Università di Pisa
Dipartimento di Informatica

Technical Report: TR-10-14

A unification of unitary similarity transforms to compressed representations

Raf Vandebril Gianna M. Del Corso

August 3, 2010

ADDRESS: Largo B. Pontecorvo 3, 56127 Pisa, Italy. TEL: +39 050 2212700 FAX: +39 050 2212726

A unification of unitary similarity transforms to compressed representations*

Raf Vandebril † Gianna M. Del Corso‡

August 3, 2010

Abstract

In this paper a new framework for transforming arbitrary matrices to compressed representations is presented. The framework provides a generic way of transforming a matrix via unitary similarity transformations to e.g. Hessenberg, Hessenberg-like and combinations of both. The new algorithms are deduced, based on the QR-factorization of the original matrix. Based on manipulations with Givens transformations, all the algorithms consists of eliminating the correct set of Givens transformations, resulting in a matrix obeying the desired structural constraints.

Based on this new reduction procedure we investigate further correspondences such as irreducibility, unicity of the reduction procedure and the link with (rational) Krylov methods.

The unitary similarity transform to Hessenberg-like form as presented here, differs significantly from the one presented in earlier work. Not only does it use less Givens transformations to obtain the desired structure, also the convergence to rational Ritz values is not observed in the standard way.

1 Introduction

The unitary similarity transformation of a matrix to Hessenberg form (see e.g. [15, 12]) is already known and used for a long time[1]. This similarity transform is e.g. essential in the development of efficient QR-algorithms [33, 22, 34]. This reduction process has also an intimite relation with three terms recurrences for orthogonal polynomials [15, 5]. In this context, also the partial reduction to Hessenberg form as carried out in iterative processes plays an important role in many iterative system solvers as well as iterative eigenvalue methods [23, 24].

Hessenberg-like matrices (also named lower semiseparable matrices) can roughly (excluding thereby singular Hessenberg-like matrices) be considered as the inverses of Hessenberg matrices[2]. Recently a unitary similarity transformation to Hessenberg-like (lower semiseparable) form was presented [27, 3]. Since Hessenberg-like matrices are the inverses of Hessenberg matrices [29, 30] the existence of such reduction came not really as a surprise. Nevertheless, it was the first constructive algorithms for obtaining such a matrix form.

In [27] the unitary similarity transformation to semiseparable form (Hermitian Hessenberg-like) was explained and initiated research towards new, alternative eigenvalue and singular value methods [29, 30]. This reduction procedure possessed a specific convergence behavior [32] revealing well separated eigenvalues. Shortly one can state that during the reduction procedure, the part

*The first author has a grant as "Postdoctoraal Onderzoeker" from the Fund for Scientific Research–Flanders (Belgium). This research was also partially supported by the Interuniversity Attraction Poles Programme, initiated by the Belgian State, Science Policy Office, Belgian Network DYSCO (Dynamical Systems, Control, and Optimization)

†KULeuven, Dept. of Computer Science, 3001 Leuven(Heverlee),Belgium, (raf.vandebril@cs.kuleuven.be)

‡Università di Pisa, Dept. of Computer Science, Largo Pontecorvo, 3, 56127 Pisa, Italy, (delcorso@di.unipi.it)

[1] Even though the results are applicable to symmetric matrices as well, i.e. a reduction to tridiagonal form, we will restrict ourselves to the generic nonsymmetric case

[2] Similarly semiseparable matrices are roughly the inverses of tridiagonal matrices

already in the correct form contains the Ritz-values and on this part extra steps of the QR-method (with or without shift) are performed. These extra QR-steps are not present in the reduction to Hessenberg form, therefore, this method uses $\mathcal{O}(n^2)$ more Givens transformations than the standard reduction to Hessenberg form.

For the classes of Hessenberg-like, semiseparable, semiseparable plus diagonal results are obtained relating the unitary similarity transforms for obtaining these matrices to orthogonal rational functions [26, 25]. Also the iterative methods as well as the corresponences with Krylov sequences are studied [14, 3, 7, 20]. In [2] it is shown that it is possible to capture all these reductions in a unified framework. The unitary matrices determining the similarity transformations can be obtained from suitably constructed Krylov matrices. The proof is, however, not algorithmic since it involves computing an orthogonal basis for an appropriate Krylov subspace.

Unfortunately, as already discussed before, the reduction to Hessenberg-like form uses $\mathcal{O}(n^2)$ more Givens transformations than the reduction to Hessenberg form. Since, however, Hessenberg and Hessenberg-like matrices are each others inverses, one wonders if it is possible to reduce an arbitrary matrix to Hessenberg-like by performing exactly the same number of Givens transformations as the algorithm for reducing a matrix to Hessenberg form. Moreover also reductions to higher order Hessenberg-like matrices are not known, whereas the unitary similarity transforms to generalized Hessenberg matrices (having more than only one subdiagonal) are known.

In this paper, we will present a unifying framework for unitary similarity transforms for reducing matrices either to band or higher order Hessenberg-like forms. It will be shown that all these reduction algorithms are in fact variations of the same idea. Moreover this design enables us also to perform unitary similarity transforms for obtaining mixed forms, which means combinations of Hessenberg and Hessenberg-like matrices. These matrices will have parts below a specified subdiagonal of a specified rank, where the rank and the subdiagonal in consideration are freely chosen parameters. This new framework enables us to derive some theoretical results related to Krylov sequences, to identify the links between proper Hessenberg and Hessenberg-like matrices and to provide theorems related to essential uniqueness of the reduction. For example, it will be proven that the new reduction to Hessenberg-like form will not inherit the standard Ritz-value convergence as in [27], but a rational Ritz-value behavior.

In the paper we assume that we want to transform a generic matrix into a similar one with an Hessenberg or Hessenberg-like structure in the lower triangular part. This is done starting from the QR representation of the original matrix and obtaining the QR representation of the transformed one. Of course, one can obtain algorithms for transforming a matrix into a similar one with for example a lower Hessenberg structure, the idea is then to use QL-factorization instead of the QR representation.

The article is organized as follows. Section 2 discusses preliminary results, essential for understanding the article. In Section 3 the unitary similarity transformations to the different structures are presented. In Section 4 it is shown that the reduction process to Hessenberg form or to Hessenberg-like form can be captured in a unifying framework. Section 5 contains the description of the reduction to mixed structures, that is the sum of a generalized Hessenberg and a generalized Hessenberg-like matrix. In Section 6 the relation of the reductions and Krylov subspaces is shown. The relation is further investigated in Section 7 were is is shown that the reduction to Hessenberg-like form inherits the convergence to the Rational-Ritz values rather than the standard convergence. Experimental evidence on the different convergence behavior of the reduction process is shown in Section 8.

2 Preliminary results

This section discusses definitions of the involved matrices; graphical schemes necessary for simplifying the understanding of the algorithms; manipulations between Givens transformations and several examples underpinning the statements.

2.1 Definitions

A $\{p\}$-Hessenberg matrix is a generalization of a standard Hessenberg matrix having all entries below the subdiagonal zero.

Definition 1 *A matrix $H = (h_{ij})_{ij} \in \mathbb{C}^{n \times n}$ is a $\{p\}$-Hessenberg matrix, $p \geq 0$, when:*

$$h_{ij} = 0 \ \textit{ for all } \ i > j + p.$$

This means that below the p-th subdiagonal, the matrix equals zero.

For symmetric matrices, one obtains a band matrix with bandwidth $2p + 1$.

The inverses of $\{p\}$-Hessenberg matrices are named $\{p\}$-Hessenberg-like matrices and are of structured rank form [29].

Definition 2 *A matrix Z is called a $\{p\}$-Hessenberg-like matrix*[3]*, $p \geq 0$, when:*

$$\operatorname{rank}\left(A(i : n, 1 : \min\{i + p - 1, n\}\right) \leq p.$$

This means that all submatrices taken out of the part below the p-th-superdiagonal have rank at most p.

These matrices are also often referred to as $\{p\}$-lower semiseparable. Note that Definitions 1 and 2 coincide for $p = 0$ and in this case we have an upper triangular matrix.

In this paper we will consider also unitary transformations to matrices having a mixed structure, which can, e.g., be expressed as the sum of a $\{p\}$-Hessenberg and a $\{q\}$-Hessenberg-like matrix.

We will refer to $\{p\}$-Hessenberg or $\{p\}$-Hessenberg-like matrices also as generalized Hessenberg or generalized Hessenberg-like matrices, respectively.

2.2 Graphical representations

The algorithms deduced in the upcoming sections are explained by graphical schemes representing the interactions between different Givens transformations and the matrix itself. The algorithms are based on the QR-factorization of the original matrix. Therefore, we introduce the graphical representation by constructing the QR-factorization of a matrix A. Without loss of generality we assume $A \in \mathbb{C}^{5 \times 5}$ as this illustrates the general case. We denote the elements of the matrix $A = (a_{ij})$ with $\times$. To construct its QR-factorization, first the element a_{51} is annihilated by a Givens transformation G_{51}^H. As a result we obtain $G_{51}^H A$ having element $(5, 1)$ zero. Graphically this is depicted as follows:

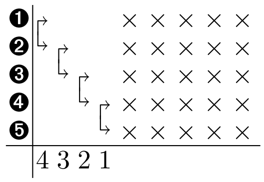
$$
\begin{array}{c|c c}
❶ & & \times \times \times \times \times \\
❷ & & \times \times \times \times \times \\
❸ & & \times \times \times \times \times \\
❹ & \lceil & \times \times \times \times \times \\
❺ & \lfloor & \times \times \times \times \times \\
\hline
 & 1 &
\end{array}
\quad = \quad
\begin{bmatrix}
\times & \times & \times & \times & \times \\
\times & \times & \times & \times & \times \\
\times & \times & \times & \times & \times \\
\times & \times & \times & \times & \times \\
0 & \times & \times & \times & \times
\end{bmatrix}.
$$

The brackets with arrows depicts the Givens transformation G_{51}^H acting on row 4 and row 5. The horizontal axis presents a sort of timeline, depicting which Givens needs to be performed first and so forth. The vertical axis simply numbers the rows. The elements of the matrix A on the left are represented with $\times$, clearly the matrix $G_{51}^H A$ on the right has already one zero.

