



# Using Implicitly Filtered RKS for generalised eigenvalue problems

*Gorik De Samblanx      Adhemar Bultheel*

*Report TW 247, November 1996*

Department of Computer Science, K.U.Leuven

## **Abstract**

The Rational Krylov Sequence (RKS) method can be seen as a generalisation of Arnoldi's method. It projects a matrix pencil onto a smaller subspace; this projection results in a small upper Hessenberg pencil. As for the Arnoldi method, RKS can be restarted implicitly, using the QR decomposition of a Hessenberg matrix. This restart comes with a projection of the subspace using a rational function. In this paper, it is shown how the restart can be worked out in practice. In a second part, it is shown when the filtering of the subspace basis can fail and how this failure can be handled.

**AMS Subject Classification.** 65F15

**Keywords:** Rational Krylov method, implicitly restarted Arnoldi, generalised eigenvalue problem, shift-invert

# 1 Introduction

The Rational Krylov Sequence (RKS) algorithm [12, 15, 13, 14] is an algorithm that finds a limited set of eigenvalues of a (generalised) eigenvalue problem

$$Ax = \lambda Bx$$

in a certain region of the complex plane. Eigenvalues in such a region are called ‘wanted’ eigenvalues, they can be rightmost eigenvalues, the largest or the smallest eigenvalues,... RKS extends the idea of shift-invert Arnoldi [10, 16, 17] by computing a rational Krylov subspace

$$\{v_1, S_1 v_1, S_2 S_1 v_1, \dots\}, \text{ with } S_i = (A - \mu_i B)^{-1} B,$$

instead of a regular Krylov subspace with a fixed  $\mu$ . The subspace is spanned by an orthogonal basis  $V_k$ . The eigenvalue problem is then projected onto this subspace. From the projected problem a sequence of approximate eigenvalues and eigenvectors is computed that, under mild assumptions, converges to a certain eigenpair.

However, the size of the subspace can become too large. E.g., if more than one eigenvalue is sought, then it is not necessary to extend  $V_k$  each time with a large number of vectors. Eigenvalues are computed one at a time. When a new, different eigenvalue is wanted,  $V_k$  contains a lot of superfluous information for the computation of that eigenvalue. On the other hand, it is a waste of computational effort to restart the algorithm totally and to begin with a completely new subspace. For Arnoldi’s method, the Implicitly Restarted Arnoldi method [5, 4, 18] was proposed as a solution to this problem. Similarly, the RKS method can be restarted implicitly [3] or explicitly. This can be done without losing too much relevant information. Implicitly restarting the RKS method then corresponds to filtering the subspace with a rational filter

$$V_{k-p} \rightarrow \prod_{i=0}^{p-1} (A - \mu_{k-i} B)^{-1} (\alpha_i A - \beta_i B) V_{k-p},$$

where the  $\alpha_i, \beta_i$  may be chosen freely. Therefore, it was called Implicitly Filtered RKS (IFRKS). In [8], it was noticed that this filtering property can be used to filter away spurious eigenvalues. Indeed, if the matrix  $B$  is singular, then the problem has an infinite eigenvalue (possibly defective) that will be approximated by large finite eigenvalues. These spurious eigenvalues can mislead the algorithm, e.g. when the largest finite eigenvalues are needed. They can also have an important influence on the accuracy of the solution. The eigenvectors of the infinite eigenvalue lie in the null-space of the columns of  $B$ . Restarting the subspace with  $\alpha_i = 0$ , will then remove these spurious eigendirections from the basis  $V_k$ .

In this text, we show how the restarting algorithm that has been presented in [3], can be used safely to shrink the subspace  $V_k$ . We also show in which cases the implicit filtering of the subspace can fail. Therefore, this text must be seen as a continuation of [3]. It reflects somewhat the downside of the IFRKS method. However, we show how the problems can be avoided that come along with it.

The text is structured as follows. In §1, we explain briefly the RKS method. We show how RKS reduces the matrix pencil and how approximate eigenvalues can be computed from the resulting, small pencil. In §2, we show how an RKS relation can be restarted implicitly in a most general way. We propose two possible practical ways to compute the restart. In §3, we focus on a possible flaw in the method: the filtering procedure can fail. We give an example of this property and show how it can be detected. §4 closes the text with some conclusions.

## 1.1 The RKS algorithm

### Algorithm 1 RKS

0. Given  $v_1 \in \mathbf{C}^n$ ,  $\|v_1\| = 1$ .

Let  $V_1 = [v_1]$ .

1. For  $i = 1, \dots, k$  do

- 1.1. Select a pole  $\mu_i$  and a continuation vector  $t_i \neq 0 \in \mathbf{C}^i$
- 1.2. Form  $w = (A - \mu_i B)^{-1} B V_i t_i$ .
- 1.3. Orthogonalise  $w$  against the columns of  $V_i$  and let  $h_i = V_i^* w$ .
- 1.4. Normalise  $v_{i+1} = w/\eta_i$  with  $\eta_i = \|w\|$ .
- 1.5. Compute the approximate eigenpair  $(\theta_i, y_i)$ .

