4\ell}|)^2 + (|g_{2\ell}| + |g_{3\ell}|)^2} \\
&\leq 3\varepsilon \max\{1, |a|\} \sqrt{2(g_{1\ell}^2 + g_{4\ell}^2 + g_{2\ell}^2 + g_{3\ell}^2)}
\end{align*}

and, finally
\begin{equation}
\|\Omega(g'_\ell) - g'_\ell\|_2 \leq \varepsilon_S \|g_\ell\|_2 := 3\sqrt{2}\varepsilon \max\{1, |a|\} \|g_\ell\|_2. \tag{3.5}
\end{equation}

### 3.3. Rounding errors of symplectic QR factorization

If we combine the results of the previous two subsections, we get the following theorem.

**Theorem 3.1.** Let
\begin{equation*}
R = \begin{bmatrix} R_1 \\ 0 \end{bmatrix} = Q^{-1} GP_2, \quad Q^T J Q = J,
\end{equation*}
be the exact symplectic QR factorization of $G$, with upper triangular $R_1$. Suppose that matrices $\tilde{Q}$ and upper triangular $\tilde{R}_1$ are computed as factors of $G$ in floating point arithmetic, with the same permutation matrix $P_2$. Then $\tilde{R}$ is the exact symplectic QR factor of some perturbed matrix $\tilde{G} = G + E$.

\begin{equation*}
\tilde{R} = \begin{bmatrix} \tilde{R}_1 \\ 0 \end{bmatrix} = \tilde{Q}^{-1}(G + E)P_2, \quad \tilde{Q}^T J \tilde{Q} = J.
\end{equation*}

Let all parameters $a_k$ for $U_{S_2}^{-1}(k)$ used in the reduction process satisfy $|a_k| \leq a_{\text{max}}$. Then
\begin{equation}
\|EP_2\varepsilon_{\ell}\|_2 \leq \tilde{\varepsilon}_{\ell}\|GP_2\varepsilon_{\ell}\|_2, \quad \ell = 2k - 1, 2k,
\end{equation}
where
\begin{equation}
\tilde{\varepsilon}_{\ell} = \gamma^{k-1} \left( \sum_{i=1}^{k-1} \text{err}(p_i) + [(1 + e_G)^{p_i} - 1] \right) \tag{3.6}
\end{equation}
for $k = 1, \ldots, n$. Here $p_i$ denotes the number of sequences of independent rotations and interlaced rotations at stage $i$, as in (3.3), and
\begin{equation*}
\text{err}(p_i) = e_S(1 + e_G)^{p_i} + \gamma \left[ (1 + e_G)^{p_i} - 1 \right],
\end{equation*}
where $e_G$ is defined by (3.1), $e_S$ by (3.5), and $\gamma = \sqrt{2 + a_{\text{max}}^2}$.

**Proof.** The proof uses the same technique as the proof of Theorem 3.1 from [8]. We apply (3.4) and (3.5) for all stages $k = 1, \ldots, n$ and note that all matrices $U_{S_2}^{-1}(k)$ satisfy
\begin{equation*}
\|U_{S_2}^{-1}(k)\|_2 \leq \sqrt{2 + a_k^2} \leq \sqrt{2 + a_{\text{max}}^2} = \gamma.
\end{equation*}
Therefore, we can substitute $e_S$ for $e_H$ and use $\gamma$ from above in Theorem 3.1 from [8]. The term in brackets in (3.6) follows from the fact that the last symplectic transformation $U^{-1}_{S_2}(k)$ introduces no errors in columns $\ell = 2k - 1, 2k$. It changes only the element in position $(2k + 1, 2k)$ which is explicitly set to 0. □

The following corollary is a generalization of Lemma 18.8 from [5]. It gives normwise and componentwise bounds for the backward perturbation $E$ of the original matrix $G$ in the symplectic QR.

**Corollary 3.1.** Under the assumptions of Theorem 3.1, $E$ satisfies normwise and componentwise bounds

$$\|E\|_F \leq \hat{\epsilon}_n \|G\|_F.$$

$$|E| \leq 2m\hat{\epsilon}_n |K| \|G\|_2,$$

where $\hat{\epsilon}_n$ is defined by (3.6) and $K = (2m)^{-1}ee^T$, with $e^T = [1, 1, \ldots, 1]$.

3.4. Rounding errors of implicit Bunch factorization

To obtain the implicit Bunch factorization, we have to apply a sequence of transformations $Y_k$ as in (2.6). Additional errors introduced by this step are described by the following result.

**Lemma 3.1.** Let $R_1$ be the triangular factor from the symplectic QR factorization of $G$. If the implicit Bunch factor $R_B$ from Theorem 2.2 is computed in floating point arithmetic, the error in the computed $\ell$th column of $R_B$ satisfies

$$\|\delta(r_B)\|_2 \leq \sqrt{59}e\|\delta r\|_2 := e_B \|r\|_2,$$

where $\omega$ is maximal absolute value of elements of $Y_k$ for $k = 1, \ldots, n$.

