STABLE AND RELIABLE COMPUTATION OF EIGENVECTORS OF LARGE PROFILE MATRICES

Martin Ruess [1]

ABSTRACT

Independent eigenvector computation for a given set of eigenvalues of typical engineering eigenvalue problems still is a big challenge for established subspace solution methods. The inverse vector iteration as the standard solution method often is not capable of reliably computing the eigenvectors of a cluster of bad separated eigenvalues.

The following contribution presents a stable and reliable solution method for independent and selective eigenvector computation of large symmetric profile matrices. The method is an extension of the well-known and well-understood QR-method for full matrices thus having all its good numerical properties. The effects of finite arithmetic precision of computer representations of eigenvalue/eigenvector solution methods are analysed and it is shown that the numerical behavior of the new method is superior to subspace solution methods.

KEY WORDS

Orthogonal Eigenvectors, Real Symmetric Matrices, Profile Structure, Clustered Eigenvalues

INTRODUCTION

The governing equation of the eigenvector computation for a given eigenvalue λ_i is the homogeneous form (1) of the special eigenvalue problem. The non-trivial solution of (1) yields the desired eigenvector $\mathbf{x}_i$.

$$\mathbf{C}\,\mathbf{x}_i \;=\; \mathbf{0} \qquad \text{with } \mathbf{C} \;:=\; \mathbf{A} - \lambda_i\,\mathbf{I} \tag{1}$$

For small matrices $(N < 50)$ of the real symmetric case the QR-method is known as the most efficient and most elegant numerical solution technique for the complete determination of eigenvalues and eigenvectors. Its numerical properties are superior compared to the very popular subspace iteration schemes that are still the method of choice for very large eigenvalue problems $(N < 500)$. These methods essentially suffer from numerical instabilities that lead to the loss of linear independence of the calculated eigenvectors. Much effort has to be done to overcome these numerical difficulties and to provide correct, accurate and orthogonal base vectors of the approximated subspace.

The basic numerical drawbacks of subspace methods are shown in this contribution. A

[1] Research Engineer, Department of Civil Engineering, Technical University of Berlin, Sekr. TIB1-B8, Gustav-Meyer-Allee 25, 13355 Berlin, Germany, Phone +49 30 31472305, martin@ifb.bv.tu-berlin.de

new and more stable solution approach based on the QR-method and therefore exploiting its good numerical properties is presented. The analysis of several examples illustrates the high quality of the computational results of the new method.

SOLUTION STRATEGY

The properties of the stable QR-method are exploited. The method is applied to large symmetric matrices with convex profile structure. It is the preservation of the profile structure during the computation process that allows an efficient use of this method. The positive convergence properties of the QR-method for eigenvalues $\lambda = 0$ is used for an independent eigenvector computation of any set of given eigenvalue approximations. Cluster of bad seperated eigenvalues of size m are approximated in subspaces of that size thus avoiding any numerically expensive orthogonalisation process for the approximate vectors and ensuring convergence to the desired subspace. For a large number of eigenvector computations the numerical effort is significantly reduced with a local interation scheme.

QR-DECOMPOSITION OF PROFILE MATRICES

The left profile of a matrix $\mathbf{C}$ indicates for each row k the index of the first nonzero entry and is denoted with $\mathbf{p}_L$. The index of the last nonzero entry per row is called right profile and is stored in a profile vector $\mathbf{p}_R$. The profile of a matrix is said to be convex if (2) is true :

$$m \geq i \;\; \Rightarrow \;\; p_L[m] \;\; \geq \;\; p_L[i] \;\; \wedge \;\; p_R[m] \;\; \geq \;\; p_R[i] \tag{2}$$

