TAIWANESE JOURNAL OF MATHEMATICS
Vol. 14, No. 3A, pp. 805-823, June 2010
This paper is available online at http://www.tjm.nsysu.edu.tw/

A RATIONAL SHIRA METHOD FOR THE HAMILTONIAN EIGENVALUE PROBLEM

Peter Benner and Cedric Effenberger

Abstract. The SHIRA method of Mehrmann and Watkins belongs among the structure preserving Krylov subspace methods for solving skew-Hamiltonian eigenvalue problems. It can also be applied to Hamiltonian eigenproblems by considering a suitable transformation. Structure-induced shift-and-invert techniques are employed to steer the algorithm towards the interesting region of the spectrum. However, the shift cannot be altered in the middle of the computation without discarding the information that has been accumulated so far. This paper shows how SHIRA can be combined with ideas from Ruhe's Rational Krylov algorithm to yield a method that permits an adjustment of shift after every step of the computation, adding greatly to the flexibility of the algorithm. We call this new method Rational SHIRA. A numerical example is presented to demonstrate its efficiency.

1. Introduction

1.1. The Hamiltonian eigenvalue problem

We consider the standard eigenvalue problem

$$Hx = \lambda x, \tag{1.1}$$

for a Hamiltonian matrix H. Hamiltonian matrices $H \in \mathbb{R}^{2n\times 2n}$ feature the explicit block structure

$$H = \begin{bmatrix} A & B \\ C & -A^T \end{bmatrix}, \qquad B = B^T, \quad C = C^T,$$

where A, B, C are real $n \times n$ matrices.

Hamiltonian matrices and eigenproblems arise in a variety of applications. They are ubiquitous in control theory, where they play an important role in various control

Received December 23, 2008, accepted March 7, 2009.
2000 *Mathematics Subject Classification*: .
Key words and phrases: .

design procedures (linear-quadratic optimal control, Kalman filtering, H_2- and H_∞-control, etc., see, e.g., [2, 17, 24, 30] and most textbooks on control theory), system analysis problems like stability radius, pseudo-spectra, H_∞-norm computations [8-10], and model reduction [1, 5, 6, 14, 26, 29]. Another source of eigenproblems exhibiting Hamiltonian structure is the linearization of certain quadratic eigenvalue problems [4, 18, 20, 28]. Further applications can be found in computational physics and chemistry, e.g. symplectic integrators for molecular dynamics [12, 16], methods for random phase approximation (RPA) [19], etc.

Hamiltonian matrices may equivalently be characterized as those that are skew-adjoint with respect to the bilinear form $\langle x, y\rangle_J := y^T Jx$ induced by the skew-symmetric matrix

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

where I_n denotes the $n \times n$ identity matrix. This definition is advantageous in that it leads us directly to two other types of structured matrices, which play an important role when dealing with Hamiltonian eigenvalue problems, namely skew-Hamiltonian and symplectic matrices. Skew-Hamiltonian matrices are those that are self-adjoint with respect to $\langle \cdot, \cdot \rangle_J$ and symplectic matrices $S \in \mathbb{R}^{2n\times 2n}$ fulfill the relation

$$S^T JS = J$$

which means they are orthogonal with respect to $\langle \cdot, \cdot \rangle_J$. Similarity transformations using symplectic matrices can be shown to preserve all of these structures.

Hamiltonian matrices arising in the aforementioned applications are often large and sparse and only a small portion of their spectrum is sought. Krylov subspace methods have proven to be a viable tool for solving this kind of task. They extract certain spectral information by projection of the original matrix onto a sequence of expanding Krylov subspaces. Recall that the Krylov subspace of order m associated with the matrix H and the starting vector u is defined to be

$$\mathcal{K}_m(H, u) := \operatorname{span}\{u, Hu, H^2u, \ldots, H^{m-1}u\}.$$

Note that the choice of the starting vector u may critically influence the outcome of the method. In practice, the process is frequently restarted to gradually improve the quality of the starting vector [3, 25].

The spectra of Hamiltonian matrices cannot be arbitrary. Real and purely imaginary eigenvalues are bound to occur in pairs $\{\lambda, -\lambda\}$, while general complex eigenvalues always occur in quadruples $\{\lambda, -\lambda, \overline{\lambda}, -\overline{\lambda}\}$. That is, the spectrum of a Hamiltonian matrix is symmetric with respect to both the real and the imaginary axis. We will refer to this property as the Hamiltonian spectral symmetry. Numerical methods that take this symmetry into account are capable of preserving the

eigenvalue pairings despite the presence of roundoff errors and thus return physically meaningful results. Moreover, exploiting the structure usually leads to more efficient and sometimes more accurate algorithms.

Krylov subspace methods can be redesigned to preserve the Hamiltonian structure. This is done by rearranging the computations such that J-orthogonal bases for the underlying Krylov subspaces are produced. Projections of the original matrix onto these spaces will then amount to partial symplectic similarity transformations and hence will inherit the Hamiltonian structure.

1.2. The isotropic Arnoldi process

Skew-Hamiltonian matrices are somewhat easier to handle in this context as we can take advantage of the following result.

Lemma 1.1. *Let $L \in \mathbb{R}^{2n\times 2n}$ be skew-Hamiltonian, then for every starting vector $u \in \mathbb{R}^{2n}$ the Krylov subspace $\mathcal{K} = \mathcal{K}_m(L, u)$ of order $m \in \mathbb{N}$, $m \leq n$, is isotropic, i.e., $y^T Jx = 0$ for all $x, y \in \mathcal{K}$.*

Proof. See [18, Prop. 3.3]. ∎

Hence, at least in theory, there is no need for J-orthogonalization and we are free to orthogonalize the basis vectors with respect to the standard inner product instead, i.e., we can apply the Arnoldi algorithm. In practice, however, this isotropy is distorted by roundoff error and has to be enforced numerically. This gives rise to the isotropic Arnoldi process of Mehrmann and Watkins [18], where each new basis vector u_{j+1} is orthogonalized not only against the previously computed basis vectors $u_1, \ldots, u_j$, but also against $Ju_1, \ldots, Ju_j$ to yield the augmented Arnoldi relation

$$LU_j = U_j T_j + JU_j S_j + u_{j+1} t_{j+1,j} e_j^T. \tag{1.2}$$

As was already mentioned, S_j would vanish in exact arithmetic. In finite precision it may happen to deviate from zero, but can still be assumed to be tiny and therefore neglected in our considerations. Moreover, since U_j is orthogonal and isotropic the extended matrix $[U, J^T U]$ is orthogonal and J-orthogonal. By projection we find

$$[U_j, J^T U_j]^T L [U_j, J^T U_j] = \begin{bmatrix} T_j & * \\ -S_j & T_j^T \end{bmatrix} \approx \begin{bmatrix} T_j & * \\ 0 & T_j^T \end{bmatrix}$$

meaning that the eigenvalues of T_j are approximations to double eigenvalues of the projected matrix.