One continues the procedure by successively eliminating the elements a_{41}, a_{31} and finally a_{21}. This results in the matrix $G_{21}^H G_{31}^H G_{41}^H G_{51}^H A = Q_1^H A$ having the entire first column except the first

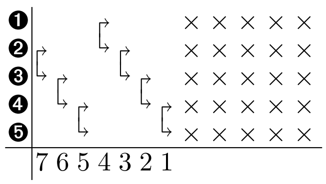
[3]With $A(i : j, \ell : k)$ we use the standard colon notion: depicting the submatrix with columns ℓ to k and rows i to j.

element equal to zero. Take $Q_1^H = G_{21}^H G_{31}^H G_{41}^H G_{51}^H$ as the unitary transformation creating zeros throughout the entire first column. Schematically we get the following:



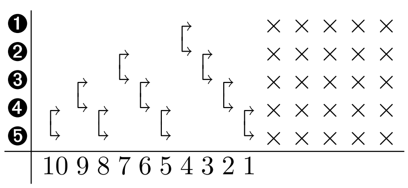
$$= \begin{bmatrix} \times & \times & \times & \times & \times \\ 0 & \times & \times & \times & \times \\ 0 & \times & \times & \times & \times \\ 0 & \times & \times & \times & \times \\ 0 & \times & \times & \times & \times \end{bmatrix}.$$

Continuing this procedure, one eliminates all elements in the second column below the subdiagonal. We get $G_{32}^H G_{42}^H G_{52}^H Q_1^H A = Q_2^H Q_1^H A$ depicted as follows:



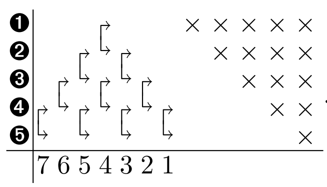
$$= \begin{bmatrix} \times & \times & \times & \times & \times \\ 0 & \times & \times & \times & \times \\ 0 & 0 & \times & \times & \times \\ 0 & 0 & \times & \times & \times \\ 0 & 0 & \times & \times & \times \end{bmatrix}.$$

This procedure can be continued and we get $Q_4^H Q_3^H Q_2^H Q_1^H A = Q^H A = R$. The product $Q^H A$ can be written as follows:



$$= \begin{bmatrix} \times & \times & \times & \times & \times \\ 0 & \times & \times & \times & \times \\ 0 & 0 & \times & \times & \times \\ 0 & 0 & 0 & \times & \times \\ 0 & 0 & 0 & 0 & \times \end{bmatrix}. \tag{1}$$

Hence if we bring the unitary transformation Q to the other side we get $A = QR$, schematically of the following form:



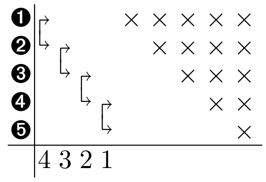
(2)

Comparing the representation of the unitary matrix Q^H in Equation (1) and Q in Equation (2), we see that they are represented in a compressed form in (2). The brackets clearly denote on which rows the Givens transformations act, hence some of the Givens transformations commute and their order of performance can be changed. In (2), e.g. two Givens transformations can be executed simultaneously in the third step, one acting on rows 2 and 3 and the other on rows 4 and 5. These schemes contain a lot of useful information, since they clearly indicate the order of the transformations and the rows they act on.

Remark 1 *Throughout the manuscript we assume to be working with rotations, i.e. having determinant equal to* 1*. Often* 2×2 *unitary matrices are also referred to as Givens transformations. In fact all results presented in this article also hold for* 2×2 *unitary matrices. More information on Givens rotations and how to compute them reliably can be found in [4].*

Let us consider the representations of some typical examples such as a Hessenberg, a Hessenberg-like matrix and combinations.

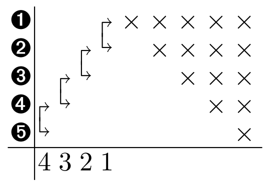
Example 1 (A Hessenberg matrix) *Suppose* H *to be of Hessenberg form with* $n = 5$*. The QR-factorization (constructed with Givens transformations)* $H = QR$ *is schematically represented* *as follows.*



(3)

The matrix Q consists of four Givens transformations, their actions and order can be deduced from the scheme. Performing these transformations on the right upper triangular matrix R will fill up the subdiagonal elements and create a Hessenberg matrix.

Example 2 (A Hessenberg-like matrix) *Suppose Z to be of Hessenberg-like form. Due to the low rank structure of this matrix, computing the QR-factorization only involves $n-1$ Givens transformations. Schematically $Z = QR$ (for $n = 5$) is depicted as follows.*

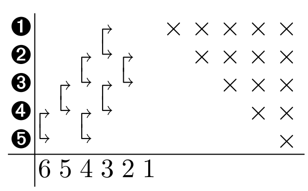


(4)

Performing the Givens transformations on the upper triangular matrix fills it up with a low rank part, all matrices taken out of the part below the diagonal are of rank at most one.

We remark that this QR-factorization of the Hessenberg-like matrix corresponds to the Givens-vector representation (see e.g. [29]). This is a compact representation making it possible to derive efficient algorithms for the class of Hessenberg-like matrices (e.g. efficient QR-factorization, QR-algorithms, Levinson-like methods, LU-factorizations, inversion, etc.).

Example 3 (A $\{2\}$-Hessenberg-like matrix) *Assume $n = 5$ and $p = 2$. Schematically $Z = QR$ is of the following form.*



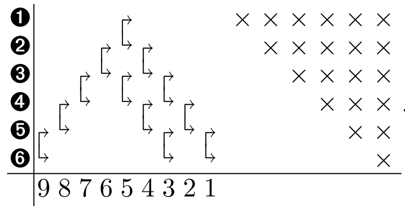
Performing the Givens transformations on the upper triangular matrix fills it up with a low rank part, such that all matrices taken out of the part including the superdiagonal have rank at most 2.

It is well known that the QR factorization of a matrix is not unique, in fact, if $A = QR$, and S is a phase matrix, i.e. S is diagonal and $|S_{ii}| = 1$, then $A = Q_1 R_1$, where $Q_1 = QS$ and $R_1 = S^H R$. In the case A is nonsingular, however the factorization is unique assuming the diagonal entries of R to be real and positive.

Example 4 (Sum of a Hessenberg and Hessenberg-like) *Suppose A to be the sum of Hessenberg and an Hessenberg-like matrix. This means that all the submatrices taken out of the part below the subdiagonal have rank at most one. For a 6×6 matrix we have the following situation*

$$A = \begin{bmatrix} \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ \boxtimes & \times & \times & \times & \times & \times \\ \boxtimes & \boxtimes & \times & \times & \times & \times \\ \boxtimes & \boxtimes & \boxtimes & \times & \times & \times \\ \boxtimes & \boxtimes & \boxtimes & \boxtimes & \times & \times \end{bmatrix},$$

where the entries denoted by $\boxtimes$ *represent the rank-one part. Due to the low rank structure, computing the QR factorization requires only* $3(n-2)$ *Givens transformations, instead of the* $n(n-1)/2$ *needed for an unstructured matrix. Schematically* $A = QR$ *is depicted as follows*



(5)

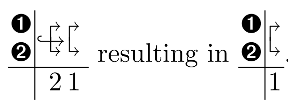
Example 5 (Generic summation) *In the general case, where* A *is the sum of a* $\{p\}$*-Hessenberg matrix and a* $\{q\}$*-Hessenberg-like matrix, we need* $q(2(n-p)-q-1)/2$ *Givens transformation to remove the rank* q *part. Once we have annihilated the rank* q *part,* q *new diagonals have been formed and we need* $\sum_{i=1}^{p+q}(n-i)-1 = (p+q)(2n-(p+q)-1)/2-1$ *more Givens transformation to obtain an upper triangular matrix* R.

2.3 Manipulating Givens transformations

The forthcoming algorithms depend heavily on manipulating Givens transformations. These operations are already described and proved in elsewhere (see e.g. [29, 30]). For completeness, however, we will briefly reconsider them in this section, without proofs.

Lemma 1 *Suppose two Givens transformations* G_1 *and* G_2 *are given. Then we have that* $G_1 G_2 = G_3$ *is again a Givens transformation. We will call this the fusion of Givens transformations in the remainder of the text.*

Graphically we depict this as follows:



Often Givens transformations of higher dimensions, say n, are considered. This means that the corresponding 2×2 Givens transformation is embedded in the identity matrix of dimension n, still acting on only two rows when applied to the left. The next lemma states that we can change the order of the Givens transformations.

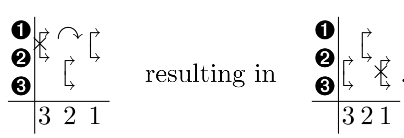
Lemma 2 (Shift-through operation) *Suppose three* 3×3 *Givens transformations* $\check{G}_1, \check{G}_2$ *and* $\check{G}_3$ *are given, such that the Givens transformations* $\check{G}_1$ *and* $\check{G}_3$ *act on the first two rows of a matrix, and* $\check{G}_2$ *acts on the second and third row (when applied on the left to a matrix).*

Then there exist 3 *Givens transformations* $\hat{G}_1, \hat{G}_2$ *and* $\hat{G}_3$ *such that*

$$\check{G}_1 \check{G}_2 \check{G}_3 = \hat{G}_1 \hat{G}_2 \hat{G}_3,$$

where $\hat{G}_1$ *and* $\hat{G}_3$ *work on the second and third row and* $\hat{G}_2$*, works on the first two rows.*

This result is well-known. The proof is based on the different ways to factorize a 3×3 unitary matrix (see [31]). Schematically we get.



The other direction (from the right to the left scheme is depicted by $\curvearrowleft$).

We remark that the fusion of Givens is a special instance of the shift-through operation.

A final important operation is the shift-through operation of length ℓ.

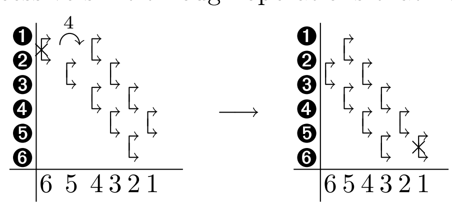
Lemma 3 (Shift-through operation of length ℓ) *Suppose we have the following matrix product GWX, in which G denotes a Givens transformation acting on row 1 and 2. The matrices W and X are both unitary matrices consisting of a descending sequence of ℓ Givens transformations. This means that both W and X consist of ℓ successive Givens transformations. The ith Givens transformation G_i^W of W acts on row $i+1$ and $i+2$. The ith Givens transformation G_i^X of X acts on row i and $i+1$.*

The matrix product GWX can then be rewritten as

$$GWX = \hat{W}\hat{X}\hat{G},$$

where $\hat{G}$ is now a Givens transformation acting on row $\ell+1$ and $\ell+2$. The unitary matrices $\hat{W}$ and $\hat{X}$ are again descending sequences of ℓ Givens transformations.

Schematically we obtain that the following two schemes represent the same action of Givens transformations. Hence this represents the same unitary transformation, but factored differently as a sequence of Givens transformations. In Scheme (6) the Givens transformation G and $\hat{G}$ are marked with a $\times$, the sequence W ranges from position 5 to 2 and sequence X ranges from position 4 to 1. A shift-through operation of a specified length is depicted by adding a super or subscript indicating the number of successive shift-through operations that need to be performed.



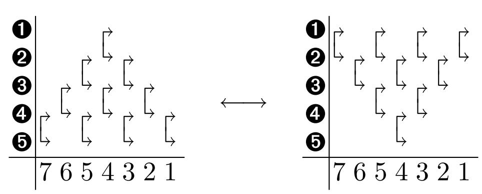
(6)

On the right we see the resulting scheme in which the new sequence $\hat{W}$ ranges from 6 to 3 and $\hat{X}$ ranges from 5 to 2.