The convex profile structure of $\mathbf{C}$ may substantially be preserved in the QR-decomposition (see Fig.1). The reduction of $\mathbf{C}$ to triangular form is carried out by premultiplying plane rotation matrices $\mathbf{P}_{ik}$ from left (3). Each matrix $\mathbf{P}_{ik}$ eliminates a coefficient $c_{ik}(k < i)$ by the choice of the rotation angle ϕ_{ik}. The sequence of the elimination process is carried out columnwise, starting with coefficient c_{21}, thus ensuring the preservation of the convex profile. The elimination of c_{ik} only affects rows i and k. The complete reduction of column k requires a temporary extension of the right profile $p_R[i]$ to $p_R[p_R[k]]$. This circumstance troubles little since it affects only a very small number of coefficients and only for a limited part of the calculation. After the reduction of column k the additional coefficients no longer affect the decomposition process and therefore are set free. More detailled information can be found in *Matrix Iteration For Large Symmetric Eigenvalue Problems* (Ruess 2002).

$$\mathbf{R} \;\; = \;\; \mathbf{P}_{n,n-1}^T \ldots \mathbf{P}_{ik}^T \ldots \mathbf{P}_{32}^T \, \mathbf{P}_{n1}^T \ldots \mathbf{P}_{31}^T \, \mathbf{P}_{21}^T \, \mathbf{A} \;\; = \;\; \mathbf{Q}^T \mathbf{A} \tag{3}$$

CONVERGENCE PROPERTIES OF THE QR-METHOD

Convergence Behavior in Exact Arithmetics

In exact arithmetics the shifted matrix $\mathbf{C}$ is semidefinite thus having a q-fold eigenvalue $\lambda = 0$. Hence the q eigenstates $(0, \mathbf{x}_i)$ are determined with a single QR-decomposition :

The semidefinite matrix $\mathbf{C}$ is singular and so is the QR-decomposition of $\mathbf{C}(= \mathbf{QR})$. The orthonormal matrix $\mathbf{Q}$ is nonsingular since it is of full column rank. Thus the triangular matrix $\mathbf{R}$ must be the singular factor of the decomposition product. Zero entries r_{ii} that occur on the diagonal of $\bar{\mathbf{R}}(= \ldots \mathbf{P}_{ik} \ldots \mathbf{P}_{21}\, \mathbf{C})$ during the decomposition process are automatically swaped down to row k with $(k > i)$ by the choice of the rotation angle ϕ_{ik} for the elimination of coefficient c_{ik}. The last q rows of the triangular matrix $\mathbf{R}$ therefore are zero (Fig. 1).

Since the last q rows of $\mathbf{R}$ equal zero the product of the recombination $(\mathbf{R}\,\mathbf{Q} = \mathbf{Q}^T\mathbf{C}\,\mathbf{Q})$ contains q rows and colums equal zero. The last q columns of $\mathbf{Q}$ contain the eigenvectors $\mathbf{x}_1, \ldots, \mathbf{x}_q$ of the q-fold eigenvalue $\lambda^{(q)}$.

Multiplying the rotation matrices $\mathbf{P}_{ik}$ from left against $\mathbf{I}_q$ in reverse order extracts the desired eigenvectors from $\mathbf{Q}$ in a locally limited calculation process.

$$\mathbf{X}_q \;=\; \mathbf{P}_{21} \ldots \mathbf{P}_{p_R[1],1}\, \mathbf{P}_{32} \ldots \mathbf{P}_{ik} \ldots \mathbf{P}_{n,n-1}\, \mathbf{I}_q \tag{4}$$