The use of this procedure is not restricted to skew-Hamiltonian eigenproblems, though. It is valuable in the Hamiltonian case as well if applied to an appropriate

modification of the original matrix. For example, one might think of employing H^2, whose eigensystem is the same as H's except that eigenvalues have been squared and which is skew-Hamiltonian whenever H is Hamiltonian. This is a proper choice as long as extremal eigenvalues of H are required, but might be rather slow when it comes to interior eigenvalues. In [18], Mehrmann and Watkins propose the use of the compound shift-and-invert operators

$$L_1(\mu) = (H - \mu I)^{-1}(H + \mu I)^{-1}$$

for real and purely imaginary target shifts μ and

$$L_2(\mu) = (H - \mu I)^{-1}(H + \mu I)^{-1}(H - \overline{\mu} I)^{-1}(H + \overline{\mu} I)^{-1}$$

for general complex target shifts. Both of these operators are skew-Hamiltonian for any Hamiltonian matrix H and accelerate the convergence of eigenvalues near the pair $\{\mu, -\mu\}$ or the quadruple $\{\mu, -\mu, \overline{\mu}, -\overline{\mu}\}$, respectively, by simultaneously mapping them to eigenvalues of large modulus.

If the isotropic Arnoldi algorithm applied to either $L_1(\mu)$ or $L_2(\mu)$ is complemented with implicit restarts [25], we obtain the SHIRA method [18].

1.3. Contributions by this paper

With conventional SHIRA the target shift μ is chosen in the beginning of the computation and maintained throughout an entire run of the algorithm. The only way to alter the shift is to terminate the current run and start a new one, thereby discarding the Krylov basis that has been assembled so far. However, if two consecutive shifts are not too far apart from each other, the old Krylov subspace will also contain some approximations to eigenvectors corresponding to eigenvalues in the vicinity of the new shift. Conversely, the new Krylov subspace will develop unwanted components in the direction of eigenvectors corresponding to eigenvalues near the old shift, which the old Krylov basis could help to filter out. Sequences of close-by shifts occur, e.g., when the first shift is an initial guess, which is successively being improved as Ritz values progress towards actual eigenvalues. For these reasons it seems beneficial to circumvent sacrificing the Krylov subspace and the information it includes.

The Rational Krylov Method of Ruhe [21-23] permits us to transform an Arnoldi relation for any shifted and inverted operator back to an Arnoldi relation for the original matrix regardless of shift. Consequently, we may vary the shift at will without being forced to rerun the algorithm. The Krylov subspace is retained and expanded at every step. The price we have to pay is that the projected smaller eigenvalue problems are now generalized upper Hessenberg/triangular instead of upper Hessenberg standard eigenvalue problems.

As the Rational Krylov Method is designed to work with the simple shift-and-invert transformation $H \mapsto (H - \mu I)^{-1}$, it might seem inappropriate for the compound operators $L_1(\mu)$ and $L_2(\mu)$ at first glance. But if we write $L_1(\mu)$ as $(H^2 - \mu^2 I)^{-1}$, we observe that it can equally well be understood as a simple shift-and-invert operator for the matrix H^2 with the target shift μ^2. Accordingly, we can apply Rational Krylov to obtain an Arnoldi relation for H^2. Based upon this idea, a rational SHIRA algorithm for the Hamiltonian eigenvalue problem will be constructed in Section 2. Numerical evidence for the efficiency of this new method will be given in section 3. The paper concludes with a brief summary and some remarks in Section 4.

2. Rational SHIRA

Throughout this section, H will denote a Hamiltonian $2n \times 2n$ matrix so that H^2 is skew-Hamiltonian. We will make use of the induced structural properties without always reminding the reader of these.

2.1. Rational Krylov transformation for real or purely imaginary shifts

Assume we have a generalized Arnoldi recurrence of the form

$$H^2 U_{j-1} T_{j-1} = U_j K_{j,j-1},$$

where the columns $u_1, \ldots, u_j$ of $U_j \in \mathbb{R}^{2n \times j}$ constitute an orthonormal and isotropic basis of the Krylov subspace $\mathcal{K}_j(H^2, u_1)$, $T_{j-1} \in \mathbb{R}^{(j-1)\times(j-1)}$ is upper triangular and $K_{j,j-1} \in \mathbb{R}^{j\times(j-1)}$ is upper Hessenberg. The components resulting from the numerical J-orthogonalization of the basis vectors in (1.2) have been neglected here as they are zero in exact arithmetic. Note that in a standard Arnoldi recurrence, $T_{j-1} = I_{j-1}$. Allowing a more general matrix T_{j-1} adds the flexibility necessary for deriving the rational Krylov transformation in the following.

Given a shift μ_j, we calculate $(H^2 - \mu_j^2 I)^{-1} u_j$ and orthogonalize it against $u_1, \ldots, u_j$ and $Ju_1, \ldots, Ju_j$ to obtain

$$t_{j+1,j} u_{j+1} = (H^2 - \mu_j^2 I)^{-1} u_j - \sum_{i=1}^{j} t_{i,j} u_i - \sum_{i=1}^{j} s_{i,j} J u_i. \tag{2.1}$$

Let us confine ourselves to real or purely imaginary shifts μ_j for now, so the computation remains real. If we again neglect the components resulting from the J-orthogonalization, (2.1) can be rewritten as

$$(H^2 - \mu_j^2 I)^{-1} u_j = U_{j+1} t_j, \qquad t_j = (t_{i,j})_{i=1}^{j+1}.$$

Premultiplication by $H^2 - \mu_j^2 I$ yields

$$(H^2 - \mu_j^2 I) U_{j+1} t_j = u_j,$$

which can be further rearranged into

$$H^2 U_{j+1} t_j = U_{j+1}(e_j + t_j \mu_j^2), \tag{2.2}$$

where e_j designates the j-th unit vector. Combining the previous Arnoldi relation with (2.2) results in

$$H^2 U_{j+1} \tilde{T}_{j+1,j} = U_{j+1} \tilde{K}_{j+1,j}, \tag{2.3}$$

where

$$\begin{aligned} \tilde{T}_{j+1,j} &= \left[\begin{array}{ccc|c} & & & t_{1,j} \\ & T_{j-1} & & \vdots \\ & & & t_{j-1,j} \\ \hline 0 & \cdots & 0 & t_{j,j} \\ 0 & \cdots & 0 & t_{j+1,j} \end{array}\right], \\ \tilde{K}_{j+1,j} &= \left[\begin{array}{ccc|c} & & & t_{1,j}\mu_j^2 \\ & K_{j,j-1} & & \vdots \\ & & & 1 + t_{j,j}\mu_j^2 \\ \hline 0 & \cdots & 0 & t_{j+1,j}\mu_j^2 \end{array}\right]. \end{aligned} \tag{2.4}$$