**Proof.** Note that each $Y_k$ modifies only two rows $(2k - 1$ and $2k)$ of $R_1$, and that $Y_k$ are independent, i.e., each row of $R_1$ is modified only by a single $Y_k$. Therefore

$$R_B([2k - 1, 2k], 1:n) = (Y_k R_1)([2k - 1, 2k], 1:n)$$

and it is enough to consider errors introduced by a single transformation $Y_k$. Let us, for simplicity, denote

$$Y_k([2k - 1, 2k], [2k - 1, 2k]) = \begin{bmatrix} y_{11} & y_{12} \\ 0 & y_{22} \end{bmatrix}.$$
From (2.5), the computed elements of this matrix satisfy

\[ f_1(y_{11}) = (1 + \varepsilon_{11})y_{11}, \quad |\varepsilon_{11}| \leq 2\varepsilon, \]
\[ f_2(y_{22}) = (1 + \varepsilon_{22})y_{22}, \quad |\varepsilon_{22}| \leq 2\varepsilon, \]
\[ f_3(y_{12}) = (1 + \varepsilon_{12})y_{12}, \quad |\varepsilon_{12}| \leq 3\varepsilon. \]

Let \( \delta(r_B)_{i,\ell} = \text{fl}((r_B)_{i,\ell}) - (r_B)_{i,\ell} \). For error in the computed elements in rows \( 2k - 1 \) and \( 2k \) in column \( (r_B)_{\ell} \) of \( R_B \), ignoring terms of order \( \varepsilon^2 \), we have

\[ \delta(r_B)_{2k-1,\ell} = (\varepsilon_1 + \varepsilon_2 + \varepsilon_{11})y_{11r_{2k-1,\ell}} + (\varepsilon_1 + \varepsilon_5 + \varepsilon_{12})y_{12r_{2k,\ell}}, \]
\[ \delta(r_B)_{2k,\ell} = (\varepsilon_4 + \varepsilon_{22})y_{22r_{2k,\ell}}. \]

By using \( (|x| + |y|)^2 \leq 2(x^2 + y^2) \) we obtain

\[
\left\lVert \begin{bmatrix} \delta(r_B)_{2k-1,\ell} \\ \delta(r_B)_{2k,\ell} \end{bmatrix} \right\rVert_2^2 \leq (4\varepsilon|y_{11}r_{2k-1,\ell}| + 5\varepsilon|y_{12}r_{2k,\ell}|)^2 + 9\varepsilon^2|y_{22}r_{2k,\ell}|^2 \\
\varepsilon^2\omega^2(32r_{2k-1,\ell}^2 + 59r_{2k,\ell}^2) \leq 59\varepsilon^2\omega^2(r_{2k-1,\ell}^2 + r_{2k,\ell}^2).
\]

After application of all \( Y_k \), the norm of the error in \( (r_B)_\ell \) has the same form

\[ \|\delta(r_B)_\ell\|_2 \leq \sqrt{59}\omega\|e\|_2. \quad \square \]

The overall backward error bound for the implicit Bunch factorization now follows from Theorem 3.1.

**Theorem 3.2.** Let \( R_B \) be the implicit Bunch factor of \( G \) computed by symplectic QR factorization and (2.6), and \( \tilde{R}_B \) the computed implicit factor of \( G \) in floating point arithmetic. Then \( \tilde{R}_B \) is the exact symplectic QR factor of some perturbed matrix \( \tilde{G} = G + E \),

\[
\begin{bmatrix} \tilde{R}_B \\ 0 \end{bmatrix} = \tilde{Q}_B^{-1}(G + E)P_2, \quad \tilde{Q}_B J \tilde{Q}_B = J
\]

with

\[ \|EP_2e\|_2 \leq \mu_k\|GP_2e\|_2, \quad \ell = 2k - 1, 2k, \]

where

\[ \mu_k = (e_B(\tilde{\delta}_k + \gamma^k) + \sqrt{3}\tilde{\delta}_k)\omega \] \hspace{1cm} (3.8)

for \( k = 1, \ldots, n \). Here \( \gamma \) and \( \tilde{\delta}_k \) are defined in Theorem 2.2, and \( e_B \) by (3.7).

**Proof.** Suppose that \( G \) and \( J \) have been prepermuted according to \( P_2 \) before the algorithm.

Let \( Y = \prod_{k=1}^n Y_k \), and let \( Q \) be the matrix of all previous symplectic transformations. Rounding errors can be interpreted as an exact symplectic QR factorization postmultiplied by \( Y \), on a slightly perturbed matrix \( \tilde{G} \), i.e.,
\[ \| \tilde{Y} \tilde{Q}^{-1} g_\ell - Y Q^{-1} g_\ell \|_2 \leq \| \tilde{Y} \tilde{Q}^{-1} g_\ell - Y \tilde{Q}^{-1} g_\ell \|_2 + \| Y \tilde{Q}^{-1} g_\ell - Y Q^{-1} g_\ell \|_2 \]
\[ \leq e_{B \omega} \| \tilde{g}_\ell \|_2 + \| Y \|_2 \| \tilde{g}_\ell - Q^{-1} g_\ell \|_2, \]  
(3.9)

where \( \tilde{g}_\ell = \tilde{Q}^{-1} g_\ell \). Inequality

\[ \| \tilde{g}_\ell \| \leq \| \tilde{Q}^{-1} g_\ell - Q^{-1} g_\ell \|_2 + \| Q^{-1} g_\ell \|_2 \leq (\tilde{e}_k + \gamma^k) \| g_\ell \|_2 \]  
(3.10)

is a consequence of Theorem 3.1 and the fact that the annihilation of the first 2\( k \) columns of \( G \) requires at most \( k \) transformations \( U_{S_2}^{-1}(i) \), \( i = 1, \ldots, k \), so

\[ \| Q^{-1} \|_2 \leq \prod_{i=1}^{k} \| U_{S_2}^{-1}(i) \|_2 \leq \gamma^k. \]