One can also perform shift-through operations from right to left and from top to bottom. To indicate where the marked Givens transformation is going to and to specify the desired number of single shift-trough to complete the task, we use the symbols $\overset{\ell}{\curvearrowright}$, $\underset{\ell}{\circlearrowleft}$, $\underset{\ell}{\circlearrowright}$ and $\overset{\ell}{\curvearrowleft}$.

Example 6 (Factorization of a unitary matrix) *An interesting result of the shift-through lemma is the factorization of a unitary $n \times n$ matrix. The shift-through lemma already illustrated sort of $\vee$ and $\wedge$-pattern for factorizing a unitary matrix.*

An arbitrary unitary matrix U can be factorized in sequences of Givens transformations as the left matrix in Scheme (7), in fact one just computes the QR-factorization and R becomes the identity matrix[4]*. Applying successive shift-through operations can change the $\wedge$-form on the left in a $\vee$-form on the right*[5]*.*



(7)

This example states in a certain sense that also the QR-factorization of an arbitrary matrix can be computed in an alternative manner. This is indeed the case, more details can be found in [31].

3 The unitary similarity transforms

The algorithms deduced in this section work on the QR-factorization of the given matrix A and unitarily transform this QR-factorization to the desired representation.

[4]For simplicity we assume the determinant of U equal to 1, otherwise, since all Givens transformations have determinant 1, we will end up with a diagonal matrix R equal to the identity except for the trailing element $r_{nn} = \det U$. The appearance of this unimodular factor does not pose any problems.

[5]Of course many other factorizations are possible, exploiting the shift-through lemma.

3.1 Unitary similarity transformation to Hessenberg form

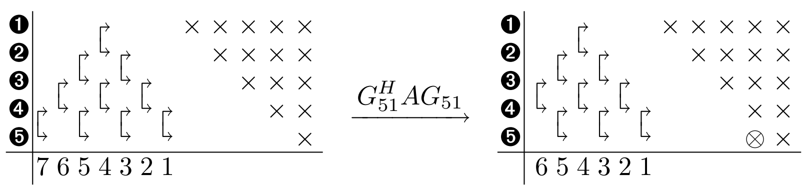
The unitary similarity transformation of an arbitrary matrix to Hessenberg form is well-known [33, 15]. Let us interpret this reduction based on the QR-factorization. Without loss of generality we will apply this procedure on a 5×5 matrix, the general reduction process proceeds identically.

Even though the description here seems very different from the standard reduction technique one can easily prove that both approaches are identical. Based on the implicit Q-theorem, it is known that the first column of the unitary similarity transformation defines the Hessenberg matrix uniquely, up to the first zero subdiagonal element.

Assume a dense matrix A is given[6] having a QR-factorization of the form (14). Assume, without loss of generality, that A has dimensions 5×5. We have $A = QR = Q_1 Q_2 Q_3 Q_4 R$, where Q_1^H annihilates elements in the first column of A, Q_2^H in the second column and so forth. Each unitary matrix Q_i is a combination of Givens transformations $Q_1 = G_{51} G_{41} G_{31} G_{21}$, transformation G_{51}^H annihilates element a_{51} and so forth.

In the end we would like to obtain a factorization as in Scheme (3). To achieve this goal all Givens transformations in the QR-factorization, except $G_{21}, G_{32}, G_{43}, G_{54}$ need to be removed. The remaining Givens transformations in the final representation of the Hessenberg matrix will not remain the same, but they will change as effect of fusion or shift-through operations.

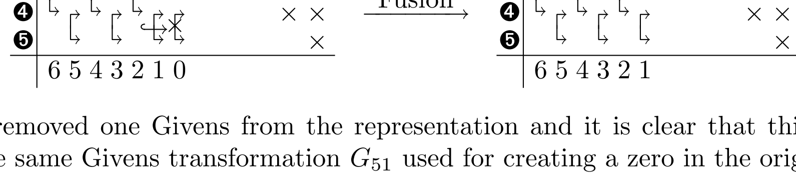
Let us start by removing one Givens transformation at a time by performing a unitary similarity transformation. We have $A^{(51)} = G_{51}^H A G_{51} = G_{41} G_{31} G_{21} Q_2 Q_3 Q_4 R G_{51}$. Schematically we get the following transition. In the right scheme R has been transformed into RG_{51} having a bulge in position $(5, 4)$.



Since our goal is to obtain the QR-factorization of a Hessenberg matrix, we want to remove the bulge $\otimes$ in the R-factor in position $(5, 4)$ immediately. A Givens transformation $\tilde{G}$ is constructed to complete this job, $R^{(51)} = \tilde{G} R G_{51}$ is upper triangular again:

$$
\begin{aligned}
A^{(51)} &= G_{41} G_{31} G_{21} Q_2 Q_3 Q_4 \tilde{G}^H \tilde{G} R G_{51} \\
&= G_{41} G_{31} G_{21} Q_2 Q_3 Q_4 \tilde{G}^H R^{(51)}.
\end{aligned}
$$

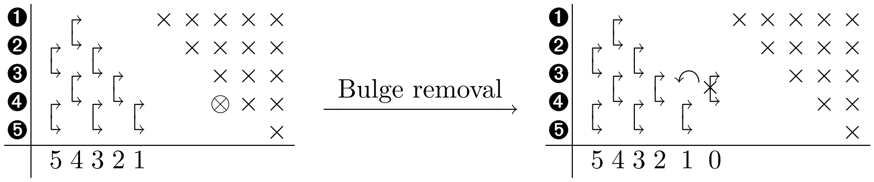
Reconsidering the schemes, we see that we can remove $\tilde{G}^H$ by a simple fusion with the lower right Givens transformation of Q_4. We remark that in the previous scheme, the left scheme represented A and the right scheme represented $A^{(51)}$. The schemes below represent twice $A^{(51)}$, with a slightly changed representation. On the left $\tilde{G}^H$ is marked with a $\times$ symbol since it is undesired. Even though it is not visibly depicted in the scheme on the right, the transformation in position 1 has changed due to the fusion.



We have removed one Givens from the representation and it is clear that this is performed by exactly the same Givens transformation G_{51} used for creating a zero in the original matrix A. Since the fusion altered a Givens transformation in the unitary factor Q_4 we have the following

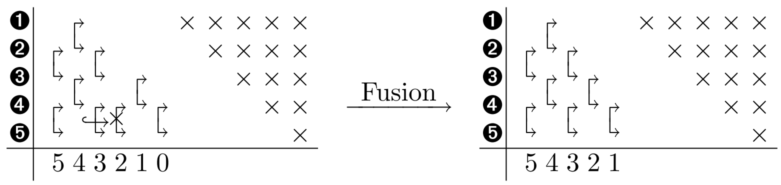
[6] In case the matrix A has some specific structure, the QR-factorization might contain less Givens transformations. This procedure will, however, remain valid but simplify since some Givens transformations will equal the identity matrix.

factorization for $A^{(51)} = G_{41}G_{31}G_{21}Q_2Q_3Q_4^{(51)}R^{(51)}$. The next unitary similarity is determined by G_{41} and removes this Givens transformation from the factorization. Performing it onto $A^{(51)}$, will create a new bulge in the matrix $R^{(51)}$. The matrix $A^{(41)} = G_{41}A^{(51)}G_{41}^H$ is depicted in the left of Scheme (8). On the right we see that another Givens transformation $\tilde{G}$ is used for annihilating the bulge.



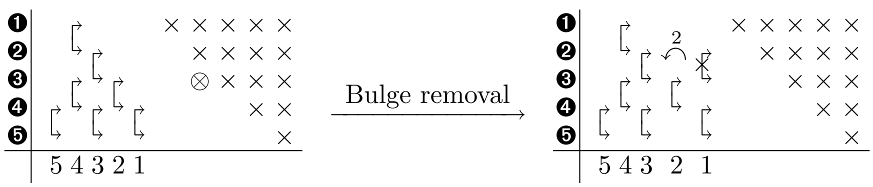
(8)

Unfortunately, a fusion cannot be applied this time, first a shift-through operation need to be performed. After the shift-through operation we get the left of Scheme (9) in which a new fusion is depicted and the Givens transformation disturbing the nice structure is again marked with $\times$. After a fusion we obtain the right figure, which represents the QR-factorization of the matrix $A^{(41)} = G_{31}G_{21}Q_2Q_3^{(41)}Q_4^{(41)}R^{(41)}$.

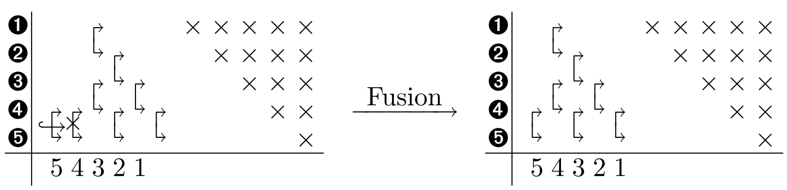


(9)

Removing the Givens transformation G_{31} is done by performing a similarity transformation $G_{31}^H A^{(41)} G_{31}$ resulting in $A^{(31)} = G_{21}Q_2Q_3^{(41)}Q_4^{(41)}R^{(41)}G_{31}$ has again a bulge created in its R factor in position $(3,2)$.

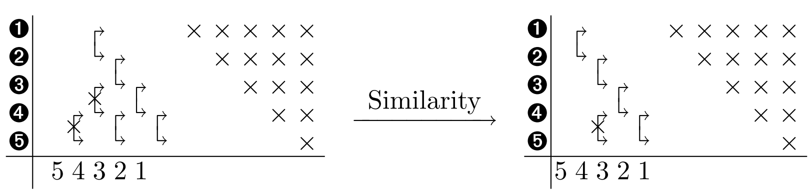


The shift-through operation of length 2 is depicted. To finish up one needs to perform another fusion.



We have now a factorization of the form $A^{(31)} = G_{21}Q_2^{(31)}Q_3^{(31)}Q_4^{(31)}R^{(31)}$. We will not remove the transformation G_{21}, since it is a part of the QR-factorization of the Hessenberg matrix.

The reader can easily verify that next, one removes all but one Givens transformations from the unitary matrix $Q_2^{(31)}$. On the left of Scheme (10) these Givens transformations are marked with a $\times$. Similar techniques as the ones discussed before can be used for obtaining the scheme on the right. To obtain the Hessenberg matrix, a final givens transformation needs to be removed, this transformation is marked on the scheme on the right of (10).



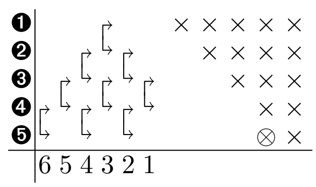
(10)

In the right figure one more Givens transformation is marked as undesired. Once this transformation is removed we obtain the QR-factorization of a Hessenberg matrix and the procedure is completed. One can easily verify that the Givens transformations used for transforming the matrix to Hessenberg form are identical to the ones that would be used when creating zeros directly in the matrix A.