In this section, we define the RKS algorithm and we define the most important matrices that are involved in the text. The RKS algorithm is presented in Algorithm 1. In each step  $i$ , it computes a vector

$$w = (A - \mu_i B)^{-1} B V_i t_i,$$

which is orthogonalised and added to the basis  $V_i$

$$\begin{aligned} V_{i+1} &= [V_i \ v_{i+1}] \\ v_{i+1} &= (w - V_i h_i) / \eta_i \\ \text{with } h_i &= V_i^* w \text{ and } \eta_i = \|w - V_i h_i\|. \end{aligned}$$

The scalar  $\mu_i$  is called the *pole*,  $t_i \in \mathbf{C}^i$  is the *continuation vector*. It is easy to see that

$$A V_{i+1} \begin{bmatrix} h_i \\ \eta_i \end{bmatrix} = B V_{i+1} (\mu_i \begin{bmatrix} h_i \\ \eta_i \end{bmatrix} + \begin{bmatrix} t_i \\ 0 \end{bmatrix}).$$

Summarising this information of the first  $k$  steps, we get

$$A V_{k+1} H_k = B V_{k+1} K_k, \quad (1)$$

where  $K_k, H_k \in \mathbf{C}^{k+1 \times k}$  are unreduced upper Hessenberg matrices and  $V_{k+1}^* V_{k+1} = I$ . (An unreduced upper Hessenberg matrix is a matrix that has nonzero subdiagonal elements.) If we collect the poles in a diagonal matrix  $M_k = \text{diag}(\mu_i) \in \mathbf{C}^{k \times k}$  and the continuation vectors in an upper triangular matrix  $T_k \in \mathbf{C}^{k+1 \times k}$ , then

$$K_k = H_k M_k + T_k. \quad (2)$$

Note that any pair of unreduced Hessenberg matrices can be decomposed in a unique way as in (2). Therefore, any relation (1) with unreduced Hessenberg matrices corresponds uniquely to a RKS process with a certain starting vector, a set of poles and continuation vectors assuming that none of the poles is an eigenvalue of  $(A, B)$ . We will call (1) a *RKS relationship* when it fulfills that condition.

It should be noted that (1) can be shifted. The following lemma shows this in a slightly more generalised way.

**Lemma 1.1** *Suppose that  $A V_{k+1} H_k = B V_{k+1} K_k + S_k$ . Given a set of scalars  $\tau, \alpha, \beta$ , then this relation can be rewritten as*

$$(A - \tau B) V_{k+1} (\alpha K_k - \beta H_k) = (\alpha A - \beta B) V_{k+1} (K_k - \tau H_k) + (\tau \alpha - \beta) S_k.$$

**Proof** Obvious. □

**Notation 1.1** *Matrices are denoted by upper case roman characters. The index of a matrix is equal to its number of columns (this is also the iteration step in which it is constructed). The  $k \times k$  leading submatrices of the (rectangular)  $K_k, H_k, T_k$  are denoted by  $\underline{K}_k, \underline{H}_k, \underline{T}_k$ . The range of the columns of a matrix  $V$  is denoted by  $\mathcal{R}(V)$ . The  $i, j$ -th element of a matrix  $H$  is denoted by  $(H)_{i,j}$ . Lower case roman characters are vectors and scalars are denoted by Greek characters,  $\bar{\alpha}$  denotes the complex conjugate of a scalar.  $\cdot^*$  denotes the Hermitian transpose and  $\|\cdot\|$  denotes the 2-norm. The machine precision is denoted by  $\mathbf{u}$ .*

## 1.2 Computation of the approximate eigenvalues

The large matrix pencil  $(A, B)$  is projected on the smaller pencil  $(K_k, H_k)$ . However, the approximate eigenvalues can not be computed directly from these matrices, since they are rectangular. There are different ways to reduce  $(K_k, H_k)$  to a (generalised) eigenvalue problem.

**Definition 1.1** [9, 11] *Given a matrix  $A$  and a subspace basis  $V_k$ . Then  $(\theta, y = V_k z)$  is called a Ritz pair of  $A$  (in  $\mathcal{R}(V_k)$ ) with respect to  $\mathcal{R}(V_k)$  if*

$$A y - \theta y \perp V_k.$$

*$(\theta, y)$  is a Harmonic Ritz pair of  $A$  with respect to  $\mathcal{R}(V_k)$  if*

$$A y - \theta y \perp A V_k.$$

Applying this definition on the reduced eigenvalue problem of the RKS relation, gives the following lemma.

**Lemma 1.2** *Suppose that  $B$  is nonsingular. Then  $(\theta, y = V_{k+1}H_k z)$  is a Ritz pair of  $B^{-1}A$  with respect to  $\mathcal{R}(V_{k+1}H_k)$  if*

$$H_k^* K_k z - \theta H_k^* H_k z = 0.$$

*$(\theta, y = V_{k+1}H_k z)$  is a Harmonic Ritz pair of  $B^{-1}A$  with respect to  $\mathcal{R}(V_{k+1}H_k)$  if*

$$K_k^* K_k z - \theta K_k^* H_k z = 0.$$

*Suppose that  $\mu_k$  is not an eigenvalue of  $(A, B)$ . Then the solution  $(\theta, y = V_{k+1}H_k z)$  of*

$$\underline{K}_k z - \theta \underline{H}_k z = 0 \tag{3}$$

*is a Harmonic Ritz pair of  $B^{-1}(A - \mu_k B)$ .*

**Proof** The first part is proven in [6]. The second part follows from

$$\begin{aligned} B^{-1}AV_{k+1}H_k z - \theta V_{k+1}H_k z &\perp B^{-1}AV_{k+1}H_k \\ V_{k+1}K_k z - \theta V_{k+1}H_k z &\perp V_{k+1}K_k \\ K_k^* K_k z - \theta K_k^* H_k z &= 0. \end{aligned}$$

