Electronic Transactions on Numerical Analysis.
Volume 23, pp. 141-157, 2006.

ISSN 1068-9613.

ETNA
Kent State University
etna@mcs.kent.edu

ON THE REDUCTION OF A HAMILTONIAN MATRIX TO HAMILTONIAN SCHUR FORM*

DAVID S. WATKINS†

Abstract. Recently Chu, Liu, and Mehrmann developed an $O(n^3)$ structure preserving method for computing the Hamiltonian real Schur form of a Hamiltonian matrix. This paper outlines an alternative derivation of the method and an alternative explanation of why the method works. Our approach places emphasis eigenvalue swapping and relies less on matrix manipulations.

Key words. Hamiltonian matrix, skew-Hamiltonian matrix, stable invariant subspace, real Schur form

AMS subject classifications. 65F15, 15A18, 93B40

1. Introduction.

In [4] Chu, Liu, and Mehrmann presented an $O(n^3)$ structure-preserving method for computing the real Hamiltonian Schur form of a Hamiltonian matrix. The current paper is the fruit of this author's attempt to understand [4]. The method sketched here differs hardly at all from the one presented in [4]; the objective of this paper is not to present a new algorithm but rather to provide an alternative explanation of the method and why it works. There are two main differences between our presentation and that of [4]. 1) We use the idea of swapping eigenvalues or blocks of eigenvalues in the real Schur form and rely less on matrix manipulations. 2) The presentation in [4] expends a great deal of effort on nongeneric cases. We focus on the generic case and dwell much less on special situations. However, we do show that the method developed for the generic case actually works in all of the nongeneric cases as well. This is true in exact arithmetic. It may turn out that in the presence of roundoff errors one cannot avoid paying special attention to nongeneric cases as in [4].

2. Definitions and Preliminary Results.

Throughout this paper we restrict our attention to matrices with real entries. Define $J \in \mathbb{R}^{2n\times 2n}$ by

$$J = \begin{bmatrix} 0 & I_n \\ -I_n & 0 \end{bmatrix}.$$

A matrix $\mathcal{H} \in \mathbb{R}^{2n\times 2n}$ is *Hamiltonian* if $J\mathcal{H} = (J\mathcal{H})^T$. $\mathcal{H}$ is *skew-Hamiltonian* if $J\mathcal{H} = -(J\mathcal{H})^T$. $\mathcal{U} \in \mathbb{R}^{2n\times 2n}$ is *symplectic* if $\mathcal{U}^T J \mathcal{U} = J$.

A number of elementary relationships are easily proved. Every symplectic matrix is nonsingular. The product of two symplectic matrices is symplectic. The inverse of a symplectic matrix is symplectic. Thus the set of symplectic matrices is a group under the operation of matrix multiplication. An important subgroup is the set of orthogonal symplectic matrices. We use two types of orthogonal symplectic transformations in this paper: If $Q \in \mathbb{R}^{n\times n}$ is orthogonal, then

$$\begin{bmatrix} Q & 0 \\ 0 & Q \end{bmatrix}$$

is orthogonal and symplectic. If $C, S \in \mathbb{R}^{n\times n}$ are diagonal and satisfy $C^2 + S^2 = I$, then

$$\begin{bmatrix} C & -S \\ S & C \end{bmatrix}$$

*Received October 14, 2005. Accepted for publication February 28, 2006. Recommended by V. Mehrmann.
†Department of Mathematics, Washington State University, Pullman, Washington 99164-3113 (watkins@math.wsu.edu).

ETNA
Kent State University
etna@mcs.kent.edu

is orthogonal and symplectic. In particular, for any k, a rotator acting in the $(k, n+k)$ plane is symplectic.

If $\mathcal{U}$ is symplectic and $\mathcal{H}$ is Hamiltonian (resp. skew-Hamiltonian), then $\mathcal{U}^{-1}\mathcal{H}\mathcal{U}$ is Hamiltonian (resp. skew-Hamiltonian).

If $\mathcal{H}$ is Hamiltonian and v is a right eigenvector of $\mathcal{H}$ with eigenvalue λ, then $v^T J$ is a left eigenvector with eigenvalue $-\lambda$. It follows that if λ is an eigenvalue of $\mathcal{H}$, then so are $\overline{\lambda}$, $-\lambda$, and $-\overline{\lambda}$. Thus the spectrum of a Hamiltonian matrix is symmetric with respect to both the real and the imaginary axes.

A subspace $\mathcal{S} \subseteq \mathbb{R}^{2n}$ is called *isotropic* if $x^T J y = 0$ for all $x, y \in \mathcal{S}$. If $X \in \mathbb{R}^{2n\times j}$ is a matrix such that $\mathcal{S} = \mathcal{R}(X)$, then $\mathcal{S}$ is isotropic if and only if $X^T J X = 0$. Given a symplectic matrix $S \in \mathbb{R}^{2n\times 2n}$, let S_1 and S_2 denote the first n and last n columns of S, respectively. Then the relationship $S^T J S = J$ implies that $\mathcal{R}(S_1)$ and $\mathcal{R}(S_2)$ are both isotropic. The following well-known fact is crucial to our development.

PROPOSITION 2.1. *Let $\mathcal{S} \subseteq \mathbb{R}^{2n}$ be a subspace that is invariant under the Hamiltonian matrix $\mathcal{H}$. Suppose that all of the eigenvalues of $\mathcal{H}$ associated with $\mathcal{S}$ satisfy $\Re(\lambda) < 0$. Then $\mathcal{S}$ is isotropic.*

Proof. Let X be a matrix with linearly independent columns such that $\mathcal{S} = \mathcal{R}(X)$. We need to show that $X^T J X = 0$. Invariance of $\mathcal{S}$ implies that $\mathcal{H}X = XC$ for some C whose eigenvalues all satisfy $\Re(\lambda) < 0$. Multiplying on the left by $X^T J$, we get $X^T J X C = X^T (J\mathcal{H}) X$. Since $J\mathcal{H}$ is symmetric, we deduce that $X^T J X C$ is symmetric, so that $X^T J X C + C^T X^T J X = 0$. Letting $Y = X^T J X$, we have $YC + C^T Y = 0$. The Lyapunov operator $Y \to YC + C^T Y$ has eigenvalues $\lambda_j + \mu_k$, where λ_j and μ_k are eigenvalues of C and C^T, respectively. Since the eigenvalues $\lambda_k + \mu_k$ all lie in the open left halfplane, they are all nonzero. Thus the Lyapunov operator is nonsingular, and the homogeneous Lyapunov equation $YC + C^T Y = 0$ has only the solution $Y = 0$. Therefore $X^T J X = 0$. ☐

If $\mathcal{H}$ is Hamiltonian, then $\mathcal{H}^2$ is skew-Hamiltonian. $\mathcal{K}$ is skew-Hamiltonian if and only if it has the block form

$$\mathcal{K} = \begin{bmatrix} A & N \\ K & A^T \end{bmatrix}, \quad N^T = -N,\ K^T = -K.$$

A matrix $B \in \mathbb{R}^{n\times n}$ is called *quasitriangular* if it is block upper triangular with 1×1 and 2×2 blocks along the main diagonal, each 2×2 block housing a conjugate pair of complex eigenvalues. Quasitriangular form is also called *real Schur form* or *Wintner-Murnaghan form*. A skew-Hamiltonian matrix $\mathcal{N} \in \mathbb{R}^{2n\times 2n}$ is said to be in *skew-Hamiltonian real Schur form* if

$$\mathcal{N} = \begin{bmatrix} B & N \\ 0 & B^T \end{bmatrix},$$

where B is quasitriangular.

THEOREM 2.2. [9] *Every skew-Hamiltonian matrix is similar, via an orthogonal symplectic similarity transformation, to a matrix in skew-Hamiltonian Schur form. That is, if $\mathcal{K} \in \mathbb{R}^{2n\times 2n}$ is skew Hamiltonian, then there is an orthogonal symplectic $\mathcal{U} \in \mathbb{R}^{2n\times 2n}$ and a skew-Hamiltonian $\mathcal{N} \in \mathbb{R}^{2n\times 2n}$ such that*

$$\mathcal{N} = \begin{bmatrix} B & N \\ 0 & B^T \end{bmatrix},$$

where B is quasitriangular, and $\mathcal{K} = \mathcal{U}\mathcal{N}\mathcal{U}^T$.

The eigenvalues of $\mathcal{K}$ and $\mathcal{N}$ are the eigenvalues of B repeated twice. We deduce that the eigenspaces of a skew-Hamiltonian matrix all have even dimension.