3.2 Unitary similarity transformation to Hessenberg-like form

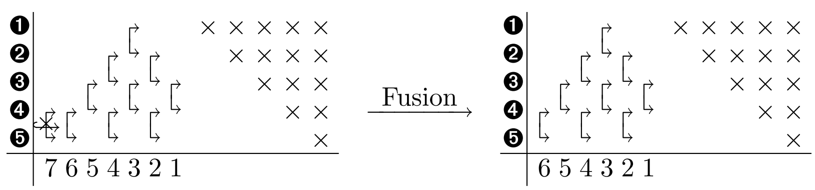
The reduction to Hessenberg-like form proceeds almost identical to the reduction to Hessenberg form. In the reduction to Hessenberg form Givens transformations were removed from the left side of the Q-factor. Each similarity transformation created a bulge in the R-factor which was then incorporated in the remaining Givens transformations of the matrix Q. Here, we will remove Givens transformations from the right-side in the unitary factor of Q. One removes these transformations by creating a bulge in the matrix R and then removing the bulge. At the end all Givens transformations except G_{51}, G_{41}, G_{31} and G_{21} are removed.

Let us illustrate the procedure. We have $A = Q_1 Q_2 Q_3 G_{54} R$. To remove the transformation G_{54} we perform it onto R. In Scheme (11) see that R has been transformed into $G_{54}R$ having a bulge in position $(5, 4)$.

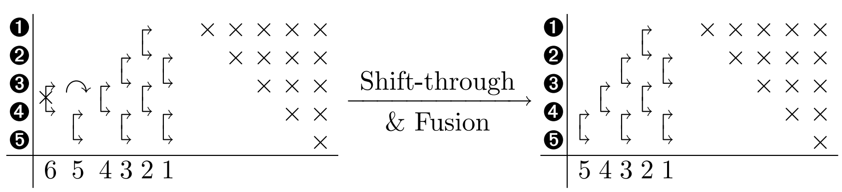


(11)

The similarity transformation is now determined by the Givens transformation $\tilde{G}_{54}$ created such that $(G_{54}R)\tilde{G}_{54} = R^{(54)}$ is again an upper triangular matrix. Mathematically we get $A^{(54)} = \tilde{G}_{54}^H A \tilde{G}_{54} = \tilde{G}_{54}^H Q_1 Q_2 Q_3 R^{(54)} = Q_1^{(54)} Q_2 Q_3 R^{(54)}$. The Givens transformation $\tilde{G}_{54}^H$ is removed by a fusion $\tilde{G}_{54}^H Q_1 = Q_1^{(54)}$ is depicted in the left scheme and executed in the right scheme.

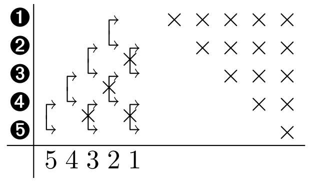


The global flow of the method should be clear now. In the next step one removes Givens transformation G_{43} from the inside of the QR-factorization of $A^{(54)} = Q_1^{(54)} Q_2 G_{53} G_{43} R^{(54)}$. This task is accomplished by a similarity transformation on $A^{(54)}$ with Givens transformation $\tilde{G}_{43}$ determined such that $(G_{43}R^{(54)})\tilde{G}_{43} = R^{(43)}$ is again an upper triangular matrix. On the left of Scheme (12) the matrix $A^{(43)} = \tilde{G}_{43}^H A^{(54)} \tilde{G}_{43}$ is shown, with a shift-through operation depicted. The shift-through operation is followed by a fusion to obtain the right scheme. The ideas are identical to the one used for transforming a matrix to Hessenberg form.



(12)

Next one removes all the Givens transformations marked with a $\times$ in the next scheme, similarly as done before.



After having removed this marked Givens transformations we obtain the QR-factorization of a Hessenberg-like matrix.

In fact one can see both the similarity transformation to Hessenberg and to Hessenberg-like form as two variants of the same method. In the first one removes the Givens on the left, and in the second variant one removes the Givens transformations on the right of the unitary matrix Q.

Remark 2 *Important to note is that when reducing a matrix A to Hessenberg form by the standard well-known similarity transformation uses $(n-1)(n-2)/2$ Givens transformations, which is the minimum number of transformations needed, because it is the same number of entries in the lower triangular part to be annihilated.*

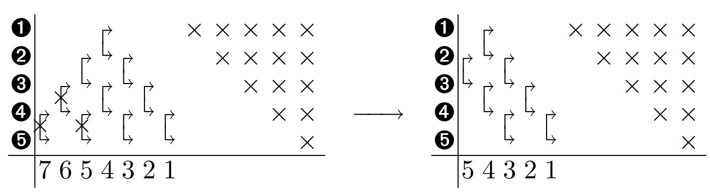
The standard reduction to Hessenberg-like form as presented in [27] uses almost twice as much Givens similarity transformations since these matrices are in fact first brought to Hessenberg form, after which a sort of chasing technique is needed before retrieving to Hessenberg-like format.

The technique presented here does, however, also needs only $(n-1)(n-2)/2$ Givens transformations. Hence the procedure deduced here is significantly different from the one in [27]. Moreover based on [27] it is not straightforward how to transform a matrix to $\{p\}$-Hessenberg-like form, especially since it is expected that this should be computationally cheaper than the reduction to Hessenberg-like form. In the Sections 3.3 and 3.4, we will see that also the unitary similarity transforms to generalized Hessenberg and Hessenberg-like forms nicely fit in this framework and are simple variations of the same idea.

3.3 Unitary similarity transformation to generalized Hessenberg form

The Q-factor in the QR-factorization of a standard Hessenberg matrix consists of a sequence of descending Givens transformations. In case of a $\{p\}$-Hessenberg matrix, we will have p descending sequences of Givens transformations. For example on the right of Scheme (13) one sees the structure of the QR-factorization of a $\{2\}$-Hessenberg matrix.

In fact we have already deduced a way to construct a similarity transformation obtaining this structure. In fact one follows the same procedure as for the reduction to Hessenberg form, the only difference is that one removes less Givens transformations. In the following QR-factorization the Givens transformations marked with $\times$, should be removed in order to obtain a $\{2\}$-Hessenberg matrix.

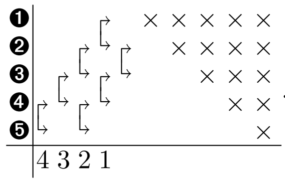


(13)

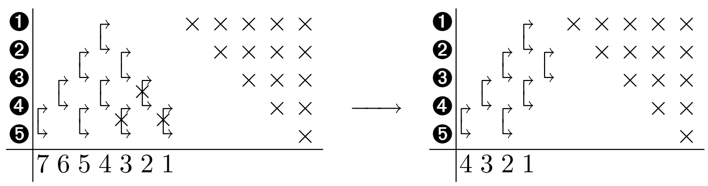
It is obvious that for retrieving a $\{p\}$-Hessenberg matrix, only $(n-p)(n-p-1)/2$ Givens similarity transformations need to be performed.

3.4 Unitary similarity transformation to generalized Hessenberg-like form

In the standard Hessenberg-like case, the Q-factor in the QR-factorization consists of an ascending sequence of Givens transformations. In the $\{p\}$-Hessenberg-like case this will be p sequences of ascending Givens transformations similarly as described in Section 3.2. Schematically, for the $p = 2$ case we get the following.



Again we have already all the ingredients necessary for constructing a unitary similarity transformation to this form. Removing the marked Givens transformations in the scheme below results in a generalized {2}-Hessenberg-like matrix.



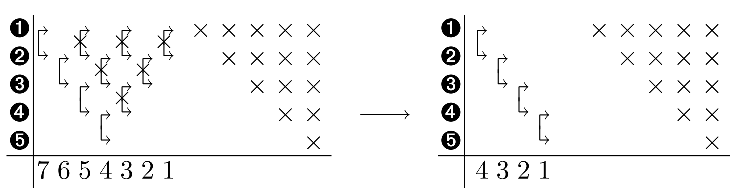
Remark 3 *The procedure for reducing a matrix to generalized Hessenberg-like form is clearly a simplification of the method for transforming the matrix to Hessenberg-like form. This is a big difference with the method proposed in [27]. In that article it is not at all obvious how to deduce a similarity transform to achieve the above structure.*

4 Inner, outer and other minimal annihilation schemes

Considering the results in the previous section, one might conclude that transforming matrices to (generalized) Hessenberg form is a sort of outer annihilation process, in which the outer Givens transformations of the unitary matrix Q are annihilated. Similarly the transformations to (generalized) Hessenberg-like form can be considered as an inner annihilation process. In this procedure the inner transformations of the unitary matrix Q are all annihilated.

In fact this is not entirely true. Whether the Givens transformations that need to be annihilated are outer or inner transformations, completely depends on the factorization of the unitary matrix Q. In Example 6 it was shown that one can factorize a unitary matrix in different ways. Consider now for example the QR-factorization of the matrix A, where the unitary matrix Q is in the $\vee$-form instead of the $\wedge$-pattern[7].

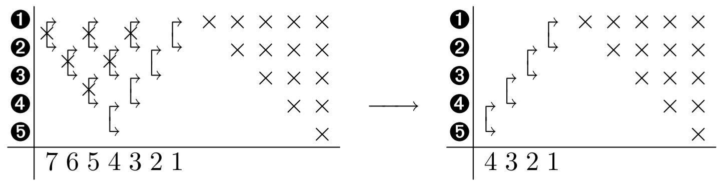
Based on this factorization, the reduction to (generalized) Hessenberg form becomes an inner annihilation process. Scheme (4) illustrates that the Givens transformations to be removed are now located in the middle of the factorization.



Similarly, the reduction to (generalized) Hessenberg-like form becomes an outer annihilation

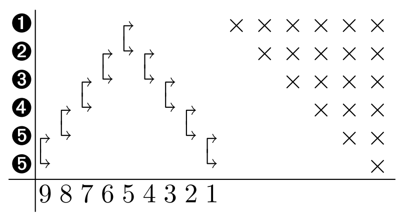
[7]Techniques exist for obtaining a QR-factorization in this form, either directly or by shift-through operations [29].

process as depicted in Scheme (4) for a reduction to Hessenberg-like form.



In Section 3 it was already noted that the above procedure reduces matrices to Hessenberg and Hessenberg-like form by the minimal number of Givens similarity transformations, each similarity transformation removing an extra element. In [10, 11, 8] it was stated how to construct the QR-factorization of a matrix in order to obtain a minimal number of Givens transformations. The fact that the unitary factor Q in the QR-factorization contains less Givens transformations has a positive impact on the complexity of the reduction to Hessenberg and Hessenberg-like form. Examples of articles discussing adapted unitary similarity transformations exploiting both rank and sparsity structures are [9, 19, 18, 13]. As a result the proposed algorithms are cheaper in complexity. Constructing, however, the minimal QR-factorization of these matrices, one can immediately note that the number of Givens transformations to be removed is essentially lower than for an arbitrary dense matrix. The idea of reducing these matrices to Hessenberg or Hessenberg-like form remains identical, simply removing the extra Givens transformations from the Q-factor in the QR-factorization will do the job. Hence, the algorithm presented here provides a unifying framework for reducing either structure to a compressed representation.