The last result can be found by shifting (1) with  $\mu_k$  and noticing that the  $(k+1, k)$ -th element of  $K_k - \mu_k H_k$  is zero. If  $\mu_k$  is not an eigenvalue of  $(A, B)$ , then  $K_k - \mu_k H_k$  has full rank (see [3]), so

$$(K_k - \mu_k H_k)^* K_k z - \theta (K_k - \mu_k H_k)^* H_k z = (\underline{K}_k - \mu_k \underline{H}_k)^* \underline{K}_k z - \theta (\underline{K}_k - \mu_k \underline{H}_k)^* \underline{H}_k z = 0,$$

so  $\underline{K}_k z - \theta \underline{H}_k z = 0$ . □

If the eigenvalues that are wanted lie in the interior of the spectrum of  $(A, B)$ , then using Harmonic Ritz values can turn out to be advantageous. However, the importance of the difference between ‘regular’ Ritz values and Harmonic Ritz values is not clear at this moment. In the remainder of this text, we will use the third option (3) to compute the approximate eigenpair, since this option seems to be a most natural choice for RKS.

We will measure the level of convergence of the algorithm by the residual norm of the approximate eigenvector

$$\|r_k\| = \|Ay_k - \theta_k By_k\| = \|AV_{k+1}H_k z_k - \theta_k BV_{k+1}H_k z_k\| = |\eta_k z_{k,k}| |\mu_k - \theta_k| \|Bv_{k+1}\|, \tag{4}$$

with  $z_{k,k}$  the  $k$ -th component of  $z_k$ . We assume that  $\|Bv_{k+1}\| = O(1)$ . The residual norm will be small if  $|\eta_k z_{k,k}|$  and/or  $|\mu_k - \theta_k|$  is small. If we do not change the pole  $\mu_k$  in every step, then  $|\mu_k - \theta_k|$  will be larger than  $|\eta_k z_{k,k}|$ . However, it is not necessary for  $|\eta_k|$  to be small when the method converges to some eigenvector. But if the method converges, then  $|\eta_k z_{k,k}|$ , which is a measure for the contribution of  $v_{k+1}$  in  $y_k$ , must clearly tend to zero.

## 2 Applying IFRKS

In this section, we show in general how the RKS subspace  $V_{k+1}$  can be filtered implicitly, using the matrices  $K_k$  and  $H_k$ . Then we derive two possible ways to compute this restart. In the next section, we will show that even when things are carefully computed, the filtering of  $V_k$  can fail. First, we recall how the RKS relation can be restarted. Theorem 2.1 is a generalised formulation of the result in [3].

**Theorem 2.1** *Consider a RKS relationship (1).*

(a) *Given an orthogonal matrix  $Q \in \mathbf{C}^{k+1 \times k}$ , a vector  $q \in \mathbf{C}^{k+1}$  that is orthogonal to  $Q$ ,  $q^* Q = 0$ . Then there exists a matrix  $Z \in \mathbf{C}^{k \times k-1}$ , with*

$$q^* K_k Z = 0 = q^* H_k Z. \tag{5}$$

*If we call  $V_k^+ = V_{k+1}Q$ ,  $K_{k-1}^+ = Q^* K_k Z$  and  $H_{k-1}^+ = Q^* H_k Z$ , then*

$$AV_k^+ H_{k-1}^+ = BV_k^+ K_{k-1}^+ \tag{6}$$

*is a new RKS relation. The corresponding shifts  $\mu_i^+$  are given by  $\mu_i^+ = \mu_{i+1}$ ,  $i = 1, \dots, k-1$ .*

(b) Given the set of Ritz vectors  $y_i = V_{k+1}H_k z_i$ ,  $i = 1, \dots, k$ . If

$$\begin{aligned}\mathcal{R}(Z) &= \mathcal{R}(z_1, \dots, z_{k-1}) \\ \mathcal{R}(Q) &= \mathcal{R}(H_k z_1, \dots, H_k z_{k-1}, u), \quad \text{where } u^* H_k = 0,\end{aligned}$$

then the new Ritz values are  $\theta_i^+ = \theta_i$  and  $z_i = Z z_i^+$ ,  $i = 1, \dots, k-1$ .

(c) If, given the scalars  $\alpha, \beta \in \mathbb{C}$ , with  $\alpha\mu_i \neq \beta$ ,  $Q$  is computed from the QR decomposition

$$QR = \alpha K_k - \beta H_k, \quad Q \in \mathbb{C}^{k+1 \times k} \text{ and } R \in \mathbb{C}^{k \times k}, \quad (7)$$

then  $\mathcal{R}(V_i^+) = \mathcal{R}((A - \mu_i B)^{-1}(\alpha A - \beta B)V_i)$ ,  $i = 1, \dots, k$ .

