BIT
2000, Vol. 40, No. 3, pp. 471–496

0006-3835/00/4003-0471 \$15.00

ERROR ANALYSIS OF THE SYMPLECTIC LANCZOS METHOD FOR THE SYMPLECTIC EIGENVALUE PROBLEM *

HEIKE FASSBENDER

Technische Universität München, Zentrum Mathematik, DE-80290 München, Germany
email: fassbender@mathematik.tu-muenchen.de

Abstract.

A rounding error analysis of the symplectic Lanczos algorithm for the symplectic eigenvalue problem is given. It is applicable when no break down occurs and shows that the restriction of preserving the symplectic structure does not destroy the characteristic feature of nonsymmetric Lanczos processes. An analog of Paige's theory on the relationship between the loss of orthogonality among the Lanczos vectors and the convergence of Ritz values in the symmetric Lanczos algorithm is discussed. As to be expected, it follows that (under certain assumptions) the computed J-orthogonal Lanczos vectors loose J-orthogonality when some Ritz values begin to converge.

AMS subject classification: 65G05, 65F15, 65F50.

Key words: Symplectic Lanczos method, symplectic matrix, eigenvalues, error analysis.

1 Introduction.

The Lanczos algorithm proposed by Lanczos in 1950 [11] is a procedure for the successive reduction of a given general matrix $A \in \mathbb{R}^{n\times n}$ to a tridiagonal matrix T. In the jth step the Lanczos algorithm generates two $n \times j$ matrices Q_j and P_j

$$Q_j = [q_1, q_2, \ldots, q_j], \qquad P_j = [p_1, p_2, \ldots, p_j]$$

which satisfy $P_j^T Q_j = I$ and

$$AQ_j = Q_jT_j + \beta_{j+1}q_{j+1}e_j^T, \tag{1.1}$$

$$A^TP_j = P_jT_j^T + \gamma_{j+1}p_{j+1}e_j^T, \tag{1.2}$$

where $e_j = [0, \ldots, 0, 1]^T \in \mathbb{R}^j$ and T_j is the tridiagonal matrix

$$T_j = \begin{bmatrix} \alpha_1 & \gamma_2 & & \\ \beta_2 & \alpha_2 & \ddots & \\ & \ddots & \ddots & \gamma_j \\ & & \beta_j & \alpha_j \end{bmatrix}.$$

*Received March 1999. Communicated by Axel Ruhe.

The eigenvalues of the intermediate tridiagonal matrices T_j of smaller dimension typically approximate some of the eigenvalues of A (often the ones largest in magnitude). During the iteration the matrix A is referenced only through matrix-vector products Ax and $A^T x$; hence the algorithm is useful for finding a few eigenvalues of very large and sparse matrices. A wide range of Lanczos papers appeared since the 1960s; see, e.g., the references in [7].

Recently, there has been considerable interest in structure-preserving Lanczos algorithms for the symplectic eigenproblem. These eigenproblems arise in applications like the problem of solving algebraic Riccati equations or H_∞-norm computations; see, e.g., [10, 12, 17]. In some of these applications the symplectic matrix is very large and sparse, and only a few eigenvalues and the corresponding invariant subspace are desired.

A structure-preserving Lanczos-like method for the symplectic eigenproblem was first proposed by Banse [2]. The symplectic matrix is reduced to a symplectic butterfly matrix. Banse presents a look-ahead version of the method which overcomes breakdown by giving up the strict butterfly form. Benner and Faßbender [3, 4] suggest to combine the idea of the symplectic Lanczos method with the idea of implicitly restarted Lanczos methods in order to deal with the numerical difficulties inherent to any nonsymmetric Lanczos-like method.

Here we give an error analysis of the symplectic Lanczos method for the symplectic eigenproblem. Numerical experiments show that, just like in the conventional Lanczos algorithm, information about the extreme eigenvalues tends to emerge long before the symplectic Lanczos process is completed. The effect of finite-precision arithmetic is discussed. Using Bai's work [1] on the nonsymmetric Lanczos algorithm, an analog of Paige's theory [13] on the relationship between the loss of orthogonality among the computed Lanczos vectors and the convergence of a Ritz value is discussed. The symplectic Lanczos algorithm is reviewed in Section 2. Stopping criteria are discussed. In Section 3 a rounding error analysis of the symplectic Lanczos algorithm in finite-precision arithmetic is presented. Section 4 discusses convergence of the symplectic Lanczos algorithm versus the loss of J-orthogonality of the computed Lanczos vectors. All proofs are deferred to the Appendix, due to their highly technical nature.

2 The symplectic Lanczos algorithm.

A matrix $M \in \mathbb{R}^{2n \times 2n}$ is called *symplectic* if

$$MJ^{2n,2n}M^T = J^{2n,2n} \tag{2.1}$$

(or equivalently, $M^T J^{2n,2n} M = J^{2n,2n}$), where

$$J^{2n,2n} = \begin{bmatrix} 0 & I^{n,n} \\ -I^{n,n} & 0 \end{bmatrix}, \tag{2.2}$$

and $I^{n,n}$ is the $n \times n$ identity matrix. If the dimension of $I^{n,n}$, or $J^{2n,2n}$, is clear from the context, we leave off the superscript. We denote by Z^{2k} the first $2k$ columns of a $2n \times 2n$ matrix Z.

The symplectic matrices form a group under multiplication. The eigenvalues of symplectic matrices occur in reciprocal pairs: if λ is an eigenvalue of M with right eigenvector x, then λ^{-1} is an eigenvalue of M with left eigenvector $(Jx)^T$.

In exact arithmetic and without breakdown, the symplectic Lanczos methods proposed by Banse [2] and Benner and Faßbender [4] reduce M to a symplectic butterfly matrix. A symplectic matrix

$$B = \begin{bmatrix} B_1 & B_2 \\ B_3 & B_4 \end{bmatrix} = \begin{bmatrix} \diagdown & \diagdown\!\!\diagdown\!\!\diagdown \\ \diagdown & \diagdown\!\!\diagdown\!\!\diagdown \end{bmatrix}$$

is called a *butterfly matrix* if $B_1, B_3 \in \mathbb{R}^{n\times n}$ are diagonal matrices and $B_2, B_4 \in \mathbb{R}^{n\times n}$ are tridiagonal matrices. An *unreduced butterfly matrix* is one for which the tridiagonal matrix B_4 is unreduced; see [4, 5]. Using the definition of a symplectic matrix, one easily verifies that if B is unreduced, then the diagonal submatrix B_3 is nonsingular. This allows the parameterization of B in the following form (see [4, 5]):

$$B = (K^{2n,2n})^{-1} N^{2n,2n}$$

$$= \left[\begin{array}{ccc|ccc} a_1^{-1} & & & b_1 & & \\ & \ddots & & & \ddots & \\ & & a_n^{-1} & & & b_n \\ \hline & & & a_1 & & \\ & & & & \ddots & \\ & & & & & a_n \end{array}\right] \left[\begin{array}{cccc|cccc} & & & & -1 & & & \\ & & & & & \ddots & & \\ & & & & & & \ddots & \\ & & & & & & & -1 \\ \hline 1 & & & & c_1 & d_2 & & \\ & \ddots & & & d_2 & \ddots & \ddots & \\ & & \ddots & & & \ddots & \ddots & d_n \\ & & & 1 & & & d_n & c_n \end{array}\right].$$

Given $s_1 \in \mathbb{R}^{2n}$ and a symplectic matrix $M \in \mathbb{R}^{2n\times 2n}$ the symplectic Lanczos algorithm generates a sequence of symplectic butterfly matrices $B^{2k,2k} \in \mathbb{R}^{2k\times 2k}$ such that (if no breakdown occurs)

$$MS^{2k} = S^{2k}B^{2k,2k} + r_{k+1}e_{2k}^T, \quad k = 1, 2, \ldots, n, \tag{2.3}$$

where $S^{2k} \in \mathbb{R}^{2n\times 2k}$, $S^{2k}e_1 = s_1$, and the columns of S^{2k} are orthogonal with respect to the indefinite inner product defined by J as in (2.2). That is, the columns of S^{2k} are J-orthogonal. The eigenvalues of the intermediate matrices $B^{2k,2k}$ are progressively better estimates of M's eigenvalues. For $k = n$ the algorithm computes a symplectic matrix S such that S transforms M into butterfly form: $S^{-1}MS = B$.

In order to simplify the notation we use in the following permuted versions of M, B, and S. Let

$$Z_P := PZP^T$$

with the permutation matrix

$$P := [e_1, e_3, \ldots, e_{2n-1}, e_2, e_4, \ldots, e_{2n}] \in \mathbb{R}^{2n\times 2n}.$$

Using the permutation matrix P, the matrix J can be permuted to the $2n \times 2n$ block diagonal matrix

$$J_P := PJP^T = \operatorname{diag}\left(\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \ldots, \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\right).$$

M_P, B_P, and S_P are permuted symplectic matrices, in other words, they are J_P-orthogonal.

Using the permuted versions of M_P, B_P, and S_P, the structure preserving Lanczos method generates a sequence of permuted symplectic matrices

$$S_P^{2k} := [v_1, w_1, v_2, w_2, \ldots, v_k, w_k] \in \mathbb{R}^{2n\times 2k}$$

satisfying

$$M_P S_P^{2k} = S_P^{2k} B_P^{2k,2k} + d_{k+1}(b_{k+1}v_{k+1} + a_{k+1}w_{k+1})e_{2k}^T. \tag{2.4}$$

The symplectic Lanczos algorithm for symplectic matrices is summarized in Table 2.1. For a derivation of the algorithm and a detailed discussion of various aspects; see [3, 4, 5]. There is some freedom in the choice of the parameters that occur in the algorithm. Essentially, the parameters b_k can be chosen freely. Here we set $b_k = 1$. A different choice of the parameters a_k and d_k is possible. Note that $M_P^{-1} = -J_P M_P^T J_P$, since M is symplectic. Thus $M_P^{-1}v_m$ is just a matrix-vector product with the transpose of M_P.

Equivalent to (2.4), as

$$B_P^{2k,2k} = (K_P^{2k,2k})^{-1} N_P^{2k,2k} \quad \text{and} \quad e_{2k}^T (N_P^{2k,2k})^{-1} = -e_{2k-1}^T,$$

we have

$$M_P S_P^{2k} (N_P^{2k,2k})^{-1} = S_P^{2k} (K_P^{2k,2k})^{-1} - d_{k+1}(b_{k+1}v_{k+1} + a_{k+1}w_{k+1})e_{2k-1}^T. \tag{2.5}$$

The vector $r_{k+1} := d_{k+1}(b_{k+1}v_{k+1} + a_{k+1}w_{k+1})$ is the *residual vector* and is J_P-orthogonal to the columns of S_P^{2k}, the *Lanczos vectors*. The matrix $B_P^{2k,2k}$ is the J_P-orthogonal projection of M_P onto the range of S_P^{2k}

$$B_P^{2k,2k} = J_P^{2k,2k} (S_P^{2k})^T J_P M_P S_P^{2k}.$$

REMARK 2.1. The usual nonsymmetric Lanczos algorithm generates two sequences of vectors $\{q_j\}$ and $\{p_j\}$ (see (1.1) and (1.2)). Adapting the usual nonsymmetric Lanczos algorithm to the situation considered here, the symplectic Lanczos process could have been stated as follows: Given $v_1, t_1 \in \mathbb{R}^{2n}$ and a

Table 2.1: Symplectic Lanczos Method for the Symplectic Eigenproblem.