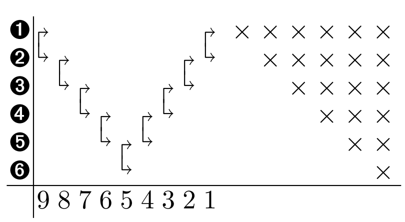
Example 7 (Semiseparable plus diagonal) *As an example, we will propose how to reduce a quasiseparable and semiseparable plus diagonal matrix to Hessenberg form*[8]*. These algorithms are proposed in [19, 13]. The QR-factorization of a quasiseparable matrix is of the following form ($\wedge$-pattern):*



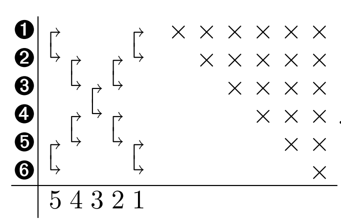
It is clear how one should reduce the matrix to Hessenberg or Hessenberg-like form, exploiting techniques from the previous sections. Removing the left leg of the $\wedge$ is always followed by removing one Givens transformation and then chasing the disturbance away, just like proposed in the articles [19, 13].

Example 8 (Semiseparable plus diagonal two-way reduction) *An extra feature of exploiting the QR-factorization for designing unitary similarity transformations to a particular structure is the fact that the different factorizations of the unitary matrix Q also lead to other variants of the reduction procedures.*

For example, reducing the semiseparable matrix to Hessenberg form via the $\wedge$-pattern leads to an outer annihilation scheme. Considering, however, the following two variants of the QR-factorization of the semiseparable plus diagonal matrix:



and



.

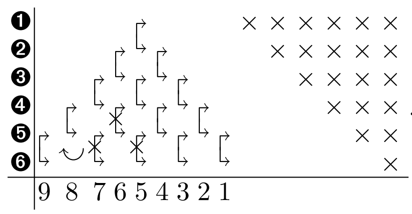
[8]There is only a small difference between semiseparable and quasiseparable matrices (see e.g. [29]). Their QR-factorization is, however, structurally identical.

The left scheme (∨-pattern) results in an inner annihilation process for obtaining a Hessenberg matrix. The right scheme (X-pattern) is a combination of both an inner and an outer annihilation process, interesting to remark is that one can start simultaneously by removing the lower left and the upper right leg of the X, which results in a two-way reduction to Hessenberg form. This can even be implemented in a parallel fashion.

5 Mixed reductions

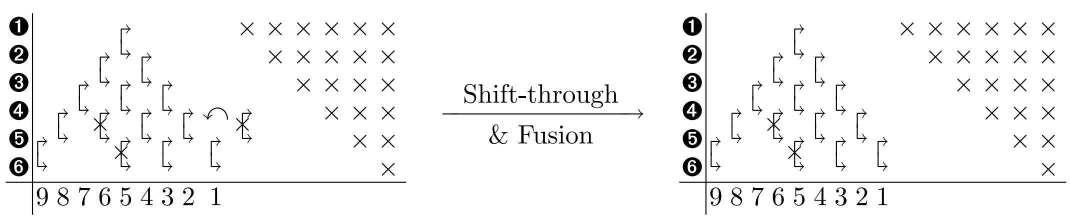
The reduction to a mixed structure, that is to the sum of a $\{p\}$-Hessenberg plus a $\{q\}$-Hessenberg-like matrix is a bit more complex since we have to remove Givens rotations appearing in the middle of the sequence of Givens transformations.

Let us describe the idea following a particular example for reducing a 6×6 matrix A to the sum of a Hessenberg-like and a Hessenberg matrix as appearing in Equation (5). In Scheme (14), the Givens transformations that need to be removed are marked.

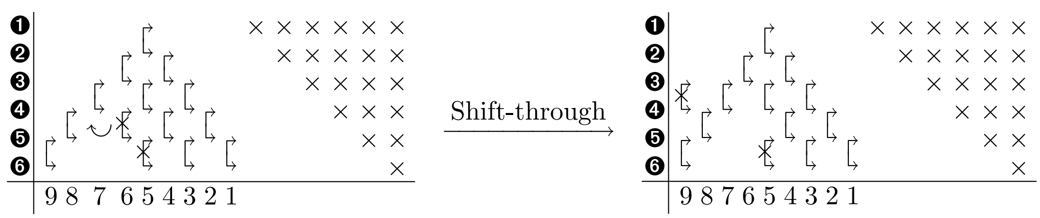


(14)

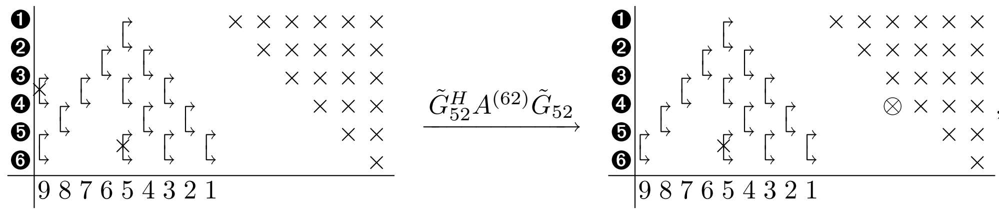
We have $A = QR = Q_1 G_{62} G_{52} G_{63} Q_4 Q_5 R$, where we denote by Q_1 the left sequence of ascending Givens rotations, and by Q_4 and Q_5 the two right most sequences of descending Givens. Once we have removed the three central Givens by unitary transformation we obtain the QR-factorization of a Hessenberg plus Hessenberg-like matrix, which consists of the Givens transformations in Q_1, Q_4 and Q_5. By means of a shift-trough operation (the operation is shown in Scheme (14)) we can remove G_{62} and have another Givens in front of the factorization, obtaining $A = \tilde{G}_{62} Q_1^{(62)} G_{52} G_{63} Q_4 Q_5 R$. Applying $\tilde{G}_{62}$ to transform A we get $A^{(62)} = \tilde{G}_{62}^H A \tilde{G}_{62}$, where $R\tilde{G}_{62}$ has a bulge in position $(5,4)$. Let us construct a Givens transformation to remove the bulge in position $(5,4)$ and recover the upper triangular structure of R.



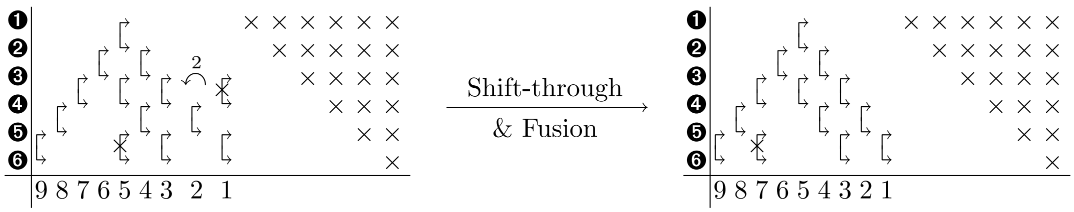
We have $A^{(62)} = Q_1^{(62)} G_{52} G_{63} Q_4^{(62)} Q_5^{(62)} R^{(62)}$. Next we will remove G_{52} similarly as was done for G_{63}. First the transformation is brought to the front by a shift-through operation.



$\tilde{G}_{52}$ denotes the leftmost Givens transformation and can be removed by a unitary similarity transformation. We have $A^{(62)} = \tilde{G}_{52} Q_1^{(52)} G_{63} Q_4^{(62)} Q_5^{(62)} R^{(62)}$.



and a bulge appears in position $(4, 3)$. To remove the bulge we need to combine rows 3 and 4 introducing a new Givens transformation to carry out this task.



We get $A^{(52)} = Q_1^{(52)} G_{63}^{(52)} Q_4^{(52)} Q_5^{(52)} R^{(52)}$. The procedure to remove the last Givens transformation $G_{63}^{(52)}$ is identical to that performed to remove G_{61} at the beginning, and consists of a shift-through operation to bring a Givens outside, a similarity transformation which produces a bulge in the factor R and a new Givens to remove the bulge, which we be fused in the last Givens of chain $Q_4^{(52)}$. At the end of this procedure we obtain the QR factorization of a Hessenberg plus rank-1 matrix as described in (5).

Remark 4 *The key idea for the reduction process is to remove Givens transformation through the fusion operation. Note that the shift-through operation allows to move up or down Givens rotations. In these schemes, where we have series of ascending or descending Givens transform, in order to apply a fusion we have to bring the undesired Givens transformation to the bottom or top two rows. Hence, if we have a Givens rotation that acts on rows $n-i-1, n-i$, we need a sequence of i descending Givens to move it down accordingly with Lemma 3, and one more Givens rotation acting on rows $n-1, n$ to be fused with. This is the reason way this scheme does not fit the reduction to semiseparable plus diagonal form, or in general the reduction to generalized Hessenberg-like plus diagonal matrices.*

6 Relations with Krylov subspaces

It is well know that Krylov subspaces are closely related to Hessenberg Matrices. In fact, in a similarity transformation to proper upper Hessenberg form, the leading columns of the unitary transforming matrix span a Krylov subspace. A complete description of properties and relationships between Krylov subspaces and Hessenberg matrices can be found in [33]. It turns out that the orthogonal transforming matrices are the Q factors of the QR factorization of the Krylov matrix constructed with starting vector $\mathbf{v}$, i.e. $K(A, \mathbf{v}) = [\mathbf{v}, A\mathbf{v}, A^2\mathbf{v}, \ldots, A^{n-1}\mathbf{v}]$. In particular, if $QR = K(A, \mathbf{v})$, then $Q^H A Q$ is an upper Hessenberg matrix. The proof, uses F, the companion matrix related to the characteristic polynomial of A showing that $AK = KF$. If $K = QR$, with simple considerations on the structure of F we get $Q^H A Q = R F R^{-1}$ which is in Hessenberg form. In [14] The Q factor of a nonsingular rational Krylov matrix was used as the transforming matrix to diagonal plus semiseparable matrix in the case A is Hermitian. In [2] this analysis has been further developed showing that with similar structural reasoning we can get other structures as the (generalized) Hessenberg-like or mixed forms. In particular, if we consider the reverse Krylov matrix, $K_r(A, \mathbf{v}) = K(A, \mathbf{v})\, J$ where J is the reverse permutation matrix, and we transform A using the Q factor of the QR factorization of K_r, we obtain the class of Hessenberg-like matrices.

In this section we will see how the QR-factorization can play an important role in proving and justifying many of the results related to Hessenberg and Hessenberg-like matrices.