Note that both $\tilde{T}_{j+1,j} \in \mathbb{R}^{(j+1)\times j}$ and $\tilde{K}_{j+1,j} \in \mathbb{R}^{(j+1)\times j}$ are rectangular upper Hessenberg matrices. In order to regain a generalized Arnoldi recurrence from (2.3), we determine orthogonal matrices $Q \in \mathbb{R}^{(j+1)\times(j+1)}$ and $Z \in \mathbb{R}^{j\times j}$, such that $Q^T \tilde{T}_{j+1,j} Z := T_{j+1,j}$ is upper triangular and $Q^T \tilde{K}_{j+1,j} Z := K_{j+1,j}$ is again upper Hessenberg. This can be achieved by the following bulge-chasing procedure, starting from the bottom and chasing the bulge off the top: use a Givens rotation applied to rows $j, j+1$ from the left to eliminate $t_{j+1,j}$. This introduces a bulge in the $(j+1, j-1)$ position in $\tilde{K}_{j+1,j}$ which in turn can be eliminated by a Givens rotation in planes $j-1, j$ applied from the right. Now the Hessenberg structure in $\tilde{K}_{j+1,j}$ is restored at the price of a bulge in the $(j, j-1)$ position of $\tilde{T}_{j+1,j}$. Eliminating this by a plane rotation applied to rows $j-1, j$ again introduces a bulge in the Hessenberg structure of $\tilde{K}_{j+1,j}$, now in the $(j, j-2)$ position. As before, this is eliminated by a rotation from the right applied to the corresponding columns, i.e., to columns $j-2, j-1$ this time. The process continues by alternately eliminating the introduced bulges in $\tilde{T}_{j+1,j}$ and $\tilde{K}_{j+1,j}$, respectively, in the same way as just described for the last two rows. In this way, the bulges move up and to the left one position during each step until the bulge is chased off the top. The accumulation of necessary rotations from the left yields Q while the rotations from the right form Z.

Now postmultiplying (2.3) by Z and utilizing these transformed matrices, we achieve

$$H^2 U_{j+1} Q T_{j+1,j} = U_{j+1,j} Q K_{j+1,j}.$$

Observe that its triangular shape causes the last row of $T_{j+1,j}$ to be zero, so we can partition

$$T_{j+1,j} = \begin{bmatrix} & T_j & \\ 0 & \cdots & 0 \end{bmatrix},$$

where $T_j \in \mathbb{R}^{j\times j}$ is upper triangular and square. Lastly, we let $V_{j+1} := U_{j+1}Q$ and our equation reduces to

$$H^2 V_j T_j = V_{j+1} K_{j+1,j}.$$

Since Q is orthogonal the columns of V_{j+1} are orthonormal as were those of U_{j+1}. Thus, we have restored a generalized Arnoldi relation. Because of

$$\|V_{j+1}^T J V_{j+1}\|_2 = \|Q^T U_{j+1}^T J U_{j+1} Q\|_2 = \|U_{j+1}^T J U_{j+1}\|_2 = 0, \tag{2.5}$$

V_{j+1} also inherits numerical isotropy from U_{j+1}. So, the Rational Krylov transformation does not interfere with the (generalized) isotropic Arnoldi process.

2.2. Rational Krylov transformation for general complex shifts

Let us now drop the assumption that μ_j be either real or purely imaginary and examine what can be done to handle a general complex target shift. We start again with the Arnoldi recurrence

$$H^2 U_{j-1} T_{j-1} = U_j K_{j,j-1},$$

where the columns $u_1,\ldots,u_j$ of U_j are orthonormal and isotropic, $T_{j-1}\in\mathbb{R}^{(j-1)\times(j-1)}$ is upper triangular and $K_{j,j-1} \in \mathbb{R}^{j\times(j-1)}$ is upper Hessenberg, once more ignoring components coming from the J-orthogonalization. In the event that μ_j is neither real nor purely imaginary the operation $L_1(\mu_j)u_j$ will fail to produce a real result. The classical SHIRA method overcomes this difficulty by employing the operator $L_2(\mu_j)$ instead, which is guaranteed to deliver a real vector regardless of the choice for μ_j. This strategy, however, is not viable in our case due to the fact that the expansion of $L_2(\mu_j)$ contains powers of H higher than H^2, which by far complicates the rational transformation. Preferably, we will adopt the approach described by Ruhe in [21] and treat the real and imaginary parts of $L_1(\mu_j)u_j$ separately.

First, we orthogonalize the real part $\Re\big[L_1(\mu_j)u_j\big]$ against $u_1,\ldots,u_j$ and Ju_1, $\ldots, Ju_j$ yielding

$$t_{j+1,j}u_{j+1} = \Re\big[L_1(\mu_j)u_j\big] - \sum_{i=1}^{j} t_{i,j}u_i - \sum_{i=1}^{j} s_{i,j}Ju_i.$$

Afterwards, we orthogonalize the imaginary part $\Im\big[L_1(\mu_j)u_j\big]$ against $u_1,\ldots,u_{j+1}$ and $Ju_1,\ldots,Ju_{j+1}$ to obtain

$$t_{j+2,j+1}u_{j+2} = \Im\big[L_1(\mu_j)u_j\big] - \sum_{i=1}^{j+1} t_{i,j+1}u_i - \sum_{i=1}^{j+1} s_{i,j+1}Ju_i.$$

If we once again neglect expressions associated with the J-orthogonalization, these can be rearranged into

$$\Re\big[(H^2-\mu_j^2 I)^{-1}u_j\big] = U_{j+2}t_j, \qquad t_j=[t_{1,j},\dots,t_{j+1,j},0]^T, \tag{2.6}$$

$$\Im\big[(H^2-\mu_j^2 I)^{-1}u_j\big] = U_{j+2}t_{j+1}, \qquad t_{j+1}=[t_{1,j+1},\dots,t_{j+2,j+1}]^T. \tag{2.7}$$