**Proof** The proof of (a) and (c) can be found in [3]. We only prove (b). If  $\theta_i$  is computed from  $M^* K_k z_i = \theta M^* H_k z_i$ , with  $M = \sigma_1 K_k - \sigma_2 H_k$ , then  $(M^* Q) Q^* K_k Z z_i^+ + 0 = \theta_i (M^* Q) Q^* H_k Z z_i^+ + 0$ . If we multiply this equation on the left by  $Z^*$ , we get  $(M^+)^* K_{k-1}^+ z_i^+ = \theta_i (M^+)^* H_{k-1}^+ z_i^+$ , with  $M^+ = \sigma_1 K_{k-1}^+ - \sigma_2 H_{k-1}^+$ .  $\square$

**Computation of  $Q$**  The computation of  $Q$  is a straightforward result of Theorem 2.1. Once we have set the parameters  $\alpha$  and  $\beta$ , then  $Q$  is the orthogonal, unreduced upper Hessenberg matrix in the (skinny) QR factorisation (7). The computation of  $Q$  is rather cheap, considered that  $Q$  has to be orthogonalised anyway – in order to get an orthogonal  $V_k^+$ . However, the QR decomposition is a process that is only backward stable, so forward errors may be introduced. In the next section, we will focus further on this problem.

**Computation of  $Z$**  In contrast to the computation of  $Q$ , there are many degrees of freedom left for the computation of  $Z$ . Indeed, any full rank upper triangular matrix  $U \in \mathbb{C}^{k-1 \times k-1}$  corresponds to a new matrix  $Z^U = ZU$  that fulfills the conditions of Theorem 2.1.  $U$  does not change the filtering properties of the restart, nor the new approximate eigenvalues, since it only replaces  $(K_{k-1}^+, H_{k-1}^+)$  by  $(K_{k-1}^+ U, H_{k-1}^+ U)$ . Therefore, we will look for a  $Z$  and a  $U$  that are cheap to compute and that make the restarting procedure as robust as possible. If we multiply (1) with  $U$  before the restart, then we get

$$K_k^U = K_k U = H_k U M_k + (T_k U + H_k M_k U - H_k U M_k) = H_k^U M_k + T_k^U,$$

with  $T_k^U = K_k U - H_k U M_k$ .

**Lemma 2.2** Given  $Q \in \mathbb{C}^{k+1 \times k}$  that fulfills (7) and say that  $\bar{Q} \in \mathbb{C}^{k \times k-1}$  is the upper left submatrix of  $Q$ . Suppose that  $\underline{T}_k$  has full rank. Then  $Z = (\alpha M_k - \beta I)(\underline{T}_k)^{-1} \bar{Q}$  fulfills condition (5) and it leads to matrices  $K_{k-1}^+, H_{k-1}^+$  that are unreduced upper Hessenberg.

**Proof** First note that, because  $Q$  is upper Hessenberg, if  $q^* Q = 0$  then  $q^* \begin{bmatrix} \bar{Q} \\ 0 \end{bmatrix} = q I_{k+1, k} Q = 0$ . If

$$QR = (\alpha K_k - \beta H_k) = H_k(\alpha M_k - \beta I) + \alpha T_k, \quad (8)$$

then  $q^*(\alpha K_k - \beta H_k)Z = 0$ . On the other hand

$$q^* H_k Z = q^* H_k(\alpha M_k - \beta I)(\underline{T}_k)^{-1} \bar{Q} = q^*(\alpha K_k - \beta H_k)(\underline{T}_k)^{-1} \bar{Q} - \alpha q^* T_k (\underline{T}_k)^{-1} \bar{Q} = 0 + 0.$$

Hence, (5) is fulfilled (if  $\alpha = 0$ , then  $K_k$  and  $H_k$  can be exchanged). By using (8), we get

$$\begin{aligned}H_{k-1}^+ &= Q^* H_k Z = R(\underline{T}_k)^{-1} \bar{Q} - \alpha I_{k, k-1} \\ K_{k-1}^+ &= Q^* K_k Z = R M_k (\underline{T}_k)^{-1} \bar{Q} + \beta I_{k, k-1},\end{aligned}$$

so these matrices are the product of a full rank upper triangular matrix and an unreduced upper Hessenberg matrix. Therefore,  $K_{k-1}^+$  and  $H_{k-1}^+$  must be unreduced upper Hessenberg.  $\square$

The choice of  $Z$  that is proposed in Lemma 2.2 is easy to apply. It is also not very expensive to implement. However, it heavily depends on the inversion of the matrix  $T_k$ . If  $T_k$  is nearly singular, then the restarting procedure can be unstable. This is often the case when the continuation vector is chosen  $t_k = z_k$ : since we hope that the method will converge, we can expect that  $z_k \simeq z_{k+1}$ . But even if  $T_k$  would be the unit matrix, then after a few steps of a repeated restart, the matrix can get a small singular value. However, there is a solution to this problem: we can multiply equation (1) by an upper triangular matrix  $U$  before restarting. If  $U$  can be chosen such that  $T_k^U$  is e.g. a diagonal matrix, then  $Z$  can be computed in a more stable way. The next proposition shows that such a  $U$  can always be found.



|                           | $Z$   | $Z^U$ | $Z^\perp$ |
|---------------------------|-------|-------|-----------|
| $\kappa(T)$               | 2e+14 | 3e+8  | 2e+14     |
| $\ q^* H_k Z\ $           | 1e+14 | 7e-9  | 1e-14     |
| $\kappa(U)$               |       | 2e+1  |           |
| $ \theta_1^+ - \theta_1 $ | 4e-6  | 8e-12 | 4e-12     |
| $ \theta_2^+ - \theta_2 $ | 3e-6  | 5e-12 | 4e-12     |
| $ \theta_3^+ - \theta_3 $ | 2e-6  | 5e-12 | 4e-12     |

Table 1: Comparison of different computations of  $Z$ . The first uses  $Z = (M - \theta_8 I) \underline{T}^{-1} \bar{Q}$ , the second transforms the problem such that  $\underline{T}$  is better conditioned, the latter computes the orthogonal  $Z$

We performed 8 steps of Algorithm 1 on this matrix, using  $v_1 = [1 \dots 1]/10$ . In the first 5 steps, we used as pole  $\mu_{1..5} = 1$  and as continuation vector the unit vector. In the last 3 steps, we used  $\mu_i = \theta_{i-1}$  and for  $t_i$  the corresponding approximate eigenvector. After these 8 steps, the error on the rightmost eigenvalue  $\lambda_1 = -0.01$  is  $|\theta_1 - \lambda_1| = 3e - 3$ .