<u>Algorithm : Symplectic Lanczos method</u>

Choose an initial vector $\widetilde{v}_1 \in \mathbb{R}^{2n}, \widetilde{v}_1 \neq 0$.
Set $v_0 = 0 \in \mathbb{R}^{2n}$.
Set $d_1 = ||\widetilde{v}_1||_2$ and $v_1 = \frac{1}{d_1}\widetilde{v}_1$.

for m = 1, 2, ... do
  (update of w_m)
    set

$$\widetilde{w}_m = M_P v_m - b_m v_m$$
$$a_m = v_m^T J_P M_P v_m$$
$$w_m = \frac{1}{a_m}\widetilde{w}_m$$

  (computation of c_m)

$$c_m = -a_m^{-1} w_m^T J_P M_P^{-1} v_m$$

  (update of v_{m+1})

$$\widetilde{v}_{m+1} = -d_m v_{m-1} - c_m v_m + w_m + a_m^{-1} M_P^{-1} v_m$$
$$d_{m+1} = ||\widetilde{v}_{m+1}||_2$$
$$v_{m+1} = \frac{1}{d_{m+1}}\widetilde{v}_{m+1}$$

symplectic matrix $M \in \mathbb{R}^{2n\times 2n}$, the symplectic Lanczos algorithm produces matrices $S_P^{2k} = [v_1,\ w_1,\ \ldots,\ v_k,\ w_k] \in \mathbb{R}^{2n\times 2k}$ and $W_P^{2k} = [t_1, \ldots, t_{2k}] \in \mathbb{R}^{2n\times 2k}$ with J_P-orthogonal columns which satisfy

$$(W_P^{2k})^T S_P^{2k} = I^{2k,2k},$$

and

$$\begin{aligned} M_P S_P^{2k} &= S_P^{2k} B_P^{2k,2k} + d_{k+1} r_{k+1} e_{2k}^T, \\ M_P^T W_P^{2k} &= W_P^{2k} (B_P^{2k,2k})^T + d_{k+1} \check{r}_{k+1} e_{2k}^T. \end{aligned}$$

As S_P is symplectic, we obtain from $(W_P^{2k})^T S_P^{2k} = I^{2k,2k}$ that

$$W_P^{2k} = J_P^{2n,2n} S_P^{2k} J_P^{2k,2k} = [-J_P w_1,\ J_P v_1, \ldots, -J_P w_k,\ J_P v_k].$$

Moreover,

$$r_{k+1} = M_P v_{k+1} \quad \text{and} \quad \check{r}_{k+1} = J_P v_{k+1}.$$

Substituting the expressions for W_P^{2k} and $\check{r}_{k+1}$ into the second recursion and pre- and post-multiplying with J_P yields that the two recursions are equivalent.

Hence one of the two sequences can be eliminated here and thus work and storage can essentially be halved. (This property is valid for a broader class of matrices; see [6].)

Assume that we have performed k steps of the symplectic Lanczos method and thus obtained the identity (after permuting back)

$$MS^{2k} = S^{2k}B^{2k,2k} + d_{k+1}(b_{k+1}\widehat{v}_{k+1} + a_{k+1}\widehat{w}_{k+1})e_{2k}^T.$$

If the norm of the residual vector is small, the $2k$ eigenvalues of $B^{2k,2k}$ are approximations to the eigenvalues of M. Numerical experiments indicate that the norm of the residual rarely becomes small by itself. Nevertheless, some eigenvalues of $B^{2k,2k}$ may be good approximations to eigenvalues of M. Let λ be an eigenvalue of $B^{2k,2k}$ with the corresponding eigenvector y. Then the vector $x = S^{2k}y$ satisfies

$$\begin{aligned} ||Mx - \lambda x||_2 &= ||(MS^{2k} - S^{2k}B^{2k,2k})y||_2 \\ &= |d_{k+1}|\,|e_{2k}^Ty|\,||b_{k+1}\widehat{v}_{k+1} + a_{k+1}\widehat{w}_{k+1}||_2. \end{aligned} \tag{2.6}$$

The vector x is referred to as *Ritz vector* and λ as *Ritz value* of M. If the last component of the eigenvector y is sufficiently small, the right-hand side of (2.6) is small and the pair $\{\lambda, x\}$ is a good approximation to an eigenvalue-eigenvector pair of M. Note that $|e_{2k}^Ty| > 0$ if $B^{2k,2k}$ is unreduced (see, e.g., [5, Lemma 3.11]). The pair $\{\lambda, x\}$ is exact for the nearby problem

$$(M+E)x = \lambda x \qquad \text{where} \quad E = -d_{k+1}(b_{k+1}\widehat{v}_{k+1} + a_{k+1}\widehat{w}_{k+1})e_k^T(S^{2k})^TJ^{2n,2n}.$$

In an actual implementation, typically the *Ritz estimate*

$$|d_{k+1}|\,|e_{2k}^Ty|\,||b_{k+1}\widehat{v}_{k+1} + a_{k+1}\widehat{w}_{k+1}||_2$$

is used in order to decide about the numerical accuracy of an approximate eigenpair. This avoids the explicit formation of the residual $(MS^{2k} - S^{2k}B^{2k,2k})y$.

A small Ritz estimate is not sufficient for the Ritz pair $\{\lambda, x\}$ to be a good approximation to an eigenvalue-eigenvector pair of M. It does not guarantee that λ is a good approximation to an eigenvalue of M. That is

$$\min_j |\lambda - \mu_j|, \quad \text{where } \mu_j \in \sigma(M) = \left\{\mu \in \mathbb{C} \mid \exists x \in \mathbb{R}^{2n}\backslash\left\{0\right\} \ni Mx = \mu x\right\}$$

is not necessarily small when the Ritz estimate is small (see, e.g., [9, Section 3]). For nonnormal matrices the norm of the residual of an approximate eigenvector is not by itself sufficient information to bound the error in the approximate eigenvalue. It is sufficient however to give a bound on the distance to the nearest matrix to which the Ritz triplet $\{\lambda, x, y\}$ is exact [9] (here y denotes the left Ritz vector of M corresponding to the Ritz value λ). In the following, we will give a computable expression for the error. Assume that $B^{2k,2k}$ is diagonalizable, i.e., there exists $Y \in \mathbb{C}^{2k\times 2k}$ such that

$$Y^{-1}B^{2k,2k}Y = \left[\begin{array}{ccc|ccc} \lambda_1 & & & & & \\ & \ddots & & & & \\ & & \lambda_k & & & \\ \hline & & & \lambda_1^{-1} & & \\ & & & & \ddots & \\ & & & & & \lambda_k^{-1} \end{array}\right] = \Lambda.$$

Let $X = S^{2k}Y = [x_1, \ldots, x_{2k}]$ and denote $b_{k+1}\widehat{v}_{k+1} + a_{k+1}\widehat{w}_{k+1}$ by $\widehat{r}_{k+1}$. Since

$$MS^{2k} = S^{2k}B^{2k,2k} + d_{k+1}\widehat{r}_{k+1}e_{2k}^T,$$

it follows that

$$MS^{2k}Y = S^{2k}YY^{-1}B^{2k,2k}Y + d_{k+1}\widehat{r}_{k+1}e_{2k}^T Y,$$

or

$$MX = X\Lambda + d_{k+1}\widehat{r}_{k+1}e_{2k}^T Y.$$

Thus

$$Mx_i = \lambda_i x_i + y_{2k,i}d_{k+1}\widehat{r}_{k+1},$$

and

$$Mx_{k+i} = \lambda_i^{-1}x_{k+i} + y_{2k,k+i}d_{k+1}\widehat{r}_{k+1},$$

for $i = 1, \ldots, k$. The last equation can be rewritten as

$$(Jx_{k+i})^T M = \lambda_i (Jx_{k+i})^T + y_{2k,k+i}\lambda_i d_{k+1}\widehat{r}_{k+1}^T JM.$$

Using Theorem 2′ of [9] we obtain that $\{\lambda_i, x_i, (Jx_{k+i})^T\}$ is an eigen-triplet of $M - F_{\lambda_i}$ where

$$||F_{\lambda_i}||_2 = |d_{k+1}| \max_i \left\{ \frac{|y_{2k,i}|\ ||\widehat{r}_{k+1}||_2}{||x_i||_2}, \frac{|y_{2k,k+i}\lambda_i|\ ||\widehat{r}_{k+1}^T JM||_2}{||Jx_{k+i}||_2} \right\}.$$

Furthermore, if $||F_{\lambda_i}||_2$ is small enough, then

$$|\theta_i - \lambda_j| \le \mathrm{cond}(\lambda_j)||F_{\lambda_i}||_2 + \mathcal{O}(||F_{\lambda_i}||_2^2),$$

where θ_i is an eigenvalue of M and $\mathrm{cond}(\lambda_j)$ is the condition number of the Ritz value λ_j

$$\mathrm{cond}(\lambda_j) = \frac{||x_i||_2||Jx_{k+i}||_2}{|x_{k+i}^T Jx_i|} = ||x_i||_2||x_{k+i}||_2.$$

Similarly, we obtain that $\{\lambda_i^{-1}, x_{k+i}, (Jx_i)^T\}$ is an eigen-triplet of $M - F_{\lambda_i^{-1}}$ where

$$||F_{\lambda_i^{-1}}||_2 = |d_{k+1}| \max_i \left\{ \frac{|y_{2k,k+i}|\ ||\widehat{r}_{k+1}||_2}{||x_{k+i}||_2}, \frac{|y_{2k,i}\lambda_i^{-1}|\ ||\widehat{r}_{k+1}^T JM||_2}{||Jx_i||_2} \right\}.$$

Consequently, as λ_i and λ_i^{-1} should be treated alike, the symplectic Lanczos algorithm should be continued until $||F_{\lambda_i}||_2$ and $||F_{\lambda_i^{-1}}||_2$ are small, and until $\mathrm{cond}(\lambda_j)||F_{\lambda_i}||_2$ and $\mathrm{cond}(\lambda_j)||F_{\lambda_i^{-1}}||_2$ are below a given threshold for accuracy. Note that as in the Ritz estimate, in the criteria derived here, the essential quantities are $|d_{k+1}|$ and the last component of the desired eigenvectors $|y_{2k,i}|$ and $|y_{2k,k+i}|$.

3 The symplectic Lanczos algorithm in finite-precision arithmetic.

In this section, we present a rounding error analysis of the symplectic Lanczos algorithm in finite-precision arithmetic. Our analysis will follow the lines of Bai's analysis of the nonsymmetric Lanczos algorithm [1]. It is in the spirit of Paige's analysis for the symmetric Lanczos algorithm [13], except that we (as Bai) carry out the analysis component-wise rather than norm-wise. The component-wise analysis allows to measure each element of a perturbation relative to its individual tolerance, so that, unlike in the norm-wise analysis, the sparsity pattern of the problem under consideration can be exploited.

We use the usual model of floating-point arithmetic, as, e.g., in [7, 8]:

$$fl(x \circ y) = (x \circ y)(1 + \varepsilon)$$

where $\circ$ denotes any of the four basic arithmetic operations $+, -, *, /$ and $|\varepsilon| \leq \mathbf{u}$ with $\mathbf{u}$ denoting the *unit roundoff.*

We summarize (as in [1]) all the results for basic linear algebra operations of sparse vectors and/or matrices that we need for our analysis:

Saxpy operation:

$$fl(\alpha x + y) = \alpha x + y + e, \quad |e| \leq \mathbf{u}\ (2|\alpha x| + |y|) + \mathcal{O}(\mathbf{u}^2),$$

Inner product:

$$fl(x^T y) = x^T y + e, \quad |e| \leq k\mathbf{u}\ |x|^T |y| + \mathcal{O}(\mathbf{u}^2),$$

Matrix-vector multiplication:

$$fl(Ax) = Ax + e, \quad |e| \leq m\mathbf{u}\ |A|\ |x| + \mathcal{O}(\mathbf{u}^2),$$

where k is the number of overlapping nonzero components in the vectors x and y, and m is the maximal number of nonzero elements of the matrix A in any row. For a vector $x = [x_1, \ldots, x_n]^T$, $|x|$ denotes the vector $[|x_1|, \ldots, |x_n|]^T$. Similar, for a matrix $A = [a_{ij}]_{i,j=1}^n$, $|A|$ denotes the $n \times n$ matrix $[|a_{ij}|]_{i,j=1}^n$.

In this section, any computed quantity will be denoted by a hat, e.g., $\widehat{\alpha}$ will denote a computed quantity that is affected by rounding errors. (Please note that in the previous section, we used hatted quantities to denote the non-permuted symplectic Lanczos vectors.)