Together, (2.6) and (2.7) imply

$$(H^2-\mu_j^2)^{-1}u_j = U_{j+2}(t_j+\mathbf{i}t_{j+1}).$$

As with the simpler case above this can be converted into

$$H^2U_{j+2}(t_j+\mathbf{i}t_{j+1}) = U_{j+2}(e_j+t_j\mu_j^2+\mathbf{i}t_{j+1}\mu_j^2)$$

Splitting $\mu_j^2=\rho_j+\mathbf{i}\theta_j$ into its real and imaginary parts, we obtain

$$H^2U_{j+2}(t_j+\mathbf{i}t_{j+1}) = U_{j+2}\big(e_j+t_j\rho_j-t_{j+1}\theta_j+\mathbf{i}(t_j\theta_j+t_{j+1}\rho_j)\big). \tag{2.8}$$

Equation (2.8), too, is decomposed into its real and imaginary parts

$$H^2U_{j+2}t_j = U_{j+2}(e_j+t_j\rho_j-t_{j+1}\theta_j), \tag{2.9}$$

$$H^2U_{j+2}t_{j+1} = U_{j+2}(t_j\theta_j+t_{j+1}\rho_j), \tag{2.10}$$

which can be combined with the initial Arnoldi relation to produce

$$H^2U_{j+2}\tilde{T}_{j+2,j+1} = U_{j+2}\tilde{K}_{j+2,j+1}, \tag{2.11}$$

where

$$\tilde{T}_{j+2,j+1} = \left[\begin{array}{ccc|cc} & & & t_{1,j} & t_{1,j+1} \\ & T_{j-1} & & \vdots & \vdots \\ & & & t_{j-1,j} & t_{j-1,j+1} \\ \hline 0 & \cdots & 0 & t_{j,j} & t_{j,j+1} \\ 0 & \cdots & 0 & t_{j+1,j} & t_{j+1,j+1} \\ 0 & \cdots & 0 & 0 & t_{j+2,j+1} \end{array}\right], \tag{2.12}$$

and

$$\begin{aligned}&\tilde{K}_{j+2,j+1}\\ &= \left[\begin{array}{ccc|cc} & & & t_{1,j}\rho_j-t_{1,j+1}\theta_j & t_{1,j}\theta_j+t_{1,j+1}\rho_j \\ & K_{j,j-1} & & \vdots & \vdots \\ & & & t_{j-1,j}\rho_j-t_{j-1,j+1}\theta_j & t_{j-1,j}\theta_j+t_{j-1,j+1}\rho_j \\ & & & 1+t_{j,j}\rho_j-t_{j,j+1}\theta_j & t_{j,j}\theta_j+t_{j,j+1}\rho_j \\ \hline 0 & \cdots & 0 & t_{j+1,j}\rho_j-t_{j+1,j+1}\theta_j & t_{j+1,j}\theta_j+t_{j+1,j+1}\rho_j \\ 0 & \cdots & 0 & -t_{j+2,j+1}\theta_j & t_{j+2,j+1}\rho_j \end{array}\right].\end{aligned} \tag{2.13}$$

Using a similar bulge chasing procedure as before, orthogonal matrices $Q \in \mathbb{R}^{(j+2)\times(j+2)}$ and $Z \in \mathbb{R}^{(j+1)\times(j+1)}$ can be found, such that $Q^T\tilde{T}_{j+2,j+1}Z := T_{j+2,j+1}$ is upper triangular and $Q^T\tilde{K}_{j+2,j+1}Z := K_{j+2,j+1}$ is upper Hessenberg, again leaving the bottom row of $T_{j+2,j+1}$ to be zero. Note that this time, in order to chase the bulges which consist of two elements each in $\tilde{T}_{j+2,j+1}, \tilde{K}_{j+2,j+1}$, we need two rotations from the left and two from the right in order to chase the bulge up one row and one column to the left.

If we postmultiply (2.11) by Z, let $V_{j+2} := U_{j+2}Q$ and choose T_{j+1} to be the leading $j+1$ rows of $T_{j+2,j+1}$, we obtain the generalized Arnoldi relation

$$H^2V_{j+1}T_{j+1} = V_{j+2}K_{j+2,j+1}.$$

Thus, we have incremented the order of our Arnoldi decomposition by 2. As

$$\Re[L_1(\mu_j)u_j] = \frac{1}{2}\big(L_1(\mu_j)u_j + L_1(\overline{\mu}_j)u_j\big),$$
$$\Im[L_1(\mu_j)u_j] = \frac{1}{2\mathbf{i}}\big(L_1(\mu_j)u_j - L_1(\overline{\mu}_j)u_j\big),$$

the above procedure can be viewed as taking both $L_1(\mu_j)u_j$ and $L_1(\overline{\mu}_j)u_j$ and then adding the space spanned by them to the Krylov subspace.

2.3. Applying the operators

In every step of the expansion phase, the operator $(H-\mu_jI)^{-1}(H+\mu_jI)^{-1}$ has to be applied to a vector u. This can be accomplished by successively solving the linear systems

$$(H-\mu_jI)y = u$$

and

$$(H+\mu_jI)x = y.$$

Thanks to the Hamiltonian structure of H, only one LU-factorization is needed to solve both linear systems for we have

$$H+\mu_jI = J(H-\mu_jI)^TJ,$$

and $(H-\mu_jI)^T = U^TL^T$ is an LU-factorization of $(H-\mu_jI)^T$ whenever $H-\mu_jI = LU$ is an LU-factorization of $H-\mu_jI$. Bear in mind, though, that this LU-factorization has to be redone whenever the shift changes. Therefore, we ought not to alter the shift too often in order to save some factorizations.

2.4. Implicit restarts, locking and purging

In practice, we cannot arbitrarily extend our Krylov sequences due to memory constraints. The remedy is to frequently restart the computation using a better starting vector, which is to be deduced from the precedent run. Polynomial filters have

successfully been utilized to purify the starting vector by de-emphasizing unwanted components. In [25], Sorensen shows how to apply them implicitly through performing QR-iterations on the projected problem. Additionally, locking and purging techniques can be pursued in order to reduce the computational effort.

As was pointed out earlier, the Arnoldi recurrences generated by the Rational SHIRA method are typically of the generalized form

$$H^2 U_j T_j = U_{j+1} K_{j+1,j}. \tag{2.14}$$

Since T_j is always invertible unless $u_1, \ldots, u_j$ span an exact invariant subspace under H^2, we may turn this into a standard Arnoldi relation by multiplying T_j^{-1} from the right. Doing so, however, is strongly discouraged as especially smaller eigenvalues may be very sensitive to the roundoff error incurred in forming the product $K_{j+1,j}T_j^{-1}$ explicitly. Preferably, we seek to adapt the restarting and deflation strategies to work directly with the pair $(K_{j+1,j}, T_j)$.