$\mathcal{M}$ is Hamiltonian if and only if it has the block form

$$\mathcal{M} = \begin{bmatrix} F & G \\ H & -F^T \end{bmatrix}, \quad G^T = G,\ H^T = H.$$

A Hamiltonian matrix $\mathcal{H} \in \mathbb{R}^{2n\times 2n}$ that has no purely imaginary eigenvalues must have exactly n eigenvalues in the left halfplane and n in the right halfplane. Such a matrix is said to be in *Hamiltonian real Schur form* if

$$\mathcal{H} = \begin{bmatrix} T & G \\ 0 & -T^T \end{bmatrix},$$

where T is quasitriangular, and the eigenvalues of T all have negative real part.

THEOREM 2.3. [8] *Let $\mathcal{M} \in \mathbb{R}^{2n\times 2n}$ be a Hamiltonian matrix that has no purely imaginary eigenvalues. Then $\mathcal{M}$ is similar, via an orthogonal symplectic similarity transformation, to a matrix in Hamiltonian real Schur form. That is, there exists an orthogonal symplectic $\mathcal{U} \in \mathbb{R}^{2n\times 2n}$ and a Hamiltonian*

$$\mathcal{H} = \begin{bmatrix} T & G \\ 0 & -T^T \end{bmatrix}$$

in Hamiltonian real Schur form, such that $\mathcal{M} = \mathcal{U}\mathcal{H}\mathcal{U}^T$.

The linear-quadratic Gaussian problem of control theory, also known as the quadratic regulator problem, can be solved by solving an algebraic Riccati equation. This is equivalent to finding the n-dimensional invariant subspace of a Hamiltonian matrix associated with the eigenvalues in the left halfplane [7]. If one can compute a Hamiltonian real Schur form as in Theorem 2.3, then the first n columns of $\mathcal{U}$ span the desired invariant subspace, leading to the solution of the linear-quadratic Gaussian problem.

The proof of Theorem 2.3 in [8] is nonconstructive. Since the publication of [8], it has been an open problem to find a backward stable $O(n^3)$ method to compute the orthogonal symplectic similarity transformation to Hamiltonian real Schur form.

We will use the following version of the symplectic URV decomposition theorem [2].

THEOREM 2.4. *Let $\mathcal{H} \in \mathbb{R}^{2n\times 2n}$ be a Hamiltonian matrix. Then there exist orthogonal, symplectic $\mathcal{U}$, $\mathcal{V} \in \mathbb{R}^{2n\times 2n}$, an upper-triangular $T \in \mathbb{R}^{n\times n}$, quasitriangular $S \in \mathbb{R}^{n\times n}$, and $C \in \mathbb{R}^{n\times n}$ such that*

$$\mathcal{H} = \mathcal{U}\mathcal{R}_1\mathcal{V}^T \quad \textit{and} \quad \mathcal{H} = \mathcal{V}\mathcal{R}_2\mathcal{U}^T,$$

where

$$\mathcal{R}_1 = \begin{bmatrix} S & C \\ 0 & T^T \end{bmatrix} \quad \textit{and} \quad \mathcal{R}_2 = \begin{bmatrix} -T & C^T \\ 0 & -S^T \end{bmatrix}.$$

There is a backward stable $O(n^3)$ algorithm, implemented in HAPACK [1], to compute this decomposition. The decompositions $\mathcal{H} = \mathcal{U}\mathcal{R}_1\mathcal{V}^T$ and $\mathcal{H} = \mathcal{V}\mathcal{R}_2\mathcal{U}^T$ together show that the skew-Hamiltonian matrix $\mathcal{H}^2$ satisfies $\mathcal{H}^2 = \mathcal{U}\mathcal{R}_1\mathcal{R}_2\mathcal{U}^T$. Since

$$\mathcal{R}_1\mathcal{R}_2 = \begin{bmatrix} -ST & -SC^T - CS^T \\ 0 & -(ST)^T \end{bmatrix},$$

this is the skew-Hamiltonian real Schur form of $\mathcal{H}^2$, computed without ever forming $\mathcal{H}^2$ explicitly. The eigenvalues of $\mathcal{H}$ are the square roots of the eigenvalues of the quasitriangular matrix $-ST$. Thus the symplectic URV decomposition is a backward stable method of computing the eigenvalues of $\mathcal{H}$.

ETNA
Kent State University
etna@mcs.kent.edu

3. The Method. Let $\mathcal{M} \in \mathbb{R}^{2n\times 2n}$ be a Hamiltonian matrix that has no eigenvalues on the imaginary axis. Our objective is to compute the Hamiltonian Schur form of $\mathcal{M}$. We begin by computing the symplectic URV decomposition of $\mathcal{M}$, as described in Theorem 2.4. Taking the orthogonal symplectic matrix $\mathcal{U}$ from that decomposition, let $\mathcal{H} = \mathcal{U}^T\mathcal{M}\mathcal{U}$. Then $\mathcal{H}$ is a Hamiltonian matrix with the property that $\mathcal{H}^2$ is in skew-Hamiltonian Schur form.

As was explained in [4], the new method produces (at $O(n^2)$ cost) an orthogonal symplectic $\mathcal{Q}$ such that the Hamiltonian matrix $\mathcal{Q}^T\mathcal{H}\mathcal{Q}$ can partitioned as

$$\mathcal{Q}^T\mathcal{H}\mathcal{Q} = \left[\begin{array}{cc|cc} \hat{F}_1 & * & * & * \\ 0 & \hat{F} & * & \hat{G} \\ \hline 0 & 0 & -\hat{F}_1^T & 0 \\ 0 & \hat{H} & * & -\hat{F}^T \end{array}\right], \tag{3.1}$$

where $\hat{F}_1$ is either 1×1 with a negative real eigenvalue or 2×2 with a complex conjugate pair of eigenvalues in the left halfplane, and the remaining Hamiltonian matrix

$$\hat{\mathcal{H}} = \begin{bmatrix} \hat{F} & \hat{G} \\ \hat{H} & -\hat{F}^T \end{bmatrix} \tag{3.2}$$

has the property that its square is in skew-Hamiltonian Schur form. Now we can apply the same procedure to $\hat{\mathcal{H}}$ to deflate out another eigenvalue or two, and repeating the procedure as many times as necessary, we get the Hamiltonian Schur form of $\mathcal{H}$, hence of $\mathcal{M}$, in $O(n^3)$ work.

4. The Method in the simplest case. To keep the discussion simple let us assume at first that all of the eigenvalues of $\mathcal{M}$ are real. Then

$$\mathcal{H}^2 = \begin{bmatrix} B & N \\ 0 & B^T \end{bmatrix},$$

where B is upper triangular and has positive entries on the main diagonal. It follows that e_1 is an eigenvector of $\mathcal{H}^2$ with associated eigenvalue b_{11}. Generically e_1 will *not* be an eigenvector of $\mathcal{H}$, and we will assume this for now. Thus the vectors e_1 and $\mathcal{H}e_1$ are linearly independent.[1] Since e_1 *is* an eigenvector of $\mathcal{H}^2$, the two-dimensional space span$\{e_1, \mathcal{H}e_1\}$ is invariant under $\mathcal{H}$. In fact

$$\mathcal{H}\begin{bmatrix} e_1 & \mathcal{H}e_1 \end{bmatrix} = \begin{bmatrix} e_1 & \mathcal{H}e_1 \end{bmatrix}\begin{bmatrix} 0 & b_{11} \\ 1 & 0 \end{bmatrix}.$$

The associated eigenvalues are $\lambda = -\sqrt{b_{11}}$ and $-\lambda = \sqrt{b_{11}}$, and the eigenvectors are $(\mathcal{H}+\lambda I)e_1$, and $(\mathcal{H}-\lambda I)e_1$, respectively.

Our objective is to build an orthogonal symplectic $\mathcal{Q}$ that causes a deflation as explained in the previous section. Let

$$x = (\mathcal{H}+\lambda I)e_1, \tag{4.1}$$

the eigenvector of $\mathcal{H}$ associated with eigenvalue λ. We will build an orthogonal symplectic $\mathcal{Q}$ that has its first column proportional to x. This will guarantee that

$$\mathcal{Q}^T\mathcal{H}\mathcal{Q} = \left[\begin{array}{cc|cc} \lambda & * & * & * \\ 0 & \hat{F} & * & \hat{G} \\ \hline 0 & 0 & -\lambda & 0 \\ 0 & \hat{H} & * & -\hat{F}^T \end{array}\right], \tag{4.2}$$

[1] All nongeneric cases,including the trivial case $\mathcal{H}e_1 = \lambda e_1$, will be resolved eventually.

thereby deflating out eigenvalues $\pm\lambda$. We have to be careful how we construct $\mathcal{Q}$, as we must also guarantee that the deflated matrix $\hat{\mathcal{H}}$ of (3.2) has the property that its square is in skew-Hamiltonian Schur form. For this it is necessary and sufficient that $\mathcal{Q}^T\mathcal{H}^2\mathcal{Q}$ be in skew-Hamiltonian Schur form. We will outline the algorithm first; then we will demonstrate that it has the desired properties.