We then used three different algorithms to restart the RKS relationship in order to remove the leftmost approximate eigenvalue of  $(\underline{K}_8, \underline{H}_8)$ . The results are displayed in Table 2.1. The first column shows the results for  $Z = (M_8 - \theta_8 I) \underline{T}_8^{-1} \bar{Q}$ . It illustrates that this straightforward choice of  $Z$  introduces not only a large error on  $\theta_i^+$ . More importantly,  $\|q^* H_k Z\| \gg 0$ , which means that a large error is introduced on the new RKS relation. This is important, because if the RKS relation is not correct, then the future approximated eigenvalues might become inaccurate. The error is caused by the badly conditioned  $\underline{T}$  matrix.

The second column shows the application of Algorithm 2 (with  $\tau = 10e-4$  and  $\hat{u}_i = [1 \dots 1]^*$ ) on this example. A good  $U$  is found that reduces the norm of  $\underline{T}^{-1}$  to some extent. However,  $\|\underline{T}^{-1}\|$  is still large. The results for  $\theta_i^+$  and  $\|q^* H_k Z\|$  are better.

The third approach seems to give the best results. The errors on the  $\theta_i^+$  are very small and  $\|q^* H_k Z\| = O(\mathbf{u})$ .

### 3 Possible errors while using IFRKS

In practice, the computed RKS relation (1) is not exact. In each step of Algorithm 1, a computational error is added to the relation. If the error is small, then it will have no important effect on the convergence properties of the algorithm. However, in [8], it has been shown for the Implicitly Restarted Arnoldi method, that this error can have an important effect on the filtering properties of the restarting algorithm. In this section, we show a similar property for the IFRKS algorithm. We show that, even if the error on the RKS relation is small, the filtering step described in Theorem 2.1(c) can be inaccurate.

There are two main sources of possible computational errors in Algorithm 1. First, the linear system in step 1.2. will only be solved to some residual error :

$$(A - \mu_i B)w = B V_i t_i + s'_i, \text{ with } \|s'_i\| \leq \mathbf{u}(\|A - \mu_i B\| \|w\| + \|B V_i t_i\|),$$

where  $\mathbf{u}$  is the machine precision. Secondly, the orthogonalisation steps 1.3. and 1.4. can be inexact :

$$w = V_i h_i + v_{i+1} \eta_i + s''_i.$$

If we call  $s'''_i$  the rounding error made on the computation of the  $i$ -th column of  $K_i$  and if we sum the errors in the vector  $s_i = s'_i - (A - \mu_i B)s''_i + B V_{i+1} s'''_i$ , then we can collect all the error vectors in one matrix  $[s_1, \dots, s_k] = S_k$ . We get the corrected RKS relationship

$$A V_{k+1} H_k = B V_{k+1} K_k + S_k. \quad (9)$$

In Theorem 3.2, we will show that if the filtering step fails, then the influence of  $S_k$  will be larger in general than the influence of the error on the computed QR decomposition. In order to prove that, we need the following Lemma.

**Lemma 3.1** *If we call  $\sigma_1(T) \geq \sigma_2(T) \geq \dots$  the singular values of a matrix  $T$  and if  $R \in \mathbf{C}^{k-1 \times k-1}$  has full rank, then given a vector  $r \in \mathbf{C}^{k-1}$  and a scalar  $\rho \in \mathbf{C}$*

$$\sigma_i(\tilde{R}) = \sigma_i \left( \begin{bmatrix} R & r \\ 0 & \rho \end{bmatrix} \right) \leq \kappa \sigma_i(R) + |\rho|, \quad i = 1, \dots, k-1,$$

where  $\kappa \leq 1 + \|R^{-1}r\|$

**Proof** See [2]. □

As a result, Lemma 3.1 says that if  $|\rho| \ll \sigma_{k-1}(\tilde{R})$  and if  $\kappa$  is not too large, then  $\sigma_{k-1}(R) = O(\sigma_{k-1}(\tilde{R}))$ . I.e., the smallest singular value of the leading submatrix  $R$  is approximately equal to the second smallest singular value of  $\tilde{R}$ . In the following theorem, we show how the error  $S_k$  can cancel the filtering property.

**Theorem 3.2** *Given an inexact RKS relation (9), with  $\|S_k\| \leq \varepsilon$ ; given the QR decomposition  $\hat{Q}R = \alpha K_k - \beta H_k$ , with  $Q$  the computed approximation of  $\hat{Q} = Q + \Delta Q$  such that  $\|QR - (\alpha K_k - \beta H_k)\|$  is small. Then there exist a matrix  $P$  of full rank such that*

$$V_k^+ = (A - \mu_k B)^{-1}(\alpha A - \beta B)V_k P + E,$$

with  $\|E\| \leq \varepsilon|\alpha\mu_k - \beta|\sigma_k^{-1}(R) + \|\Delta Q\|$ . Moreover, if  $(R)_{k,k} < \sigma_{k-1}(R)$ , then  $\|\Delta Q\|_F \leq \Omega\|R\|\sigma_{k-1}^{-1}(R)\mathbf{u}$ , for some  $\Omega \geq 0$ .