Hence, a QZ-like approach must be taken to effect an implicit restart. Assume we want to apply the filter polynomial $p(H^2) = (H^2 - \rho_1 I)\cdots(H^2 - \rho_d I)$ to the decomposition (2.14). It suffices to investigate how a single linear factor $(H^2 - \rho I)$ of the polynomial can be applied. First, we split up the right hand side of (2.14) into

$$H^2 U_j T_j = U_j K_j + u_{j+1} k_{j+1,j} e_j^T, \qquad K_{j+1,j} = \begin{bmatrix} K_j \\ k_{j+1,j} e_j^T \end{bmatrix}, \tag{2.15}$$

which is another common way of writing down an Arnoldi relation. Afterwards, the equation is shifted by $-\rho U_j T_j$ to give

$$(H^2 - \rho I) U_j T_j = U_j (K_j - \rho T_j) + u_{j+1} k_{j+1,j} e_j^T. \tag{2.16}$$

Postmultiply (2.16) by the first unit vector e_1 and let Q_0 be a Householder reflector, such that $Q_0^T (K_j - \rho T_j) e_1 = \alpha e_1$, $\alpha \in \mathbb{R}$. Then, we have for $j > 1$

$$t_{1,1}(H^2 - \rho I) U_j e_1 = \alpha U_j Q_0 e_1 \tag{2.17}$$

since T_j is upper triangular. Thus, the first column of $U_j Q_0$ is some multiple of $(H^2 - \rho I)u_1$. Now, apply Q_0^T from the left to T_j and K_j. This will introduce a bulge in the $(2,1)$-position of T_j while leaving K_j's upper Hessenberg structure intact. Using the standard Hessenberg-triangular bulge-chasing procedure employed in a QZ step [13], we can annihilate this bulge by performing a Givens rotation Z_1 from the right on both T_j and K_j. T_j is now again in upper triangular form, but the Hessenberg form of K_j is disturbed by a bulge in the $(3,1)$-position. Eliminate it using a Givens rotation Q_1^T from the left on both T_j and K_j. Keep on applying Givens rotations Z_i from the right and Q_i^T from the left to chase the bulge out of these matrices. The resulting matrices $\hat{T}_j$ and $\hat{K}_j$ will again be of upper triangular

and upper Hessenberg form, respectively. If we accumulate the transformations $Q := Q_0 Q_1 \cdots Q_s$ and $Z := Z_1 Z_2 \cdots Z_{s+1}$, we can equate

$$\hat{T}_j = Q^T T_j Z, \quad \hat{K}_j = Q^T K_j Z.$$

Postmultiply (2.15) by Z and utilize our newly acquired matrices $\hat{T}_j$ and $\hat{K}_j$ together with $V_j := U_j Q$ to arrive at

$$H^2 V_j \hat{T}_j = V_j \hat{K}_j + u_{j+1} k_{j+1,j} e_j^T Z.$$

Observe that $e_j^T Z_i = e_j^T$ for all Z_i except the last one. Hence, $e_j^T Z$ has nonzeros only in its last two components. Observe furthermore that $Q_1^T, \ldots, Q_s^T$ do not act upon the first rows of T_j or K_j. Consequently, $Q_1, \ldots, Q_s$ do not touch the first column of Q_0, which means $Qe_1 = Q_0 e_1$. If we combine this with (2.17), we find that v_1, i.e. the first column of V_j, is a multiple of $(H^2 - \rho I)u_1$. Thus, we have successfully applied a linear factor to the starting vector. These steps may be repeated to process more linear factors until the complete filter polynomial has been applied. Note, however, that the last vector of the decomposition has to be truncated every time due to the fill-in in e_j^T.

For the purposes of locking and purging we propose a slight modification of the Krylov-Schur approach [27], which is capable of dealing with the expanded generalized Arnoldi relation (2.15). First, K_j and T_j are reduced to generalized real Schur form by means of the QZ-algorithm. Accordingly, orthogonal matrices, $Q \in \mathbb{R}^{j \times j}$ and $Z \in \mathbb{R}^{j \times j}$, are determined, such that $Q^T T_j Z := \hat{T}_j$ is upper triangular and $Q^T K_j Z := \hat{K}_j$ is upper quasi-triangular, i.e. block upper triangular having only 1×1- or 2×2-blocks on its diagonal. If we multiply (2.15) by Z from the right, let $V_j = U_j Q$ and deploy the transformed matrices $\hat{T}_j$ and $\hat{K}_j$, we obtain the generalized Krylov-Schur decomposition

$$H^2 V_j \hat{T}_j = V_j \hat{K}_j + u_{j+1} b^T, \tag{2.18}$$

where $b^T = k_{j+1,j} e_j^T Z$ is now, in general, a full vector. Partitioning

$$\hat{T}_j = \begin{bmatrix} \hat{T}_{1,1} & \hat{T}_{1,2} \\ 0 & \hat{T}_{2,2} \end{bmatrix}, \quad \hat{K}_j = \begin{bmatrix} \hat{K}_{1,1} & \hat{K}_{1,2} \\ 0 & \hat{K}_{2,2} \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix},$$

where both, $\hat{T}_{1,1}$ and $\hat{K}_{1,1}$, are $k \times k$ and $b_1 \in \mathbb{R}^k$, we may infer

$$H^2 V_k \hat{T}_{1,1} = V_k \hat{K}_{1,1} + u_{j+1} b_1^T \tag{2.19}$$

by looking only at the first k columns of (2.18). Since V_k has orthonormal columns and $\hat{T}_{1,1}$ and $\hat{K}_{1,1}$ are upper triangular and upper quasi-triangular, respectively, (2.19) itself poses a generalized Krylov-Schur decomposition. This fact can be

exploited for purging by using the reordering procedure from [15] to move unwanted Ritz values into the trailing part of the generalized Schur decomposition and then truncating. Moreover, after rearranging and taking norms, (2.19) provides us with a convergence criterion,

$$\|H^2 V_k \hat{T}_{1,1} - V_k \hat{K}_{1,1}\|_F = \|u_{j+1} b_1^T\|_F \leq \|b_1\|_2, \tag{2.20}$$

for the column space of V_k towards an invariant subspace under H^2. Hence, we may reorder (again using the procedure from [15]) every diagonal block of $\hat{K}_j$ into the position immediately following the leading block of previously locked eigenvalues while monitoring the elements of b. Once the leading components b_1 of b are found to be negligible, we may consider the corresponding eigenvalue (in case the block is 1×1) or pair of complex conjugate eigenvalues (in case of a 2×2-block) as converged. Subsequently, we may safely set the already tiny elements of b_1 to zero without sacrificing backward stability. Doing so deflates the problem because it enables us to transform only the trailing part back into an Arnoldi decomposition leaving the converged block as well as the associated H^2-invariant column space of V_k untouched and removing them from the active part of the computation. This constitutes our locking procedure.

