A stable cubically convergent GR algorithm and Krylov subspace methods for non-Hermitian matrix eigenvalue problems

Dissertation

der Mathematischen Fakultät

der Eberhard-Karls-Universität Tübingen

zur Erlangung des Grades eines Doktors

der Naturwissenschaften

vorgelegt von

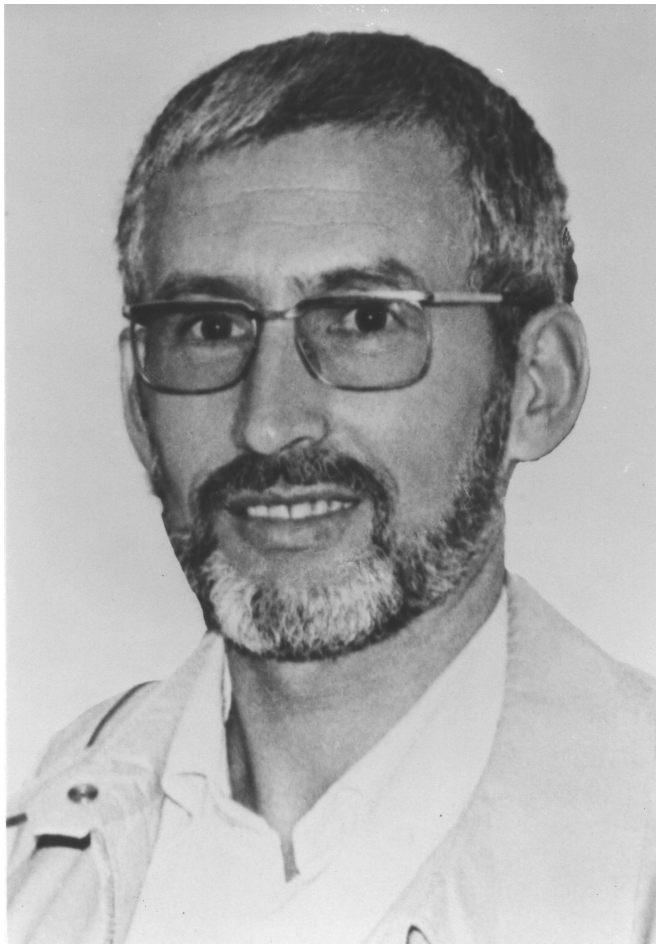
Markus Ziegler

aus

Rosenfeld

2001

Tag der mündlichen Prüfung:	11. Juli 2001
Dekan:	Prof. Dr. Ch. Lubich
Erster Berichterstatter:	Prof. Dr. K.P. Hadeler
Zweiter Berichterstatter:	Prof. Dr. M. Gutknecht



This thesis is dedicated to the memory of my father

Ernst Ziegler

(29th December 1936 - 14th December 1981)

I express my deep gratitude to my wife for her understanding and support during all the years in which I taught at school and spent every free minute working on this thesis.

Special thanks also go to Prof. Hadeler for all his support and advice and the pleasant working atmosphere he helped to create.

Zusammenfassung

In dieser Dissertation werden Krylov-Verfahren und „restarted Krylov-Verfahren“ zur Eigenwertberechnung von beliebigen Matrizen untersucht. Es wird gezeigt, dass das allgemeine restarted Krylov-Verfahren mathematisch äquivalent zum allgemeinen Zerlegungsalgorithmus (GR-Algorithmus) ist. Es wird bewiesen, dass ein spezielles restarted look-ahead-Lanczos-Verfahren, angewandt auf Matrizen mit paarweise verschiedenen Eigenwerten, unter sehr schwachen Voraussetzungen kubisch konvergiert (IRL-Verfahren). Ausgehend von diesen Ergebnissen wird ein neues, numerisch stabiles GR-Verfahren zur Eigenwertberechnung von beliebigen Matrizen entwickelt (LGR-Verfahren). Dieses Verfahren ist mathematisch äquivalent zu dem IRL-Verfahren. Folglich konvergiert das LGR-Verfahren, angewandt auf Matrizen mit paarweise verschiedenen Eigenwerten, unter sehr schwachen Voraussetzungen ebenfalls kubisch. Man beachte, dass der QR-Algorithmus und alle anderen numerisch stabilen GR-Verfahren (die dem Autor bekannt sind) i.a. quadratisch konvergieren, wenn sie auf beliebige Matrizen mit paarweise verschiedenen Eigenwerten angewandt werden. Schließlich wird gezeigt, dass das LGR-Verfahren i.a.

$$\frac{8}{3}N^3 + O(N^2)$$

Gleitkomma-Operationen für die Berechnung aller Eigenwerte von vollbesetzten $N \times N$ Matrizen und

$$O(N^2)$$

Gleitkomma-Operationen für die Bestimmung aller Eigenwerte von Tridiagonalmatrizen benötigt.
Wir erinnern daran, dass der (implizite) QR-Algorithmus

$$10N^3 + O(N^2)$$

Gleitkomma-Operationen für die Bestimmung aller Eigenwerte von vollbesetzten Matrizen und

$$\frac{20}{3}N^3 + O(N^2)$$

Gleitkomma-Operationen für die Berechnung aller Eigenwerte von Tridiagonalmatrizen und Hessenberg-Matrizen braucht.
Numerische Beispiele (28 Matrizen mit $N \leq 250$, IEEE-Arithmetik „double“) schließen die Arbeit ab. Bei der Durchführung dieser Beispiele wurden folgende Beobachtungen gemacht:

Konvergenz

In allen Beispielen konvergiert das LGR-Verfahren kubisch und der QR-Algorithmus quadratisch.

Gleitkomma-Operationen

Für vollbesetzte Matrizen und obere Hessenberg-Matrizen (mit $N \leq 100$) benötigt das LGR-Verfahren im Durchschnitt halb so viele Gleitkomma-Operationen wie der QR-Algorithmus.
Im Falle von Tridiagonalmatrizen braucht das LGR-Verfahren
für $N = 50$ ungefähr $1/5$ der Gleitkomma-Operationen wie das QR Verfahren,
für $N = 100$ ungefähr $1/10$ der Gleitkomma-Operationen wie das QR Verfahren und
für $N = 200$ ungefähr $1/20$ der Gleitkomma-Operationen wie das QR Verfahren.

Approximationsfehler

Das LGR-Verfahren liefert im Durchschnitt Eigenwertnäherungen mit einem maximalen relativen Fehler von $2 \cdot 10^{-8}$.
Das QR-Verfahren erzeugt im Durchschnitt Eigenwertnäherungen mit einem maximalen relativen Fehler von $3 \cdot 10^{-14}$.

Contents

1 Introduction

In 1950, Lanczos [51] proposed a method for the reduction of a given, in general non-Hermitian, $N \times N$ matrix A to tridiagonal form. The Lanczos algorithm generates in step j a $j \times j$ tridiagonal matrix $H^{(j)}$ and a biorthogonal basis of the Krylov subspaces $\mathcal{K}(j, A^T, w) = \text{span}\{w, A^T w, \ldots, (A^T)^{(j-1)} w\}$ and $\mathcal{K}(j, A, v) = \text{span}\{v, Av, \ldots, A^{(j-1)} v\}$ where v and w are the given starting vectors. Lanczos recognized that this algorithm can break down if it is applied to nonsymmetric matrices, i.e., division by 0 may occur before a basis of an A or A^T invariant Krylov subspace is constructed. If no breakdown occurs then the Lanczos method terminates after at most $s \leq N$ steps with a tridiagonal matrix $H^{(s)}$ that represents the restriction of A or A^T to the A-invariant subspace $\mathcal{K}(s, A, v)$ or the A^T-invariant subspace $\mathcal{K}(s, A^T, w)$, respectively. In particular, all eigenvalues of $H^{(s)}$ are eigenvalues of A. The possibility of breakdowns and numerical instabilities has brought the nonsymmetric Lanczos method into discredit, and it received little attention until the 1980s.

Several attempts have been made to develop algorithms for tridiagonalization by elimination. In 1954, Givens [25] presented a method for reducing real symmetric matrices to tridiagonal form by plane rotations. Five years later Householder and Bauer [41] studied a method for the reduction of symmetric matrices to tridiagonal form using Householder transformations $I - xx^T$. Bauer [3] (in 1959) and La Budde [50] (in 1963) introduced methods for the reduction of arbitrary matrices to tridiagonal form using elementary matrices $I - \gamma xy^T$. The transformation method of Strachey and Francis [71], suggested in 1961, first reduces the matrix A to a lower Hessenberg matrix H by orthogonal transformations. Then the matrix H is reduced to tridiagonal form via Gauss transformations.

We point out that the reduction of an arbitrary given square matrix to tridiagonal form is in exact arithmetic always possible (see, e.g.,[24], [61], [42], [9]). But all known methods, that transform general matrices to tridiagonal form, are in a certain sense mathematically equivalent to the nonsymmetric Lanczos method (see Parlett [58]). Consequently, breakdowns (division by zero) and near breakdowns (divison by very small numbers) of these algorithms occur, which leads to numerical instabilities.
Such breakdowns can be avoided if we do not transform to tridiagonal form but to Hessenberg form. This can be done with unitary transformations (e.g., Householder transformations [28]) or non-unitary transformations (e.g., Gauss transformations with pivoting [28]).
In 1951, Arnoldi [1] studied a method which generates, when applied to a general $N \times N$ matrix A, in step j a $j \times j$ upper Hessenberg matrix $H^{(j)}$ and an orthogonal basis of the Krylov subspace $\mathcal{K}(j, A, v)$. Here v is the given starting vector. The Arnoldi method terminates with an upper Hessenberg matrix $H^{(s)}$ which represents (with respect to the computed orthogonal basis) the restriction of A to the A-invariant subspace $\mathcal{K}(s, A, v)$. If

$s = N$, then the Arnoldi method is mathematically equivalent to a unitary transformation to upper Hessenberg form.

In 1954, Rutishauser [62] presented his quotient-difference (QD) algorithm for computing poles of various classes of functions. Furthermore he considered the QD algorithm in its application to the function

$$f(\lambda) = \langle w, (\lambda I - A)^{-1} v \rangle = \sum_{i=0}^{\infty} \frac{\langle w, A^i v \rangle}{\lambda^{i+1}}$$

and found that the data of the QD algorithm can be organized in products of bidiagonal matrices. This led to his famous LR algorithm [63]. Rutishauser [62] and Henrici [37] further recognized that the QD scheme can also be obtained by restarting the Lanczos method (for more details see [56]).
The LR algorithm can break down (division by zero). To avoid breakdowns and near breakdowns Wilkinson suggested in [79] an LR algorithm with pivoting.
In 1961, Francis [19] and Kublanovskaya [49] presented the famous QR method which is the unitary analogue to the LR algorithm. In the following years the implicitly shifted QR method ([79], [28]) became the state of the art as a stable and fast method for the eigenvalue computation of full matrices. The QR method preserves the upper Hessenberg structure of the input matrix A but neither preserves the tridiagonal form of non-Hermitian matrices nor the symplectic structure of matrices. For matrices with such special structures new eigenvalue algorithms preserving these structures were developed recently. Bunse-Gerstner studied in [7, 8, 10] an eigenvalue algorithm (HR algorithm) which preserves sign-symmetric tridiagonal matrices. The SR algorithm, developed by Della-Dora [16] and Bunse-Gerstner, Mehrmann and Watkins [12] preserves Hamiltonian and symplectic matrices. In 1991, Watkins and Elsner [76, 77] introduced and studied the concept of a generic GR algorithm which includes the QR, LR, SR and HR algorithms as special cases. The convergence theory of the GR algorithm is based on the idea that the GR algorithm performs nested subspace iteration with a change of coordinate system at each step. Earlier Buurema [13] used this idea for a geometric convergence proof of the QR algorithm. In 1997, Uhlig [74] developed a specil GR algorithm, called DQR algorithm, which is based on hyperbolic Givens rotations. This method computes the eigenvalues of tridiagonal matrices and preserves the tridiagonal structure.
We point out that all GR algorithms that preserve the tridiagonal structure are in a certain sense the mathematical equivalent of a restarted Lanczos method (see Chapter 6). Hence, the instabilities of the Lanczos method transfer to these methods.
In 1999, Geist, Howell and Watkins [23] presented the BR algorithm. This is a special GR algorithm which uses Gauss transformations and row and column pivoting. To avoid instabilities by near-breakdowns a row-column pair is eliminated only if the multipliers are smaller than a tolerance. The BR algorithm generates a sequence of banded upper Hessenberg matrices. The band width depends on the choice of the tolerance and the size of the starting matrix.

In the 1980s eigenvalue problems with large, sparse, not necessarily Hermitian $N \times N$ matrices came into focus. Such eigenvalue problems arise, e.g., in physics, chemistry and biology by discretising differential equations (see, e.g., [2], [67]). Typically, only few eigenvalues with specified properties (e.g. maximal real part) are wanted and the matrix A is too large for explicit storage, but matrix vector products can be easily computed. In this situation the known GR algorithms are not applicable or at least not efficient because these methods require explicit storage of the matrix A and destroy in general the sparseness of the matrix A. It was soon recognized that the Lanczos algorithm and the Arnoldi algorithm are useful for sparse matrix computations because the matrix A appears only in matrix vector products.

In 1980, Saad [64] pointed out that the Arnoldi method can be seen as an iterative Rayleigh-Ritz projection method for the Krylov subspaces $\mathcal{K}(j, A, v)$ $(j = 1, 2, \ldots, s)$. In other words, the Arnoldi matrix $H^{(j)}$ represents the orthogonal projection of A onto the Krylov subspace $\mathcal{K}(j, A, v)$ with respect to the computed orthogonal basis. Furthermore the Krylov subspaces 'converge' to an invariant subspace of A. Typically, after few iteration steps the eigenvalues of A with the largest real part are well-approximated by eigenvalues of $H^{(j)}$ (see, e.g., Saad [67], Jia [43]).

Saad [65] (in 1982) and Parlett, Taylor and Liu [59] (in 1985) pointed out that the nonsymmetric Lanczos method is an oblique projection method. More precisely, the Lanczos matrices $H^{(j)}$ $(j = 1, 2, \ldots, s)$ represent oblique projections of the matrix A onto the Krylov subspace $\mathcal{K}(j, A, v)$. Like the Arnoldi method the Lanczos method can be seen as an iterative method where the eigenvalues of $H^{(j)}$ 'converge' to the eigenvalues of A (if no breakdown occurs in an early step).

In 1985 Parlett, Taylor and Liu [59] classified the breakdowns of the nonsymmetric Lanczos method. They found that the breakdowns fall into two classes which they called curable and incurable. If an incurable breakdown occurs, then every eigenvalue of the computed tridiagonal matrix $H^{(s)}$ is an eigenvalue of the given matrix A although no invariant subspace is found. This is the famous 'Mismatch Theorem' of Taylor [59, 73]. Furthermore they developed a so-called 'look-ahead Lanczos algorithm' which can skip over curable breakdowns and near-breakdowns of length 2. In 1992, Gutknecht [33, 34] derived, via orthogonal polynomials and Padé approximation, a look-ahead Lanczos algorithm that can skip over curable breakdowns and near breakdowns of any length. At the same time, Parlett [58] presented an extended two-sided Gram-Schmidt method which led to the same look-ahead Lanczos method as Gutknecht suggested in [33, 34]. One year later Freund, Gutknecht und Nachtigal [20] presented a practical implementation of this look-ahead Lanczos method. In the context of formally orthogonal polynomials the look-ahead Lanczos recursions for overcoming exact curable breakdowns were mentioned already by Gragg [30] (in 1974) and by Draux [18] (in 1983). Gragg and Lindquist [31] used these recursions when looking for a solution of the partial realization problem of systems theory. The relations to systems theory were also worked out by Parlett [58], Boley and Golub [5], and Golub, Kågström and Van Dooren [29]. An extensive treatment of Padé approxima-

tion and its connection to look-ahead Lanczos methods has been presented by Hochbruck [39], together with suggestions for making look-ahead Lanczos methods more stable and efficient. In the review article [35] of Gutknecht and in Hochbruck's Habilitationsschrift [39] various (look-ahead) Lanczos-type solvers for large linear systems are discussed in detail and ample references to earlier work are given.
Recently, Hochbruck and Lubich [40], Liesen [53] and Koch and Liesen [47] derived new error bounds for the solution of linear systems via Krylov methods.

The main problem of Arnoldi's method is that it becomes increasingly expensive per iteration step. For this reason Saad [64, 66, 67] suggested to restart the Arnoldi method after every $k << N$ steps. There are different possibilities to choose the new starting vector. A typical scheme is to compute the eigenvalues of the generated matrix $H^{(k)}$ and divide them into wanted eigenvalues (e.g. the l rightmost eigenvalues if we are looking for the l rightmost eigenvalues) and unwanted eigenvalues μ_j $(j = 1, 2, \ldots, k-l)$. In order to damp out the unwanted eigenvalues the new starting vector is computed by $v_1^{(i+1)} = \psi_i(A)v_1^{(i)}$ where ψ_i is a polynomial with zeros μ_j $(j = 1, 2 \ldots, k-l)$.
Even for Lanczos methods, restarting is a good choice because reorthogonalization is in general not necessary.
In 1992, Sorensen [70] recognized that the restarted Arnoldi method is a truncated QR-algorithm. In our notation, the matrices $H_i^{(s)}$, generated by the restarted Arnoldi method with shift polynomials ψ_{i-1}, are the leading submatrices of the upper Hessenberg matrices A_i, generated by a QR algorithm using the same shift polynomials ψ_{i-1}:

$$A_1 = Q_0^{-1}AQ_0, \qquad \psi_{i-1}(A_i) = Q_iR_i, \qquad A_{i+1} = Q_i^{-1}A_iQ_j = \begin{bmatrix} H_i^{(s)} & \star \\ \star & \star \end{bmatrix} \tag{1.1}$$

for all $i \in \mathbb{N}$. Using this connection, Sorensen derived the so-called implicitly restarted Arnoldi method. This method is mathematically equivalent to the restarted Arnoldi method, described above, but requires much less computational effort.
In 1995, Lehoucq [52] analysed the implicitly restarted Arnoldi method and presented an implementation of this algorithm.
The 'implicit restart' idea of Sorensen has been transferred to the nonsymmetric Lanczos process by Grimme, Sorensen and Van Dooren [32] (in 1996), Benner and Faßbender [4] (in 1997) and De Samblanx and Bultheel [17] (in 1998). Grimme, Sorensen and Van Dooren used a relation between the HR algorithm and a restarted sign-symmetric Lanczos method similar to (1.1). Benner and Faßbender used a connection between the SR algorithm and a restarted symplectic Lanczos method similar to (1.1).

1.1 Summary of the thesis

The main goals of this thesis are:

a) To provide a useful setting for the construction of stable and efficient Krylov subspace projection methods and restarted Krylov subspace projection methods.

b) To discuss the connections between restarted Krylov subspace projection methods and GR algorithms. It turns out that GR algorithms and Krylov subspace projection methods are in the following sense mathematically equivalent: Let $(H_i)_{i\in\mathbb{N}}$ be the sequence of $s \times s$ matrices generated by a restarted Krylov subspace projection method using the shift polynomials ψ_i. Then there is a GR algorithm (using the same shift polynomials ψ_i) which generates a sequence of upper Hessenberg matrices $(A_i)_{i\in\mathbb{N}}$ such that for all $i \in \mathbb{N}$ the matrix H_i is the leading $s \times s$ submatrix of A_{i+1} (similar to (1.1)).
The reverse of this statement is also valid.

c) To prove that the restarted look-ahead Lanczos method, using Rayleigh-quotient shifts, converges cubically under some mild conditions, when applied to an arbitrary given $N \times N$ matrix with distinct eigenvalues.

d) To present a stable and efficient GR algorithm which is mathematically equivalent to the restarted look-ahead Lanczos method (see b)). Notice, that (by c)) this GR algorithm converges cubically under mild conditions when applied to any given $N \times N$ matrix with distinct eigenvalues. We point out that the QR algorithm and all other stable GR algorithms (known to the author) converge typically quadratically when applied to a given $N \times N$ matrix with distinct eigenvalues.

The thesis is organized as follows: In Chapter 2 we describe the main idea of general subspace projection methods and describe the Arnoldi method and the Lanczos method in detail. The connection of the Arnoldi method to orthogonal projection methods and orthogonal bases in Krylov chains is discussed. The relation between the Lanczos method and oblique projection methods and biorthogonal basis of Krylov chains is described. A classification of the breakdowns of the Lanczos method is given and serious breakdowns are explained by the non-existence of biorthogonal basis in the underlying Krylov chain. Chapter 3 follows the line of [81] presented by the author in 1997. In this chapter we introduce a generalization of biorthogonal bases in a chain of subspaces and examine their existence and their degrees of freedom. It turns out that these general biorthogonal bases are fixed points of Parlett's [58] extended Gram-Schmidt biorthogonalization algorithm. Then we examine in particular general biorthogonal bases of Krylov chains. We find that these general biorthogonal bases exist even when a curable breakdown of the Lanczos method occurs. Furthermore we show that every general biorthogonal basis of a

Krylov chain can be constructed by special block-three-term recursions. These recursions are identical to the most general form of Gutknecht's [33] (exact) look-ahead Lanczos recursions. Whereas Gutknecht favours three-term recursions for the inner vectors, we propose a special recursion (minimal look-ahead Lanczos recursion) which requires minimal computational effort and storage. The construction of any generalized biorthogonal basis requires exact knowledge of the degree indices. To get these indices one has to decide, after every step, whether a certain matrix is invertible. Thus, in finite-precision arithmetic only a subsequence of the degree indices can be determined. For this reason the theory of generalized biorthogonal bases is extended to the case that only a subsequence of the degree indices is known. This is done in Chapter 4. The recursions found with this ansatz are identical to the practical look-ahead Lanczos recursions of Freund, Gutknecht and Nachtigal [20] and Gutknecht [34]. These algorithms can skip over curable and near breakdowns. In Chapter 4 we further present an implementation of the practical minimal look-ahead Lanczos method and compare this method with the practical monomial look-ahead Lanczos method.
The aim of Chapter 5 is to provide a general concept for Krylov subspace projection methods. Let an arbitrary $N \times N$ matrix A be given. We call an algorithm Krylov subspace projection method if it generates matrices $V \in \mathrm{Mat}_{N\times s}$ and $H \in \mathrm{Mat}_{s\times s}$ such that the column vectors of V form a basis of the Krylov chain

$$\mathcal{K}(1, A, Ve_1) \subset \mathcal{K}(2, A, Ve_1) \subset \cdots \subset \mathcal{K}(s, A, Ve_1)$$

and there is a (not necessarily orthogonal) projection Π onto the subspace $\mathcal{K}(s, A, Ve_1)$ such that H represents the projection of A onto this Krylov subspace (i.e. $VH = \Pi AV$). For simplicity and clarity we collect these objects in a five tuple (A, V, Π, H, s) and call this tuple 'Krylov tuple'. We study how to construct Krylov tuples with less effort and discuss their degrees of freedom. Estimates for the condition numbers of the Krylov matrices V and eigenvalue estimates are derived. A very short and instructive proof of Taylor's Mismatch Theorem is presented. It is discussed how to reduce the effort of practical look-ahead Lanczos methods by using adapted non-degenerate biorthogonal forms in the look-ahead Lanczos process.

In Chapter 6 we study a general polynomial restarted Krylov subspace projection method and its connection to the general GR algorithm. We prove that these methods are mathematically equivalent in the sense described in b) above. The convergence theorems of Watkins and Elsner [76] for GR algorithms apply and yield (under mild conditions) the convergence of the restarted Krylov subspace projection methods. Under some conditions quadratic convergence is shown if the Rayleigh-quotient shift strategy (typically used for GR algorithms) is used. Furthermore, we prove that the LR algorithm with exact row-pivoting is, under mild conditions, mathematically equivalent to the restarted monomial look-ahead Lanczos method when the left starting vectors $w_1^{(i+1)}$ are not changed:

$$w_1^{(i+1)} = w_1^{(0)} \qquad \text{and} \qquad v_1^{(i+1)} = (A - \mu_i)v_1^{(i)}. \tag{1.2}$$

It is well known that the LR algorithm with Rayleigh-quotient shifts converges typically quadratically. Consequently, the restarted monomial look-ahead Lanczos method, using Rayleigh-quotient shifts and starting vectors of the form (1.2), converges also typically quadratically (in exact arithmetic). But we prove in Chapter 7 that restarted practical look-ahead Lanczos methods converge even cubically, under some mild conditions, if the Rayleigh-quotient shift strategy is used to determine the shift polynomial ψ_i and both starting vectors are updated with the polynomial ψ_i:

$$w_1^{(i+1)} = \psi_i(A^T)w_1^{(i)} \qquad \text{and} \qquad v_1^{(i+1)} = \psi_i(A)v_1^{(i)}. \tag{1.3}$$

A numerical example is given which shows that the practical restarted look-ahead Lanczos methods are too unstable for the accurate computation of all eigenvalues. To avoid such instabilities we develop in Chapter 8 a new stable method for the computation of look-ahead Lanczos matrices via elimination. In each step k it is checked whether the row g_k^T and the column $\tilde{g}_k$, which have to be eliminated in this step, satisfy the condition

$$\frac{|g_k^T \tilde{g}_k|}{||g_k||\ ||\tilde{g}_k||} \geq tol. \tag{1.4}$$

The tolerance $tol \in (0, 1]$ is chosen by the user. If the inequality (1.4) holds then the column-row pair is eliminated by a similarity transformation with an elementary matrix $I - \gamma xy^T$ satisfying the inequality

$$\kappa_2(I - \gamma xy^T) \leq 18 \left(\frac{||g_k||\ ||\tilde{g}_k||}{|g_k^T \tilde{g}_k|} + \frac{1}{2} \right)^2 + 2. \tag{1.5}$$

There is a degree of freedom in the choice of the elementary matrix. We have used this degree of freedom to find an elementary matrix with small condition number. If the condition (1.4) is not fulfilled then we use unitary Householder transformations in the following l steps until the condition number of a certain $l \times l$ matrix D is smaller than $1/tol$ ('inner steps'). If this occurs a linear equation with the matrix D has to be solved. If the inequality (1.4) holds we continue with an row-column elimination as described above ('regular step').
In Chapter 9 we suggest a restarted look-ahead Lanczos method (called RSL algorithm) for the computation of all eigenvalues of a matrix.
In Chapter 10 we present an implicit GR algorithm which is mathematically equivalent to the RSL algorithm. We call this method LGR algorithm. The RSL algorithm and the LGR algorithm are based on the stable similarity look-ahead Lanczos method (described above) and use Rayleigh-quotient shifts, starting vectors of the form (1.3), and deflation techniques. We point out that both methods cannot break down and that an upper bound for the condition numbers of all occuring transformations is determined by the user (see (1.4) and (1.5)). The cubic convergence theorem for restarted look-ahead Lanczos methods, proved in Chapter 7, transfers to the LGR and the RSL algorithm. More

precisely, the RSL algorithm converges cubically for any given square matrix with distinct eigenvalues if it converges at all and if the number *tol* is chosen (by the user) small enough. If the chosen number *tol* is too large then it converges at least quadratically if it converges at all. These convergence statements are also valid for the LGR algorithm if the condition numbers of the product of all used transformation matrices remain bounded when the iteration proceeds (i.e. $i \longrightarrow \infty$).
We recall that the QR algorithm (using Rayleigh-quotient shifts) converges quadratically for any given square matrix with distinct eigenvalues if it converges. We furthermore recall that there are examples [80] for which the QR algorithm (using Rayleigh-quotient shifts) fails to converge.
In Chapter 11 we give flop counts and discuss numerical examples to compare the LGR and the RSL algorithms with the QR algorithm. There are different definitions of 'flop'. In this thesis we use the new definiton of 'flop' (see [28]): One flop is one floating point operation and not an arithmetic operation of the form $a + b \cdot c$.
We show that the LGR and the RSL algorithm typically require

$$\frac{8}{3}N^3 + O(N^2)$$

flops for the computation of all eigenvalues of any given $N \times N$ matrix and

$$O(N^2)$$

flops for the computation of all eigenvalues of $N \times N$ Hessenberg matrices which are nearly tridiagonal.
More precise flop estimates are also derived in Chapter 11.
In comparison we recall that the implicitly shifted QR algorithm requires typically

$$10N^3 + O(N^2)$$

flops for the computation of all eigenvalues of an arbitrary given $N \times N$ matrix, and

$$\frac{20}{3}N^3 + O(N^2)$$

flops for the computation of all eigenvalues of $N \times N$ Hessenberg matrices (see [28]).
The most important observations we have made in our numerical examples (28 matrices with $N \leq 250$, double precision arithmetic) are the following:
In all examples we have observed cubic convergence of the LGR algorithm and quadratic convergence of the QR algorithm. Furthermore we have observed that the RSL algorithm does not converge in 12 out of 28 examples.
For full and upper Hessenberg matrices with $N \leq 100$ the LGR algorithm requires about 1/2 the number of flops as compared to the QR method.

For tridiagonal matrices the LGR method requires
for $N = 50$ about 1/5 the number of flops of the QR method,
for $N = 100$ about 1/10 the number of flops of the QR method,
for $N = 200$ about 1/20 the number of flops of the QR method.
The LGR algorithm computes eigenvalue approximations with an average maximal relative error of $2 \cdot 10^{-8}$.
The QR algorithm generates eigenvalue approximations with an average maximal relative error of $3 \cdot 10^{-14}$.

1.2 Notation

Throughout the paper we use the notation:

- $\Re\lambda$ and $\Im\lambda$ denote the real part and imaginary part of $\lambda \in \mathbb{C}$, respectively.
- $x^T = [x_1, x_2, \ldots, x_N]$ denotes the transpose and $x^* = [\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_N]$ the complex conjugate transpose of
$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \in \mathbb{C}^N.$$
- $\langle x, y\rangle = x^T y$ and $\langle x, y\rangle_* = x^* y$.
- Let $\mathcal{W}$ be a subspace of $\mathbb{C}^N$.
Then
$$\mathcal{W}^{\perp} = \{x \in \mathbb{C}^N \mid \langle w, x\rangle = 0 \quad \text{for all} \quad w \in \mathcal{W}\}$$
and
$$\mathcal{W}^{\perp_*} = \{x \in \mathbb{C}^N \mid \langle w, x\rangle_* = 0 \quad \text{for all} \quad w \in \mathcal{W}\}$$
denote the orthogonal complements of $\mathcal{W}$ with respect to $\langle\cdot,\cdot\rangle$ and $\langle\cdot,\cdot\rangle_*$, respectively.
- $||x|| = \sqrt{x^* x}$ and $||x||_\infty = \max_{1\le i\le N} |x_i|$ for $x \in \mathbb{C}^N$.
- Mat_N denotes the set of complex $N \times N$ matrices and $\mathrm{Mat}_{k\times l}$ denotes the set of complex $k \times l$ matrices.
- Let $A = [a_{ij}] \in \mathrm{Mat}_{k\times l}$. Then $A^T = [a_{ji}] \in \mathrm{Mat}_{l\times k}$ denotes the transpose and $A^* = \overline{A}^T := [\overline{a}_{ji}] \in \mathrm{Mat}_{l\times k}$ denotes the complex conjugate transpose.
- For $v \in \mathbb{C}^N \setminus \{0\}$ and $A \in \mathrm{Mat}_N$ the i-th Krylov subspace is defined by
$$\mathcal{K}(i, A, v) = \mathrm{span}\{v, Av, \ldots, A^{i-1}v\}.$$
- $\sigma_{min}(V)$ denotes the minimal singular value and $\sigma_{max}(V)$ the maximal singular value of a matrix $V \in \mathrm{Mat}_{N\times k}$.

- For matrices $V \in \mathrm{Mat}_{N\times k}$ with full rank the condition number $\kappa_2(V)$ is defined by $\kappa_2(V) = \sigma_{max}(V)/\sigma_{min}(V)$.
- $||V|| = \max_{||x||=1} ||Vx||$ and $||V||_\infty = \max_{||x||_\infty=1} ||Vx||_\infty$ for $V \in \mathrm{Mat}_{N\times k}$.
- For every matrix $V = [v_{ij}] \in \mathrm{Mat}_{N\times k}$ the matrix $|V| \in \mathrm{Mat}_{N\times k}$ is defined by $|V| = [|v_{ij}|]$.
- Pol_i denotes the set of all polynomials with degree less than $i+1$ and coefficients in $\mathbb{C}$.
  Pol denotes the set of all polynomials with coefficients in $\mathbb{C}$.

2 Introduction to projection methods

We consider the problem of finding k eigenvalues of a very large, sparse, not necessarily Hermitian $N \times N$ matrix A where $N >> k$. Efficient algorithms for solving such eigenvalue problems are subspace projection methods. Subspace projection methods are based on the idea of reducing the N-dimensional eigenvalue problem to an s-dimensional eigenvalue problem ($k \leq s << N$):

(1) Choose an s-dimensional subspace $\mathcal{K}$ of $\mathbb{C}^N$.

(2) Choose a projection $\mathbf{\Pi}$ onto $\mathcal{K}$.

(3) Compute the eigenvalues and eigenvectors of the linear operator $\mathbf{H} : \mathcal{K} \longrightarrow \mathcal{K}$ defined by

$$\mathbf{H} = \mathbf{\Pi}\mathbf{A}|_{\mathcal{K}}.$$

Here $\mathbf{A}$ is the linear operator related to the matrix $A \in \mathrm{Mat}_N$.

The quality of the approximation depends on the subspace $\mathcal{K}$ and the projection $\mathbf{\Pi}$. If $\mathcal{K}$ is an $\mathbf{A}$-invariant subspace, then every eigenvector and eigenvalue of $\mathbf{H}$ is an eigenvector and eigenvalue of $\mathbf{A}$, respectively. If $\mathcal{K}$ is close to an $\mathbf{A}$-invariant subspace and the norm of the projection is not too large, then the eigenvalues of $\mathbf{A}$ are well approximated (see Section 5.4).

For the computation of the eigenvectors and eigenvalues of $\mathbf{H}$ we require a matrix representation $H \in \mathrm{Mat}_s$ of the linear operator $\mathbf{H}$. Therefore we have to choose a basis $v_1, v_2, \ldots, v_s$ of $\mathcal{K}$ and determine H as the unique solution of the equation

$$VH = \Pi AV \tag{2.1}$$

where $V = [v_1, v_2, \ldots, v_s]$ and Π is a matrix representation of the projection $\mathbf{\Pi}$ with respect to the standard basis $e_1, e_2, \ldots, e_N$ of $\mathbb{C}^N$.

Consequently the matrix H depends on:

- The choice of the subspace $\mathcal{K}$.
- The choice of the basis $v_1, v_2, \ldots, v_s$ of the subspace $\mathcal{K}$.
- The choice of the kernel of Π.

In the following we choose

$$\mathcal{K} = \mathcal{K}(s, A, v_1)$$

where $v_1 \in \mathbb{C}^N \setminus \{0\}$. The Krylov subspace $\mathcal{K}(s, A, v_1)$ is a good choice for the following reasons:

- The construction of a basis of the Krylov subspace only requires matrix vector products. Thus, the sparseness of the matrix A can be used.

- Krylov subspaces are good approximations to suitably chosen invariant subspaces of A (see e.g. [66], [67], [43]).

2.1 The Arnoldi method and orthonormal bases

Now we turn to the construction of a basis of $\mathcal{K}(s, A, v_1)$ where s is chosen such that $s \leq \dim \mathcal{K}(N, A, v_1)$. One possibility is to construct an orthonormal basis $\tilde{v}_1, \tilde{v}_2, \ldots, \tilde{v}_s$. Theoretically, this can be done by applying the Gram-Schmidt orthonormalisation method [68] to a basis of $\mathcal{K}(s, A, v_1)$, for example to

$$v_1, Av_1, A^2v_1, \ldots, A^{s-1}v_1. \tag{2.2}$$

In this case we obtain the formulas:

For $i = 1, 2, \ldots, s-1$

$$\begin{aligned} \tilde{v}_1 &:= \frac{v_1}{||v_1||} \\ \check{v}_{i+1} &:= A^i v_1 - \sum_{j=1}^{i} \langle \tilde{v}_j, A^i v_1 \rangle_* \tilde{v}_j = (I - \Pi_i) A^i v_1 \\ \tilde{v}_{i+1} &:= \frac{\check{v}_{i+1}}{||\check{v}_{i+1}||}. \end{aligned} \tag{2.3}$$

Here Π_i denotes the orthogonal projection onto $\mathcal{K}(i, A, v_1)$. In view of the power method one can expect that the matrix $[v_1, Av_1, A^2v_1, \ldots, A^{s-1}v_1]$ is not well-conditioned. Hence the computation of $\tilde{v}_1, \tilde{v}_2, \ldots, \tilde{v}_s$ by (2.3) is, in general, not stable. A better choice is the recursive computation of an orthonormal basis:

For $i = 1, 2, \ldots, s-1$

$$\begin{aligned} v_1 &:= \frac{v_1}{||v_1||} \\ \hat{v}_{i+1} &:= Av_i - \sum_{j=1}^{i} \langle v_j, Av_i \rangle_* v_j = (I - \Pi_i) Av_i \\ v_{i+1} &:= \frac{\hat{v}_{i+1}}{||\hat{v}_{i+1}||}. \end{aligned} \tag{2.4}$$

This is the Arnoldi method [1]. It generates for every $j \in \{1, 2, \ldots, s\}$ an orthonormal

basis $v_1, v_2, \ldots, v_j$ of $\mathcal{K}(i, A, v_1)$ and the upper Hessenberg matrix

$$H^{(j)} = \begin{bmatrix} \langle v_1, Av_1\rangle_* & \langle v_1, Av_2\rangle_* & \cdots & \langle v_1, Av_j\rangle_* \\ \|\hat{v}_2\| & \langle v_2, Av_2\rangle_* & \cdots & \langle v_2, Av_j\rangle_* \\ & \ddots & \ddots & \vdots \\ 0 & & \|\hat{v}_j\| & \langle v_j, Av_j\rangle_* \end{bmatrix} \in \mathrm{Mat}_j$$

satisfying the equation

$$V^{(j)}H^{(j)} = \Pi_j A V^{(j)} \tag{2.5}$$

where $V^{(j)} = [v_1, v_2, \ldots, v_j] \in \mathrm{Mat}_{N\times j}$ and Π_j is the orthogonal projection on $\mathcal{K}(j, A, v_1)$.

Remark 2.1 The equation (2.5) can be rewritten in the more familiar form

$$V^{(j)}H^{(j)} = AV^{(j)} - \hat{v}_{j+1}e_j^T$$

as described below: Using that

$$\mathrm{Im}\, AV^{(j-1)} \subseteq \mathcal{K}(j, A, v_1) = \mathrm{Im}\, \Pi_j$$

we conclude

$$\begin{aligned} V^{(j)}H^{(j)} &= \Pi_j AV^{(j)} = [\Pi_j AV^{(j-1)},\, \Pi_j Av_j] = [AV^{(j-1)},\, \Pi_j Av_j] \\ &= [AV^{(j-1)},\, Av_j] - [0, \ldots, 0,\, (I - \Pi_j)Av_j] \\ &= AV^{(j)} - [0, \ldots, 0,\, (I - \Pi_j)Av_j] \\ &= AV^{(j)} - \{(I - \Pi_j)Av_j\}e_j^T \\ &= AV^{(j)} - \hat{v}_{j+1}e_j^T. \end{aligned}$$

In the following we show that the Arnoldi vectors $v_1, v_2, \ldots, v_s$ are (up to scalar factors of modulus 1) identical with the vectors $\tilde{v}_1, \tilde{v}_2, \ldots, \tilde{v}_s$ generated by the Gram-Schmidt method applied to the sequence (2.2). For this purpose and for later use we examine the uniqueness of orthogonal bases in a chain of subspaces

$$\mathcal{V}_1 \subset \mathcal{V}_2 \subset \cdots \subset \mathcal{V}_s$$

with $\dim \mathcal{V}_1 = i$ for all $i \in \{1, 2, \ldots, s\}$.

Definition 2.1 (Basis of $(\mathcal{V}_i)_{1\le i\le s}$) *A sequence of vectors $(v_i)_{1\le i\le s}$ with the property*

$$\mathcal{V}_i = \mathrm{span}\{v_1, v_2, \ldots, v_i\}$$

for all $1 \le i \le s$ is called a basis of the chain $(\mathcal{V}_i)_{1\le i\le s}$.

Definition 2.2 (Orthogonal basis) *A basis $(v_i)_{1\leq i\leq s}$ of $(\mathcal{V}_i)_{1\leq i\leq s}$ is called orthogonal if*

$$\langle v_i, v_j\rangle_* = 0 \qquad \text{for} \quad i \neq j.$$

The following Lemma states the degrees of freedom in the choice of a basis in a chain of subspaces.

Lemma 2.1 *Let $(v_i)_{1\leq i\leq s}$ and $(\tilde{v}_i)_{1\leq i\leq s}$ be bases of a chain $(\mathcal{V}_i)_{1\leq i\leq s}$. Then there is an invertible upper triangular matrix $R \in \mathrm{Mat}_s$ such that*

$$[\tilde{v}_1,\, \tilde{v}_2, \ldots, \tilde{v}_s] = [v_1,\, v_2, \ldots, v_s]R.$$

Lemma 2.1 follows immediately from the definition of a basis of a chain of subspaces.

Theorem 2.1 (Existence and uniqueness) *There is (up to scalar factors) exactly one orthogonal basis of $(\mathcal{V}_i)_{1\leq i\leq s}$.*

Proof. The Gram-Schmidt orthogonalisation applied to any basis of $(\mathcal{V}_i)_{1\leq i\leq s}$ yields an orthogonal basis. Thus, the existence is shown. Let $(v_i)_{1\leq i\leq s}$ and $(\tilde{v}_i)_{1\leq i\leq s}$ be orthogonal bases of $(\mathcal{V}_i)_{1\leq i\leq s}$. Define

$$V = [v_1, v_2, \ldots, v_s] \quad \text{and} \quad \tilde{V} = [\tilde{v}_1, \tilde{v}_2, \ldots, \tilde{v}_s].$$

Then by Lemma 2.1 there is an invertible upper triangular matrix R such that

$$\tilde{V} = VR. \tag{2.6}$$

By assumption there are invertible diagonal matrices $\tilde{D}$ and D such that

$$V^*V = D \qquad \text{and} \qquad \tilde{V}^*\tilde{V} = \tilde{D}. \tag{2.7}$$

Combining (2.6) and (2.7) we obtain the equation

$$R = D^{-1}(R^*)^{-1}\tilde{D}.$$

The matrix R is upper triangular and the matrix $D^{-1}(R^*)^{-1}\tilde{D}$ is lower triangular. Thus, the matrix R is diagonal. ∎

Theorem 2.1 implies that the Gram-Schmidt orthonormalisation method, applied to any basis of $(\mathcal{V}_i)_{1\leq i\leq s}$, generates the (up to scalar factors of modulus one) uniquely determined orthonormal basis of $(\mathcal{V}_i)_{1\leq i\leq s}$.