Outline of the algorithm, real case. Partition the eigenvector x from (4.1) as

$$x = \begin{bmatrix} v \\ w \end{bmatrix},$$

where v, $w \in \mathbb{R}^n$. We begin by introducing zeros into the vector w

Let $U_{1,2}$ be a rotator in the $(1,2)$ plane such that $U_{1,2}^T w$ has a zero in position 1.[2] Then let $U_{2,3}$ be a rotator in the $(2,3)$ plane such that $U_{2,3}^T U_{1,2}^T w$ has zeros in positions 1 and 2. Continuing in this manner, produce $n-1$ rotators such that

$$U_{n-1,n}^T \cdots U_{2,3}^T U_{1,2}^T w = \gamma e_n,$$

where $\gamma = \pm\| w \|_2$. Let

$$U = U_{1,2}U_{2,3}\cdots U_{n-1,n}.$$

Then $U^T w = \gamma e_n$. Let

$$\mathcal{Q}_1 = \begin{bmatrix} U & \\ & U \end{bmatrix},$$

and let

$$x^{(1)} = \begin{bmatrix} v^{(1)} \\ w^{(1)} \end{bmatrix} = \mathcal{Q}_1^T \begin{bmatrix} v \\ w \end{bmatrix} = \begin{bmatrix} U^T v \\ \gamma e_n \end{bmatrix}.$$

Let $\mathcal{Q}_2$ be an orthogonal symplectic rotator acting in the $(n, 2n)$ plane that annihilates the γ. In other words,

$$\mathcal{Q}_2 x^{(1)} = \begin{bmatrix} v^{(2)} \\ 0 \end{bmatrix}.$$

Next let $Y_{n-1,n}$ be a rotator in the $(n-1,n)$ plane such that $Y_{n-1,n}^T v^{(2)}$ has a zero in the nth position. Then let $Y_{n-2,n-1}$ be a rotator in the $(n-2, n-1)$ plane such that $Y_{n-2,n-1}^T Y_{n-1,n}^T v^{(2)}$ has zeros in positions $n-1$ and n. Continuing in this manner, produce rotators $Y_{n-3,n-2}, \ldots, Y_{1,2}$ such that $Y_{1,2}^T \cdots Y_{n-1,n}^T v^{(2)} = \beta e_1$, where $\beta = \pm\| v^{(2)} \|$. Letting $Y = Y_{n-1,n}\cdots Y_{1,2}$, we have $Y^T v^{(1)} = \beta e_1$. Let

$$\mathcal{Q}_3 = \begin{bmatrix} Y & \\ & Y \end{bmatrix},$$

and

$$\mathcal{Q} = \mathcal{Q}_1\mathcal{Q}_2\mathcal{Q}_3.$$

This is the desired orthogonal symplectic transformation matrix.

Each of U and Y is a product of $n-1$ rotators. Thus $\mathcal{Q}$ is a product of $2(n-1)$ symplectic double rotators and one symplectic single rotator. The transformation $\mathcal{H} \to \mathcal{Q}^T\mathcal{H}\mathcal{Q}$ is effected by applying the rotators in the correct order.

[2]Generically it will be the case that $w_1 \neq 0$, and the rotator $U_{1,2}$ is nontrivial. However, it could happen that $w_1 = 0$. If $w_2 = 0$ as well, then $U_{1,2}$ could be an arbitrary rotator in principle. However, we will insist on the following rule: In any situation where the entry that is to be made zero is already zero to begin with, then the trivial rotator (the identity matrix) must be used.

ETNA
Kent State University
etna@mcs.kent.edu

5. Why the algorithm works (real-eigenvalue case). The transforming matrix $\mathcal{Q}$ was designed so that $\mathcal{Q}^T x = \mathcal{Q}_3^T \mathcal{Q}_2^T \mathcal{Q}_1^T x = \beta e_1$. Thus $\mathcal{Q} e_1 = \beta^{-1} x$; the first column of $\mathcal{Q}$ is proportional to x. Therefore we get the deflation shown in (4.2).

The more challenging task is to show that $\mathcal{Q}^T \mathcal{H}^2 \mathcal{Q}$ is in skew-Hamiltonian Schur form. To this end, let

$$\mathcal{H}_1 = \mathcal{Q}_1^T \mathcal{H} \mathcal{Q}_1, \qquad \mathcal{H}_2 = \mathcal{Q}_2^T \mathcal{H}_1 \mathcal{Q}_2, \qquad \text{and} \qquad \mathcal{H}_3 = \mathcal{Q}_3^T \mathcal{H}_2 \mathcal{Q}_3.$$

Then

$$\mathcal{H}_3 = \mathcal{Q}^T \mathcal{H} \mathcal{Q},$$

so our objective is to show that $\mathcal{H}_3^2$ is in skew-Hamiltonian Schur form. We will do this by showing that the skew-Hamiltonian Schur form is preserved at every step of the algorithm.

We start with the transformation from $\mathcal{H}$ to $\mathcal{H}_1$. Since

$$\mathcal{H}_1^2 = \mathcal{Q}_1^T \mathcal{H}^2 \mathcal{Q}_1 = \begin{bmatrix} U^T B U & U^T N U \\ 0 & U^T B^T U \end{bmatrix},$$

it suffices to study the transformation $B \rightarrow U^T B U$. Since U is a product of rotators, the transformation from B to $U^T B U$ can be effected by applying the rotators one after another. We will show that each of these rotators preserves the upper-triangular structure of B, effecting a swap of two eigenvalues.

We will find it convenient to use computer programming notation here. We will treat B as an array whose entries can be changed. As we apply the rotators, B will gradually be transformed to $U^T B U$, but we will continue to refer to the matrix as B rather than giving it a new name each time we change it.

Our argument uses the fact that w is an eigenvector of B^T: The equation $\mathcal{H} x = x\lambda$ implies $\mathcal{H}^2 x = x\lambda^2$, which implies $B^T w = w\lambda^2$. (It can happen that $w = 0$. Our argument covers that case.) The transformations in B correspond to transformations in w. For example, when we change B to $U_{1,2}^T B U_{1,2}$ to get a new "B", we also change w to $U_{1,2}^T w$, and this is our new "w". Thus we are also treating w as an array. It will start out as the true w and be transformed gradually to the final "w", which is γe_n.

Let us start with the original B and w. It can happen that $w_1 = 0$, in which case $U_{1,2} = I$. Thus the new B, after the (trivial) transformation is still upper triangular. Now assume $w_1 \neq 0$. Then the rotator $U_{1,2}$ is non trivial, and the operation $w \leftarrow U_{1,2}^T w$ gives a new w that satisfies $w_1 = 0$ and $w_2 \neq 0$. The transformation $B \leftarrow U_{1,2}^T B U_{1,2}$ recombines the first two rows of B and the first two columns, resulting in a matrix that is upper triangular, except that its $(2,1)$ entry could be nonzero. The new B and w still satisfy the relationship $B^T w = w\lambda^2$. Looking at the first component of this equation, remembering that $w_1 = 0$, we have $b_{2,1} w_2 = 0$. Thus $b_{2,1} = 0$, and B is upper triangular. Moreover, the second component of the equation implies $b_{2,2} w_2 = w_2 \lambda^2$, so $b_{2,2} = \lambda^2$. Since we originally had $b_{1,1} = \lambda^2$, the similarity transformation has reversed the positions of the first two eigenvalues in B.

Now consider the second step: $B \leftarrow U_{2,3}^T B U_{2,3}$. Immediately before this step, we may have $w_2 = 0$, in which case $U_{2,3} = I$, the step is trivial, and triangular form is preserved. If, on the other hand, $w_2 \neq 0$, then $U_{2,3}$ is a nontrivial rotator, and the transformation $w \leftarrow U_{2,3}^T w$ gives a new w that satisfies $w_1 = w_2 = 0$ and $w_3 \neq 0$. The corresponding transformation $B \leftarrow U_{2,3}^T B U_{2,3}$ operates on rows and columns 2 and 3, and gives a new B that is upper triangular, except that the $(3,2)$ entry could be nonzero. The new B and w continue to satisfy the relationship $B^T w = w\lambda^2$. Looking at the second component of this equation, we see that $b_{3,2} w_3 \neq 0$. Therefore $b_{3,2} = 0$, and B is upper triangular. Looking at

ETNA
Kent State University
etna@mcs.kent.edu

the third component of the equation, we find that $b_{3,3} = \lambda^2$. Thus another eigenvalue swap has occurred.

Clearly this same argument works for all stages of the transformation. We conclude that the final B, which is really $U^T BU$, is upper triangular. We have $b_{nn} = \lambda^2$. In the generic case, in which the original w satisfies $w_1 \neq 0$, each of the transforming rotators is nontrivial and moves the eigenvalue λ^2 downward. In the end, λ^2 is at the bottom of the matrix, and each of the other eigenvalues has been moved up one position.