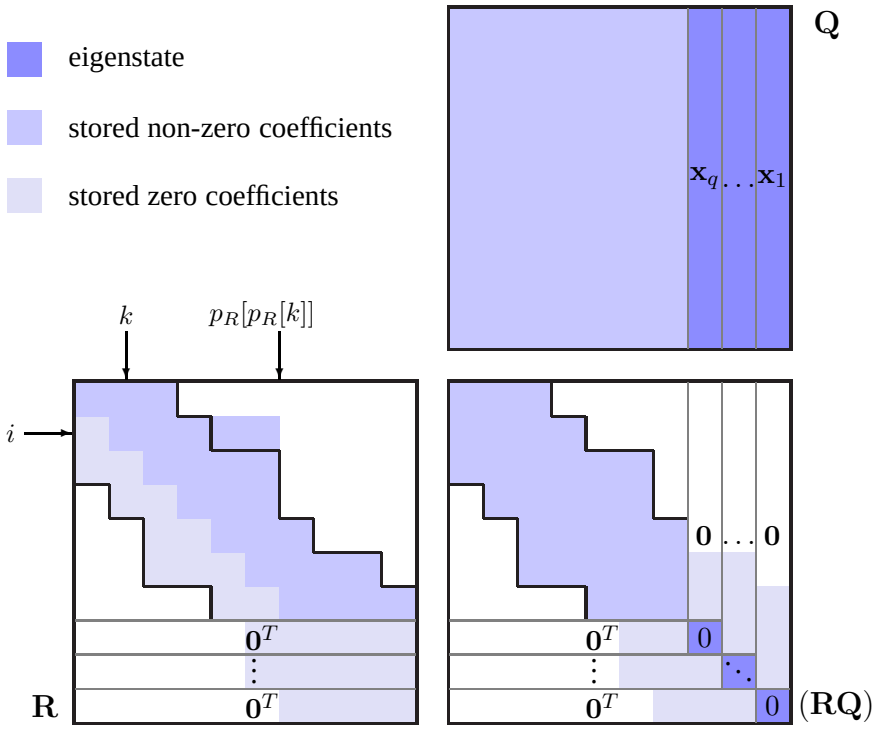

Figure 1: q eigenstates $(0, \mathbf{x}_i)$ after a single transformation $\mathbf{Q}^T\mathbf{C}\,\mathbf{Q} = \mathbf{R}\,\mathbf{Q}$

Convergence Behavior in Finite Precision Arithmetics

In finite precision arithmetics we are faced with roundoff errors that may dramatically influence the theoretically straightforward solution process for the eigenvector computation. Though the QR-decomposition process is known as numerically backward stable that is the method computes the exact eigenvalues and eigenvectors of a perturbed matrix $\check{\mathbf{C}}$, the strategy for

selective and independent eigenvector computation may fail without further numerical effort in a few cases where eigenvalues are very close.

$$\check{\mathbf{C}} \;=\; \mathbf{C}+\mathbf{E} \qquad\qquad \text{with } \|\mathbf{E}\|_2 \;\leq\; m\,\epsilon\,\|\mathbf{C}\|_2 \tag{5}$$

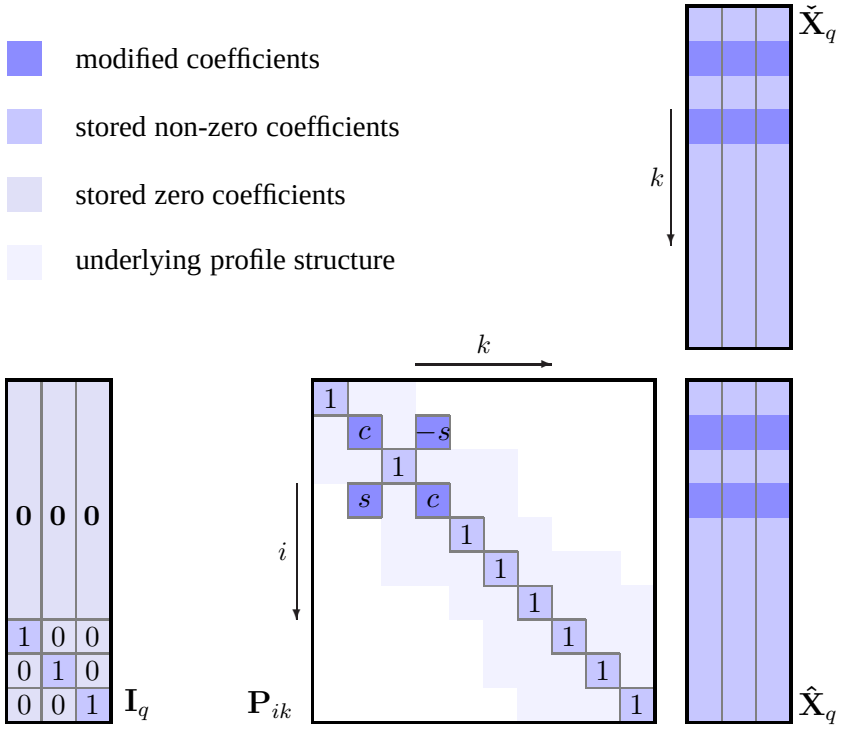

Figure 2: Computation of q eigenvectors $\mathbf{x}_i$ of a q-fold eigenvalue $\lambda_i^{(q)}$