Analyzing one step of the symplectic Lanczos algorithm to see the effects of the finite-precision arithmetic we obtain the following theorem:

THEOREM 3.1. *Let $M \in \mathbb{R}^{2n \times 2n}$ be a symplectic matrix with at most m nonzero entries in any row or column. If no breakdown occurs during the execution of k steps of the symplectic Lanczos algorithm as given in Table* 2.1*, the* *computed Lanczos vectors satisfy*

$$\widehat{a}_j \widehat{w}_j = M_P \widehat{v}_j - \widehat{v}_j + h_j, \tag{3.1}$$

$$\widehat{d}_{j+1} \widehat{v}_{j+1} = -\widehat{d}_j \widehat{v}_{j-1} - \widehat{c}_j \widehat{v}_j + \widehat{w}_j + \widehat{a}_j^{-1} M_P^{-1} \widehat{v}_j + g_{j+1}, \tag{3.2}$$

where

$$|h_j| \leq (m+2)\mathbf{u}\ |M_P|\ |\widehat{v}_j| + 2\mathbf{u}\ |\widehat{v}_j| + \mathcal{O}(\mathbf{u}^2), \tag{3.3}$$

$$|g_{j+1}| \leq (m+5)\mathbf{u}\ |\widehat{a}_j^{-1}|\ |M_P^{-1}|\ |\widehat{v}_j| + 4\mathbf{u}\ |\widehat{w}_j| + 4\mathbf{u}\ |\widehat{c}_j \widehat{v}_j| \tag{3.4}$$

$$+\, 3\mathbf{u}\ |\widehat{d}_j \widehat{v}_{j-1}| + \mathcal{O}(\mathbf{u}^2). \tag{3.5}$$

The computed matrices $\widehat{S}_P^{2k}, \widehat{N}_P^{2k,2k}$, *and* $\widehat{K}_P^{2k,2k}$ *satisfy*

$$M_P \widehat{S}_P^{2k} (\widehat{N}_P^{2k,2k})^{-1} = \widehat{S}_P^{2k} (\widehat{K}_P^{2k,2k})^{-1} - \widehat{d}_{k+1} M_P \widehat{v}_{k+1} e_{2k-1}^T + E_k, \tag{3.6}$$

where

$$\begin{aligned} ||E_k||_F \leq \mathbf{u}\ ||\widehat{S}^{2k}||_F \Big[&(m+5)\ ||\widehat{K}^{2k,2k}||_F ||M||_F^2 + 4\ ||\widehat{N}^{2k,2k}||_F ||M||_F \\ &+\ (m+6)\ ||M||_F + 2||\widehat{K}^{2k,2k}||_F \Big] + \mathcal{O}(\mathbf{u}^2). \end{aligned} \tag{3.7}$$

PROOF. See Appendix. □

This indicates that the recursion equation

$$M_P S_P^{2k} (N_P^{2k,2k})^{-1} = S_P^{2k} (K_P^{2k,2k})^{-1} - d_{k+1} M_P v_{k+1} e_{2k-1}^T$$

is satisfied to working precision, if $||\widehat{N}^{2k,2k}||_F ||M||_F$, $||M||_F^2$, $||\widehat{S}^{2k}||_F$, and $||\widehat{K}^{2k,2k}||_F$ are of moderate size. But, unfortunately, $||\widehat{S}^{2k}||_F$ may grow unboundedly in the case of near breakdown.

While equation (3.1) is given by the $(2j)$th column of $M_P S_P N_P^{-1} = S_P K_P^{-1}$, equation (3.2) corresponds to the $(2j-1)$th column of $S_P N_P^{-1} = M_P^{-1} S_P K_P^{-1}$. The upper bounds associated with (3.1) and (3.2) involve only $||M||_F$ as to be expected; see (3.3) and (3.5). Recall that $M_P^{-1} = -J_P M_P^T J_P$, since M is symplectic. Thus $|M_P^{-1}|$ does not introduce any problems usually involved by forming the inverse of a matrix. In order to summarize these results into one single equation, we define

$$E_k = [M_P g_2,\ -h_1,\ M_P g_3,\ -h_2, \ldots, M_P g_{k+1},\ -h_k]\,. \tag{3.8}$$

Then (3.6) holds. Using the component-wise upper bounds for $|h_j|$ and $|g_{j+1}|$, we obtain the upper bound for E_k as given in (3.7). As we summarize our results in terms of the equation $M_P S_P N_P^{-1} = S_P K_P^{-1}$, we have to pre-multiply the error bound associated with (3.2) by M_P, resulting in an artificial $||M||_F^2$ term here. Hence combining all our findings into one single equation forces the $||M||_F^2$ term.

For the nonsymmetric Lanczos algorithm, Bai obtains a similar result in [1]. The equations corresponding to our equations (3.1) and (3.2) are (see (1.1) and (1.2))

$$\begin{aligned}\widehat{\beta}_{j+1}\widehat{q}_{j+1} &= A\widehat{q}_j - \widehat{\alpha}_j\widehat{q}_j - \widehat{\gamma}_j\widehat{q}_j + h_j^{nonsymLan},\\ \widehat{\gamma}_{j+1}\widehat{p}_{j+1} &= A^T\widehat{p}_j - \widehat{\alpha}_j\widehat{p}_j - \widehat{\beta}_j\widehat{p}_{j-1} + g_{j+1}^{nonsymLan}.\end{aligned}$$

The errors associated are given by

$$\begin{aligned}\left|h_j^{nonsymLan}\right| &\leq (3+m)\mathbf{u}\ |A|\ |\widehat{q}_j| + 4\mathbf{u}\ |\widehat{\alpha}_j|\ |\widehat{q}_j| + 3\mathbf{u}\ |\widehat{\gamma}_j|\ |\widehat{q}_{j-1}| + \mathcal{O}(\mathbf{u}^2),\\ \left|g_{j+1}^{nonsymLan}\right| &\leq (3+m)\mathbf{u}\ |A|\ |\widehat{p}_j| + 4\mathbf{u}\ |\widehat{\alpha}_j|\ |\widehat{p}_j| + 3\mathbf{u}\ |\widehat{\gamma}_j|\ |\widehat{p}_{j-1}| + \mathcal{O}(\mathbf{u}^2).\end{aligned}$$

Hence, the symplectic Lanczos algorithms behaves essentially like the nonsymmetric Lanczos algorithm. The additional restriction of preserving the symplectic structure does not pose any additional problems concerning the rounding error analysis, the results of the analysis are essentially the same.

Remark 3.1. In Remark 2.1 we have noted that the usual nonsymmetric Lanczos algorithm generates two sequences of vectors, but that due to the symplectic structure, the two recurrence relations of the standard nonsymmetric Lanczos algorithm are equivalent for the situation discussed here. It was noted that the equation which is not used is given by

$$M_P^T W_P^{2k} (K_P^{2k,2k})^T = W_P^{2k} (N_P^{2k,2k})^T + d_{k+1} J_P v_{k+1} e_{2k}^T,$$

where

$$W_P^{2k} = J_P^{2n,2n} S_P^{2k} J_P^{2k,2k} = \left[-J_P w_1,\ J_P v_1, \ldots, -J_P w_k,\ J_P v_k\right].$$

Instead of summarizing our findings into equation (3.6), we could have summarized

$$M_P^T \widehat{W}_P^{2k} (\widehat{K}_P^{2k,2k})^T = \widehat{W}_P^{2k} (\widehat{N}_P^{2k,2k})^T + \widehat{d}_{k+1} J_P \widehat{v}_{k+1} e_{2k}^T + F_k \tag{3.9}$$

where

$$\begin{aligned}\widehat{W}_P^{2k} &= J_P^{2n,2n} \widehat{S}_P^{2k} J_P^{2k,2k},\\ F_k &= \left[M_P^T J_P h_1,\ J_P g_2, \ldots, M_P^T J_P h_k,\ J_P g_{k+1}\right].\end{aligned} \tag{3.10}$$

As an upper bound for $||F_k||_F$ we obtain

$$\begin{aligned}||F_k||_F \leq \mathbf{u}\ ||\widehat{S}^{2k}||_F \Big[&(m+2)\ ||M||_F^2 + (m+7)\ ||\widehat{K}^{2k,2k}||_F ||M||_F\\ &+ 4\ ||\widehat{N}^{2k,2k}||_F + 4\Big] + \mathcal{O}(\mathbf{u}^2).\end{aligned} \tag{3.11}$$

As before, the term $||M||_F^2$ is introduced because we summarize all our findings into one single equation.

It is well-known that in finite-precision arithmetic, orthogonality between the computed Lanczos vectors in the symmetric Lanczos process is lost. This loss of orthogonality is due to cancellation and is not the result of the gradual accumulation of roundoff error (see, e.g., [15, 16]). What can we say about the J-orthogonality of the computed symplectic Lanczos vectors? Obviously, rounding errors, once introduced into some computed Lanczos vectors, are propagated to future steps. Such error propagation for the nonsymmetric Lanczos process is analyzed by Bai [1].

Let us take a closer look at the J_P-orthogonality of the computed symplectic Lanczos vectors. Define

$$L = [\widehat{v}_1, \widehat{w}_1, \ldots, \widehat{v}_k, \widehat{w}_k]^T J_P [\widehat{v}_1, \widehat{w}_1, \ldots, \widehat{v}_k, \widehat{w}_k].$$

That is,

$$\begin{aligned} \ell_{2j-1,2m-1} &= \widehat{v}_j^T J_P \widehat{v}_m, & \ell_{2j-1,2m} &= \widehat{v}_j^T J_P \widehat{w}_m, \\ \ell_{2j,2m-1} &= \widehat{w}_j^T J_P \widehat{v}_m, & \ell_{2j,2m} &= \widehat{w}_j^T J_P \widehat{w}_m. \end{aligned}$$

In exact arithmetic we would have $L = J_P$, where J_P is block diagonal; each diagonal block is of the form $\left[\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right]$. As $x^T J_P x = 0$ for any vector x, we have

$$\ell_{2j,2j} = \ell_{2j-1,2j-1} = 0,$$

not depending on the loss of J_P-orthogonality between the computed symplectic Lanczos vectors. Examining the other elements of K we obtain the following lemma:

LEMMA 3.2. *The elements ℓ_{jm} of L satisfy the following equations:*

$$\begin{aligned} \ell_{jj} &= 0, & j &= 1, \ldots, 2k, \\ -\ell_{j,j+1} = \ell_{j+1,j} &= -1 + \kappa_j, & j &= 1, \ldots, 2k-1, \end{aligned} \tag{3.12}$$

where

$$|\kappa_j| \le \mathbf{u} \, \frac{|\widehat{\widetilde{v}}_j|^T |J_P| \, \{2(m+n+2) \, |M_P| + 5I\} \, |\widehat{v}_j|}{|\widehat{\widetilde{w}}_j^T J_P \widehat{\widetilde{v}}_j|} + \mathcal{O}(\mathbf{u}^2), \tag{3.13}$$

and

$$\begin{aligned} \widehat{d}_m \widehat{a}_m \ell_{2j,2m} = {} & \widehat{a}_j^{-1} \ell_{2j-1,2m-2} - \widehat{d}_m \ell_{2j,2m-1} - \widehat{c}_{m-1} \widehat{a}_{m-1} \ell_{2j,2m-2} - \widehat{c}_{m-1} \ell_{2j,2m-3} \\ & + \widehat{a}_{m-1}^{-1} \ell_{2j,2m-3} - \widehat{d}_{m-1} \widehat{a}_{m-2} \ell_{2j,2m-4} - \widehat{d}_{m-1} \ell_{2j,2m-5} \\ & + \widehat{d}_j \ell_{2m-2,2j-3} + \widehat{c}_j \ell_{2m-2,2j-1} + \widehat{d}_{j+1} \ell_{2m-2,2j+1} + \ell_{2m-2,2j} \\ & + \widehat{d}_{m-1} \widehat{w}_j^T J_P h_{m-2} + \widehat{c}_{m-1} \widehat{w}_j^T J_P h_{m-1} + \widehat{d}_m \widehat{w}_j^T J_P h_m \\ & + \widehat{a}_j^{-1} h_j^T J_P M_P \widehat{w}_{m-1} - \widehat{w}_{m-1}^T J_P g_{j+1} + \widehat{w}_j^T J_P M_P g_m. \end{aligned} \tag{3.14}$$

Similar expressions can be derived for $\ell_{2j,2m-1}$, and $\ell_{2j-1,2m-1}$.