In the nongeneric case, suppose $w_1 = w_2 = \cdots = w_k = 0$, and $w_{k+1} \neq 0$. Then the $(k+1)$st component of the equation $B^T w = w\lambda^2$ implies that $b_{k+1,k+1} = \lambda^2$. Since also $b_{1,1} = \lambda^2$, we see that the nongeneric case (with $w \neq 0$) cannot occur unless λ is a multiple eigenvalue of $\mathcal{H}$. In this case the first k transformations are trivial, then the subsequent nontrivial transformations moves the eigenvalue $b_{k+1,k+1} = \lambda^2$ to the bottom.

In any event, we have

$$\mathcal{H}_1^2 = \begin{bmatrix} B^{(1)} & N^{(1)} \\ 0 & B^{(1)T} \end{bmatrix},$$

where $B^{(1)}$ is upper triangular.

Now consider the transformation from $\mathcal{H}_1$ to $\mathcal{H}_2 = \mathcal{Q}_2^T \mathcal{H}_1 \mathcal{Q}_2$, which is effected by a single symplectic rotator in the $(n, 2n)$ plane. The 2×2 submatrix of $\mathcal{H}_1^2$ extracted from rows and columns n and $2n$ is

$$\begin{bmatrix} b_{nn} & 0 \\ 0 & b_{nn} \end{bmatrix}.$$

(Recall that the "N" matrix is skew symmetric.) This submatrix is a multiple of the identity matrix, so it remains unchanged under the transformation from $\mathcal{H}_1^2$ to $\mathcal{H}_2^2$. An inspection of the other entries in rows and columns n and $2n$ of $\mathcal{H}_1^2$ shows that no unwanted new nonzero entries are produced by the transformation. We have

$$\mathcal{H}_2 = \begin{bmatrix} B^{(2)} & N^{(2)} \\ 0 & B^{(2)T} \end{bmatrix},$$

where $B^{(2)}$ is upper triangular.

Now consider the transformation from $\mathcal{H}_2$ to $\mathcal{H}_3$. Since

$$\mathcal{H}_3^2 = \mathcal{Q}_3^T \mathcal{H}_2^2 \mathcal{Q}_3 = \begin{bmatrix} Y^T B^{(2)} Y & Y^T N^{(2)} Y \\ 0 & Y^T B^{(2)T} Y \end{bmatrix},$$

it suffices to study the transformation of $B^{(2)}$ to $Y^T B^{(2)} Y$.

At this point the eigenvector equation $\mathcal{H}^2 x = x\lambda^2$ has been transformed to $\mathcal{H}_2^2 x^{(2)} = x^{(2)}\lambda^2$ or

$$\begin{bmatrix} B^{(2)} & N^{(2)} \\ 0 & B^{(2)T} \end{bmatrix} \begin{bmatrix} v^{(2)} \\ 0 \end{bmatrix} = \begin{bmatrix} v^{(2)} \\ 0 \end{bmatrix} \lambda^2,$$

which implies $B^{(2)} v^{(2)} = v^{(2)}\lambda^2$. The nth component of this equation is $b_{n,n}^{(2)} v_n^{(2)} = v_n^{(2)}\lambda^2$, implying that $b_{n,n}^{(2)} = \lambda^2$ if $v_n^{(2)} \neq 0$. In fact, we already knew this. The entry $v_n^{(2)}$ cannot be zero, unless the original w was zero. We showed above that if $w \neq 0$, then we must have $b_{n,n}^{(2)} = \lambda^2$.

Returning to our computer programming style of notation, we now write B and v in place of $B^{(2)}$ and $v^{(2)}$ and consider the transformation from B to $Y^T BY$, where $Bv = v\lambda^2$.

Since Y is a product of rotators, the transformation from B to Y^TBY can be effected by applying the rotators one after another. We will show that each of these rotators preserves the upper-triangular structure of B, effecting a swap of two eigenvalues.

The first transforming rotator is $Y_{n-1,n}$, which acts in the $(n-1,n)$ plane and is designed to set v_n to zero. If $v_n = 0$ already, we have $Y_{n-1,n} = I$, and B remains upper triangular. If $v_n \neq 0$, then $Y_{n-1,n}$ is nontrivial, and the transformation $B \leftarrow Y_{n-1,n}^T B Y_{n-1,n}$ leaves B in upper-triangular form, except that the $(n,n-1)$ entry could be nonzero. The transformation $v \leftarrow Y_{n-1,n}^T v$ gives a new v with $v_n = 0$ and $v_{n-1} \neq 0$. The new B and v still satisfy $Bv = v\lambda^2$. The nth component of this equation is $b_{n,n-1}v_{n-1} = 0$, which implies $b_{n,n-1} = 0$. Thus B is upper triangular. Moreover, the $(n-1)$st component of the equation is $b_{n-1,n-1}v_{n-1} = v_{n-1}\lambda^2$, which implies $b_{n-1,n-1} = \lambda^2$. Thus the eigenvalue λ^2 has been swapped upward.

The next rotator, $Y_{n-2,n-1}$, acts in the $(n-2,n-1)$ plane and sets v_{n-1} to zero. If $v_{n-1} = 0$, the rotator is trivial, and B remains upper triangular. If $v_{n-1} \neq 0$, then $Y_{n-2,n-1}$ is nontrivial, and the transformation $B \leftarrow Y_{n-2,n-1}^T B Y_{n-2,n-1}$ leaves B in upper-triangular form, except that the $(n-1,n-2)$ entry could be nonzero. The transformation $v \leftarrow Y_{n-2,n-1}^T v$ gives a new v with $v_n = v_{n-1} = 0$ and $v_{n-2} \neq 0$. The equation $Bv = v\lambda^2$ still holds. Its $(n-1)$st component is $b_{n-1,n-2}v_{n-2} = 0$, which implies $b_{n-1,n-2} = 0$. Thus B is still upper triangular. Furthermore the $(n-2)$nd component of the equation is $b_{n-2,n-2}v_{n-2} = v_{n-2}\lambda^2$, which implies $b_{n-2,n-2} = \lambda^2$. Thus the eigenvalue λ^2 has been swapped one more position upward.

Clearly this argument works at every step, and our final B, which is actually $B^{(3)}$, is upper triangular. Thus $\mathcal{H}_3^2$ has skew-Hamiltonian Hessenberg form.

In the generic case $v_n^{(2)} \neq 0$ (which holds as long as $w \neq 0$) each of the transforming rotators is nontrivial and moves the eigenvalue λ^2 upward. In the end, λ^2 has been moved to the top of B, and each of the other eigenvalues has been moved down one position.

In the generic case $w_1 \neq 0$, the first part of the algorithm marches λ^2 from top to bottom of B, moving each other eigenvalue upward. Then the last part of the algorithm exactly reverses the process, returning all eigenvalues of B to their original positions.

We already discussed above the nongeneric cases $w_1 = \cdots = w_k = 0$, $w_{k+1} \neq 0$. Now consider nongeneric cases in which $w = 0$. Then $\mathcal{Q}_1 = \mathcal{Q}_2 = I_{2n}$, and $\mathcal{H}_2 = \mathcal{H}$. There must be some largest k for which $v_k \neq 0$. If $k = 1$, then $\mathcal{Q}_3 = I_{2n}$ and $\mathcal{H}_3 = \mathcal{H}$; the whole transformation is trivial. Now assume $k > 1$. The kth component of the equation $Bv = v\lambda^2$ implies that $b_{k,k} = \lambda^2$. Since also $b_{1,1} = \lambda^2$, we see that this nongeneric case cannot occur unless $\mathcal{H}$ has a multiple eigenvalue. In this case the first $n-k$ rotators $Y_{n-1,n}$, $\ldots Y_{n-k,n-k+1}$ are all trivial and the rest of the rotators push the eigenvalue $b_{k,k} = \lambda^2$ to the top of B.

6. Accommodating complex eigenvalues. Now we drop the assumption that the eigenvalues are all real. Then in the real skew-Hamiltonian Schur form

$$\mathcal{H}^2 = \begin{bmatrix} B & N \\ 0 & B^T \end{bmatrix},$$

B is quasitriangular; that is, it is block triangular with 1×1 and 2×2 blocks on the main diagonal. Each 2×2 block houses a complex conjugate pair of eigenvalues of $\mathcal{H}^2$. Each 1×1 block is a positive real eigenvalue of $\mathcal{H}^2$. There are no non-positive real eigenvalues because, by assumption, $\mathcal{H}$ has no purely imaginary eigenvalues. Suppose there are l diagonal blocks of dimensions $n_1, \ldots, n_l$. Thus n_i is either 1 or 2 for each i.

As before, our objective is to build an orthogonal symplectic $\mathcal{Q}$ that causes a deflation as explained in section 3. How we proceed depends upon whether n_1 is 1 or 2. If $n_1 = 1$, we

ETNA
Kent State University
etna@mcs.kent.edu

do exactly as described in section 4.