It is easy to compute that

\[ \| Y \|_2 = \max_{k=1, \ldots, n} \sqrt{\lambda_{\max}(Y_k^T Y_k)} \leq \sqrt{3} \omega. \]  
(3.11)

Substitution of (3.10) and (3.11) into (3.9), together with the Theorem 3.1 proves the statement of the theorem. \( \square \)

The size of elements of \( Y \) (measured by \( \omega \)) is directly linked to element growth in the Bunch factorization of \( A \).

For the implicit Bunch factorization we have the following analog of Corollary 3.1.

**Corollary 3.2.** Under the assumptions of Theorem 3.2, \( E \) satisfies normwise and componentwise bounds

\[ \| E \|_F \leq \mu_n \| G \|_F, \]
\[ |E| \leq 2m \mu_k K |G|, \quad \| K \|_F = 1, \]

where \( \mu_k \) is defined by (3.8) and \( K = (2m)^{-1}ee^T \), with \( e^T = [1, 1, \ldots, 1] \).

### 4. Perturbation analysis

One of our objectives in constructing the symplectic QR was to compute the eigenvalues of \( A \) without computing \( A \), when \( A \) is given implicitly by its factors \( G \) and \( J \). To access the accuracy of computed eigenvalues, we need a perturbation theory for the symplectic QR which can be used, together with the error analysis already done, to provide the required eigenvalue bounds.
As can be expected, the results are similar to those for the ordinary QR and the triangular indefinite QR, and some common parts of the proofs will be omitted.

The Bauer–Skeel condition number of nonsingular square matrix is defined by (see, e.g., [10])

$$\kappa_{BS}(S) = ||S^{-1}|| ||S||.$$  

The following theorem is a complete analog of Theorem 4.1 in [8].

**Theorem 4.1.** Let $G, \tilde{G} \in \mathbb{C}^{2m \times 2n}$, $m \geq n$, be such that

$$\text{rank}(G^T JG) = \text{rank}(\tilde{G}^T J\tilde{G}) = 2n$$

with $J \in \mathbb{C}^{2m \times 2m}$ given by (1.2). Also, let

$$G = Q_B \begin{bmatrix} R_B & 0 \\ 0 & 0 \end{bmatrix} P_2^T = [Q_1, Q_2] \begin{bmatrix} R_B & 0 \\ 0 & 0 \end{bmatrix} P_2^T, \quad Q_B^T J Q_B = J$$

and

$$\tilde{G} = \tilde{Q}_B \begin{bmatrix} \tilde{R}_B & 0 \\ 0 & 0 \end{bmatrix} P_2^T = [\tilde{Q}_1, \tilde{Q}_2] \begin{bmatrix} \tilde{R}_B & 0 \\ 0 & 0 \end{bmatrix} P_2^T, \quad \tilde{Q}_B^T J \tilde{Q}_B = J$$

be the symplectic QR factorizations of $G$ and $\tilde{G}$, respectively, with the Bunch form of upper triangular $R_B$ and $\tilde{R}_B$, and let $J_{11} = J(1 : 2n, 1 : 2n)$. Let $E = \tilde{G} - G$,

$$F = \begin{bmatrix} F_B \\ 0 \end{bmatrix} = \begin{bmatrix} \tilde{R}_B - R_B \\ 0 \end{bmatrix}$$

$$W = [W_1, W_2] = \tilde{Q}_B - Q_B = [\tilde{Q}_1 - Q_1, \tilde{Q}_2 - Q_2]$$

and

$$|E| \leq \varepsilon |G|$$

for some matrix $K$ with nonnegative elements.

If $\| \cdot \|$ is a consistent and monotone norm and

$$\eta = \max \left\{ \| Q_1^T J Q_1 \|, \| Q_2^T J Q_2 \|, \| K \| \right\}$$

and

$$\varepsilon \eta \| J_{11} \|(\kappa_{BS}(R_B^{-1}) + \kappa_{BS}(R_B^T)) < 1,$$

then we have

$$\frac{\| F_B \|}{\| R_B \|} \leq \varepsilon \eta \| J_{11} \|(\kappa_{BS}(R_B^{-1}) + \kappa_{BS}(R_B^T)) + O(\varepsilon^2)$$

and

$$\frac{\| W_1 \|}{\| Q_1 \|} \leq \frac{\varepsilon (\| R \| \kappa_{BS}(R_B^{-1}) + \eta (\kappa_{BS}(R_B^{-1}) + \kappa_{BS}(R_B^T)))}{1 - \varepsilon \eta \| J_{11} \|(\kappa_{BS}(R_B^{-1}) + \kappa_{BS}(R_B^T))} + O(\varepsilon^2).$$
Proof. The beginning of the proof is similar to proof of Theorem 4.1 from [8]. We obtain that

\[ |W_1| \leq |EP_2| |R_B^{-1}| + |Q_1| |F_B R_B^{-1}| + |W_1| |F_B R_B^{-1}| \]

and up to elements of order \( \varepsilon^2 \)

\[ J_{11} F_B R_B^{-1} + R_B^{-T} F_B^T J_{11} = R_B^{-T} P_2^T E J_1 + Q_1^T E P_2 R_B^{-1}. \]

The matrix \( J_{11} F_B R_B^{-1} + R_B^{-T} F_B^T J_{11} = J_{11} F_B R_B^{-1} - (J_{11} F_B R_B^{-1})^T \) has a special structure.