Two effects of finite precision arithmetic mainly influence the eigenvector computation :

1. From (5) follows that instead of an exact eigenvalue λ we have to deal with an eigenvalue approximation μ of a perturbed Matrix $\check{\mathbf{C}}$ as shift for the desired eigenvector. The maximum error for μ is of order (6) (see Parlett 1997).

$$|\mu-\lambda| \;\leq\; m\,\epsilon\,\|\mathbf{C}\|_2 \tag{6}$$

The error in μ particularly affects the computation if eigenvalues in the neigborhood of μ form a bad separated cluster. Typically the pairing-up problem of eigenvalues may lead to wrong eigenvector approximations and miss the desired one. The corresponding eigenvector approximation for μ is denoted with $\hat{\mathbf{x}}$.

Pairing-up Problem : Figure 3 illustrates a typical distribution of the exact eigenvalues λ_i and their numerical approximation μ_i. The eigenvalue approximations μ_3 and μ_5 clearly may not be identified as approximations for λ_3 and λ_5, since they are closer to

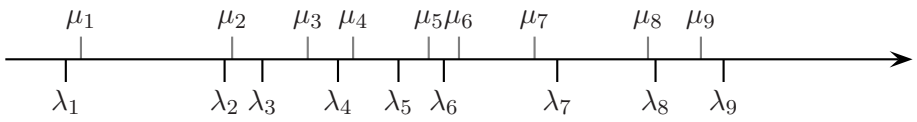

Figure 3: Pairing-up (approximation μ / exact value λ)

neighboring eigenvalues than to their assigned exact values. Successive computing of the eigenvector approximations $\hat{\mathbf{x}}_3$ and $\hat{\mathbf{x}}_5$ for these insufficient paired up values result in identical eigenvector approximations for μ_2, μ_3 and μ_4, μ_5, respectively.

2. The QR-decomposition of $\mathbf{C}$ comes with an error matrix $\mathbf{M}$ that is caused by the determination of the rotation parameters $cos(\phi_{ik})$ and $sin(\phi_{ik})$ ($cos(\phi_{ik})^2 + sin(\phi_{ik})^2 \neq 1$) and the operations of the reduction process.

Example 1 : The pairing-up problem and further consequences of finite precision arithmetic are illustrated with a numerical example. The four largest eigenvalue approximations of a matrix of order 567 are given from a separate eigenvalue computation (see Tab.1). The eigenvalues form a cluster with a bad seperation of the values and embeds an eigenvalue of multiplicity 2, ($\lambda_{565} = \lambda_{566}$). The approximation μ_{564} seems to be another multiple of λ_{565}. A Sturm sequence check identified this value as a close but different eigenvalue.

Table 1: Four largest eigenvalues of $\mathbf{A}_{567}$

Eigenvalue index	**Approximation** μ	$\lvert\lambda - \mu\rvert \leq \lVert\mathbf{r}\rVert_2^2/\delta$
λ_{564}	74250.477760425110	2.300e-11
λ_{565}	74250.477760425500	8.290e-10
λ_{566}	74250.477760425540	1.768e-09
λ_{567}	74250.477760435570	1.500e-10

$$\mathbf{r} := \mathbf{A}\,\hat{\mathbf{x}}_i - \mu_i\,\hat{\mathbf{x}}_i \quad : \text{ restnorm} \tag{7}$$

$$\delta = |\rho - \mu| - \|\mathbf{r}\|_2 \qquad \rho := \mathbf{x}_i^T\mathbf{A}\,\mathbf{x}_i \,/\, \mathbf{x}_i^T\mathbf{x}_i \quad : \text{ Rayleigh quotient} \tag{8}$$

Multiplicity 2 of eigenvalue λ_{565} can not be determined by inspection since roundoff introduces an error of order $O(1.e-10)$. Without knowledge about multiplicity of eigenvalues the start matrix $\mathbf{I}_q$ cannot be initialized correctly with $\mathbf{e}_{565}$ and $\mathbf{e}_{566}$, respectively thus avoiding the correct determination of the desired eigenvectors.