6.1 Properness of Hessenberg and Hessenberg-like matrices

The definitions of properly Hessenberg and properly Hessenberg-like matrices are slightly different from each other. A matrix is said to be properly (or unreduced) Hessenberg if all of its subdiagonal elements are nonzero. A matrix is said to be properly (or unreduced) Hessenberg-like if none of the subdiagonal elements is zero and if none of the superdiagonal elements is includable in the lower triangular rank structure.

In fact these definitions find their origin in the QR-algorithm, they provide sort of easy verifiable criterion such that an implicit QR-step can run to completion and will theoretically not breakdown. The underlying idea is that in the QR-factorization of the matrix in consideration, the first $n-1$ columns of Q are uniquely defined, and therefore also the trailing column of Q is unique. Putting these constraints on a Hessenberg matrix gives the straightforward criterion were all the entries in the strictly lower triangular part need to be zero, for a Hessenberg-like matrix it is slightly more involved. Considering, however, the QR-factorization of these matrices, we obtain an almost identical requirement for both types of matrices. Considering the QR-factorization of a properly Hessenberg matrix H, the assumption that the subdiagonal entries are nonzero, implies that each of the Givens transformations for bringing the Hessenberg matrix to upper triangular form are uniquely defined[9] and different from the identity. As a result, the diagonal elements of the resulting upper triangular matrix R are all nonzero, except possibly for the trailing element r_{nn}. We then have the following necessary and sufficient condition for testing if a matrix is of proper upper Hessenberg form.

Lemma 4 *A Hessenberg matrix $H \in \mathbb{C}^{n\times n}$, having QR-factorization $H = QR$, is properly Hessenberg if and only if, the Givens transformations appearing in the factorization of the unitary matrix Q are different from the identity and the matrix $R(1:n-1, 1:n-1)$ is nonsingular.*

Let us simply convey this condition to the Hessenberg-like case.

Lemma 5 *A Hessenberg-like matrix $Z \in \mathbb{C}^{n\times n}$, having QR-factorization $Z = QR$, is properly Hessenberg-like if and only if, the Givens transformations appearing in the factorization of the unitary matrix Q are different from the identity and the matrix $R(1:n-1, 1:n-1)$ is nonsingular.*

Note that the nonsingularity of a Hessenberg or Hessenberg-like matrix does not guarantee properness. As an example, consider the identity matrix which is a non-proper Hessenberg matrix (and either an Hessenberg-like matrix) and is also invertible. However, in the case the matrix is nonsingular but not proper, it means that at least one of the Givens transformations is the identity matrix.

Moreover, if H is a nonsingular Hessenberg non-proper matrix with $h_{i+1,i} = 0$, that is

$$H = \begin{bmatrix} H_1 & H_3 \\ O & H_2 \end{bmatrix}$$

then $Z = H^{-1}$ is a non-proper Hessenberg-like matrix of the following from

$$Z = \begin{bmatrix} Z_1 & Z_3 \\ O & Z_2 \end{bmatrix} = \begin{bmatrix} H_1^{-1} & Z_3 \\ O & H_2^{-1} \end{bmatrix}.$$

On the contrary, assume Z is non-proper and nonsingular. As proven in [28], in the case Z is non-proper because of a superdiagonal entry includable in the lower triangular rank structure, then Z is singular, hence if Z is nonsingular then Z has a block of zeros in the lower triangular corner, and then $H = Z^{-1}$ is non-proper as well.

6.2 Unitary similarity transformations

Based on the lemmas above, we can provide an easy description of the unitary similarity transformations to Hessenberg and Hessenberg-like form, exploiting some properties from Krylov subspaces. Results on the uniqueness of the reduction procedures can be found in [2, 3] and an implicit

[9] Uniquely defined up to a unitary diagonal scaling.

Q-theorem for Hessenberg-like matrices is given in [30, 3]. The presentation here is simple, based on Krylov sequences and draws from [33].

Denote Krylov subspaces as follows $\mathcal{K}_k(A, \mathbf{v}) = \text{span}\{\mathbf{v}, A\mathbf{v}, A^2\mathbf{v}, \ldots, A^{k-1}\mathbf{v}\}$. The relationship between upper Hessenberg matrices and Krylov subspaces is shown in [33]. In the same way, we can prove that the same relationship holds for Hessenberg-like matrices when one considers rational Krylov subspaces, that is the Krylov subspace related to A^{-1}.

$$\mathcal{K}_k(A^{-1}, \mathbf{v}) = \text{span}\{\mathbf{v}, A^{-1}\mathbf{v}, A^{-2}\mathbf{v}, \ldots, A^{-(k-1)}\mathbf{v}\}.$$

We have the following Theorem, homologous of Theorem 3.3.2 which proves that staring from a nonsingular matrix A, all the unitary transformations to Hessenberg-like form, are the Q-factor of a rational Krylov matrix. of [33].

Theorem 6 *Let $A \in \mathbb{C}^{n\times n}$ and $Z \in \mathbb{C}^{n\times n}$ be nonsingular matrices, and let $G \in \mathbb{C}^{n\times n}$ be a unitary matrix , such that $Z = G^H AG$. Let $\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_n$ be the columns of G. Let $\mathbf{v}$ be any vector proportional to $\mathbf{g}_1$.*

a) If Z is a non proper Hessenberg-like matrix, then there is a index k such that $z(i, j) = 0,\ i \geq k,\ 1 \leq j \leq k$. We then have

$$\text{span}\{\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g_j}\} = \mathcal{K}_j(A^{-1}, \mathbf{v}), \quad j = 1, \ldots, k.$$

b) If Z is a proper Hessenberg-like,

$$\text{span}\{\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g_j}\} = \mathcal{K}_j(A^{-1}, \mathbf{v}), \quad j = 1, \ldots, n. \tag{15}$$

In this case $\mathcal{K}_n(A^{-1}, \mathbf{v})$ has dimension n.

c) Conversely, if (15) holds, Z is Hessenberg-like.

Proof. To prove the Theorem, we just apply Theorem 3.3.2 of [33] to matrix A^{-1}. As pointed out in Section 6.1, any non-proper nonsingular Hessenberg-like matrix is the inverse of a non-proper Hessenberg matrix and viceversa.

A more elegant proof can be obtained using the QR factorization as follows. Let $A = QR$ and $Z = \hat{Q}\hat{R}$. Providing $\mathbf{v}$ proportional to $\mathbf{g}_1$, and knowing that $\hat{Q}$ is a lower unitary Hessenberg matrix, we have

$$\begin{aligned}\mathcal{K}_j(A^{-1}, \mathbf{v}) &= \text{span}\{\mathbf{v}, G\,(\hat{Q}\hat{R})^{-1}G^H\mathbf{v}, G\,(\hat{Q}\hat{R})^{-2}G^H\mathbf{v}, \ldots, G\,(\hat{Q}\hat{R})^{-j+1}G^H\mathbf{v}\}\\ &= \text{span}\{\mathbf{v}, G\,(\hat{Q}\hat{R})^{-1}\mathbf{e}_1, G\,(\hat{Q}\hat{R})^{-2}\mathbf{e}_1, \ldots, G\,(\hat{Q}\hat{R})^{-j+1}\mathbf{e}_1\}\\ &= \text{span}\{\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_j\}.\end{aligned}$$

If Z is a proper Hessenberg-like the equality holds for each $j = 1, \ldots, n$, otherwise it holds only up to the first index k such that $\hat{r}_{kk} = 0$.

Remark 5 *In the case A is singular however we cannot compute A^{-1} and the rational Krylov space is not defined. Let Z be a singular, proper Hessenberg-like matrix, and let $G \in \mathbb{C}^{n\times n}$ the unitary matrix such that $Z = G^H A G$. If we perturb Z in the entry (n, n), $\tilde{Z} = Z + \epsilon \mathbf{e}_n\mathbf{e}_n^T$, we get $\tilde{A} = G\,\tilde{Z}\,G^H$ which is nonsingular. We can apply Theorem 6 to $\tilde{A}$ and $\tilde{Z}$, obtaining that*

$$\text{span}\{\mathbf{g}_1, \ldots, \mathbf{g}_j\} = \mathcal{K}_j(\tilde{A}^{-1}, \mathbf{v}), \quad j = 1, \ldots, n,$$

providing $\mathbf{v}$ proportional to $\mathbf{g}_1$. If Z is a not proper Hessenberg-like singular matrix, we have to perturb the entry (k, k) where k is the index of properness[10]. In this case the first j columns of G span the space $\mathcal{K}_j(\tilde{A}^{-1}, \mathbf{v})$, for $j = 1, \ldots, k$.

[10] That is the first index such that $z(i, j) = 0$, for $i \geq k$ and $1 \leq j \leq k$.

It is possible to relate the columns of the transformation to Hessenberg-like form also to the standard Krylov subspace, as underlined by the following theorem.

Theorem 7 *Let $A \in \mathbb{C}^{n\times n}$ be a nonderogatory matrix, and let $\mathbf{v} \in \mathbb{C}^n$ be such that the Krylov matrix $K(A, \mathbf{v}) = [\mathbf{v}, A\mathbf{v}, \ldots, A^{n-1}\mathbf{v}]$ is nonsingular. If $G \in \mathbb{C}^{n\times n}$ is the Q factor of the QR-factorization of $K(A, \mathbf{v})\, J$, where J is the reversion matrix, then $Z = G^H AG$ is an Hessenberg-like matrix. As a consequence*

$$\operatorname{span}\{\mathbf{g}_1, \ldots, \mathbf{g}_j\} = \operatorname{span}\{A^{n-1}\mathbf{v}, \ldots, A^{n-j}\mathbf{v}\}, \quad j = 1, \ldots, n-1.$$

Proof. Since A is nonderogatory the existence of vectors $\mathbf{v}$ such that the matrix $K(a, \mathbf{v})$ is nonsingular is guaranteed. As proved in [2] $Z = G^H AG$ is Hessenberg-like. The relation between the Krylov subspaces and the space spanned by the columns of G follows from the equality $KJ = GR$. Note that if A is nonsingular, we have the same result obtained in the previous theorem, just taking $\mathbf{v}$ as the solution of $A^{n-1}\mathbf{v} = \mathbf{g}_1$.

7 Ritz-values and the new unitary similarity transforms

It is already known [14, 7] that semiseparable matrices have a close relation with rational Krylov sequences, and we have presented similar results also for Hessenberglike matrices. Recently some results on the convergence of rational Ritz values were given based on potential theoretical results [1, 6]. There are however some problems related to the computational aspects with rational Krylov, most important: solving systems of equations with some shifted matrices is not only time consuming, but can also lead to ill-conditioning in case the shifts are getting close to eigenvalues [21].