Now we turn to the Arnoldi method. The Arnoldi method is the Gram-Schmidt orthonormalisation method applied to the sequence

$$v_1, Av_1, Av_2, \ldots, Av_{s-1}. \tag{2.8}$$

It is evident that this sequence forms a basis of the Krylov chain $(\mathcal{K}(i, A, v_1))_{1\leq i\leq s}$. Consequently, by Theorem 2.1, the Arnoldi vectors are, up to scalar factors of modulus 1, identical with the vectors generated by the Gram-Schmidt orthonormalisation method applied to the sequence (2.8).

2.2 The Lanczos method and biorthogonal bases

In the last section we have seen that the Arnoldi vectors are (up to scalar factors of modulus 1) identical with the vectors generated by the Gram-Schmidt orthonormalisation method applied to any basis of the Krylov chain $(\mathcal{K}(i, A, v_1))_{1\le i\le s}$. Here v_1 is the starting vector of the Arnoldi method. In this section we show that the vectors generated by the Lanczos method are (up to scalar factors) identical with the vectors generated by the Gram-Schmidt biorthogonalisation method applied to any basis of the pair of chains

$$\begin{aligned} \mathcal{K}(1, A, v_1) &\subset \mathcal{K}(2, A, v_1) \subset \cdots \subset \mathcal{K}(s, A, v_1) \\ \mathcal{K}(1, A^T, w_1) &\subset \mathcal{K}(2, A^T, w_1) \subset \cdots \subset \mathcal{K}(s, A^T, w_1) \end{aligned}$$

where v_1 and w_1 are the starting vectors of the Lanczos method.

Definition 2.3 *A sequence of pairs of subspaces $((\mathcal{W}_i, \mathcal{V}_i))_{1\le i\le s}$ is called a chain of subspaces if*

$$\begin{aligned} \mathcal{V}_1 &\subset \mathcal{V}_2 \subset \cdots \subset \mathcal{V}_s \\ \mathcal{W}_1 &\subset \mathcal{W}_2 \subset \cdots \subset \mathcal{W}_s \end{aligned}$$

and $\dim \mathcal{V}_i = i = \dim \mathcal{W}_i$ *for all* $i \in \{1, 2, \ldots, s\}$.

Definition 2.4 (Basis of $((\mathcal{W}_i, \mathcal{V}_i))_{1\le i\le s}$) *A sequence of pairs of vectors $((w_i, v_i))_{1\le i\le s}$ with the property*

$$\mathcal{V}_i = \operatorname{span}\{v_1, v_2, \ldots, v_i\} \quad \textit{and} \quad \mathcal{W}_i = \operatorname{span}\{w_1, w_2, \ldots, w_i\}$$

for all $i \in \{1, 2, \ldots, s\}$ is called basis of $((\mathcal{W}_i, \mathcal{V}_i))_{1\le i\le s}$.

Definition 2.5 (Biorthogonal basis) *A basis $((w_i, v_i))_{1\le i\le s}$ of $((\mathcal{W}_i, \mathcal{V}_i))_{1\le i\le s}$ is called biorthogonal if*

$$\langle w_i, v_j\rangle = 0 \quad \textit{for} \quad i \neq j \quad \textit{and} \quad \langle w_i, v_i\rangle \neq 0.$$

Theorem 2.2 (Uniqueness) *Let $((w_i, v_i))_{1\le i\le s}$ and $((\tilde{w}_i, \tilde{v}_i))_{1\le i\le s}$ be biorthogonal bases of $((\mathcal{W}_i, \mathcal{V}_i))_{1\le i\le s}$. Then there are numbers $\alpha_i, \beta_i \in \mathbb{C} \setminus \{0\}$ such that*

$$w_i = \alpha_i \tilde{w}_i \quad \textit{and} \quad v_i = \beta_i \tilde{v}_i \qquad \textit{for all } i.$$

Proof. Define

$$\begin{aligned} V &= [v_1, v_2, \ldots, v_s], \qquad \tilde{V} = [\tilde{v}_1, \tilde{v}_2, \ldots, \tilde{v}_s] \\ W &= [w_1, w_2, \ldots, w_s] \quad \text{and} \quad \tilde{W} = [\tilde{w}_1, \tilde{w}_2, \ldots, \tilde{w}_s]. \end{aligned}$$

Then by Lemma 2.1 there are invertible upper triangular matrix R and $\tilde{R}$ such that

$$\tilde{V} = VR \qquad \text{and} \qquad \tilde{W} = W\tilde{R}. \tag{2.9}$$

By assumption there are invertible diagonal matrices $\tilde{D}$ and D such that

$$W^T V = D \qquad \text{and} \qquad \tilde{W}^T \tilde{V} = \tilde{D}. \tag{2.10}$$

Combining (2.9) and (2.10) we find the identity

$$R = D^{-1}(\tilde{R}^T)^{-1}\tilde{D}.$$

The matrix R is upper triangular and the matrix $D^{-1}(\tilde{R}^T)^{-1}\tilde{D}$ is lower triangular. Hence the matrices R and $\tilde{R}$ are diagonal. ■

Next we discuss the existence of a biorthogonal basis of $((\mathcal{W}_i, \mathcal{V}_i))_{1\le i\le s}$. For this purpose we require some results about the representation and existence of projections.

Lemma 2.2 (Representation of a projection) *Let $\Pi \in \mathrm{Mat}_N$ be a projection onto an i-dimensional subspace. Let $V, W \in \mathrm{Mat}_{N\times i}$ such that $\operatorname{Im} V = \operatorname{Im}\Pi$ and $\operatorname{Im} W = (\operatorname{Ker}\Pi)^\perp$. Then $W^T V$ is invertible and Π has the representation*

$$\Pi = V(W^T V)^{-1} W^T.$$

Proof. We first show that $W^T V$ is invertible. Combining the equations

$$\operatorname{Ker}\Pi \cap \operatorname{Im}\Pi = \{0\} \qquad \text{and} \qquad \operatorname{Ker}\Pi = (\operatorname{Im} W)^\perp = \operatorname{Ker} W^T$$

we find

$$\operatorname{Ker} W^T \cap \operatorname{Im} V = \operatorname{Ker}\Pi \cap \operatorname{Im}\Pi = \{0\}.$$

Thus, for every $x \in \mathbb{C}^i \setminus \{0\}$ the vector $\tilde{x} := Vx$ is not in $\operatorname{Ker} W^T$. Consequently $W^T V x$ does not vanish for all vectors $x \in \mathbb{C}^i \setminus \{0\}$. This yields the invertibility of $W^T V$.
Define $\tilde{\Pi} := V(W^T V)^{-1} W^T$.
The identity

$$\tilde{\Pi}^2 = V(W^T V)^{-1} W^T V (W^T V)^{-1} W^T = V(W^T V)^{-1} W^T = \tilde{\Pi}$$

shows that $\tilde{\Pi}$ is a projection. Finally the equations

$$\begin{aligned} \operatorname{Ker}\Pi &= (\operatorname{Im} W)^\perp = \operatorname{Ker} W^T = \operatorname{Ker}\tilde{\Pi} \\ \operatorname{Im}\Pi &= \operatorname{Im} V = \operatorname{Im}\tilde{\Pi} \end{aligned}$$

yield $\Pi = \tilde{\Pi}$, which proves the lemma. ■

Lemma 2.3 (Existence of a projection) *Let $\mathcal{V}$ and $\mathcal{W}$ be subspaces of $\mathbb{C}^N$ of the same dimension i. Then the following statements are equivalent:*

a) There is a projection Π with $\operatorname{Im}\Pi = \mathcal{V}$ and $\operatorname{Ker}\Pi = \mathcal{W}^\perp$.

b) The matrix $[w_1, w_2, \ldots, w_i]^T[v_1, v_2, \ldots, v_i]$ is invertible for every choice of bases $v_1, v_2, \ldots, v_i$ of $\mathcal{V}$ and $w_1, w_2, \ldots, w_i$ of $\mathcal{W}$.

c) There are bases $v_1, v_2, \ldots, v_i$ of $\mathcal{V}$ and $w_1, w_2, \ldots, w_i$ of $\mathcal{W}$ such that the matrix $[w_1, w_2, \ldots, w_i]^T[v_1, v_2, \ldots, v_i]$ is invertible.

d) The bilinear form $\langle\cdot,\cdot\rangle|_{\mathcal{W}\times\mathcal{V}}$ is non-degenerate.

Proof. The implication a) $\Longrightarrow$ b) is a direct consequence of Lemma 2.2. b) $\Longrightarrow$ c) is evident. Defining $\Pi := V(W^TV)^{-1}W^T$ where $V := [v_1, v_2, \ldots, v_i]$ and $W := [w_1, w_2, \ldots, w_i]$ the implication c) $\Longrightarrow$ a) is proved. Finally we prove the equivalence of c) and d). Let $x \in \mathcal{W}$ and $y \in \mathcal{V}$. Then there are numbers α_i and β_i such that

$$x = \sum_{k=1}^{i} \alpha_k w_k \qquad \text{and} \qquad y = \sum_{j=1}^{i} \beta_j v_j.$$

Considering the representation

$$\langle x, y\rangle|_{\mathcal{W}\times\mathcal{V}} = \sum_{j=1}^{i}\sum_{k=1}^{i} \alpha_k\beta_j\langle x, y\rangle = \alpha^T W^T V\beta$$

where $\alpha^T = (\alpha_1, \alpha_2, \ldots, \alpha_i)$ and $\beta^T = (\beta_1, \beta_2, \ldots, \beta_i)$ we find that $\langle\cdot,\cdot\rangle|_{\mathcal{W}\times\mathcal{V}}$ is non-degenerate if and only if

$$\alpha^T W^T V \neq 0 \qquad \text{and} \qquad W^T V\beta \neq 0$$

for all $\alpha, \beta \in \mathbb{C}^i \setminus \{0\}$. This is equivalent to the invertibility of W^TV. ∎

First we derive a necessary condition for the existence of a biorthogonal basis. Let $((w_i, v_i))_{1\leq i\leq s}$ be a biorthogonal basis. Set

$$W^{(i)} := [w_1, w_2, \ldots, w_i] \qquad \text{and} \qquad V^{(i)} := [v_1, v_2, \ldots, v_i].$$

Then, by definition, the matrix $D^{(i)} = (W^{(i)})^TV^{(i)}$ is invertible and diagonal. Consequently, by Lemma 2.2, the Gram-Schmidt projection $\Pi_i = V^{(i)}(D^{(i)})^{-1}(W^{(i)})^T$ with

$$\operatorname{Im}\Pi_i = \mathcal{V}_i \qquad \text{and} \qquad \operatorname{Ker}\Pi_i = \mathcal{W}_i^\perp \tag{2.11}$$

exists for all $i \in \{1, 2, \ldots, s\}$. Next we show that the Gram-Schmidt biorthogonalisation method applied to an arbitrary basis $((\tilde{w}_i, \tilde{v}_i))_{1\leq i\leq s}$ of $((\mathcal{W}_i, \mathcal{V}_i))_{1\leq i\leq s}$ does not break down if the Gram-Schmidt projections defined by (2.11) exist for all $i \in \{1, 2, \ldots, s\}$.

Gram-Schmidt biorthogonalisation method

$v_1 := \tilde{v}_1$ and $w_1 := \tilde{w}_1$

for $i = 2, 3, \ldots, s$

$$v_i := \tilde{v}_i - \sum_{j=1}^{i-1} \frac{\langle w_j, \tilde{v}_i \rangle}{\langle w_j, v_j \rangle} v_j = (I - \Pi_{i-1})\tilde{v}_i \tag{2.12}$$

$$w_i := \tilde{w}_i - \sum_{j=1}^{i-1} \frac{\langle w_i, \tilde{v}_j \rangle}{\langle w_j, v_j \rangle} w_j = (I - \Pi_{i-1}^T)\tilde{w}_i. \tag{2.13}$$

The Gram-Schmidt biorthogonalisation method breaks down after the i-th step if and only if

$$\langle w_1, v_1 \rangle \neq 0, \quad \langle w_2, v_2 \rangle \neq 0, \quad \ldots, \quad \langle w_{i-1}, v_{i-1} \rangle \neq 0 \tag{2.14}$$

and

$$\langle w_i, v_i \rangle = 0. \tag{2.15}$$

Thus, the breakdown occurs if and only if i is the smallest number such that the matrix

$$[w_1, w_2, \ldots, w_i]^T [v_1, v_2, \ldots, v_i]$$

is not invertible.
Consequently, by Lemma 2.3, the Gram-Schmidt biorthogonalisation method breaks down after the i-th step if and only if the Gram-Schmidt projections $\Pi_1, \Pi_2, \ldots, \Pi_{i-1}$ exist and the projection Π_i does not exist.

These observations are summarized in Lemma 2.4 below.

Lemma 2.4 (Existence of biorthogonal bases) *The following statements are equivalent:*

a) There is a biorthogonal basis of $((\mathcal{W}_i, \mathcal{V}_i))_{1 \leq i \leq s}$.

b) The Gram-Schmidt projections $\Pi_1, \Pi_2, \ldots, \Pi_s$ *exist.*

c) The Gram-Schmidt biorthogonalisation method applied to an arbitrary basis of the chain $((\mathcal{W}_i, \mathcal{V}_i))_{1 \leq i \leq s}$ *generates a biorthogonal basis.*

d) The matrix $[w_1, w_2, \ldots, w_j]^T [v_1, v_2, \ldots, v_j]$ *is invertible for every choice of bases* $((w_i, v_i))_{1 \leq \imath \leq s}$ *of* $((\mathcal{W}_i, \mathcal{V}_i))_{1 \leq i \leq s}$ *and for all* $j \in \{1, 2, \ldots, s\}$.

e) There is a basis $((w_i, v_i))_{1 \leq \imath \leq s}$ *of* $((\mathcal{W}_i, \mathcal{V}_i))_{1 \leq i \leq s}$ *such that the matrix* $[w_1, w_2, \ldots, w_j]^T [v_1, v_2, \ldots, v_j]$ *is invertible for all* $j \in \{1, 2, \ldots, s\}$.

f) The bilinear form $\langle \cdot, \cdot \rangle|_{\mathcal{W}_j \times \mathcal{V}_j}$ *is non-degenerate for all* $j \in \{1, 2, \ldots, s\}$.

Now we specialize to the generation of biorthogonal bases in Krylov chains. In a similar way as we have derived the Arnoldi method from the Gram-Schmidt orthonormalisation method we now derive the Lanczos method from the Gram-Schmidt biorthogonalisation method.

Let two starting vectors $v_1, w_1 \in \mathbb{C}^N \setminus \{0\}$ be given.

Algorithm 2.1

For $i = 1, 2, \ldots, s-1$

$$v_{i+1} := Av_i - \sum_{j=1}^{i} \frac{\langle w_j, Av_i \rangle}{\langle w_j, v_j \rangle} v_j = (I - \Pi_i) A v_i \tag{2.16}$$

$$w_{i+1} := A^T w_i - \sum_{j=1}^{i} \frac{\langle A^T w_i, v_j \rangle}{\langle w_j, v_j \rangle} w_j = (I - \Pi_i^T) A^T w_i. \tag{2.17}$$

Algorithm 2.1 is the Gram-Schmidt biorthogonalisation method applied to the basis

$$\begin{array}{l} v_1, Av_1, Av_2, \ldots, Av_{s-1} \\ w_1, A^T w_1, A^T w_2, \ldots, A^T w_{s-1} \end{array}$$

of the Krylov chain $((\mathcal{K}(i, A^T, w_1), \mathcal{K}(i, A, v_1)))_{1 \le i \le s}$. Thus, by Lemma 2.4, Algorithm 2.1 generates the up to scalars uniquely determined biorthogonal basis of the Krylov chain $((\mathcal{K}(i, A^T, w_1), \mathcal{K}(i, A, v_1)))_{1 \le i \le s}$.

Now we show that the recursions (2.16) and (2.17) can be reduced to three-term recursions. Indeed, the relations

$$\begin{aligned} A^T w_j \in A^T \mathcal{K}(j, A^T, w_1) &= A^T \mathrm{span}\{w_1, w_2, \ldots, w_j\} \\ &\subseteq \mathrm{span}\{w_1, w_2, \ldots, w_{j+1}\} \end{aligned}$$

and

$$v_i \perp \mathrm{span}\{w_1, w_2, \ldots, w_{j+1}\} \qquad \text{for} \quad i \ge j+2$$

imply

$$\langle w_j, Av_i \rangle = \langle A^T w_j, v_i \rangle = 0 \qquad \text{for} \quad i \ge j+2.$$

By duality

$$\langle A^T w_i, v_j \rangle = \langle w_i, Av_j \rangle = 0 \qquad \text{for} \quad i \ge j+2.$$

Hence the recursions (2.16) and (2.17) reduce to the three-term recursions

$$\begin{aligned} v_{i+1} &= Av_i - \frac{\langle w_i, Av_i \rangle}{\langle w_i, v_i \rangle} v_i - \frac{\langle w_{i-1}, Av_i \rangle}{\langle w_{i-1}, v_{i-1} \rangle} v_{i-1} \\ w_{i+1} &= A^T w_i - \frac{\langle A^T w_i, v_i \rangle}{\langle w_i, v_i \rangle} w_i - \frac{\langle A^T w_i, v_{i-1} \rangle}{\langle w_{i-1}, v_{i-1} \rangle} w_{i-1}. \end{aligned}$$

Using that

$$\begin{aligned}\langle w_{i-1}, Av_i\rangle &= \langle A^T w_{i-1}, v_i\rangle = \langle A^T w_{i-1}, (I-\Pi_{i-1})v_i\rangle \\ &= \langle (I-\Pi_{i-1})^T A^T w_{i-1}, v_i\rangle \\ &= \langle w_i, v_i\rangle\end{aligned}$$

and (by duality)

$$\langle A^T w_i, v_{i-1}\rangle = \langle w_i, v_i\rangle$$

we obtain

$$\begin{aligned}v_{i+1} &= Av_i - \alpha_i v_i - \beta_i v_{i-1} \\ w_{i+1} &= A^T w_i - \alpha_i w_i - \beta_i w_{i-1}\end{aligned}$$

with

$$\beta_1 = 0, \qquad v_0 = w_0 = 0$$

$$\alpha_i = \frac{\langle w_i, Av_i\rangle}{\langle w_i, v_i\rangle} \qquad \text{and} \qquad \beta_i = \frac{\langle w_i, v_i\rangle}{\langle w_{i-1}, v_{i-1}\rangle}.$$

These are the famous Lanczos three-term recursions.
Suppose that there is a biorthogonal basis of the chain $((\mathcal{K}(i, A^T, w_1), \mathcal{K}(i, A, v_1)))_{1\leq i\leq s}$. Then the Lanczos recursions generate a biorthogonal basis $((w_i, v_i))_{1\leq i\leq s}$ of this chain. The Lanczos method further generates implicitly for every $j \in \{1, 2, \ldots, s\}$ the tridiagonal matrix

$$H^{(j)} = \begin{bmatrix} \alpha_1 & \beta_2 & & 0 \\ 1 & \alpha_2 & \ddots & \\ & \ddots & \ddots & \beta_j \\ 0 & & 1 & \alpha_j \end{bmatrix}$$

satisfying the equations

$$V^{(j)}H^{(j)} = \Pi_j A V^{(j)} \qquad \text{and} \qquad W^{(j)}H^{(j)} = \Pi_j^T A^T W^{(j)} \tag{2.18}$$

where $V^{(j)} := [v_1, v_2, \ldots, v_j]$, $W^{(j)} := [w_1, w_2, \ldots, w_j] \in \mathrm{Mat}_{N\times j}$ and

$$\Pi_j = V^{(j)}((W^{(j)})^T V^{(j)})^{-1}(W^{(j)})^T$$

is the Gram-Schmidt projection with

$$\operatorname{Im}\Pi_j = \mathcal{K}(j, A, v_1) \qquad \text{and} \qquad \operatorname{Ker}\Pi_j = \mathcal{K}(j, A^T, w_1)^\perp.$$

Remark 2.2 The equations (2.18) can be rewritten in the more familiar form

$$V^{(j)}H^{(j)} = AV^{(j)} - v_{j+1}e_j^T \tag{2.19}$$

$$W^{(j)}H^{(j)} = A^T W^{(j)} - w_{j+1}e_j^T. \tag{2.20}$$

2.2.1 Breakdowns of the Lanczos method

In the following we use the notation

$$m := \dim \mathcal{K}(N, A^T, w_1), \qquad \hat{m} := \dim \mathcal{K}(N, A, v_1)$$

and we define for every $l \in \mathbb{N}$ the Hankel matrix Γ_l by

$$\Gamma_l := [\langle w_1, A^{i+j} v_1 \rangle]_{0 \le i,j \le l-1}.$$

The Lanczos method terminates after step s_L if

$$\langle w_1, v_1 \rangle \neq 0, \langle w_2, v_2 \rangle \neq 0, \ldots, \langle w_{s_L}, v_{s_L} \rangle \neq 0 \tag{2.21}$$

and

$$\langle w_{s_L+1}, v_{s_L+1} \rangle = 0. \tag{2.22}$$

Furtheron s_L denotes the termination index of the Lanczos method. The breakdowns of the Lanczos method can be classified as follows:

1) *Regular termination:* ($v_{s_L+1} = 0$ or $w_{s_L+1} = 0$)
If $v_{s_L+1} = 0$ then by equation (2.19) the subspace $\text{span}\{v_1, v_2, \ldots, v_{s_L}\} = \mathcal{K}(s_L, A, v_1)$ is A-invariant and the matrix $H^{(s_L)}$ represents the restriction of A to this subspace. $\mathcal{K}(N, A, v_1)$ is the smallest A-invariant subspace containing v_1. Hence we can conclude

$$\mathcal{K}(N, A, v_1) = \mathcal{K}(s_L, A, v_1).$$

Hence the equality $s_L = m$ holds. Similarly, if $w_{s_L+1} = 0$, then by equation (2.20) the subspace $\text{span}\{w_1, w_2, \ldots, w_{s_L}\} = \mathcal{K}(s_L, A^T, w_1)$ is A^T-invariant and $H^{(s_L)}$ is the restriction of A^T to this subspace. Further the identity $s_L = \hat{m}$ holds.

2) *Serious breakdown:* If the termination condition (2.22) is fulfilled with $v_{s_L+1} \neq 0$ and $w_{s_L+1} \neq 0$, then the breakdown is called 'serious'. Notice that in this case the Lanczos vectors do not span an A-invariant nor an A^T-invariant subspace of $\mathbb{C}^N$. There are two different cases of a serious breakdown:

   2.1) *Curable breakdown:* If there is a $k \geq 1$ such that the Hankel matrix Γ_{k+s_L} is invertible then the breakdown is called 'curable'. In general the eigenvalues of the generated tridiagonal matrix $H^{(s_L)}$ are not eigenvalues of A, however some eigenvalues of A are approximated (see Section 5.4).

   2.2) *Incurable breakdown:* If the Hankel matrix Γ_{k+s_L} is not invertible for all $k \in \mathbb{N}$ then the breakdown is called 'incurable'. In this case each eigenvalue of $H^{(s_L)}$ is an eigenvalue of A (see Theorem 5.12).

The possibility of a curable breakdown has brought the Lanczos method into discredit. How can we overcome such a breakdown ? By Lemma 2.4 a serious breakdown occurs if and only if there exists no biorthogonal basis in the Krylov chain

$$((\mathcal{K}(i, A^T, w_1), \mathcal{K}(i, A, v_1)))_{1 \leq i \leq m_L}$$

where

$$m_L := \min\{m, \hat{m}\}. \tag{2.23}$$

Hence it is natural to develop an appropriate concept of a generalized biorthogonal basis which even exists in the case of an curable breakdown. This is done in the following chapter.

3 Exact look-ahead Lanczos methods and generalized biorthogonal bases

Chapter 3 follows the lines of [81] presented by the author in 1997.

3.1 Generalized biorthogonal bases of chains of subspaces

In the following let $((\mathcal{W}_i, \mathcal{V}_i))_{1\le i\le s}$ be an arbitrary chain of subspaces of $\mathbb{C}^N$.

Definition 3.1 (Generalized biorthogonal bases) *A basis $((w_i, v_i))_{1\le i\le s}$ of the chain $((\mathcal{W}_i, \mathcal{V}_i))_{1\le i\le s}$ is called generalized biorthogonal, if $D := [\langle w_i, v_j\rangle]_{1\le i,j\le s}$ is an invertible block diagonal matrix with maximal number of diagonal blocks.*

In this context "maximal" means: If $((\tilde w_i, \tilde v_i))_{1\le i\le s}$ is a basis of $((\mathcal{W}_i, \mathcal{V}_i))_{1\le i\le s}$ and $\tilde D := [\langle \tilde w_i, \tilde v_j\rangle]_{1\le i,j\le s}$ is invertible, then the number of diagonal blocks of $\tilde D$ is not greater than the number of diagonal blocks in D.

This definition in fact generalizes biorthogonality because if there exists a conventional biorthogonal basis in $((\mathcal{W}_i, \mathcal{V}_i))_{1\le i\le s}$ then any basis of this chain is in the conventional sense biorthogonal if and only if it is generalized biorthogonal.
Now we turn to the existence and construction of generalized biorthogonal bases.

The extended Gram-Schmidt biorthogonalisation method below was introduced by Parlett [58] in 1992.

Extended Gram-Schmidt biorthogonalisation method (EGS)

$k := 0$; $\nu(0) := 0$; $\Pi_0 := 0$;
repeat

$v_{\nu(k)+1} := (I - \Pi_{\nu(k)})\tilde v_{\nu(k)+1}$;
$w_{\nu(k)+1} := (I - \Pi^T_{\nu(k)})\tilde w_{\nu(k)+1}$;
if $v_{\nu(k)+1} = 0$ **or** $w_{\nu(k)+1} = 0$ **then** *stop* := true
else

$V_k := [v_{\nu(k)+1}]$; $W_k := [w_{\nu(k)+1}]$; $D_k := V_k^T W_k$;
$i := 1$;
while D_k not invertible **and** $\nu(k) + i < s$ **do**

$i := i + 1$;
$v_{\nu(k)+i} := (I - \Pi_{\nu(k)})\tilde v_{\nu(k)+i}$;
$w_{\nu(k)+i} := (I - \Pi^T_{\nu(k)})\tilde w_{\nu(k)+i}$;
$V_k := [V_k, v_{\nu(k)+1}]$; $W_k := [W_k, w_{\nu(k)+1}]$; $D_k := V_k^T W_k$;

if D_k is invertible **then**

$\nu(k+1) := \nu(k) + i;$
$k := k + 1;$

until $\nu(k) + i \geq s$ **or** *stop*.
$b := k;$

Here

$$\Pi_{\nu(k)} = \sum_{j=1}^{k} V_j D_j^{-1} W_j^T$$

is the Gram-Schmidt projection with

$$\operatorname{Im} \Pi_{\nu(k)} = \mathcal{V}_{\nu(k)} \qquad \text{and} \qquad \operatorname{Ker} \Pi_{\nu(k)} = \mathcal{W}_{\nu(k)}^{\perp}.$$

By construction and Lemma 2.3 the *EGS* method applied to a basis $((\tilde{w}_i, \tilde{v}_i))_{1 \leq i \leq s}$ of $((\mathcal{W}_i, \mathcal{V}_i))_{1 \leq i \leq s}$ generates

- the degree indices $(\nu(k))_{k \in K}$ of $((\mathcal{W}_i, \mathcal{V}_i))_{1 \leq i \leq s}$ which are defined as those indices $l = \nu(0) := 0 < \nu(1) < \nu(2) < \ldots < \nu(b)$ for which the bilinear form $\langle \cdot, \cdot \rangle \mid_{W_l \times V_l}$ is non-degenerate
- a basis $((w_i, v_i))_{1 \leq i \leq \nu(b)}$ of $((\mathcal{W}_i, \mathcal{V}_i))_{1 \leq i \leq \nu(b)}$ with the property:

$$D^{(\nu(b))} := [W_1, W_2, \ldots, W_b]^T [V_1, V_2, \ldots, V_b] = \operatorname{diag}(D_1, D_2, \ldots, D_b) \tag{3.1}$$

where

$$V_k := [v_{\nu(k-1)+1}, v_{\nu(k-1)+2}, \ldots, v_{\nu(k)}] \tag{3.2}$$
$$W_k := [w_{\nu(k-1)+1}, w_{\nu(k-1)+2}, \ldots, w_{\nu(k)}] \tag{3.3}$$
$$D_k := W_k^T V_k \tag{3.4}$$

for all $k \in \{1, 2, \ldots, b\}$.

Now we show that the *EGS* method generates a generalized biorthogonal basis if there exists one. Suppose there is a generalized biorthogonal basis $((\hat{w}_i, \hat{v}_i))_{1 \leq i \leq s}$ of $((\mathcal{W}_i, \mathcal{V}_i))_{1 \leq i \leq s}$. Then by definiton there are indices

$$l = \hat{\nu}(0) := 0 < \hat{\nu}(1) < \hat{\nu}(2) < \ldots < \hat{\nu}(\hat{b})$$

and invertible matrices $\hat{D}_i \in \operatorname{Mat}_{\hat{\nu}(i) - \hat{\nu}(i-1)}$ such that

$$\hat{D}^{(\nu(k))} := [\hat{w}_1, \hat{w}_2, \ldots, \hat{w}_{\hat{\nu}(k)}]^T [\hat{v}_1, \hat{v}_2, \ldots, \hat{v}_{\hat{\nu}(k)}] = \operatorname{diag}(\hat{D}_i, \hat{D}_2, \ldots, \hat{D}_k)$$

for all $k \in \{1, 2 \ldots, \hat{b}\}$. Lemma 2.3 shows that the bilinear forms $\langle \cdot, \cdot \rangle|_{\mathcal{W}_{\hat{\nu}(k)} \times \mathcal{V}_{\hat{\nu}(k)}}$ are non-degenerate. Therefore the inclusion

$$(\hat{\nu}(k))_{0 \le k \le \hat{b}} \subseteq (\nu(k))_{0 \le k \le b} \tag{3.5}$$

holds. Comparing (3.1) and (3.5), we find that

$$(\hat{\nu}(k))_{0 \le k \le \hat{b}} = (\nu(k))_{0 \le k \le b}.$$

Thus, we have proved the following theorem.

Theorem 3.1 (Existence and size)

A) Existence: *The following statements are equivalent:*
A.1) There is a generalized biorthogonal basis of $((\mathcal{W}_i, \mathcal{V}_i))_{1 \le i \le s}$.
A.2) The extendend Gram-Schmidt biorthogonalisation method applied to an arbitrary basis of the chain $((\mathcal{W}_i, \mathcal{V}_i))_{1 \le i \le s}$ *generates a generalized biorthogonal basis.*
A.3) The bilinear form $\langle \cdot, \cdot \rangle \,|_{\mathcal{W}_s \times \mathcal{V}_s}$ *is non-degenerate.*

B) Size of the diagonal blocks: *Let* $((w_i, v_i))_{1 \le i \le s}$ *be a generalized biorthogonal basis of* $((\mathcal{W}_i, \mathcal{V}_i))_{1 \le i \le s}$. *Then the maximal number* c *of diagonal blocks of the matrix* $D := [\langle w_i, v_j \rangle]_{1 \le i,j \le s} = diag\,(D_1, D_2, \ldots, D_c)$ *is equal to* b, *the argument of the maximal degree index. The* k*-th diagonal block* D_k *has the size* $\nu(k) - \nu(k-1)$ *for all* $1 \le k \le c$.

Theorem 3.2 states the degrees of freedom in the choice of generalized biorthogonal bases and explains how every generalized biorthogonal basis can be constructed from any given generalized biorthogonal basis. By Theorem 3.1 the *EGS* generates a generalized biorthogonal basis. Throughout the rest of Section 3 let b denote the argument of the maximal degree index and $dim_k := \nu(k) - \nu(k-1)$.

Theorem 3.2 (Degrees of freedom) *Let* $((w_i, v_i))_{1 \le i \le s}$ *be a generalized biorthogonal basis of* $((\mathcal{W}_i, \mathcal{V}_i))_{1 \le i \le s}$. *Let* $((\tilde{w}_i, \tilde{v}_i))_{1 \le i \le s}$ *be a basis of* $((\mathcal{W}_i, \mathcal{V}_i))_{1 \le i \le s}$. *Then the sequence* $((\tilde{w}_i, \tilde{v}_i))_{1 \le i \le s}$ *is generalized biorthogonal if and only if the following relations hold for all* $0 \le k \le b-1$, $1 \le i \le dim_{k+1}$:

$$\begin{aligned} \tilde{v}_{\nu(k)+i} &\in \operatorname{span}\{v_{\nu(k)+1}, v_{\nu(k)+2}, \ldots, v_{\nu(k)+i}\} \\ \tilde{w}_{\nu(k)+i} &\in \operatorname{span}\{w_{\nu(k)+1}, w_{\nu(k)+2}, \ldots, w_{\nu(k)+i}\}. \end{aligned}$$

Proof. By assumption $\tilde{v}_{\nu(k)+i} \in \mathcal{V}_{\nu(k)+i} = \operatorname{span}\{v_1, v_2, \ldots, v_{\nu(k)+i}\}$. Hence there exist coefficients $a_j \in \mathbb{C}$ such that $\tilde{v}_{\nu(k)+i} = \sum_{j=1}^{\nu(k)+i} a_j v_j$. Theorem 3.1 provides the desired identity

$$\tilde{v}_{\nu(k)+i} = (I - \Pi_{\nu(k)})\tilde{v}_{\nu(k)+i} \;=\; \sum_{j=1}^{\nu(k)+i} a_j (I - \Pi_{\nu(k)}) v_j = \sum_{j=\nu(k)+1}^{\nu(k)+i} a_j v_j.$$

Replacing v with w, $\mathcal{V}$ with $\mathcal{W}$ and Π with Π^T, one gets the dual statement

$$\tilde{w}_{\nu(k)+i} \in \operatorname{span}\{w_{\nu(k)+1}, w_{\nu(k)+2}, \ldots, w_{\nu(k)+i}\}.$$

■

3.2 Generalized biorthogonal bases of Krylov chains

In this section we specialize to generalized biorthogonal bases of Krylov chains.

3.2.1 Notation

- $(\nu(k))_{k\in K}$ denotes the degree indices of $((\mathcal{K}(i, A^T, w_1), \mathcal{K}(i, A, v_1)))_{1\le i\le N}$.
- $r := \nu(b) = \max\{\nu(k) \mid k \in K\}$.
- $m_L := \min\{\dim \mathcal{K}(N, A, v_1), \dim \mathcal{K}(N, A^T, w_1)\}$.
- If $((w_i, v_i))_{1\le i\le s}$ is a basis of $((\mathcal{K}(i, A^T, w_1), \mathcal{K}(i, A, v_1)))_{1\le i\le s}$ then define

  $$V^{(i)} := [v_1, v_2, \ldots, v_i], \qquad W^{(i)} := [w_1, w_2, \ldots, w_i], \qquad D^{(i)} := (W^{(i)})^T V^{(i)}$$

  $$V_k := [v_{\nu(k-1)+1},\ v_{\nu(k-1)+2},\ \ldots,\ v_{\nu(k)}], \qquad W_k := [w_{\nu(k-1)+1},\ w_{\nu(k-1)+2},\ \ldots,\ w_{\nu(k)}],$$

  and $D_k := (W_k)^T V_k$.
- For $l \le 0$ let $w_l := 0$, $v_l := 0$, $D_l := I$, $W_l := 0$, and $V_l := 0$.

3.2.2 Construction and properties of generalized biorthogonal bases of Krylov chains

To get a generalized biorthogonal basis of $((\mathcal{W}_i, \mathcal{V}_i))_{1\le i\le r}$ the EGS can be applied to any basis of $((\mathcal{W}_i, \mathcal{V}_i))_{1\le i\le r}$ (see Theorem 3.1). For Krylov chains it seems quite natural to choose $(((A^T)^{i-1}w_1, A^{i-1}v_1))_{1\le i\le r}$ as a basis of $((\mathcal{K}(i, A^T, w_1), \mathcal{K}(i, A, v_1)))_{1\le i\le r}$. But on reconsidering the conventional biorthogonal case, one would prefer another basis. Indeed, let $((w_i, v_i))_{1\le i\le s_L}$ be the uniquely determined (up to scalar factors) biorthogonal basis of $((\mathcal{K}(i, A^T, w_1), \mathcal{K}(i, A, v_1)))_{1\le i\le s_L}$. Then the ordinary Gram-Schmidt biorthogonalisation method applied to the basis

$$(w_1, v_1),\ (A^T w_1, A v_1),\ (A^T w_2, A v_2), \ldots,\ (A^T w_{s_L-1}, A v_{s_L-1})$$

reduces to the ordinary Lanczos three-term recursion (see Section 2.2). To imitate this structure in the general case we define the 'monomial look-ahead Lanczos method' as follows:

Monomial look-ahead Lanczos method

For all $0 \leq k \leq b-1$
for all $0 \leq i \leq \nu(k+1) - \nu(k) - 1$ if $\nu(k) + i \neq 0$

$$\begin{aligned} v_{\nu(k)+i+1} &:= (I - \Pi_{\nu(k)}) A v_{\nu(k)+i} \\ w_{\nu(k)+i+1} &:= (I - \Pi^T_{\nu(k)}) A^T w_{\nu(k)+i}. \end{aligned}$$

Lemma 3.1 *The monomial look-ahead Lanczos method generates a generalized biorthogonal basis* $((w_i, v_i))_{1 \leq i \leq r}$ *of* $((\mathcal{K}(i, A^T, w_1), \mathcal{K}(i, A, v_1)))_{1 \leq i \leq r}$.

Evidently Lemma 3.1 is a consequence of the definition of the monomial look-ahead Lanczos method.

The following example shows that, in contrast to the biorthogonal case, the monomial look-ahead Lanczos method need not generate the same basis as the EGS, applied to the Krylov sequence $(((A^T)^{i-1} w_1, A^{i-1} v_1))_{1 \leq i \leq r}$.

Example 3.1

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad w_1 = v_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

The matrices on the left are generated by the EGS applied to $(((A^T)^{i-1} w_1, A^{i-1} v_1))_{1 \leq i \leq s}$. The matrices on the right are computed by the monomial look-ahead Lanczos method.

$$\begin{aligned} \tilde{V}^{(4)} &= [\tilde{v}_1, \tilde{v}_2, \tilde{v}_3, \tilde{v}_4] & V^{(4)} &= [v_1, v_2, v_3, v_4] \\ &= \begin{bmatrix} 1 & 1/4 & 0 & 0 \\ 1 & 1/4 & 0 & -1 \\ 1 & 1/4 & -1 & -1 \\ 1 & -3/4 & 1 & 2 \end{bmatrix} & &= \begin{bmatrix} 1 & 1/4 & 0 & 0 \\ 1 & 1/4 & 0 & -1 \\ 1 & 1/4 & -1 & 1 \\ 1 & -3/4 & 1 & 0 \end{bmatrix} \end{aligned}$$

$$\tilde{v}_4 \neq \gamma v_4 \quad \text{for all} \quad \gamma \in \mathbb{C}$$

$$\tilde{D}^{(4)} = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & -1/4 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix} \qquad D^{(4)} = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & -1/4 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & -2 \end{bmatrix}.$$

Lemma 3.2 (Properties of the monomial Lanczos vectors)
The monomial look-ahead Lanczos method generates a generalized biorthogonal basis $((w_i, v_i))_{1\le i\le r}$ *of* $((\mathcal{K}(i, A^T, w_1), \mathcal{K}(i, A, v_1)))_{1\le i\le r}$ *with the following properties:*

a) Diagonal blocks: *The diagonal blocks* D_k *of the matrix*

$$D^{(r)} := (\langle w_i, v_j\rangle)_{1\le i,j\le r} = diag\,(D_1, D_2, \ldots, D_b)$$

are lower antitriangular Hankel matrices

$$D_k = \begin{bmatrix} & & & * \\ & 0 & \cdot & \\ & \cdot & & \\ * & & & \end{bmatrix}.$$

b) Inner Vectors:

For all $0 \le k \le b-1$

For all $0 \le i \le \dim_{k+1} -1$ *the following identities hold:*

$$v_{\nu(k)+i+1} = A^i v_{\nu(k)+1}\,, \qquad w_{\nu(k)+i+1} = (A^T)^i w_{\nu(k)+1}.$$

c) Incurable breakdown: *An incurable breakdown occurs after step* j *if and only if*

$$\langle w_{j+1}, A^i v_{j+1}\rangle = 0 \qquad \textit{for all} \quad i \in \mathbb{N}$$

or, equivalently, if $j = r := \max\{\nu(k)|k \in K\} < m_L$.