A stepwise calculation strategy for this cluster with a subsequent application of the approximate values $\mu_{564}, \ldots, \mu_{567}$ produces the same eigenvector approximation for μ_{565} and μ_{566} since for each QR-decompostion convergence is reached only in the last row of matrix $\mathbf{R}$

thus providing the necessary rotation information for just one eigenvector as shown in Fig.4. Moreover the orthogonality of the four eigenvector approximations is far beyond the required value (9). Figure 4 shows a submatrix with the columns 564 to 567 of the last four rows of the decomposition factor $\mathbf{R}$ using μ_{564} as spectral shift in matrix $\mathbf{C}$.

$$|\mathbf{x}_i^T\,\mathbf{x}_j| \;\leq\; O(\epsilon\, m) \qquad\qquad (i \neq j) \tag{9}$$

[564]	[565]	[566]	[567]
−305.2875822916	−1182.7046563114	6757.7101084395	−2612.7342116394
0.0000000000	−4877.6601864348	17622.4093530023	−20289.3353582083
0.0000000000	0.0000000000	8097.8561548511	8097.8561548511
0.0000000000	0.0000000000	0.0000000000	−0.0000000014

Figure 4: Submatrix $\mathbf{R}_1 = \mathbf{Q}^T(\mathbf{A} - \mu_{565}\mathbf{I})$

The application of any other value of table 1 as spectral shift gives the same picture after the first QR-decomposition. We clearly need an additional QR-decomposition for the reliable determination of the missing eigenvectors for this cluster.

[564]	[565]	[566]	[567]
−0.81507759316372	2.34736594819012	−4.70951765401011	−0.00000000009086
0.00000000000000	0.00000000667527	−0.00000000601069	0.00000000002844
0.00000000000000	0.00000000000000	−0.00000000006852	0.00000000000080
0.00000000000000	0.00000000000000	0.00000000000000	0.00000000000476

Figure 5: Submatrix $\mathbf{R}_2$

Figure 5 shows the same matrix clip after a second QR-decomposition. Three of the eigenvalues show a sufficient convergence for a reliable and accurate set of corresponding eigenvector approximations. The resulting subspace clearly has the good and required orthogonality properties that are known from the original QR-method for dense and fully occupied matrices since its base vectors are determined in the same step with the same rotation informations from the two decompositions.

Submatrix $\mathbf{R}_2$ even shows a tendency for convergence to the forth eigenvalue λ_{567}. The off-diagonal elements in row 564 are relatively small compared to the rows < 564. A third QR-decomposition would bring sufficient convergence for λ_{567} and improve the results for the eigenvectors $\hat{\mathbf{x}}_{564}$ to $\hat{\mathbf{x}}_{566}$ but is not carried out for efficiency reason. Instead a new eigenvector computation is started with the shifted matrix $\mathbf{C} = (\mathbf{A} - \lambda_{567}\,\mathbf{I})$. The seperation of μ_{567} from μ_{564} that was used as spectral shift for the first eigenvector calculation of this cluster is smaller than $(600\,\epsilon\,\|\mathbf{C}\|_2)$ and therefore sufficient to avoid convergence to the subspace $\mathcal{S}^3$ determined

in the foregoing step.

Table 2 shows the euclidian restnorm of the eigenvalue problem for the approximated eigenstates $(\mu, \hat{\mathbf{x}})$ and the difference between the eigenvalue approximation μ and the Rayleigh qoutient ρ. The spectral norm of the standard criterion $\|\mathbf{r}\|_2 \leq m\epsilon \|\mathbf{A}\|_2$ for testing the quality of the solution is usually replaced by the more generous and simpler Frobenius norm $\|\mathbf{A}\|_F$.