To simplify the notation, let \( \text{bdiag} \) be the block matrix analog of the \( \text{diag} \) operator in the following sense: \( \text{bdiag}(A_{11}, \ldots, A_{nn}) \) is a block diagonal matrix with diagonal blocks \( A_{11}, \ldots, A_{nn} \), and \( \text{bdiag}(A) \) is a block diagonal matrix with diagonal blocks taken from \( A \). Dimensions of blocks will be clear from the context. Subsequently, we will use only blocks of order 2 in \( \text{bdiag} \). Also, let

\[ \text{boffdiag}(A) = A - \text{bdiag}(A) \]

be the block off-diagonal part of \( A \)—the matrix \( A \) with diagonal blocks (of order 2) equal to 0.

Since \( F_B R_B^{-1} \) and \( J_{11} F_B R_B^{-1} \) are block upper triangular, we conclude that

\[ \text{boffdiag}(|J_{11} F_B R_B^{-1}|) \leq \text{boffdiag}(|J_{11} F_B R_B^{-1} - (J_{11} F_B R_B^{-1})^T|). \] (4.1)

The diagonal blocks of \( R_B \) are

\[ \text{bdiag}(R_B) = \text{bdiag} \left( \begin{bmatrix} r_{11} & 0 \\ 0 & r_{11} \end{bmatrix}, \ldots, \begin{bmatrix} r_{nn} & 0 \\ 0 & r_{nn} \end{bmatrix} \right). \]

so the diagonal blocks of \( F_B \) and \( F_B R_B^{-1} \) have a similar scalar structure. Therefore,

\[ \text{bdiag}(J_{11} F_B R_B^{-1}) = \text{bdiag} \left( \begin{bmatrix} 0 & z_{11} \\ -z_{11} & 0 \end{bmatrix}, \ldots, \begin{bmatrix} 0 & z_{nn} \\ -z_{nn} & 0 \end{bmatrix} \right) \] (4.2)

and

\[ \text{bdiag}(J_{11} F_B R_B^{-1} - (J_{11} F_B R_B^{-1})^T) = \text{bdiag} \left( \begin{bmatrix} 0 & 2z_{11} \\ -2z_{11} & 0 \end{bmatrix}, \ldots, \begin{bmatrix} 0 & 2z_{nn} \\ -2z_{nn} & 0 \end{bmatrix} \right). \] (4.3)

By taking absolute values in (4.2) and (4.3) we see that

\[ \text{bdiag}(|J_{11} F_B R_B^{-1}|) \leq \text{bdiag}(|J_{11} F_B R_B^{-1} - (J_{11} F_B R_B^{-1})^T|). \] (4.4)

Relations (4.1) and (4.4) together yield

\[ |J_{11} F_B R_B^{-1}| \leq |J_{11} F_B R_B^{-1} - (J_{11} F_B R_B^{-1})^T|. \] (4.5)
Now, as in the proof of Theorem 4.1 [8], we obtain
\[ |J_1F_B R_B^{-1} - (J_1F_B R_B^{-1})^T| \leq \varepsilon (\|R_B^{-T}| |R_B^T| |Q_1^T| J| |Q_1| + |Q_1^T| J| K| |Q_1| + |R_B|||R_B^{-1}|). \] (4.6)

From \( |F_B| \leq |J_1||J_1F_B R_B^{-1}| |R_B| \) and (4.5) and (4.6), we obtain
\[
\begin{align*}
\|F_B\| &\leq \|J_1\| \|J_1F_B R_B^{-1}\| \|R_B\| \\
&\leq \|J_1\| \|J_1F_B R_B^{-1} - (J_1F_B R_B^{-1})^T\| \|R_B\| \\
&\leq \varepsilon \|J_1\| (\|R_B^{-T}\| \|R_B^T\| \|Q_1\| \|J| K| |Q_1|) \\
&\quad + \|Q_1^T| K| J| |Q_1| \|R_B\| \|R_B^{-1}\| \|R_B\| \\
&= \varepsilon \eta \|J_1\| (\kappa_{BS}(R_B^{-1}) + \kappa_{BS}(R_B^T)) R_B.
\end{align*}
\]