Proof. First we present and proof the following three statements A,B,C:
A) If the following equations are valid for all $j \in \{0, 1, \ldots, i\}$

$$\begin{aligned} v_{\nu(k)+j+1} &= A^j v_{\nu(k)+1} \\ w_{\nu(k)+j+1} &= (A^T)^j w_{\nu(k)+1} \\ \langle p_{\nu(k)+j+1}, q_{\nu(k)+1}\rangle &= 0, \end{aligned}$$

then the identities

$$A^{i+1} v_{\nu(k)+1} = (I - \Pi_{\nu(k)}) A^{i+1} v_{\nu(k)+1} \tag{3.6a}$$
$$(A^T)^{i+1} w_{\nu(k)+1} = (I - \Pi^T_{\nu(k)})(A^T)^{i+1} w_{\nu(k)+1} \tag{3.6b}$$

hold.
Proof of A). (3.6a) and (3.6b) are dual statements. Therefore it is sufficient to show (3.6a). Identity (3.6a) is a consequence of following relations:

i)

$$\begin{aligned} & v_{\nu(k)+1} \in \{(A^T)^{i+1}\mathcal{K}(\nu(k), A^T, w_1)\}^\perp \\ \iff \quad & A^{i+1}v_{\nu(k)+1} \in \mathcal{K}(\nu(k), A^T, w_1)^\perp = \operatorname{Ker}\Pi_{\nu(k)} \\ \iff \quad & A^{i+1}v_{\nu(k)+1} = (I - \Pi_{\nu(k)})A^{i+1}v_{\nu(k)+1} \end{aligned}$$

ii)

$$\begin{aligned} (A^T)^{i+1}\mathcal{K}(\nu(k), A^T, w_1) &= \operatorname{span}\{(A^T)^{i+1}w_1, (A^T)^{i+2}w_1, \ldots, (A^T)^{\nu(k)+i}w_1, \} \\ &\subseteq \mathcal{K}(\nu(k)+i+1, A^T, w_1) \\ &= \operatorname{span}\{w_1, w_2, \ldots, w_{\nu(k)+i+1}\} \end{aligned}$$

iii)

$$v_{\nu(k)+1} \in \mathcal{K}(\nu(k), A^T, w_1)^\perp = \operatorname{span}\{w_1, w_2, \ldots, w_{\nu(k)}\}^\perp.$$

B) The following implications hold:

$$\left.\begin{array}{l} A^{i+1}v_{\nu(k)+1} = (I - \Pi_{\nu(k)})A^{i+1}v_{\nu(k)+1} \\ A^i v_{\nu(k)+1} = v_{\nu(k)+i+1} \end{array}\right\} \Rightarrow A^{i+1}v_{\nu(k)+1} = v_{\nu(k)+i+2}$$

$$\left.\begin{array}{l} (A^T)^{i+1}w_{\nu(k)+1} = (I - \Pi^T_{\nu(k)})(A^T)^{i+1}w_{\nu(k)+1} \\ (A^T)^i w_{\nu(k)+1} = w_{\nu(k)+i+1} \end{array}\right\} \Rightarrow (A^T)^{i+1}w_{\nu(k)+1} = w_{\nu(k)+i+2}.$$

C) Let $k+1 \in K$ and $i+3 \leq \nu(k+1) - \nu(k)$. If the equations

$$v_{\nu(k)+j+1} = A^j v_{\nu(k)+1} \quad \text{and} \quad w_{\nu(k)+j+1} = (A^T)^j w_{\nu(k)+1},$$

hold for all $j \in \{1, 2, \ldots, i+1\}$ then $\langle w_{\nu(k)+1+l}, v_{\nu(k)+1}\rangle$ vanishes for all

$$l \in \{0, 1, 2, \ldots, i+1\}.$$

Proof of C). Assuming there is an $l \in \{0, 1, 2, \ldots, i+1\}$ such that

$$\langle w_{\nu(k)+1+l}, v_{\nu(k)+1}\rangle \neq 0,$$

then there exists a minimal $l \in \{1, 2, \ldots, i+1\}$ with this property. For this l the Hankel matrix

$$D^l_k := (\langle w_i, v_j\rangle)_{\nu(k)+1 \leq i,j \leq \nu(k)+1+l}$$

is invertible because all entries above the antidiagonal are vanishing and all entries on the antidiagonal are different from zero. The invertibility of D_k^l guarantees the non-degeneracy of the bilinear form

$$\langle \cdot, \cdot \rangle \mid_{\mathcal{K}(\nu(k)+l+1, A^T, w_1) \times \mathcal{K}(\nu(k)+l+1, A, v_1)} .$$

Therefore the identity

$$l + 1 = \nu(k+1) - \nu(k)$$

holds which contradicts the assumption.

After these preliminary steps Lemma 3.2 can be proved easily by induction on

$$i \in \{0, 1, \ldots, \nu(k+1) - \nu(k) - 3\}.$$

For $i = 0$ we have: If $\nu(k+1) - \nu(k) > 1$, then $\langle w_{\nu(k)+1}, v_{\nu(k)+1} \rangle = 0$

$$\begin{aligned} v_{\nu(k)+1} &= A^0 v_{\nu(k)+1} \\ w_{\nu(k)+1} &= (A^T)^0 w_{\nu(k)+1}. \end{aligned}$$

Assume the statement

$$\begin{aligned} v_{\nu(k)+j+1} &= A^j v_{\nu(k)+1} \\ w_{\nu(k)+j+1} &= (A^T)^j w_{\nu(k)+1} \\ \langle w_{\nu(k)+j+1}, v_{\nu(k)+1} \rangle &= 0 \end{aligned}$$

is true for some $i \geq 0$ and all $j \in \{0, 1, \ldots, i\}$.
Now proceed to $i + 1$: By the induction hypothesis and statement A) we have

$$\begin{aligned} A^{i+1} v_{\nu(k)+1} &= (I - \Pi_{\nu(k)}) A^{i+1} v_{\nu(k)+1} \\ (A^T)^{i+1} w_{\nu(k)+1} &= (I - \Pi_{\nu(k)}^T)(A^T)^{i+1} w_{\nu(k)+1}. \end{aligned}$$

From B) and the induction hypothesis it follows that

$$\begin{aligned} v_{\nu(k)+i+2} &= A^{i+1} v_{\nu(k)+1} \\ w_{\nu(k)+i+2} &= (A^T)^{i+1} w_{\nu(k)+1}. \end{aligned}$$

Finally, statement C) together with the induction hypothesis yields

$$\langle w_{\nu(k)+j+1}, v_{\nu(k)+1} \rangle = 0$$

for all $j \in \{0, 1, \ldots, i+1\}$ as long as $i + 2 < \nu(k+1) - \nu(k)$. ∎

This result can be easily transferred to arbitrary generalized biorthogonal bases. Lemma 3.2 essentially reduces the construction of generalized biorthogonal bases of Krylov chains to the calculation of the vectors $w_{\nu(k)+1}$ and $v_{\nu(k)+1}$, called 'regular vectors'. The next theorem shows that the construction of $w_{\nu(k)+1}$ and $v_{\nu(k)+1}$ can be reduced to the 'look-ahead Lanczos' block-three-term recursions proposed by Gragg [30], Draux [18] and Gutknecht [33].

Theorem 3.3 (Construction) *Let $((w_i, v_i))_{1\le i\le r}$ be a generalized biorthogonal basis of the Krylov chain $((\mathcal{K}(i, A^T, w_1), \mathcal{K}(i, A, v_1)))_{1\le i\le r}$. Then there are coefficients $g_{i,j}^{(k)}, f_{i,j}^{(k)} \in \mathbb{C}$, $\beta_i^{(k)}, \alpha_i^{(k)} \in \mathbb{C}\setminus\{0\}$ such that the following identities hold:*
For all $0 \le k \le b-1$

$$v_{\nu(k)+1} = \beta_1^{(k)} \left(A v_{\nu(k)} - V_k (D_k)^{-1} (W_k)^T A v_{\nu(k)} - \psi_k \, v_{\nu(k-2)+1} \right) \tag{3.7}$$

$$w_{\nu(k)+1} = \alpha_1^{(k)} \left(A^T w_{\nu(k)} - W_k (D_k^T)^{-1} (V_k)^T A^T w_{\nu(k)} - \xi_k \, w_{\nu(k-2)+1} \right) \tag{3.8}$$

for all $1 \le i \le dim_{k+1} - 1$

$$v_{\nu(k)+i+1} = \beta_{i+1}^{(k)} \left(A v_{\nu(k)+i} - \sum_{l=1}^{i} g_{i+1,l}^{(k)} \, v_{\nu(k)+l} \right) \tag{3.9}$$

$$w_{\nu(k)+i+1} = \alpha_{i+1}^{(k)} \left(A^T w_{\nu(k)+i} - \sum_{l=1}^{i} f_{i+1,l}^{(k)} \, w_{\nu(k)+l} \right) \tag{3.10}$$

where for $k \ge 2$

$$\psi_k := \frac{\langle p_{\nu(k-1)+1}, q_{\nu(k)} \rangle}{\langle p_{\nu(k-1)}, q_{\nu(k-2)+1} \rangle} \cdot \frac{1}{\alpha_1^{(k-1)}}, \qquad \xi_k := \frac{\langle p_{\nu(k)}, q_{\nu(k-1)+1} \rangle}{\langle p_{\nu(k-2)+1}, q_{\nu(k-1)} \rangle} \cdot \frac{1}{\beta_1^{(k-1)}}$$

else $\psi_k := 0$ and $\xi_k := 0$.

Conversely, these recurrence relations generate a generalized biorthogonal basis for any choice of the coefficients $g_{i,j}^{(k)}, f_{i,j}^{(k)} \in \mathbb{C}$ and $\beta_i^{(k)}, \alpha_i^{(k)} \in \mathbb{C}\setminus\{0\}$.

Proof. Lemma 3.2 together with Theorem 3.2 yield the representations (3.9) and (3.10). (3.7) and (3.8) are dual statements. Therefore it is sufficient to show (3.7). There are scalars $\beta_{i+1}^{(k)} \in \mathbb{C}\setminus\{0\}$ such that for every k the following identities hold:

$$\begin{aligned}
\frac{1}{\beta_1^{(k)}} v_{\nu(k)+1} &= (I - \Pi_{\nu(k)}) A v_{\nu(k)} \\
&= (I - V^{(\nu(k))} (D^{\nu(k)})^{-1} (W^{\nu(k)})^T) A v_{\nu(k)} \\
&= \left(I - V_k (D_k)^{-1} (W_k)^T - V_{k-1} (D_{k-1})^{-1} (W_{k-1})^T \right) A v_{\nu(k)} \\
&\quad - \sum_{l=0}^{k-2} V_l (D_l)^{-1} (W_l)^T A v_{\nu(k)} \\
&= (I - V_k (D_k)^{-1} (W_k)^T - V_{k-1} (D_{k-1})^{-1} (W_{k-1})^T) A v_{\nu(k)} \\
&\quad - (I - \Pi_{\nu(k-2)}) A v_{\nu(k)}.
\end{aligned}$$

In addition we have

$$\begin{aligned} Av_{\nu(k)} \in A\mathrm{Ker}\,\Pi_{\nu(k-1)} &= A(\mathcal{K}(\nu(k-1), A^T, w_1))^\perp \\ &\subseteq A(\mathrm{span}\{A^T w_1, (A^T)^2 w_1, \ldots, (A^T)^{\nu(k-1)-1} w_1\})^\perp \\ &= \mathcal{K}(\nu(k-1)-1, A^T, w_1)^\perp \qquad (3.11) \\ &\subseteq \mathcal{K}(\nu(k-2), A^T, w_1)^\perp \\ &= \mathrm{Ker}\,\Pi_{\nu(k-2)}. \qquad (3.12) \end{aligned}$$

Therefore it remains to show that

$$\psi_k v_{\nu(k-2)+1} = V_{k-1}(D_{k-1})^{-1}(W_{k-1})^T A v_{\nu(k)}.$$

Theorem 3.2 and (3.11) yield

$$V_{k-1}(D_{k-1})^{-1}(W_{k-1})^T A v_{\nu(k)} = \sum_{i=\nu(k-2)+1}^{\nu(k-1)} a_i v_i$$

where the coefficients a_i are determined by the identity

$$\begin{aligned} (a_i)_{\nu(k-2)+1\le i\le\nu(k-1)} &= (D_{k-1})^{-1}(\langle w_i, Av_{\nu(k)}\rangle)_{\nu(k-2)+1\le i\le\nu(k-1)} \\ &= (D_{k-1})^{-1} e_{(\nu(k-1)-\nu(k-2))}\langle p_{\nu(k-1)}, Aq_{\nu(k)}\rangle \\ &\overset{(3.11)}{=} (D_{k-1})^{-1} e_{(\nu(k-1)-\nu(k-2))}\langle p_{\nu(k-1)}, q_{\nu(k-2)+1}\rangle\gamma_k \\ &= (D_{k-1})^{-1} D_{k-1} e_1 \gamma_k \\ &= e_1\gamma_k. \end{aligned}$$

Here

$$\gamma_k := \frac{\langle w_{\nu(k-1)}, Av_{\nu(k)}\rangle}{\langle w_{\nu(k-1)}, v_{\nu(k-2)+1}\rangle}.$$

Hence

$$V_{k-1}(D_{k-1})^{-1}(W_{k-1})^T A v_{\nu(k)} = \frac{\langle w_{\nu(k-1)}, Av_{\nu(k)}\rangle}{\langle w_{\nu(k-1)}, v_{\nu(k-2)+1}\rangle} v_{\nu(k-2)+1}.$$

Using

$$\begin{aligned} \langle w_{\nu(k-1)}, Av_{\nu(k)}\rangle &= \frac{1}{\alpha_1^{(k)}}\langle w_{\nu(k-1)+1}, v_{\nu(k)}\rangle + \langle \Pi_{\nu(k-1)}^T A^T w_{\nu(k-1)}, v_{\nu(k)}\rangle \\ &= \frac{1}{\alpha_1^{(k)}}\langle w_{\nu(k-1)+1}, v_{\nu(k)}\rangle + \langle A^T w_{\nu(k-1)}, \Pi_{\nu(k-1)} v_{\nu(k)}\rangle \end{aligned}$$

and $\Pi_{\nu(k-1)} v_{\nu(k)} = 0$, the assertion follows. ∎

Remark 3.1 (Ordinary Lanczos three-term recursion) If there exists a classical biorthogonal basis in $((\mathcal{K}(i, A^T, w_1), \mathcal{K}(i, A, v_1)))_{1\leq i\leq r}$, then the recursions in Theorem 3.3 represent the ordinary Lanczos three-term recursions.

Remark 3.2 (Breakdowns of the Lanczos and look-ahead Lanczos methods) In terms of the indices r, s_L and m_L the breakdowns of the Lanczos and the look-ahead Lanczos methods can be classified in a simple way. First we recall the definitions of the indices s_L, r, m_L. The index s_L denotes the termination index of the ordinary Lanczos method. By Lemma 2.4 the identity

$$s_L = \max\{k \in \{0, 1, \ldots, b\} \,|\, \nu(k) = k\}$$

holds.
The sequence $(\nu(k))_{0\leq k\leq b}$ denotes the degree indices of $(\mathcal{K}(i, A^T, w_1), \mathcal{K}(i, A, v_1))_{1\leq i\leq N}$. The number r is defined as

$$r := \max\{\nu(k) \,|\, 0 \leq k \leq b\} = \nu(b)$$

and $m_L := \min\{\dim \mathcal{K}(N, A, v_1), \dim \mathcal{K}(N, A^T, w_1)\}$.
As soon as one of the weak inequalities

$$s_L \leq r \leq m_L$$

becomes strict a serious breakdown of the Lanczos method occurs. A classical biorthogonal basis of $((\mathcal{K}(i, A^T, w_1), \mathcal{K}(i, A, v_1)))_{1\leq i\leq l}$ exists for $l \leq s_L$, but not for $l > s_L$ (see Lemma 2.4). However, generalized biorthogonal bases exist up to the index $l = r$.

$s_L < r$: *(Curable breakdown)* The breakdown of the ordinary Lanczos method can be cured by using the look-ahead Lanczos recursions. A block of higher dimension is constructed.

$r < m_L$: *(Incurable breakdown)* The look-ahead Lanczos process terminates after step r. By Lemma 2.3 an incurable breakdown occurs if and only if there is no projection Π with

$$\operatorname{Im}\Pi = \mathcal{K}(N, A, v_1) \qquad \text{and} \qquad \operatorname{Ker}\Pi = \mathcal{K}(N, A^T, w_1)^{\perp}.$$

The recursion formulas in Theorem 3.3 are identical with the most general case of Gutknecht's [33] recursion relations for formal orthogonal polynomials. However, Gutknecht, [33], Freund, Gutknecht and Nachtigal [20], and Gragg and Lindquist [31] favour the special case where the inner vectors $((v_i, w_i))_{\nu(k)+2\leq i\leq \nu(k+1)}$ are constructed in a symmetric fashion,

$$f_{i,j}^{(k)} = g_{i,j}^{(k)} \quad \text{for all} \quad k, i, j\,, \tag{3.13}$$

or with symmetric three-term recursions only:

$$g_{i,j}^{(k)} = f_{i,j}^{(k)} \quad \text{for all} \quad k,i,j \quad \text{and} \quad g_{i,j}^{(k)} = 0 = f_{i,j}^{(k)} \quad \text{for} \quad i-j > 2\,. \tag{3.14}$$

For the computation of the regular vectors $w_{\nu(k)+1}$ and $v_{\nu(k)+1}$ the matrix D_k has to be inverted. By Theorem 3.3 and Lemma 3.2 the best structure for D_k would be antidiagonal

$$D_k = \begin{bmatrix} & & & * \\ & 0 & \cdot & \\ & & \cdot & \\ & \cdot & 0 & \\ * & & & \end{bmatrix}.$$

Gutknecht showed in [33] how to construct inner vectors which fulfil the symmetry condition (3.13) and form antidiagonal blocks. The computational work for the construction of this basis is larger than the work for other generalized biorthogonal bases. Therefore we do not require symmetry (3.13). Our goal is to find a generalized biorthogonal basis which can be constructed with minimal computational work and minimal storage requirements. We will see that a special generalized biorthogonal basis with antidiagonal blocks (minimal Lanczos vectors) fulfils these minimality conditions and that this minimum can only be reached giving up condition (3.13).

3.2.3 Generalized biorthogonal bases with antidiagonal blocks

The next theorem shows how to construct bases such that the diagonal blocks belonging to these bases have antidiagonal structure.

Theorem 3.4 (Blocks with antidiagonal structure) *Let $((\tilde{w}_i, \tilde{v}_i))_{1\le i\le r}$ be a generalized biorthogonal basis of $((\mathcal{K}(i, A^T, w_1), \mathcal{K}(i, A, v_1)))_{1\le i\le r}$. Then there exists (up to scalar factors) exactly one generalized biorthogonal basis $((w_i, v_i))_{1\le i\le r}$ with the properties:*

a) The vectors v_i and $\tilde{v}_i$ are identical for all $1 \le i \le r$.

b) The matrix D_k has antidiagonal structure for all $k \in K$.

Let $g_{i,j}^{(k)}$ be the coefficients of the $\tilde{v}$-vectors (from Theorem 3.3). Then the following recursion generates the generalized biorthogonal basis $((w_i, v_i))_{1\le i\le r}$ with the properties a) and b):

For all $0 \le k \le b-1$, $1 \le i \le dim_{k+1} - 1$

$$w_{\nu(k)+1} := \tilde{w}_{\nu(k)+1}\,, \qquad w_{\nu(k)+i+1} := A^T w_{\nu(k)+i} - \sum_{l=1}^{i} f_{i+1,l}^{(k)} w_{\nu(k)+l}$$

where

$$f_{i+1,1}^{(k)} := \frac{\langle w_{\nu(k)+i}, Av_{\nu(k+1)}\rangle}{\langle w_{\nu(k)+1}, v_{\nu(k+1)}\rangle}, \quad f_{i+1,l}^{(k)} := \frac{\langle w_{\nu(k)+i}, v_{\nu(k+1)-i+1}\rangle}{\langle w_{\nu(k)+l}, v_{\nu(k+1)-l+1}\rangle} g_{dim_{k+1}-l+2,dim_{k+1}-i+1}^{(k)}.$$

Proof. By Lemma 3.2 the matrix $D^{k+1} := (\langle \tilde{w}_{\nu(k)+i}, \tilde{v}_{\nu(k+1)-j+1}\rangle)_{1\le i,j\le dim_{k+1}}$ is invertible and lower triangular. Therefore the ordinary Gram-Schmidt biorthogonalisation method applied to the basis $(\tilde{w}_{\nu(k)+i}, \tilde{v}_{\nu(k+1)-i+1})_{1\le i\le dim_{k+1}}$ $(0 \le k \le b-1)$ generates the uniquely determined (up to scalar factors) generalized biorthogonal basis $((w_i, v_i))_{1\le i\le s}$ that fulfils the conditions a) and b). Let $\beta_i^{(k)}, g_{i,j}^{(k)}$ be the coefficients of the $\tilde{v}$-vectors from Theorem 3.3. Then the following identity holds for $2 \le l \le i \le dim_{k+1}$:

$$\begin{aligned}\langle w_{\nu(k)+i}, Av_{\nu(k+1)-l+1}\rangle &= \frac{\langle w_{\nu(k)+i}, v_{\nu(k+1)-l+2}\rangle}{\beta_{dim_{k+1}-l+2}^{(k)}} + \sum_{j=1}^{dim_{k+1}-l+1} g_{dim_{k+1}-l+2,j}^{(k)} \langle w_{\nu(k)+i}, v_{\nu(k)+j}\rangle \\ &= g_{dim_{k+1}-l+2,dim_{k+1}-i+1}^{(k)} \langle w_{\nu(k)+i}, v_{\nu(k+1)-i+1}\rangle. \qquad (3.15)\end{aligned}$$

Finally, equation (3.15) combined with the ordinary Gram-Schmidt biorthogonalisation method proves the assertion

$$\begin{aligned}\frac{1}{\alpha_{i+1}^{(k)}} w_{\nu(k)+i+1} &= A^T w_{\nu(k)+i} - \sum_{l=1}^{i} \frac{\langle A^T w_{\nu(k)+i}, v_{\nu(k+1)-l+1}\rangle}{\langle w_{\nu(k)+l}, v_{\nu(k+1)-l+1}\rangle} w_{\nu(k)+l} \\ &= A^T w_{\nu(k)+i} \\ &\quad - \sum_{l=2}^{i} \frac{\langle w_{\nu(k)+i}, v_{\nu(k+1)-i+1}\rangle}{\langle w_{\nu(k)+l}, v_{\nu(k+1)-l+1}\rangle} g_{dim_{k+1}-l+2)(dim_{k+1}-i+1)}^{(k)} w_{\nu(k)+l} \\ &\quad - \frac{\langle w_{\nu(k)+i}, Av_{\nu(k+1)}\rangle}{\langle w_{\nu(k)+1}, v_{\nu(k+1)}\rangle} w_{\nu(k)+1}.\end{aligned}$$

■

Let $w_{\nu(k)+1}$ and $v_{\nu(k)+1}$ be regular vectors. By Theorem 3.4 the recursion

$$v_{\nu(k)+i+1} = Av_{\nu(k)+i} \qquad (3.16)$$

for all $0 \le k \le b-1$, $1 \le i \le dim_{k+1} - 1$ determines uniquely (up to scalar factors) a generalized biorthogonal basis with antidiagonal blocks. We call this basis minimal Lanczos vectors. Theorem 3.4 shows that the computation of the minimal Lanczos basis requires minimal computational work (as compared to any other basis with diagonal blocks) for the construction of the inner vectors. A further substantial advantage of the minimal Lanczos vectors is that the block three-term recursions for the construction of the regular vectors $w_{\nu(k)+1}$ can be reduced to ordinary three-term recursions (see Theorem 3.5). The minimal Lanczos vectors are the uniquely determined generalized biorthogonal basis with this property.

Theorem 3.5 (Minimal look-ahead Lanczos method) *Let $\alpha_i^{(k)}, \beta_i^{(k)} \in \mathbb{C} \setminus \{0\}$. The following recursions generate the minimal Lanczos vectors:*

For all $0 \leq k \leq b-1$

$$\begin{aligned}
\frac{1}{\beta_1^{(k)}} v_{\nu(k)+1} &= A v_{\nu(k)} - \sum_{l=0}^{dim_k - 1} \frac{\langle w_{\nu(k)-l}, A v_{\nu(k)} \rangle}{\langle w_{\nu(k)-l}, v_{\nu(k-1)+l+1} \rangle} v_{\nu(k-1)+l+1} \\
&\quad - \frac{1}{\alpha_1^{(k-1)}} \cdot \frac{\langle w_{\nu(k-1)+1}, v_{\nu(k)} \rangle}{\langle w_{\nu(k-1)}, v_{\nu(k-2)+1} \rangle} v_{\nu(k-2)+1} \\
\frac{1}{\alpha_1^{(k)}} w_{\nu(k)+1} &= A^T w_{\nu(k)} - \frac{\langle w_{\nu(k)}, A v_{\nu(k)} \rangle}{\langle w_{\nu(k-1)+1}, v_{\nu(k)} \rangle} w_{\nu(k-1)+1} \\
&\quad - \frac{1}{\beta_1^{(k-1)}} \cdot \frac{\langle w_{\nu(k)}, v_{\nu(k-1)+1} \rangle}{\langle w_{\nu(k-2)+1}, v_{\nu(k-1)} \rangle} w_{\nu(k-2)+1}
\end{aligned}$$

for all $1 \leq i \leq dim_{k+1} - 1$

$$v_{\nu(k)+i+1} = \beta_{i+1}^{(k)} A v_{\nu(k)+i}$$

for all $1 \leq i \leq dim_{k+1} - 1$

$$w_{\nu(k)+i+1} = \alpha_{i+1}^{(k)} \left(A^T w_{\nu(k)+i} - \frac{\langle p_{\nu(k)+i}, A q_{\nu(k+1)} \rangle}{\langle p_{\nu(k)+1}, q_{\nu(k+1)} \rangle} w_{\nu(k)+1} \right).$$

Proof. The minimal Lanczos vectors satisfy, for $1 \leq l \leq dim_{k+1} - 1$, the equations

$$\beta_{dim_k - l + 1}^{(k)} \langle A^T w_{\nu(k)}, v_{\nu(k)-l} \rangle = \langle w_{\nu(k)}, v_{\nu(k)-l+1} \rangle = 0 \tag{3.17}$$

because of the antidiagonal structure of D_{k+1}. Identity (3.17), Theorem 3.4 and Theorem 3.3 yields

$$\begin{aligned}
\frac{1}{\alpha_{i+1}^{(k)}} w_{\nu(k)+i+1} &= A^T w_{\nu(k)+i} \\
&\quad - \sum_{l=1}^{i} \frac{\langle A^T p_{\nu(k)+i}, q_{\nu(k+1)-l+1} \rangle}{\langle p_{\nu(k)+l}, q_{\nu(k+1)-l+1} \rangle} w_{\nu(k)+l} \\
&= A^T w_{\nu(k)+i} - \frac{\langle p_{\nu(k)+i}, A q_{\nu(k+1)} \rangle}{\langle p_{\nu(k)+1}, q_{\nu(k+1)} \rangle} w_{\nu(k)+1}
\end{aligned}$$

the desired representation of the w vectors. ∎

Estimate of the cost of the minimal look-ahead Lanczos method: Independently of the block size two matrix-vector multiplications and two inner products are required per step. All other inner products can be computed recursively

$$\beta^{(k-1)}_{dim_k-i+1}\langle w_{\nu(k-1)+i+1}, v_{\nu(k)-i}\rangle = \alpha^{(k-1)}_{i+1}\langle w_{\nu(k-1)+i}, v_{\nu(k)-i+1}\rangle$$

for all $1 \leq i \leq dim_k - 1$. To compute the regular vectors $w_{\nu(k)+1}$ and $v_{\nu(k)+1}$ three scalar-vector multiplications and two vector additions are required. The construction of the inner vectors $w_{\nu(k)+i}$ and $v_{\nu(k)+i}$ requires two more scalar-vector multiplications and one vector addition.
Therefore the computational work for generating the minimal Lanczos vectors increases as block size decreases and reaches its maximum (= computational work of the ordinary Lanczos method) if and only if there exists an ordinary biorthogonal basis of the Krylov chain $((\mathcal{K}(i, A^T, w_1), \mathcal{K}(i, A, v_1)))_{1\leq i\leq r}$.

3.2.4 Structure of the generated matrices

Every look-ahead Lanczos method generates a generalized biorthogonal basis of $((\mathcal{K}(i, A^T, w_1), \mathcal{K}(i, A, v_1)))_{1\leq i\leq r}$ where $r := \nu(b) = \max\{\nu(k)|k \in K\}$ (see Theorem 3.3). It further generates implicitly for every $j \in \{\nu(k)|k \in K\}$ upper Hessenberg matrices $H^{(j)}$ and $\hat{H}^{(j)}$ satisfying the equations

$$V^{(j)}H^{(j)} = \Pi_j A V^{(j)} \tag{3.18}$$
$$W^{(j)}\hat{H}^{(j)} = \Pi_j^T A^T W^{(j)} \tag{3.19}$$

where Π_j is the Gram-Schmidt projection with

$$\operatorname{Im}\Pi_j = \mathcal{K}(j, A, v_1) \qquad \text{and} \qquad \operatorname{Ker}\Pi_j = \mathcal{K}(j, A^T, w_1)^{\perp}.$$

Before we discuss how to construct the matrices $H^{(j)}$, from data computed by the look-ahead Lanczos process, we show that $H^{(j)}$ and $\hat{H}^{(j)}$ are similar matrices.
Using the identity

$$\Pi_j = V^{(j)}(D^{(j)})^{-1}(W^{(j)})^T$$

where $D^{(j)} := (W^{(j)})^T V^{(j)}$ the equations (3.18) and (3.19) become

$$H^{(j)} = (D^{(j)})^{-1}(W^{(j)})^T A V^{(j)} \tag{3.20}$$
$$\hat{H}^{(j)} = ((D^{(j)})^{-1})^T (V^{(j)})^T A^T W^{(j)}. \tag{3.21}$$

Transposing the equation (3.21) and multiplying the new equation with $(D^{(j)})^{-1}$ from the left and $D^{(j)}$ from the right, we obtain the identity

$$(D^{(j)})^{-1}\hat{H}^{(j)}D^{(j)} = (D^{(j)})^{-1}(W^{(j)})^T A V^{(j)} = H^{(j)}. \tag{3.22}$$

The following Theorem 3.6 provides information about the structure of the matrix $H^{(j)}$.

Theorem 3.6 (Block tridiagonal matrices) *The matrix $H^{(r)}$, generated by a look-ahead Lanczos method, is block tridiagonal and unreduced upper Hessenberg*

$$H^{(r)} = \begin{bmatrix} E_1 & C_1 & & 0 \\ B_1 & E_2 & \ddots & \\ & \ddots & \ddots & C_{b-1} \\ 0 & & B_{b-1} & E_b \end{bmatrix}.$$

Each of the non-diagonal blocks $B_k \in \mathrm{Mat}_{dim_{k+1} \times dim_k}$ and $C_k \in \mathrm{Mat}_{dim_k \times dim_{k+1}}$ has exactly one non-zero element in the upper right corner. Let $g_{i,j}^{(k)}$, $\beta_i^{(k)}$ be the coefficients from Theorem 3.3. Then the entries of the i-th column of E_{k+1} are the coefficients of the $(i+1)$-st inner vector from the k-th block:

For all $k+1 \in K$

for all $i \in \{1, 2, \ldots, dim_{k+1} - 1\}$

$$E_{k+1} e_i = (g_{i+1,1}^{(k)},\ g_{i+1,2}^{(k)},\ \ldots,\ g_{i+1,i}^{(k)},\ (\beta_{i+1}^{(k)})^{-1},\ 0,\ \ldots,\ 0)^T.$$

The entries of the $\nu(k)$-th column of $H^{(r)}$ are the projection coefficients of $Av_{\nu(k)}$, in other words: Let $(a_j^{(k)})_{\nu(k-1)+1 \le j \le \nu(k)}$ be the last column of E_k, c_{k-1} the non-zero element in C_{k-1} and b_k $(= (\beta_1^{(k)})^{-1})$ the non-zero element in E_k. Then

$$b_k v_{\nu(k)+1} = (I - \Pi_{\nu(k)}) A v_{\nu(k)} = A v_{\nu(k)} - \sum_{j=\nu(k-1)+1}^{\nu(k)} a_j^{(k)} v_j - c_{k-1} v_{\nu(k-2)+1}\,.$$

Theorem 3.6 follows directly from the equation

$$V^{(j)} H^{(j)} = A V^{(j)} - \{(I - \Pi_j) A v_j\} e_j^T$$

which is evidently equivalent to the equation (3.18).

Remark 3.3 Let $((w_i, v_i))_{1 \le i \le r}$ and $((\tilde{w}_i, \tilde{v}_i))_{1 \le i \le r}$ be bases with identical v vectors. Then the generated matrices $H^{(r)}$ and $\tilde{H}^{(r)}$ are identical, too.

In the special case of inner three-term recursions ($f_{i,j}^{(k)} = 0 = g_{i,j}^{(k)}$ for $i - j > 2$) the diagonal blocks have the structure

$$E_k = \begin{bmatrix} * & * & & 0 & * \\ * & \ddots & \ddots & & \vdots \\ & \ddots & \ddots & \ddots & \vdots \\ & & \ddots & \ddots & * \\ & 0 & & * & * \end{bmatrix}.$$

The minimal look-ahead Lanczos method generates diagonal blocks of the form

$$E_k = \begin{bmatrix} & & & * \\ * & & 0 & \vdots \\ & \ddots & & \vdots \\ 0 & & * & * \end{bmatrix}.$$

The structures of the various matrices $H^{(r)}$, $D^{(r)}$ and $W^{(r)}H^{(r)}AV^{(r)}$ generated by different look-ahead Lanczos methods are illustrated in the following example.

Example 3.2

$$A = \begin{bmatrix} 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3/2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 5/4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 6/5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 7/6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 8/7 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 9/8 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad w_1 = \frac{1}{8}\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{bmatrix}, \quad v_1 = \begin{bmatrix} 1 \\ 1/2 \\ 1/3 \\ 1/4 \\ 1/5 \\ 1/6 \\ 0 \\ 1/8 \end{bmatrix}.$$

In this example the vectors w_i and v_i are normalised using $\alpha_i^{(k)}$ and $\beta_i^{(k)}$ and the maximum norm. The degree indices of $((\mathcal{K}(i, A^T, w_1), \mathcal{K}(i, A, v_1)))_{1 \le i \le 8}$ are

$$\nu(0) = 0,\ \nu(1) = 1,\ \nu(2) = 2,\ \nu(3) = 6,\ \nu(4) = 7 \quad \text{and} \quad \nu(5) = 8.$$

The diagonal blocks associated with every generalized biorthogonal basis are

$$D_1 = [7/8], \quad D_2 = [-7/8], \quad D_4 = [3/4], \quad D_5 = [-1].$$

a) *Ordinary Lanczos method:* The ordinary Lanczos method breaks down curably after step 2 and generates a tridiagonal matrix $H^{(2)}$ with eigenvalues that are not eigenvalues of A:

$$H^{(2)} = \begin{bmatrix} 6/7 & -1/7 \\ 1/7 & 8/7 \end{bmatrix}, \qquad D^{(2)} = \begin{bmatrix} 7/8 & 0 \\ 0 & -7/8 \end{bmatrix}, \qquad (W^{(2)})^T AV^{(2)} = \begin{bmatrix} 3/4 & -1/8 \\ -1/8 & -1 \end{bmatrix}.$$

b) *Minimal look-ahead Lanczos method:* The minimal look-ahead Lanczos method generates the following matrices:

$$D^{(3)} = \begin{bmatrix} 0 & 0 & 0 & 3/4 \\ 0 & 0 & 1/2 & 0 \\ 0 & 5/12 & 0 & 0 \\ 3/8 & 0 & 0 & 0 \end{bmatrix}$$

$$H^{(8)} = \left[\begin{array}{cc|cccc|cc} 6/7 & -1/7 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 1/7 & 8/7 & 0 & 0 & 0 & -3/2 & 0 & 0 \\ \hline 0 & 7/3 & 0 & 0 & 0 & -3 & 2 & 0 \\ 0 & 0 & 6/5 & 0 & 0 & -15/4 & 0 & 0 \\ 0 & 0 & 0 & 5/4 & 0 & -18/5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4/3 & -3 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 1/4 & 1 & -7/6 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 6/7 & 0 \end{array}\right]$$

$$(W^{(8)})^T A V^{(8)} = \left[\begin{array}{cc|ccccc|c} 3/4 & -1/8 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline -1/8 & -1 & 0 & 0 & 0 & 21/16 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & -9/4 & 0 & 0 \\ 0 & 0 & 0 & 5/8 & 0 & -15/8 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 0 & -3/2 & 0 & 0 \\ 0 & 7/8 & 0 & 0 & 0 & -9/8 & 3/4 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 3/16 & 3/4 & -7/8 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & -6/7 & 0 \end{array}\right].$$

c) *Inner three-term recursion:* Let the coefficients $f_{ij}^{(k)} = g_{ij}^{(k)}$ in Theorem 3.3 be chosen as follows:

$$\begin{aligned} f_{2,1}^{(k)} &= g_{2,1}^{(k)} = 0.3, & f_{3,1}^{(k)} &= g_{3,1}^{(k)} = 0.2, & f_{3,2}^{(k)} &= g_{3,2}^{(k)} = 0.6 \\ f_{4,1}^{(k)} &= g_{4,1}^{(k)} = 0.0, & f_{4,2}^{(k)} &= g_{4,2}^{(k)} = 0.4, & f_{4,3}^{(k)} &= g_{4,3}^{(k)} = 0.9. \end{aligned}$$

The look-ahead Lanczos method with this symmetric inner three-term recursion generates the following matrices:

$$D_3 = \begin{bmatrix} 0 & 0 & 0 & 0.35 \\ 0 & 0 & 0.45 & -1.20 \\ 0 & 0.38 & -1.14 & 1.69 \\ 0.27 & -0.99 & 1.63 & -1.67 \end{bmatrix}$$

$$H^{(8)} = \left[\begin{array}{cc|cccc|cc} 6/7 & -1/7 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 1/7 & 8/7 & 0 & 0 & 0 & -0.71 & 0 & 0 \\ \hline 0 & 7/3 & 0.30 & 0.20 & 0 & -2.64 & 2 & 0 \\ 0 & 0 & 1.38 & 0.60 & 0.40 & -5.71 & 0 & 0 \\ 0 & 0 & 0 & 1.61 & 0.90 & -7.09 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1.91 & -4.80 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0.12 & 1 & -7/6 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 6/7 & 0 \end{array}\right]$$

$$(W^{(8)})^T A V^{(8)} = \left[\begin{array}{cc|cccc|cc} 3/4 & -1/8 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline -1/8 & -1 & 0 & 0 & 0 & 0.62 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0.68 & -1.69 & 0 & 0 \\ 0 & 0 & 0 & 0.73 & -1.89 & 2.57 & 0 & 0 \\ 0 & 0 & 0.53 & -1.60 & 2.37 & -2.24 & 0 & 0 \\ 0 & 0.63 & -1.29 & 2.09 & -2.12 & 1.37 & 0.54 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0.09 & 3/4 & -7/8 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & -6/7 & 0 \end{array}\right].$$

The following example shows the decisive influence of the starting vectors v_1 and w_1 on the degree indices.

Example 3.3 Let A be defined as in Example 3.2. In the following we vary the starting vectors to illustrate their influence on the degree indices.

a) Let $i \in \{2, 3, 4, 5, 6, 7\}$. For every choice

$$v_1 \in \operatorname{span}\{e_1, e_2, \ldots, e_i\}, \qquad w_1 \in \operatorname{span}\{e_{i+1}, e_{i+2}, \ldots, e_8\}$$

there exists no generalized biorthogonal basis because $r = 0$.

b) *Early termination:* The starting vectors $w_1 = e_1$, $v_1 = e_1$ yield the indices $s_L = r = m_L = 1$. Hence early termination after step 1.

c) *Curable breakdown:* If one chooses $w_1 = e_1$ and $v_1 = e_8$ then the indices are $r = m_L = \nu(1) = 8$, $s_L = 0$. Therefore no classical but a general biorthogonal basis exists.

d) *No breakdown:* In contrast to c) the ordinary Lanczos method applied to $w_1 = v_1 = \sum_{j=1}^{N} e_j$ generates a tridiagonal matrix ($s_L = r = m_L = 8$).

e) *Incurable breakdown:* The Lanczos method applied to the starting vectors $w_1 = v_1 = e_j$ ($j \in \{2, 3, 4, 5, 6, 7\}$) breaks down incurably after step 1 ($s_L = r = 1 < m$).

4 Practical look-ahead Lanczos methods

The construction of any generalized biorthogonal basis of a Krylov chain requires exact knowledge of the degree indices $(\nu(k))_{k\in K}$. To get these indices one has to decide, after every step, whether D_k is invertible. Hence in finite precision arithmetic only a subsequence $(\hat{\nu}(k))_{k\in\hat{K}}$ of the degree indices $(\nu(k))_{k\in K}$ can be determined. This leads to the following definition of a basis to a subsequence $(\hat{\nu}(k))_{k\in\hat{K}}$. In this section let $\hat{b} := \max\{k \mid k \in \hat{K}\}$ and $dim_k := \hat{\nu}(k) - \hat{\nu}(k-1)$.

Definition 4.1 (Basis to $(\hat{\nu}(k))_{k\in\hat{K}}$) *Let $((\mathcal{W}_i, \mathcal{V}_i))_{1\le i\le s}$ be a chain of subspaces with degree indices $(\nu(k))_{k\in K}$ and let $(\hat{\nu}(k))_{k\in\hat{K}} \subseteq (\nu(k))_{k\in K}$. A basis $((w_i, v_i))_{1\le i\le s}$ of the chain $((\mathcal{W}_i, \mathcal{V}_i))_{1\le i\le s}$ is called a basis to $(\hat{\nu}(k))_{k\in\hat{K}}$ if the matrix $D^{(s)} := (\langle w_i, v_j\rangle)_{1\le i,j\le s}$ is block diagonal*

$$D^{(s)} = diag\,(D_1, D_2, \ldots, D_{\hat{b}})$$

with invertible diagonal blocks D_k of order dim_k (for $k = 1, 2, \ldots, \hat{b}$).

Now we turn to the construction of a basis to $(\hat{\nu}(k))_{k\in\hat{K}}$.

Suppose there is a basis to $(\hat{\nu}(k))_{k\in\hat{K}}$. A slight modification of the extended Gram-Schmidt biorthogonalisation method (see $MEGS$ below) applied to an arbitrary basis $((\tilde{w}_i, \tilde{v}_i))_{1\le i\le s}$ of $((\mathcal{W}_i, \mathcal{V}_i))_{1\le i\le s}$ yields a basis to $(\hat{\nu}(k))_{k\in\hat{K}}$.

Modified extended Gram-Schmidt look-ahead Lanczos method ($MEGS$)

For $k = 0, 1, 2 \ldots, s-1$
  for $i = 1, 2, \ldots, dim_k$

$$\begin{aligned} v_{\hat{\nu}(k)+i} &:= (I - \Pi_{\hat{\nu}(k)})\tilde{v}_{\hat{\nu}(k)+i} \\ w_{\hat{\nu}(k)+i} &:= (I - \Pi^T_{\hat{\nu}(k)})\tilde{w}_{\hat{\nu}(k)+i}. \end{aligned}$$

Here $\Pi_{\hat{\nu}(k)}$ is the projection onto $\mathcal{V}_{\hat{\nu}(k)}$ with kernel $\mathcal{W}^{\perp}_{\hat{\nu}(k)}$.

Using the $MEGS$ algorithm the results about generalized biorthogonal bases can be easily transferred to bases to $(\hat{\nu}(k))_{k\in\hat{K}}$.