**Proof** Given (9) and Lemma 1.1, we can write

$$\begin{aligned} (A - \mu_k B)V_{k+1}(\alpha K_k - \beta H_k) &= (\alpha A - \beta B)V_{k+1}(K_k - \mu_k H_k) + (\alpha\mu_k - \beta)S_k \\ &= (\alpha A - \beta B)V_k(\underline{K}_k - \mu_k \underline{H}_k) + (\alpha\mu_k - \beta)S_k \\ (A - \mu_k B)V_{k+1}QR &= (\alpha A - \beta B)V_k(\underline{K}_k - \mu_k \underline{H}_k) + (\alpha\mu_k - \beta)S_k \\ V_{k+1}Q &= (A - \mu_k B)^{-1}(\alpha A - \beta B)V_k[(\underline{K}_k - \mu_k \underline{H}_k)R^{-1}] + (\alpha\mu_k - \beta)S_k R^{-1}, \end{aligned}$$

with  $\|(\alpha\mu_k - \beta)S_k R^{-1}\| \leq \varepsilon|\alpha\mu_k - \beta|\sigma_k^{-1}(R)$ . In [1] is proved that  $\|\Delta Q\|_F \leq \sqrt{2}\|R\|\|R_{k-1}^{-1}\|\mathbf{u} + O(\mathbf{u}^2)$ , where  $R_{k-1}$  is the  $k-1 \times k-1$  leading submatrix of  $R$ . If we neglect the second order term in  $\mathbf{u}$  and if we combine this result with Lemma 3.1, we get

$$\|\Delta Q\|_F \leq \sqrt{2}\|R\|(\kappa + (R)_{k,k}/\sigma_{k-1}(R))\sigma_{k-1}^{-1}(R) \leq \sqrt{2}\|R\|(\kappa + 1)\sigma_{k-1}^{-1}(R),$$

with  $\kappa$  as in Lemma 3.1 and  $(R)_{k,k}/\sigma_{k-1}(R) < 1 \leq \kappa$ . □

Theorem 3.2 says that when a RKS relation is restarted and it has a  $R$  that is almost singular, then the filtering property can be lost. However, the influence of the (possibly small) error matrix  $S_k$  will be dominant to the inaccuracy of the computed matrix  $Q$ , assuming that  $0 \simeq \sigma_k(R) \ll \sigma_{k-1}(R)$ .

Let us consider a case where  $\|R^{-1}\|$  can be large. This means that  $R$  is nearly singular, so there exists a vector  $z$  of unit length, such that

$$\sigma_k(R) = \|Rz\| = \|\alpha K_k z - \beta H_k z\| \simeq 0.$$

If  $\beta/\alpha = \theta_k$ , then with  $z$  equal to the corresponding eigenvector of  $(\underline{K}_k, \underline{H}_k)$ , we get

$$\sigma_k(R) \leq \|Rz\| = (\alpha\mu_k - \beta)|\eta_k z_{k,k}|.$$

A small  $|\eta_k z_{k,k}|$ , corresponds, following (4), to the residual of an eigenvector that has converged well. Therefore, using IFRKS to remove a converged eigenvector from the subspace  $V_{k+1}$  seems not to be a good idea, because the filtering will be inaccurate. A second filtering step can then be necessary.

It is clear that unlike the problems with the computation of  $Z$ , this problem can not be cured by switching to a different set of  $K_k$  and  $H_k$  matrices. In that case, a more explicit procedure should be employed (see e.g. [2]). It should be stressed that, however the filtering can fail, the norm of the ‘new’ error matrix  $S_k^+$  will be of the same order of  $\|S_k\|$  if the condition (5) is fulfilled to computational accuracy.

Before we illustrate this with an example, we must keep in mind the following warning. There is a second possibility to have a large  $\|R^{-1}\|$ . If  $H_k$  is singular, then  $K_k$  must be (in general) singular too, since both sides of (1) must have the same rank and share the same null space. Any combination of  $K_k$  and  $H_k$  will then have a singular  $R$ . Moreover, the small eigenvalue problem will have an arbitrary eigenvalue. We can aspect a similar behaviour if  $H_k$  and  $K_k$  have a small singular value and  $R$  becomes near-singular. Therefore, this situation must be avoided by the RKS algorithm, since it will lead to wrong results that, unfortunately, are hard to identify.