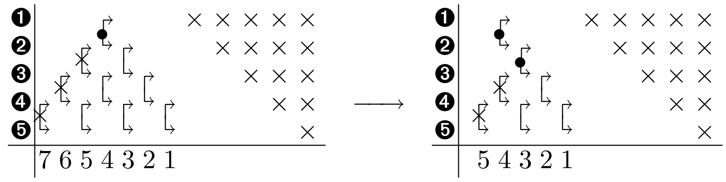
In this section, we will prove that the procedure based on unitary similarity transforms for bringing the matrix A to Hessenberg-like form inherits the rational Krylov behavior, i.e. that the eigenvalues in that part of the matrix already in Hessenberg-like form are rational Ritz-values. The standard reductions to Hessenberg-like [32] and Hessenberg form, have, on the contrary, the Ritz-values in the part already of Hessenberg or Hessenberg-like form.

7.1 Part of the matrix in the desired structure

When transforming a matrix to Hessenberg form, each Householder transformation (or some successive Givens transformations) brings a column to Hessenberg form and in fact the part of the matrix already in Hessenberg form grows at each step of the process. This part will have as eigenvalues the Ritz-values.

Let us clarify which parts of the matrices during the unitary similarity transformations proposed in this article are already of the desired structure. We will indicate, by graphical schemes, which Givens transformations remain unaltered during the reduction process. Since they will not alter anymore, these transformations make up the part of the matrix in final form. In the forthcoming schemes, the Givens transformations that will not alter anymore are marked with a $\bullet$, this means that they will not be affected by any other fusions, shift-throughs and so forth. The Givens transformations to be removed in the first step are marked with a $\times$.

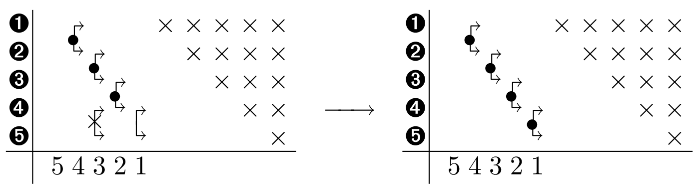
We start with the reduction to Hessenberg form. In the first phase of the reduction process three Givens transformations are removed (marked with $\times$), this corresponds to bringing the first column of the matrix A in Hessenberg form. Important to note is that even in the removal process of these three Givens transformations, the top transformation will not change. This implies that the upper left element of the matrix a_{11} was already in the correct Hessenberg form before performing the starting similarity transformations. The result is the matrix $A^{(31)}$ depicted in the right of Scheme (16), another Givens transformation is now in final form and will not change when removing the next two Givens transformations, marked by $\times$.



(16)

Denote the global reduction to Hessenberg form with $V^H AV = H$, these schemes indicate the following equalities[11]: for the left scheme we get that $a_{11} = \mathbf{e}_1^T A\mathbf{e}_1 = \mathbf{e}_1^T H\mathbf{e}_1$ and for the right scheme we get that $A^{(31)}(1:2,1:2) = [\mathbf{e}_1,\mathbf{e}_2]^T A^{(31)}[\mathbf{e}_1,\mathbf{e}_2] = [\mathbf{e}_1,\mathbf{e}_2]^T V^H AV[\mathbf{e}_1,\mathbf{e}_2] = [\mathbf{e}_1,\mathbf{e}_2]^T H[\mathbf{e}_1,\mathbf{e}_2]$ are already of Hessenberg form. More precisely we get $a_{11} = h_{11}$ and $A^{(31)}(1:2,1:2) = H(1:2,1:2)$.

Remove now the two Givens transformations marked in the right of Scheme (16), we get the left of Scheme (17).



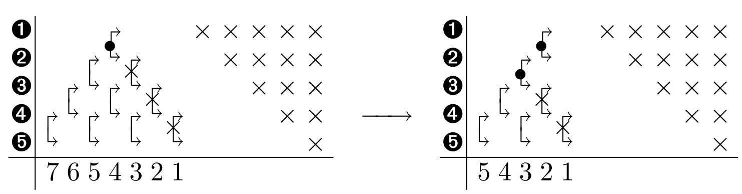
(17)

Mathematically the left of Scheme (17) (the first two columns are already brought to Hessenberg form) is equivalent to $A^{(42)}(1:3,1:3) = [\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3]^T A^{(42)}[\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3] = [\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3]^T H[\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3] = H(1:3,1:3)$ and the right scheme (the first three columns are brought to Hessenberg form) implies

$$\begin{aligned} A^{(53)}(1:4,1:4) &= [\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3,\mathbf{e}_4]^T A^{(53)}[\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3,\mathbf{e}_4] \\ &= [\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3,\mathbf{e}_4]^T V^H AV[\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3,\mathbf{e}_4] = H(1:4,1:4). \end{aligned}$$

Since $A^{(53)} = H$, we get full equality here.

For the reduction to Hessenberg-like form, we get similar results[12]. Let us illustrate compactly which parts of the matrix will reach final form after a number of Givens unitary similarity transforms. Assume that in the end we have that $V^H AV = Z$ is in Hessenberg-like form.



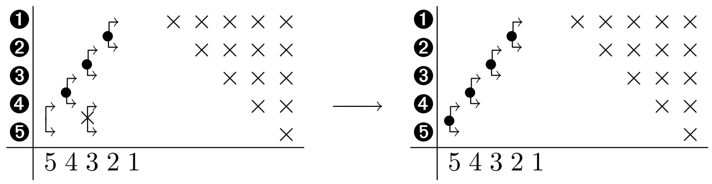
(18)

When starting the reduction procedure (left of Scheme (18)) the upper left part of A is already in final form, this means that the part of $\mathbf{e}_1^T V^H AV\mathbf{e}_1$ is fixed throughout the remainder of the procedure. In the right of Scheme (18) we note that the upper left 2×2 subblock of the matrix corresponds to $[\mathbf{e}_1,\mathbf{e}_2]^T V^H AV[\mathbf{e}_1,\mathbf{e}_2]$, is of the desired Hessenberg-like structure and will not alter anymore.

[11] In the procedures discussed above $V\mathbf{e}_1 = \mathbf{e}_1$, an arbitrary initial similarity transformation to obtain a generic $V\mathbf{e}_1$ is possible and does not complicate matters. It coincides with a different starting vector for obtaining the Ritz-values.

[12] Remark that this is also different from [27], in this article the Hessenberg-like part of the matrix keeps changing due to a chasing procedure and reaches only the final form after the last Givens transformation is performed.

In the following two schemes clearly the part that will remain unaltered grows



In the forthcoming subsection we will prove that the part of the matrix in Hessenberg-like form will have as eigenvalues the rational Ritz-values.

7.2 Relation with the Ritz-values

The Ritz-values of a matrix A are defined as the eigenvalues of the projected counterpart $V^H AV$, where the columns of $V = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k]$ form an orthogonal basis for the Krylov subspace $\mathcal{K}_k(A, \mathbf{v}_1) = \text{span}\{\mathbf{v}_1, A\mathbf{v}_1, A^2\mathbf{v}_1, \ldots, A^{k-1}\mathbf{v}_1\}$, i.e. that

$$\text{span}\{\mathbf{v}_1, A\mathbf{v}_1, A^2\mathbf{v}_1, \ldots, A^{k-1}\mathbf{v}_1\} = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}.$$

For rational Ritz-values (without shifts) we consider the Krylov subspace $\mathcal{K}_k(A^{-1}, \mathbf{v})$. In [16, 17] the convergence of Ritz-values to eigenvalues is investigated, it is proven that under mild conditions (on e.g. the starting vector, the distribution of the eigenvalues) the Ritz-values approximate those eigenvalues well-separated from the rest of the spectrum first. In some circumstances these are not the eigenvalues one wants to find first, hence one applies some shift and invert techniques resulting in Rational-Ritz values.

When transforming a matrix to Hessenberg form, we know that the part of the matrix already in Hessenberg form has as eigenvalues the Ritz-values. We will show now that in the reduction to Hessenberg-like form, the part of the matrix already in Hessenberg-like form will have as eigenvalues the Rational-Ritz values.

Let us first consider the unitary similarity transformation to Hessenberg form. Assume we have a matrix with its QR-factorization $A = QR$. As proved in Section 3.1 the reduction to Hessenberg form coincides with a transformation of the QR-factorization to the form $H = V^H AV = \hat{Q}\hat{R}$, where $\hat{Q}$ is a unitary Hessenberg and $\hat{R}$ is upper triangular. Assume that the unitary Hessenberg matrix[13] is proper and hence the upper triangular matrix $R(1 : n-1, 1 : n-1)$ is nonsingular (see Section 6.1). In this case we see that the part already in Hessenberg form corresponds to $[\mathbf{e}_1, \ldots, \mathbf{e}_k]^T V^H AV[\mathbf{e}_1, \ldots, \mathbf{e}_k] = [\mathbf{v}_1, \ldots, \mathbf{v}_k]^H A[\mathbf{v}_1, \ldots, \mathbf{v}_k]$. As proven in [33] the vectors $\mathbf{v}_i$ $(1 \leq i \leq k)$ span the Krylov space $\mathcal{K}_k(A, \mathbf{v}_1)$, that is

$$\mathcal{K}_k(A, \mathbf{v}_1) = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}.$$

Consider now the reduction to Hessenberg-like form. We start with the QR-factorization of the matrix $A = QR$. The reduction to Hessenberg-like form results in a factorization of $V^H AV = \hat{Q}\hat{R}$, where $\hat{Q}$ is now a lower unitary Hessenberg matrix, and $\hat{R}$ is still upper triangular. Assume the unitary matrix to be irreducible and the upper triangular matrix to be nonsingular.

We see now that the part of the matrix already in final Hessenberg-like form $[\mathbf{v}_1, \ldots, \mathbf{v}_k]^H A[\mathbf{v}_1, \ldots, \mathbf{v}_k]$ has the rational Ritz values since from Theorem 6 we have

$$\mathcal{K}_k(A^{-1}, \mathbf{v}_1) = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}.$$

As already stressed in Section 6.2, the properness of the Hessenberg-like matrix does not imply necessarily nonsingularity of the matrix R. In case of a singular R, however, computing A^{-1} is not possible, and rational Krylov does not exist. The fact that R might be singular does, however, not pose any constraints on the unitary similarity transformation to Hessenberg-like form. The convergence behavior resembles still the convergence behavior of Rational Krylov sequences (see Remark 5).

[13] Otherwise we have a reducible Hessenberg matrix and then we need a suitable vector for reconstructing the corresponding Krylov sequence.

8 Numerical example related to the Rational Ritz-value convergence

In this section we will briefly illustrate the rational Ritz-value convergence behavior of the reduction to Hessenberg-like form. We remark that the unitary similarity transformation as presented in [27] did not inherit this behavior, but inherited the standard Ritz-value convergence.