Theorem 4.1 (Existence) *The following statements are equivalent:*

a) There is a basis to $(\hat{\nu}(k))_{k\in\hat{K}}$ of $((\mathcal{W}_i, \mathcal{V}_i))_{1\le i\le s}$.

b) There is a generalized biorthogonal basis of $((\mathcal{W}_i, \mathcal{V}_i))_{1\le i\le s}$.

c) The $MEGS$ method applied to an arbitrary basis of the chain $((\mathcal{W}_i, \mathcal{V}_i))_{1\le i\le s}$ generates a basis to $(\hat{\nu}(k))_{k\in\hat{K}}$.

d) The bilinear form $\langle \cdot, \cdot \rangle \mid_{\mathcal{W}_s \times \mathcal{V}_s}$ *is non-degenerate.*

e) $s = \max\{\nu(k) \mid k \in K\}$.

f) $s = \max\{\hat{\nu}(k) \mid k \in \hat{K}\}$.

g) There is a projection Π *with*

$$\operatorname{Im}\Pi = \mathcal{V}_s \qquad \textit{and} \qquad \operatorname{Ker}\Pi = \mathcal{W}_s^{\perp}.$$

Theorem 4.2 (Degrees of freedom) *Let* $((w_i, v_i))_{1 \le i \le s}$ *be a basis to* $(\hat{\nu}(k))_{k \in \hat{K}}$ *of the chain* $((\mathcal{W}_i, \mathcal{V}_i))_{1 \le i \le s}$. *Let* $((\tilde{w}_i, \tilde{v}_i))_{1 \le i \le s}$ *be a basis of* $((\mathcal{W}_i, \mathcal{V}_i))_{1 \le i \le s}$. *Then the sequence* $((\tilde{w}_i, \tilde{v}_i))_{1 \le i \le s}$ *is a basis to* $(\hat{\nu}(k))_{k \in \hat{K}}$ *if and only if*

$$\begin{aligned} \tilde{v}_{\hat{\nu}(k)+i} &\in \operatorname{span}\{v_{\hat{\nu}(k)+1}, v_{\hat{\nu}(k)+2}, \ldots, v_{\hat{\nu}(k)+i}\} \\ \tilde{w}_{\hat{\nu}(k)+i} &\in \operatorname{span}\{w_{\hat{\nu}(k)+1}, w_{\hat{\nu}(k)+2}, \ldots, w_{\hat{\nu}(k)+i}\} \end{aligned}$$

for all $0 \le k \le \hat{b} - 1$, $1 \le i \le dim_{k+1}$.

Notice that, by Theorem 4.2 and Theorem 3.2, the regular vectors $w_{\hat{\nu}(k)+1}, v_{\hat{\nu}(k)+1}$ of a basis to $(\hat{\nu}(k))_{k \in \hat{K}}$ are regular vectors of a generalized biorthogonal basis.

In the following we specialize to Krylov chains $((\mathcal{K}(i, A^T, w_1), \mathcal{K}(i, A, v_1)))_{1 \le i \le s}$.
The relation

$$Av_{\hat{\nu}(k)+i} \in A\operatorname{Ker}\Pi_{\hat{\nu}(k-1)} \subseteq \operatorname{Ker}\Pi_{\hat{\nu}(k-2)} \tag{4.1}$$

yields the following recursion for the regular vector $v_{\hat{\nu}(k)+1}$:

$$\begin{aligned} v_{\hat{\nu}(k)+1} &= \beta_1^{(k)}(I - \Pi_{\hat{\nu}(k)})Av_{\hat{\nu}(k)} \\ &= \beta_1^{(k)}(Av_{\hat{\nu}(k)} - V_k D_k^{-1} W_k^T Av_{\hat{\nu}(k)} - V_{k-1} D_{k-1}^{-1} (W_{k-1})^T Av_{\hat{\nu}(k)}) \end{aligned}$$

where $\beta_1^{(k)} \in \mathbb{C} \setminus \{0\}$ and the projection $\Pi_{\hat{\nu}(k)}$ is defined by

$$\operatorname{Im}\Pi_{\hat{\nu}(k)} = \mathcal{K}(\hat{\nu}(k), A, v_1) \quad \text{and} \quad \operatorname{Ker}\Pi_{\hat{\nu}(k)} = \mathcal{K}(\hat{\nu}(k), A^T, w_1)^{\perp}.$$

In the generalized biorthogonal case the term

$$V_{k-1} D_{k-1}^{-1} (W_{k-1})^T Av_{\hat{\nu}(k)}$$

is equal to the vector $v_{\hat{\nu}(k-1)+1}$ up to a scalar factor. This is not true in the case of an arbitrary basis to $(\hat{\nu}(k))_{k \in \hat{K}}$ because in general the space $\mathcal{K}(dim_{k-1}, A, v_{\hat{\nu}(k-2)+1})$ is not invariant under $(I - \Pi_{\hat{\nu}(k-2)})$ and the space $\mathcal{K}(dim_{k-1}, A^T, w_{\hat{\nu}(k-2)+1})$ is not invariant under $(I - \Pi_{\hat{\nu}(k-2)}^T)$. As a consequence for the chain $((\mathcal{K}(i, A^T, w_{\hat{\nu}(k)+1}), \mathcal{K}(i, A, v_{\hat{\nu}(k)+1})))_{1 \le i \le dim_{k+1}}$ a basis of inner vectors $((w_i, v_i))_{\hat{\nu}(k)+2 \le i \le \hat{\nu}(k+1)}$ cannot be chosen in an arbitrary manner.

To form inner vectors one has to choose any basis of this Krylov chain and then project its elements by $((I - \Pi^T_{\hat{\nu}(k)}) \ ,(I - \Pi_{\hat{\nu}(k)}))$. Hence, we construct the inner vectors by the recursion

$$v_{\hat{\nu}(k)+i+1} := \beta^{(k)}_{i+1}(I - \Pi_{\hat{\nu}(k)}) \left(A v_{\hat{\nu}(k)+i} - \sum_{j=1}^{i} g^{(k)}_{i+1,j} \, v_{\hat{\nu}(k)+j} \right).$$

Using the relation (4.1) we obtain the representation (4.2) below which are the **practical look-ahead Lanczos recursions** introduced by Freund, Gutknecht and Nachtigal [20] and Gutknecht [34].

Theorem 4.3 (Construction of bases to $(\hat{\nu}(k))_{k\in\hat{K}}$)
Let $((w_i, v_i))_{1\le i\le s}$ be a basis to $(\hat{\nu}(k))_{k\in\hat{K}}$. Then there are coefficients $g^{(k)}_{i,j}, f^{(k)}_{i,j} \in \mathbb{C}$ and $\beta^{(k)}_i, \alpha^{(k)}_i \in \mathbb{C} \setminus \{0\}$ such that the following identities hold:

For all $0 \le k \le \hat{b} - 1$

$$\begin{aligned} v_{\hat{\nu}(k)+1} &= \beta^{(k)}_1 (A v_{\hat{\nu}(k)} - V_k D_k^{-1} W_k^T A v_{\hat{\nu}(k)} - V_{k-1} D_{k-1}^{-1} (W_{k-1})^T A v_{\hat{\nu}(k)} \\ w_{\hat{\nu}(k)+1} &= \alpha^{(k)}_1 (A^T w_{\hat{\nu}(k)} - W_k (D_k^T)^{-1} V_k^T A^T w_{\hat{\nu}(k)} - W_{k-1} (D_{k-1}^T)^{-1} V_{k-1}^T A^T w_{\hat{\nu}(k)} \end{aligned}$$

for all $1 \le i \le dim_{k+1} - 1$

$$\begin{aligned} v_{\hat{\nu}(k)+i+1} &= \beta^{(k)}_{i+1} \left(A v_{\hat{\nu}(k)+i} - \sum_{j=1}^{i} g^{(k)}_{i+1,j} \, v_{\hat{\nu}(k)+j} - V_k D_k^{-1} W_k^T A v_{\hat{\nu}(k)+i} \right) \qquad (4.2) \\ w_{\hat{\nu}(k)+i+1} &= \alpha^{(k)}_{i+1} \left(A^T w_{\hat{\nu}(k)+i} - \sum_{j=1}^{i} f^{(k)}_{i+1,j} \, w_{\hat{\nu}(k)+j} - W_k (D_k^T)^{-1} V_k^T A^T w_{\hat{\nu}(k)+i} \right). \end{aligned}$$

Conversely these recursions generate a basis $((w_i, v_i))_{1\le i\le s}$ to $(\hat{\nu}(k))_{k\in\hat{K}}$ for any choice of the coefficients $g^{(k)}_{i,j}, f^{(k)}_{i,j} \in \mathbb{C}$, $\beta^{(k)}_i, \alpha^{(k)}_i \in \mathbb{C} \setminus \{0\}$.

Every practical look-ahead Lanczos method generates implicitly, for every number $j \in \{\hat{\nu}(k) | k \in \hat{K}\}$, upper Hessenberg matrices $H^{(j)}$ and $\hat{H}^{(j)}$ satisfying the equations

$$V^{(j)} H^{(j)} = \Pi_j A V^{(j)} \qquad (4.3)$$

$$W^{(j)} \hat{H}^{(j)} = \Pi_j^T A^T V^{(j)}. \qquad (4.4)$$

In the following theorem we describe how the matrices $H^{(j)}$ can be constructed from data computed in the practical look-ahead Lanczos process.

Theorem 4.4 (Block tridiagonal matrices) *The matrix $H^{(s)}$, generated by a practical look-ahead Lanczos method, is block tridiagonal and unreduced upper Hessenberg*

$$H^{(s)} = \begin{bmatrix} E_1 & C_1 & & 0 \\ B_1 & E_2 & \ddots & \\ & \ddots & \ddots & C_{\hat{b}-1} \\ 0 & & B_{\hat{b}-1} & E_{\hat{b}} \end{bmatrix}.$$

Every subdiagonal block $B_k \in \mathrm{Mat}_{dim_{k+1} \times dim_k}$ has exactly one non-zero element in the upper right corner. In general $C_k = (c_{ij})$ is a full matrix. Let $g_{i,j}^{(k)}$, $f_{i,j}^{(k)}$ be the coefficients used in Theorem 4.3. Then the following relations hold:

a) The entries of the $\hat{\nu}(k)$-th column of $H^{(s)}$ are the projection coefficients of $Av_{\hat{\nu}(k)}$. In other words: Let $(a_j^{(k)})_{\hat{\nu}(k-1)+1 \le j \le \hat{\nu}(k)}$ be the last column of E_k and b_k the non-zero element in B_k. Then

$$b_k v_{\hat{\nu}(k)+1} = (I - \Pi_{\hat{\nu}(k)}) A v_{\hat{\nu}(k)} = A v_{\hat{\nu}(k)} - \sum_{j=\hat{\nu}(k-1)+1}^{\hat{\nu}(k)} a_j^{(k)} v_j - \sum_{i=\hat{\nu}(k-2)+1}^{\hat{\nu}(k-1)} c_{i\hat{\nu}(k)} v_i \,.$$

b) The projection coefficients of

$$v := A v_{\hat{\nu}(k)+i} - \sum_{j=1}^{i} g_{i+1,j}^{(k)} \, v_{\hat{\nu}(k)+j}$$

are the entries of the $(\hat{\nu}(k)+i)$-st column of $H^{(s)}$:

$$\frac{1}{\beta_{i+1}^{(k)}} v_{\hat{\nu}(k)+i+1} = (I - \Pi_{\hat{\nu}(k)}) v = A v_{\hat{\nu}(k)+i} - \sum_{j=1}^{i} g_{i+1,j}^{(k)} \, v_{\hat{\nu}(k)+j} - \sum_{j=\hat{\nu}(k-2)+1}^{\hat{\nu}(k-1)} c_{j(\hat{\nu}(k)+i)} v_j .$$

Remark 4.1 In a similar manner as described in Section 3.2.4 we obtain the identity

$$H^{(s)} = (D^{(s)})^{-1} \hat{H}^{(s)} D^{(s)}. \tag{4.5}$$

Here $D^{(s)} = (W^{(s)})^T V^{(s)}$.

In the following we denote the projection coefficients of $A^T w_{\hat{\nu}(k)}$ by $\check{a}_j^{(k)}, \check{c}_{i\hat{\nu}(k)}$.

The practical monomial look-ahead Lanczos method uses the recursions from Theorem 4.3 with the coefficients $g_{i,j}^{(k)} = 0$ and $f_{i,j}^{(k)} = 0$ for all k, i, j. Throughout the rest of the paper we assume that all vectors w_i and v_i are normalized in Euclidean norm.

4.1 Improvement of stability for look-ahead Lanczos methods

The block biorthogonality ($(W_i)^T V_j = 0$ for $i \neq j$) is used in Theorem 4.3 and Theorem 4.4. Unfortunately, in finite precision arithmetic this biorthogonality can be lost after few steps.
In the following examples the IEEE-arithmetic 'extended' has been used for the computation of the inner products and the IEEE-arithmetic 'double' for all other computations. The starting vectors v_1 and w_1 are given without normalization.

Example 4.1

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 6 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 8 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 10 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 12 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 14 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 16 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 18 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 20 \end{bmatrix}, \quad w_1 = \begin{bmatrix} 1.00 \\ 0.28 \\ 0.55 \\ 0.64 \\ 0.74 \\ 0.16 \\ 0.14 \\ 0.12 \\ 0.17 \\ 0.45 \end{bmatrix}, \quad v_1 = \begin{bmatrix} 0.06 \\ 0.41 \\ 0.66 \\ 0.32 \\ 0.85 \\ 0.95 \\ 0.32 \\ 0.59 \\ -1.00 \\ 1.00 \end{bmatrix}.$$

The degree indices of (w_1, A, v_1) are: $\nu(k) = k$ for $1 \leq k \leq 10$. The ordinary Lanczos method generates the diagonal blocks

$$\begin{array}{llllllllll} D_1 = & 0.541 & D_2 = & -0.006 & D_3 = & 0.006 & D_4 = & -0.658 & D_5 = & 0.493 \\ D_6 = & -0.479 & D_7 = & 0.653 & D_8 = & -0.328 & D_9 = & -0.483 & D_{10} = & 0.284. \end{array}$$

The first column of $D^{(10)}$ shows that the block biorthogonality has been nearly lost after few steps

$$\begin{aligned} (D^{(10)} e_1)^T &= [5 \cdot 10^{-1}, 2 \cdot 10^{-17}, -7 \cdot 10^{-18}, 4 \cdot 10^{-13}, -3 \cdot 10^{-13}, \\ &\quad 8 \cdot 10^{-10}, 3 \cdot 10^{-8}, 4 \cdot 10^{-7}, 1 \cdot 10^{-5}, 1 \cdot 10^{-4}]. \end{aligned}$$

A slow loss of block biorthogonality cannot be avoided in finite precision arithmetic. However, a quick loss can be avoided in many cases. There are two possible reasons for a quick loss of block biorthogonality:

1) The nearly linear dependence of vectors from different blocks.

2) The nearly linear dependence of the two vectors $Av_{\hat{\nu}(k)}$ and $\Pi_{\hat{\nu}(k)}Av_{\hat{\nu}(k)}$.

The projection coefficients $a_j^{(k)}, c_{i\hat{\nu}(k)}, \check{a}_j^{(k)}, \check{c}_{i\hat{\nu}(k)}$ from Theorems 4.3 and 4.4 provides a simple measure for the norm of the projection $\Pi_{\hat{\nu}(k)}$:

$$\begin{aligned} ||\Pi_{\hat{\nu}(k)}Av_{\hat{\nu}(k)}|| &\leq \underbrace{\sum_{j=\hat{\nu}(k-1)+1}^{\hat{\nu}(k)} |a_j^{(k)}|}_{=:\ \zeta_v^1(k)} + \underbrace{\sum_{i=\hat{\nu}(k-2)+1}^{\hat{\nu}(k-1)} |c_{i\hat{\nu}(k)}|}_{=:\ \zeta_v^2(k)} \\ ||\Pi_{\hat{\nu}(k)}^T A^T w_{\hat{\nu}(k)}|| &\leq \underbrace{\sum_{j=\hat{\nu}(k-1)+1}^{\hat{\nu}(k)} |\check{a}_j^{(k)}|}_{=:\ \zeta_w^1(k)} + \underbrace{\sum_{i=\hat{\nu}(k-2)+1}^{\hat{\nu}(k-1)} |\check{c}_{i\hat{\nu}(k)}|}_{=:\ \zeta_w^2(k)}. \end{aligned}$$

We define

$$\mu_v^1(k) := \frac{\zeta_v^1(k)}{||Av_{\hat{\nu}(k)}||}, \quad \mu_v^2(k) := \frac{\zeta_v^2(k)}{||Av_{\hat{\nu}(k)}||} \quad \mu_w^1(k) := \frac{\zeta_w^1(k)}{||A^T w_{\hat{\nu}(k)}||}, \quad \mu_w^2(k) := \frac{\zeta_w^2(k)}{||A^T w_{\hat{\nu}(k)}||}.$$

In general the linear independence of vectors from different blocks can be improved by the following procedure: A regular vector is constructed only if the condition

$$\xi \geq \mu(k) := \max\{\mu_w^1(k), \mu_w^2(k), \mu_v^1(k), \mu_v^2(k)\} \tag{4.6}$$

(where ξ is a constant) is fulfilled. Instead of condition (4.6) Freund, Gutknecht and Nachtigal [20] use the check

$$C \geq \zeta(k) := \max\{\zeta_w^1(k), \zeta_w^2(k), \zeta_v^1(k), \zeta_v^2(k)\}. \tag{4.7}$$

Here the constant C can be obtained from a previous run, or by setting

$$C := \max\{||Av_1||, ||A^T w_1||\}. \tag{4.8}$$

This avoids the computation of $||Av_{\hat{\nu}(k)}||$ and $||A^T w_{\hat{\nu}(k)}||$. However, $\zeta(k)/C$ is a rather crude measure for the linear independence of vectors from different blocks (see Example 4.2).

Example 4.2 Let $\gamma_2, \gamma_3, \gamma_4 \in \mathbb{C}$.

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & \gamma_2 & 1 & 0 \\ 0 & 1 & \gamma_3 & 1 \\ 0 & 0 & 1 & \gamma_4 \end{bmatrix}, \qquad w_1 = v_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

The Lanczos method applied to (w_1, A, v_1) generates the exact result even in finite precision arithmetic:

$$W^{(4)} = V^{(4)} = D^{(4)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \qquad H^{(4)} = A.$$

a) Let C be defined as in (4.8) and let $|\gamma_2| > 1$. Then check (4.7) fails after the second step.
b) Choose $\xi \geq 1$. Then the condition (4.6) is fulfilled for all $\gamma_2, \gamma_3, \gamma_4 \in \mathbb{C}$:

$$\mu(1) = 0, \quad \mu(4) = \frac{\max\{1, |\gamma_4|\}}{\sqrt{1 + |\gamma_4|^2}}, \quad \mu(k) = \frac{\max\{1, |\gamma_k|\}}{\sqrt{2 + |\gamma_k|^2}}, \quad k = 2, 3.$$

If condition (4.7) cannot be fulfilled by increasing the block size then the obvious thing to do is continuing with that block size which has the minimal $\zeta(k)$-value. Then one can continue using the same constant C as before or, as Freund, Gutknecht and Nachtigal [20] proposed, one can replace C with the minimal $\zeta(k)$-value. Both possibilities have advantages: By updating C it can be prevented in many cases that the condition (4.7) fails in a later step. Hence storage and computational work may be saved. On the other hand, if one does not change C then the block biorthogonality may be improved in a later step. As a compromise we propose the following procedure:
If the minimal value $\zeta(k)$ is below the upper bound $\varsigma \cdot C$ (where $\varsigma > 1$ is a constant) then C is replaced by $\zeta(k)$, otherwise C is chosen as $\varsigma \cdot C$. A similar procedure can be used to update the value ξ. In the rest of this section we restrict our attention to check (4.6). To measure the block biorthogonality of a 'finite precision basis' to $(\hat{\nu}(k))_{k \in \hat{K}}$ (i.e. a sequence generated in finite precision arithmetic by the recursions of Theorem 4.3) we use the measure τ defined as

$$\tau_s := \max\{|\delta_{ij}| : (i, j) \notin \{\nu(k-1)+1, \nu(k-1)+2, \ldots, \nu(k)\}^2 \quad \text{for all} \quad 1 \leq k \leq \hat{b}\}$$

where $D^{(s)} = (\delta_{ij})_{1 \leq i,j \leq s} := (\langle w_i, v_j \rangle)_{1 \leq i,j \leq s}$. In Example 4.1 the block biorthogonality can be improved considerably by using check (4.6).

Example 4.3 (Continuing Example 4.1) We choose $\xi = 5$. Then the monomial look-ahead Lanczos method together with check (4.6) applied to the triple (w_1, A, v_1) generates the diagonal blocks

$$D_1 = \quad 0.541, \quad D_3 = -0.658, \quad D_4 = \quad 0.493, \quad D_5 = -0.479$$
$$D_6 = \quad 0.653, \quad D_7 = -0.328, \quad D_8 = -0.482, \quad D_9 = \quad 0.285$$

$$D_2 = \begin{bmatrix} -0.006 & 0.489 \\ 0.722 & 0.571 \end{bmatrix}.$$

In this example the check (4.6) offers a considerable improvement of the block biorthogonality:

$$\tau_{10}(\text{without (4.6)}) = 2.5 \cdot 10^{-3} \qquad \tau_{10}(\text{with (4.6)}) = 1.0 \cdot 10^{-5}.$$

Condition (4.6) can be fulfilled although the block biorthogonality is nearly lost. Consider the case where $Av_{\hat{\nu}(k)}$ and $\Pi_{\hat{\nu}(k)}Av_{\hat{\nu}(k)}$ are nearly linearly dependent and have about the same norm. Then $\mu(k) \approx 1$ although the block biorthogonality is nearly lost after few steps (see Example 4.4 below).

Example 4.4 Let $\xi \geq 1$. Then the monomial look-ahead Lanczos method together with (4.6) applied to

$$A = [\sqrt{i+j}]_{1 \leq i,j \leq 5}\,, \qquad w_1 = v_1 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \end{bmatrix}^T$$

generates the diagonal blocks $D_k = (1)$ for $1 \leq k \leq 5$. In this process the condition (4.6) is fulfilled at every step

$$\mu(1) = 9.9 \cdot 10^{-1}, \quad \mu(2) = 9.9 \cdot 10^{-1}, \quad \mu(3) = 9.6 \cdot 10^{-1}, \quad \mu(4) = 9.7 \cdot 10^{-1}, \quad \mu(5) = 5.2 \cdot 10^{-3}$$

although $\tau_5 = 5.2 \cdot 10^{-3}$.

For $k \in \{1, 2, \ldots, \hat{b}-1\}$ we propose a further check (4.9) to guarantee that the subtraction $Av_{\hat{\nu}(k)} - \Pi_{\hat{\nu}(k)}Av_{\hat{\nu}(k)}$ is not ill-conditioned:

$$\Psi \geq \psi(k) := \max\{\alpha_1^{(k)}, \beta_1^{(k)}\} \tag{4.9}$$

where $\beta_1^{(k)} := ||(I - \Pi_{\hat{\nu}(k)})Av_{\hat{\nu}(k)}||^{-1}$, $\alpha_1^{(k)} := ||(I - \Pi_{\hat{\nu}(k)}^T)A^T w_{\hat{\nu}(k)}||^{-1}$ and Ψ is a constant. The values $\beta_1^{(k)}$ and $\alpha_1^{(k)}$ are used to normalize the vectors $w_{\hat{\nu}(k)+1}$ and $v_{\hat{\nu}(k)+1}$ if the conditions (4.6) and (4.9) are fulfilled. Hence, check (4.9) does not cause any computational costs if the checks (4.6) and (4.9) are passed.

Example 4.5 (Continuing Example 4.4) Let (w_1, A, v_1) be defined as in Example 4.4. By using check (4.9) the maximal ψ-value can be reduced. This leads to an improvement in block biorthogonality:

$$\tau_5(\text{without (4.9)}) = 5.2 \cdot 10^{-3} \qquad \tau_5(\text{with (4.9)}) = 2.3 \cdot 10^{-4}.$$

Summarizing the checks: The k-th block is closed (i.e., a regular vector is generated) only if

$$\sigma_{min}(D_k) \geq tol \tag{4.10}$$

(the constant $tol > 0$ is chosen by the user) and the check

$$1 \geq \Upsilon(k) := \max\left\{\frac{\psi(k)}{\Psi}, \frac{\mu(k)}{\xi}\right\} \tag{4.11}$$

holds ($\psi(\hat{b}) := 0$). If condition (4.11) cannot be fulfilled by increasing the block size we continue with that block size dim_k which produces the minimal $\Upsilon(k)$ value. In this case we update the values Ψ and ξ as follows: Let $\psi(k)$ and $\mu(k)$ be the values belonging to the block size dim_k. If the value $\mu(k)$ ($\psi(k)$) is below the upper bound $\varsigma \cdot \xi$ ($\hat{\varsigma} \cdot \Psi$) and greater than $\mu(k-1)$ ($\psi(k-1)$) then ξ (Ψ) is replaced by the number $\mu(k)$ ($\psi(k)$). If the upper bound $\varsigma \cdot \xi$ ($\hat{\varsigma} \cdot \Psi$) is smaller than the number $\mu(k)$ ($\psi(k)$) then we update ξ by $\xi := \varsigma \cdot \xi$ and Ψ by $\Psi := \hat{\varsigma} \cdot \Psi$.

4.2 A stable version of the minimal look-ahead Lanczos method

Notation:

- $\epsilon :=$ relative rounding error for the computation of $\langle w, v \rangle$.
- $D_{k+1}^l = (\delta_{ij}^k)_{1 \leq i,j \leq l+1} := (\langle w_{\hat{\nu}(k)+i}, v_{\hat{\nu}(k)+j} \rangle)_{1 \leq i,j \leq l+1}$.
- $\mathcal{A}_{ij}^k := \langle w_{\hat{\nu}(k)+i}, A v_{\hat{\nu}(k)+j} \rangle$.
- max is a variable which is determined by the user (or by available storage; see procedure 'inner step').
- The constants Ψ, ξ, tol and a are chosen by the user.

Now we turn to the construction of a 'stable' version of the minimal look-ahead Lanczos method.
Idea: If an inner step $\hat{\nu}(k)+i$ is numerically an 'exact inner step' (i.e. D_{k+1}^l is numerically singular for all $0 \leq l \leq i-1$) then the minimal Lanczos vectors are used. If a block has to be continued because check (4.10) of (4.11) fails the antidiagonal structure of D_{k+1} is lost. For the residual vectors of this block any basis can be chosen. In the following we use the monomial look-ahead Lanczos vectors as residual vectors.

4.2.1 The stable minimal look-ahead Lanczos algorithm

procedure inner step;
procedure compute vector $v(k,i)$;

$$v_{\hat{\nu}(k)+i+1} := Av_{\hat{\nu}(k)+i} - V_k D_k^{-1} W_k^T A v_{\hat{\nu}(k)+i}\,;$$
$$\beta_{i+1}^{(k)} := 1/\|v_{\hat{\nu}(k)+i+1}\|; \qquad v_{\hat{\nu}(k)+i+1} := \beta_{i+1}^{(k)} v_{\hat{\nu}(k)+i+1}\,;$$

procedure compute minimal Lanczos vector $w(k,i)$;

$$w_{\hat{\nu}(k)+i+1} := A^T w_{\hat{\nu}(k)+i} - \mathcal{A}_{i(\varrho+1)}^k / \delta_{1(\varrho+1)}^k w_{\hat{\nu}(k)+1} - W_k D_k^{-t} V_k^T A^T w_{\hat{\nu}(k)+i}\,;$$
$$\alpha_{i+1}^{(k)} := 1/\|w_{\hat{\nu}(k)+i+1}\|; \qquad w_{\hat{\nu}(k)+i+1} := \alpha_{i+1}^{(k)} w_{\hat{\nu}(k)+i+1}\,;$$

procedure compute monomial Lanczos vector $w(k,i)$;

$$w_{\hat{\nu}(k)+i+1} := A^T w_{\hat{\nu}(k)+i} - W_k D_k^{-t} V_k^T A^T w_{\hat{\nu}(k)+i}\,;$$
$$\alpha_{i+1}^{(k)} := 1/\|w_{\hat{\nu}(k)+i+1}\|; \qquad w_{\hat{\nu}(k)+i+1} := \alpha_{i+1}^{(k)} w_{\hat{\nu}(k)+i+1}\,;$$

procedure compute minimal Lanczos vectors;

$\varrho := 0$; $\delta_{1(\varrho+1)}^{k+1} := \langle w_{\hat{\nu}(k)+1}, v_{\hat{\nu}(k)+\varrho+1}\rangle$;
while $|\delta_{1(\varrho+1)}^{k+1}| \le \epsilon$ **and** $\hat{\nu}(k) + \varrho + 1 < N$ **do**

$\varrho := \varrho + 1$; compute vectors $v(k,\varrho)$; $\delta_{1(\varrho+1)}^{k+1} := \langle w_{\hat{\nu}(k)+1}, v_{\hat{\nu}(k)+\varrho+1}\rangle$;

if $|\delta_{1(\varrho+1)}^{k+1}| \le \epsilon$ **then** *incurable* := true
else for $i := 1$ **to** ϱ **do** compute minimal Lanczos vectors $w(k,\varrho)$;

procedure compute D_{k+1}^{ϱ};

Compute the entries of D_{k+1}^{ϱ}. If the absolute values of all entries above and below the antidiagonal are not greater than ϵ then $minimal_{k+1}$:= true else $minimal_{k+1}$:= false. If $minimal_{k+1}$ = true and the absolute value of all entries on the antidiagonal of D_{k+1}^{ϱ} are greater than $a \cdot \epsilon$ then *singular* := false else *singular* := true and $minimal_{k+1}$:= false.

($\star$ inner step $\star$)
incurable := false; compute minimal Lanczos vectors;
if not *incurable* **then**

compute D_{k+1}^{ϱ} ; $j := 0$;
if *singular* **then** *number* := 0 **else** *number* := 1;
if $\sigma_{min}(D_{k+1}^{\varrho+j}) < tol$ **then** *singular* :=true
stop:=$((\hat{\nu}(k) + \varrho + j + 1 \geq N)$ or $(max \leq number))$;
while (*singular* or (4.11) not true) and **not** stop **do**

stop:=$((\hat{\nu}(k) + \varrho + j + 1 \geq N)$ or $(max \leq number))$;
repeat

$j := j + 1$; compute vector $v(k, \varrho + j)$;
compute monomial vector $w(k, \varrho + j)$
if $\sigma_{min}(D_{k+1}^{\varrho+j}) < tol$ **then** *singular* :=true **else** *singular* :=false

until not *singular* **or** $\hat{\nu}(k) + \varrho + j + 1 = N$
if not *singular* **then** *number* := *number* + 1;

if *number* ≥ 1 **then**

$j :=$ block size for which $\Upsilon(k)$ is minimal;
update the values ξ and Ψ; $\hat{\nu}(k+1) := \hat{\nu}(k) + \varrho + j + 1$; $k := k + 1$;
$D_k := D_k^{\varrho+j}$;

else *incurable* := true;

procedure regular step;

procedure regular step for minimal Lanczos vectors;

$$\begin{aligned} v_{\hat{\nu}(k)+1} &:= Av_{\hat{\nu}(k)} - \sum_{i=0}^{dim_k-1} (\mathcal{A}^k_{(-i)0}/\delta^{k-1}_{(dim_k-i)(i+1)}) v_{\hat{\nu}(k-1)+i+1} \\ &\quad -V_{k-1}(D_{k-1})^{-1}(W_{k-1})^T A v_{\hat{\nu}(k)} \,; \\ w_{\hat{\nu}(k)+1} &:= A^T w_{\hat{\nu}(k)} - (\mathcal{A}^k_{00}/\delta^{k-1}_{1dim_k}) w_{\hat{\nu}(k-1)+1} - W_{k-1}(D_{k-1})^{-t}(V_{k-1})^T A^T w_{\hat{\nu}(k)} \,; \end{aligned}$$

procedure regular step else;

$$v_{\hat{\nu}(k)+1} := Av_{\hat{\nu}(k)} - V_k(D_k)^{-1}(W_k)^T A v_{\hat{\nu}(k)} - V_{k-1}(D_{k-1})^{-1}(W_{k-1})^T A v_{\hat{\nu}(k)} \,;$$

$$w_{\hat{\nu}(k)+1} := A^T w_{\hat{\nu}(k)} - W_k(D_k)^{-t}(V_k)^T A^T w_{\hat{\nu}(k)} - W_{k-1}(D_{k-1})^{-t}(V_{k-1})^T A^T w_{\hat{\nu}(k)} \,;$$

($\star$ regular step $\star$)
if $minimal_k$ **then** regular step for minimal Lanczos vectors **else** regular step else;
$\beta_1^{(k)} := 1/\|v_{\hat{\nu}(k)+1}\|$; $\quad v_{\hat{\nu}(k)+1} := \beta_1^{(k)} v_{\hat{\nu}(k)+1}$;
$\alpha_1^{(k)} := 1/\|w_{\hat{\nu}(k)+1}\|$; $\quad w_{\hat{\nu}(k)+1} := \alpha_1^{(k)} w_{\hat{\nu}(k)+1}$;
($\star$ Stable version of the minimal look-ahead Lanczos method $\star$)
$k := 0$; $\quad \hat{\nu}(k) := 0$;
repeat

if $k \geq 1$ **then** regular step;
inner step

until *incurable* **or** $\hat{\nu}(k) = N$.

4.2.2 Remarks on the implementation

In the following we give some implementation details for the computation of $V_k D_k^{-1}(W_k)^T A v_{\hat{\nu}(k)+i}, W_k(D_k^T)^{-1}(V_k)^T A^T w_{\hat{\nu}(k)+i}$ and $D_{k+1}^{\varrho+j}$. Our goal is to reduce the computational effort such that only two inner products have to be computed explicitly per step. (The computational effort for the check (4.11) is omitted.)

4.2.2.1 Computation of D_{k+1}^{ϱ}

All entries of D_{k+1}^{ϱ} can be computed by recursion from the first row of D_{k+1}^{ϱ} and the values

$$\mathcal{A}_{i(\varrho+1)}^{k} = \langle w_{\hat{\nu}(k)+i}, A v_{\hat{\nu}(k)+\varrho+1} \rangle \qquad 1 \leq i \leq \varrho \tag{4.12}$$

which already have been computed in the procedure 'compute minimal Lanczos vectors'.

Computation of the antidiagonal and all entries above the antidiagonal

For $sum := 2$ **to** $\varrho + 1$ **do**

for $i := 2$ **to** sum **do**

$$j := sum - i + 1;\ \delta_{ij}^{k} := \alpha_i^{(k)} (\delta_{(i-1)(j+1)}^{k} / \beta_{j+1}^{(k)} - \mathcal{A}_{(i-1)(\varrho+1)}^{k} \delta_{1j}^{k} / \delta_{1(\varrho+1)}^{k}).$$

Computation of the entries below the antidiagonal

All entries from the last column of D_{k+1}^{ϱ} vanish except the first entry. Therefore we can compute all entries below the antidiagonal by:

$l := 1$;
for $sum := \varrho$ **to** $2\varrho - 2$ **do**

$l := l + 1$;
for $j := \varrho$ **downto** l **do**

$$i := sum - j + 3; \quad u := -\mathcal{A}_{(i-1)(\varrho+1)}^{k} \delta_{1j}^{k} / \delta_{1(\varrho+1)}^{k};$$
if $j < \varrho$ **then** $u := u + \delta_{(i-1)(j+1)}^{k} / \beta_{j+1}^{(k)}$;
$\delta_{ij}^{k} := u\, \alpha_i^{(k)}$.

4.2.2.2 Computation of $D_{k+1}^{\varrho+j}$ for $j \geq 1$

The construction of $D_{k+1}^{\varrho+j}$ requires only the explicit computation of the inner products

$$\langle w_{\hat{\nu}(k)+\varrho+j}, v_{\hat{\nu}(k)+\varrho+j+1} \rangle \quad \text{and} \quad \langle w_{\hat{\nu}(k)+\varrho+j+1}, v_{\hat{\nu}(k)+\varrho+j+1} \rangle.$$

All other entries can be constructed by recursion from $D_{k+1}^{\varrho+j-1}$ and the values (4.12).

Computation of the last column
If $j = 1$ **then for** $i := 1$ **to** ϱ **do** $\delta^k_{i(\varrho+2)} := \mathcal{A}^k_{i(\varrho+1)}\beta^{(k)}_{\varrho+2}$
else for $i := 1$ **to** ϱ **do**

$$\delta^k_{i(\varrho+1+j)} := \beta^{(k)}_{j+1+\varrho}(\delta^k_{(i+1)(\varrho+j)}/\alpha^{(k)}_{i+1} + \delta^k_{1(\varrho+j)}/\delta^k_{1(\varrho+1)}\mathcal{A}^k_{i(\varrho+1)})\,;$$

for $i := \varrho+1$ **to** $\varrho+j-1$ **do** $\delta^k_{i(\varrho+j+1)} := \delta^k_{(i+1)(\varrho+j)}\beta^{(k)}_{\varrho+j+1}/\alpha^{(k)}_{i+1}$.

Computation of the last row
For $i := 1$ **to** $\varrho+j$ **do** $\delta^k_{(\varrho+1+j)i} := \delta^k_{(\varrho+j)(i+1)}\alpha^{(k)}_{\varrho+j+1}/\beta^{(k)}_{i+1}$.

4.2.2.3 Computation of $(W_{k-1})^T A v_{\hat{\nu}(k)}$, $(V_{k-1})^T A^T w_{\hat{\nu}(k)}$, $W_k^T A v_{\hat{\nu}(k)+i}$, $V_k^T A^T w_{\hat{\nu}(k)+i}$ for $i \geq 1$

In view of the identities

$$V_k^T A^T w_{\hat{\nu}(k)+i} = \frac{\delta^k_{i1}}{\beta^{(k)}_1} e_{dim_k}\,, \quad W_k^T A v_{\hat{\nu}(k)+i} = \frac{\delta^k_{1i}}{\alpha^{(k)}_1} e_{dim_k}$$

$$(V_{k-1})^T A^T w_{\hat{\nu}(k)} = \frac{\delta^{k-1}_{dim_{k-1}1}}{\beta^{(k-1)}_1} e_{dim_{k-1}}\,, \quad (W_{k-1})^T A v_{\hat{\nu}(k)} = \frac{\delta^{k-1}_{1dim_{k-1}}}{\alpha^{(k-1)}_1} e_{dim_{k-1}}$$

further inner products need not to be computed.

4.2.2.4 Computation of $V_k^T A^T w_{\hat{\nu}(k)}$

Only $\langle w_{\hat{\nu}(k)}, A v_{\hat{\nu}(k)}\rangle$ have to be computed explicitly:

$$V_k^T A^T w_{\hat{\nu}(k)} = \sum_{i=2}^{dim_k} \frac{\delta^{k-1}_{dim_k i}}{\beta^{(k-1)}_i} e_{i-1} + \langle w_{\hat{\nu}(k)}, A v_{\hat{\nu}(k)}\rangle e_{dim_k}.$$

4.2.2.5 Computation of $W_k^T A v_{\hat{\nu}(k)}$

If $j = 0$ (i.e. $D_k = D_k^\varrho$) then all entries have been computed before. If $j \neq 0$ then the construction of the entries does not require the computation of any further inner product:
For $i := 1$ **to** ϱ **do** $\mathcal{A}^{k-1}_{idim_k} := \delta^{k-1}_{(i+1)dim_k}/\alpha^{(k-1)}_{i+1} + \mathcal{A}^{k-1}_{i(\varrho+1)}\delta^{k-1}_{1dim_k}/\delta^{k-1}_{1(\varrho+1)}$;
for $i := \varrho+1$ **to** j **do** $\mathcal{A}^{k-1}_{idim_k} := \delta^{k-1}_{(i+1)dim_k}/\alpha^{(k-1)}_{i+1}$.

4.3 Comparision of the minimal look-ahead Lanczos method and the monomial look-ahead Lanczos method

4.3.1 Computational work

Both methods require the same number of explicit computations of inner products per step as the ordinary Lanczos method. If no exact block is constructed then the minimal

look-ahead Lanczos method requires about the same computational effort as the monomial look-ahead Lanczos method. However, if an exact block is constructed then the minimal look-ahead Lanczos method requires less storage and computational work (see Section 3.2.3).

4.3.2 Numerical quality of the generated bases

In this subsection let $A \in \mathrm{Mat}_N(\mathbb{R})$ and v_1, $w_1 \in \mathbb{R}^N$. Assume that the minimal look-ahead Lanczos and the monomial look-ahead Lanczos method terminate after step N. If no exact block is constructed then the two methods generate bases with similar numerical properties. The following discussion shows that the minimal look-ahead Lanczos method has better stability properties than the monomial look-ahead Lanczos method if an exact block is constructed.

Theorem 4.5 shows that the left minimal Lanczos vectors w_i are the normal vectors to the hyperplanes generated by the right minimal Lanczos vectors v_i.

Theorem 4.5 *Let $(w_i, v_i)_{1\le i\le N}$ be a basis to $(\hat\nu(k))_{k\in L}$ and let D_{k+1} be antidiagonal. Further let R_l be the orthogonal projection onto $W_l^\perp$*

$$W_l := \mathrm{span}\{v_i : 1 \le i \le N, \quad i \ne l\}\,.$$

Then the equation

$$\langle w_{\hat\nu(k)+i+1}, v_{\hat\nu(k+1)-i}\rangle w_{\hat\nu(k)+i+1} = R_{\hat\nu(k+1)-i} v_{\hat\nu(k+1)-i} \tag{4.13}$$

holds for all $0 \le l, k \le \hat b - 1$, $0 \le j \le dim_{l+1} - 1$, $0 \le i \le dim_{k+1} - 1$.

Proof. By construction $w_{\hat\nu(k)+i+1} \in W^\perp_{\hat\nu(k+1)-i}$ and $\dim W^\perp_{\hat\nu(k+1)-i} = N - 1$. Hence the equation $\mathrm{span}\{w_{\hat\nu(k)+i+1}\} = W^\perp_{\hat\nu(k+1)-i}$ holds, which proves the assertion. ∎

The vectors generated by the monomial look-ahead Lanczos method do not have the property (4.13). This observation is the key for understanding the different numerical properties of the two methods.