PROOF. See Appendix. □

Table 3.1: Upper bound for $|\kappa_m|$ from Lemma 3.2, random starting vector.

m	bound for $\lvert\kappa_m\rvert$	$\lvert\ell_{m+1,m}+1\rvert$	$\lvert\ell_{m,m+1}-1\rvert$
1	$2.8599 \cdot 10^{-14}$	$2.2204 \cdot 10^{-16}$	$4.4409 \cdot 10^{-16}$
2	$8.3207 \cdot 10^{-14}$	$2.2204 \cdot 10^{-16}$	$1.1102 \cdot 10^{-16}$
3	$7.0045 \cdot 10^{-14}$	$2.2204 \cdot 10^{-16}$	$1.1102 \cdot 10^{-16}$
4	$5.5521 \cdot 10^{-14}$	$2.2204 \cdot 10^{-16}$	$2.2204 \cdot 10^{-16}$
5	$6.6807 \cdot 10^{-14}$	$2.2204 \cdot 10^{-16}$	$2.2204 \cdot 10^{-16}$
6	$6.4046 \cdot 10^{-14}$	$2.2204 \cdot 10^{-16}$	$2.2204 \cdot 10^{-16}$
7	$6.8944 \cdot 10^{-14}$	$2.2204 \cdot 10^{-16}$	$2.2204 \cdot 10^{-16}$
8	$5.2126 \cdot 10^{-14}$	$2.2204 \cdot 10^{-16}$	$2.2204 \cdot 10^{-16}$
9	$1.0334 \cdot 10^{-13}$	$2.2204 \cdot 10^{-16}$	$3.3307 \cdot 10^{-16}$
10	$2.0037 \cdot 10^{-13}$	$3.3307 \cdot 10^{-16}$	$3.3307 \cdot 10^{-16}$
11	$2.7005 \cdot 10^{-11}$	$3.3307 \cdot 10^{-16}$	$3.5527 \cdot 10^{-15}$
12	$3.6116 \cdot 10^{-14}$	$3.5527 \cdot 10^{-15}$	$5.3291 \cdot 10^{-15}$
13	$4.3575 \cdot 10^{-14}$	$3.5527 \cdot 10^{-15}$	$5.3291 \cdot 10^{-15}$

Lemma 3.2 describes how J-orthogonality between the computed symplectic Lanczos vectors is lost. Especially, (3.14) shows how the error is propagated to future steps. Moreover, in case $\widehat{d}_m$ and/or $\widehat{a}_m$ is tiny (which may indicate that either a J_P-orthogonal $2m$ dimensional invariant subspace of M_P or an invariant subspace of dimension $2m-1$ has been found) then we can expect $\ell_{2j,2m}$ to become large.

Example 3.1. In order to illustrate the findings of Lemma 3.2 some numerical experiments were done using a 100×100 symplectic block-diagonal matrix

$$M = \operatorname{diag}\bigl(D, D^{-1}\bigr), \qquad D = \operatorname{diag}\bigl(200, 100, 50, 47, \ldots, 4, 3, \left[\begin{smallmatrix} 2 & 1 \\ -1 & 2 \end{smallmatrix}\right]\bigr). \tag{3.15}$$

All computations were done using Matlab[1] Version 5.3 on a Sun Ultra 1 with IEEE double-precision arithmetic and machine precision $\epsilon = 2.2204 \times 10^{-16}$. Our code implements exactly the algorithm as given in Table 2.1.

The symplectic Lanczos process generates a sequence of symplectic butterfly matrices $B^{2k,2k}$ whose eigenvalues are increasingly better approximates to eigenvalues of M. The largest Ritz value approximates the largest eigenvalue $\lambda_1 = 200$ of M.

For the first set of tests a random starting vector v_1 was used. Table 3.1 lists the upper bound κ_m for the deviation of $\ell_{m+1,m}$ from 1 and of $\ell_{m,m+1}$ from -1 for $m = 1, \ldots, 13$. Due to roundoff errors in the computation of L these deviations are not the same, as they should be theoretically. The bound for $|\kappa_m|$ is typically one order of magnitude larger than the computed values of κ_m.

The propagation of the roundoff error in L, described by (3.14), can nicely be seen in Figure 3.1. In order to follow the error propagation we have computed $Z = L - J_P$. In each step of the symplectic Lanczos method two symplectic

[1] Matlab is a registered trademark of The MathWorks, Inc.

Lanczos vectors are computed. Hence looking at the principal submatrices of Z of dimension $2m, m = 2, 3, 4, \ldots,$ we can follow the error propagation as these submatrices grow by two additional rows and columns representing the error associated with the two new Lanczos vectors. In Figure 3.1 the absolute values of the entries of the principal submatrices of Z of dimension $2m, m = 3, 4, 5, 6,$ are shown. For $m = 3$, the entries of the 6×6 principal submatrix of Z are of the order of 10^{-16}. The same is true for the entries of the 8×8 principal submatrix, but it can be seen that the error in the newly computed entries is slightly larger than before. The next two figures for $m = 5$ and $m = 6$ show that the error associated with the new computed entries is increasing slowly.

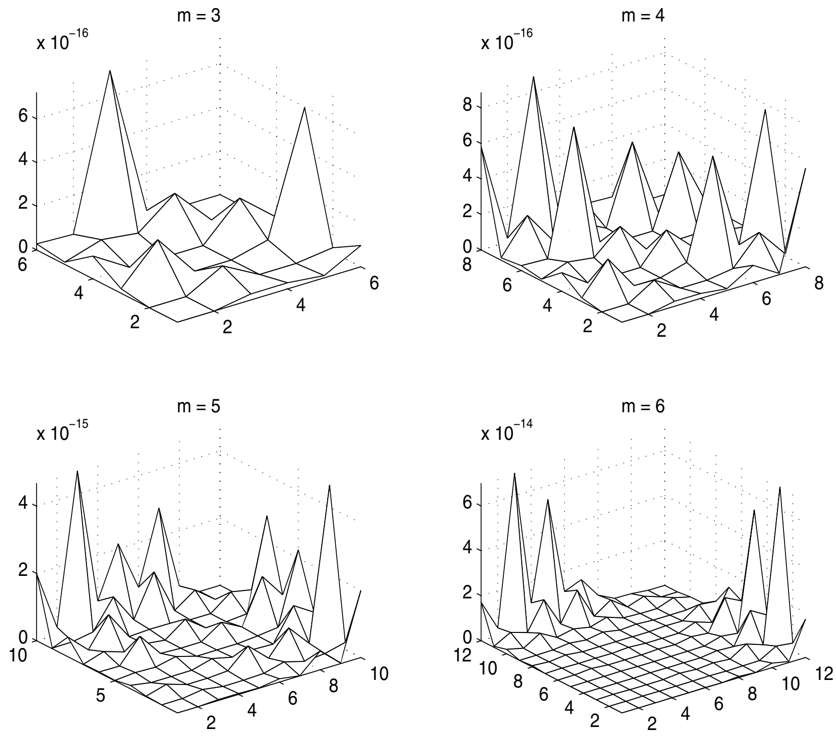

Figure 3.1: Error propagation, random starting vector.

The next test reported here was done using the starting vector

$$v_1 = [1, 1, 10^{-11}, \ldots, 10^{-11}]^T \in \mathbb{R}^{100}.$$

This starting vector is close to the sum of the eigenvectors corresponding to the largest and the smallest eigenvalue of M_P. Hence, it can be expected that an invariant subspace corresponding to these eigenvalues is detected soon. Table 3.2 and Figure 3.2 give the same information as the Table 3.1 and Figure 3.1.

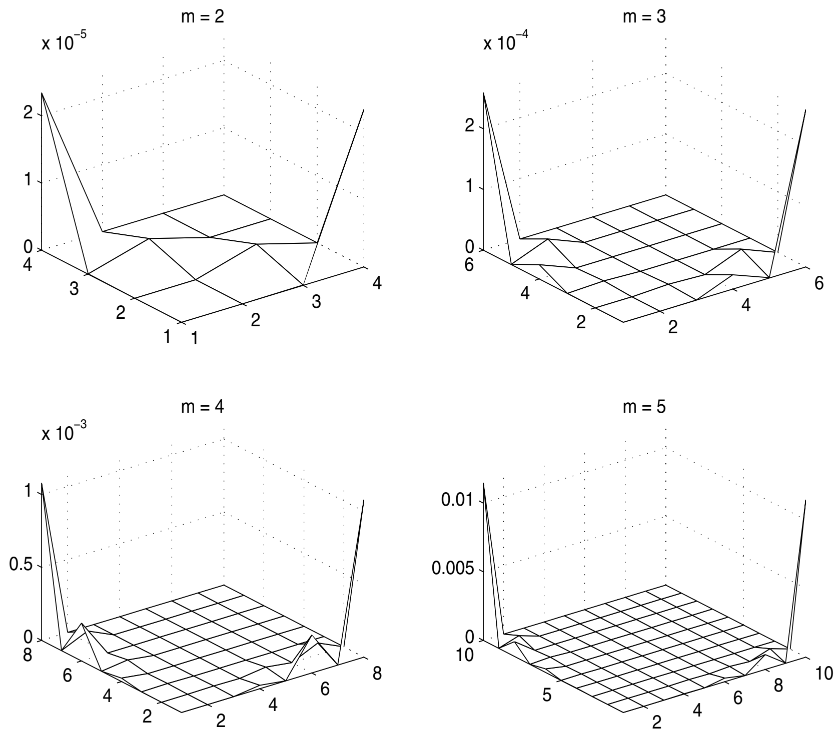

Figure 3.2: Error propagation, special starting vector.

Obviously, the J-orthogonality is lost already after two steps of the symplectic Lanczos method. This has almost no effect on the computed values for κ_m.

The next section will discuss loss of J-orthogonality versus convergence. It will be seen that under certain conditions, loss of J-orthogonality is accompanied by the convergence of Ritz values, just as in the last example.

4 Convergence versus loss of *J*-orthogonality.

It is well-known that in the symmetric Lanczos procedure, loss of orthogonality between the computed Lanczos vectors implies convergence of a Ritz pair to an eigenpair; see, e.g., [14]. Here we will discuss the situation for the symplectic Lanczos algorithm, following the lines of Section 4 of Bai's analysis of the nonsymmetric Lanczos algorithm in [1]. We will see that a conclusion similar to the one for the symmetric Lanczos process holds here, subject to a certain condition.