This proves the first statement of the theorem. The second statement follows from
\[
\begin{align*}
|W_1| &\leq \|E P_2\| \|R_B^{-1}\| + |Q_1||J_1||J_1F_B R_B^{-1}| + |W_1||J_1||J_1F_B R_B^{-1}| \\
&\leq \varepsilon \eta \|J_1\| (\|R_B\| \|R_B^{-1}\| + |Q_1||J_1||J_1F_B R_B^{-1}| + |W_1||J_1||J_1F_B R_B^{-1}|)
\end{align*}
\]

and \( |J_1F_B R_B^{-1}| \leq \varepsilon \eta (\kappa_{BS}(R_B^{-1}) + \kappa_{BS}(R_B^T)) \). By monotonicity of the norm, we conclude
\[
\begin{align*}
\|W_1\| &\leq \varepsilon \|J_1\| \|Q_1\| + \|R_B\| \|R_B^{-1}\| \\
&\quad + \|J_1\| (|Q_1| \|J_1F_B R_B^{-1}| + |W_1| \|J_1F_B R_B^{-1}|) \\
&\leq \varepsilon (\|K\| \|Q_1\| + \|R_B\| \|R_B^{-1}\|) \\
&\quad + \eta \|J_1\| (|Q_1| + \|W_1\|)(\kappa_{BS}(R_B^{-1}) + \kappa_{BS}(R_B^T)).
\end{align*}
\]

Rearrangement of terms completes the proof. \( \square \)

Now we can link rounding-error bounds and perturbation bounds. The componentwise bound for \( E \) from Corollary 3.2 can be used in Theorem 4.1.

**Corollary 4.1.** Under the assumptions of Theorem 4.1 with \( \eta = \|Q_1\|^2_F \) and \( \varepsilon = 2m\mu_n \), the matrix \( F_B \) satisfies
\[
\frac{\|F_B\|_F}{\|R_B\|_F} \leq \varepsilon \eta \sqrt{2m(\kappa_{BS}(R_B^{-1}) + \kappa_{BS}(R_B^T))}.
\]

**Proof.** Take \( K = (2m)^{-1}ee^T \) from Corollary 3.2. Since \( |J|K = K \), we have
\[
\eta = (2m)^{-1}\|Q_1^T|ee^T|Q_1|\|_F \leq \tilde{\eta}.
\] \( \square \)
This Frobenius norm result can be used to obtain a more useful rowwise perturbation bound.

**Corollary 4.2.** Under the assumptions of Theorem 4.1 and Corollary 4.1
\[ \| \bar{F}^T B x \|_2 \leq v \| R_B^T x \|_2, \]
for all \( x \in \mathbb{R}^n \), where
\[ v = 2m \sqrt{2} \mu_k \tilde{\eta} (\kappa_{BS}(R_B^{-1}) + \kappa_{BS}(R_B^T)) \| R_B^{-1} \|_F \| R_B \|_F. \]

**Proof.** We have
\[ \| F^T B x \|_2 = \| F^T R_B^{-T} R_B^T x \|_2 \leq \| R_B^{-1} F_B \|_F \| R_B^T x \|_2. \]
The assertion follows from Corollary 4.1 and \( \| R_B^{-1} F_B \|_F \leq \| R_B^{-1} \|_F \| F_B \|_F. \)

Finally, this result can be used to bound the floating point perturbations of eigenvalues of factorized skew-symmetric matrices after the symplectic QR reduction.

**Corollary 4.3.** Let \( A = P_2 R_B^T J_1 R_B P_2^T \) and \( \tilde{A} = P_2 \tilde{R}_B^T J_1 \tilde{R}_B P_2^T \). A nonsingular matrix \( A \) with
\[ \| F^T B x \|_2 \leq v \| R_B^T x \|_2 \]
for all \( x \in \mathbb{R}^n \) and \( v < 1 \). Then eigenvalues \( \pm i \lambda_k \) of \( A \) and \( \pm i \lambda_k' \) of \( \tilde{A} \) satisfy the inequalities
\[ (1 - v)^2 \leq \frac{\lambda_k'}{\lambda_k} \leq (1 + v)^2. \]

**Proof.** By a direct consequence of Theorem 1.0.4 from [6] and the previous corollary the result follows.

**Remark 4.1.** All the results of this section have been stated in terms of the implicit Bunch factor \( R_B = Y R_1 \). In practice, we can also use \( R_1 \) as an input for the Pietzsch algorithm, and thus avoid all the errors introduced by forming \( R_B \).

Unfortunately, the technique used in the proof of Theorem 4.1 cannot be applied on \( R_1 \), since it crucially depends on the scalar structure of diagonal blocks.

5. Numerical examples

The following examples have been constructed with two goals in mind:
- To illustrate the symplectic QR factorization.
- To show the advantage of this approach over the "multiply and factorize" approach.
Starting from the same initial matrix $G$, with $J$ as in (1.2), we will compare three different factorizations:

- **SQR**—symplectic QR factorization (Theorem 2.1).
- **SQRB**—implicit Bunch factorization (2.6).
- **SSF**—explicit Bunch factorization (2.4) of computed $A := G^T J G$.

The computed triangular factors $R$ by these factorizations will be denoted by $R_1$, $R_B$ and $R'_B$, respectively.

In theory, at least, $A$ and $R^T J R$ have the same eigenvalues. Since each of the computed factors $R_1$, $R_B$ and $R'_B$ can be used as an input for the Pietzsch algorithm to compute the eigenvalues of $A = G^T J G$, the quality of these computed factors can be measured by comparing the computed eigenvalues.