4.3.2.1 Numerical properties of D_k

Assume that the k-th block is exact. Then the matrices D_k (minimal Lanczos) and D_k (monomial Lanczos) have the properties:

a) The condition number of the matrix D_k (minimal Lanczos) is determined by the quotient

$$\kappa_2(D_k) = \max\left\{ \frac{\cos(\Theta(W_{\hat\nu(k)-i}, v_{\hat\nu(k)+i}))}{\cos(\Theta(W_{\hat\nu(k)-j}, v_{\hat\nu(k)+j}))} \;: 1 \le i, j \le dim_k \right\}.$$

Here the angle $\Theta(W_{\hat{\nu}(k)-i}, v_{\hat{\nu}(k)+i})$ is defined as

$$\Theta(W_{\hat{\nu}(k)-i}, v_{\hat{\nu}(k)+i}) := \min\{\arccos|\langle x, v_{\hat{\nu}(k)+i}\rangle| \,:\, x \in W_{\hat{\nu}(k)-i}\,,\ ||x|| = 1\}.$$

In general the values of the angles $\Theta(W_{\hat{\nu}(k)-i}, v_{\hat{\nu}(k)+i})$ and $\Theta(W_{\hat{\nu}(k)-j}, v_{\hat{\nu}(k)+j})$ are of the same magnitude.

b) The singular values $\{\sigma_i \,:\, 1 \leq i \leq \hat{\nu}(k) - \hat{\nu}(k-1)\}$ of the matrix D_k (minimal Lanczos)

$$\sigma_i = ||R_{\hat{\nu}(k)-i} v_{\hat{\nu}(k)-i}|| = \cos(\Theta(W_{\hat{\nu}(k)-i}, v_{\hat{\nu}(k)+i}))$$

are bounded above by 1. The singular values of D_k (monomial) can become much greater.

a) and b) indicate that in general the condition number of D_k (minimal Lanczos) is smaller than the condition number of D_k (monomial Lanczos).

Convergence to an eigenvector of A^T

The monomial inner vectors w_i can converge to an eigenvector of A^T although the right vectors v_i are well-conditioned (i.e. the matrix $V^{(N)}$ is well-conditioned). This leads to a nearly singular matrix D_k (monomial Lanczos) and a very quick loss of block biorthogonality. However, in the minimal look-ahead Lanczos process (or any look-ahead Lanczos method which generates an antidiagonal block) this cannot happen if an exact block is constructed. This observation can be stated precisely as follows.

Corollary 4.1 *Let $(w_i, v_i)_{1\leq i\leq N}$ be a basis to $(\hat{\nu}(k))_{k\in\hat{K}}$ and let D_{k+1} be antidiagonal. Further let R_l and W_l be defined as in Theorem 4.5. Then the following identity holds for all $0 \leq l, k \leq \hat{b} - 1$, $0 \leq j \leq dim_{l+1} - 1$, $0 \leq i \leq dim_{k+1} - 1$:*

$$\begin{aligned} |\langle w_{\hat{\nu}(l)+j+1}, w_{\hat{\nu}(k)+i+1}\rangle| &= \frac{|\langle R_{\hat{\nu}(l+1)-j} v_{\hat{\nu}(l+1)-j}, R_{\hat{\nu}(k+1)-i} v_{\hat{\nu}(k+1)-i}\rangle|}{||R_{\hat{\nu}(l+1)-j} v_{\hat{\nu}(l+1)-j}||\,||R_{\hat{\nu}(k+1)-i} v_{\hat{\nu}(k+1)-i}||} \\ &= \cos(\Theta(W_{\hat{\nu}(k+1)-i}, W_{\hat{\nu}(k+1)-j})). \end{aligned}$$

Here the angle $\Theta(W_{\hat{\nu}(k+1)-i}, W_{\hat{\nu}(k+1)-j})$ between the hyperplanes is defined as

$$\Theta(W_{\hat{\nu}(k+1)-i}, W_{\hat{\nu}(k+1)-j}) := \arccos\{|\frac{\langle x, y\rangle}{||x||\,||y||}| \,:\, x \perp W_{\hat{\nu}(k+1)-i}\,, y \perp W_{\hat{\nu}(k+1)-j}\}.$$

Corollary 4.1 shows that in the general case the condition numbers of the matrices $V^{(N)}$ and $W^{(N)}$ are of the same magnitude. Corollary 4.1 also shows that the left vectors w_i are orthogonal if the right vectors v_i are orthogonal.

Example 4.6 (Convergence to an eigenvector of A^T) Let A be defined as in Example 4.1 and let $w_1 = e_{10}$, $v_1 = e_1 - e_3$. The monomial inner vectors w_i converge to an eigenvector of A^T (see Table 4.6.2).

	Monomial Lanczos $\Psi \geq 100$, $\xi \geq 11$ or $\Psi \leq 10^9$, $\xi \geq 5$	Monomial Lanczos $5 \leq \xi \leq 10$ without (4.9)	Minimal Lanczos $\Psi \geq 100$, $\xi \geq 11$	Minimal Lanczos $5 \leq \xi \leq 10$ $\Psi \geq 100$
$\hat{\nu}(1)$	8	9	8	9
$\hat{\nu}(2)$	9	10	9	10
$\hat{\nu}(3)$	10	–	10	–
$\tau_1 0$	$2.6 \cdot 10^{-05}$	$5.8 \cdot 10^{-05}$	$3.6 \cdot 10^{-15}$	$6.0 \cdot 10^{-14}$
$\kappa_2(D_1)$	$6.5 \cdot 10^{09}$	$3.5 \cdot 10^{11}$	$2.0 \cdot 10^{00}$	$3.4 \cdot 10^{02}$
$\kappa_2(W_1)$	$6.5 \cdot 10^{09}$	$1.4 \cdot 10^{11}$	$5.4 \cdot 10^{00}$	$1.3 \cdot 10^{02}$
$\kappa_2(V_1)$	$3.1 \cdot 10^{00}$	$3.5 \cdot 10^{00}$	$3.1 \cdot 10^{00}$	$3.5 \cdot 10^{00}$
$\kappa_2(W^{(10)})$	$9.4 \cdot 10^{09}$	$1.9 \cdot 10^{11}$	$9.0 \cdot 10^{00}$	$1.8 \cdot 10^{02}$
$\kappa_2(V^{(10)})$	$8.1 \cdot 10^{00}$	$8.8 \cdot 10^{00}$	$8.1 \cdot 10^{00}$	$8.8 \cdot 10^{00}$
$\mu_w(1)$	$1.0 \cdot 10^{00}$	$1.0 \cdot 10^{00}$	$9.8 \cdot 10^{-01}$	$1.0 \cdot 10^{00}$
$\mu_v(1)$	$1.1 \cdot 10^{01}$	$1.2 \cdot 10^{00}$	$1.1 \cdot 10^{01}$	$1.2 \cdot 10^{00}$
$\mu_w(2)$	$1.0 \cdot 10^{00}$	$8.0 \cdot 10^{09}$	$1.0 \cdot 10^{00}$	$1.1 \cdot 10^{01}$
$\mu_v(2)$	$1.1 \cdot 10^{00}$	$7.6 \cdot 10^{00}$	$1.1 \cdot 10^{01}$	$7.6 \cdot 10^{00}$
$\mu_w(3)$	$8.2 \cdot 10^{-01}$	–	$8.2 \cdot 10^{-01}$	–
$\mu_v(3)$	$1.3 \cdot 10^{00}$	–	$1.3 \cdot 10^{00}$	–
$\alpha_1^{(1)}$	$7.5 \cdot 10^{08}$	–	$8.9 \cdot 10^{-01}$	–
$\beta_1^{(1)}$	$1.8 \cdot 10^{-02}$	–	$1.8 \cdot 10^{-02}$	–
$\alpha_1^{(2)}$	$7.7 \cdot 10^{-01}$	$1.2 \cdot 10^{10}$	$7.8 \cdot 10^{-01}$	$1.7 \cdot 10^{01}$
$\beta_1^{(2)}$	$3.9 \cdot 10^{00}$	$1.9 \cdot 10^{00}$	$3.9 \cdot 10^{00}$	$1.9 \cdot 10^{00}$

Tab.4.6

This leads to nearly singular matrices W_1(monomial) and D_1(monomial). The matrix $V^{(10)}$ (the common right vectors v_i) is well-conditioned. Therefore the minimal Lanczos vectors w_i cannot converge to an eigenvector of A^T. A consequence of this behaviour is that the minimal look-ahead Lanczos method provides a better block biorthogonality (measured by τ_N) and well-conditioned matrices $D^{(N)}$ and $W^{(N)}$ (see Table 4.6). Table

4.6 also shows that the properties of the monomial Lanczos vectors can be improved by using check (4.11) instead of check (4.6).

	Monomial Lanczos method	Minimal Lanczos method
$\langle w_1, w_2 \rangle$	0.98752338877845	0.00000000000000
$\langle w_2, w_3 \rangle$	0.99994838016513	0.00000000000000
$\langle w_3, w_4 \rangle$	0.99999988173873	0.00000000000000
$\langle w_4, w_5 \rangle$	0.99999999972924	0.00000000000000
$\langle w_5, w_6 \rangle$	0.99999999999938	0.00000000000000
$\langle w_6, w_7 \rangle$	1.00000000000000	0.00000000000000
$\langle w_7, w_8 \rangle$	1.00000000000000	0.00000000000000

Tab.4.6.2

Convergence to an eigenvector of A

If the inner right vectors v_i converge to an eigenvector of A then, in general, the relation 4.13 between the left and right minimal Lanczos vectors the condition number of the matrix W_k (minimal Lanczos) is not much better than the condition number of V_k. However the condition number of D_k (minimal Lanczos) is less than the condition numbers of V_k and D_k (monomial). The following example demonstrates this behaviour.

Example 4.7 (Convergence to an eigenvector of A) Let $(\tilde{w}_1, \tilde{A}, \tilde{v}_1)$ be the dual system to (w_1, A, v_1) defined in Example 4.6 (i.e. $\tilde{w}_1 = v_1$, $\tilde{v}_1 = w_1$, $\tilde{A} = A^T$). The bases generated by the monomial look-ahead Lanczos method applied to $(\tilde{w}_1, \tilde{A}, \tilde{v}_1)$ and (w_1, A, v_1), respectively, are identical if the right vectors v_i are interchanged with the left vectors w_i. Hence (see Example 4.6) the common minimal Lanczos and monomial $\tilde{v}$-vectors converge to an eigenvector of $\tilde{A}$ and the matrix $\tilde{W}$ (monomial) is well-conditioned. Although the left vectors $\tilde{w}$ of the minimal look-ahead Lanczos method are nearly linearly dependent, the condition number of the matrix $\tilde{D}_1$ (minimal Lanczos) is less than the condition number of the matrix generated by the monomial look-ahead Lanczos method. The values of τ show that the minimal look-ahead Lanczos method yields a better block biorthogonality than the monomial look ahead Lanczos method (see Table 4.7).

	Monomial Lanczos $\Psi \geq 100$, $\xi \geq 11$ or $\Psi \leq 10^9$, $\xi \geq 5$	Monomial Lanczos $5 \leq \xi \leq 10$ without (4.9)	Minimal Lanczos $\xi \geq 5$ $\Psi \geq 100$
$\hat{\nu}(1)$	8	9	8
$\hat{\nu}(2)$	9	10	9
$\hat{\nu}(3)$	10	–	10
$\kappa_2(D_1)$	$6.5 \cdot 10^{09}$	$3.5 \cdot 10^{11}$	$3.0 \cdot 10^{07}$
$\kappa_2(W_1)$	$3.1 \cdot 10^{00}$	$3.5 \cdot 10^{00}$	$7.3 \cdot 10^{09}$
$\kappa_2(V_1)$	$6.5 \cdot 10^{09}$	$1.4 \cdot 10^{11}$	$6.5 \cdot 10^{09}$
$\kappa_2(W^{(10)})$	$8.1 \cdot 10^{00}$	$8.8 \cdot 10^{00}$	$8.9 \cdot 10^{09}$
$\kappa_2(V^{(10)})$	$9.4 \cdot 10^{09}$	$1.9 \cdot 10^{11}$	$9.4 \cdot 10^{09}$
$\tau_1 0$	$5.4 \cdot 10^{-05}$	$8.0 \cdot 10^{-05}$	$3.5 \cdot 10^{-08}$
$\mu_w(1)$	$1.1 \cdot 10^{01}$	$1.2 \cdot 10^{00}$	$8.7 \cdot 10^{-01}$
$\mu_v(1)$	$1.0 \cdot 10^{00}$	$1.0 \cdot 10^{00}$	$1.0 \cdot 10^{00}$
$\mu_w(2)$	$1.1 \cdot 10^{00}$	$7.6 \cdot 10^{00}$	$1.1 \cdot 10^{00}$
$\mu_v(2)$	$1.0 \cdot 10^{00}$	$8.1 \cdot 10^{09}$	$1.0 \cdot 10^{00}$
$\mu_w(3)$	$1.3 \cdot 10^{00}$	–	$1.3 \cdot 10^{00}$
$\mu_v(3)$	$8.2 \cdot 10^{-01}$	–	$8.2 \cdot 10^{-01}$
$\alpha_1^{(1)}$	$1.8 \cdot 10^{-02}$	–	$1.0 \cdot 10^{00}$
$\beta_1^{(1)}$	$7.5 \cdot 10^{08}$	–	$7.5 \cdot 10^{08}$
$\alpha_1^{(2)}$	$3.9 \cdot 10^{00}$	$1.9 \cdot 10^{00}$	$3.9 \cdot 10^{00}$
$\beta_1^{(2)}$	$7.7 \cdot 10^{-01}$	$1.2 \cdot 10^{10}$	$7.7 \cdot 10^{-01}$

Tab.4.7

The values $\zeta(k)$, $\mu(k)$ and $\psi(k)$

If the minimal look-ahead Lanczos method and the monomial look-ahead Lanczos method generate bases to the same indices $(\hat{\nu}(k))_{k \in \hat{K}}$ then the following holds:

a) The minimal look-ahead Lanczos method and the monomial look-ahead Lanczos method yield the same values $\zeta_v^1(k)$, $\zeta_v^2(k)$, $\mu_v^1(k)$, $\mu_v^2(k)$ and $\beta_1^{(k)}$.

b) Let the k-th block be exact. By construction of the left vectors w_i of the minimal

look-ahead Lanczos method, the vector $A^T w_{\hat{\nu}(k)}$ is orthogonal to the vectors

$$v_{\hat{\nu}(k-1)+1}, v_{\hat{\nu}(k-1)+2}, \ldots v_{\hat{\nu}(k)-1}.$$

Hence in the k-th regular step (construction of $w_{\hat{\nu}(k)+1}$) we only have to force the orthogonality against $v_{\hat{\nu}(k)}$ and the $(k-1)$-st block. This leads in many cases to a value of $\alpha_1^{(k)}$ which is smaller than the the value of $\alpha_1^{(k)}$(monomial).

c) $\zeta_w^1(k)$ is constructed as an approximation of $\|W_k D_k^{-t} V_k^T A^T w_{\hat{\nu}(k)}\|$. Because of the inner biorthogonality the values $\zeta_w^1(k)$(minimal Lanczos) and $\|W_k D_k^{-t} V_k^T A^T w_{\hat{\nu}(k)}\|$ are identical if an exact block is constructed. In general, the value $\zeta_w^1(k)$(monomial) do not have this property.

5 Krylov subspace projection methods

5.1 A general concept

Up to know we have discussed the Arnoldi method and the look-ahead Lanczos methods. What do these methods have in common ?
Both methods applied to a matrix $A \in \mathrm{Mat}_N$ and a starting vector $v_1 \in \mathbb{C}^N$ generate matrices $V \in \mathrm{Mat}_{N\times s}$, $H \in \mathrm{Mat}_{s\times s}$ (for some $s \leq N$) with the properties:

a) The column vectors of V form a basis of the Krylov chain $(\mathcal{K}(i, A, v_1))_{1\leq i\leq s}$.

b) There is a projection Π onto $\mathcal{K}(s, A, v_1)$ such that the equation

$$VH = \Pi AV \tag{5.1}$$

holds.

In the case of the Arnoldi method the matrix Π is the orthogonal projection onto $\mathcal{K}(s, A, v_1)$ and in the case of the (look-ahead) Lanczos method Π is the projection defined by

$$\mathrm{Im}\,\Pi = \mathcal{K}(s, A, v_1) \qquad \text{and} \qquad \mathrm{Ker}\,\Pi = \mathcal{K}(s, A^T, w_1)^\perp$$

where w_1 is the left starting vector of the (look-ahead) Lanczos method. In the following we call methods which generate matrices $V \in \mathrm{Mat}_{N\times s}$ and $H \in \mathrm{Mat}_{s\times s}$, satisfying the conditions a) and b) above, Krylov subspace projection methods.
The Lanczos method and the Arnoldi method are 'extremal points' in the set of Krylov subspace projection methods. Indeed, the Lanczos method requires minimal computational effort and storage but the generated basis can be ill-conditioned. The Arnoldi method generates an orthogonal basis but it requires much more computational effort and storage than the Lanczos method. Therefore we ask how to construct a Krylov subspace projection method 'lying' between the Lanczos and the Arnoldi method in the sense that it requires small computational effort and storage and the generated basis is not ill-conditioned. The look-ahead Lanczos method is such a method. But is this really the best method or are there other Krylov subspace projection methods which have much better numerical properties ?
This question was the starting point for studying general Krylov subspace projection methods. Here we do not provide an ultimate answer. But we present an appropriate setting for the construction of stable and efficient Krylov subspace projection methods and offer some theoretical results which are useful to distinguish good methods from bad ones. Further we derive results which are used in Chapter 6 for convergence proofs of restarted Krylov subspace projection methods.

First we show that every matrix H generated by a Krylov subspace projection method is an unreduced upper Hessenberg matrix.
As described in Remark 2.1 the equation (5.1) can be rewritten in the form

$$VH = AV - \{(I - \Pi)AVe_s\}e_s^T$$

which is equivalent to

$$\{(I - \Pi)AVe_s\}e_s^T = AV - VH.$$

In component notation, this equation reads

$$0 = Av_i - \sum_{j=1}^{s} h_{j,i}v_j, \qquad i = 1, 2, \ldots, s-1 \tag{5.2}$$

$$(I - \Pi)Av_s = Av_s - \sum_{j=1}^{s} h_{j,s}v_j \tag{5.3}$$

where $H = [h_{i,j}]$ and $V = [v_1,\, v_2,\, \ldots,\, v_s]$.
Using

$$\begin{aligned} v_{i+1} &\in \mathcal{K}(i+1, A, v_1) = \mathrm{span}\{v_1, v_2, \ldots, Av_i\} \\ Av_i &\notin \mathcal{K}(i, A, v_1) = \mathrm{span}\{v_1, v_2, \ldots, v_i\} \end{aligned}$$

we conclude

$$h_{i+1,i} \neq 0 \qquad \text{and} \qquad h_{j,i} = 0 \quad \text{for} \quad j > i+1.$$

Hence we have proved Lemma 5.1 below.

Lemma 5.1 *Every matrix $H = [h_{i,j}]$, generated by a Krylov subspace projection method, is an unreduced upper Hessenberg matrix. Furthermore the equations*

$$\begin{aligned} h_{i+1,i}v_{i+1} &= Av_i - \sum_{j=1}^{i} h_{j,i}v_j, \qquad i = 1, 2, \ldots, s-1 \\ (I - \Pi)Av_s &= Av_s - \sum_{j=1}^{s} h_{j,s}v_j \end{aligned}$$

hold.

For simplicity and clarity we introduce 'Krylov tuples'.

Definition 5.1 (Krylov tuple) *A five-tuple (A, V, Π, H, s) with A, $\Pi \in \mathrm{Mat}_N$, $s \in \{1, 2, \ldots, N\}$, $V \in \mathrm{Mat}_{N\times s}$ and $H \in \mathrm{Mat}_s$ is called a Krylov tuple if the column vectors of V form a basis of the Krylov chain $(\mathcal{K}(j, A, v_1))_{1\leq j\leq s}$, Π is a projection onto $\mathcal{K}(s, A, v_1)$ and $VH = \Pi AV$.*

Hence a Krylov subspace projection method is a method which generates matrices $V \in \mathrm{Mat}_{N\times s}$ and $H \in \mathrm{Mat}_s$ such that (A, V, Π, H, s) is a Krylov tuple for an appropriate choice of s and Π.
Theorem 5.1 below clarifies the connection between matrices H and $\hat{H}$ generated by different Krylov subspace projection methods.

Theorem 5.1 *Let (A, V, Π, H, s) and $(A, \hat{V}, \hat{\Pi}, \hat{H}, s)$ be Krylov tuples.*
Let $Ve_1 \in \mathrm{span}\{\hat{V}e_1\}$. Then the following holds:

a) There is an invertible upper triangular matrix $R \in \mathrm{Mat}_N$ such that $V = \hat{V}R$ and $He_i = R^{-1}\hat{H}Re_i$ for $i = 1, 2, \ldots, s-1$. If also $\Pi = \hat{\Pi}$ holds, then $H = R^{-1}\hat{H}R$.

b) If $s = m := \dim \mathcal{K}(N, A, Ve_1)$ then $H = R^{-1}\hat{H}R$.

Proof. By definition of a basis of a chain of subspaces there is an invertible upper triangular matrix $R \in \mathrm{Mat}_N$ such that $V = \hat{V}R$. Using $VH = \Pi AV$ we obtain $\hat{V}RH = \Pi A\hat{V}R$. Hence $\hat{V}RHR^{-1} = \Pi A\hat{V}$. Finally the relations

$$A\hat{v}_j \in \mathcal{K}(j+1, A, v_1) = \mathrm{Im}\,\Pi = \mathrm{Im}\,\hat{\Pi}, \qquad \text{for} \quad j = 1, 2 \ldots, s-1$$

yield

$$\begin{aligned} \hat{V}RHR^{-1}e_i &= \Pi A\hat{V}e_i = [A\hat{v}_1, A\hat{v}_2, \ldots, A\hat{v}_{s-1}, \Pi A\hat{v}_s]e_i \\ &= [A\hat{v}_1, A\hat{v}_2, \ldots, A\hat{v}_{s-1}, \hat{\Pi}A\hat{v}_s]e_i = \hat{\Pi}A\hat{V}e_i \end{aligned}$$

for all $i \in \{1, 2, \ldots, s-1\}$ which proves a) and b). ∎

Now we discuss how to construct a Krylov subspace projection method or, equivalently, how to generate a Krylov tuple step by step.
Suppose that we have already generated a Krylov tuple (A, V, Π, H, s). In the next step we construct a Krylov tuple $(A, \tilde{V}, \tilde{\Pi}, \tilde{H}, r)$ with the property

$$r \geq s, \qquad \tilde{V} = [V, \star] \qquad \text{and} \qquad \tilde{H} = \begin{bmatrix} H & \star \\ \star & \star \end{bmatrix}.$$

Such a Krylov tuple $(A, \tilde{V}, \tilde{\Pi}, \tilde{H}, r)$ is called a continuation of (A, V, Π, H, s). The following theorem characterizes continuations of a Krylov tuple.

Theorem 5.2 *Let $(A, \tilde{V}, \tilde{\Pi}, \tilde{H}, r)$ and (A, V, Π, H, s) be Krylov tuples and $s < r$. Suppose that $\tilde{V} = [V, \star]$. Then $(A, \tilde{V}, \tilde{\Pi}, \tilde{H}, r)$ is a continuation of (A, V, Π, H, s) if and only if there is a number $\alpha \in \mathbb{C}$ such that $\tilde{V}e_{s+1} = \alpha(I - \Pi)AVe_s$.*

Proof. By Lemma 5.1 the identities

$$\begin{aligned} \tilde{h}_{s+1,s}\tilde{v}_{s+1} &= A\tilde{v}_s - \sum_{j=1}^{s} \tilde{h}_{j,s}\tilde{v}_j \\ (I-\Pi)Av_s &= Av_s - \sum_{j=1}^{s} h_{j,s}v_j \;=\; A\tilde{v}_s - \sum_{j=1}^{s} \tilde{h}_{j,s}\tilde{v}_j \end{aligned}$$

hold. Thus,

$$\tilde{h}_{s+1,s}\tilde{v}_{s+1} = (I-\Pi)Av_s$$

which proves the theorem. ∎

Theorem 5.2 yields the following algorithm for the construction of a continuation $(A, \tilde{V}, \tilde{\Pi}, \tilde{H}, r)$ of a Krylov tuple (A, V, Π, H, s).

Algorithm 5.1

(1) Choose $r \in \mathbb{N}$ such that $r > s$ and $r \leq \dim \mathcal{K}(N, A, v_1)$.

(2) Choose $\beta_{s+1} \in \mathbb{C} \setminus \{0\}$ and compute $v_{s+1} := \beta_{s+1}(I-\Pi)Av_s$.

(3) For $j = s+1, s+2, \ldots, r-1$ choose coefficients $\tilde{h}_{l,j} \in \mathbb{C}$, $\beta_{j+1} \in \mathbb{C} \setminus \{0\}$ and compute

$$v_{j+1} := \beta_{j+1} \left(Av_j - \sum_{l=1}^{j} \tilde{h}_{l,j} v_l \right).$$

(4) $V_2 := [v_{s+1}, v_{s+2}, \ldots, v_r]$, $\tilde{V} := [V, V_2]$.

(5) Choose an arbitrary full rank matrix $\tilde{W} \in \mathrm{Mat}_{N\times r}$ with the property that $\tilde{D} := \tilde{W}^*\tilde{V}$ is invertible. Define $\tilde{\Pi} := \tilde{V}\tilde{D}^{-1}\tilde{W}^*$.

(6) Put $(e_{s+1}^T \tilde{H} e_s) := \beta_{s+1}^{-1}$. For $j = s+1, \ldots, r-1$ set $\tilde{H}e_j := \beta_{j+1}^{-1} e_{j+1} + \sum_{l=1}^{j} \tilde{h}_{l,j} e_l$.

(7) $\tilde{H}e_r := \sum_{l=1}^{r} \tilde{h}_{l,r} e_l$ where $\tilde{\Pi}Av_r = \sum_{l=1}^{r} \tilde{h}_{l,r} v_l$.

The degrees of freedom in the construction of a continuation are the coefficients $\tilde{h}_{l,j}$ (for $j < r$), the matrix $\tilde{W}$ and the scaling factors β_{s+1}, β_{j+1}. The factors β_{s+1}, β_{j+1} are usually used to normalize the vectors v_{j+1}.

Choice of the matrix $\tilde{W}$

To reduce the costs for the computation of a full rank matrix $\tilde{W} \in \mathrm{Mat}_{N\times r}$ with the property $\det(\tilde{W}^*\tilde{V}) \neq 0$ one can choose the following strategy:

a) Define $\tilde{W} := [W, W_2]$, where the matrix $W \in \mathrm{Mat}_{N\times s}$ has already been computed to represent the projection $\Pi = V(W^*V)^{-1}W^*$ of the Krylov tuple (A, V, Π, H, s).

b) The full rank matrix $W_2 \in \mathrm{Mat}_{N\times(r-s)}$ is chosen such that $\det(\tilde{W}^*\tilde{V}) \neq 0$.

The result is a representation $\tilde{V}\tilde{D}^{-1}\tilde{W}^*$ of the projection $\tilde{\Pi}$ with

$$\mathrm{Im}\,\tilde{\Pi} = \mathcal{K}(r, A, Ve_1) \qquad \text{and} \qquad \mathrm{Ker}\,\tilde{\Pi} = \mathrm{Ker}\,\tilde{W}^*$$

where

$$\tilde{D} := \tilde{W}^*\tilde{V} = \begin{bmatrix} W^*V & W^*V_2 \\ W_2^*V & W_2^*V_2 \end{bmatrix}.$$

To reduce the costs for the inversion of the matrix $\tilde{D}$ one can choose the full rank matrix W_2 from above such that

$$W_2^*V = 0 \qquad \text{and} \qquad \det(W_2^*V_2) \neq 0. \tag{5.4}$$

Indeed, the condition (5.4) implies that $\tilde{W} = [W, W_2]$ has full rank and $\tilde{D}$ is invertible with the representation

$$\tilde{D}^{-1} = \begin{bmatrix} D^{-1} & -D^{-1}W^*V_2\check{D}^{-1} \\ 0 & \check{D}^{-1} \end{bmatrix}$$

where $D := W^*V$ and $\check{D} := W_2^*V_2$.

Choice of the coefficients $\tilde{h}_{l,j}$

A further reduction of the costs of the inversion of $\tilde{D}$ can be achieved by choosing the coefficients $\tilde{h}_{l,j}$ such that $W^*V_2 = 0$. The conditon $W^*V_2 = 0$ is equivalent to $\Pi\, V_2 = 0$. Thus, we obtain the following recursion for the construction of the matrix $V_2 = [v_{s+1}, v_{s+1}, \ldots, v_r]$ with the property $W^*V_2 = 0$:

For $j = s, s+1, \ldots, r-1$ choose coefficients $\tilde{h}_{l,j} \in \mathbb{C}$, $\beta_j \in \mathbb{C} \setminus \{0\}$ and compute

$$v_{j+1} := \beta_{j+1}\left(Av_j - \sum_{l=s+1}^{j} \tilde{h}_{l,j}v_l - \Pi Av_j\right). \tag{5.5}$$

The following theorem summarizes these observations.

Theorem 5.3 *Let (A, V, Π, H, s) be a Krylov tuple and let $V(W^*V)^{-1}W^*$ be a representation of the projection Π. If a continuation $(A, \tilde{V}, \tilde{\Pi}, \tilde{H}, r)$ is constructed by* Algorithm 5.1 *and the relations* (5.5) *and* (5.4) *are used instead of* (3) *and* (5) *then the matrices $\tilde{V}$, $\tilde{W}$, $\tilde{W}^*\tilde{V}$ and $\tilde{\Pi}$ have the representations:*

a) $\tilde{W} = [W, W_2]$ *and* $\tilde{V} = [V, V_2]$.

b) $\tilde{W}^* \tilde{V} = \begin{bmatrix} W^*V & 0 \\ 0 & W_2^* V_2 \end{bmatrix}$.

c) $\tilde{\Pi} = \Pi + \check{\Pi}$ *where* $\check{\Pi} := V_2 (W_2^* V_2)^{-1} W_2^*$.

Remark 5.1 Theorem 5.3 c) is equivalent to $\Pi V_2 = 0$ and $\check{\Pi} V = 0$.

Remark 5.2 It is evident that in Algorithm 5.1 and Theorem 5.3 the conjugate and transposed sign '$*$' can be replaced by the transposed sign 'T'.

The Arnoldi method, the Lanczos method and the look-ahead Lanczos methods are the most popular methods which generate continuations as described in Theorem 5.3 and Remark 5.2, respectively.

We call a Krylov tuple normalized if the column vectors of V are normalized in Euclidean norm $|| \cdot ||$.

Example 5.1 (Arnoldi Krylov tuples) The Arnoldi method (see Section 2.1) generates a sequence of normalized Krylov tuples

$$(A, V^{(1)}, \Pi_1, H^{(1)}, 1), \quad (A, V^{(2)}, \Pi_2, H^{(2)}, 2), \quad \ldots, \quad (A, V^{(m)}, \Pi_m, H^{(m)}, m)$$

where $m = \dim \mathcal{K}(N, A, v_1)$ and Π_i is the orthogonal projection onto $\mathcal{K}(i, A, v_1)$. Π_{i+1} has the representation $\Pi_{i+1} = \Pi_i + v_{i+1} v_{i+1}^*$. Hence the Arnoldi method generates in the i-th step a continuation $(A, V^{(i+1)}, \Pi_{i+1}, H^{(i+1)}, i+1)$ of $(A, V^{(i)}, \Pi_i, H^{(i)}, i)$ with the properties described in Theorem 5.3:

$$\tilde{V}^* \tilde{V} = \begin{bmatrix} V^*V & 0 \\ 0 & V_2^* V_2 \end{bmatrix} = I$$

and

$$\tilde{\Pi} = \Pi + \check{\Pi} = V (V^*V)^{-1} V^* + V_2 (V_2^* V_2)^{-1} V_2^* = V V^* + v_{i+1} v_{i+1}^*.$$

Here $V := V^{(i)}$, $V_2 := [v_{i+1}]$, $\tilde{V} := [V, V_2]$.

Example 5.2 (Lanczos Krylov tuples) The Lanczos method (see Section 2.2) generates two sequences of Krylov tuples

$$(A, V^{(1)}, \Pi_1, H^{(1)}, 1), \quad (A, V^{(2)}, \Pi_2, H^{(2)}, 2), \quad \ldots, \quad (A, V^{(s_L)}, \Pi_{s_L}, H^{(s_L)}, s_L)$$
$$(A^T, W^{(1)}, \Pi_1^T, \hat{H}^{(1)}, 1), \quad (A^T, W^{(2)}, \Pi_2^T, \hat{H}^{(2)}, 2), \quad \ldots, \quad (A^T, W^{(s_L)}, \Pi_{s_L}^T, \hat{H}^{(s_L)}, s_L)$$

where Π_i is the projection onto $\mathcal{K}(i, A, v_1)$ with $\mathrm{Ker}\,\Pi = \mathcal{K}(i, A^T, w_1)^\perp$. The projection Π_{i+1} has the representation

$$\Pi_{i+1} = \Pi_i + \frac{v_{i+1} w_{i+1}^T}{\langle w_{i+1}, v_{i+1} \rangle}.$$

The Lanczos tuple $(A, V^{(i+1)}, \Pi_{i+1}, H^{(i+1)}, i+1)$ is a continuation of $(A, V^{(i)}, \Pi_i, H^{(i)}, i)$ and $(A^T, W^{(i+1)}, \Pi_{i+1}, \hat{H}^{(i+1)}, i+1)$ is a continuation of $(A^T, W^{(i)}, \Pi_i^T, \hat{H}^{(i)}, i)$.
The identity

$$\tilde{W}^T \tilde{V} = \begin{bmatrix} W^T V & 0 \\ 0 & W_2^T V_2 \end{bmatrix} = \mathrm{diag}\,(\delta_1, \delta_2, \ldots, \delta_{s_L})$$

shows that the Lanczos method generates continuations as described in Remark 5.2. Here $V := V^{(i)}$, $V_2 := [v_{i+1}]$, $\tilde{V} := [V, V_2]$, $W := W_i$, $W_2 := [w_{i+1}]$ and $\tilde{W} := [V, V_2]$.

Example 5.3 (Practical look-ahead Lanczos Krylov tuples)
Every (practical) look-ahead Lanczos method (see Theorem 4.3) generates two sequences

$$(A, V^{(\hat{\nu}(1))}, \Pi_{\hat{\nu}(1)}, H^{(\hat{\nu}(1))}, \hat{\nu}(1)), (A, V^{(\hat{\nu}(2))}, \Pi_{\hat{\nu}(2)}, H^{(\hat{\nu}(2))}, \hat{\nu}(2)), \ldots, (A, V^{(r)}, \Pi_r, H^{(r)}, r)$$

$$(A^T, W^{(\hat{\nu}(1))}, \Pi_{\hat{\nu}(1)}^T, \hat{H}^{(\hat{\nu}(1))}, \hat{\nu}(1)), (A^T, W^{(\hat{\nu}(2))}, \Pi_{\hat{\nu}(2)}^T, \hat{H}^{(\hat{\nu}(2))}, \hat{\nu}(2)), \ldots, (A^T, W^{(r)}, \Pi_r^T, \hat{H}^{(r)}, r)$$

of Krylov tuples. The projection $\Pi_{\hat{\nu}(k)}$ is defined as in Example 5.2 and has the representation

$$\Pi_{\hat{\nu}(k)} = \Pi_{\hat{\nu}(k-1)} + V_k (D_k)^{-1} (W_k)^T.$$

$(A, V^{(\hat{\nu}(k))}, \Pi_{\hat{\nu}(k)}, H^{(\hat{\nu}(k))}, \hat{\nu}(k))$ is a continuation of $(A, V^{(\hat{\nu}(k-1))}, \Pi_{\hat{\nu}(k-1)}, H^{(\hat{\nu}(k-1))}, \hat{\nu}(k-1))$ and $(A^T, W^{(\hat{\nu}(k))}, \Pi_{\hat{\nu}(k)}^T, \hat{H}^{(\hat{\nu}(k))}, \hat{\nu}(k))$ is a continuation of the Krylov tuple $(A^T, W^{(\hat{\nu}(k-1))}, \Pi_{\hat{\nu}(k-1)}^T, \hat{H}^{(\hat{\nu}(k-1))}, \hat{\nu}(k-1))$. The following equation shows that every look-ahead Lanczos method generates continuations as described in Remark 5.2:

$$(W^{(\nu(k))})^T V^{(\nu(k))} = \begin{bmatrix} (W^{(\nu(k-1))})^T V^{(\nu(k-1))} & 0 \\ 0 & W_k^T V_k \end{bmatrix} = \mathrm{diag}\,(D_1, D_2, \ldots, D_r).$$

5.2 Reducing the computational effort in the look-ahead Lanczos process

Suppose that $A^T = A$. Choose $w_1 = v_1$. Then the Lanczos recursions for the left and right vectors are identical. Consider the Gram-Schmidt projection Π_i used in the (look-ahead) Lanczos process:

$$\mathrm{Im}\,\Pi_i = \mathcal{K}(i, A, v_1) \tag{5.6}$$

$$\mathrm{Ker}\,\Pi_i = \mathcal{K}(i, A^T, w_1)^\perp = \mathcal{K}(i, A, v_1)^\perp. \tag{5.7}$$

Using that $(\operatorname{Im}\Pi_i)^\perp = \operatorname{Ker}\Pi_i^T$ and $(\operatorname{Ker}\Pi_i)^\perp = \operatorname{Im}\Pi_i^T$ we conclude

$$\operatorname{Ker}\Pi_i^T = \mathcal{K}(i, A, v_1)^\perp \tag{5.8}$$
$$\operatorname{Im}\Pi_i^T = \mathcal{K}(i, A, v_1). \tag{5.9}$$

Bringing the equations (5.6)-(5.9) together we find that $\Pi_i^T = \Pi_i$. Consequently the look-ahead Lanczos recursions for the left and right vectors are identical if the scaling coefficients $\alpha_i^{(k)}$ and $\beta_i^{(k)}$ and the coefficients for the inner vectors $f_{i.j}^{(k)}$ and $g_{i,j}^{(k)}$ are chosen such that

$$\alpha_i^{(k)} = \beta_i^{(k)} \qquad \text{and} \qquad f_{i.j}^{(k)} = g_{i,j}^{(k)}.$$

Is such a reduction also possible for an arbitrary matrix A ? The answer to this question is the subject of this section. Remember that the ordinary Lanczos method generates a biorthogonal basis with respect to the standard bilinear form $\langle x, y\rangle := x^T y$ on $\mathbb{C}^N \times \mathbb{C}^N$. First we discuss how we have to modify the Lanczos recursions to obtain a method which generates a biorthogonal basis $((w_i, v_i))_{1\le i\le s_L}$ with respect to a given non-degenerate bilinear form $\langle\cdot,\cdot\rangle_S$ on $\mathbb{C}^N \times \mathbb{C}^N$.

For this purpose we recall the definitions of non-degenerate bilinear forms and biorthogonal bases with respect to a form $\langle\cdot,\cdot\rangle_S$. For later use we further generalize the concept of degree indices $(\nu(k))_{k\in K}$, generalized biorthogonal bases and basis to $(\hat\nu(k))_{k\in\hat K}$ and list some fundamental results about adjoint operators and representation of projections.

Definition 5.2 (Non-degenerate bilinear form) *A bilinear form $\langle\cdot,\cdot\rangle_S : X \times Y \longrightarrow \mathbb{C}$ (where X and Y are vector spaces over $\mathbb{C}$) is called non-degenerate if the following implications hold:*

$$(\langle x, y\rangle_S = 0 \text{ for all } y \in Y) \Longrightarrow x = 0, \qquad (\langle x, y\rangle_S = 0 \text{ for all } x \in X) \Longrightarrow y = 0.$$

In the following $\langle\cdot,\cdot\rangle_S$ denotes a non-degenerate bilinear form on $\mathbb{C}^N \times \mathbb{C}^N$.

Lemma 5.2 (Representation) *The form $\langle\cdot,\cdot\rangle_S$ can be represented by an invertible matrix $S \in \operatorname{Mat}_N$:*

$$\langle x, y\rangle_S = x^T S y \qquad \text{for all} \quad x, y \in \mathbb{C}^N.$$

Definition 5.3 (Biorthogonal basis) *A basis $((w_i, v_i))_{1\le i\le s}$ of $((\mathcal{W}_i, \mathcal{V}_i))_{1\le i\le s}$ is called biorthogonal with respect to $\langle\cdot,\cdot\rangle_S$ if*

$$\langle w_i, v_j\rangle_S = 0 \quad \text{for} \quad i \ne j \quad \text{and} \quad \langle w_i, v_i\rangle_S \ne 0$$

for all $i, j \in \{1, 2, \ldots, s\}$.

Definition 5.4 (Orthogonal complement) *The subspace*

$$\mathcal{W}^{\perp_S} := \{v \in \mathbb{C}^N \mid \quad \langle w, v\rangle_S = 0 \quad \textit{for all} \quad w \in \mathcal{W}\} = (S^T\mathcal{W})^\perp$$

is called the (right) orthogonal complement of the subspace $\mathcal{W}$ of $\mathbb{C}^N$ with respect to $\langle\cdot,\cdot\rangle_S$.

Definition 5.5 (Adjoint matrix) *A matrix $A^+ \in \mathrm{Mat}_N$ is called (left-) adjoint of $A \in \mathrm{Mat}_N$ with respect to $\langle\cdot,\cdot\rangle_S$ if*

$$\langle x, Ay\rangle_S = \langle A^+x, y\rangle_S \qquad \textit{for all } x, y \in \mathbb{C}^N.$$

Lemma 5.3 *a) Let A^+ be the (left-) adjoint of an arbitrary matrix $A \in \mathrm{Mat}_N$ with respect to $\langle\cdot,\cdot\rangle_S$. Then A^+ has the representation $A^+ = (SAS^{-1})^T$.*

b) Let $\Pi \in \mathrm{Mat}_N$ be a projection onto an s-dimensional subspace. Let $W, V \in \mathrm{Mat}_{N\times s}$ be matrices of full rank such that $\mathrm{Ker}\,\Pi = (\mathrm{Im}\,W)^{\perp_S}$ and $\mathrm{Im}\,V = \mathrm{Im}\,\Pi$. Then $\Pi = VD^{-1}W^TS$ where $D := W^TSV$.