From the previous section, we know that the computed symplectic Lanczos

Table 3.2: Upper bound for $|\kappa_m|$ from Lemma 3.2, special starting vector.

j	bound for $\|\kappa_j\|$	$\|\ell_{j+1,j}+1\|$	$\|\ell_{j,j+1}-1\|$
1	$4.4787 \cdot 10^{-14}$	$1.1102 \cdot 10^{-16}$	$1.1102 \cdot 10^{-16}$
2	$7.1240 \cdot 10^{-04}$	$2.2204 \cdot 10^{-16}$	$1.1102 \cdot 10^{-16}$
3	$1.2347 \cdot 10^{-13}$	$3.3307 \cdot 10^{-16}$	$1.1102 \cdot 10^{-16}$
4	$5.4242 \cdot 10^{-14}$	$3.3307 \cdot 10^{-16}$	$4.4409 \cdot 10^{-16}$
5	$6.6991 \cdot 10^{-14}$	$3.3307 \cdot 10^{-16}$	$4.4409 \cdot 10^{-16}$
6	$6.6742 \cdot 10^{-14}$	$3.3307 \cdot 10^{-16}$	$4.4409 \cdot 10^{-16}$
7	$1.1444 \cdot 10^{-13}$	$3.3307 \cdot 10^{-16}$	$4.4409 \cdot 10^{-16}$
8	$4.9490 \cdot 10^{-14}$	$3.3307 \cdot 10^{-16}$	$4.4409 \cdot 10^{-16}$
9	$6.0268 \cdot 10^{-14}$	$3.3307 \cdot 10^{-16}$	$4.4409 \cdot 10^{-16}$
10	$8.5459 \cdot 10^{-14}$	$3.3307 \cdot 10^{-16}$	$4.4409 \cdot 10^{-16}$
11	$1.4171 \cdot 10^{-13}$	$3.3307 \cdot 10^{-16}$	$4.4409 \cdot 10^{-16}$
12	$2.0743 \cdot 10^{-10}$	$2.1316 \cdot 10^{-14}$	$4.4409 \cdot 10^{-16}$
13	$3.5934 \cdot 10^{-14}$	$4.9738 \cdot 10^{-14}$	$5.6843 \cdot 10^{-14}$

vectors obey the following equalities:

$$M_P \widehat{S}_P^{2k} = \widehat{S}_P^{2k} \widehat{B}_P^{2k,2k} - \left[\widehat{d}_{k+1}\widehat{r}_{k+1}e_{2k-1}^T - E_k\right] \widehat{N}_P^{2k,2k}, \tag{4.1}$$

$$M_P^T \widehat{W}_P^{2k} = \widehat{W}_P^{2k} (\widehat{B}_P^{2k,2k})_P^T + \left[\widehat{d}_{k+1} J_P \widehat{v}_{k+1} e_{2k}^T + F_k\right] (\widehat{K}_P^{2k,2k})^{-T}, \tag{4.2}$$

with

$$(\widehat{S}_P^{2k})^T J_P^{2n,2n} \widehat{S}_P^{2k} = K = J_P^{2k,2k} + C_k + \Delta_k - C_k^T, \tag{4.3}$$

where $\widehat{B}_P^{2k,2k} = (\widehat{K}_P^{2k,2k})^{-1}\widehat{N}_P^{2k,2k}$, $\widehat{W}_P^{2k} = J_P^{2n,2n}\widehat{S}_P^{2k}J_P^{2k,2k}$, the rounding error matrices E_k and F_k are as in (3.8) and, (3.10), resp., Δ_k is a block diagonal matrix with 2×2 block on the diagonal,

$$\Delta_k = \operatorname{diag}\left(\begin{bmatrix} 0 & \kappa_1 \\ -\kappa_1 & 0 \end{bmatrix}, \ldots, \begin{bmatrix} 0 & \kappa_k \\ -\kappa_k & 0 \end{bmatrix}\right),$$

and C_k is a strictly lower block triangular matrix with block size 2. That is $(C_k)_{\ell,j} = 0$ for $\ell = 1, \ldots, 2k, j = \ell, \ldots, 2k$, and $(C_k)_{2\ell,2\ell-1} = 0$ for $\ell = 1, \ldots, k$.

To simplify our discussion, we make two assumptions, which are also used in the analysis of the symmetric Lanczos process [15, p. 265] and in the analysis of the nonsymmetric Lanczos process [1]. The first assumption is *local J-orthogonality*, that is, the computed symplectic Lanczos vectors are J-orthogonal to their neighboring Lanczos vectors:

$$\begin{bmatrix} \widehat{v}_{j-1}^T \\ \widehat{w}_{j-1}^T \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} [\widehat{v}_j \ \widehat{w}_j] = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}. \tag{4.4}$$

This implies that the 2×2 block on the sub-diagonal of C_k are zero, yielding the following block-structure:

$$C_k = \begin{bmatrix} 0 & 0 & 0 & \cdots & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & \cdots & 0 & 0 & 0 \\ X & 0 & 0 & \cdots & \cdots & 0 & 0 & 0 \\ X & X & 0 & \cdots & \cdots & 0 & 0 & 0 \\ X & X & X & \ddots & & 0 & 0 & 0 \\ & \vdots & & \ddots & \ddots & & \vdots & \\ X & X & X & \cdots & X & 0 & 0 & 0 \\ X & X & X & \cdots & X & X & 0 & 0 \end{bmatrix},$$

where the X denote 2×2 blocks.

The second assumption is that the eigenvalue problem for the $2k \times 2k$ butterfly matrix $\widehat{B}_P^{2k,2k} = (\widehat{K}_P^{2k,2k})^{-1} \widehat{N}_P^{2k,2k}$ is solved exactly, that is,

$$Y_k^{-1} \widehat{B}_P^{2k,2k} Y_k = \operatorname{diag}\big(\lambda_1, \lambda_1^{-1}, \ldots, \lambda_k, \lambda_k^{-1}\big). \tag{4.5}$$

This implies that the computed Ritz vector for λ_j is given by $z_j = \widehat{S}_P^{2k} y_{2j-1}$, while the computed Ritz vector for λ_j^{-1} is given by $x_j = \widehat{S}_P^{2k} y_{2j}$.

THEOREM 4.1. *Assume that the symplectic Lanczos algorithm in finite-precision arithmetic satisfies* (4.1)–(4.3). *Let*

$$\begin{aligned} L_k^{(2)} + U_k^{(2)} &= J_P^{2k,2k} \Delta_k \widehat{B}_P^{2k,2k} - \widehat{B}_P^{2k,2k} J_P^{2k,2k} \Delta_k, \\ L_k^{(4)} + U_k^{(4)} &= (\widehat{K}_P^{2k,2k})^{-1} F_k^T \widehat{S}_P^{2k} - (\widehat{W}_P^{2k})^T E_k \widehat{N}_P^{2k,2k}, \end{aligned}$$

where $L_k^{(2)}$ and $L_k^{(4)}$ are strictly lower block triangular matrices, and $U_k^{(2)}$ and $U_k^{(4)}$ are strictly upper block triangular matrices with block size 2. *Then the computed Ritz vectors $x_j = \widehat{S}_P^{2k} y_{2j}$ and $z_j = \widehat{S}_P^{2k} y_{2j-1}$ satisfy*

$$x_j^T J_P^{2n,2n} \widehat{r}_{k+1} = \frac{y_{2j}^T J_P^{2k,2k} \left[U_k^{(4)} - U_k^{(2)}\right] y_{2j-1}}{\widehat{d}_{k+1}(e_{2k}^T y_{2j-1})} =: \frac{\psi_1}{\widehat{d}_{k+1}(e_{2k}^T y_{2j-1})}, \tag{4.6}$$

$$z_j^T J_P^{2n,2n} \widehat{r}_{k+1} = \frac{y_{2j-1}^T J_P^{2k,2k} \left[U_k^{(4)} - U_k^{(2)}\right] y_{2j}}{\widehat{d}_{k+1}(e_{2k}^T y_{2j})} =: \frac{\psi_2}{\widehat{d}_{k+1}(e_{2k}^T y_{2j})}. \tag{4.7}$$

PROOF. See Appendix. ∎

The derived equations are similar to those obtained by Bai for the nonsymmetric Lanczos process. Hence we can interpret our findings analogously: equations (4.6) and (4.7) describe the way in which the J-orthogonality is lost. Recall that the scalar d_{k+1} and the last eigenvector components $(e_{2k}^T y_{2j-1})$ and $(e_{2k}^T y_{2j})$ are also essential quantities used as the backward error criteria for the computed Ritz triplets $\{\lambda_i, z_i, (Jx_i)^T\}$ and $\{\lambda_i^{-1}, x_i, (Jz_i)^T\}$ discussed in Section 2. (Also recall that $|e_{2k}^T y_\ell| > 0$ if $B^{2k,2k}$ is unreduced.) Hence, if the quantities $|\psi_1|$ and

Table 4.1: Loss of J-orthogonality versus convergence of Ritz value, random starting vector.

Lanczos step	$z_1^T J_P \widehat{r}_{k+1}$	$\widehat{d}_{k+1}(e_{2k}^T y_2)$
1	$-9.1290 \cdot 10^{-17}$	$-6.4278 \cdot 10^{-01}$
2	$1.6751 \cdot 10^{-17}$	$3.5949 \cdot 10^{-01}$
3	$-8.4297 \cdot 10^{-18}$	$-8.1016 \cdot 10^{-02}$
4	$2.6983 \cdot 10^{-17}$	$-1.7984 \cdot 10^{-02}$
5	$2.8513 \cdot 10^{-16}$	$-1.3822 \cdot 10^{-03}$
6	$4.7089 \cdot 10^{-15}$	$-8.3119 \cdot 10^{-05}$
7	$6.8569 \cdot 10^{-14}$	$-5.7074 \cdot 10^{-06}$
8	$-8.3995 \cdot 10^{-13}$	$-4.6590 \cdot 10^{-07}$
9	$-9.3850 \cdot 10^{-12}$	$-4.1698 \cdot 10^{-08}$
10	$9.0525 \cdot 10^{-11}$	$-4.3229 \cdot 10^{-09}$
11	$-4.1822 \cdot 10^{-10}$	$9.3571 \cdot 10^{-10}$
12	$6.8361 \cdot 10^{-09}$	$-5.7230 \cdot 10^{-11}$
13	$-2.9881 \cdot 10^{-07}$	$1.3010 \cdot 10^{-12}$
14	$5.7946 \cdot 10^{-06}$	$6.5210 \cdot 10^{-14}$
15	$-1.0299 \cdot 10^{-04}$	$2.6478 \cdot 10^{-15}$
16	$1.5128 \cdot 10^{-03}$	$-1.0915 \cdot 10^{-15}$

$|\psi_2|$ are bounded and bounded away from zero, then (4.6) and (4.7) reflect the reciprocal relation between the convergence of the symplectic Lanczos process (i.e., tiny $\widehat{d}_{k+1}(e_{2k}^T y_{2j-1})$ and $\widehat{d}_{k+1}(e_{2k}^T y_{2j})$) and the loss of J-orthogonality (i.e., large $\widehat{r}_{k+1}^T J_P x_j$ and $\widehat{r}_{k+1}^T J_P z_j$).

Example 4.1. Here we continue the numerical tests with the test matrix (3.15). The first test reported was done using a random starting vector v_1. Table 4.1 illustrates the loss of J-orthogonality among the symplectic Lanczos vectors in terms of $z_1^T J_P \widehat{r}_{k+1}$ and the convergence of a Ritz value in terms of the residual $\widehat{d}_{k+1}(e_{2k}^T y_2)$. As predicted by Theorem 4.1, the loss of J-orthogonality accompanies the convergence of a Ritz value to the largest eigenvalue λ_1 (and the convergence of a Ritz value to the smallest eigenvalue λ_1^{-1}) in terms of small residuals.

When the symplectic Lanczos process is stopped at $k = 16$, the computed largest Ritz value λ_1 has the relative accuracy

$$\frac{|200 - \lambda_1|}{200} \approx 1.5632 \cdot 10^{-15}.$$

We note that in this example, the Ritz value corresponding to the largest eigenvalue of M is well conditioned, while the condition number for all eigenvalues of M is one, the condition number of the largest Ritz value is ≈ 1.08. The results for $w_1^T J_P \widehat{r}_{k+1}$ and $\widehat{d}_{k+1}(e_{2k}^T y_1)$ are almost the same.

Using the special starting vector $v_1 = [1, 1, 10^{-11}, \ldots, 10^{-11}]^T$, the results presented in Table 4.2 are obtained. As already seen in Example 3.1, J-orthogonality

Table 4.2: Loss of J-orthogonality versus convergence of Ritz value, special starting vector.

Lanczos step	$z_1^T J_P \widehat{r}_{k+1}$	$\widehat{d}_{k+1}(e_{2k}^T y_2)$
1	$1.2839 \cdot 1^{-08}$	$-5.4906 \cdot 1^{-11}$
2	$1.1915 \cdot 1^{-07}$	$-5.9165 \cdot 1^{-12}$
3	$4.0409 \cdot 1^{-07}$	$1.7423 \cdot 1^{-12}$
4	$-2.3952 \cdot 1^{-06}$	$-2.9338 \cdot 1^{-13}$
5	$-3.0405 \cdot 1^{-05}$	$2.3785 \cdot 1^{-14}$
6	$-4.0010 \cdot 1^{-04}$	$3.2861 \cdot 1^{-15}$
7	$1.5373 \cdot 1^{-03}$	$-1.1173 \cdot 1^{-14}$

is lost fast. This is accompanied by the convergence of a Ritz value to the largest eigenvalue λ_1 (and the convergence of a Ritz value to the smallest eigenvalue λ_1^{-1}) in terms of small residuals.