It should be highlighted that only orthogonal transformations are applied to the Krylov basis U_j, and, therefore, all of the above techniques maintain their numerical isotropy, compare (2.5).

2.5. Postprocessing

As our method assembles a sequence of generalized, isotropic Arnoldi decompositions of H^2, the Ritz values θ_i we calculate from the projected problems will, of course, be approximations to eigenvalues of H^2. Taking into account the Hamiltonian structure and resulting spectral symmetry of H, though, we can readily conclude that adequate Ritz values for H are given by both $\sqrt{\theta_i}$ and $-\sqrt{\theta_i}$. It is safe to use them simultaneously (thereby doubling the number of Ritz values), because the enforced isotropy ensures only one Ritz value is picked up by the Arnoldi process for every positive/negative pair of H's eigenvalues. Hence, no spurious double eigenvalues of H can be forged this way. Note also that the different manner in which Rational SHIRA handles general complex shifts removes the ambiguity occurring with the original method, see [18, sect. 5.1], when inferring the eigenvalues of H.

When it comes to eigenvectors more work has to be done. Since the eigenspaces corresponding to a pair $\{\lambda, -\lambda\}$ of eigenvalues of H are merged into a single eigenspace associated with the eigenvalue λ^2 when H is squared, an eigenvector of H^2 will, in general, turn out to be a linear combination of several eigenvectors of H belonging to different eigenvalues and, therefore, will fail to be invariant under H. The easiest way of acquiring eigenvectors is probably to perform a few steps of

inverse iteration with the computed eigenvalue estimates. It might also be possible to derive a Hamiltonian Schur form from the skew-Hamiltonian Krylov-Schur relation (2.18) using a variant of the algorithm presented in [11] provided none of the real Ritz values are negative, although this remains to be further researched.

2.6. Complete algorithm

Putting the pieces together, we obtain the following algorithm.

INPUT: Hamiltonian matrix H, starting vector u_1, number of requested eigenvalue pairs p.
OUTPUT: p eigenvalue pairs, optionally corresponding eigenvectors.

$j := 1$
normalize $u_1 := \frac{1}{\|u_1\|} u_1$
$T_0 := []$
$K_{1,0} := []$
repeat
  pick a shift μ_j
  compute $w := (H - \mu_j I)^{-1}(H + \mu_j I)^{-1} u_j$ as described in Section 2.3
  if μ_j^2 is real **then**
    orthogonalize $\tilde{u}_{j+1} := w - U_j t_j$, where $t_j := U_j^T w$ (Gram-Schmidt)
    repeat orthogonalization as necessary
    normalize $u_{j+1} := \frac{1}{t_{j+1,j}} \tilde{u}_{j+1}$, where $t_{j+1,j} := \|\tilde{u}_{j+1}\|$
    form $\tilde{T}_{j+1,j}$ and $\tilde{K}_{j+1,j}$ as in (2.4)
    determine orthogonal matrices $Q \in \mathbb{R}^{(j+1)\times(j+1)}$, $Z \in \mathbb{R}^{j\times j}$, such that
    $T_{j+1,j} := Q^T \tilde{T}_{j+1,j} Z$ is upper triangular and
    $K_{j+1,j} := Q^T \tilde{K}_{j+1,j} Z$ is upper Hessenberg
    define T_j to be the first j rows of $T_{j+1,j}$
    set $U_{j+1} := U_{j+1} Q$
    $j := j + 1$
  else
    orthogonalize $\tilde{u}_{j+1} := \Re(w) - U_j t_j$, where $t_j := U_j^T \Re(w)$ (Gram-Schmidt)
    repeat orthogonalization as necessary
    normalize $u_{j+1} := \frac{1}{t_{j+1,j}} \tilde{u}_{j+1}$, where $t_{j+1,j} := \|\tilde{u}_{j+1}\|$
    orthogonalize $\tilde{u}_{j+2} := \Im(w) - U_{j+1} t_{j+1}$,
    where $t_{j+1} := U_{j+1}^T \Im(w)$ (Gram-Schmidt)
    repeat orthogonalization as necessary
    normalize $u_{j+2} := \frac{1}{t_{j+2,j+1}} \tilde{u}_{j+2}$, where $t_{j+2,j+1} := \|\tilde{u}_{j+2}\|$
    split $\mu_j^2 := \rho_j + \mathbf{i}\theta_j$
    form $\tilde{T}_{j+2,j+1}$ as in (2.12) and $\tilde{K}_{j+2,j+1}$ as in (2.13)

26 : determine orthogonal matrices $Q \in \mathbb{R}^{(j+2)\times(j+2)}$, $Z \in \mathbb{R}^{(j+1)\times(j+1)}$ such that,
$T_{j+2,j+1} := Q^T \tilde{T}_{j+2,j+1} Z$ is upper triangular and
$K_{j+2,j+1} := Q^T \tilde{K}_{j+2,j+1} Z$ is upper Hessenberg
27 : define T_{j+1} to be the first $j+1$ rows of $T_{j+2,j+1}$
28 : $U_{j+2} := U_{j+2} Q$
29 : $j := j + 2$
30 : **end if**
31 : reduce T_{j-1} and $K_{j,j-1}$ to generalized Schur form using the QZ algorithm
32 : test every Ritz value for convergence by appropriately reordering the Schur form
33 : lock converged Ritz values
34 : transform active part back into an Arnoldi decomposition
35 : optionally apply a filter polynomial
36 : **Until** p Ritz values have been locked
37 : take positive and negative roots of locked Ritz values as eigenvalue estimates
38 : determine corresponding eigenvectors by inverse iteration if requested

3. Numerical Experiments

In order to assess the numerical behavior of the Rational SHIRA method as it is described in Section 2, we have applied it to the "intelligent highway" problem taken from the benchmark collection [7, Ex. 15] using MATLAB 6.1. We chose the number of vehicles to be 500 resulting in a Hamiltonian matrix H of dimension 1998×1998. The matrix itself is very sparse as is shown in Figure 3.1, which allows for cheap solutions of linear systems involving this matrix or shifted versions thereof. The spectrum of the matrix (as returned by MATLAB's `eig` command) has the appearance of Figure 3.2 featuring both real and general complex, but no purely imaginary eigenvalues.