Definition 5.6 (Degree indices) *The degree indices $(\nu(k))_{k\in K}$ of $((\mathcal{W}_i, \mathcal{V}_i))_{1\le i\le s}$ with respect to $\langle\cdot,\cdot\rangle_S$ are defined as those indices*

$$l = \nu(0) := 0 < \nu(1) < \nu(2) < \ldots < \nu(b)$$

for which the bilinear form $\langle\cdot,\cdot\rangle_S|_{W_l\times V_l}$ is non-degenerate.

Remark 5.3 Let A^+ be the adjoint of $A \in \mathrm{Mat}_N$ with respect to $\langle\cdot,\cdot\rangle_S$. The degree indices $(\nu(k))_{k\in K}$ of a Krylov chain $((\mathcal{K}(i, A^+, w_1), \mathcal{K}(i, A, v_1)))_{1\le i\le s}$ with respect to $\langle\cdot,\cdot\rangle_S$ are exactly those indices l for which the matrix $[\langle w_i, v_j\rangle_S]_{1\le i,j\le l}$ is invertible. Here $((w_i, v_i))_{1\le i\le s}$ is an arbitrary basis of $((\mathcal{K}(i, A^+, w_1), \mathcal{K}(i, A, v_1)))_{1\le i\le s}$.

Definition 5.7 (Basis to $(\hat\nu(k))_{k\in\hat K}$) *Let $((\mathcal{W}_i, \mathcal{V}_i))_{1\le i\le s}$ be a chain of subspaces with degree indices $(\nu(k))_{k\in K}$ with respect to $\langle\cdot,\cdot\rangle_S$. Let $(\hat\nu(k))_{k\in\hat K} \subseteq (\nu(k))_{k\in K}$. A basis $((w_i, v_i))_{1\le i\le s}$ of the chain $((\mathcal{W}_i, \mathcal{V}_i))_{1\le i\le s}$ is called a basis to $(\hat\nu(k))_{k\in\hat K}$ with respect to $\langle\cdot,\cdot\rangle_S$ if the matrix $D^{(s)} := [\langle w_i, v_j\rangle_S)_{1\le i,j\le s}$ is block diagonal*

$$D^{(s)} = \mathit{diag}\,(D_1, D_2, \ldots, D_{\hat b})$$

with invertible diagonal blocks D_k of order dim_k (for $k = 1, 2, \ldots, \hat b$).
Here $\hat b := \max\{k \mid k \in \hat K\}$ and $dim_k := \hat\nu(k) - \hat\nu(k-1)$.

Definition 5.8 (Generalized biorthogonal basis) *A basis to $(\hat\nu(k))_{k\in\hat K}$ with respect to $\langle\cdot,\cdot\rangle_S$ is called generalized biorthogonal with respect to $\langle\cdot,\cdot\rangle_S$ if $(\hat\nu(k))_{k\in\hat K} = (\nu(k))_{k\in K}$.*

Now we turn to the construction of a biorthogonal basis of $((\mathcal{W}_i, \mathcal{V}_i))_{1\le i\le s}$ with respect to $\langle\cdot,\cdot\rangle_S$. Let $(\tilde{w}_i, \tilde{v}_i)_{1\le i\le s}$ be a basis of a chain of subspaces $((\mathcal{W}_i, \mathcal{V}_i))_{1\le i\le s}$. Consider the slight variation of the Gram-Schmidt biorthogonalisation method (VGS method) below.

VGS method

$v_1 := \tilde{v}_1$ and $w_1 := \tilde{w}_1$
for $i = 2\ldots, s$

$$\begin{aligned} v_i &:= (I - \Pi_{i-1})\tilde{v}_i \\ w_i &:= (I - \Pi_{i-1}^{+})\tilde{w}_i. \end{aligned}$$

Here

$$\Pi_i x = \sum_{j=1}^{i} \frac{\langle w_j, x\rangle_S}{\langle w_j, v_j\rangle_S} v_j$$

is the projection onto $\mathcal{V}_i$ with kernel $\mathcal{W}_i^{\perp_S}$. One easily verifies (see Section 2.2) that the VGS method yields the (up to scalar factors) uniquely determined biorthogonal basis $((w_i, v_i))_{1\le i\le s}$ of $((\mathcal{W}_i, \mathcal{V}_i))_{1\le i\le s}$ with respect to $\langle\cdot,\cdot\rangle_S$, if there is one. This leads to the following extension of Theorem 4.1:

Lemma 5.4 (Existence of biorthogonal bases) *The following statements are equivalent:*

a) There is a biorthogonal basis of $((\mathcal{W}_i, \mathcal{V}_i))_{1\le i\le s}$ with respect to $\langle\cdot,\cdot\rangle_S$.

b) The Gram-Schmidt projections $\Pi_1, \Pi_2, \ldots, \Pi_s$ exist.

c) The Gram-Schmidt biorthogonalisation method applied to an arbitrary basis of $((\mathcal{W}_i, \mathcal{V}_i))_{1\le i\le s}$ generates a biorthogonal basis with respect to $\langle\cdot,\cdot\rangle_S$.

d) The matrix $[w_1, w_2, \ldots, w_j]^T S[v_1, v_2, \ldots, v_j]$ is invertible for every choice of bases $((w_i, v_i))_{1\le i\le s}$ of $((\mathcal{W}_i, \mathcal{V}_i))_{1\le i\le s}$ and for all $j \in \{1, 2, \ldots, s\}$.

e) There is a basis $((w_i, v_i))_{1\le i\le s}$ of $((\mathcal{W}_i, \mathcal{V}_i))_{1\le i\le s}$ such that the matrix $[w_1, w_2, \ldots, w_j]^T S[v_1, v_2, \ldots, v_j]$ is invertible for all $j \in \{1, 2, \ldots, s\}$.

f) The bilinear form $\langle\cdot,\cdot\rangle_S|_{\mathcal{W}_j\times\mathcal{V}_j}$ is non-degenerate for all $j \in \{1, 2, \ldots, s\}$.

g) $\nu(k) = k$ for all $k \in \{1, 2, \ldots, s\}$.

Assume there exists a biorthogonal basis of a given chain $((\mathcal{K}(i, A^T, w_1), \mathcal{K}(i, A, v_1)))_{1\le i\le s}$ with respect to a given bilinear form $\langle\cdot,\cdot\rangle_S$. Computing

for $i = 1, 2 \ldots, s-1$

$$\begin{aligned} v_{i+1} &:= (I - \Pi_i) A v_i && (5.10) \\ w_{i+1} &:= (I - \Pi_i^+) A^+ w_i && (5.11) \end{aligned}$$

we obtain a biorthogonal basis $((w_i, v_i))_{1 \le i \le s}$ of $((\mathcal{K}(i, A^T, w_1), \mathcal{K}(i, A, v_1)))_{1 \le i \le s}$ with respect to $\langle \cdot, \cdot \rangle_S$. In a similar manner as described in Section 2.2 the recursions (5.10) and (5.11) reduce to the (modified) Lanczos three-term-recursions:

$$\begin{aligned} v_{i+1} &= A v_i - \frac{\langle w_i, A v_i \rangle_S}{\langle w_i, v_i \rangle_S} v_i - \frac{\langle w_i, v_i \rangle_S}{\langle w_{i-1}, v_{i-1} \rangle_S} v_{i-1} \\ w_{i+1} &= A^+ w_i - \frac{\langle w_i, A v_i \rangle_S}{\langle w_i, v_i \rangle_S} w_i - \frac{\langle w_i, v_i \rangle_S}{\langle w_{i-1}, v_{i-1} \rangle_S} w_{i-1}. \end{aligned}$$

Next we show how to generate a basis to $(\hat{\nu}(k))_{k \in \hat{K}}$ with respect to $\langle \cdot, \cdot \rangle_S$. Let $((\tilde{w}_i, \tilde{v}_i))_{1 \le i \le s}$ be a basis of $((\mathcal{W}_i, \mathcal{V}_i))_{1 \le i \le s}$. Suppose there is a basis to $(\hat{\nu}(k))_{k \in \hat{K}}$ with respect to $\langle \cdot, \cdot \rangle_S$. Consider the slightly modified extended Gram-Schmidt biorthogonalisation method defined as follows:

VMEGS

For $k = 0, 1, 2 \ldots, s-1$
  for $i = 1, 2, \ldots, dim_k$

$$\begin{aligned} v_{\hat{\nu}(k)+i} &:= (I - \Pi_{\hat{\nu}(k)}) \tilde{v}_{\hat{\nu}(k)+i} \\ w_{\hat{\nu}(k)+i} &:= (I - \Pi_{\hat{\nu}(k)}^+) \tilde{w}_{\hat{\nu}(k)+i}. \end{aligned}$$

Here $\Pi_{\hat{\nu}(k)}$ is the projection with $\operatorname{Im} \Pi_{\hat{\nu}(k)} = \mathcal{V}$ and $\operatorname{Ker} \Pi_{\hat{\nu}(k)} = \mathcal{W}^{\perp_S}$. The $VMEGS$ algorithm applied to $((\tilde{w}_i, \tilde{v}_i))_{1 \le i \le s}$ yields a basis $((w_i, v_i))_{1 \le i \le s}$ to $(\hat{\nu}(k))_{k \in \hat{K}}$.

Theorem 4.1 can be easily transferred to the situation of bases to $(\hat{\nu}(k))_{k \in \hat{K}}$ with respect to $\langle \cdot, \cdot \rangle_S$:

Theorem 5.4 (Existence) *Let a sequence of subspaces $((\mathcal{W}_i, \mathcal{V}_i))_{1 \le i \le s}$ be given. The following statements are equivalent:*

a) There is a basis to $(\hat{\nu}(k))_{k \in \hat{K}}$ with respect to $\langle \cdot, \cdot \rangle_S$.

b) There is a generalized biorthogonal basis with respect to $\langle \cdot, \cdot \rangle_S$.

c) The $VMEGS$ method applied to an arbitrary basis of the chain $((\mathcal{W}_i, \mathcal{V}_i))_{1 \le i \le s}$ generates a basis to $(\hat{\nu}(k))_{k \in \hat{K}}$ with respect to $\langle \cdot, \cdot \rangle_S$.

d) The bilinear form $\langle \cdot, \cdot \rangle_S \mid_{\mathcal{W}_s \times \mathcal{V}_s}$ *is non-degenerate.*

e) $s = \max\{\nu(k) \mid k \in K\}$.

f) $s = \max\{\hat{\nu}(k) \mid k \in \hat{K}\}$.

g) There is a projection Π *with*

$$\operatorname{Im}\Pi = \mathcal{V}_s \qquad \textit{and} \qquad \operatorname{Ker}\Pi = \mathcal{W}_s^{\perp_S}.$$

Suppose there is a basis to $(\hat{\nu}(k))_{k\in\hat{K}}$. The sequence of vectors $((w_i, v_i))_{1\le i\le s}$ computed by the recursions

for all $0 \le k \le \hat{b} - 1$

$$v_{\hat{\nu}(k)+1} := \beta_1^{(k)}(I - \Pi_{\hat{\nu}(k)})Av_{\hat{\nu}(k)} \tag{5.12}$$

$$w_{\hat{\nu}(k)+1} := \alpha_1^{(k)}(I - \Pi^+_{\hat{\nu}(k)})A^+ w_{\hat{\nu}(k)} \tag{5.13}$$

for all $1 \le i \le dim_{k+1} - 1$

$$v_{\hat{\nu}(k)+i+1} = \beta_{i+1}^{(k)}(I - \Pi_{\hat{\nu}(k)}) \left(Av_{\hat{\nu}(k)} - \sum_{j=1}^{i} g_{i+1,j}^{(k)}\, v_{\hat{\nu}(k)+j} \right) \tag{5.14}$$

$$w_{\hat{\nu}(k)+i+1} = \alpha_{i+1}^{(k)}(I - \Pi^+_{\hat{\nu}(k)}) \left(A^+ w_{\hat{\nu}(k)} - \sum_{j=1}^{i} f_{i+1,j}^{(k)}\, w_{\hat{\nu}(k)+j} \right) \tag{5.15}$$

forms a basis to $(\hat{\nu}(k))_{k\in\hat{K}}$ with respect to $\langle \cdot, \cdot \rangle_S$. Here the coefficients $f_{i+1,l}^{(k)}, g_{i+1,l}^{(k)} \in \mathbb{C}$ and $\alpha_{i+1}^{(k)}, \beta_{i+1}^{(k)} \in \mathbb{C} \setminus \{0\}$ can be chosen arbitrarily. For numerical stability the numbers $\beta_{i+1}^{(k)}$ and $\alpha_{i+1}^{(k)}$ are usually chosen such that $\|v_j\| = 1$ and $\|w_j\| = 1$ for all j.
In a similar manner as described in Chapter 4 the recursions (5.12)-(5.15) reduce to the **(modified) practical look-ahead Lanczos recursions**:

For all $0 \le k \le \hat{b} - 1$

$$v_{\hat{\nu}(k)+1} := \beta_1^{(k)}(Av_{\hat{\nu}(k)} - V_k D_k^{-1} W_k^T S A v_{\hat{\nu}(k)} - V_{k-1} D_{k-1}^{-1} (W_{k-1})^T S A v_{\hat{\nu}(k)})$$

$$w_{\hat{\nu}(k)+1} := \alpha_1^{(k)}(A^+ w_{\hat{\nu}(k)} - W_k (D_k^T)^{-1} V_k^T S^T A^+ w_{\hat{\nu}(k)} - W_{k-1} (D_{k-1}^T)^{-1} V_{k-1}^T S^T A^+ w_{\hat{\nu}(k)})$$

for all $1 \le i \le dim_{k+1} - 1$

$$v_{\hat{\nu}(k)+i+1} = \beta_{i+1}^{(k)} \left(Av_{\hat{\nu}(k)+i} - \sum_{j=1}^{i} g_{i+1,j}^{(k)}\, v_{\hat{\nu}(k)+j} - V_k D_k^{-1} W_k^T S A v_{\hat{\nu}(k)+i} \right)$$

$$w_{\hat{\nu}(k)+i+1} = \alpha_{i+1}^{(k)} \left(A^+ w_{\hat{\nu}(k)+i} - \sum_{j=1}^{i} f_{i+1,j}^{(k)}\, w_{\hat{\nu}(k)+j} - W_k (D_k^T)^{-1} V_k^T S^T A^+ w_{\hat{\nu}(k)+i} \right)$$

where $D_k := W_k^T S V_k$.

The following Theorem shows that the choice of the form $\langle \cdot, \cdot \rangle_S$ is essentially equivalent to the choice of the starting vector w_1.

Theorem 5.5 *Let (A, V, Π, H, s) and $(A^+, W, \Pi^+, \hat{H}, s)$ be Krylov tuples generated by a modified practical look ahead Lanczos method (using an arbitrary non-degenerate bilinear form $\langle \cdot, \cdot \rangle_S$). Then the following holds:*

a) The degree indices of $(\mathcal{K}(i, A^+, We_1), \mathcal{K}(i, A, Ve_1))_{1 \le i \le N}$ with respect to the form $\langle \cdot, \cdot \rangle_S$ are identical with the degree indices of $(\mathcal{K}(i, A^T, S^T We_1), \mathcal{K}(i, A, Ve_1))_{1 \le i \le N}$ with respect to the form $\langle x, y \rangle = x^T y$.

b) The ordinary practical look-ahead Lanczos method (i.e. using $\langle \cdot, \cdot \rangle = x^T y$) applied to the starting vectors $\tilde{v}_1 = Ve_1$ and $\tilde{w}_1 = S^T We_1$ generates the Krylov tuples (A, V, Π, H, s) and $(A^T, S^T W, \Pi^T, \hat{H}, s)$ if the coefficients $\alpha_i^{(k)}, \beta_i^{(k)}, f_{i,j}^{(k)}, g_{i,j}^{(k)}$ and the subsequence of the degree indeces $(\hat{\nu}(k))_{k \in \hat{K}}$ (used in the process) are chosen as in the modified practical look-ahead Lanczos process.

Proof. The identity

$$\begin{aligned} \operatorname{Im} S^T W_i &= S^T \operatorname{Im} W_i = S^T \mathcal{K}(i, A^T, We_1) \\ &= \mathcal{K}(i, S^T A^T (S^T)^{-1}, We_1) \\ &= \mathcal{K}(i, A^+, \tilde{w}_1) \end{aligned}$$

combined with the equations

$$(\langle We_i, Ve_j \rangle_S)_{1 \le i,j \le l} = W^T S V = (\langle S^T We_i, Ve_j \rangle)_{1 \le i,j \le l}, \qquad l = 1, 2, \ldots, s$$

yields statement a). In the following $(\hat{\nu}(k))_{k \in \hat{K}}$ denotes the subsequence used in the practical look-ahead Lanczos processes.
Let

$$(A, V^{(\hat{\nu}(1))}, \Pi_{\hat{\nu}(1)}, H^{(\hat{\nu}(1))}, \hat{\nu}(1)),\ (A, V^{(\hat{\nu}(2))}, \Pi_{\hat{\nu}(2)}, H^{(\hat{\nu}(2))}, \hat{\nu}(2)),\ \ldots,\ (A, V^{(s)}, \Pi_s, H^{(s)}, s)$$

and

$$(A^+, W^{(\hat{\nu}(1))}, \Pi^+_{\hat{\nu}(1)}, \hat{H}^{(\hat{\nu}(1))}, \hat{\nu}(1)),\ \ldots,\ (A^+, W^{(s)}, \Pi^+_s, \hat{H}^{(s)}, s)$$

be the Krylov tuples generated by the modified look-ahead Lanczos method.
Let

$$(A, \tilde{V}^{(\hat{\nu}(1))}, \tilde{\Pi}_{\hat{\nu}(1)}, \tilde{H}^{(\hat{\nu}(1))}, \hat{\nu}(1)),\ (A, \tilde{V}^{(\hat{\nu}(2))}, \tilde{\Pi}_{\hat{\nu}(2)}, \tilde{H}^{(\hat{\nu}(2))}, \hat{\nu}(2)),\ \ldots,\ (A, \tilde{V}^{(s)}, \tilde{\Pi}_s, \tilde{H}^{(s)}, s)$$

and

$$(A^T, \tilde{W}^{(\hat{\nu}(1))}, \tilde{\Pi}^T_{\hat{\nu}(1)}, \check{H}^{(\hat{\nu}(1))}, \hat{\nu}(1)),\ \ldots,\ (A^T, \tilde{W}^{(s)}, \tilde{\Pi}^T_s, \check{H}^{(s)}, s)$$

be the Krylov tuples generated by the ordinary practical look-ahead Lanczos method applied to the starting vectors $\tilde{v}_1 = Ve_1$ and $\tilde{w}_1 = S^T We_1$. Now we show by induction on j that $\tilde{W}^{(j)} = S^T W^{(j)}$, $\tilde{V}^{(j)} = V^{(j)}$ and $\Pi_{\hat{\nu}(k)} = \tilde{\Pi}_{\hat{\nu}(k)}$.
By choice of the starting vectors the identities $\tilde{v}_1 = v_1$ and $\tilde{w}_1 = w_1$ hold.
Assume that

$$\tilde{V}^{(j)} = V^{(j)} \qquad \text{and} \qquad \tilde{W}^{(j)} = S^T W^{(j)}$$

for some j. Choose $k \in \hat{K}$ such that $\hat{\nu}(k) \leq j \leq \hat{\nu}(k+1)$. Choose $i \in \mathbb{N}$ such that $j = \hat{\nu}(k) + i$. Then

$$\begin{aligned} \Pi_{\hat{\nu}(k)} &= V^{(\hat{\nu}(k))}((W^{(\hat{\nu}(k))})^T S V^{(\hat{\nu}(k))})^{-1}(W^{(\hat{\nu}(k))})^T S \\ &= V^{(\hat{\nu}(k))}((\tilde{W}^{(\hat{\nu}(k))})^T V^{(\hat{\nu}(k))})^{-1}(\tilde{W}^{(\hat{\nu}(k))})^T = \tilde{\Pi}_{\hat{\nu}(k)}. \end{aligned}$$

Hence the desired equations

$$\begin{aligned} Ve_{j+1} &= \beta_{i+1}^{(k)}(I - \Pi_{\hat{\nu}(k)})\left(AVe_j - \sum_{l=1}^{i} g_{i+1,l}^{(k)} Ve_{\hat{\nu}(k)+l}\right) \\ &= \beta_{i+1}^{(k)}(I - \tilde{\Pi}_{\hat{\nu}(k)})\left(A\tilde{V}e_j - \sum_{l=1}^{i} g_{i+1,l}^{(k)} \tilde{V}e_{\hat{\nu}(k)+l}\right) \\ &= \tilde{V}e_{j+1} \end{aligned}$$

and

$$\begin{aligned} We_{j+1} &= \alpha_{i+1}^{(k)}(I - \Pi_{\hat{\nu}(k)}^{+})\left(A^+ We_j - \sum_{l=1}^{i} f_{i+1,l}^{(k)} We_{\hat{\nu}(k)+l}\right) \\ &= \alpha_{i+1}^{(k)}(I - (SV^{(\hat{\nu}(k))}((W^{(\hat{\nu}(k))})^T S V^{(\hat{\nu}(k))})^{-1}(W^{(\hat{\nu}(k))})^T SS^{-1})^T)\Big((SAS^{-1})^T We_j \\ &\quad - \sum_{l=1}^{i} f_{i+1,l}^{(k)} We_{\hat{\nu}(k)+l}\Big) \\ &= (S^T)^{-1}\alpha_{i+1}^{(k)}(I - \tilde{\Pi}_{\hat{\nu}(k)}^T)\left(A^T S^T We_j - \sum_{l=1}^{i} f_{i+1,l}^{(k)} S^T We_{\hat{\nu}(k)+l}\right) \\ &= (S^T)^{-1}\tilde{W}e_{j+1} \end{aligned}$$

hold.
Finally the identity

$$\tilde{V}^{(\hat{\nu}(k))}\tilde{H}^{(\hat{\nu}(k))} = \tilde{\Pi}_{(\hat{\nu}(k))}A\tilde{V}^{(\hat{\nu}(k))} = \Pi_{(\hat{\nu}(k))}AV^{(\hat{\nu}(k))} = V^{(\hat{\nu}(k))}H^{(\hat{\nu}(k))}$$

yields $\tilde{H}^{(\hat{\nu}(k))} = H^{(\hat{\nu}(k))}$ which proves the theorem. ∎

Remark 5.4 (Breakdowns of the modified Lanczos and look-ahead Lanczos methods) By Theorem 4.1 the classification of breakdowns of the the (look-ahead) Lanczos methods (see Section 2.2.1 and Remark 3.2) transfers to the modified (look-ahead) Lanczos methods simply by replacing $\langle \cdot, \cdot \rangle$ with $\langle \cdot, \cdot \rangle_S$, '$\perp$' with '$\perp_S$' and the transpose sign 'T' with the adjoint sign '+'.

For later use we note the following Lemma.

Lemma 5.5 *Let (A, V, Π, H, s) and $(A^+, W, \Pi^+, \hat{H}, s)$ be generated by a (modified) practical look-ahead Lanczos method. Let $(\hat{\nu}(k))_{0 \le k \le \hat{b}}$ denote the subsequence of the degree indices used in the practical look-ahead Lanczos process. Then*

$$H = D^{-1} \hat{H}^T D$$

where $D := W^T S V = diag\,(D_1, D_2, \ldots, D_{\hat{b}})$ and $D_k \in \mathrm{Mat}_{\hat{\nu}(k) - \hat{\nu}(k-1)}$.

Proof. Theorem 5.5 together with identity (4.5) proves the lemma. ∎

Considering the recursions (5.12)-(5.15) we find that the vectors v_j and w_j, generated by a modified practical look-ahead Lanczos method, are identical (i.e. $v_j = w_j$ for every j) if the matrices A and $\Pi_{\hat{\nu}(k)}$ are self-adjoint

$$A = A^+ = (SAS^{-1})^T \qquad \text{and} \qquad \Pi_{\hat{\nu}(k)} = \Pi^+_{\hat{\nu}(k)} = (S \Pi_{\hat{\nu}(k)} S^{-1})^T$$

with respect to a non-degenerate bilinear form $\langle x, y \rangle_S = x^T S y$ and the coefficients and starting vectors are chosen as follows:

$$g^{(k)}_{i+1,j} = f^{(k)}_{i+1,j}, \qquad \beta^{(k)}_i = \alpha^{(k)}_i \qquad \text{and} \qquad v_1 = w_1. \tag{5.16}$$

The following Lemma 5.6 is due to Taussky and Zassenhaus [72]. It shows that for every matrix $A \in \mathrm{Mat}_N$ there is an invertible matrix $S \in \mathrm{Mat}_N$ such that $S = S^T$ and A is self-adjoint with respect to the form $\langle x, y \rangle_S = x^T S y$. Of course, S is not positive definite, in general.

Lemma 5.6 *Let $A \in \mathrm{Mat}_N$. There is an invertible matrix $S \in \mathrm{Mat}_N$ such that $S^T = S$ and $A = (SAS^{-1})^T$.*

Proof. Let $J = \mathrm{diag}\,(J_1, J_2, \ldots, J_k)$, $J_i \in \mathrm{Mat}_{l_i}$, be a Jordan normal form of A. Then there is an invertible matrix F such that $J = FAF^{-1}$. Define

$$X := \mathrm{diag}\,(X_1, X_2, \ldots, X_k) \qquad \text{where} \qquad X_i = \begin{bmatrix} 0 & & 1 \\ & \iddots & \\ 1 & & 0 \end{bmatrix} \in \mathrm{Mat}_{l_i} \quad \text{for } l_i \ge 2$$

and $X_i = 1$ for $l_i = 1$. Using the relations $J^T = XJX$ and $X^2 = I$ we obtain

$$A^T = (FJF^{-1})^T = F^T J^T (F^T)^{-1} = F^T XJX(F^T)^{-1} = (F^T XV)A(F^{-1}X(F^T)^{-1}).$$

Define $S := F^T XF$. The equation $S^T = (F^T XF)^T = F^T XF = S$ proves the lemma. ∎

Suppose that $S \in \mathrm{Mat}_N$ is chosen such that $A = (SAS^{-1})^T$ and $S = S^T$. Let $(\nu(k))_{k\in K}$ denote the degree indices of $((\mathcal{K}(i, A^+, v_1), \mathcal{K}(i, A, v_1)))_{1\le i\le N}$ with respect to $\langle x, y\rangle_S = x^T Sy$. Then the Gram-Schmidt projection $\Pi_{\nu(k)}$ is self-adjoint with respect to $\langle\cdot,\cdot\rangle_S$. This follows from the equations (5.6)-(5.9) by replacing the transpose sign 'T' with the adjoint sign '$+$' and replacing '$\perp$' with '$\perp_S$'. Hence the vectors v_j and w_j are identical (i.e. $v_j = w_j$ for every j) if the coefficients in the modified look-ahead Lanczos process are chosen as described in (5.16).

Let $\langle x, y\rangle = x^T Sy$ be the form used in the modified look-ahead Lanczos process. Then the look-ahead Lanczos vectors v_j and w_j can be chosen identical if there is a scalar $\gamma \in \mathbb{C}$ such that the equation

$$\gamma A = (SAS^{-1})^T \tag{5.17}$$

holds. This is the content of Lemma 5.7 below.

Lemma 5.7 *Choose $\gamma \in \mathbb{C}$ and $S \in \mathrm{Mat}_N$ such that the equation (5.17) holds. Choose a starting vector $v_1 \in \mathbb{C}^N \setminus \{0\}$. Set $w_1 := v_1$. Choose the coefficients $f^{(k)}_{i+1,l}$, $g^{(k)}_{i+1,l}$, $\beta^{(k)}_{i+1}$ and $\alpha^{(k)}_{i+1}$ in the practical modified look-ahead Lanczos process such that the equations*

$$\beta^{(k)}_{i+1} = \gamma \cdot \alpha^{(k)}_{i+1} \qquad \text{and} \qquad f^{(k)}_{i+1,l} = \gamma \cdot g^{(k)}_{i+1,l}$$

are valid. Then the left and right vectors generated by the practical modified look-ahead Lanczos method (using $\langle\cdot,\cdot\rangle_S$) are identical (i.e. $v_j = w_j$ for all j).

Proof.

Set $V^{(j)} := [v_1, v_2, \ldots, v_j]$ and $W^{(j)} := [w_1, w_2, \ldots, w_j]$ where v_j and w_j are the vectors generated by the practical modified look-ahead Lanczos method.
We show by induction that $V^{(j)} = W^{(j)}$ for all j.
By assumption the equation $V^{(1)} = W^{(1)}$ holds.
Assume the statement $V^{(j)} = W^{(j)}$ is true for some j.
Now proceed to $j+1$: There are numbers $k \in K$ and $i \in \mathbb{N}\setminus\{0\}$ such that $j = \nu(k)+i+1$ and $j \le \nu(k+1)$.
The following equations prove the claim:

$$\begin{aligned} w_{\hat\nu(k)+i+1} &= \alpha^{(k)}_{i+1}(I - \Pi^+_{\hat\nu(k)})\left(A^+ w_{\hat\nu(k)+i} + \sum_{l=1}^{i} f^{(k)}_{i+1,l}\, w_{\hat\nu(k)+l}\right) \\ &= \alpha^{(k)}_{i+1}(I - (S\Pi_{\hat\nu(k)}S^{-1})^T)\left(\gamma A v_{\hat\nu(k)+i} + \sum_{l=1}^{i} f^{(k)}_{i+1,l}\, v_{\hat\nu(k)+l}\right) \end{aligned}$$

$$
\begin{aligned}
&= \alpha_{i+1}^{(k)}(I - (SV^{\hat{\nu}(k)}((W^{\hat{\nu}(k)})^T SV^{\hat{\nu}(k)})^{-1}(W^{\hat{\nu}(k)})^T SS^{-1})^T) \\
&\quad \left(\gamma A v_{\hat{\nu}(k)+i} + \sum_{l=1}^{i} \gamma g_{i+1,l}^{(k)}\, v_{\hat{\nu}(k)+l}\right) \\
&= \gamma\alpha_{i+1}^{(k)}(I - (V^{\hat{\nu}(k)}((V^{\hat{\nu}(k)})^T SV^{\hat{\nu}(k)})^{-1}(V^{\hat{\nu}(k)})^T S)\left(Av_{\hat{\nu}(k)+i} + \sum_{l=1}^{i} g_{i+1,l}^{(k)}\, v_{\hat{\nu}(k)+l}\right) \\
&= \beta_{i+1}^{(k)}(I - \Pi_{\hat{\nu}(k)})(Av_{\hat{\nu}(k)+i} + \sum_{l=1}^{i} g_{i+1,l}^{(k)}\, v_{\hat{\nu}(k)+l}) \\
&= v_{\hat{\nu}(k)+i+1}.
\end{aligned}
$$

■

Example 5.4 (Choice of γ and S)

a) If $A = A^T$ then choose $\gamma = 1$ and $\langle x, y\rangle_S = x^T y$.

b) If $A = -A^T$ then choose $\gamma = -1$ and $\langle x, y\rangle_S = x^T y$.

c) If $A = iA^T$ then choose $\gamma = i$ and $\langle x, y\rangle_S = x^T y$.

d) Let $C, D \in \mathrm{Mat}_N$ such that $C^T = C$.

$$\text{If} \quad A = \begin{bmatrix} C & D \\ -D^T & C \end{bmatrix} \text{ then choose } \gamma = 1, \quad S = \begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix}, \quad \langle x, y\rangle_S = x^T S y.$$

e) If A is symmetric with respect to the antidiagonal then choose $\gamma = 1$ and

$$S = \begin{bmatrix} 0 & & 1 \\ & \iddots & \\ 1 & & 0 \end{bmatrix}.$$

Using Lemma 5.7 we obtain the following practical modified look-ahead Lanczos algorithm:

Algorithm 5.2

(1) Compute $S \in \mathrm{Mat}_N$ such that there is a scalar $\gamma \in \mathbb{C}$ satisfying the equation $\gamma A = (SAS^{-1})^T$.

(2) Choose a starting vector $v_1 \in \mathbb{C}^N$ with $\|v_1\| = 1$.

(3) For all $0 \leq k \leq \hat{b} - 1$

$$\begin{aligned} v_{\hat{\nu}(k)+1} &:= Av_{\hat{\nu}(k)} - V_k D_k^{-1} V_k^T S A v_{\hat{\nu}(k)} - V_{k-1} D_{k-1}^{-1} (V_{k-1})^T S A v_{\hat{\nu}(k)} \\ v_{\hat{\nu}(k)+1} &:= v_{\hat{\nu}(k)+1} / ||v_{\hat{\nu}(k)+1}|| \end{aligned}$$

for all $1 \leq i \leq dim_{k+1} - 1$

$$\begin{aligned} v_{\hat{\nu}(k)+i+1} &= Av_{\hat{\nu}(k)+i} - \sum_{j=1}^{i} g_{i+1,j}^{(k)} \, v_{\hat{\nu}(k)+j} - V_k D_k^{-1} V_k^T S A v_{\hat{\nu}(k)+i} \\ v_{\hat{\nu}(k)+i+1} &:= v_{\hat{\nu}(k)+1} / ||v_{\hat{\nu}(k)+i+1}||. \end{aligned}$$

Here $D_k := V_k^T S V_k$. The coefficients $g_{i+1,j}^{(k)} \in \mathbb{C}$ can be chosen arbitrarily.

Using Theorem 5.5 b) Algorithm 5.2 above can be rewritten in a form which requires less computational effort:

Algorithm 5.3

(1) Compute $S \in \mathrm{Mat}_N$ such that there is a scalar $\gamma \in \mathbb{C}$ satisfying the equation $\gamma A = (SAS^{-1})^T$.

(2) Choose a starting vector $v_1 \in \mathbb{C}^N$ with $||v_1|| = 1$. Compute $w_1 := S^T v_1$.

(3) For all $0 \leq k \leq \hat{b} - 1$

$$\begin{aligned} v_{\hat{\nu}(k)+1} &:= Av_{\hat{\nu}(k)} - V_k D_k^{-1} W_k^T A v_{\hat{\nu}(k)} - V_{k-1} D_{k-1}^{-1} (W_{k-1})^T A v_{\hat{\nu}(k)} \\ v_{\hat{\nu}(k)+1} &:= v_{\hat{\nu}(k)+1} / ||v_{\hat{\nu}(k)+1}|| \\ w_{\hat{\nu}(k)+1} &:= S^T v_{\hat{\nu}(k)+1} \\ w_{\hat{\nu}(k)+1} &:= w_{\hat{\nu}(k)+1} / ||w_{\hat{\nu}(k)+1}|| \end{aligned}$$

for all $1 \leq i \leq dim_{k+1} - 1$

$$\begin{aligned} v_{\hat{\nu}(k)+i+1} &= Av_{\hat{\nu}(k)+i} - \sum_{j=1}^{i} g_{i+1,j}^{(k)} \, v_{\hat{\nu}(k)+j} - V_k D_k^{-1} W_k^T A v_{\hat{\nu}(k)+i} \\ v_{\hat{\nu}(k)+i+1} &:= v_{\hat{\nu}(k)+i+1} / ||v_{\hat{\nu}(k)+i+1}|| \\ w_{\hat{\nu}(k)+i+1} &:= S^T v_{\hat{\nu}(k)+i+1} \\ w_{\hat{\nu}(k)+i+1} &:= w_{\hat{\nu}(k)+i+1} / ||w_{\hat{\nu}(k)+i+1}||. \end{aligned}$$

Here $D_k := W_k^T V_k$. The coefficients $g_{i+1,j}^{(k)} \in \mathbb{C}$ can be chosen arbitrarily.

The special case of Algorithm 5.3 where $\gamma = 1$ has been derived by Freund and Zha [21] in 1991. They called this algorithm transpose-free look-ahead Lanczos algorithm.

Remark 5.5 Algorithms 5.2 and 5.3 are only useful for those special classes of matrices A, for which one can find a matrix S and a scalar γ satisfying the equation $\gamma A = (SAS^{-1})^T$ with small effort, and for which the matrix-vector products with S^T can be computed cheaply.

Let H be the matrix generated by a look-ahead Lanczos method. The following lemma characterizes coordinate transformations leaving the matrix H unchanged.

Lemma 5.8 *Let $A \in \mathrm{Mat}_N$ and two starting vectors $v_1, w_1 \in \mathbb{C}^N$ be given. Choose an invertible matrix $C \in \mathrm{Mat}_N$. Define*

$$\tilde{v}_1 = Cv_1, \qquad \tilde{w}_1 = (C^T)^{-1}w_1 \qquad \text{and} \qquad \tilde{A} = CAC^{-1}.$$

Let (A, V, Π, H, s) and $(A^T, W, \Pi^T, \hat{H}, s)$ be generated by a look-ahead Lanczos method applied to the triple (w_1, A, v_1).
Let $(CAC^{-1}, \tilde{V}, \tilde{\Pi}, \tilde{H}, s)$ and $((CAC^{-1})^T, \tilde{W}, \tilde{\Pi}^T, \check{H}, s)$ be generated by the same look-ahead Lanczos method applied to the triple $(\tilde{w}_1, \tilde{A}, \tilde{v}_1)$.
Then the following identities hold:

$$\tilde{H} = H, \qquad \check{H} = \hat{H}, \qquad \tilde{\Pi} = \Pi$$

$$\tilde{V} = CV \qquad \text{and} \qquad \tilde{W} = (C^T)^{-1}W.$$

Lemma 5.8 can be easily proved by induction.

5.3 Condition numbers of bases

Let (A, V, Π, H, s) be generated by a Krylov subspace projection method. It is of practical interest to have estimates for the condition number $\kappa_2(V)$. As discussed in Section 5.1, a Krylov subspace projection method generates, step by step, a sequence of continuations of Krylov tuples. Hence it is natural to discuss how the condition numbers of the bases change from one step to the next.

Estimates for the condition numbers of continuations of Krylov tuples are presented in Theorem 5.6 below for general Krylov subspace projection methods and in Corollary 5.1 for Krylov subspace projection methods which generate continuations $(A, [V, V_2], \tilde{\Pi}, \tilde{H}, r)$ with the property

$$\Pi V_2 = 0. \tag{5.18}$$

Recall that (practical) look-ahead Lanczos methods compute continuations with the property (5.18) (see Section 5.1). As discussed in Section 5.1, such special Krylov subspace projection methods are very important in practice because they require much less computational effort than the methods which do not have the property (5.18).

Theorem 5.6 *Let $(A, [V, V_2], \tilde{\Pi}, \tilde{H}, r)$ be a continuation of a normalized Krylov tuple (A, V, Π, H, s). Let $\hat{\Pi}$ be an arbitrary projection with $\operatorname{Im}\hat{\Pi} \supseteq \operatorname{Im} V_2$ and $\operatorname{Ker}\hat{\Pi} \supseteq \operatorname{Im} V$. Then the following relations hold:*

$$\frac{1}{\sigma_{\min}([V, V_2])} \leq \frac{\|\Pi(I - \hat{\Pi})\|}{\sigma_{\min}(V)} + \frac{\|\hat{\Pi}\|}{\sigma_{\min}(V_2)} \qquad \textit{and} \qquad \sigma_{\max}([V, V_2]) \leq \sqrt{r}.$$

To prove Theorem 5.6 we need some preparations.

Lemma 5.9 *Let $C, D \in \operatorname{Mat}_{N\times i}$. If $C^*D = I$ then $1 \leq \|C^*\| \min_{\|x\|=1} \|Dx\|$.*

Lemma 5.9 is evident.

Lemma 5.10 *Let $(\mathcal{W}_i)_{1\leq i\leq r}$ be a chain of subspaces with $\dim \mathcal{W}_i = i$ for all i. Let $(v_i)_{1\leq i\leq s}$ be a basis of $(\mathcal{W}_i)_{1\leq i\leq s}$ with $s < r$. Then there is a basis $(\tilde{q}_i)_{1\leq i\leq r}$ of $(\mathcal{W}_i)_{1\leq i\leq r}$ with the properties:*

a) $\tilde{q}_i = v_i$ for $1 \leq i \leq s$.

b) $(\tilde{\sigma}_1, \tilde{\sigma}_2, \ldots, \tilde{\sigma}_r) = (\sigma_1, \sigma_2, \ldots, \sigma_s, 1, \ldots, 1)$ where $\tilde{\sigma}_i$ are the singular values of the matrix $[\tilde{q}_1, \tilde{q}_2, \ldots, \tilde{q}_r]$ and σ_i are the singular values of the matrix $[q_1, q_2, \ldots, q_s]$ listed in an appropriate order.

Proof. Choose an arbitrary basis $(w_i)_{1\leq i\leq r}$ of $(\mathcal{W}_i)_{1\leq i\leq r}$. Let N_i be the orthogonal projection on W_i. Define $\tilde{v}_i := (I - N_i)w_i/\|(I - N_i)w_i\|$ for $i = s+1, s+2, \ldots, r$. Putting $V := [v_1, \ldots, v_s]$, $V_2 := [\tilde{v}_{s+1}, \ldots, \tilde{v}_r]$ and $\tilde{V} := [V, V_2]$, we obtain the representation

$$\tilde{V}^*\tilde{V} \quad = \quad \begin{bmatrix} V^*V & V^*V_2 \\ V_2{}^*V & V_2{}^*V_2 \end{bmatrix} \quad = \quad \begin{bmatrix} V^*V & 0 \\ 0 & I \end{bmatrix}$$

which proves Lemma 5.10. ∎

After these preliminary steps we are now set to prove Theorem 5.6.

Proof of Theorem 5.6.

Put $k := \dim \operatorname{Ker}\hat{\Pi}$. Choose matrices $W \in \operatorname{Mat}_{N\times s}$ and $\check{W} \in \operatorname{Mat}_{N\times k}$ such that $\operatorname{Ker} W^* = \operatorname{Ker}\Pi$ and $\operatorname{Ker}\check{W}^* = \operatorname{Ker}\hat{\Pi}$. The following relations a)-c) imply that the matrix $\tilde{W} := [W, \check{W}]$ has full rank.

a) $(\operatorname{Im}\check{W})^\perp = \operatorname{Ker}\check{W}^* = \operatorname{Ker}\hat{\Pi} \supseteq \operatorname{Im} V = \operatorname{Im}\Pi \quad (\Rightarrow \quad \operatorname{Im}\check{W} \subseteq (\operatorname{Im}\Pi)^\perp)$.

b) $(\operatorname{Im} W)^\perp = \operatorname{Ker} W^* = \operatorname{Ker}\Pi \quad (\Rightarrow \quad \operatorname{Im} W = (\operatorname{Ker}\Pi)^\perp)$.

c) $(\operatorname{Ker}\Pi)^\perp \cap (\operatorname{Im}\Pi)^\perp = (\operatorname{Ker}\Pi + \operatorname{Im}\Pi)^\perp = \{0\}$.