To minimize the effect of rounding errors introduced by the Pietzsch algorithm, the actual computation is performed in two steps. In the first step, all three factorizations are computed in two different precisions:

- **low**—IEEE single precision, with unit roundoff $\varepsilon_s = 2^{-24} \approx 5.96 \times 10^{-8}$,
- **high**—IEEE extended precision, with $\varepsilon_e = 2^{-64} \approx 5.42 \times 10^{-20}$,

giving a total of six computed factors. For each factor, the eigenvalues are computed in high precision only, to avoid additional roundoff contamination.

More precisely, the Pietzsch algorithm computes only the positive imaginary parts $\lambda_j$, $j = 1, \ldots, n$, of the actual eigenvalues $\lambda_j(A) = \pm \lambda_ji$ of $A$. To simplify the notation, we will compare the computed values of $\lambda_j$. Those computed from the extended precision $R_1$ will be taken as reference values for relative errors.

To compute the symplectic QR of $G$, we have used a column pivoting strategy that is equivalent to a complete pivoting strategy in the skew-symmetric factorization of $A$. In all the examples below, the computed column permutation matrix $P_2$ is the same for all six factorizations, so $R_B$ and $R'_B$ can also be compared elementwise.

Finally, an additional information about the quality of factors $R_1$ and $R_B$ can be provided by monitoring the condition of symplectic matrices $Q$ and $Q_B$ in respective symplectic QR factorizations of $G$. Note that $Q'_B$ can also be computed, once we have $R'_B$, but there is no point in doing so.

The first example shows that both symplectic QR factors $R_1$ and $R_B$ are very good, while $R'_B$ is not, even though $A$ is computed correctly.

**Example 5.1.** Let

$$G = \begin{bmatrix} 2.0e+2 & 1.0e+1 & 1.0e-1 & -1.0e-1 \\ -1.0e+1 & 1.0e+2 & 1.0e-1 & 2.0e-1 \\ 1.0e-3 & 1.0e-3 & -1.0e-4 & 1.0e-5 \\ 1.0e-3 & -1.0e-3 & 2.0e-5 & 1.0e-4 \end{bmatrix}.$$ 

The computed matrix $A = G^T J G$ in single precision is
The elements of $A$ have small relative errors, so $A$ is computed accurately. The symplectic QR factorization of $G$ in single precision gives

$$R_1 = \begin{bmatrix} 2.0024985e+02 & 4.9937620e+00 & -1.0986276e-01 & 9.4881475e-02 \\ 1.0037461e+02 & 1.9475672e-01 & 1.0486900e-01 & 1.1985863e-05 \\ 2.0290208e-04 & 1.0290208e+00 & -1.1.0286549e-04 \end{bmatrix}.$$ 

with small relative error in each element. The implicit Bunch factor $R_B$ computed from $R_1$ is

$$R_B = \begin{bmatrix} 1.4177448e+02 & 0.0000000e+00 & -8.4641472e-02 & 6.3481107e-02 \\ 1.4177448e+02 & 2.7508482e-01 & 1.4812259e-01 & 6.7301320e-05 \\ 0.0000000e+00 & 0.0000000e+00 & 1.0288378e-04 & 6.7301320e-05 \end{bmatrix}.$$

again with small relative error in each element. In addition, the conditions of matrices $Q$ and $Q_B = QY^{-1}$ are low: $\kappa_2(Q) \approx 1$ and $\kappa_2(Q_B) \approx \kappa_2(Y) \approx 2$, which confirms how good both factors $R_1$ and $R_B$ are.

On the other hand, the computed explicit Bunch factor $R_B'$ from $A$ in single precision is

$$R_B' = \begin{bmatrix} 1.4177448e+02 & 0.0000000e+00 & -8.4641472e-02 & 6.3481107e-02 \\ 1.4177448e+02 & 2.7508482e-01 & 1.4812259e-01 & 6.7301320e-05 \\ 6.7301320e-05 & 0.0000000e+00 & 1.0288378e-04 & 6.7301320e-05 \end{bmatrix}.$$

While the first two rows are good, the last two diagonal elements have high relative errors. Their true value (from extended) is $1.0288379e-04$, just as in $R_B$. Even though $A$ is computed correctly, it is ill conditioned, and SSF suffers from severe cancellation in the last step.

The computed eigenvalues by the Pietzsch algorithm for all six factorizations are given in Tables 1 and 2, along with relative errors.

As one can expect, $\lambda_1$ is fully accurate.

Surprisingly, since $A$ is so ill conditioned, $\lambda_2$ is almost as accurate as $\lambda_1$, when computed from $R_1$ and $R_B$. The trick is simply not to compute $A$, and $\lambda_2$ is well determined by the initial $G$.