We start out performing two steps of Rational SHIRA with shift $\mu_1 = 0.7$. We keep track of the residuals of the individual Ritz values by means of (2.20) utilizing the Krylov-Schur approach outlined in 2.4. We then choose the new shift to be the Ritz value with smallest residual not less than 10^{-5} in order to speed up convergence for this Ritz value, but prevent the shifted operator from becoming nearly singular. We carry out two more steps of Rational SHIRA and then repeat this strategy of selecting another shift from the Ritz values. Continuing this scheme, where every shift is kept for two iterations and then replaced, yields the convergence history depicted in Figure 3.3. A Ritz value whose residual becomes less than 10^{-9} is considered converged and deflated from the active part of the computation as explained in 2.4. As doing so inhibits our residual estimate (2.20) (causing it to be zero), the residuals of deflated (locked) eigenvalues are always listed to be machine precision $u \approx 2.22 \cdot 10^{-16}$.

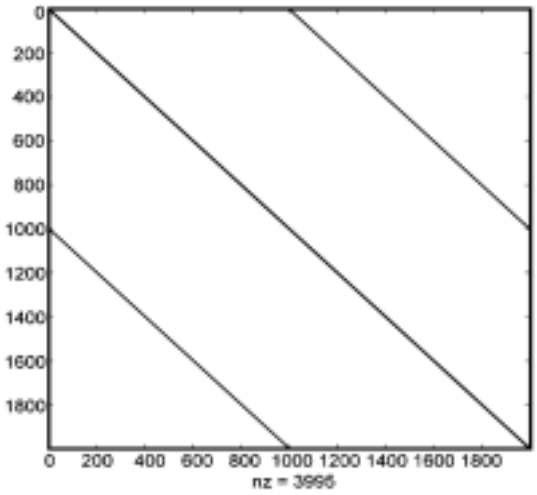

Fig. 3.1. Sparsity pattern of test matrix.

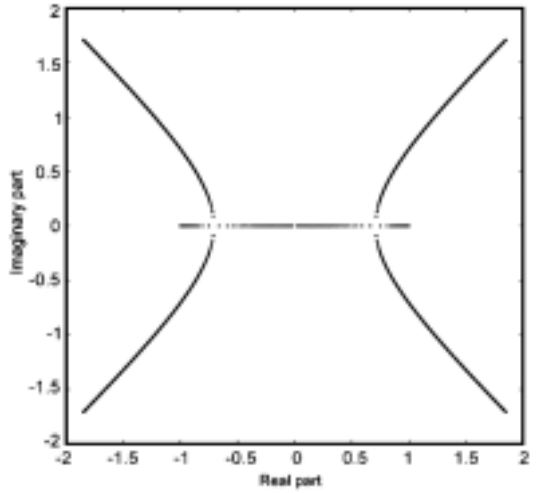

Fig. 3.2. Spectrum of the test matrix.

The plot in Figure 3.3 indicates that the residuals of a total of 18 real Ritz values or pairs of complex conjugate Ritz values could be driven below the threshold of 10^{-9} within 40 steps of the iteration. Bear in mind, however, that the Krylov sequence has been extended by two vectors in those steps where complex shifts were used. Figure 3.4 shows the computed Ritz values in the first quadrant in comparison to the eigenvalues determined by `eig`. Because of the Hamiltonian spectral symmetry we also have corresponding Ritz values within the other three quadrants.

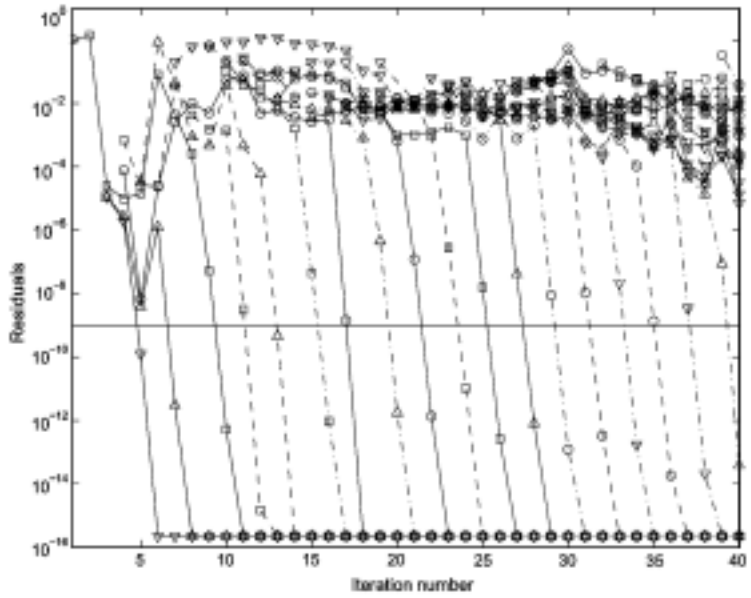

Fig. 3.3. Convergence history of the Rational SHIRA method for the test matrix.

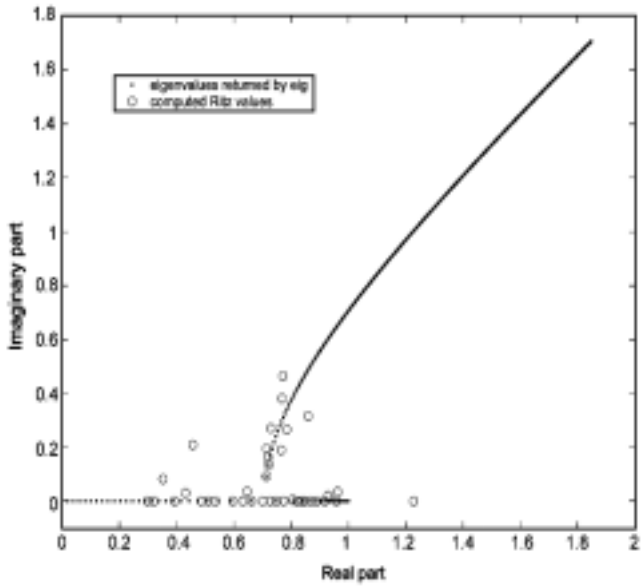

Fig. 3.4. Ritz values for the test problem computed by Rational SHIRA.

Even though the computed Krylov basis is all real, MATLAB's complex arithmetic has been taken advantage of when solving linear systems for operators with complex shifts.

4. Conclusions

We have presented a detailed derivation of an algorithm for solving Hamiltonian eigenproblems, which combines the SHIRA method of Mehrmann and Watkins with Ruhe's Rational Krylov algorithm in pursuit of equipping the former with the possibility of changing the shifts at runtime. We, therefore, call this new method Rational SHIRA. Preliminary numerical results confirm the efficiency of the method, although the need for more numerical experience persists at the time of writing this paper.