By Lemma 5.10 there are matrices $\tilde{V} \in \mathrm{Mat}_{N\times(k+s)}$, $\check{V} \in \mathrm{Mat}_{N\times k}$ such that $\tilde{V} = [V, \check{V}] = [V, V_2, \star]$, $\sigma_{\min}(\tilde{V}) = \sigma_{\min}([V, V_2])$, $\sigma_{\max}(\tilde{V}) = \sigma_{\max}([V, V_2])$ and $\operatorname{Im}\check{V} = \operatorname{Im}\hat{\Pi}$.
We obtain the representation

$$\Pi = VD^{-1}W^*, \qquad \hat{\Pi} = \check{V}\check{D}^{-1}\check{W}^* \quad \text{and} \quad \tilde{D}^{-1}\tilde{W}^*\tilde{V}^* = I$$

where $D := W^*V$, $\check{D} := \check{W}^*\check{V}$ and $\tilde{D} := \tilde{W}^*\tilde{V}$. Lemma 5.9 yields:

$$1 \le ||\tilde{D}^{-1}\tilde{W}^*|| \min_{||x||=1} ||\tilde{V}x|| = ||\tilde{D}^{-1}\tilde{W}^*||\sigma_{\min}(\tilde{V}) = ||\tilde{D}^{-1}\tilde{W}^*||\,\sigma_{\min}([V, V_2]). \tag{5.19}$$

Using the relation $\check{W}^*V = 0$ we have

$$\tilde{D} = \begin{bmatrix} D & W^*\check{V} \\ 0 & \check{D} \end{bmatrix}, \qquad \tilde{D}^{-1} = \begin{bmatrix} D^{-1} & -D^{-1}W^*\check{V}\check{D}^{-1} \\ 0 & \check{D}^{-1} \end{bmatrix}.$$

One easily verifies that

$$||\tilde{D}^{-1}\tilde{W}^*|| \le ||\check{D}^{-1}\check{W}^*|| + ||D^{-1}W^*(I - \hat{\Pi})||. \tag{5.20}$$

The Gram-Schmidt orthogonalisation method applied to the columns of V and $\check{V}$ generates the decompositions $V = QR$ and $\check{V} = \check{Q}\check{R}$ where Q and $\check{Q}$ are matrices with orthonormal columns and R and $\check{R}$ are invertible upper triangular matrices. Using the fact that $\operatorname{Im}Q = \operatorname{Im}V$, $\operatorname{Im}\check{Q} = \operatorname{Im}\check{V}$, $\sigma_{\min}(V) = \sigma_{\min}(R)$, $\sigma_{\min}(V_2) = \sigma_{\min}(\check{V}) = \sigma_{\min}(\check{R})$, $||QA|| = ||A||$ and $||\check{Q}\check{A}|| = ||\check{A}||$ for arbitrary matrices A and $\check{A}$, we conclude

$$\begin{aligned} ||\check{D}^{-1}\check{W}^*|| &= ||\check{R}^{-1}(\check{W}^*\check{Q})^{-1}\check{W}^*|| \le ||\check{R}^{-1}||\,||(\check{W}^*\check{Q})^{-1}\check{W}^*|| \\ &= \frac{||(\check{W}^*\check{Q})^{-1}\check{W}^*||}{\sigma_{\min}(\check{V})} = \frac{||\check{Q}(\check{W}^*\check{Q})^{-1}\check{W}^*||}{\sigma_{\min}(V_2)} = \frac{||\hat{\Pi}||}{\sigma_{\min}(V_2)} \end{aligned} \tag{5.21}$$

and, in a similar manner

$$||D^{-1}W^*(I - \check{\Pi})|| \le \frac{||\Pi(I - \hat{\Pi})||}{\sigma_{\min}(V)}. \tag{5.22}$$

Combination of equation (5.19), (5.20), (5.21) and (5.22) yields the assertion. ∎

For Krylov subspace methods which generate continuations $(A, [V, V_2], \tilde{\Pi}, \tilde{H}, r)$ with the property (5.18) we have the following estimate.

Corollary 5.1 *Let $(A, [V, V_2], \tilde{\Pi}, \tilde{H}, r)$ be a continuation of the normalized Krylov tuple (A, V, Π, H, s) with $\Pi V_2 = 0$. Then*

$$\frac{1}{\sigma_{\min}([V, V_2])} \leq \frac{||\Pi||}{\sigma_{\min}(V)} + \frac{||I - \Pi||}{\sigma_{\min}(V_2)} \qquad \textit{and} \qquad \sigma_{\max}([V, V_2]) \leq \sqrt{r}.$$

Proof. Define $\hat{\Pi} := (I - \Pi)$. Then $\operatorname{Ker} \hat{\Pi} = \operatorname{Im} \Pi = \operatorname{Im} V$ and $\operatorname{Im} \hat{\Pi} \supseteq \operatorname{Ker} \Pi \supseteq \operatorname{Im} V_2$. Thus, Theorem 5.6 proves the corollary. ∎

Remark 5.6 Let (A, V, Π, H, s) be a normalized Krylov tuple with $s < N$. Choose $r \in \{s+1, \ldots, N\}$. A continuation $(A, [V, V_2], \tilde{\Pi}, \tilde{H}, r)$ with the properties

a) $\Pi V_2 = 0$ and $V_2^* V_2 = I$

b) $$\frac{1}{\sigma_{\min}([V, V_2])} \leq \frac{||\Pi||}{\sigma_{\min}(V)} + ||I - \Pi|| \qquad \text{and} \qquad \sigma_{\max}([V, V_2]) \leq \sqrt{r}.$$

can be constructed as follows:

(1) $v_{s+1} := (I - \Pi)AVe_s / ||(I - \Pi)AVe_s||$.

(2) Apply the Gram-Schmidt orthonormalisation method to a basis of the chain $((I - \Pi)\mathcal{K}(j, A, v_{s+1}))_{1 \leq j \leq r-s}$ to obtain an orthonormal basis $(v_j)_{s+1 \leq j \leq r}$. Put $V_2 := [v_{s+1}, v_{s+2}, \ldots, v_r]$.

(3) Choose an arbitrary full rank matrix $\tilde{W} \in \mathrm{Mat}_{N \times r}$ such that $\tilde{D} := \tilde{W}^*[V, V_2]$ is invertible. Define $\tilde{\Pi} := [V, V_2]\tilde{D}^{-1}\tilde{W}^*$.

(4) Continue as described in Algorithm 5.1 (6) and (7).

We consider again a Krylov tuple (A, V, Π, H, s). Using a representation

$$\Pi = V(W^T S V)^{-1} W^T S \tag{5.23}$$

of the projection Π, the condition number of V can be estimated in terms of $\kappa_2(W^T S V)$. This is the content of Theorem 5.7 below. Estimates in terms of $\kappa_2(W^* S V)$ have been derived by Parlett [58] for the look-ahead Lanczos process using the form $\langle \cdot, \cdot \rangle_*$.

Theorem 5.7 *Choose an invertible matrix $S \in \mathrm{Mat}_N$. Let (A, V, Π, H, s) be a normalized Krylov tuple. Let $W \in \mathrm{Mat}_{N \times s}$ be a full rank matrix with normalized columns satisfying the equation* (5.23)*. Then the following estimates hold:*

$$\sigma_{min}(W^T S V) \leq \sqrt{s}\, ||S||\, \sigma_{min}(V) \qquad \textit{and} \qquad \sigma_{min}(W^T S V) \leq \sqrt{s}\, ||S||\, \sigma_{min}(W). \tag{5.24}$$

Theorem 5.7 can be proved in an similar manner as [58] Theorem 10.1.

Remark 5.7 Let a matrix $A \in \mathrm{Mat}_N$ and two starting vectors be given. Choose a tolerance $tol \in (0,1]$. Then there is a practical look-ahead Lanczos method which generates normalized Krylov tuples

$$(A, V^{(\hat{\nu}(1))}, \Pi_{\hat{\nu}(1)}, H^{(\hat{\nu}(1))}, \hat{\nu}(1)),\ (A, V^{(\hat{\nu}(2))}, \Pi_{\hat{\nu}(2)}, H^{(\hat{\nu}(2))}, \hat{\nu}(2)),\ \ldots,\ (A, V^{(s)}, \Pi_s, H^{(s)}, s)$$

and

$$(A^+, W^{(\hat{\nu}(1))}, \Pi^+_{\hat{\nu}(1)}, \hat{H}^{(\hat{\nu}(1))}, \hat{\nu}(1)),\ \ldots,\ (A^+, W^{(s)}, \Pi^+_s, \hat{H}^{(s)}, s)$$

such that the matrix $D^{(s)} := (W^{(s)})^T S V^{(s)}$ is block diagonal

$$D^{(s)} = \mathrm{diag}\,(D_1, D_2, \ldots, D_{\hat{b}})$$

and the inequality

$$\sigma_{min}(D_k) \geq tol \tag{5.25}$$

holds for all k. Using Theorem 5.7 and the inequalities

$$\sigma_{min}(D^{(\hat{\nu}(k))}) = \min_{1 \leq j \leq k} \sigma_{min}(D_j) \geq tol \tag{5.26}$$

and

$$\begin{aligned} ||\Pi_{\hat{\nu}(k)}|| &= ||V^{\hat{\nu}(k)} (D^{\hat{\nu}(k)})^{-1} (W^{\hat{\nu}(k)})^T S|| \leq ||V^{\hat{\nu}(k)}||\, ||(D^{\hat{\nu}(k)})^{-1}||\, ||(W^{\hat{\nu}(k)})^T||\, ||S|| \\ &\leq s \frac{1}{\sigma_{min}(D^{\hat{\nu}(k)})}\, ||S|| \end{aligned}$$

we find

$$\sigma_{min}(V^{\hat{\nu}(k)}) \geq \frac{tol}{||S||\sqrt{s}}, \qquad \sigma_{min}(W^{\hat{\nu}(k)}) \geq \frac{tol}{||S||\sqrt{s}} \tag{5.27}$$

$$\sigma_{max}(V^{\hat{\nu}(k)}) \leq \sqrt{s}, \qquad \sigma_{max}(W^{\hat{\nu}(k)}) \leq \sqrt{s} \qquad \text{and} \qquad ||\Pi_{\hat{\nu}(k)}|| \leq \frac{s\,||S||}{tol} \tag{5.28}$$

for all $k \in \hat{K}$.

5.4 Eigenvalues of the matrix H

In this section we discuss the connection between the eigenvalues of H and the eigenvalues of A. Let a matrix A and a starting vector v_1 be given. The following theorem shows that for any choice of numbers $\lambda_1, \lambda_2, \ldots, \lambda_s \in \mathbb{C}$ there is a Krylov subspace projection method which generates (applied to A and v_1) a matrix H such that $\lambda_1, \lambda_2, \ldots, \lambda_s$ are the eigenvalues of H.

Theorem 5.8 *Let $(v_i)_{1\le i\le s}$ be an arbitrary basis of a Krylov chain $(\mathcal{K}(i,A,v_1))_{1\le i\le s}$ with $s < m$. Choose arbitrary numbers $\lambda_1,\lambda_2,\ldots,\lambda_s \in \mathbb{C}$. Then there is a Krylov tuple (A,V,Π,H,s) such that $\lambda_1,\lambda_2,\ldots,\lambda_s$ are the eigenvalues of H and $V=[v_1,v_2,\ldots,v_s]$.*

The following Lemmas are used in the proof of Theorem 5.8

Lemma 5.11 *Let $\mathcal{V}_2 \neq \mathbb{C}^N$ be a subspace of $\mathbb{C}^N$. Choose arbitrary vectors $q \in \mathbb{C}^N \setminus \mathcal{V}_2$ and $u \in \mathcal{V}_2$. Then there is a projection Π on $\mathcal{V}_2$ such that $\Pi q = u$.*

Proof. By choice of q and u the relation $w := (q-u) \notin \mathcal{V}_2$ holds. Hence there is a subspace $\mathcal{W} \subseteq \mathbb{C}^N$ such that $w \in \mathcal{W}$ and $\mathcal{V}_2 \oplus \mathcal{W} = \mathbb{C}^N$. Define a projection Π by $\operatorname{Im}\Pi = \mathcal{V}_2$ and $\operatorname{Ker}\Pi = \mathcal{W}$. Hence

$$0 = \Pi w = \Pi(q-u) = \Pi q - \Pi u = \Pi q - u$$

which proves the assertion. ∎

Lemma 5.12 *Let (A,V,Π,H,s) be a Krylov tuple. Let ϕ be the characteristic polynomial of $H=[h_{i,j}]$. Then the identity*

$$\begin{aligned} \phi(A)Ve_1 &= (I-\Pi)A^sVe_1 && (5.1)\\ &= h_{2,1}h_{3,2}\cdots h_{s,s-1}(I-\Pi)AVe_s && (5.2)\end{aligned}$$

holds.

Proof. Using that

$$v_i = \frac{1}{h_{i,i-1}}\left(AV_{i-1} - \sum_{l=1}^{i-1} h_{l,i-1}v_l\right)$$

we obtain inductively the representation

$$v_s = \frac{1}{h_{2,1}h_{3,2}\cdots h_{s,s-1}}\left(A^{s-1}Ve_1 - \sum_{l=1}^{s-1}\rho_l v_l\right)$$

where $\rho_l \in \mathbb{C}$. Consequently

$$\begin{aligned}(I-\Pi)AVe_s &= \frac{1}{h_{2,1}h_{3,2}\cdots h_{s,s-1}}\left((I-\Pi)A^sVe_1 - \sum_{l=1}^{s-1}\rho_l(I-\Pi)Av_l\right)\\ &= \frac{1}{h_{2,1}h_{3,2}\cdots h_{s,s-1}}(I-\Pi)A^sVe_1.\end{aligned}$$

Define $\hat{V} := [Ve_1, AVe_1, A^2Ve_1, \ldots, A^{s-1}Ve_1]$ and define $\hat{H}$ by $\hat{V}\hat{H} = \Pi A\hat{V}$. Then the tuple $(A, \hat{V}, \Pi, \hat{H}, s)$ is a Krylov tuple. $\hat{H}$ is a Frobenius matrix

$$\hat{H} = \begin{bmatrix} 0 & \cdots & 0 & \hat{h}_{1s} \\ 1 & \ddots & \vdots & \hat{h}_{2s} \\ & \ddots & 0 & \vdots \\ 0 & & 1 & \hat{h}_{ss} \end{bmatrix}$$

with characteristic polynomial $\phi(x) = x^s - \sum_{l=1}^{s} \hat{h}_{ls} x^{l-1}$. Using Lemma 5.1 we obtain

$$(I - \Pi)A^sVe_1 = A^sVe_1 - \sum_{l=1}^{s} \hat{h}_{ls} A^{l-1}Ve_1 = \phi(A)Ve_1.$$

By Theorem 5.1 a) the matrices $\hat{H}$ and H are similar. Hence Lemma 5.12 is proved. ∎

Now we are set to prove Theorem 5.8.

Proof of Theorem 5.8. Define a polynomial ϕ by $\phi(x) := (x - \lambda_1)(x - \lambda_2) \cdots (x - \lambda_s)$. By Lemma 5.11 there is a projection Π on $\mathcal{K}(s, A, v_1)$ such that $\phi(A)v_1 = (I - \Pi)A^s v_1$. Define H by $VH = \Pi AV$. By Lemma 5.12 the polynomial ϕ is the characteristic polynomial of H and (A, V, Π, H, s) is a Krylov tuple. ∎

Corollary 5.2 *Let $A \in \mathrm{Mat}_N$ and $v_1 \in \mathbb{C}^N$ be given. Choose an arbitrary unreduced upper Hessenberg matrix $H = [h_{i,j}] \in \mathrm{Mat}_s$ with $s < m := \dim \mathcal{K}(N, A, v_1)$. Then there are matrices Π and V such that (A, V, Π, H, s) is a Krylov tuple and $Ve_1 = v_1$.*

Proof. For $i = 1, 2, \ldots, s-1$ set

$$v_{i+1} := \frac{1}{h_{i+1,i}} \left(AV_i - \sum_{l=1}^{i} h_{l,i} v_l \right).$$

Choose $V := [v_1, v_2, \ldots, v_s]$. By Lemma 5.11 there is a pojection Π with $\mathrm{Im}\,\Pi = \mathcal{K}(s, A, v_1)$ and

$$\Pi A v_s = \sum_{l=1}^{s} h_{l,s} v_l.$$

By construction the vectors $v_1, v_2, \ldots, v_s$ form a basis of the Krylov chain $(\mathcal{K}(i, A, v_1))_{1 \le i \le s}$ and the equation

$$VH = [Av_1, Av_2, \ldots, \Pi A v_s] = \Pi AV$$

holds. Hence (A, V, Π, H, s) is the desired Krylov tuple. ∎

If the projection Π is not "very oblique" (i.e., if $||\Pi||$ is of modest size), the matrix V is not ill-conditioned and the gap (see (5.3) below) between $\mathcal{K}(s, A, v_1)$ and an eigenvector u of A is small, then u and the corresponding eigenvalue λ are well approximated by an eigenpair of H. This is the content of Theorem 5.9 below. We recall that the gap $d(u, V)$ between a vector u and a subspace $\mathcal{V}$ is defined as

$$d(u, \mathcal{V}) := \frac{||(I - \hat{\Pi})u||}{||u||} \tag{5.3}$$

where $\hat{\Pi}$ is the orthogonal projection onto $\mathcal{V}$.

Theorem 5.9 (Approximation of eigenpairs) *Let (A, V, Π, H, s) be a Krylov tuple. Let λ be an eigenvalue of A with corresponding eigenvector u. Suppose that u is not orthogonal to $\mathcal{K}(s, A, v_1)$. Then there is a vector y with $||y|| = 1$ such that the estimate*

$$||(H - \lambda I)y|| \leq \kappa_2((V^*V)^{-1}V^*)||\Pi||\,||(A - \lambda I)||\,d(u, \mathcal{K}(s, A, v_1))$$

holds.

Proof. Define $\hat{\Pi} = V(V^*V)^{-1}V^*$. Using $H = (V^*V)^{-1}V^*\Pi AV$, $\Pi\hat{\Pi} = \hat{\Pi}$ and $(A - \lambda I)u = 0$ we obtain the equation

$$\begin{aligned}
(H - \lambda I)(V^*V)^{-1}V^*u &= ((V^*V)^{-1}V^*\Pi AV - \lambda I)(V^*V)^{-1}V^*u \\
&= (V^*V)^{-1}V^*(\Pi AV - \lambda V)(V^*V)^{-1}V^*u \\
&= (V^*V)^{-1}V^*(\Pi A - \lambda I)V(V^*V)^{-1}V^*u \\
&= (V^*V)^{-1}V^*\Pi(A - \lambda I)\hat{\Pi}u \\
&= (V^*V)^{-1}V^*\Pi(A - \lambda I)(\hat{\Pi} - I)u.
\end{aligned}$$

Hence

$$||(H - \lambda I)(V^*V)^{-1}V^*u|| \leq ||(V^*V)^{-1}V^*||\,||\Pi||\,||(A - \lambda I)||\,||(I - \hat{\Pi})u||.$$

The assumption $u \not\perp \mathcal{K}(s, A, v_1)$ implies that $||(V^*V)^{-1}V^*u||$ does not vanish.
Set $y := (V^*V)^{-1}V^*u/||(V^*V)^{-1}V^*u||$. Using that

$$\frac{(V^*V)^{-1}V^*u}{||u||} \geq \sigma_{min}((V^*V)^{-1}V^*) \qquad \text{and} \qquad ||(V^*V)^{-1}V^*|| = \sigma_{max}((V^*V)^{-1}V^*)$$

we conclude

$$\sigma_{min}((V^*V)^{-1}V^*)||(H - \lambda I)y|| \leq \sigma_{max}((V^*V)^{-1}V^*)\,||\Pi||\,||(A - \lambda I)||\,\frac{||(I - \hat{\Pi})u||}{||u||}$$

which proves the theorem. ∎

Saad [66, 67] and Jia [43] showed that the gap (see (5.3)) between eigenvectors of A and the Krylov subspace $\mathcal{K}(s, A, v_1)$ tends to zero for increasing s. They also showed that the convergence is particular fast for eigenvectors corresponding to eigenvalues with maximal real part.

In view of the relation

$$H = \Pi A|_{\mathcal{K}(s,A,v_1)}$$

the eigenvalues of H do not depend on the choice of the basis V of $(\mathcal{K}(i, A, v_1))_{1\leq i\leq s}$. Therefore it is possible to derive an estimate for the approximation of the eigenvalues which does not depend on the condition number of V. This is done in Corollary 5.3 and Theorem 5.10 below.

Corollary 5.3 (Approximation of eigenvalues) *Let (A, V, Π, H, s) be a Krylov tuple. Let λ be an eigenvalue of A with corresponding eigenvector u. Suppose that u is not orthogonal to $\mathcal{K}(s, A, v_1)$. Then there is a vector y with $||y|| = 1$ and a matrix $\tilde{H}$, similar to H, such that the estimate*

$$||(\tilde{H} - \lambda I)y|| \leq ||\Pi||\ ||(A - \lambda I)||\, d(u, \mathcal{K}(s, A, v_1))$$

holds.

Proof. Choose an orthonormal basis $v_1, v_2, \ldots, v_s$ of $(\mathcal{K}(s, A, v_1))_{1\leq i\leq s}$.
Put $\tilde{V} := [v_1\, v_2\, \ldots, v_s]$.
Define $\tilde{H} \in \mathrm{Mat}_s$ by

$$\tilde{V}\tilde{H} = \Pi A\tilde{V}.$$

Then $(A, \tilde{V}, \tilde{\Pi}, \tilde{H}, s)$ is a Krylov tuple and $\kappa_2((V^*V)^{-1}V^*) = 1$.
Theorem 5.9 yields the estimate

$$||(\tilde{H} - \lambda I)y|| \leq ||\Pi||\ ||(A - \lambda I)||\, d(u, \mathcal{K}(s, A, v_1)).$$

Finally, by Theorem 5.1, there is an invertible matrix R such that $\tilde{H} = R^{-1}HR$. ∎

Theorem 5.10 (Approximation of eigenvalues) *Let (A, V, Π, H, s) be a Krylov tuple. Let $||Ve_1|| = 1$. Let ϕ be the characteristic polynomial of $H = [h_{i,j}]$. Let $J \in \mathrm{Mat}_N$ be to Jordan normal form of A. Let F be the transformation matrix on Jordan normal form (i.e. $J = FAF^{-1}$). Set $y := FVe_1/||FVe_1|| = (y_i)_{1\leq i\leq N}$*

a) Then the estimate

$$||\phi(J)y|| \leq \kappa_2(F)|h_{2,1}h_{3,2}\cdots h_{s,s-1}|\ ||(I - \Pi)AVe_s||$$

holds.

b) Suppose that A is diagonalizable.

$$J = diag(\lambda_1, \lambda_2, \ldots, \lambda_N).$$

Put $m := \dim \mathcal{K}(N, A, Ve_1)$. Then there is a permuation π of the indices $1, 2, \ldots, N$ such that

$$\sqrt{\sum_{i=1}^{m} |y_{\pi(i)}\phi(\lambda_{\pi(i)})|^2} \leq \kappa_2(F)|h_{2,1}h_{3,2}\cdots h_{s,s-1}|\|(I-\Pi)AVe_s\|,$$

$$\sum_{i=1}^{m} |y_{\pi(i)}|^2 = 1$$

and $y_{\pi(i)} \neq 0$ for $i = 1, 2, \ldots, m$.
Further $\lambda_{\pi(1)}, \lambda_{\pi(2)}, \ldots, \lambda_{\pi(m)}$ are the eigenvalues of of $A|_{\mathcal{K}(N,A,Ve_1)}$.

Proof. Combining the equation (5.2) and the inequality

$$\begin{aligned} \sigma_{max}(F)\|\phi(A)Ve_1\| &= \|F\|\,\|\phi(A)Ve_1\| \\ &\geq \|F\phi(A)F^{-1}FVe_1\| = \|\Phi(J)FVe_1\| \\ &\geq \|\Phi(J)FVe_1\|\,\sigma_{min}(F) \end{aligned}$$

we obtain statement a).
Suppose that A is diagonalizable. Then the vector $y := FVe_1/\|FVe_1\|$ has exactly m non-vanishing entries. Thus, there is a permutation π such that

$$\begin{aligned} y_{\pi(i)} &\neq 0 \quad \text{for} \quad i = 1, 2, \ldots, m \\ y_{\pi(i)} &= 0 \quad \text{for} \quad i = m+1, m+2, \ldots, N. \end{aligned}$$

By construction, the identity

$$1 = \|y\|^2 = \sum_{i=1}^{m} |y_{\pi(i)}|^2$$

holds. Finally the equation

$$\|\phi(J)y\| = \|\mathrm{diag}\,(\phi(\lambda_1), \phi(\lambda_2), \ldots, \phi(\lambda_N))\| = \sqrt{\sum_{i=1}^{m} |y_{\pi(i)}\phi(\lambda_{\pi(i)})|^2}$$

proves the theorem. ∎

Remark 5.8 By Lemma 5.12 the equation

$$(I-\Pi)A^sVe_1 = h_{2,1}h_{3,2}\cdots h_{s,s-1}(I-\Pi)AVe_s$$

holds. Hence the product $|h_{2,1}h_{3,2}\cdots h_{s,s-1}|\|(I-\Pi)AVe_s\|$ does not depend on the choice of V. It just depends on the choice of Π, A, s and the vector Ve_1.

We recall that by the Bauer-Fike Theorem (Lemma 5.13 below) the conditon number $\kappa_2(F)$ is a measure for the sensitivity of the eigenvalues of a diagonalizable matrix

$$A = F^{-1}\mathrm{diag}\,(\lambda_1, \lambda_2, \ldots, \lambda_N)F.$$

Lemma 5.13 (Bauer-Fike) *If μ is an eigenvalue of $A + E \in \mathrm{Mat}_N$ and $F^{-1}AF = diag\,(\lambda_1, \lambda_2, \ldots, \lambda_N)$, then there is an eigenvalue λ_i of A such that*

$$|\mu - \lambda_i| \leq \kappa_2(F)||E||.$$

Notice, that the right hand side of the estimates of Theorem 5.10 can be used as a cheap stopping criterion

$$|h_{2,1}h_{3,2}\cdots h_{s,s-1}|\|(I-\Pi)AVe_s\| \leq \epsilon$$

in a Krylov subspace projection method. Indeed, the numbers $h_{2,1}, h_{3,2}, \ldots, h_{s,s-1}$ have already been computed while generating the Krylov tuple (A, V, Π, H, s) and the number $||(I-\Pi)AVe_s||$ is used for the construction of a continuation of (A, V, Π, H, s).

If (A, V, Π, H, s) is a maximal Krylov tuple (i.e. $s = m := \dim \mathcal{K}(N, A, Ve_1)$) then H is the restriction of A to the A-invariant subspace $\mathcal{K}(s, A, Ve_1)$. Thus, the characteristic polynomial of H is a divisor of the minimal polynomial of A. The following theorem shows in which cases the characteristic polynomial of H is a divisor of the minimal polynomial of A and $\mathcal{K}(s, A, Ve_1)$ is not A-invariant.

Theorem 5.11 *Let (A, V, Π, H, s) be a Krylov tuple. Define $v_{s+1} := (I-\Pi)AVe_s$. Then there exists a maximal continuation $(A, [V, V_2], \tilde{\Pi}, \tilde{H}, m)$ with the property*

$$\tilde{H} = \begin{bmatrix} H & 0 \\ \star & H_2 \end{bmatrix} \qquad \text{for some } H_2 \in \mathrm{Mat}_{m-s} \tag{5.4}$$

if and only if

$$\dim \mathcal{K}(N, A, v_{s+1}) \leq m - s. \tag{5.5}$$

In this case the following holds:

a) The product of the characteristic polynomials of H and H_2 is equal to the minimal polynomial of $A|_{\mathcal{K}(N,A,Ve_1)}$.

b) $\dim \mathcal{K}(N, A, v_{s+1}) = m - s$.

c) The columns of the matrix V_2 form a basis of the chain $(\mathcal{K}(j, A, v_{s+1}))_{1\leq j\leq m-s}$.

In the proof of Theorem 5.11 we use the following Lemma.

Lemma 5.14 *Let* $q \in \mathbb{C}^N \setminus \{0\}$ *and* $\tilde{q} \in \mathcal{K}(s+1, A, q) \setminus \mathcal{K}(s, A, q)$ *for some* $s+1 \leq m := \dim \mathcal{K}(N, A, q)$. *Then* $\dim \mathcal{K}(N, A, \tilde{q}) \geq m - s$.

Proof of Lemma 5.14. Assume $\dim \mathcal{K}(N, A, \tilde{q}) < m - s$. Then there are polynomials ψ and $\phi \neq 0$ such that $\psi(A)q = \tilde{q}$, $\phi(A)q = 0$, $\deg \psi = s$ and $\deg \phi < m - s$. Hence $0 = \psi(A)\phi(A)q = (\psi\phi)(A)q$ and $\deg(\psi\phi) < m$ which contradicts the definition of m. ∎

Proof of Theorem 5.11.

i) Let $(A, [V, V_2], \tilde{\Pi}, \tilde{H}, m)$ be a continuation with the property (5.4). Define $\tilde{V} := [V, V_2]$.The identity $\tilde{V}\tilde{H} = \tilde{\Pi}A\tilde{V}$ combined with the property (5.4) yields

$$\begin{aligned} Av_m &\in \operatorname{span}\{v_{s+1}, v_{s+1}, \ldots, v_m\} = \operatorname{Im} V_2 \\ Av_{s+j} &\in \operatorname{span}\{v_{s+1}, v_{s+1}, \ldots, v_{s+j+1}\} \qquad \text{for} \quad j = 1, 2, \ldots, m-s-1. \end{aligned}$$

Thus, $\operatorname{Im} V_2$ is A-invariant. Therefore

$$\mathcal{K}(N, A, v_{s+1}) \subseteq \operatorname{Im} V_2 \quad \text{and} \quad \dim \mathcal{K}(N, A, v_{s+1}) \leq \dim \operatorname{Im} V_2 = m - s.$$

ii) Let $\dim \mathcal{K}(N, A, v_{s+1}) \leq m - s$. By Lemma 5.14, $\dim \mathcal{K}(N, A, v_{s+1}) = m - s$. Define $V_2 := [v_{s+1}, Av_{s+1}, \ldots, A^{m-s-1}v_{s+1}]$. The column vectors of the matrix $\tilde{V} := [V, V_2]$ form a basis of the Krylov chain $(\mathcal{K}(j, A, v_1))_{1 \leq j \leq m}$. Define a matrix $\tilde{H}$ by the equation $\tilde{V}\tilde{H} = A\tilde{V}$. Property (5.5) yields $A^{m-s}v_{s+1} \in \operatorname{Im} V_2$. Hence, the matrix $\tilde{H}$ has the property (5.4). Choose an arbitrary projection $\tilde{\Pi}$ onto $\mathcal{K}(N, A, v_1)$. Then the tuple $(A, \tilde{V}, \tilde{\Pi}, \tilde{H}, m)$ is a continuation of the Krylov tuple (A, V, Π, H, s) with the desired property (5.4). ∎

Corollary 5.4 *Let* (A, V, Π, H, s) *be a Krylov tuple. Define* $v_{s+1} := (I - \Pi)AVe_s$. *Let*

$$\Pi\, \mathcal{K}(N, A, v_{s+1}) = 0. \tag{5.6}$$

Then there is a continuation $(A, \tilde{V}, \tilde{\Pi}, \tilde{H}, m)$ *with the property* (5.4).

Proof. By assumption the following relations hold:

$$\operatorname{Im} \Pi = \mathcal{K}(s, A, v_1) \quad \text{and} \quad \operatorname{Ker} \Pi \supseteq \mathcal{K}(N, A, v_{s+1}).$$

Hence $\mathcal{K}(s, A, v_1) \cap \mathcal{K}(N, A, v_{s+1}) = \{0\}$. Using the inclusions $\mathcal{K}(N, A, v_1) \supseteq \mathcal{K}(s, A, v_1)$ and $\mathcal{K}(N, A, v_1) \supseteq \mathcal{K}(N, A, v_{s+1})$ we obtain the inequality

$$\dim \mathcal{K}(N, A, v_1) \geq \dim \mathcal{K}(s, A, v_1) + \dim \mathcal{K}(N, A, v_{s+1}).$$

Theorem 5.11 proves the corollary. ∎

Example 5.5 shows that the condition (5.6) is not necessary for the existence of a continuation with the property (5.4).

Example 5.5 Choose arbitrary numbers $a, b, c, d \in \mathbb{C}$ and $\alpha \in \mathbb{C} \setminus \{0\}$. Put

$$A = \begin{bmatrix} a & 0 & 0 \\ 1 & b & c \\ 0 & 1 & d \end{bmatrix}, \qquad V = [e_1], \qquad H = [a]$$

and define a projection Π by $\operatorname{Im}\Pi = \operatorname{span}\{e_1\}$ and $\operatorname{Ker}\Pi = \operatorname{span}\{e_2, \alpha e_1 + e_3\}$. Then $(A, V, \Pi, H, 1)$ is a Krylov tuple. Choosing $\tilde{V} := I$, $\tilde{H} := A$ and $\tilde{\Pi} := I$, we obtain a continuation $(A, \tilde{V}, \tilde{\Pi}, 3)$ with the property (5.4). But $\Pi\mathcal{K}(N, A, (I - \Pi)Ae_1) = \operatorname{span}\{\alpha e_1\} \neq \{0\}$.

With the help of Corollary 5.4 we can prove Taylor's Mismatch Theorem in a very short and instructive manner. Taylor's Mismatch Theorem has been proved by Taylor [73] in 1982. Parlett [58] explained this amazing result using Kalman's normal form [45] in 1992. He found that an incurable breakdown occurs only if a minimal realization of the system (w_1, A, v_1) is found. Ziegler [81] proved Taylor's Mismatch Theorem using Parlett's idea and a result of Kronecker [48] about the rank of infinite Hankel matrices. Gutknecht [33] gave another proof of the Mismatch Theorem. The proof of the Mismatch Theorem suggested below is much shorter than the proofs of Taylor [73], Gutknecht [33] and Ziegler [81].

Theorem 5.12 (Mismatch Theorem) *Let (A, V, Π, H, r) and $(A^T, W, \Pi^T, \hat{H}, r)$ be generated by a (look-ahead) Lanczos algorithm. Suppose that an incurable breakdown occurs after step r (i.e. $r < m_L$, see Remark 3.2). Then the characteristic polynomial of the generated matrix H is a divisor of the minimal polynomial of the matrix A.*

Proof. Recall that an incurable breakdown occurs after step s if and only if $\Gamma_l := (\langle w_1, A^{i+j}v_1\rangle)_{0\leq i,j\leq l}$ is not invertible for all $l \geq s+1$. We show by induction on i that $\Pi A^i v_{r+1} = 0$ for all $i \in \mathbb{N}$. Then Corollary 5.4 proves the Mismatch Theorem.
For $i = 0$ we have: $(I - \Pi)A^0 v_{r+1} = (I - \Pi)v_{r+1} = v_{r+1}$.
Assume the statement $\Pi A^j v_{r+1} = 0$ is true for some $i \geq 0$ and all $j \in \{0, 1, \ldots, i\}$.
Now we proceed to $i+1$: We have to show that the term $\langle (A^T)^l w_1, A^{i+1}v_{r+1}\rangle$ vanishes for all $l \in \{0, 1, \ldots, r-1\}$. By the induction hypothesis we have

$$\langle (A^T)^l w_1, A^{i+1}v_{r+1}\rangle = \langle (A^T)^{l+1} w_1, A^i v_{r+1}\rangle = \langle (I - \Pi^T)(A^T)^{l+1} w_1, A^i v_{r+1}\rangle.$$

By the definiton of the projection Π the vector $(I - \Pi^T)(A^T)^{l+1}w_1$ vanishes for all $l \in \{0, 1, \ldots, r-2\}$. Finally the identity $0 = \langle w_{r+1}, A^i v_{r+1}\rangle = \langle (I - \Pi^T)(A^T)^r w_1, A^i v_{r+1}\rangle$ yields the assertion. ∎

Taylor's Mismatch Theorem can be easily extended to modified (look-ahead) Lanczos methods. This is done in Corollary 5.5

Corollary 5.5 (Extended Mismatch Theorem)
Let (A, V, Π, H, r) *and* $(A^+, W, \Pi^+, \hat{H}, r)$ *be the Krylov tuples generated by a modified (look-ahead) Lanczos algorithm. Suppose that an incurable breakdown occurs after step* r *(i.e* $r < m_L$*, see Remark 5.4). Then the characteristic polynomial of the generated matrix* H *is a divisor of the minimal polynomial of the matrix* A*.*
Further there is a maximal continuation $(A, [V, V_2], \tilde{\Pi}, \tilde{H}, m)$ *of* (A, V, Π, H, r) *and a maximal continuation* $(A^+, [W, W_2], \check{\Pi}, \check{H}, \hat{m})$ *of* $(A^+, W, \Pi^+, \hat{H}, r)$ *such that*

$$\tilde{H} = \begin{bmatrix} H & 0 \\ \star & H_2 \end{bmatrix} \qquad \text{and} \qquad \check{H} = \begin{bmatrix} \hat{H} & 0 \\ \star & \check{H}_2 \end{bmatrix} \tag{5.7}$$

where $\tilde{H} \in \mathrm{Mat}_m$ *and* $\check{H} \in \mathrm{Mat}_{\hat{m}}$ *are unreduced upper Hessenberg matrices and*

$$m := \dim \mathcal{K}(N, A, Ve_1), \qquad \hat{m} := \dim \mathcal{K}(N, A^+, We_1).$$

Proof. By Theorem 5.5 there is a (look-ahead) Lanczos method which generates the Krylov tuples (A, V, Π, H, r) and $(A^T, S^T W, \Pi^T, \hat{H}, r)$ and breaks down incurably after step r. Theorem 5.12 applies and it yields that the characteristic polynomial of H is a divisor of the minimal polynomial of A. In the proof of Theorem 5.12 we have shown that the relation (5.6) holds if an incurable breakdown occurs. Finally, by Corollary 5.4 there are continuations with the properties (5.7). ∎

By Corollary 5.5 an incurable breakdown of the (modified) (look-ahead) Lanczos method can be overcome. Theorem 5.13 below shows how to construct such a continuation.

Theorem 5.13 *Let* (A, V, Π, H, s) *be a Krylov tuple with the property* (5.5) *or* (5.6)*. Then a continuation with the property* (5.4) *can be constructed as follows:*

(1) $v_{s+1} := \beta(I - \Pi)Av_s$ *for an arbitrary* $\beta \in \mathbb{C} \setminus \{0\}$*.*

(2) Choose an arbitrary Krylov subspace projection method with starting vector v_{s+1} *to generate a Krylov tuple* $(A, V_2, \check{\Pi}, H_2, m - s)$*.*

(3) Put

$$\tilde{V} := [V, V_2], \qquad G := \begin{bmatrix} 0 & \beta^{-1} \\ 0 & 0 \end{bmatrix}, \qquad \tilde{H} := \begin{bmatrix} H & 0 \\ G & H_2 \end{bmatrix}.$$

(4) Choose an arbitrary projection $\tilde{\Pi}$ *on* $\mathcal{K}(N, A, v_1)$*.*

A slight modification of part ii) of the proof of Theorem 5.11 proves Theorem 5.13.

Theorem 5.14 and Corollary 5.6 below state conditions which imply that no incurable breakdown of the (modified) look-ahead Lanczos method occurs.

Theorem 5.14 (Existence of an incurable breakdown) *Let $A \in \mathrm{Mat}_N$. Let $v_1, w_1 \in \mathbb{C}^N \setminus \{0\}$. Choose an arbitrary non-degenerate bilinear form $\langle \cdot, \cdot \rangle_S$. Suppose that*

$$N = \max\{\dim \mathcal{K}(N, A, v_1), \dim \mathcal{K}(N, A^+, w_1)\}.$$

Then no incurable breakdown of the modified look-ahead Lanczos method occurs. In other words: Every modified look-ahead Lanczos method, applied to v_1, w_1, A, yields two Krylov tuples (A, V, Π, H, r) and $(A^+, W, \Pi^+, \hat{H}, r)$ where

$$r = m_L := \min\{\dim \mathcal{K}(N, A, v_1), \dim \mathcal{K}(N, A^+, w_1)\}.$$

Proof. By duality it is sufficient to consider the case where $N = m := \dim \mathcal{K}(N, A, v_1)$ is assumed.
We prove Theorem 5.14 by contradiction.
Suppose that an incurable breakdown of a modified look-ahead Lanczos method occurs (i.e. $r < m_L$). Then by Corollary 5.5 there is a maximal continuation $(A, [V, V_2], \tilde{\Pi}, \tilde{H}, N)$ of (A, V, Π, H, r) such that

$$[V, V_2]^{-1} A [V, V_2] = \tilde{H} = \begin{bmatrix} H & 0 \\ \star & H_2 \end{bmatrix}. \tag{5.8}$$

Defining $D := W^T S V$, $\tilde{D} := \mathrm{diag}\,(D, I_{N-r}) \in \mathrm{Mat}_N$ and $\tilde{W} := (S^{-1})^T ([V, V_2]^{-1})^T D^T$ we obtain the equations

$$\tilde{W}^T S [V, V_2] = \tilde{D} \qquad \text{and} \qquad \tilde{W} = [W, \star]. \tag{5.9}$$

From there, from $A^T = S^T A^+ (S^T)^{-1}$ (see Lemma 5.3) and $H = D^{-1} \hat{H}^T D$ (see Lemma 5.5) we find the identity

$$\begin{aligned} \tilde{W}^{-1} A^+ \tilde{W} &= (\tilde{D}^T)^{-1} [V, V_2]^T S^T A^+ (S^{-1})^T ([V, V_2]^{-1})^T \tilde{D}^T && (5.10) \\ &= (\tilde{D}^T)^{-1} [V, V_2]^T A^T ([V, V_2]^{-1})^T \tilde{D}^T = (\tilde{D}^T)^{-1} \tilde{H}^T \tilde{D}^T && (5.11) \\ &= \begin{bmatrix} (D^T)^{-1} H^T D^T & \star \\ 0 & H_2^T \end{bmatrix} = \begin{bmatrix} \hat{H} & \star \\ 0 & H_2^T \end{bmatrix} && (5.12) \end{aligned}$$

which implies

$$r = m_L = \dim \mathcal{K}(N, A^+, w_1). \tag{5.13}$$

Identity (5.13) contradicts the assumption $r < m_L$ which proves the theorem. ∎

Corollary 5.6 *Let $A \in \mathrm{Mat}_N$ be an unreduced upper Hessenberg matrix. Choose $v_1 := e_1$ and $w_1 \in \mathbb{C}^N \setminus \{0\}$. Choose an arbitrary non-degenerate bilinear form $\langle \cdot, \cdot \rangle_S$. Then no incurable breakdown of the modified look-ahead Lanczos method occurs.*

Proof. The unreduced upper Hessenberg form of A implies the identity

$$\mathcal{K}(N, A, e_1) = \mathbb{C}^N.$$

Hence Theorem 5.14 applies and proves the corollary. ∎

Using Corollary 5.5 and Lemma 5.5 we find immediately Corollary 5.7 below.