When the symplectic Lanczos process is stopped at $k = 10$, the computed butterfly matrix has two eigenvalues close to 200 and one eigenvalue close to 100. Hence the loss of J-orthogonality results, as in the standard nonsymmetric Lanczos algorithm, in ghost eigenvalues. That is, multiple eigenvalues of $B^{2k,2k}$ correspond to simple eigenvalues of M. The eigenvalues close to 200 have relative accuracy $1.4211 \cdot 10^{-16}$, resp. $4.0767 \cdot 10^{-07}$ for its ghost, the one close to 100 has the relative accuracy $9.4163 \cdot 10^{-10}$.

Let us conclude our analysis by estimating $|\psi_1|$ and $|\psi_2|$. Let us assume (again analogous to Bai's analysis) that $\Delta_k = 0$, i.e., $\widehat{w}_j^T J_P \widehat{v}_j = -1$, which simplifies the technical details of the analysis and appears to be the case in practice, up to the order of machine precision. Under this assumption, we have $U_k^{(2)} = 0$. Moreover, we have $|\psi_\ell| \leq ||U_k^{(4)}||_F ||y_{2j}||_2 ||y_{2j-1}||_2$, for $\ell = 1, 2$. Let us derive an estimate for $||U_k^{(4)}||_F$. $U_k^{(4)}$ is the strictly upper block triangular part of

$$(\widehat{K}_P^{2k,2k})^{-1} F_k^T \widehat{S}_P^{2k} - (\widehat{W}_P^{2k})^T E_k \widehat{N}_P^{2k,2k}. \tag{4.8}$$

A generous upper bound is therefore given by

$$\begin{aligned}
||U_k^{(4)}||_F &\leq ||\widehat{K}^{2k,2k}||_F ||F_k^T||_F ||\widehat{S}^{2k}||_F + ||\widehat{W}^{2k}||_F ||E_k||_F ||\widehat{N}^{2k,2k}||_F \\
&\leq ||\widehat{S}^{2k}||_F \Big[||\widehat{K}^{2k,2k}||_F ||F_k||_F + ||E_k||_F ||\widehat{N}^{2k,2k}||_F \Big] \\
&\leq \mathbf{u} ||\widehat{S}^{2k}||_F^2 \Big\{ (m+5) ||\widehat{K}^{2k,2k}||_F ||\widehat{N}^{2k,2k}||_F ||M||_F^2 \\
&\quad + 7 ||\widehat{K}^{2k,2k}||_F ||\widehat{N}^{2k,2k}||_F + 4 ||\widehat{K}^{2k,2k}||_F + (m+2) ||\widehat{K}^{2k,2k}||_F ||M||_F^2 \\
&\quad + (m+8) ||\widehat{K}^{2k,2k}||_F^2 ||M||_F + 4 ||\widehat{N}^{2k,2k}||_F^2 ||M||_F \\
&\quad + (m+6) ||\widehat{N}^{2k,2k}||_F ||M||_F \Big\} + \mathcal{O}(\mathbf{u}^2).
\end{aligned}$$

Summarizing, we obtain the following corollary, which gives an upper bound for $|\psi_1|$ and $|\psi_2|$:

COROLLARY 4.2. *Assume that* $\Delta_k = 0$ *in Theorem* 4.1. *Then*

$$\begin{aligned}|\psi| \leq \mathbf{u}\,\mathrm{cond}(\lambda_j)\Big\{&(m+5)||\widehat{K}^{2k,2k}||_F||\widehat{N}^{2k,2k}||_F||M||_F^2 + 7||\widehat{K}^{2k,2k}||_F||\widehat{N}^{2k,2k}||_F\\ &+(m+2)||\widehat{K}^{2k,2k}||_F||M||_F^2 + (m+8)||\widehat{K}^{2k,2k}||_F^2||M||_F + 4||\widehat{K}^{2k,2k}||_F\\ &+4||\widehat{N}^{2k,2k}||_F^2||M||_F + (m+6)||\widehat{N}^{2k,2k}||_F||M||_F\Big\} + \mathcal{O}(\mathbf{u}^2),\end{aligned}$$

where $\psi \in \{\psi_1, \psi_2\}$ *and*

$$\mathrm{cond}(\lambda_j) = \mathrm{cond}(\lambda_j^{-1}) = ||\widehat{S}^{2k}||_F^2||y_{2j}||_2||y_{2j-1}||_2$$

is the condition number of the Ritz values λ_j *and* λ_j^{-1}.

Note that this bound is too pessimistic. In order to derive an upper bound for $||U_k^{(4)}||_F$, an upper bound for the matrix (4.8) is used, as $U_k^{(4)}$ is the strictly upper block triangular part of that matrix. This is a very generous upper bound for $||U_k^{(4)}||_F$. Moreover, the term

$$||\widehat{K}^{2k,2k}||_F||\widehat{N}^{2k,2k}||_F$$

is an upper bound for the norm of $\widehat{B}_P^{2k,2k}$. The squared terms $||\widehat{K}^{2k,2k}||_F^2$ and $||\widehat{N}^{2k,2k}||_F^2$ are introduced as the original equations derived in (3.6) and (3.9) are given in terms of $\widehat{K}^{2k,2k}$ and $\widehat{N}^{2k,2k}$, but not in terms of $\widehat{B}_P^{2k,2k}$.

Table 4.3: $|\psi_2|$ and its upper bound from Corollary 4.2, random starting vector.

Lanczos step	ψ_2	bound for $\lvert\psi\rvert$
1	$5.8679 \cdot 1^{-17}$	$4.0869 \cdot 1^{-08}$
2	$6.0217 \cdot 1^{-18}$	$3.6108 \cdot 1^{-07}$
3	$6.8293 \cdot 1^{-19}$	$8.4543 \cdot 1^{-07}$
4	$-4.8526 \cdot 1^{-19}$	$1.5293 \cdot 1^{-06}$
5	$-3.9411 \cdot 1^{-19}$	$2.3792 \cdot 1^{-06}$
6	$-3.9140 \cdot 1^{-19}$	$3.5814 \cdot 1^{-06}$
7	$-3.9135 \cdot 1^{-19}$	$4.7835 \cdot 1^{-06}$
8	$3.9133 \cdot 1^{-19}$	$6.0405 \cdot 1^{-06}$
9	$3.9133 \cdot 1^{-19}$	$8.5821 \cdot 1^{-06}$
10	$-3.9133 \cdot 1^{-19}$	$1.4113 \cdot 1^{-05}$
11	$-3.9133 \cdot 1^{-19}$	$1.0338 \cdot 1^{-04}$
12	$-3.9123 \cdot 1^{-19}$	$4.9373 \cdot 1^{-04}$
13	$-3.8875 \cdot 1^{-19}$	$8.9149 \cdot 1^{-04}$
14	$3.7786 \cdot 1^{-19}$	$9.3192 \cdot 1^{-04}$
15	$-2.7270 \cdot 1^{-19}$	$9.5668 \cdot 1^{-04}$
16	$-1.6512 \cdot 1^{-18}$	$9.7898 \cdot 1^{-04}$

Unfortunately, for the symplectic Lanczos process (as for any nonsymmetric Lanczos-like process), near breakdown may cause the norms of the symplectic Lanczos vectors $||\widetilde{v}_j||_2$ and $||w_j||_2$ to grow unboundedly. Accumulating the quantity $\sum_{j=1}^{k}(||\widehat{\widetilde{v}}_j||_2^2 + ||\widehat{w}_j||_2^2)$, which costs about $4nk$ flops, we can obtain a computable bound for $\text{cond}(\lambda_j)$ and $\text{cond}(\lambda_j^{-1})$ in practise. Theorem 4.1 and Corollary 4.2 indicate that if the J-orthogonality between $\widehat{r}_{k+1}$ and x_j (and z_j) is lost, then the value $\widehat{d}_{k+1}(e_{2k}^T y_{2j-1})$ is proportional to $|\psi_1|$ (and the value $\widehat{d}_{k+1}(e_{2k}^T y_{2j})$ is proportional to $|\psi_2|$). Given the upper bound from Corollary 4.2, and supposing that $\text{cond}(\lambda_j)$ is reasonably bounded, the loss of J-orthogonality implies that $\widehat{d}_{k+1}(e_{2k}^T y_{2j-1})$ (and $\widehat{d}_{k+1}(e_{2k}^T y_{2j})$) are small. Therefore, in the best case we can state that if the effects of finite-precision arithmetic, E_k and F_k in (4.1) and (4.2), are small, then small residuals tell us that the computed eigenvalues are eigenvalues of matrices close to the given matrix.

Example 4.2. Example 3.1 and 4.1 are continued. Table 4.3 reports the value for ψ_2 and its upper bound from Corollary 4.2 using a random starting vector. The upper bound $|\psi|$ is too pessimistic, as already discussed above.

When using the special starting vector $v_1 = [1, 1, 10^{-11}, \ldots, 10^{-11}]^T$ the results are similar.

A Proofs for Theorem 3.1, Lemma 3.2, and Theorem 4.1.

Proof of Theorem 3.1: We need to analyze one step of the symplectic Lanczos algorithm. After $j-1$ steps of the symplectic Lanczos algorithm, we have computed $\widehat{a}_{j-1}$, $\widehat{w}_{j-1}$, $\widehat{c}_{j-1}$, $\widehat{d}_j$, $\widehat{v}_j$. During the jth step we will compute $\widehat{a}_j$, $\widehat{w}_j$, $\widehat{c}_j$, $\widehat{d}_{j+1}$ and $\widehat{v}_{j+1}$. Recall that we set $b_k = 1$, hence b_k is not a computed quantity.

As the analysis is standard rounding error analysis, we will present the details of the rounding error analysis only for the computation of $a_j = v_j^T J_P M_P v_j$. Due to its special structure, multiplication by J_P does not cause any roundoff-error; hence it will not influence our analysis. Let M_P have at most m nonzero entries in any row or column. Then for the matrix-vector multiplication $J_P M_P v_j$ we have

$$\widehat{s}_1 = fl(J_P M_P \widehat{v}_j) = J_P M_P \widehat{v}_j + \widehat{e}_1,$$

where

$$|\widehat{e}_1| \leq m\mathbf{u}\ |J_P M_P|\ |\widehat{v}_j| + \mathcal{O}(\mathbf{u}^2).$$

Then a_j is computed by an inner product

$$\widehat{s}_2 = fl(\widehat{v}_j^T \widehat{s}_1) = \widehat{v}_j^T \widehat{s}_1 + \widehat{e}_2,$$

where

$$|\widehat{e}_2| \leq 2n\mathbf{u}\ |\widehat{v}_j|^T |\widehat{s}_1| + \mathcal{O}(\mathbf{u}^2),$$

assuming that $\widehat{v}_j$ and $\widehat{s}_1$ are full vectors. Overall, we have

$$\widehat{a}_j = \widehat{v}_j^T J_P M_P \widehat{v}_j + \widehat{f}_j^{[1]}, \tag{A.1}$$

where the roundoff error $\widehat{f}_j^{[1]} = \widehat{v}_j^T \widehat{e}_1 + \widehat{e}_2$ is bounded by

$$\begin{aligned} |\widehat{f}_j^{[1]}| &\le m\mathbf{u}\ |\widehat{v}_j|^T |J_P M_P|\ |\widehat{v}_j| + 2n\mathbf{u}\ |\widehat{v}_j|^T |\widehat{s}_1| + \mathcal{O}(\mathbf{u}^2) \\ &\le (m+2n)\mathbf{u}\ |\widehat{v}_j|^T |J_P M_P|\ |\widehat{v}_j| + \mathcal{O}(\mathbf{u}^2). \end{aligned}$$