References

1. A. C. Antoulas, A new result on positive real interpolation and model reduction, *Systems and Control Letters*, **54** (2005), 361-374.
2. P. Benner, Computational Methods for Linear-Quadratic Optimization, *Supplemento ai Rendiconti del Circolo Matematico di Palermo, Serie II*, **58** (1999), 21-56.
3. P. Benner and H. Faβbender, An implicitly restarted symplectic Lanczos method for the Hamiltonian eigenvalue problem, *Linear Algebra Appl.*, **263** (1997), 75-111.
4. P. Benner, H. Faβbender and M. Stoll, Solving Large-Scale Quadratic Eigenvalue Problems with Hamiltonian Eigenstructure using a Structure-Preserving Krylov Subspace Method, *Electr. Trans. Num. Anal.*, **29** (2008), 212-229.
5. P. Benner and H. Faβbender, Numerische Methoden zur passivitätserhaltenden Modellreduktion, *at-Automatisierungstechnik*, **54(4)** (2006), 153-160.
6. P. Benner and H. Faβbender, *Passivity Preserving Model Reduction via a Structured Lanczos Method*, Proc. 2006 IEEE Conf. CACSD, Munich, Germany, Oct. 4-6, IEEE, 2006, pp. 8-13.
7. P. Benner, A. J. Laub and V. Mehrmann, A Collection of Benchmark Examples for the Numerical Solution of Algebraic Riccati Equations I: Continuous-Time Case, Fakultät für Mathematik, TU Chemnitz-Zwickau, **SPC 95_22**, 1995.
8. S. Boyd, V. Balakrishnan and P. Kabamba, A bisection method for computing the H_∞ norm of a transfer matrix and related problems, *Math. Control, Signals, Sys.*, **2** (1989), 207-219.
9. J. V. Burke, A. S. Lewis and M. L. Overton, Robust stability and a criss-cross algorithm for pseudospectra, *IMA J. Numer. Anal.*, **23** (2003), 359-375.
10. R. Byers, A bisection method for measuring the distance of a stable to unstable matrices, *SIAM J. Sci. Statist. Comput.*, **9** (1988), 875-881.
11. D. Chu, X. Liu and V. Mehrmann, A numerical method for computing the Hamiltonian Schur form, *Numer. Math.*, **105(3)** (2006), 375-412.
12. T. Eirola, *Krylov integrators for Hamiltonian systems*, Workshop on Exponential Integrators, Innsbruck, 20.-23.10.2004, 2004.

13. G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press third edition, Baltimore, 1996.

14. K. Ito and K. Kunisch, Reduced order control based on approximate inertial manifolds, *Linear Algebra Appl.*, **415(2-3)** (2006), 531-541.

15. B. Kagström, *A Direct Method for Reordering Eigenvalues in the Generalized Real Schur Form of a Regular Matrix Pair* (A, B), Linear Algebra for Large Scale and Real-Time Applications, M. S. Moonen, G. H. Golub and B. L. R. De Moor, eds., Kluwer Academic Publishers, Amsterdam, 1993, pp. 195-218.

16. L. Lopez and V. Simoncini, Preserving geometric properties of the exponential matrix by block Krylov subspace methods, *BIT*, **46** (2006), 813-830.

17. V. Mehrmann, *The Autonomous Linear Quadratic Control Problem, Theory and Numerical Solution*, Lecture Notes in Control and Information Sciences, No. 163, M. Thomas and A. Wyner, eds., Springer-Verlag, Heidelberg, 1991.

18. V. Mehrmann and D. S. Watkins, Structure-preserving methods for computing eigenpairs of large sparse skew-Hamiltonian/Hamiltonian pencils, *SIAM J. Sci. Statist. Comput.*, **22** (2001), 1905-1925.

19. J. Olson, H. J. A. Jensen and P. Jørgensen, Solution of large matrix equations which occur in response theory, *J. Comput. Phys.*, **74** (1988), 265-282.

20. C. Pester, Hamiltonian eigenvalue symmetry for quadratic operator eigenvalue problems, *J. Integral Equations Appl.*, **17(1)** (2005), 71-89.

21. A. Ruhe, The Rational Krylov algorithm for nonsymmetric Eigenvalue problems. III: Complex shifts for real matrices, *BIT*, **34** (1994), 165-176.

22. A. Ruhe, Rational Krylov algorithms for nonsymmetric Eigenvalue problems, Recent Advances in Iterative Methods, IMA Volumes in Mathematics and its Applications 60, G. Golub, A. Greenbaum and M. Luskin eds., Springer-Verlag, New York, 1994, pp. 149-164.

23. A. Ruhe, Rational Krylov algorithms for nonsymmetric eigenvalue problems, II: Matrix pairs, *Linear Algebra Appl.*, **197/198** (1994), 283-296.

24. V. Sima, *Algorithms for Linear-Quadratic Optimization*, Pure and Applied Mathematics, 200, Marcel Dekker, Inc., New York, NY., 1996.

25. D. C. Sorensen, Implicit application of polynomial filters in a k-step Arnoldi method, *SIAM J. Matrix Anal. Appl.*, **13(1)** (1992), 357-385.

26. D. C. Sorensen, Passivity Preserving Model Reduction via Interpolation of Spectral Zeros, *Systems & Control Letters*, **54** (2005), 347-360.

27. G. W. Stewart, A Krylov-Schur algorithm for large eigenproblems, *SIAM J. Matrix Anal. Appl.*, **23(3)** (2001), 601-614.

28. F. Tisseur and K. Meerbergen, The quadratic eigenvalue problem, *SIAM Review*, **43** (2001), 235-286.

29. N. Wong, V. Balakrishnan and C.-K. Koh, *Passivity-Preserving Model Reduction Via A Computationally Efficient Project-And-Balance Scheme*, Proc. Design Automation Conference, San Diego, CA (June), 2004, pp. 369-374.

30. K. Zhou, J. C. Doyle and K. Glover, *Robust and Optimal Control*, Prentice-Hall, Upper Saddle River, NJ, 1996.

Peter Benner
TU Chemnitz, Fakultät für Mathematik,
Mathematik in Industrie und Technik,
09107 Chemnitz,
Germany
E-mail: benner@mathematik.tu-chemnitz.de

Cedric Effenberger
Seminar für Angewandte Mathematik,
ETH Zürich, Rämistrasse 101,
Switzerland
E-mail: cedric.effenberger@sam.math.ethz.ch