Now suppose $n_1 = 2$. In this case span$\{e_1\}$ is not invariant under $\mathcal{H}^2$ but span$\{e_1, e_2\}$ is. It follows that span$\{e_1, e_2, \mathcal{H}e_1, \mathcal{H}e_2\}$ is invariant under $\mathcal{H}$. For now we will make the generically valid assumption that this space has dimension 4. Letting $E = \begin{bmatrix} e_1 & e_2 \end{bmatrix}$, we have

$$\mathcal{H} \begin{bmatrix} E & \mathcal{H}E \end{bmatrix} = \begin{bmatrix} E & \mathcal{H}E \end{bmatrix} \begin{bmatrix} 0 & C \\ I & 0 \end{bmatrix},$$

where $C = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$ has complex conjugate eigenvalues μ and $\overline{\mu}$. Thus the eigenvalues of $\mathcal{H}$ associated with the invariant subspace span$\{e_1, e_2, \mathcal{H}e_1, \mathcal{H}e_2\}$ are λ, $-\lambda$, $\overline{\lambda}$, and $-\overline{\lambda}$, where $\lambda^2 = \mu$. Without loss of generality assume $\Re(\lambda) < 0$.

From span$\{e_1, e_2, \mathcal{H}e_1, \mathcal{H}e_2\}$ extract a two-dimensional subspace $\mathcal{S}_2$, invariant under $\mathcal{H}$, associated with the eigenvalues λ and $\overline{\lambda}$. This subspace is necessarily isotropic by Proposition 2.1. We will build a symplectic orthogonal $\mathcal{Q}$ whose first two columns span $\mathcal{S}_2$. This will guarantee that we get a deflation as in (3.1), where $\hat{F}_1$ is 2×2 and has λ and $\overline{\lambda}$ as its eigenvalues. Of course $\mathcal{Q}$ must have other properties that guarantee that the skew-Hamiltonian form of the squared matrix is preserved. We will outline the algorithm first; then we will explain why it works.

Outline of the algorithm, complex-eigenvalue case. Our procedure is very similar to that in the case of a real eigenvalue. Let

$$X = \begin{bmatrix} V \\ W \end{bmatrix}$$

denote a $2n \times 2$ matrix whose columns are an orthonormal basis of $\mathcal{S}_2$. We can choose the basis so that $w_{11} = 0$. Isotropy implies $X^T J X = 0$. The first part of the algorithm applies a sequence of rotators that introduce zeros into W. We will use the notation $U_{i,i+1}^{(j)}$, $j = 1, 2$, to denote a rotator acting in the $(i, i+1)$ plane such that $U_{i,i+1}^{(j)T}$ introduces a zero in position $w_{i,j}$. (Here we are using the computer programming notation again, referring to the (i, j) position of an array that started out as W but has subsequently been transformed by some rotators.) The rotators are applied in the order $U_{2,3}^{(1)}$, $U_{1,2}^{(2)}$, $U_{3,4}^{(1)}$, $U_{2,3}^{(2)}$, $\ldots$, $U_{n-1,n}^{(1)}$, $U_{n-2,n-1}^{(2)}$. The rotators are carefully ordered so that the zeros that are created are not destroyed by subsequent rotators. We can think of the rotators as coming in pairs. The job of the pair $U_{i+1,i+2}^{(1)}$, $U_{i,i+1}^{(2)}$ together is to zero out the ith row of W.[3] Let

$$U = U_{2,3}^{(1)} U_{1,2}^{(2)} \cdots U_{n-1,n}^{(1)} U_{n-2,n-1}^{(2)}.$$

Then $U^T W = W^{(1)}$, where

$$W^{(1)} = \begin{bmatrix} 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & w_{n-1,2}^{(1)} \\ w_{n,1}^{(1)} & w_{n,2}^{(1)} \end{bmatrix}.$$

[3] Actually $U_{i+1,i+2}^{(1)}$ zeros out the entry $w_{i+1,1}$. This makes it possible for $U_{i,i+1}^{(2)}$ to zero out the entry $w_{i,1}$, thereby completing the task of forming zeros in row i.

Let

$$\mathcal{Q}_1 = \begin{bmatrix} U & \\ & U \end{bmatrix},$$

and let

$$X^{(1)} = \mathcal{Q}_1^T X = \begin{bmatrix} V^{(1)} \\ W^{(1)} \end{bmatrix}.$$

Next let $\mathcal{Q}_2$ denote an orthogonal symplectic matrix acting rows $n-1$, n, $2n-1$, $2n$ such that $\mathcal{Q}_2^T$ zeros out the remaining nonzero entries in $W^{(1)}$, that is,

$$X^{(2)} = \mathcal{Q}_2^T X^{(1)} = \begin{bmatrix} V^{(2)} \\ 0 \end{bmatrix}.$$

This can be done stably by a product of four rotators. Starting from the configuration

$$\begin{bmatrix} v_{n-1,1} & v_{n-1,2} \\ v_{n,1} & v_{n,2} \\ 0 & w_{n-1,2} \\ w_{n,1} & w_{n,2} \end{bmatrix},$$

(leaving off the superscripts to avoid notational clutter) a symplectic rotator in the $(n, 2n)$ plane sets the $w_{n,1}$ entry to zero. Then a symplectic double rotator acting in the $(n-1, n)$ and $(2n-1, 2n)$ planes transforms the $w_{n-1,2}$ and $v_{n,1}$ entries to zero simultaneously, as we explain below. Finally a symplectic rotator acting in the $(n, 2n)$ plane zeros out the $w_{n,2}$ entry. $\mathcal{Q}_2^T$ is the product of these rotators.

The simultaneous appearance of two zeros is a consequence of isotropy. The condition $X^T JX = 0$, which was valid initially, continues to hold as X is transformed, since each of the transforming matrices is symplectic. The isotropy forces $w_{n-1,2}$ to zero when $v_{n,1}$ is set to zero, and vice versa.

The final part of the algorithm applies a sequence of rotators that introduce zeros into $V^{(2)}$. Notice that the $(n, 1)$ entry of $V^{(2)}$ is already zero. We will use the notation $Y_{i-1,i}^{(j)}$, $j = 1, 2$, to denote a rotator acting in the $(i-1, i)$ plane such that $Y_{i-1,i}^{(j)T}$ introduces a zero in position $v_{i,j}$ (actually the transformed $v_{i,j}$). The rotators are applied in the order $Y_{n-2,n-1}^{(1)}$, $Y_{n-1,n}^{(2)}$, $Y_{n-3,n-2}^{(1)}$, $Y_{n-2,n-1}^{(2)}, \ldots, Y_{1,2}^{(1)}, Y_{2,3}^{(2)}$. Again the rotators are ordered so that the zeros that are created are not destroyed by subsequent rotators, and again the rotators come in pairs. The job of the pair $Y_{i-2,i-1}^{(1)}$, $Y_{i-1,i}^{(2)}$ together is to zero out the ith row of V. Let

$$Y = Y_{n-2,n-1}^{(1)} Y_{n-1,n}^{(2)} \cdots Y_{1,2}^{(1)} Y_{2,3}^{(2)}.$$

Then $Y^T V^{(2)} = V^{(3)}$, where

$$V^{(3)} = \begin{bmatrix} v_{11}^{(3)} & v_{12}^{(3)} \\ 0 & v_{22}^{(3)} \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{bmatrix}.$$

ETNA
Kent State University
etna@mcs.kent.edu

Let

$$\mathcal{Q}_3 = \begin{bmatrix} Y & \\ & Y \end{bmatrix},$$

and let

$$X^{(3)} = \mathcal{Q}_3^T X^{(2)} = \begin{bmatrix} V^{(3)} \\ 0 \end{bmatrix}.$$

Let

$$\mathcal{Q} = \mathcal{Q}_1 \mathcal{Q}_2 \mathcal{Q}_3.$$

This is the desired orthogonal symplectic transformation matrix. $\mathcal{Q}$ is a product of approximately $8n$ rotators; the transformation $\mathcal{H} \to \mathcal{Q}^T \mathcal{H} \mathcal{Q}$ is effected by applying the rotators in order.

7. Why the algorithm works. The transforming matrix $\mathcal{Q}$ was designed so that $\mathcal{Q}^T X = \mathcal{Q}_3^T \mathcal{Q}_2^T \mathcal{Q}_1^T X = X^{(3)}$. Thus $\mathcal{Q} X^{(3)} = X$. Since $\mathcal{R}(X^{(3)}) = \text{span}\{e_1, e_2\}$, we have $\mathcal{Q}\text{span}\{e_1, e_2\} = \mathcal{S}_2$; the first two columns of $\mathcal{Q}$ span the invariant subspace of $\mathcal{H}$ associated with λ and $\overline{\lambda}$. Therefore we get the deflation shown in (3.1).