Table 2: computational results for the four largest eigenvalues of matrix $\mathbf{A}_{567}$

Eigenvalue	**Restnorm** $\|\mathbf{r}\|_2$	$\|\rho - \mu\|$
λ_{564}	1.34e-10	2.90e-10
λ_{565}	2.97e-10	5.82e-09
λ_{566}	8.99e-09	5.53e-09
λ_{567}	1.04e-08	1.04e-08

The restnorms in table 2 are much better than $4.2e-07 (= 1.\,\epsilon\,\|\mathbf{A}\|_F)$. The values in the third column validate the accuracy level of this solution. The orthogonality level for all eigenvector approximations is better than $1.e-15$ that is nearly the machine precision ϵ. The determined subspace $\mathcal{S}^4$ is completely orthogonal to any other subspace of $\mathbb{R}^N$. The method treats the complete set of clustered eigenvalues in the same eigenvector calculation therefore avoiding the pairing-up problem or loss of orthogonality even when the eigenvalue approximations have little accuracy.

A drawback of the method is a somewhat numerically costly decomposition of about $(6b^2\,N)$ multiplications (b : mean bandwidth) and additional storage requirements of order (bN) for each additional QR-decomposition (see (10)). Thus the efficiency of the method seems to be the bottleneck for large matrices. But the comparison to other standard solution methods shows a different picture and will be discussed in the last chapter. Before, the next section shows how efficiency may significantly be improved.

Improvement of Efficiency With a Local Iteration Scheme

In general eigenvalue clusters are not always as tight as shown in example 1. But sometimes we only have approximations of low accuracy to a multiple eigenvalue or eigenvalues of a cluster. As shown in the last section this results in a triangular matrix $\mathbf{R}$ that is significantly perturbed by a triangular matrix $\mathbf{E}$ thus not providing the desired zero entries in the last row or the last p rows, respectively.

Even so the situation may be much better than in example 1, since further conververgence e.g. for a p-fold eigenvalue is clearly observable. The submatrix in Fig. 6 is taken from the same example problem but shows a different eigenvector calculation.

Figure 6 illustrates the situation after a first QR-decomposition for an eigenvalue of multiplicity 2. Convergence has settled in the last two rows. The applied eigenvalue approximation $\mu_{500} = \mu_{501} = 23881.581266\ldots$ is properly seperated from the remaining eigenvalue set. The

[564]	[565]	[566]	[567]
5242.9794501483	2011.4617871989	−27175.1950832849	−7122.9497996383
0.0000000000	3621.8952269670	−7683.2000272067	−29900.8188398782
0.0000000000	0.0000000000	−0.0001430867	−0.0001400880
0.0000000000	0.0000000000	0.0000000000	−0.0000290795

Figure 6: Submatrix $\mathbf{R}_1 = \mathbf{Q}^T(\mathbf{A} - \mu_{500}\mathbf{I})$

two values μ_{500} and μ_{501} are identical in 11 decimal places and their error is bounded by the error estimate $7.e-11$. Without further numerical effort the error for the eigenvector approximates is expected to be large. For this common situation an essential improvement of the computational result is achieved by a local iteration. Since the lower part of the decomposed matrix already seems to be rich of information about the desired eigenvalues/eigenvectors further computation is limited to this part, accepting a negligible error.

For the local iteration a submatrix $\dot{\mathbf{C}}^{(q\times q)}$ of moderate size $(q < 0.25b)$ is extracted from the diagonal of the iterated matrix $\mathbf{C}_1(=\mathbf{R}_1\,\mathbf{Q})$ (see Fig. 7). The eigenstates $(\eta_k, \mathbf{v}_k), (k = 1,\ldots,q)$ of $\dot{\mathbf{C}}$ are completely determined with a stable Jacobi algorithm in a separate calculation. The eigenmatrix $\mathbf{V}^{(q\times q)}$ is used for the definition of an orthonormal transformation matrix $\mathbf{U}$ as shown in Fig. 7. A similarity transformation of $\mathbf{C}_1$ with $\mathbf{U}$ yields the missing data for a stable and accurate computation of the corresponding eigenvectors $\hat{\mathbf{x}}_{500}$ and $\hat{\mathbf{x}}_{501}$. Instead of the product (10), the much cheaper and acceptable accurate product (11) is computed.