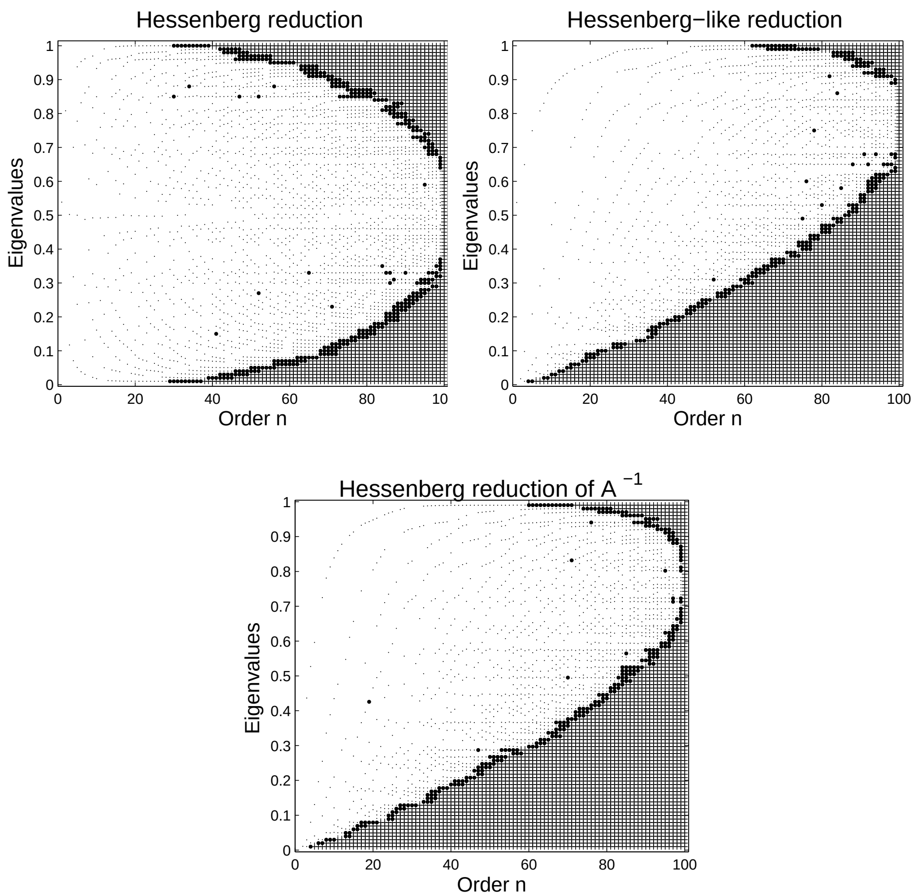


Figure 1: Convergence plots of Ritz-values

Example 9 *In this example we will plot Ritz-value behavior for a symmetric matrix having* 100 *eigenvalues equally spaced in the interval* $[1/n, 1]$*. In the Figures 1, the horizontal axis denotes the order of the part of the matrix already in Hessenberg or Hessenberg-like form. The vertical axis denotes the range of the eigenvalues of the original matrix. In the figures a small dot is depicted if a Ritz-value approximates an eigenvalue in the range* $[10^{-2.5}, 10^{-5}]$*, a bigger dot is depicted if the approximation lies within* $[10^{-5}, 10^{-7.5}]$ *and a plus sign is plotted if the approximation is better than* $10^{-7.5}$.

Three convergence plots of the Ritz-values are shown in Figure 1. The leftmost figure shows the standard Ritz-value convergence behavior, the rightmost the convergence related to the reduction to Hessenberg-like form and one can see that this convergence pattern is almost identitical to the pattern received when performing the Hessenberg reduction on A^{-1}*, where convergence of the inverse of the eigenvalues is plotted towards the eigenvalues of* A*.*

It can be seen from the figures that the convergence to small eigenvalues is faster when reducing a matrix to Hessenberg-like form, rather than Hessenberg form where we get the same kind of *convergence for the extreme eigenvalues.*

9 Conclusions and future research

In this paper a new alternative approach was discussed for constructing unitary similarity transformations to matrices representable in compact form. The proposed approach unifies the reductions to Hessenberg-like and Hessenberg form. With respect to the initial reduction procedure proposed in [27], there are two main differences. First, the new procedure uses less Givens transformations and is closer to the reduction to Hessenberg form. Second, the new reduction procedure inherits the rational Krylov convergence behavior, whereas the old procedure has the traditional Krylov convergence behavior, with respect to the Ritz values.

It was proven in e.g. [14], that Rational Krylov with shifts has a close relation with semiseparable plus diagonal matrices. Unfortunately the semiseparable plus diagonal case does not fit in properly in the framework provided in this article. Future research will focus on this aspect and search for an alternative unitary similarity transformation to semiseparable plus diagonal form, inheriting the rational Krylov behavior, with shifts.

References

[1] Beckermann, B., Güttel, S., Vandebril, R.: On the convergence of rational Ritz-values. SIAM Journal on Matrix Analysis and Applications **31**, 1740–1774 (2010)

[2] Bevilacqua, R., Bozzo, E., Del Corso, G.M.: Transformations to rank structures by unitary similarity. Linear Algebra and its Applications **402**, 126–134 (2005)

[3] Bevilacqua, R., Del Corso, G.M.: Structural properties of matrix unitary reduction to semiseparable form. Calcolo **41**(4), 177–202 (2004)

[4] Bindel, D., Demmel, J.W., Kahan, W., Marques, O.A.: On computing Givens rotations reliably and efficiently. ACM Transactions on Mathematical Software **28**(2), 206–238 (2002)

[5] Bultheel, A., Van Barel, M.: Linear Algebra, Rational Approximation and Orthogonal Polynomials, *Studies in computational mathematics*, vol. 6. North-Holland, Elsevier Science B.V., Amsterdam, Netherlands (1997)

[6] Chesnokov, A., Deckers, K., Van Barel, M.: A numerical solution of the constrained weighted energy problem. Journal of Computational and Applied Mathematics (2010). Accepted for publication

[7] Deckers, K., Bultheel, A.: Rational Krylov sequences and orthogonal rational functions. Poster presented at conference (2008)

[8] Delvaux, S.: Rank structured matrices. Ph.D. thesis, Department of Computer Science, Katholieke Universiteit Leuven, Celestijnenlaan 200A, 3000 Leuven (Heverlee), Belgium (2007)

[9] Delvaux, S., Van Barel, M.: A Hessenberg reduction algorithm for rank structured matrices. SIAM Journal on Matrix Analysis and Applications **29**(3), 895–926 (2007)

[10] Delvaux, S., Van Barel, M.: A *QR*-based solver for rank structured matrices. SIAM Journal on Matrix Analysis and Applications **30**(2), 464–490 (2008)

[11] Delvaux, S., Van Barel, M.: Unitary rank structured matrices. Journal of Computational and Applied Mathematics **215**(1), 268–287 (2008)

[12] Demmel, J.W.: Applied Numerical Linear Algebra. SIAM, Philadelphia, Pennsylvania, USA (1997)

[13] Eidelman, Y., Gemignani, L., Gohberg, I.C.: On the fast reduction of a quasiseparable matrix to Hessenberg and tridiagonal forms. Linear Algebra and its Applications **420**(1), 86–101 (2007)

[14] Fasino, D.: Rational Krylov matrices and QR-steps on Hermitian diagonal-plus-semiseparable matrices. Numerical Linear Algebra with Applications **12**(8), 743–754 (2005)

[15] Golub, G.H., Van Loan, C.F.: Matrix Computations, third edn. Johns Hopkins University Press, Baltimore, Maryland, USA (1996)

[16] Kuijlaars, A.B.J.: Which eigenvalues are found by the Lanczos method? SIAM Journal on Matrix Analysis and Applications **22**(1), 306–321 (2000)

[17] Kuijlaars, A.B.J.: Convergence analysis of Krylov subspace iterations with methods from potential theory. SIAM Review **48**(1), 3–40 (2006)

[18] Lemmerling, P., Mastronardi, N., Van Huffel, S.: Fast algorithms for structured total least squares in speech compression (2001). In english

[19] Mastronardi, N., Chandrasekaran, S., Van Huffel, S.: Fast and stable reduction of diagonal plus semi-separable matrices to tridiagonal and bidiagonal form. BIT **41**(1), 149–157 (2003)

[20] Mastronardi, N., Schuermans, M., Van Barel, M., Vandebril, R., Van Huffel, S.: A Lanczos-like reduction of symmetric structured matrices into semiseparable ones. Calcolo **42**(3–4), 227–241 (2005)

[21] Meerbergen, K.: Changing poles in the rational Lanczos method for the hermitian eigenvalue problem. Numerical Linear Algebra with Applications **8**(1), 33–52 (2001)

[22] Parlett, B.N.: The Symmetric Eigenvalue Problem, *Classics in Applied Mathematics*, vol. 20. SIAM, Philadelphia, Pennsylvania, USA (1998)

[23] Saad, Y.: Numerical Methods for Large Eigenvalue Problems. Manchester University Press, Manchester, United Kingdom (1992)

[24] Saad, Y.: Iterative Methods for Sparse Linear Systems, second edn. SIAM, Philadelphia, Pennsylvania, USA (2003)

[25] Van Barel, M., Fasino, D., Gemignani, L., Mastronardi, N.: Orthogonal rational functions and diagonal plus semiseparable matrices. In: F.T. Luk (ed.) Advanced Signal Processing Algorithms, Architectures, and Implementations XII, *Proceedings of SPIE, Bellingham, Washington, USA*, vol. 4791, pp. 167–170 (2002)

[26] Van Barel, M., Fasino, D., Gemignani, L., Mastronardi, N.: Orthogonal rational functions and structured matrices. SIAM Journal on Matrix Analysis and Applications **26**(3), 810–829 (2005)

[27] Van Barel, M., Vandebril, R., Mastronardi, N.: An orthogonal similarity reduction of a matrix into semiseparable form. SIAM Journal on Matrix Analysis and Applications **27**(1), 176–197 (2005)

[28] Van Barel, M., Vandebril, R., Mastronardi, N.: An Implicit Q-theorem for Hessenberg-like matrices Mediterr. J. Math. **2**, 259–275 (2005)

[29] Vandebril, R., Van Barel, M., Mastronardi, N.: Matrix Computations and Semiseparable Matrices, Volume I: Linear Systems. Johns Hopkins University Press, Baltimore, Maryland, USA (2008)

[30] Vandebril, R., Van Barel, M., Mastronardi, N.: Matrix Computations and Semiseparable Matrices, Volume II: Eigenvalue and Singular Value Methods. Johns Hopkins University Press (2008)

[31] Vandebril, R., Van Barel, M., Mastronardi, N.: A parallel QR-factorization/solver of structured rank matrices. Electronic Transactions on Numerical Analysis **30**, 144–167 (2008)

[32] Vandebril, R., Van Camp, E., Van Barel, M., Mastronardi, N.: On the convergence properties of the orthogonal similarity transformations to tridiagonal and semiseparable (plus diagonal) form. Numerische Mathematik **104**, 205–239 (2006)

[33] Watkins, D.S.: The Matrix Eigenvalue Problem: GR and Krylov Subspace Methods. SIAM, Philadelphia, Pennsylvania, USA (2007)

[34] Wilkinson, J.H.: The Algebraic Eigenvalue Problem. Numerical Mathematics and Scientific Computation. Oxford University Press, New York, USA (1999)