Again the more challenging task is to show that $\mathcal{Q}^T \mathcal{H}^2 \mathcal{Q}$ is in skew-Hamiltonian Schur form. Let

$$\mathcal{H}_1 = \mathcal{Q}_1^T \mathcal{H} \mathcal{Q}_1, \qquad \mathcal{H}_2 = \mathcal{Q}_2^T \mathcal{H}_1 \mathcal{Q}_2, \qquad \text{and} \qquad \mathcal{H}_3 = \mathcal{Q}_3^T \mathcal{H}_2 \mathcal{Q}_3,$$

so that

$$\mathcal{H}_3 = \mathcal{Q}^T \mathcal{H} \mathcal{Q},$$

as before. We must show that $\mathcal{H}_3^2$ is in skew-Hamiltonian real Schur form.

We will cover the cases $n_1 = 2$ and $n_1 = 1$. In the case $n_1 = 2$, X denotes an $n \times 2$ matrix such that the two-dimensional space $\mathcal{R}(X)$ is invariant under $\mathcal{H}$. In the case $n_1 = 1$, X will denote the eigenvector x from sections 4 and 5. For this X, the one-dimensional space $\mathcal{R}(X)$ is invariant under $\mathcal{H}$.

Since $\mathcal{R}(X)$ is invariant under $\mathcal{H}$, we have $\mathcal{H}X = X\Lambda$, for some $n_i \times n_i$ matrix Λ. If $n_i = 1$, then $\Lambda = \lambda$, a negative real eigenvalue. If $n_i = 2$, Λ has complex eigenvalues λ and $\overline{\lambda}$ lying in the left half-plane. Consequently $\mathcal{H}^2 X = X\Lambda^2$ or

$$\begin{bmatrix} B & N \\ 0 & B^T \end{bmatrix} \begin{bmatrix} V \\ W \end{bmatrix} = \begin{bmatrix} V \\ W \end{bmatrix} \Lambda^2,$$

which implies

$$B^T W = W \Lambda^2.$$

Thus $\mathcal{R}(W)$ is invariant under B^T.

Recalling that B is quasitriangular with main-diagonal blocks of dimension n_i, $i = 1, \ldots, l$, we introduce the partition

$$B = \begin{bmatrix} B_{11} & B_{12} & \cdots & B_{1l} \\ & B_{22} & & B_{2l} \\ & & \ddots & \vdots \\ & & & B_{ll} \end{bmatrix},$$

ETNA
Kent State University
etna@mcs.kent.edu

where B_{ii} is $n_i \times n_i$. Partitioning W conformably with B, we have

$$W = \begin{bmatrix} W_1 \\ \vdots \\ W_l \end{bmatrix},$$

where W_i is $n_i \times n_1$.

We start with the transformation from $\mathcal{H}$ to $\mathcal{H}_1$, for which it suffices to study the transformation from B to U^TBU. Again we will use computer programming notation, treating B and W as arrays whose entries get changed as the rotators are applied.

The equation $B^TW = W\Lambda^2$ implies that

$$\begin{bmatrix} B_{11}^T & 0 \\ B_{12}^T & B_{22}^T \end{bmatrix} \begin{bmatrix} W_1 \\ W_2 \end{bmatrix} = \begin{bmatrix} W_1 \\ W_2 \end{bmatrix} \Lambda^2. \tag{7.1}$$

In particular $B_{11}^TW_1 = W_1\Lambda^2$, which reminds us that the eigenvalues of B_{11} are the same as those of Λ^2. Generically W_1 will be nonsingular, implying that the equation $B_{11}^TW_1 = W_1\Lambda^2$ is a similarity relationship between B_{11}^T and Λ^2, but it could happen that $W_1 = 0$. Lemma 9 of [4] implies that these are the only possibilities: If W_1 is not zero, then it must be nonsingular. For now we will make the generically valid assumption that W_1 is nonsingular. The nongeneric cases will be discussed later.

The algorithm begins by creating zeros in the top of W. Proceed until the first n_2 (1 or 2) rows of W have been transformed to zero. In the case $n_1 = 1$ (resp. $n_1 = 2$), this will require n_2 rotators (resp. *pairs* of rotators). We apply these rotators to B as well, effecting an orthogonal symplectic similarity transformation. The action of these rotators is confined to the first $n_1 + n_2$ rows and columns of B. The resulting new "B" remains quasitriangular, except that the block B_{21} could now be nonzero. At this point the relationship (7.1) will have been transformed to

$$\begin{bmatrix} B_{11}^T & B_{21}^T \\ B_{12}^T & B_{22}^T \end{bmatrix} \begin{bmatrix} 0 \\ W_2 \end{bmatrix} = \begin{bmatrix} 0 \\ W_2 \end{bmatrix} \Lambda^2. \tag{7.2}$$

Here we must readjust the partition in order to retain conformability. The block of zeros is $n_2 \times n_1$, so the new W_2 must be $n_1 \times n_1$. Since the original W_1 was nonsingular, the new W_2 must also be nonsingular. The new B_{11} and B_{22} are $n_2 \times n_2$ and $n_1 \times n_1$, respectively. The top equation of (7.2) is $B_{21}^TW_2 = 0$, which implies $B_{21} = 0$. Thus the quasitriangularity is preserved. Moreover, the second equation of (7.2) is $B_{22}^TW_2 = W_2\Lambda^2$, which implies that B_{22} has the eigenvalues of Λ^2 (and the new B_{11} must have the eigenvalues of the old B_{22}). Thus the eigenvalues have been swapped.

In the language of invariant subspaces, equation (7.2) implies that the space $\text{span}\{e_{n_1+1}, \ldots, e_{n_1+n_2}\}$ is invariant under B^T. Therefore the block B_{21}^T must be zero.

Now, proceeding inductively, assume that the first $n_2 + n_3 \cdots + n_k$ rows of W have been transformed to zero. At this point we have a transformed equation $B^TW = W\Lambda^2$. Assume inductively that B has quasitriangular form, where B_{kk} is $n_1 \times n_1$ and has the eigenvalues of Λ^2. For $i = 1, \ldots, k-1$, B_{ii} is $n_{i+1} \times n_{i+1}$ and has the same eigenvalues as the original $B_{i+1,i+1}$ had. For $i = k+1, \ldots, l$, B_{ii} has not yet been touched by the algorithm. Our current W has the form

$$W = \begin{bmatrix} 0 \\ W_k \\ \vdots \\ W_l \end{bmatrix},$$

ETNA
Kent State University
etna@mcs.kent.edu

where the zero block has $n_2 + \cdots + n_k$ rows, W_k is $n_1 \times n_1$ and nonsingular, and W_{k+1}, $\ldots$, W_l have not yet been touched by the algorithm. From the current version of the equation $B^T W = W\Lambda^2$ we see that

$$\left[\begin{array}{cc} B_{kk}^T & 0 \\ B_{k,k+1}^T & B_{k+1,k+1}^T \end{array}\right]\left[\begin{array}{c} W_k \\ W_{k+1} \end{array}\right] = \left[\begin{array}{c} W_k \\ W_{k+1} \end{array}\right]\Lambda^2.$$

The next step of the algorithm is to zero out the next n_{k+1} rows of W. The situation is exactly the same as it was in the case $k = 1$. Arguing exactly as in that case, we see that the quasitriangularity is preserved, and the eigenvalues are swapped.

At the end of the first stage of the algorithm we have

$$W = \left[\begin{array}{c} 0 \\ \vdots \\ 0 \\ W_l \end{array}\right],$$

where W_l is $n_1 \times n_1$ and nonsingular. This W is what we previously called $W^{(1)}$. The corresponding B (which is actually $B^{(1)}$) is quasitriangular. B_{ll} is $n_1 \times n_1$ and has the eigenvalues of Λ^2. Each other B_{ii} is $n_{i+1} \times n_{i+1}$ and has the eigenvalues of the original $B_{i+1,i+1}$. At this point we have transformed $\mathcal{H}$ to $\mathcal{H}_1 = \mathcal{Q}_1^T \mathcal{H} \mathcal{Q}_1$.

Before moving on to the next stage of the algorithm, we pause to consider what happens in the nongeneric cases. Assume W_1 is not nonsingular. Then W_1 must be zero. Suppose $W_1, \ldots, W_{j-1}$ are all zero and $W_j \neq 0$. Then, by Lemma 9 of [4], W_j must be $n_1 \times n_1$ and nonsingular. If we partition the equation $B^T W = W\Lambda^2$, then the jth block equation is $B_{jj}^T W_j = W_j \Lambda^2$. Thus B_{jj} is similar to Λ^2 and must therefore have the same eigenvalues as B_{11}. We conclude that this nongeneric case can occur only if the eigenvalues of Λ have multiplicity greater than one as eigenvalues of $\mathcal{H}$. In this case the first $n_2 + \ldots + n_{j-1}$ rotators (or pairs of rotators) are trivial. Once we reach W_j, the rotators become nontrivial, and the eigenvalues of the block B_{jj} (i.e. the eigenvalues of Λ^2) get swapped downward to the bottom, just as in the generic case. Another nongeneric case occurs when the entire matrix W is zero to begin with. We will say more about that case later.