$$\mathbf{X}_q = \mathbf{Q}_2\,\mathbf{Q}_1\,\mathbf{I}_q \tag{10}$$

$$\mathbf{X}_q = \mathbf{Q}_1\,\mathbf{U}\,\mathbf{I}_q \tag{11}$$

Figure 7 shows the accepted error of the similarity transformation in the red shaded part. Zero elements in this part are replaced with numbers $\neq 0$ of very small magnitude. The accepted error is of the order of the applied error border for the off-diagonal elements of the eigenvector computation. The additional computational effort is negligibly small compared to a second QR-decomposition ($< 2\%$ of QR).

Table 3 shows the restnorm error of the eigenstates $(\mu_{500}, \hat{\mathbf{x}}_{500})$ and $(\mu_{501}, \hat{\mathbf{x}}_{501})$ determined with equation (4) (one QR-step) in the first column, equation (10) and equation (11) in the second and third column, respectively. The local iteration clearly improved the restnorm error. Orthogonality between the vectors is maintained.

The numerical results for the small test matrix of dimension 567 may also be transferred to large eigenvalue problems. Figure 8 shows the relative restnorm error for an oscillation problem of an unsupported square plate with 11163 degrees of freedom. The smallest 1000 eigenstates were determined. About a forth of the eigenvalues have multiplicity 2. The 67 critical eigenvectors that required additional effort with local iteration or a second QR-decomposition are marked in the diagramm with blue and orange dots. The remaining eigenstates are rep-

Table 3: Restnorm error

	$\|\mathbf{r}\|_F - (\mathbf{Q}_1\,\mathbf{C})$	$\|\mathbf{r}\|_F - (\mathbf{Q}_2\,\mathbf{Q}_1\,\mathbf{C})$	$\|\mathbf{r}\|_F - (\mathbf{Q}_2\,\mathbf{U}\,\mathbf{C})$
$(\mu_{500}, \hat{\mathbf{x}}_{500})$	$2.002e-04$	$2.443e-09$	$3.121e-07$
$(\mu_{501}, \hat{\mathbf{x}}_{501})$	$2.910e-05$	$2.453e-09$	$2.706e-07$

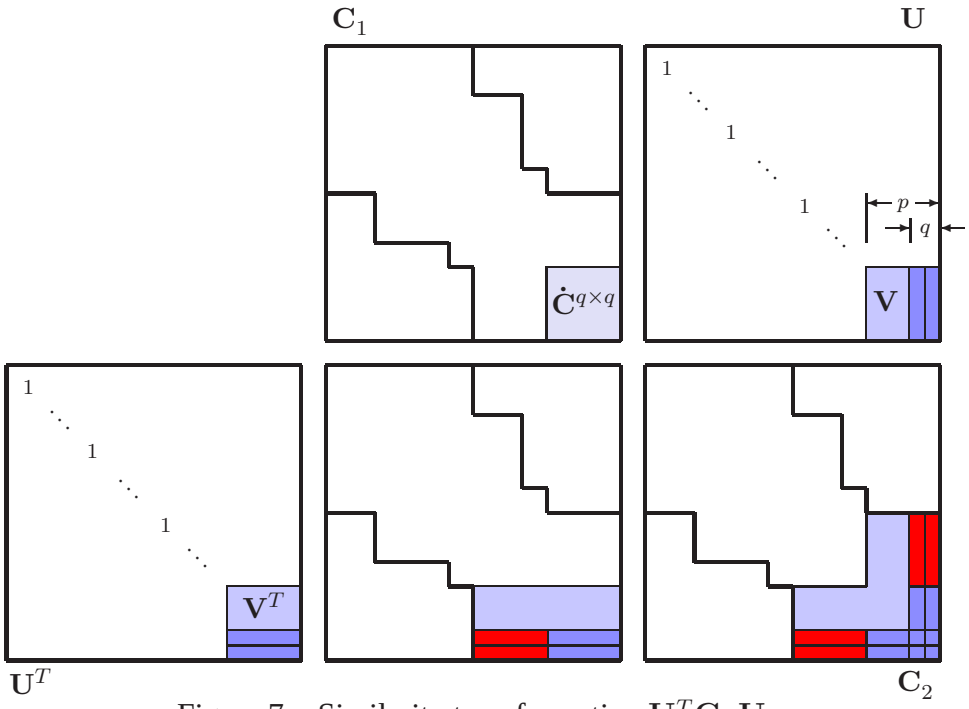

Figure 7: Similarity transformation $\mathbf{U}^T\mathbf{C}_1\,\mathbf{U}$

resented by their mean value, marked with a green line. The additional effort for a second QR-decomposition reflects a relative high accuracy.