Our next example shows that, as long as the condition of $Q$ is low, the computed eigenvalues are also good. There is a strong numerical evidence that $\kappa_2(Y)$ has almost no effect on eigenvalues.
Table 1

<table>
<thead>
<tr>
<th></th>
<th>$\lambda_1$</th>
<th>Relative error</th>
</tr>
</thead>
<tbody>
<tr>
<td>extended SQR</td>
<td>2.0100054409342457e+04</td>
<td>5.3025e−19</td>
</tr>
<tr>
<td>extended SQRB</td>
<td>2.0100054409342457e+04</td>
<td>5.3025e−19</td>
</tr>
<tr>
<td>extended SSF</td>
<td>2.01000549198904817e+04</td>
<td>2.5819e−08</td>
</tr>
<tr>
<td>single SQR</td>
<td>2.01000561921609765e+04</td>
<td>8.9116e−08</td>
</tr>
<tr>
<td>single SSF</td>
<td>2.01000561921419000e+04</td>
<td>8.9115e−08</td>
</tr>
</tbody>
</table>

Table 2

<table>
<thead>
<tr>
<th></th>
<th>$\lambda_2$</th>
<th>Relative error</th>
</tr>
</thead>
<tbody>
<tr>
<td>extended SQR</td>
<td>1.0585045978289041e−08</td>
<td>2.3658e−18</td>
</tr>
<tr>
<td>extended SQRB</td>
<td>1.05850459782928210e−08</td>
<td>3.5700e−13</td>
</tr>
<tr>
<td>extended SSF</td>
<td>1.05850438728025921e−08</td>
<td>1.9891e−07</td>
</tr>
<tr>
<td>single SQR</td>
<td>1.05850434589461458e−08</td>
<td>2.3801e−07</td>
</tr>
<tr>
<td>single SSF</td>
<td>4.52945542592941665e−09</td>
<td>5.7209e−01</td>
</tr>
</tbody>
</table>

Example 5.2. Let

$$G = \begin{bmatrix} 1.0e+7 & 1.0e+4 & 1.0e+3 & 1.0e+1 \\ 1.0e+4 & -1.0e+3 & 1.0e+0 & 1.0e-1 \\ 1.0e+2 & 1.0e+0 & 1.0e-1 & 1.0e-3 \\ 1.0e-2 & -1.0e-3 & 0.0e+0 & 1.0e-4 \end{bmatrix}$$

The symplectic factorization of $G$ in single precision gives

$$R_1 = \begin{bmatrix} 1.0049876e+04 & 9.9493770e+06 & 9.9493768e+02 & 9.9404221e+00 \\ 1.0049876e+06 & 1.0049876e+02 & 1.0945410e+00 & 9.0000000e-02 \\ 9.0000000e-02 & 9.0196999e-04 & 9.9910896e-05 \end{bmatrix}$$

with $\kappa_2(G) \approx 1$. Successive pairs of diagonal elements of $R_1$ have widely varying orders of magnitude, even within each pair. This is reflected in $\kappa_2(Y) \approx 9.9 \times 10^3$ to obtain pairs of equal diagonal elements in

$$R_B = \begin{bmatrix} 1.0049876e+05 & 0.0000000e+00 & -5.3405762e-05 & -8.9553337e+00 \\ 1.0049876e+05 & 1.0049876e+01 & 1.0945410e-01 & 0.0000000e+00 \\ 2.9986631e-03 & 0.0000000e+00 & 2.9986631e-03 \end{bmatrix}$$

Once again, $A$ is ill-conditioned and computed correctly, so $R_B'$ suffers in the same way as before
Table 3

<table>
<thead>
<tr>
<th></th>
<th>( \lambda_2 )</th>
<th>Relative error</th>
</tr>
</thead>
<tbody>
<tr>
<td>extended SQR</td>
<td>8.99198011725668472e−06</td>
<td></td>
</tr>
<tr>
<td>extended SQRB</td>
<td>8.99198011725668472e−06</td>
<td>0.0000e+00</td>
</tr>
<tr>
<td>extended SSF</td>
<td>8.99198011725056505e−06</td>
<td>6.8057e−13</td>
</tr>
<tr>
<td>single SQR</td>
<td>8.99198092535999536e−06</td>
<td>8.9869e−08</td>
</tr>
<tr>
<td>single SQRB</td>
<td>8.99198038172997935e−06</td>
<td>2.9412e−08</td>
</tr>
<tr>
<td>single SSF</td>
<td>1.13550501275070634e−05</td>
<td>2.6280e−01</td>
</tr>
</tbody>
</table>

\[
R'_B = \begin{bmatrix}
1.0049876e+05 & 0.0000000e+00 & 9.9503721e−09 & -8.9553347e+00 \\
1.0049876e+05 & 1.0049875e+01 & 1.0945409e−01 & 3.3697255e−03 \\
3.6972555e−03 & 0.0000000e+00 & 0.0000000e+01 & 3.6972555e−03 \\
\end{bmatrix}
\]

The first eigenvalue is \( \lambda_1 = 1.01000000907150009e+10 \), and all algorithms compute it with high relative accuracy. The second one is given in Table 3.

Finally, we want to show that the eigenvalues computed from \( R_1 \) and \( R_B \) are sensitive to the condition of \( Q \). This time, just for illustration, we will work with rectangular \( G \).