Whether we are in the generic case or not, we have

$$\mathcal{H}_1^2 = \left[\begin{array}{cc} B^{(1)} & N^{(1)} \\ 0 & B^{(1)T} \end{array}\right],$$

where $B^{(1)}$ is quasi-triangular.

Now consider the transformation from $\mathcal{H}_1$ to $\mathcal{H}_2 = \mathcal{Q}_2^T \mathcal{H}_1 \mathcal{Q}_2$. We have to show that the skew-Hamiltonian Schur form of

$$\mathcal{H}_1^2 = \left[\begin{array}{cc} B^{(1)} & N^{(1)} \\ 0 & B^{(1)T} \end{array}\right]$$

is preserved by the transformation. If we write $\mathcal{H}_1^2$ in block form conformably with the blocking of $B^{(1)}$, we have $2l$ block rows and block columns. The action of the transformation affects only block rows and columns l and $2l$. If we consider the block-triangular form of $B^{(1)}$, we find that zero blocks are combined only with zero blocks and no unwanted nonzero entries are introduced except possibly in block $(2l, l)$.

To see that the $(2l, l)$ block also remains zero, examine our main equation

$$\left[\begin{array}{cc} B^{(1)} & N^{(1)} \\ 0 & B^{(1)T} \end{array}\right]\left[\begin{array}{c} V^{(1)} \\ W^{(1)} \end{array}\right] = \left[\begin{array}{c} V^{(1)} \\ W^{(1)} \end{array}\right]\Lambda^2,$$

ETNA
Kent State University
etna@mcs.kent.edu

bearing in mind that almost all of the entries of $W^{(1)}$ are zeros. Writing the equation in partitioned form and examining the lth and $2l$th block rows, we see that they amount to just

$$\begin{bmatrix} B_{ll}^{(1)} & N_{ll}^{(1)} \\ 0 & B_{ll}^{(1)T} \end{bmatrix} \begin{bmatrix} V_l^{(1)} \\ W_l^{(1)} \end{bmatrix} = \begin{bmatrix} V_l^{(1)} \\ W_l^{(1)} \end{bmatrix} \Lambda^2. \tag{7.3}$$

Every matrix in this picture has dimension $n_1 \times n_1$. The similarity transformation by Q_2 changes (7.3) to

$$\begin{bmatrix} B_{ll}^{(2)} & N_{ll}^{(2)} \\ M & B_{ll}^{(2)T} \end{bmatrix} \begin{bmatrix} V_l^{(2)} \\ 0 \end{bmatrix} = \begin{bmatrix} V_l^{(2)} \\ 0 \end{bmatrix} \Lambda^2, \tag{7.4}$$

where M is the block in question. Notice that $V_l^{(2)}$ cannot avoid being nonsingular. The second block equation in (7.4) is $MV_l^{(2)} = 0$, which implies $M = 0$, as desired. The first block equation is $B_{ll}^{(2)} V_l^{(2)} = V_l^{(2)} \Lambda^2$, which says that $B_{ll}^{(2)}$ is similar to Λ^2 and therefore has the same eigenvalues as Λ^2. If we prefer, we can use the language of invariant subspaces. Equation (7.4) shows that $\operatorname{span}\{e_1, \ldots, e_{n_1}\}$ is an invariant subspace of $\begin{bmatrix} B_{ll}^{(2)} & N_{ll}^{(2)} \\ M & B_{ll}^{(2)T} \end{bmatrix}$, forcing M to be zero. Either way, we conclude that the skew-Hamiltonian Schur form is preserved in stage 2 of the algorithm.

Now consider the transformation from $\mathcal{H}_2$ to $\mathcal{H}_3$, for which it suffices to study the transformation from $B^{(2)}$ to $B^{(3)} = Y^T B^{(2)} Y$. At this point the main equation $\mathcal{H}^2 X = X\Lambda^2$ has been transformed to $\mathcal{H}_2^2 X^{(2)} = X^{(2)} \Lambda^2$ or

$$\begin{bmatrix} B^{(2)} & N^{(2)} \\ 0 & B^{(2)T} \end{bmatrix} \begin{bmatrix} V^{(2)} \\ 0 \end{bmatrix} = \begin{bmatrix} V^{(2)} \\ 0 \end{bmatrix} \Lambda^2,$$

which implies $B^{(2)} V^{(2)} = V^{(2)} \Lambda^2$, so $\mathcal{R}(V^{(2)})$ is invariant under $B^{(2)}$.

Returning to our computer programming style of notation, we now write B and V in place of $B^{(2)}$ and $V^{(2)}$ and consider the transformation from B to $Y^T BY$, where $BV = V\Lambda^2$. For now we assume that we are in the generic case; we will discuss the remaining nongeneric cases later. Thus B is quasi-triangular. The block B_{ii} is $n_{i+1} \times n_{i+1}$ for $i = 1, \ldots, l-1$. The block B_{ll} is $n_1 \times n_1$ and has the eigenvalues of Λ^2. If we partition V conformably with B, we have

$$V = \begin{bmatrix} V_1 \\ \vdots \\ V_l \end{bmatrix},$$

where each V_i has its appropriate dimension. In particular V_l is $n_1 \times n_1$ and is, in the generic case, nonsingular. The equation $BV = V\Lambda^2$ and the quasitriangularity of B imply that

$$\begin{bmatrix} B_{l-1,l-1} & B_{l-1,l} \\ 0 & B_{ll} \end{bmatrix} \begin{bmatrix} V_{l-1} \\ V_l \end{bmatrix} = \begin{bmatrix} V_{l-1} \\ V_l \end{bmatrix} \Lambda^2. \tag{7.5}$$

The third stage of the algorithm is wholly analogous with the first stage, except that we work upward instead of downward. We begin by transforming the bottom n_l rows of V to zero. In the case $n_1 = 1$ (resp. $n_1 = 2$) this requires n_l rotators (resp. *pairs* of rotators). When we apply these rotators to B, their action is confined to the bottom $n_l + n_1$ rows and

ETNA
Kent State University
etna@mcs.kent.edu

columns, so the new B remains quasitriangular, except that the block $B_{l,l-1}$ could now be nonzero. At this point the relationship (7.5) will have been transformed to

$$\begin{bmatrix} B_{l-1,l-1} & B_{l-1,l} \\ B_{l,l-1} & B_{ll} \end{bmatrix} \begin{bmatrix} V_{l-1} \\ 0 \end{bmatrix} = \begin{bmatrix} V_{l-1} \\ 0 \end{bmatrix} \Lambda^2. \tag{7.6}$$

Just as in stage 1, we must readjust the partition in order to retain conformability. The block of zeros is $n_l \times n_1$, so the new V_{l-1} must be $n_1 \times n_1$, and it is nonsingular. The new $B_{l-1,l-1}$ and B_{ll} are $n_1 \times n_1$ and $n_l \times n_l$, respectively. The bottom equation of (7.6) is $B_{l,l-1}V_{l-1} = 0$, which implies $B_{l,l-1} = 0$. Thus the quasitriangularity is preserved. Moreover, the top equation of (7.6) is $B_{l-1,l-1}V_{l-1} = V_{l-1}\Lambda^2$, which implies that $B_{l-1,l-1}$ has the eigenvalues of Λ^2 (and the new B_{ll} must have the eigenvalues of the old $B_{l-1,l-1}$). Thus the eigenvalues have been swapped.

In the language of invariant subspaces, equation (7.6) implies that the space $\text{span}\{e_1, \ldots, e_{n_1}\}$ is invariant under $\begin{bmatrix} B_{l-1,l-1} & B_{l-1,l} \\ B_{l,l-1} & B_{ll} \end{bmatrix}$. Therefore the block $B_{l,l-1}$ must be zero.

This first step sets the pattern for all of stage 3. We will skip the induction step, which is just like the first step. Stage 3 preserves the quasitriangular form of B and swaps the eigenvalues of Λ^2 back to the top. In the end we have

$$\mathcal{H}_3^2 = \begin{bmatrix} B^{(3)} & N^{(3)} \\ 0 & B^{(3)T} \end{bmatrix},$$

where

$$B^{(3)} = \begin{bmatrix} B_{11}^{(3)} & B_{12}^{(3)} & \cdots & B_{1l}^{(3)} \\ 0 & B_{22}^{(3)} & & B_{2l}^{(3)} \\ \vdots & & \ddots & \vdots \\ 0 & & & B_{ll}^{(3)} \end{bmatrix},$$

where each $B_{ii}^{(3)}$ is $n_i \times n_i$ and has the same eigenvalues as the original B_{ii}.