CONCLUSIONS - QR vs. Inverse Iteration

The presented method is very reliable and accurate since it keeps the good numerical properties of the original QR-method for dense matrices. Compared to the standard solution method for independent eigenvector computation, the inverse vector iteration, the QR-method has in the general case a higher numerical effort.

Dhillon (1998) systematically describes the drawbacks of inverse iteration that also completely occurred in the linear algebra package that was used as analysis platform for this contribution. The pairing-up problem may result in a small error but a completely wrong eigenvector approximation. For more than one eigenvalue approximation in the neighborhood of the applied spectral shift the system of equations that is solved as part of the iteration is

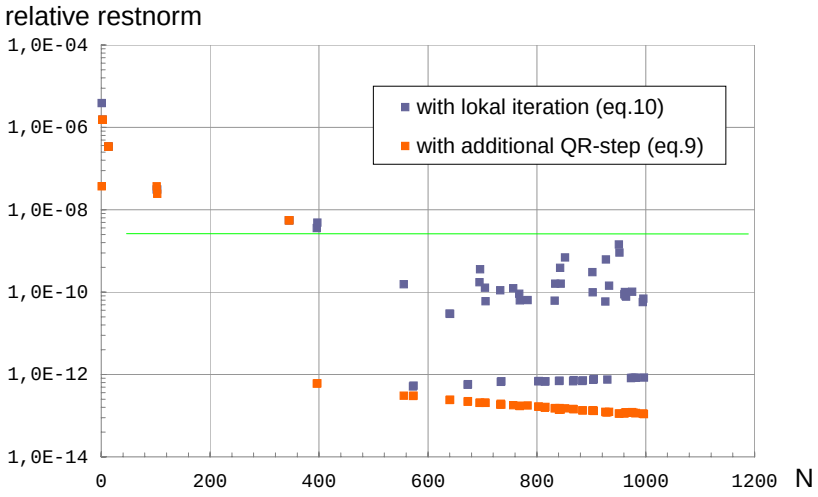

Figure 8: A posteriori error for $\mathbf{K}_{11163}$

highly sensitive for small perturbations and has therefore be artificially perturbed which may lead to loss of accuracy.

The QR-method does not depend on a well-seperated shift parameter and even may profit from multiple or clustered eigenvalues, since their eigenvectors may be computed in many cases with a single decompositon, maintaining orthogonality at machine precision level.

Orthogonality between the eigenvector approximates is often lost with inverse iteration when a larger number of eigenvectors is calculated. Especially in the case of bad seperated eigenvalue clusters a reorthogonalization procedure for the complete calculated subspace is obligatory and must be frequently repeated. Thus the complete set of vectors are kept in the main storage. Unfortunately reorthogonalisation does not necessarily bring the expected success, since strongly parallel vectors may lead to desastrous cancellation of significant directions in the iterated vector and finally completely fails. The result is a random vector, despite a small restnorm.

None of the calculations with the QR-method used any numerically expensive orthogonalisation process. Hence the reorthogonalisation problem of inverse iteration and the lack of a reliable computation strategy are the criteria that make the QR-method superior to subspace methods even with a higher calculation effort.

REFERENCES

Dhillon I.S. (1998). *Current inverse iteration software can fail*, IBM Almaden Research Center, San Jose, California.

Parlett B.N. (1998). *The Symmetric Eigenvalue Problem*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia.

Ruess M. (2002). *Matrix Iteration For Large Symmetric Eigenvalue Problems*, ICCCBE IX, Taipei, Taiwan