Example 5.3. Let

\[
G = \begin{bmatrix}
1.0000e+1 & 2.0000e+1 & 1.0000e+0 & 1.1000e+0 \\
1.0000e+0 & 2.0000e+0 & 1.0000e+0 & 1.0000e+0 \\
1.0000e−1 & 2.0000e−1 & 1.1000e−2 & 1.1005e−2 \\
1.0005e−1 & 2.0005e−1 & 1.0000e−2 & 1.0000e−2 \\
1.0000e−2 & 2.0000e−2 & 1.0000e−3 & 2.0000e−3 \\
1.0000e−2 & 1.0050e−2 & 1.1000e−3 & 2.1000e−3 \\
\end{bmatrix}
\]

The computed \( R_1 \) in single precision is

\[
R_1 = \begin{bmatrix}
2.0101748e+01 & 1.0946424e+00 & 1.0050874e+01 & 1.1941583e+00 \\
8.9543468e−01 & 2.3880710e−07 & 8.8549519e−01 & -7.0534544e−03 \\
8.8543468e−01 & 2.3880710e−07 & 8.8549519e−01 & -7.0534544e−03 \\
-7.0534544e−03 & -1.3484017e−03 & -1.1009070e−05 & -1.1009070e−05 \\
\end{bmatrix}
\]

with moderate condition of \( Q \), \( \kappa_2(Q) \approx 4.08 \times 10^4 \), which is almost entirely due to the high condition of the last nonunitary symplectic transformation \( U_{S_2}(2) \). The last diagonal element has low relative accuracy of about \( 10^{-4} \), and this spoils the smaller eigenvalue. The element in position \((2, 3)\) is even less accurate, but without any effect.
Table 4

<table>
<thead>
<tr>
<th></th>
<th>$\lambda_2$</th>
<th>Relative error</th>
</tr>
</thead>
<tbody>
<tr>
<td>extended SQR</td>
<td>4.93818201248208904e−08</td>
<td></td>
</tr>
<tr>
<td>extended SQRB</td>
<td>4.93818201248208902e−08</td>
<td>2.4864e−18</td>
</tr>
<tr>
<td>extended SSF</td>
<td>4.93818201421374008e−08</td>
<td>1.3847e−11</td>
</tr>
<tr>
<td>single SQR</td>
<td>4.93881928032005695e−08</td>
<td>1.2905e−04</td>
</tr>
<tr>
<td>single SQRB</td>
<td>4.9388188850786122e−08</td>
<td>1.2897e−04</td>
</tr>
<tr>
<td>single SSF</td>
<td>1.42321746618694365e−07</td>
<td>1.8821e+00</td>
</tr>
</tbody>
</table>

The condition of $Y$ is not low, $\kappa_2(Y) \approx 6.6 \times 10^2$, but this barely affects the condition of $Q_B$, as we have $\kappa_2(Q_B) \approx 4.15 \times 10^4$. The computed $R_B$ is

$$R_B = \begin{bmatrix}
4.2426171e+00 & 0.0000000e+00 & 2.1213086e+00 & 2.3568025e−02 \\
4.2426171e+00 & 1.1314808e−06 & 4.1955323e+00 & 2.7866716e−02
\end{bmatrix}.$$

The last two diagonal elements still have four correct decimal digits.

When $A$ is computed, a cancellation of about four decimal digits occurs in positions $(1, 2)$ and $(2, 1)$, reflecting the condition of $Q$.

$$A = \begin{bmatrix}
0.0000000e+00 & -4.5008305e−06 & 8.9999008e+00 & 8.8999996e+00 \\
-4.5008305e−06 & 0.0000000e+00 & 1.7999800e+01 & 1.7799997e+01 \\
-8.9999008e−06 & -1.7999800e+01 & 0.0000000e+00 & -9.9990174e−02 \\
-8.8999996e+00 & -1.7799997e+01 & 9.9990174e−02 & 0.0000000e+00
\end{bmatrix}.$$

All other (larger) elements of $A$ are good, so $\lambda_1$ is well determined by $G$, but $\lambda_2$ is not. Since $A$ is ill conditioned again, this is reflected in the computed $R_B'$.

$$R_B' = \begin{bmatrix}
4.2426171e+00 & 0.0000000e+00 & 2.1213086e+00 & 2.3568042e−02 \\
4.2426171e+00 & 1.0608618e−06 & 4.1955228e+00 & 4.7306065e−04
\end{bmatrix}.$$

The last two diagonal elements have no correct digits at all.

The first eigenvalue is $\lambda_1 = 2.8302844366088339e+01$, and all algorithms compute it with full relative accuracy. The second one is given in Table 4.

To be fair, we can easily construct an example of $G$ which badly determines all eigenvalues of $A$. The only thing needed is severe cancellation in all elements of $A$, and then even $\lambda_1$ suffers.

On the other hand, if $A$ is not so ill conditioned as in our examples, the eigenvalues computed from $R_B'$ can be just as good or even slightly better then those computed from $R_1$ or $R_B$.
6. Conclusion

In this paper we have developed a symplectic analog of the QR factorization, which can be used when a nonsingular skew-symmetric matrix is given by its factors $G$ and $J$. We have also presented rounding error and perturbation bounds for this factorization.

The symplectic QR factorization, combined with the Pietzsch algorithm, can be used as an accurate eigensystem solver. In some cases, this approach has distinct advantages over the “multiply and factorize” approach, as it computes eigenvalues with higher accuracy.

Some interesting questions still remain to be answered. For example, we have seen that matrices $Y$ which produce $R_B$ from $R_1$ can have relatively high conditions. This can be regarded as potentially dangerous, even though the quality of eigenvalues computed from $R_B$ remains more or less the same as from $R_1$.

Acknowledgement

We wish to thank Eberhard Pietzsch who kindly provided us with all the software used in his thesis.

References