It seems as if we are back where we started. Indeed we are, as far as $\mathcal{H}^2$ is concerned. But we must not forget that our primary interest is in $\mathcal{H}$, not $\mathcal{H}^2$. Thanks to this transformation, $\text{span}\{e_1\}$ (in the case $n_1 = 1$) or $\text{span}\{e_1, e_2\}$ (in the case $n_1 = 2$) is now invariant under $\mathcal{H}_3$ as well as $\mathcal{H}_3^2$, so we get the deflation (3.1).

The discussion is now complete, except for some more nongeneric cases. Suppose that at the beginning of stage 3, V_l is singular. Then, by Lemma 6 of [4], V_l must be zero. In fact this can happen only if the original W was zero. In this case the transformations $\mathcal{Q}_1$ and $\mathcal{Q}_2$ are both trivial, and we have $\mathcal{H}_2 = \mathcal{H}$. Thus the B and V with which we begin stage 3 are the original B and V. It must be the case that $V \neq 0$; in fact V has full rank n_1. If we partition V conformably with B, we have

$$V = \begin{bmatrix} V_1 \\ \vdots \\ V_l \end{bmatrix},$$

where V_i is $n_i \times n_1$. Suppose $V_j \neq 0$ and $V_{j+1}, \ldots, V_l$ are all zero, and suppose $j > 1$. Then, by Lemma 6 of [4], V_j is $n_1 \times n_1$ and nonsingular. Moreover, the equation $BV = V\Lambda^2$ implies that $B_{jj}V_j = V_j\Lambda^2$, so we see that B_{jj} is similar to Λ^2, so it has the same eigenvalues

ETNA
Kent State University
etna@mcs.kent.edu

as B_{11}. Thus this nongeneric case cannot arise unless the eigenvalues of Λ have multiplicity greater than one as eigenvalues of $\mathcal{H}$.

In this nongeneric case all of the rotators are trivial until the block V_j is reached. From that point on, all of the rotators are nontrivial, and the algorithm proceeds just as in the generic case. Because V_j is nonsingular, the arguments that worked in the generic case are valid here as well. The eigenvalues of $B_{jj} = V_j \Lambda^2 V_j^{-1}$ get swapped to the top of B.

Finally we consider the nongeneric case $j = 1$. Here V_1 is nonsingular, and all other V_i are zero. This is the lucky case, in which $\mathcal{R}(X) = \text{span}\{e_1\}$ (when $n_1 = 1$) or $\mathcal{R}(X) = \text{span}\{e_1, e_2\}$ (when $n_1 = 2$). In this case all rotators are trivial, $\mathcal{Q} = I_{2n}$, and $\mathcal{H}_3 = \mathcal{H}$.

Now the story is nearly complete.

8. One last difficulty with the nongeneric case. Our stated goal has been to get the left-half-plane eigenvalues into the $(1,1)$ block of the Hamiltonian Schur form. In the nongeneric cases we may fail to achieve this goal. First of all, in the "lucky" nongeneric case that we just mentioned above, $\text{span}\{e_1\}$ or $\text{span}\{e_1, e_2\}$ is already invariant under $\mathcal{H}$, and we get an immediate deflation for free. If the eigenvalue(s) associated with this invariant subspace happen to be in the right half plane, then we get a deflation of one or two unwanted eigenvalues in the $(1,1)$ block.

Something even worse happens in the other nongeneric cases. For clarity let us consider the case $n_1 = 1$ first. Then $X = x = \begin{bmatrix} v \\ w \end{bmatrix}$, an eigenvector of $\mathcal{H}$. In all nongeneric cases, $w_1 = 0$. Suppose this holds, and e_1 is not an eigenvector. Then $\text{span}\{e_1, \mathcal{H}e_1\}$ is two dimensional, and we will show immediately below that the space spanned by the first two columns of $\mathcal{Q}$ is exactly $\text{span}\{e_1, \mathcal{H}e_1\}$.[4] Since the eigenvalues associated with this invariant space are λ and $-\lambda$, the effect of this is that the first two columns of $\mathcal{H}_3 = \mathcal{Q}^T \mathcal{H} \mathcal{Q}$ have the form

$$\begin{bmatrix} \lambda & * \\ 0 & -\lambda \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{bmatrix}.$$

The desired deflation of λ is followed immediately by a trivial, undesired deflation of $-\lambda$.

Let $\mathcal{V}$ denote the space spanned by the first two columns of $\mathcal{Q}$. To see that $\text{span}\{e_1, \mathcal{H}e_1\} = \mathcal{V}$, recall that $\mathcal{Q}$ is the product of a large number of rotators, the first pair of which are $\text{diag}\{U_{12}, U_{12}\}$, and the last pair of which are $\text{diag}\{Y_{12}, Y_{12}\}$. These are the only rotators that act in the $(1,2)$ plane, and they are the only rotators that affect directly the first column of $\mathcal{Q}$. Let $Z = \text{diag}\{Y_{12}, Y_{12}\}$, and let $\mathcal{Q}_p = \mathcal{Q}Z^{-1}$. Then $\mathcal{Q}_p$ is the same as $\mathcal{Q}$, except that the final pair of rotators has not yet been applied. Since $w_1 = 0$, the rotator pair $\text{diag}\{U_{12}, U_{12}\}$ is trivial. Therefore $\mathcal{Q}_p$ is a product of rotators that do not act on the first column; the first column of $\mathcal{Q}_p$ is thus e_1. Notice also that the relationship $\mathcal{Q} = \mathcal{Q}_p Z$ implies that $\mathcal{V}$, the space spanned by the first two columns of $\mathcal{Q}$, is the same as the space spanned by the first two columns of $\mathcal{Q}_p$. Therefore $e_1 \in \mathcal{V}$. Recall also that $\mathcal{Q}$ is designed so that its first column is proportional to x, so $x \in \mathcal{V}$. But $x = \mathcal{H}e_1 + \lambda e_1$. Therefore $\mathcal{H}e_1 = x - \lambda e_1 \in \mathcal{V}$. We conclude that $\text{span}\{e_1, \mathcal{H}e_1\} = \mathcal{V}$.

The exact same difficulty arises in the case of complex eigenvalues. First of all, there can be the undesired trivial deflation of a right-half-plane pair of complex eigenvalues in the $(1,1)$

[4]Recalling that $x = \mathcal{H}e_1 + \lambda e_1$, notice that the condition $w_1 = 0$ means exactly that $\text{span}\{e_1, \mathcal{H}e_1\}$ is an isotropic subspace.

block. If that doesn't happen but we are in the nongeneric case $W_1 = 0$, then an argument just like the one we gave in the real-eigenvalue case shows that the space spanned by the first four columns of $\mathcal{Q}$ is exactly the isotropic invariant subspace span$\{e_1, e_2, \mathcal{H}e_1, \mathcal{H}e_2\}$. This means that the desired deflation of a pair λ, $\overline{\lambda}$ is immediately followed by the unwanted trivial deflation of the pair $-\lambda$, $-\overline{\lambda}$.

Thus in nongeneric cases we frequently end up with right-half-plane eigenvalues in the $(1, 1)$ block, and we are faced with the additional postprocessing task of swapping them out. Methods for swapping eigenvalues in the Hamiltonian real Schur form have been outlined by Byers [3] and Kressner [5, 6].

REFERENCES

[1] P. Benner and D. Kressner, *Fortran 77 subroutines for computing the eigenvalues of Hamiltonian matrices*, ACM Trans. Math. Software, 32 (2006), pp. 352–373.
[2] P. Benner, V. Mehrmann, and H. Xu, *A numerically stable, structure preserving method for computing the eigenvalues of real Hamiltonian or symplectic pencils*, Numer. Math., 78 (1998), pp. 329–358.
[3] R. Byers, *Hamiltonian and Symplectic Algorithms for the Algebraic Riccati Equation*, PhD thesis, Cornell University, 1983.
[4] D. Chu, X. Liu, and V. Mehrmann, *A numerically strongly stable method for computing the Hamiltonian Schur form*, Tech. Rep. 24-2004, TU Berlin, Institut für Mathematik, 2004.
[5] D. Kressner, *Numerical Methods and Software for General and Structured Eigenvalue Problems*, PhD thesis, TU Berlin, Institut für Mathematik, Berlin, Germany, 2004.
[6] D. Kressner, *Numerical Methods for General and Structured Eigenproblems*, Springer, 2005.
[7] A. J. Laub, *A Schur method for solving algebraic Riccati equations*, IEEE Trans. Auto. Control, AC24 (1979), pp. 913–921.
[8] C. C. Paige and C. Van Loan, *A Schur decomposition for Hamiltonian matrices*, Linear Algebra Appl., 41 (1981), pp. 11–32.
[9] C. F. Van Loan, *A symplectic method for approximating all the eigenvalues of a Hamiltonian matrix*, Linear Algebra Appl., 61 (1984), pp. 233–252.