Analyzing the other necessary computations in a similar fashion (see [5] for details) we obtain

$$\widehat{d}_{j+1}\widehat{v}_{j+1} = -\widehat{d}_j\widehat{v}_{j-1} - \widehat{c}_j\widehat{v}_j + \widehat{w}_j + \widehat{a}_j^{-1} M_P^{-1}\widehat{v}_j + g_{j+1} \tag{A.2}$$

with

$$\begin{aligned} |g_{j+1}| \le (m+5)\mathbf{u}\ |\widehat{a}_j^{-1}|\ |M_P^{-1}|\ |\widehat{v}_j| + 4\mathbf{u}\ |\widehat{w}_j| + 4\mathbf{u}\ |\widehat{c}_j\widehat{v}_j| \\ + 3\mathbf{u}\ |\widehat{d}_j\widehat{v}_{j-1}| + \mathcal{O}(\mathbf{u}^2), \end{aligned} \tag{A.3}$$

and

$$\widehat{a}_j\widehat{w}_j = M_P\widehat{v}_j - \widehat{v}_j + h_j, \tag{A.4}$$

with

$$|h_j| \le (m+2)\mathbf{u}\ |M_P|\ |\widehat{v}_j| + 2\mathbf{u}\ |\widehat{v}_j| + \mathcal{O}(\mathbf{u}^2). \tag{A.5}$$

While the equation $a_j w_j = M_P v_j - v_j$ is given by the $(2j)$th column of $M_P S_P N_P^{-1} = S_P K_P^{-1}$, the equation $d_{j+1}v_{j+1} = -d_j v_{j_1} - c_j v_j + w_j + a_j^{-1} M_P^{-1} v_j$ corresponds to the $(2j-1)$th column of $S_P N_P^{-1} = M_P^{-1} S_P K_P^{-1}$. Hence, in order to summarize the results obtained so far into one single equation, let

$$E_k = [M_P g_2,\ -h_1,\ M_P g_3,\ -h_2, \ldots, M_P g_{k+1},\ -h_k].$$

Then we have from (A.2) and (A.4)

$$M_P \widehat{S}_P^{2k} (\widehat{N}_P^{2k,2k})^{-1} = \widehat{S}_P^{2k} (\widehat{K}_P^{2k,2k})^{-1} - \widehat{d}_{k+1} M_P \widehat{v}_{k+1} e_{2k-1}^T + E_k. \tag{A.6}$$

Using the component-wise upper bounds for $|g_{j+1}|$ and $|h_j|$, let us derive an upper bound for $||E_k||_F$. Clearly,

$$||E_k||_F \le ||\,[h_1,\ h_2, \ldots, h_k]\,||_F + ||M_P||_F||\,[g_2,\ g_3, \ldots, g_{k+1}]\,||_F.$$

From (A.5) we have

$$||\,[h_1,\ h_2, \ldots, h_k]\,||_F \le \mathbf{u}||\widehat{S}^{2k}||_F \left[(m+2)||M||_F + 2||\widehat{K}^{2k,2k}||_F\right] + \mathcal{O}(\mathbf{u}^2). \tag{A.7}$$

Using (A.3) we obtain

$$\begin{aligned} ||\,[g_2,\ g_3, \ldots, g_{k+1}]\,||_F \le \mathbf{u}||\widehat{S}^{2k}||_F \Big[&(m+5)||\widehat{K}^{2k,2k}||_F||M||_F \\ &+ 4 + 4||\widehat{N}^{2k,2k}||_F\Big] + \mathcal{O}(\mathbf{u}^2). \end{aligned} \tag{A.8}$$

Hence, summarizing we obtain as an upper bound for the error matrix E_k of (A.6) the bound given in (3.7). □

Proof of Lemma 3.2: Obviously,

$$\ell_{2j,2j} = \ell_{2j-1,2j-1} = 0$$

for $j = 1, \ldots, k$ as $x^T J_P x = 0$ for any vector x. Moreover, as $\ell_{2m,2j-1} = -\ell_{2j-1,2m}$, we only need to examine $\ell_{2j,2j-1}$ for $j = 1, \ldots, k$, and $\ell_{2j,2m-1}$, $\ell_{2j-1,2m-1}$ and $\ell_{2j,2m}$ for $j, m = 1, \ldots, k$, $j < m$.

Let us start with $\ell_{2j,2j-1}$. Using the result of the rounding error analysis for $\widehat{w}_j$ and $\widehat{v}_j$ we have (see [5] for details)

$$\ell_{2j,2j-1} = \frac{\widehat{\widetilde{w}}_j^T J_P \widehat{\widetilde{v}}_j + \zeta_1}{\widehat{a}_j \widehat{d}_j} + \mathcal{O}(\mathbf{u}^2), \tag{A.9}$$

where

$$|\zeta_1| \leq 2\mathbf{u}\, |\widehat{\widetilde{v}}_j|^T |J_P|\, (|M_P|\, |\widehat{v}_j| - |\widehat{v}_j|).$$

We would like to be able to rewrite $\ell_{2j,2j-1} = \widehat{w}_j^T J_P \widehat{v}_j = -1 + \textit{some small error}$. In order to do so, we rewrite $\widehat{a}_j \widehat{d}_j$ suitably. From (A.1), and the result of the rounding error analysis for $\widehat{\widetilde{w}}_j$ and $\widehat{v}_j$ we have (see [5] for details)

$$\widehat{a}_j \widehat{d}_j = -\widehat{\widetilde{w}}_j^T J_P \widehat{\widetilde{v}}_j + \zeta_2, \tag{A.10}$$

where the roundoff error ζ_2 is bounded by

$$|\zeta_2| \leq (2m + 2n + 2)\mathbf{u}\, |\widehat{\widetilde{v}}_j|^T |J_P|\, |M_P|\, |\widehat{v}_j| + 3\mathbf{u}\, |\widehat{\widetilde{v}}_j|^T |J_P|\, |\widehat{v}_j| + \mathcal{O}(\mathbf{u}^2).$$

Combining (A.9) and (A.10) we have

$$\begin{aligned}
\ell_{2j,2j-1} &= \frac{\widehat{\widetilde{w}}_j^T J_P \widehat{\widetilde{v}}_j + \zeta_1}{-\widehat{\widetilde{w}}_j^T J_P \widehat{\widetilde{v}}_j + \zeta_2} + \mathcal{O}(\mathbf{u}^2) \\
&= -1 + \frac{\zeta_1 + \zeta_2}{-\widehat{\widetilde{w}}_j^T J_P \widehat{\widetilde{v}}_j + \zeta_2} + \mathcal{O}(\mathbf{u}^2) \\
&=: -1 + \kappa_j + \mathcal{O}(\mathbf{u}^2).
\end{aligned}$$

Using the Taylor expansion of $f(x) = \frac{\zeta_1+\zeta_2}{x+\zeta_2}$ at $t = x - \zeta_2$, we obtain (see [5] for details)

$$\begin{aligned}
|\kappa_j| &\leq \frac{|\zeta_1| + |\zeta_2|}{|\widehat{\widetilde{w}}_j^T J_P \widehat{\widetilde{v}}_j|} + \mathcal{O}(\mathbf{u}^2) \\
&\leq \frac{2(m+n+2)\mathbf{u}\, |\widehat{\widetilde{v}}_j|^T |J_P|\, |M_P|\, |\widehat{v}_j| + 5\mathbf{u}\, |\widehat{\widetilde{v}}_j|^T |J_P|\, |\widehat{v}_j|}{|\widehat{\widetilde{w}}_j^T J_P \widehat{\widetilde{v}}_j|} + \mathcal{O}(\mathbf{u}^2). \qquad \text{(A.11)}
\end{aligned}$$

Next we turn our attention to the terms $\ell_{2j,2m-1}, \ell_{2j-1,2m-1}$, and $\ell_{2j,2m}$. The analysis of these three terms will be demonstrated by considering $\ell_{2j,2m}$ =

$\widehat{w}_j^T J_P \widehat{w}_m$. Let us assume that we have already analyzed all previous terms, that is, all the terms

$$\begin{aligned} \ell_{i,\ell} \quad & i = 1, \ldots, 2m-1, \quad \ell = 1, \ldots, 2m-1, \\ \ell_{2m,\ell} \quad & \ell = 1, \ldots, 2j-1, \\ \ell_{i,2m} \quad & i = 1, \ldots, 2j-1. \end{aligned}$$

Our goal is to rewrite $\ell_{2j,2m}$ in terms of any of these already analyzed terms. First of all, note, that for $j = m$ we have $\ell_{2m,2m} = 0$. Hence for the following discussion we assume $j < m$. From (A.4) we have

$$\begin{aligned} \widehat{a}_m \ell_{2j,2m} &= \widehat{w}_j^T J_P (M_P \widehat{v}_m - \widehat{v}_m + h_m) \\ &= \widehat{w}_j^T J_P M_P \widehat{v}_m - \ell_{2j,2m-1} + \widehat{w}_j^T J_P h_m. \end{aligned}$$

Using (A.2) and (A.4) to analyze $\widehat{w}_j^T J_P M_P \widehat{v}_m$ we obtain (see [5] for details)

$$\begin{aligned} \widehat{d}_m \widehat{a}_m \ell_{2j,2m} = {} & \widehat{a}_j^{-1} \ell_{2j-1,2m-2} - \widehat{d}_m \ell_{2j,2m-1} - \widehat{c}_{m-1} \widehat{a}_{m-1} \ell_{2j,2m-2} - \widehat{c}_{m-1} \ell_{2j,2m-3} \\ & + \widehat{a}_{m-1}^{-1} \ell_{2j,2m-3} - \widehat{d}_{m-1} \widehat{a}_{m-2} \ell_{2j,2m-4} - \widehat{d}_{m-1} \ell_{2j,2m-5} \\ & + \widehat{d}_j \ell_{2m-2,2j-3} + \widehat{c}_j \ell_{2m-2,2j-1} + \widehat{d}_{j+1} \ell_{2m-2,2j+1} + \ell_{2m-2,2j} \\ & + \widehat{d}_{m-1} \widehat{w}_j^T J_P h_{m-2} + \widehat{c}_{m-1} \widehat{w}_j^T J_P h_{m-1} + \widehat{d}_m \widehat{w}_j^T J_P h_m \\ & + \widehat{a}_j^{-1} h_j^T J_P M_P \widehat{w}_{m-1} - \widehat{w}_{m-1}^T J_P g_{j+1} + \widehat{w}_j^T J_P M_P g_m. \end{aligned}$$

A similar analysis can be done for $\ell_{2j,2m-1}$ and $\ell_{2j-1,2m-1}$. □

Proof of Theorem 4.1: Our goal is to derive expressions for $z_j^T J_P \widehat{r}_{k+1}$ and $x_j^T J_P \widehat{r}_{k+1}$ that describe the way in which J-orthogonality is lost. In exact arithmetic, these expressions are zero. Our approach follows Bai's derivations in [1, Proof of Theorem 4.1]. Pre-multiplying (4.2) by $(\widehat{S}_P^{2k})^T$ and taking the transpose yields

$$(\widehat{W}_P^{2k})^T M_P \widehat{S}_P^{2k} = \widehat{B}_P^{2k,2k} (\widehat{W}_P^{2k})^T \widehat{S}_P^{2k} + (\widehat{K}_P^{2k,2k})^{-1} \left[\widehat{d}_{k+1} J_P \widehat{v}_{k+1} e_{2k}^T + F_k \right]^T \widehat{S}_P^{2k}.$$

Pre-multiplying (4.1) by $(\widehat{W}_P^{2k})^T$ we obtain

$$(\widehat{W}_P^{2k})^T M_P \widehat{S}_P^{2k} = (\widehat{W}_P^{2k})^T \widehat{S}_P^{2k} \widehat{B}_P^{2k,2k} - (\widehat{W}_P^{2k})^T \left[\widehat{d}_{k+1} \widehat{r}_{k+1} e_{2k-1}^T - E_k \right] \widehat{N}_P^{2k,2k}.$$