|                            | $k = 10$   | Filter  | Filter          | $k = 4$         | Filter          | Remove               | Remove                |
|----------------------------|------------|---------|-----------------|-----------------|-----------------|----------------------|-----------------------|
| $\mu$                      | 1          |         |                 | -9.49           |                 |                      |                       |
| $(\alpha, \beta)$          |            | (0, 1)  | (0, 1)          |                 | (0, 1)          | $(1, \theta_{left})$ | $(1, \theta_{right})$ |
| $\theta_1$                 | 50.5       | 1369    | -9.49           | -9.4883         | -9.4883         | -9.4883              | $-10 + 0.3i$          |
| $\theta_2$                 | $13 - 33i$ | -9.49   | $-11.4 - 14.6i$ | $-11.4 - 14.6i$ | $-11.4 - 14.6i$ | $-11.4 - 14.6i$      | $-11.6 - 14.7i$       |
| $\theta_{best}$            | -9.38      | -9.49   | -9.49           | -9.4883         | -9.4883         | -9.4883              |                       |
| $\ Ay_1 - \theta_1 By_1\ $ | $1e-6$     | $8e-4$  | $3e-5$          | $1e-10$         | $1e-10$         | $4e-10$              | $4e-10$               |
| $\ AVH - BVK\ /\ H\ $      | $1e-14$    | $2e-12$ | $2e-12$         | $3e-14$         | $3e-14$         | $3e-14$              | $1e-13$               |
| $\sigma_k(R)$              |            | $1e-4$  | $6e-4$          |                 | $9e-10$         | $8e-8$               | $6e-19$               |
| $\sigma_{k-1}(R)$          |            | $9e-4$  | $1e-3$          |                 | $2e-6$          | $1e-4$               | $1e-4$                |
| $\#\theta_{sp}$            | 3          | 1       | 0               | 0               | 0               | 0                    | 0                     |

Table 2: Illustrations of aspects of filtering with IFRKS

**Example 3.1** Let us recall the example in [3].

The example comes from a model of viscous free-surface fluid flow on a tilted plane [7]. The Navier-Stokes equations were discretised by a finite element approach leading to an eigenvalue problem  $Ax = \lambda Bx$  of size  $n = 536$ . The matrices  $A$  and  $B$  are nonsymmetric,  $B$  is singular ( $B$  has rank 429) and  $A$  is not. The goal is to find the rightmost eigenvalues used for the stability analysis of a steady state solution of the Navier-Stokes equations.

The main problem is that  $B$  is singular. This means that  $Ax = \lambda Bx$  has an infinite eigenvalue. Moreover, for this example, this eigenvalue is defective. When we apply an iterative eigenvalue solver to this problem, this infinite eigenvalue may emerge as finite spurious eigenvalues. The calculation of spurious eigenvalues can be avoided as follows. Let  $\mathcal{R}$  denote the eigenspaces (and the generalised eigenspaces if the eigenvalues are defective) corresponding to the finite eigenvalues and  $\mathcal{N}$  the eigenspace (and generalised eigenspace) corresponding to the infinite eigenvalue. Since  $\mathbf{R}^n = \mathcal{R} + \mathcal{N}$ , it is sufficient to remove the  $\mathcal{N}$  component in  $V$ , in order to avoid the calculation of ‘infinity’. Since an infinite eigenvalue of  $Ax = \lambda Bx$  corresponds to a zero eigenvalue of  $A^{-1}B$ ,  $\mathcal{N} = \text{nullspace}((A^{-1}B)^\nu)$  where  $\nu$  is the index of this eigenvalue. If  $\nu = 1$ , then the eigenvalue is non-defective. Thus, the calculation of spurious eigenvalues due to singular  $B$  can be avoided by filtering

$$V^+ \leftarrow \text{orth}((A^{-1}B)^\nu V)$$

or even by shifting the problem, e.g. as

$$V^+ \leftarrow \text{orth}((A - \mu_\nu B)^{-1} B \cdots (A - \mu_1 B)^{-1} B V).$$

Such a filtering is easily done by the IFRKS method with shifts  $(\alpha, \beta) = (0, 1)$ .

For this example, the rightmost eigenvalues are known, so it is easy to check whether a computed rightmost eigenvalue is spurious. These eigenvalues are  $\lambda_1 = -9.4883$ ,  $\lambda_{2,3} = -11.6062 \pm 14.6602i$ , and  $\lambda_{4,5} = -15.9689 \pm 3.2343i$ . The index of the infinite eigenvalue is not known, but we found that a small number of QZ steps were sufficient to avoid the calculation of spurious eigenvalues in this problem. We ran the following test in Matlab on a Ultra Sparc-2 station.

First, we performed 10 steps of Algorithm 1 with pole  $\mu_{1..10} = 1$  and  $v_1 = [1 \cdots 1]/\sqrt{n}$ . Then, we filtered 2 times the subspace  $V_k$  with  $(\alpha, \beta) = (0, 1)$  in order to remove the spurious eigenvalues from the approximated spectrum. We resumed Algorithm 1 with pole  $\mu_i = \theta_1 = -9.49$  for 4 iterations. We then filtered 1 time with  $(\alpha, \beta) = (0, 1)$ . Finally, we restarted the subspace twice with an exact shift, i.e.  $(\alpha, \beta) = (1, \theta)$ . The first time, we removed the leftmost Ritz value. Since this Ritz value lies far from the second shift, it has not converged. Finally, we tried to remove the converged Ritz value from the approximate spectrum. Table 3.1 shows the results.

The example illustrates several points. First, it shows that it is important to filter away possible spurious eigenvalues before changing the pole  $\mu$ . Indeed, the first column of Table 3.1 shows that if we set the new  $\mu$  equal to the rightmost eigenvalue,  $\theta_1 = 50.5$ , then this will slow down the convergence. This pole lies far from the true rightmost eigenvalue. The second and third column show that it can be necessary to filter more than once, if the zero eigenvalue of  $B$  is defective. Indeed, it takes two filtering steps to remove all spurious eigenvalues ( $\#\theta_{sp}$  displays the number of Ritz values that have a positive real part. These Ritz values are certainly spurious). If we resume the RKS algorithm with a new pole, then the rightmost eigenvalue converges well. The filtering step in column five turns out to be not necessary, since there are

no spurious eigenvalues (with positive real parts) left. We filtered once anyway, to be sure. The last two columns illustrate that it is safe to remove Ritz values that have not converged yet. However, it is not safe to remove the converged, rightmost eigenvalue. The last column of Table 3.1 shows how such a filtering step fails: the Ritz eigenvalue  $-9.4883$  is not removed, but a bad approximation remains present in the subspace. This failure is caused by the fact that  $R$  is almost singular ( $\sigma_k(R) = 6e-19$ ).

## 4 Conclusions

In this text, we showed how the RKS relation can be restarted in two different ways. The first option was to transform the matrices  $(K_k, H_k)$  with an upper triangular matrix  $U$  to the set  $(K_k^U, H_k^U)$ , such that the corresponding  $T_k^U$  is well conditioned. The second option was to compute  $Z$  directly such that it is orthogonal. The latter option seems to be the most effective. It also has the advantage that it is easy to implement. In order to introduce no large errors in the RKS relation, it is important for both methods to make sure that the  $\|q^* K_k Z\|$  and  $\|q^* H_k Z\|$  in (5) are as small as possible.

Restarting the RKS relation can be combined with an implicit filtering step of the subspace that is spanned by  $V_k$ . However, similarly to the IRA method, this filtering can fail if the matrix  $QR = \alpha K_k - \beta H_k$  is almost singular. We showed that if this near-singularity is caused by a well converged eigenvalue, then the influence of the computed  $Q$  on the filtering step is smaller than the influence of the near singularity.

## 5 Acknowledgements

This research was supported by the National Fund for Scientific Research (NFWO), project Lanczos, grant #2.0042.93 and by the Human Capital and Mobility project ROLLS of the European Community under contract ERBCHRXCT930416.

## References

- [1] X-W Chang, C. Paige, and G.W. Stewart. Perturbation analyses for the QR factorization.
- [2] G. De Samblanx and A. Bultheel. On the convergence an the restarting of inexact eigenvalue solvers. TW-REPORT, Dept. Computing Science, K.U.Leuven, 1996.
- [3] G. De Samblanx, K. Meerbergen, and A. Bultheel. The implicit application of a rational filter in the RKS method. Report TW239, Department of Computing Science, K.U.Leuven, 1996.
- [4] R.B. Lehoucq. Restarting an arnoldi reduction. Preprint MCS-P591-0496, Argonne National Laboratory, Argonne, IL 60439, 1996. Submitted to SISC.
- [5] R.B. Lehoucq, D.C. Sorensen, and P. Vu. ARPACK: *An implementation of the Implicitly Re-started Arnoldi Iteration that computes some of the eigenvalues and eigenvectors of a large sparse matrix.* Available from netlib@ornl.gov under the directory scalapack, 1995.
- [6] K. Meerbergen and R. Lehoucq. The inexact Rational Krylov Sequence method. Tech. Report 978, Dept. of Math., University of Utrecht, 1996.
- [7] K. Meerbergen and D. Roose. Matrix transformations for computing rightmost eigenvalues of real nonsymmetric matrices. *IMA J. Numer. Anal.*, 16:297–346, 1996.
- [8] K. Meerbergen and A. Spence. Implicitly restarted Arnoldi and purification for the shift-invert transformation. *Math. Comp.*, 1996. Accepted for publication.
- [9] R.B. Morgan. Computing Interior Eigenvalues of Large Matrices. *Linear Alg. Appl.*, 154–156:289–309, 1991.
- [10] R. Natarajan. An Arnoldi-based iterative scheme for nonsymmetric matrix pencils arising in finite element stability problems. *J. Comput. Phys.*, 100:128–142, 1992.
- [11] C. Paige, B.N. Parlett, and H.A. Van der Vorst. Approximate solutions and eigenvalue bounds from Krylov subspaces. *Num. Lin. Alg. Appl.*, 2:115–133, 1995.

- [12] A. Ruhe. Rational Krylov sequence methods for eigenvalue computation. *Linear Alg. Appl.*, 58:391–405, 1984.
- [13] A. Ruhe. Rational Krylov algorithms for nonsymmetric eigenvalue problems. In G. Golub, M. Luskin, and A. Greenbaum, editors, *IMA volumes in Mathematics and its Applications*, volume 60, pages 149–164. Springer, 1993.
- [14] A. Ruhe. The Rational Krylov algorithm for nonsymmetric eigenvalue problems, III: Complex shifts for real matrices. *BIT*, 34:165–176, 1994.
- [15] A. Ruhe. Rational Krylov algorithms for nonsymmetric Eigenvalue problems, II: Matrix pairs. *Linear Alg. Appl.*, 197/198:283–296, 1994.
- [16] Y. Saad. Numerical solution of large nonsymmetric eigenvalue problems. *Comput. Phys. Comm.*, 53:71–90, 1989.
- [17] Y. Saad. *Numerical methods for large eigenvalue problems*. Algorithms and Architectures for Advanced Scientific Computing. Manchester University Press, Manchester, 1992.
- [18] D.C. Sorensen. Implicitly restarted Arnoldi/Lanczos methods for large scale eigenvalue calculations. Dept. Computational and Applied Mathematics, Rice University, Houston, T X 77251, USA, 1995.