Subtracting these two equations, we obtain

$$\begin{aligned} & (\widehat{W}_P^{2k})^T \widehat{S}_P^{2k} \widehat{B}_P^{2k,2k} - \widehat{B}_P^{2k,2k} (\widehat{W}_P^{2k})^T \widehat{S}_P^{2k} = \widehat{d}_{k+1} (\widehat{K}_P^{2k,2k})^{-1} e_{2k} \widehat{v}_{k+1}^T J_P^T \widehat{S}_P^{2k} \\ & - \widehat{d}_{k+1} (\widehat{W}_P^{2k})^T \widehat{r}_{k+1} e_{2k}^T + (\widehat{K}_P^{2k,2k})^{-1} F_k^T \widehat{S}_P^{2k} - (\widehat{W}_P^{2k})^T E_k \widehat{N}_P^{2k,2k}. \end{aligned} \tag{A.12}$$

We are most interested in deriving an expression for

$$(\widehat{S}_P^{2k})^T J_P \widehat{r}_{k+1} e_{2k-1}^T \qquad (\text{or } (\widehat{W}_P^{2k})^T J_P \widehat{r}_{k+1} e_{2k-1}^T)$$

from the above equation. From this we can easily obtain expressions for $z_j^T J_P r_{k+1}$ or $x_j^T J_P r_{k+1}$ as desired. In order to do so, we note that most of the matrices in (A.12) have a very special form. Let us start with the left-hand side. From (4.3) we have $(\widehat{S}_P^{2k})^T J_P \widehat{S}_P^{2k} = K = J_P^{2k,2k} + C_k + \Delta_k - C_k^T$. This implies

$$(\widehat{W}_P^{2k})^T \widehat{S}_P^{2k} = -I^{2k,2k} + J_P^{2k,2k} C_k + J_P^{2k,2k} \Delta_k - J_P^{2k,2k} C_k^T,$$

where $J_P^{2k,2k} C_k$ and $(J_P^{2k,2k} C_k^T)^T$ have the same form as C_k, and $J_P^{2k,2k}\Delta_k$ is a diagonal matrix,

$$J_P^{2k,2k}\Delta_k = \operatorname{diag}\left(\begin{bmatrix} -\kappa_1 & 0 \\ 0 & -\kappa_1 \end{bmatrix}, \ldots, \begin{bmatrix} -\kappa_k & 0 \\ 0 & -\kappa_k \end{bmatrix} \right).$$

Therefore, we can rewrite the left-hand side of (A.12) as

$$\begin{aligned} (\widehat{W}_P^{2k})^T \widehat{S}_P^{2k} \widehat{B}_P^{2k,2k} - \widehat{B}_P^{2k,2k} (\widehat{W}_P^{2k})^T \widehat{S}_P^{2k} = {} & \left[J_P^{2k,2k} C_k \widehat{B}_P^{2k,2k} - \widehat{B}_P^{2k,2k} J_P^{2k,2k} C_k \right] \\ & + \left[J_P^{2k,2k} \Delta_k \widehat{B}_P^{2k,2k} - \widehat{B}_P^{2k,2k} J_P^{2k,2k} \Delta_k \right] \\ & + \left[J_P^{2k,2k} C_k^T \widehat{B}_P^{2k,2k} - \widehat{B}_P^{2k,2k} J_P^{2k,2k} C_k^T \right]. \end{aligned}$$

By the local J-orthogonality assumption (and, therefore, by the special form of $J_P^{2k,2k} C_k$), it follows that

$$L_k^{(1)} := J_P^{2k,2k} C_k \widehat{B}_P^{2k,2k} - \widehat{B}_P^{2k,2k} J_P^{2k,2k} C_k$$

is a strictly lower block triangular matrix with block size 2. With the same argument we have that

$$U_k^{(1)} := J_P^{2k,2k} C_k^T \widehat{B}_P^{2k,2k} - \widehat{B}_P^{2k,2k} J_P^{2k,2k} C_k^T$$

is a strictly upper block triangular matrix with block size 2. Since the 2×2 diagonal blocks of $J_P^{2k,2k} \Delta_k \widehat{B}_P^{2k,2k} - \widehat{B}_P^{2k,2k} J_P^{2k,2k} \Delta_k$ are zero, we can write

$$J_P^{2k,2k} \Delta_k \widehat{B}_P^{2k,2k} - \widehat{B}_P^{2k,2k} J_P^{2k,2k} \Delta_k = L_k^{(2)} + U_k^{(2)},$$

where $L_k^{(2)}$ is strictly lower block triangular and $U_k^{(2)}$ strictly upper block triangular. Hence,

$$(\widehat{W}_P^{2k})^T \widehat{S}_P^{2k} \widehat{B}_P^{2k,2k} - \widehat{B}_P^{2k,2k} (\widehat{W}_P^{2k})^T \widehat{S}_P^{2k} = L_k^{(1)} + L_k^{(2)} + U_k^{(1)} + U_k^{(2)}.$$

Now let us turn our attention to the right-hand side of (A.12). The row vector

$$\widehat{v}_{k+1}^T J_P^T \widehat{S}_P^{2k} = [* \cdots * \ 0 \ 0]$$

has nonzero elements in its first $2n-2$ positions. As $(\widehat{K}_P^{2k,2k})^{-1} e_{2k} = b_k e_{2k-1} + a_k e_{2k}$ we have that

$$L_k^{(3)} := \widehat{d}_{k+1} (\widehat{K}_P^{2k,2k})^{-1} e_{2k} \widehat{v}_{k+1}^T J_P^T \widehat{S}_P^{2k}$$

is a strictly lower block triangular matrix with block size 2. Similarly we have that

$$U_k^{(3)} := \widehat{d}_{k+1}(\widehat{W}_P^{2k})^T \widehat{r}_{k+1} e_{2k}^T$$

is a strictly upper block triangular matrix with block size 2. Hence, we can rewrite (A.12) as

(A.13)
$$L_k^{(1)} + L_k^{(2)} - L_k^{(3)} + U_k^{(1)} + U_k^{(2)} - U_k^{(3)} = (\widehat{K}_P^{2k,2k})^{-1} F_k^T \widehat{S}_P^{2k} - (\widehat{W}_P^{2k})^T E_k \widehat{N}_P^{2k,2k}.$$

This implies that the diagonal blocks of $(\widehat{K}_P^{2k,2k})^{-1} F_k^T \widehat{S}_P^{2k} - (\widehat{W}_P^{2k})^T E_k \widehat{N}_P^{2k,2k}$ must be zero. Therefore, we can write

$$(\widehat{K}_P^{2k,2k})^{-1} F_k^T \widehat{S}_P^{2k} - (\widehat{W}_P^{2k})^T E_k \widehat{N}_P^{2k,2k} = L_k^{(4)} + U_k^{(4)}$$

where $L_k^{(4)}$ is strictly lower block triangular and $U_k^{(4)}$ is strictly upper block triangular. By writing down only the strictly upper block triangular part of (A.13) we have

$$\widehat{d}_{k+1}(\widehat{W}_P^{2k})^T \widehat{r}_{k+1} e_{2k}^T = J_P^{2k,2k} C_k^T \widehat{B}_P^{2k,2k} - \widehat{B}_P^{2k,2k} J_P^{2k,2k} C_k^T + U_k^{(2)} - U_k^{(4)}.$$

This is equivalent to

(A.14)
$$\widehat{d}_{k+1}(\widehat{S}_P^{2k})^T J_P^{2n,2n} \widehat{r}_{k+1} e_{2k}^T = C_k^T \widehat{B}_P^{2k,2k} - (\widehat{B}_P^{2k,2k})^{-T} C_k^T - J_P^{2k,2k}\left[U_k^{(2)} - U_k^{(4)}\right],$$

where we have used the fact that $\widehat{B}_P^{2k,2k}$ is symplectic.

Pre-multiplying (A.14) by y_{2j}^T and post-multiplying by y_{2j-1} yields (using (4.5))

$$\widehat{d}_{k+1} y_{2j}^T (\widehat{S}_P^{2k})^T J_P^{2n,2n} \widehat{r}_{k+1}(e_{2k}^T y_{2j-1}) = y_{2j}^T J_P^{2k,2k}\left[U_k^{(4)} - U_k^{(2)}\right] y_{2j-1}.$$

Similarly, pre-multiplying (A.14) by y_{2j-1}^T and post-multiplying by y_{2j} yields

$$\widehat{d}_{k+1} y_{2j-1}^T (\widehat{S}_P^{2k})^T J_P^{2n,2n} \widehat{r}_{k+1} e_{2k}^T y_{2j} = y_{2j-1}^T J_P^{2k,2k}\left[U_k^{(4)} - U_k^{(2)}\right] y_{2j}.$$

With the assumptions (4.4) and (4.5) this concludes the proof of Theorem 4.1. □

Acknowledgments.

The author would like to thank the anonymous referees and Peter Benner for many helpful comments which greatly improved the presentation of the results.

REFERENCES

1. Z. Bai, *Error analysis of the Lanczos algorithm for the nonsymmetric eigenvalue problem*, Math. Comp., 62 (1994), pp. 209–226.
2. G. Banse, *Symplektische Eigenwertverfahren zur Lösung zeitdiskreter optimaler Steuerungs Probleme*, Ph.D Thesis, Fachbereich 3—Mathematik und Informatik, Universität Bremen, Bremen, Germany, 1995.
3. P. Benner and H. Faßbender, *An implicitly restarted symplectic Lanczos method for the symplectic eigenvalue problem*, Berichte aus der Technomathematik, Report 98-01, Fachbereich 3—Mathematik und Informatik, Universität Bremen, Bremen, Germany, 1998; to appear in SIAM. J. Matrix Anal. Appl.
4. P. Benner and H. Faßbender, *The symplectic eigenvalue problem, the butterfly form, the SR algorithm, and the Lanczos method*, Linear Algebra Appl., 275–276 (1998), pp. 19–47.
5. H. Faßbender, *Symplectic Methods for Symplectic Eigenproblems*, Habilitationsschrift, Fachbereich 3—Mathematik und Informatik, Universität Bremen, Bremen, Germany, 1998.
6. R. Freund, *Transpose-free quasi-minimal residual methods for non-Hermitian linear systems*, in Recent Advances in Iterative Methods, IMA Workshop on Iterative Methods for Sparse and Structured Problems, Minneapolis, MN, February 24–March 1, 1992., G. Golub et al., eds., Vol. 60 of IMA Vol. Math. Appl., Springer–Verlag, New York, NY, 1994, pp. 69–94.
7. G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed., The Johns Hopkins University Press, Baltimore, MD, 1996.
8. N. Higham, *Accuracy and Stability of Numerical Algorithms*, SIAM Publications, Philadelphia, PA, 1996.
9. W. Kahan, B. Parlett, and E. Jiang, *Residual bounds on approximate eigensystems of nonnormal matrices*, SIAM J. Numer. Anal., 19 (1982), pp. 470–484.
10. P. Lancaster and L. Rodman, *The Algebraic Riccati Equation*, Oxford University Press, Oxford, 1995.
11. C. Lanczos, *An iteration method for the solution of the eigenvalue problem of linear differential and integral operators*, J. Res. Nat. Bur. Standards, 45 (1950), pp. 255–282.
12. V. Mehrmann, *The Autonomous Linear Quadratic Control Problem, Theory and Numerical Solution*, no. 163 in Lecture Notes in Control and Information Sciences, Springer-Verlag, Heidelberg, July 1991.
13. C. Paige, *Error analysis of the Lanczos algorithm for tridiagonalizing a symmetric matrix*, J. Inst. Math. Appl., 18 (1976), pp. 341–349.
14. B. Parlett, *A new look at the Lanczos algorithm for solving symmetric systems of linear equations*, Linear Algebra Appl., 29 (1980), pp. 323–346.
15. B. Parlett, *The Symmetric Eigenvalue Problem*, Prentice-Hall, Englewood Cliffs, NJ, 1980.
16. H. Simon, *Analysis of the symmetric Lanczos algorithm with reorthogonalization methods*, Linear Algebra Appl., 61 (1984), pp. 101–132.
17. K. Zhou, J. C. Doyle, and K. Glover, *Robust and Optimal Control*, Prentice-Hall, Upper Saddle River, NJ, 1996.