Corollary 5.7 *Let (A, V, Π, H, r) and $(A^+, W, \Pi^+, \hat{H}, r)$ be the Krylov tuples generated by a (modified) look-ahead Lanczos algorithm. Suppose that this algorithm terminates after step r. Then the characteristic polynomial of the generated matrix H is a divisor of the minimal polynomial of the matrix A. Further the matrices H and $\hat{H}$ are similar.*

6 Restarted Krylov subspace projection methods

6.1 Restarted Krylov methods and GR algorithms

We consider again a Krylov tuple (A, V, Π, H, s) with $s < m := \dim \mathcal{K}(N, A, Ve_1)$. In Section 5.4 we have seen that the eigenvalues of H are not eigenvalues of A in general. However the eigenvalues of largest real part are typically well approximated for a relatively small number s (see [43]). If k is the number of eigenvalues that are wanted, then the number s must be chosen much larger. In case of the Arnoldi method large s leads to considerable computational effort, in case of the symmetric Lanczos method reorthogonalization is necessary. For these reasons Saad [64] and Sorensen [70] suggested to restart the Arnoldi method after every $s << N$ steps with a new starting vector $v_1^{(i+1)} = \psi_i(A)v_1^{(i)}$. Here ψ_i is a polynomial used to filter out the unwanted eigenvalues.
In this Chapter we generalize the idea of Saad and Sorensen to general Krylov subspace projection methods. Furthermore we show that the general GR method, introduced by Watkins and Elsner [76], is in a certain sense mathematically equivalent to our restarted general Krylov subspace projection method. First we recall the definition of the general GR algorithm.

GR algorithm

Let $A \in \mathrm{Mat}_N$ be given. A GR algorithm generates a sequence of similar matrices $(A_i)_{i\in\mathbb{N}}$ as follows:

The matrix A_1 is taken to be A, or some convenient matrix similar to A, say

$$A_1 = G_0 A G_0^{-1}$$

(e.g. transformation to upper Hessenberg form or tridiagonal form).
Let A_i be given. Choose a shift polynomial ψ_{i-1} which approximates a divisor of the characteristic polynomial of A_i (e.g. choose ψ_{i-1} to be the characteristic polynomial of the lower right-hand 2×2 submatrix of A_i).
Next compute an invertible matrix $G_i \in \mathrm{Mat}_N$ and an upper triangular matrix R_i such that

$$\psi_{i-1}(A_i) = G_i R_i.$$

Set

$$A_{i+1} := G_i^{-1} A_i G_i.$$

Under some suitable conditions the sequence $(A_i)_{i\in\mathbb{N}}$ tends to block triangular form or even to upper triangular form. The QR (see [19]), LR (see [63]), HR (see [8], [10]) and SR (see [11], [12]) algorithms are examples of GR algorithms.

Now we turn to the definition of restarted Krylov subspace projection methods.

Restarted Krylov subspace projection method

Let $A \in \mathrm{Mat}_N$ be given. A restarted Krylov subspace projection method generates a sequence of matrices $(H_{(i)})_{i\in\mathbb{N}_0}$ as follows:

Choose a starting vector $v_1^{(0)}$.
Choose a number $s_0 \in \{1, 2, \ldots, m_0\}$ where $m_0 := \dim \mathcal{K}(N, A, v_1^{(0)})$.
Apply a Krylov subspace projection method to A and $v_1^{(0)}$ to obtain a Krylov tuple $(A, V_{(0)}, \Pi_{(0)}, H_{(0)}, s_0)$.
Let a Krylov tuple $(A, V_{(i)}, \Pi_{(i)}, H_{(i)}, s_i)$ be given.
Choose a shift polynomial ψ_i which approximates a divisor of the characteristic polynomial of A (e.g. choose ψ_i to be a divisor of the characteristic polynomial of $H_{(i)}$).
Next compute a new starting vector

$$v_1^{(i+1)} := \psi_i(A)v_1^{(i)}.$$

Choose a number $s_{i+1} \in \{1, 2 \ldots, m_{i+1}\}$ where $m_{i+1} := \dim \mathcal{K}(N, A, v_1^{(i+1)})$.
Apply a Krylov subspace projection method to A and $v_1^{(i+1)}$ to obtain a new Krylov tuple $(A, V_{(i+1)}, \Pi_{(i+1)}, H_{(i+1)}, s_{i+1})$.

Notice that the numbers $s_i \in \{1, 2 \ldots, m_i\}$ can be chosen arbitrarily. We will show that there is a GR algorithm such that for all $i \in \mathbb{N}$ the matrix $H_{(i-1)}$ is the leading $s_i \times s_i$ submatrix of the upper Hessenberg matrix A_i generated by the same GR algorithm, in other words:

$$A_1 = G_0^{-1}AG_0, \qquad \psi_{i-1}(A_i) = G_iR_i, \qquad A_{i+1} = G_i^{-1}A_iG_i = \begin{bmatrix} H_{(i)} & \star \\ \star & \star \end{bmatrix}. \tag{6.1}$$

For shortness of notation and clarity we introduce the notion of a restarted Krylov tuple.

Definition 6.1 (Restarted Krylov tuple)
A sequence of six-tuples $((A, V_{(i)}, \Pi_{(i)}, H_{(i)}, s_i, \psi_i))_{i\in\mathbb{N}_0}$ *is called a restarted Krylov tuple if for every* $i \in \mathbb{N}_0$ *the five-tuple* $(A, V_{(i)}, \Pi_{(i)}, H_{(i)}, s_i)$ *is a Krylov tuple,* ψ_i *is a polynomial and* $V_{(i+1)}e_1 = \psi_i(A)V_{(i)}e_1$.

Let $v_1^{(0)} \in \mathbb{C}^N \setminus \{0\}$ and $A \in \mathrm{Mat}_N$ be given. Then a restarted Krylov subspace projection method is a method which generates matrices $V_{(i)}$ and $H_{(i)}$ such that
$((A, V_{(i)}, \Pi_{(i)}, H_{(i)}, s_i, \psi_i))_{i\in\mathbb{N}_0}$ is a restarted Krylov tuple for an appropriate choice of $\Pi_{(i)}$ and ψ_i.

The following Lemma shows that the computation of the $(i+1)$-st starting vector $V_{(i+1)}e_1$ can be reduced to the computation of $V_{(i)}\psi_i(H_{(i)})e_1$.

Lemma 6.1 *Let $((A, V_{(i)}, \Pi_{(i)}, H_{(i)}, s_i, \psi_i))_{i\in\mathbb{N}_0}$ be a restarted Krylov tuple. Then the following statements hold:*

a) $\mathcal{K}(N, A, V_{(i+1)}e_1) \subseteq \mathcal{K}(N, A, V_{(i)}e_1)$ *for all* $i \in \mathbb{N}_0$.

b) If $\deg\psi_i < s_i$ *for some* i, *then* $V_{(i+1)}e_1 = V_{(i)}\psi_i(H_{(i)})e_1$.

c) If $s_i = m_i := \dim\mathcal{K}(N, A, V_{(i)}e_1)$ *for some* i, *then* $V_{(i+1)}e_1 = V_{(i)}\psi_i(H_{(i)})e_1$.

Proof. The identity

$$\begin{aligned}\mathcal{K}(N, A, V_{(i+1)}e_1) &= \mathcal{K}(N, A, \psi_i(A)V_{(i)}e_1)\\ &= \{\phi(A)\psi_i(A)V_{(i)}e_1 \mid \phi \in \mathrm{Pol}\}\\ &\subseteq \{\tilde{\phi}(A)V_{(i)}e_1 \mid \tilde{\phi} \in \mathrm{Pol}\}\\ &= \mathcal{K}(N, A, V_{(i)}e_1)\end{aligned}$$

yields a). Let $(A, \tilde{V}_{(i)}, \tilde{\Pi}_{(i)}, \tilde{H}_{(i)}, m_i)$ be a maximal continuation of $(A, V_{(i)}, \Pi_{(i)}, H_{(i)}, s_i)$. Using that $A^l\tilde{V}_{(i)} = \tilde{V}_{(i)}\tilde{H}^l_{(i)}$ for all $l \in \mathbb{N}$ we obtain the equation

$$V_{(i+1)}e_1 = \psi_i(A)V_{(i)}e_1 = \psi_i(A)\tilde{V}_{(i)}e_1 = \tilde{V}_{(i)}\psi_i(\tilde{H}_{(i)})e_1$$

which proves c). The upper Hessenberg structure of the matrix $\tilde{H}_{(i)}$ yields, for all polynomials ψ with $\deg\psi < s_i$, the representation

$$\psi(\tilde{H}_{(i)})e_1 = \begin{bmatrix} \psi(H_{(i)})e_1 \\ 0 \end{bmatrix}$$

which proves b). ∎

Now we are going to show that every restarted Krylov subspace projection method can be seen as a GR algorithm in the sense described in (6.1). To prove this we need some preliminary steps.

Definition 6.2 *A matrix $A \in \mathrm{Mat}_N$ is called k-Hessenberg if A has the form*

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

where $A_{11} \in M_k$ is upper Hessenberg and A_{21} consists entirely of zeros, with the possible exception of the single entry $a_{k+1,k}$ in the upper right corner.
We call a k-Hessenberg matrix A unreduced, if the matrix A_{11} is an unreduced upper Hessenberg.

Remark 6.1 The concept of k-Hessenberg matrices has been introduced by Watkins and Elsner [77]. Note that a k-Hessenberg matrix is not a band matrix.

Lemma 6.2 *Let $(A, \tilde{V}, \tilde{\Pi}, \tilde{H}, m)$ be a maximal Krylov tuple and $m < N$. Let $C \in \mathrm{Mat}_N$ be an invertible matrix such that $F := C^{-1}AC$ has the form*

$$F = \begin{bmatrix} \hat{F} & \star \\ 0 & \star \end{bmatrix} \tag{6.2}$$

where $\hat{F} \in \mathrm{Mat}_m$ is an unreduced upper Hessenberg matrix. Suppose there is a polynomial ψ with the property $\tilde{V}e_1 = C\psi(F)e_1$. Then there is an invertible matrix $\hat{G} \in \mathrm{Mat}_m$ such that

$$C\tilde{V} = \begin{bmatrix} \hat{G} \\ 0 \end{bmatrix}. \tag{6.3}$$

Further there is a matrix $V^{(2)} \in M_{N\times(N-m)}$ with the following properties:

a) The matrix $V^{(2)}$ has orthogonal columns and $\operatorname{Im} V^{(2)} \perp_* \operatorname{Im} \tilde{V}$.

b) There is an upper triangular matrix R such that

$$[\tilde{V}, V^{(2)}]R = C\psi(F) \qquad \text{and} \qquad R = \begin{bmatrix} \hat{R} & \star \\ 0 & \star \end{bmatrix}. \tag{6.4}$$

Here $\hat{R} \in \mathrm{Mat}_m$ is an invertible upper triangular matrix.

c)

$$[\tilde{V}, V^{(2)}]^{-1} A [\tilde{V}, V^{(2)}] = \begin{bmatrix} \tilde{H} & \star \\ 0 & \star \end{bmatrix}.$$

d) The matrices $[\tilde{V}, V^{(2)}]$ and $\tilde{V}$ have the same singular values.

Proof. Let the columns of $\tilde{V}$ be denoted by $q_1, q_2, \ldots, q_m$.
i) First we show that there is an invertible matrix $\hat{R} \in \mathrm{Mat}_m$ such that

$$\tilde{V}\hat{R} = C\psi(F)[e_1,\, e_2, \ldots\, e_m]. \tag{6.5}$$

By assumption the column vectors of $C^{-1}\tilde{V}$ form a basis of the Krylov chain $(C^{-1}\mathcal{K}(j, A, C\psi(F)e_1))_{1\le j\le m} = (\mathcal{K}(j, F, \psi(F)e_1))_{1\le j\le m}$. The matrix $\hat{F}$ is unreduced upper Hessenberg. Hence $\mathcal{K}(i, F, e_1) = \mathrm{span}\{e_1, e_2, \ldots, e_i\}$ for $i = 1, 2 \ldots, m$. We obtain the relation

$$\begin{aligned}\psi(F)\mathrm{span}\{e_1, e_2, \ldots, e_i\} &= \psi(F)\mathcal{K}(i, F, e_1)\\ &= \mathcal{K}(i, F, \psi(F)e_1) = C^{-1}\mathrm{span}\{q_1, \ldots, q_i\}\end{aligned}$$

which proves the statement (6.5).

ii) Next we construct the column vectors $v_1, v_2, \ldots, v_{N-m}$ of $V^{(2)}$. Let σ_1 denote the maximal singular value of $\tilde{V}$. Set $v_0 := q_m$. Assume that for some $i \ge 0$ the vectors $v_0, v_1, \ldots, v_i$ have been constructed such that

$$C\psi(F)e_{j+m} \in W_j := \mathrm{span}\{q_1, \ldots, q_{m-1}, v_0, \ldots, v_j\}$$

for $j = 0, 1 \ldots, i$ and $v_j \perp_* W_{j-1}$ for $j = 1, 2, \ldots, i$.
We construct the vector v_{i+1} as follows:
If $C\psi(F)e_{i+1+m} \notin W_i$ then define $\tilde{v}_{i+1} := C\psi(F)e_{i+1+m}$,
otherwise choose an arbitrary $\tilde{v}_{i+1} \in \mathbb{C}^N \setminus W_i$.
Set

$$v_{i+1} := \sqrt{\sigma_1}(I - O_i)\tilde{v}_{i+1}/\|(I - O_i)\tilde{v}_{i+1}\|$$

where O_i is the orthogonal projection on W_i.
By construction and equation (6.5) the matrix $V^{(2)}$ has orthogonal columns, there is an upper triangular matrix R such that the relations (6.4) hold and $\mathrm{Im}\, V^{(2)} \perp_* \mathrm{Im}\, \tilde{V}$. By assumption the equation $\tilde{V}\tilde{H} = A\tilde{V}$ holds. Hence the equation c) is proved. Assertion d) is a consequence of the relation

$$[\tilde{V}, V^{(2)}]^*[\tilde{V}, V^{(2)}] = \begin{bmatrix} (\tilde{V}^*\tilde{V}) & 0 \\ 0 & (\sigma_1 I) \end{bmatrix}.$$

Finally the representation (6.3) follows from the identities (6.2) and (6.5):

$$\begin{aligned} C^{-1}\tilde{V}\hat{R} &= \psi(F)[e_1\, e_2, \ldots\, e_m]\\ &= \begin{bmatrix} \psi(\hat{F}) & \star \\ 0 & \star \end{bmatrix} [e_1\, e_2, \ldots\, e_m]\\ &= \begin{bmatrix} \psi(\hat{F}) \\ 0 \end{bmatrix}. \end{aligned}$$

■

Lemma 6.3 *Let $(A, \tilde{V}, \tilde{\Pi}, \tilde{H}, m)$ be a maximal Krylov tuple. Then there is a matrix $V^{(2)} \in M_{N\times(N-m)}$ with the properties a),c),d) stated in Lemma 6.2.*

Proof. Define $\psi := 0$, $C := I$. Part ii) of the proof of Lemma 6.2 yields the assertion. ■

Now we are ready to prove that for every $i \in \mathbb{N}$ the matrix $H_{(i-1)}$, generated by a restarted Krylov subspace projection method, is a submatrix of the matrix A_i generated by an appropriately chosen GR algorithm.
Theorem 6.1 below states further that for every $i \in \mathbb{N}$ the condition number of the matrix $G_0G_1\cdots G_i$ depends only on the norm of the projection $\Pi_{(i)}$ and the condition number of $V_{(i)}$. These quantities can be bounded above by the user (e.g. in the practical look-ahead Lanczos process this is done by using checks (see Theorem 5.7 and Section 4).

Theorem 6.1 *Let $((A, V_{(i)}, \Pi_{(i)}, H_{(i)}, s_i, \psi_i))_{i\in\mathbb{N}_0}$ be a sequence of six-tuples generated by a restarted Krylov subspace projection method. Then there is a GR algorithm which generates a sequence of invertible matrices $(G_i)_{i\in\mathbb{N}_0}$, a sequence of upper triangular matrices $(R_i)_{i\in\mathbb{N}}$ and a sequence of matrices $(A_i)_{i\in\mathbb{N}}$ with the following properties (for $i \in \mathbb{N}$):*

a) $A_1 = G_0^{-1}AG_0, \qquad \psi_{i-1}(A_i) = G_iR_i, \qquad A_{i+1} = G_i^{-1}A_iG_i.$

b) Define $m_i := \dim \mathcal{K}(N, A, V_{(i)}e_1)$. The matrix A_i is unreduced m_i-Hessenberg and $H_{(i-1)}$ is the leading $s_i \times s_i$ submatrix of

$$A_i = \begin{bmatrix} H_{(i-1)} & \star \\ \star & \star \end{bmatrix}.$$

(Notice that $m_i \geq s_i$).

c) There is an invertible upper triangular matrix $\tilde{R}_i \in \mathrm{Mat}_{m_i}$ and a full rank matrix $\tilde{G}_{i+1} \in \mathrm{Mat}_{m_i\times m_{i-1}}$ such that

$$G_i = \begin{bmatrix} \tilde{G}_i & \star \\ 0 & \star \end{bmatrix}, \qquad R_i = \begin{bmatrix} \tilde{R}_i & \star \\ 0 & \star \end{bmatrix}.$$

d) $G_0G_1\cdots G_ie_l = V_ie_l$ for $l = 1, 2, \ldots, s_i$.

e) The following estimates hold if the column vectors of $V_{(i)}$ are normalized:

$$\frac{1}{\sigma_{\min}(G_0G_1\cdots G_i)} \leq \frac{\|\Pi_{(i)}\|}{\sigma_{\min}(V_{(i)})} + \|I - \Pi_{(i)}\| \tag{6.6}$$

$$\sigma_{\max}(G_0G_1\cdots G_i) \leq \sqrt{m_i} \leq \sqrt{m_0} \leq \sqrt{N}. \tag{6.7}$$

f) There is, for every $i \in \mathbb{N}$, a maximal continuation $(A, \tilde{V}_{(i-1)}, \tilde{\Pi}_{(i-1)}, \tilde{H}_{(i-1)}, m_{i-1})$ of $(A, V_{(i-1)}, \Pi_{(i-1)}, H_{(i-1)}, s_{i-1})$ with the following properties:

f1) A_i has the representation

$$A_i = \begin{bmatrix} \tilde{H}_{(i-1)} & \star \\ 0 & \star \end{bmatrix}, \qquad \tilde{H}_{(i-1)} = \begin{bmatrix} H_{(i-1)} & \star \\ \star & \star \end{bmatrix}.$$

f2) $G_0 G_1 \cdots G_{i-1} e_l = \tilde{V}_{(i-1)} e_l$ for $l = 1, 2, \ldots, m_{i-1}$,

$$\sigma_{min}(\tilde{V}_{(i-1)}) = \sigma_{min}(G_1 G_2 \cdots G_{i-1})$$

and

$$\sigma_{max}(\tilde{V}_{(i-1)}) = \sigma_{max}(G_1 G_2 \cdots G_{i-1}).$$

f3) Let $\tilde{G}_i$ be the matrix defined in c). Then

$$\tilde{G}_i \tilde{H}_{(i)} = \tilde{H}_{(i-1)} \tilde{G}_i \qquad \text{and} \qquad m_i \leq m_{i-1}.$$

f4) Let m be defined as $\min_{i \in \mathbb{N}_0} m_i$. Choose $i_0 \geq 1$ such that $m_{i_0} = m$. Then the identities

$$\psi_i(H_{(i-1)}) = \tilde{G}_i \tilde{R}_i \qquad \text{and} \qquad \tilde{H}_{(i)} = \tilde{G}_i^{-1} \tilde{H}_{(i-1)} \tilde{G}_i$$

hold for all $i \geq i_0 + 1$.

In the following m_i denotes the dimension of $\dim \mathcal{K}(N, A, V_{(i)} e_1)$.
Notice that in Theorem 6.1 the numbers $s_i \in \{1, 2, \ldots, m_i\}$ can be chosen arbitrarily. We remark further that in general the identity $m_i = m_{i+1}$ holds for all $i \in \mathbb{N}_0$. Indeed, by Lemma 6.4 below the case $m_i \neq m_{i+1}$ occurs if and only if there is a common divisor of the minimal polynomial of A and the shift polynomial ψ_i.

Proof of Theorem 6.1. We prove a), b), c), d), f1) and f2) by induction on i.
By Remark 5.6 and Lemma 6.1 there is, for every $i \in \mathbb{N}$, a maximal continuation $(A, \tilde{V}_{(i)}, \tilde{\Pi}_{(i)}, \tilde{H}_{(i)}, m_i)$ of $(A, V_{(i)}, \Pi_{(i)}, H_{(i)}, s_i)$ with the properties

$$\frac{1}{\sigma_{\min}(\tilde{V}_{(i)})} \leq \frac{||\Pi_{(i)}||}{\sigma_{\min}(V_{(i)})} + ||I - \Pi_{(i)}|| \tag{6.8}$$

$$\sigma_{\max}(G_0 G_1 \cdots G_i) \leq \sqrt{m_i} \leq \sqrt{m_0} \leq \sqrt{N}. \tag{6.9}$$

Applying Lemma 6.3 to the Krylov tuple $(A, \tilde{V}_{(0)}, \tilde{\Pi}_{(0)}, \tilde{H}_{(0)}, m_0)$ we obtain the matrices $G_0 := [\tilde{V}_{(0)}, V^{(2)}]$ and $A_1 := G_0^{-1} A G_0$ which have the desired properties a)-d), f1) and f2).

Assume there are invertible matrices G_j and matrices A_j which have the properties a)-d), f1) and f2) for some $i \geq 0$ and all $j = 0, 1, \ldots, i$.
Now we proceed to $i+1$:
Define $C := G_0 G_1 \cdots G_i$, $\tilde{V} := V_{(i+1)}$, $\psi := \psi_i$ and $F := C^{-1}AC = A_{i+1}$. Then we have

$$\tilde{V}e_1 = \psi(A)\tilde{V}_{(i)}e_1 = \psi(A)Ce_1 = C\psi(F)e_1.$$

Hence Lemma 6.2 is applicable and yields a matrix $V^{(2)} \in M_{n\times(N-m_{i+1})}$ and an upper tringular matrix R such that the following equations hold:

$$[\tilde{V}, V^{(2)}]R = C\psi(F) \tag{6.10}$$

$$A_{i+2} := [\tilde{V}, V^{(2)}]^{-1}A[\tilde{V}, V^{(2)}] = \begin{bmatrix} \tilde{H}_{(i+1)} & \star \\ \star & \star \end{bmatrix} \tag{6.11}$$

$$\sigma_{min}(\tilde{V}) = \sigma_{min}([\tilde{V}, V^{(2)}]) \qquad \text{and} \qquad \sigma_{max}(\tilde{V}) = \sigma_{max}([\tilde{V}, V^{(2)}]). \tag{6.12}$$

Equations (6.10) and (6.11) imply the relations

$$[C^{-1}\tilde{V}, C^{-1}V^{(2)}]R = \psi(A_{i+1})$$

and

$$\begin{aligned} A_{i+2} &= [\tilde{V}, V^{(2)}]^{-1}CC^{-1}ACC^{-1}[\tilde{V}, V^{(2)}] \\ &= [C^{-1}\tilde{V}, C^{-1}V^{(2)}]^{-1}C^{-1}AC[C^{-1}\tilde{V}, C^{-1}V^{(2)}] \\ &= [C^{-1}\tilde{V}, C^{-1}V^{(2)}]^{-1}A_{i+1}[C^{-1}\tilde{V}, C^{-1}V^{(2)}]. \end{aligned}$$

Set $G_{i+1} := [C^{-1}\tilde{V}, C^{-1}V^{(2)}]$. Thus,

$$\begin{aligned} G_0G_1\cdots G_{i+1} &= G_0G_1\cdots G_i[C^{-1}\tilde{V}, C^{-1}V^{(2)}] \\ &= C[C^{-1}\tilde{V}, C^{-1}V^{(2)}] = [\tilde{V}, V^{(2)}] \end{aligned}$$

which proves a)-d), f1) and f2). Combining equation (6.12) and the inequalities (6.8) and (6.9) yields e).
By Lemma 6.1 a) the inequality $m_i \leq m_{i-1}$ holds. Combining a), c) and f1) we obtain the equation

$$\begin{bmatrix} \tilde{G}_i & \star \\ 0 & \star \end{bmatrix} \begin{bmatrix} \tilde{H}_{(i)} & \star \\ 0 & \star \end{bmatrix} = G_iA_{i+1} = A_iG_i = \begin{bmatrix} \tilde{H}_{(i-1)} & \star \\ 0 & \star \end{bmatrix} \begin{bmatrix} \tilde{G}_i & \star \\ 0 & \star \end{bmatrix}$$

which yields the identity $\tilde{G}_i\tilde{H}_{(i)} = \tilde{H}_{(i-1)}\tilde{G}_i$. Thus, f3) is proved.
f3) implies $\tilde{H}_{(i)} = \tilde{G}_i^{-1}\tilde{H}_{(i-1)}\tilde{G}_i$ for $i \geq i_0$.
Finally a), b) and f1) imply the identity

$$\begin{bmatrix} \psi_i(H_{(i-1)}) & \star \\ 0 & \star \end{bmatrix} = \psi_{i-1}(A_i) = G_iR_i = \begin{bmatrix} \tilde{G}_i & \star \\ 0 & \star \end{bmatrix} \begin{bmatrix} \tilde{R}_i & \star \\ 0 & \star \end{bmatrix}$$

which yields $\psi_i(H_{(i-1)}) = \tilde{G}_i\tilde{R}_i$. ∎

Lemma 6.4 *Let $\phi \neq 0$ be an arbitrary polynomial, $v \in \mathbb{C}^N \setminus \{0\}$ and define $m := \dim \mathcal{K}(N, A, v)$. Then the following statements are equivalent for every $l \in \{0, 1, \ldots, m\}$:*

a) $\dim \mathcal{K}(N, A, \phi(A)v) = m - l$

b) The maximal degree of a common divisor of ϕ and the minimal polynomial of $A|_{\mathcal{K}(N,A,v)}$ is equal to l.

Lemma 6.4 is evident.

6.2 Connections between the QR and the restarted Arnoldi algorithms

A restarted Krylov subspace projection method is called restarted Arnoldi method if every Krylov tuple $(A, V_{(i)}, \Pi_{(i)}, H_{(i)}, s_i)$ is generated by performing s_i steps of the Arnoldi method. Here $s_i \in \{1, 2, \ldots, m_i\}$ can be chosen arbitrarily. Theorem 6.2 below describes the relations between restarted Arnoldi methods and the QR algorithm. To prove this theorem we use the technical Lemmas 6.5 and 6.6 below.

Lemma 6.5 *Let ψ be a non-constant polynomial. Let $k, s \in \{1, 2, \ldots, N\}$ be arbitrary. Let $A \in \mathrm{Mat}_N$ be an unreduced k-Hessenberg matrix. Define $m := \dim \mathcal{K}(N, A, \psi(A)e_1)$, $r := \min\{s, k, m\}$. Let $G \in \mathrm{Mat}_N$ be an invertible matrix and let $R \in \mathrm{Mat}_{N\times s}$ be upper triangular. Suppose that*

$$\psi(A)e_l = GRe_l \qquad \textit{for} \quad l = 1, 2 \ldots, s.$$

Then $(Ge_i)_{1\leq i\leq r}$ is a basis of the Krylov chain $(\mathcal{K}(i, A, \psi(A)e_1))_{1\leq i\leq r}$.

Proof. For $i = 1, 2, \ldots, \min\{s, k\}$ we have

$$\begin{aligned} \mathcal{K}(i, A, \psi(A)e_1) &= \psi(A)\,\mathrm{span}\{e_1, e_2, \ldots, e_i\} \\ &= GR\,\mathrm{span}\{e_1, e_2, \ldots, e_i\} \subseteq G\,\mathrm{span}\{e_1, e_2, \ldots, e_i\}. \end{aligned}$$

Using that $\dim \mathcal{K}(i, A, \psi(A)e_1) = i$ for $i \in \{1, 2 \ldots, r\}$, we obtain

$$\mathcal{K}(i, A, \psi(A)e_1) = G\,\mathrm{span}\{e_1, e_2, \ldots, e_i\}$$

which proves the lemma. ∎

Lemma 6.6 *Let $k \leq m \leq l \leq r \leq N$ be arbitrary integers. Let $C \in \mathrm{Mat}_{N\times l}$, $G \in \mathrm{Mat}_{l\times r}$ and $V \in \mathrm{Mat}_{N\times m}$ be matrices of full rank. Let G be structured as follows:*

$$G = [G_1, \star], \qquad \text{and} \qquad G_1^T = [G_{11}^T, 0]$$

with $G_1 \in \mathrm{Mat}_{l\times m}$, and $G_{11} \in \mathrm{Mat}_{m\times m}$.
Suppose that $CGe_l = Ve_l$ for $1 \leq l \leq k$.

a) Consider the partition $C = [C_1, C_2]$ with $C_1 \in \mathrm{Mat}_{N\times m}$. Then

$$C_1 G_{11} e_l = Ve_l \qquad \text{for } 1 \leq l \leq k.$$

b) Let the column vectors of C_1 and the column vectors of V be orthonormal (i.e. $V^*V = I$, $C_1^*C_1 = I$). Let $k = m$. Then G_1 and G_{11} have orthonormal columns.

Proof. a) is evident. Recall that $\langle x, y\rangle_* := x^*y$. Statement b) follows from the relations:

$$\langle e_i, e_j\rangle_* = \langle Ve_i, Ve_j\rangle_* = \langle C_1G_{11}e_i, C_1G_{11}e_j\rangle_* = \langle G_{11}e_i, C_1^*C_1G_{11}e_j\rangle_* = \langle G_{11}e_i, G_{11}e_j\rangle_*$$

for all $1 \leq i, j \leq m$. ∎

Theorem 6.2 (QR-Arnoldi) *Let the tuple $((A, V_{(i)}, \Pi_{(i)}, H_{(i)}, s_i, \psi_i))_{i\in\mathbb{N}_0}$ be generated by a restarted Arnoldi method. Let $(A, \tilde{V}_{(0)}, \tilde{\Pi}_{(0)}, \tilde{H}_{(0)}, m_0, \psi_0)$ be a maximal continuation of the Krylov tuple $(A, V_{(0)}, \Pi_{(0)}, H_{(0)}, s_0, \psi_0)$ generated by the Arnoldi subspace projection method. Let $(\hat{A}_i)_{i\in\mathbb{N}}$, $(\hat{G}_i)_{i\in\mathbb{N}}$, $(\hat{R}_i)_{i\in\mathbb{N}}$ be the matrices generated by the QR algorithm with shift polynomials ψ_i applied to the matrix $\hat{A}_1 := \hat{G}_0^{-1}A\hat{G}_0$, where $\hat{G}_0 = [\tilde{V}_{(0)}, \star]$. Then for each $i \in \mathbb{N}$ there is a diagonal matrix $D_i = diag\,(d_{i1}, d_{i2}, \ldots, d_{iN})$ with $|d_{ij}| = 1$ for all j such that the following equations hold for all $i \in \mathbb{N}_0$:*

$$D_i^{-1}\hat{A}_{i+1}D_i = \begin{bmatrix} H_{(i)} & \star \\ \star & \star \end{bmatrix} \qquad \text{and} \qquad \hat{G}_0\hat{G}_1\cdots\hat{G}_iD_ie_l = V_{(i)}e_l \quad \text{for} \quad 1 \leq l \leq s_i. \tag{6.13}$$

Proof. In the following the tuple $(A, \tilde{V}_{(i)}, \tilde{\Pi}_{(i)}, \tilde{H}_{(i)}, m_i)$ denotes a maximal continuation of $(A, V_{(i)}, \Pi_{(i)}, H_{(i)}, s_i)$ obtained by continuing the Arnoldi subspace projection method. Let $(A_i)_{i\in\mathbb{N}_0}$, $(G_i)_{i\in\mathbb{N}_0}$ and $(R_i)_{i\in\mathbb{N}}$ be the matrices associated with the restarted Krylov tuple $((A, \tilde{V}_{(i)}, \tilde{\Pi}_{(i)}, \tilde{H}_{(i)}, m_i, \psi_i))_{i\in\mathbb{N}_0}$ as stated in Theorem 6.1. We prove Theorem 6.2 by induction on i:
By definition the identities $\hat{A}_1 = \hat{G}_0^{-1}A\hat{G}_0$ and $\hat{G}_0e_l = V_{(0)}e_l$ hold for $l = 1, 2, \ldots, m_0$.
Assume the equations (6.13) hold for some $i \in \mathbb{N}_0$ and all $0 \leq j \leq i$.

Now we proceed to $i+1$:
By the induction hypothesis we have

$$G_{i+1}R_{i+1}e_l = \psi(A_{i+1})e_l = \psi(\hat{A}_{i+1})e_l = \hat{G}_{i+1}\hat{R}_{i+1}e_l \qquad \text{for } l = 1, 2, \ldots, m_i.$$

Thus by Lemma 6.5 the sequences $(G_{i+1}e_l)_{1\le l\le m_{i+1}}$ and $(\tilde{G}_{i+1}e_l)_{1\le l\le m_{i+1}}$ are bases of the Krylov chain $(\mathcal{K}(j, A_{i+1}, G_{i+1}e_1))_{1\le j\le m_{i+1}}$. Choose a matric $C_2 \in \mathrm{Mat}_{N\times(N-m_i)}$ such that $C := [\tilde{V}_{(i)}, C_2]$ is invertible. Put $G := G_{i+1}$ and $V := \tilde{V}_{i+1}$. By Theorem 6.1 the equation $CGe_l = Ve_l$ holds for $l = 1, 2, \ldots, m_{i+1}$. Lemma 6.6 b), applied to the matrices C, G and V, yields the orthonormality of the vectors

$$G_{i+1}e_1, G_{i+1}e_2, \ldots, G_{i+1}e_{m_{i+1}}.$$

Thus we have two orthonormal bases of the same Krylov chain. By Theorem 2.1 orthonormal bases of Krylov chains are unique up to scalar factors d_{ij} with $|d_{ij}| = 1$. ∎

Remark 6.2 Sorensen [70] developed and discussed special restarted Arnoldi methods in 1992. He used the connection between restarted Arnoldi methods and the QR algorithm described in Theorem 6.2.

6.3 Connections between the QR and the power method

The **power method** is defined by

(1) Choose a starting vector $v_1^{(0)} \in \mathbb{C}^N \setminus \{0\}$ with $||v_1^{(0)}|| = 1$.

(2) For $i = 0, 1, 2, \ldots$

$$\begin{aligned} H_{(i)} &:= (v_1^{(i)})^* A v_1^{(i)} \\ v_1^{(i+1)} &:= \frac{Av_1^{(i)}}{||Av_1^{(i)}||}. \end{aligned}$$

Setting

$$V_{(i)} := [v_1^{(i)}], \qquad \Pi_{(i)} := v_1^{(i)}(v_1^{(i)})^*$$

and

$$\psi_i(x) := \frac{x}{||Av_1^{(i)}||}, \tag{6.14}$$

one easily verifies that the power method and the restarted Arnoldi method generate identical restarted Krylov tuples if in the restarted Arnoldi process s_i is chosen equal to one and the shift polynomials ψ_i are chosen as described in (6.14). Hence Theorem 6.2 describes also the relations between the power method and the QR algorithm.

6.4 Connections between the LR and restarted look-ahead Lanczos algorithms

In this section we describe the connections between the restarted monomial look-ahead Lanczos method (see Section 3.2.2) and the LR algorithm with exact row-pivoting. As described in Section 5.2, a look-ahead Lancos method generates two Krylov tuples

$$(A, V, \Pi, H, s) \qquad \text{and} \qquad (A^+, W, \Pi^+, \hat{H}, s). \tag{6.15}$$

In view of (6.15) we call a restarted Krylov subspace projection method a restarted look-ahead Lanczos method if it generates two restarted Krylov tuples

$$((A, V_{(i)}, \Pi_{(i)}, H_{(i)}, s_i, \psi_i))_{i\in\mathbb{N}_0} \qquad \text{and} \qquad ((A^+, W_{(i)}, \Pi^+_{(i)}, \hat{H}_{(i)}, s_i, \hat{\psi}_i))_{i\in\mathbb{N}_0} \tag{6.16}$$

where for every i

$$(A, V_{(i)}, \Pi_{(i)}, H_{(i)}, s_i) \qquad \text{and} \qquad (A^+, W_{(i)}, \Pi^+_{(i)}, \hat{H}_{(i)}, s_i) \tag{6.17}$$

are generated by performing s_i steps of a look-ahead Lanczos method. Notice that $s_i \in \{\nu_i(k) \mid k \in K_i\}$ can be chosen arbitrarily. $(\nu_i(k))_{k\in K_i}$ denote the degree indices of $((\mathcal{K}(l, A^+, W_{(i)}e_1), \mathcal{K}(l, A, V_{(i)}e_1))_{1\le l\le N}$.

The LR **algorithm with exact row-pivoting** is given by

$$A_1 := A, \qquad \psi_{i-1}(A_i) = X_i^T L_i R_i, \qquad A_{i+1} := (X_i^T L_i)^{-1} A_i (X_i^T L_i).$$

L_i is an invertible lower triangular matrix, R_i is upper triangular and ψ_{i-1} is a polynomial. X_i is the permutation matrix used in the Gauss elimination process with 'exact pivoting' (i.e. two rows are interchanged if and only if the corresponding principal submatrix is singular [79]).

Definition 6.3 (Gragg matrix) *An unreduced upper Hessenberg matrix $H \in \mathrm{Mat}_N$ is called a Gragg matrix with indices $(\mu(k))_{0\le k\le b}$ if H has the form*

$$H = \begin{bmatrix} E_1 & C_1 & & 0 \\ B_1 & E_2 & \ddots & \\ & \ddots & \ddots & C_{b-1} \\ 0 & & B_{b-1} & E_b \end{bmatrix} \qquad \textit{where} \qquad E_k = \begin{bmatrix} 0 & \cdots & 0 & a_{k1} \\ \alpha_{k2} & \ddots & \vdots & a_{k2} \\ & \ddots & 0 & \vdots \\ 0 & & \alpha_{kl_k} & a_{kl_k} \end{bmatrix} \in \mathrm{Mat}_{l_k},$$

$l_k = \mu(k) - \mu(k-1)$, $\mu(0) := 0$ *and each of the non-diagonal blocks $B_k \in \mathrm{Mat}_{l_{k+1}\times l_k}$ and $C_k \in \mathrm{Mat}_{l_k\times l_{k+1}}$ has exactly one non-zero element in the upper right corner.*

Remark 6.3 The exact monomial look-ahead Lanczos method applied to an arbitrary triple (w_1, A, v_1), where w_1, v_1 denotes the starting vectors, generates a Gragg matrix H with indices $(\mu(k))_{0\leq k\leq b} = (\nu(k))_{0\leq k\leq b}$. Here $(\nu(k))_{k\in K}$ are the degree indices of the Krylov chain $((\mathcal{K}(j, A^T, w_1), \mathcal{K}(j, A, v_1)))_{1\leq j\leq N}$. Gragg [30, 31] was the first who described the exact monomial look-ahead Lanczos recursion and the structure of the generated matrix H.

The next Theorem shows that the LR algorithm with exact row-pivoting and the restarted exact monomial look-ahead Lanczos method (with a special choice of shift polynomials) generate identical sequences of matrices if some mild conditions are fulfilled.

Theorem 6.3 (LR and Lanczos) *Let $A \in \mathrm{Mat}_N$ be given. Choose arbitrary starting vectors $w_1^{(0)}, v_1^{(0)} \in \mathbb{C}^N \setminus \{0\}$. Choose shift polynomials*

$$\hat{\psi}_i(x) = 1 \qquad \text{for all } i \in \mathbb{N}_0 \tag{6.18}$$

$$\psi_i(x) = x - \gamma_i \qquad \text{for all } i \in \mathbb{N}_0 \tag{6.19}$$

where $\gamma_i \in \mathbb{C}$.
Let $(H_{(i)})_{i\in\mathbb{N}_0}$ denote the sequence of matrices generated by the restarted monomial look-ahead Lanczos method with shift polynomials $\hat{\psi}_i$ and ψ_i defined in (6.18) *and* (6.19)*, respectively. Let $(\mu_i(k))_{0\leq k\leq b_i}$ denote the indices of the Gragg matrix $H_{(i)}$ and let $H_{(i)}^{(k)}$ denote the principal $\mu_i(k) \times \mu_i(k)$ submatrix of $H_{(i)}$ (i.e. $H_{(i)}^{(k)} = (e_l^T H_{(i)} e_j)_{1\leq l,j\leq \mu_i(k)}$).*
Assume that $H_{(0)} \in \mathrm{Mat}_N$. Let $(A_i)_{i\in\mathbb{N}}$ be the sequence of matrices generated by the LR algorithm with exact pivoting applied to $A_1 := H_{(0)}$ using the shift polynomials ψ_i.
Suppose γ_i is not an eigenvalue of $H_{(i)}^{(k)}$ for all $k \in \{1, 2, \ldots, b_i\}$ and for all $i \in \mathbb{N}_0$. Then for each $i \in \mathbb{N}_0$ there is an invertible diagonal matrix O_i such that

$$A_{i+1} = O_i^{-1} H_{(i)} O_i. \tag{6.20}$$

The indices of the Gragg matrices A_{i+1} and $H_{(i)}$ are identical for every $i \in \mathbb{N}_0$. The indices $(\mu_i(k))_{0\leq k\leq b_i}$ of $H_{(i)}$ is given by

$$\{\mu_i(k) \mid 0 \leq k \leq b_i\} = \{\mu_{i-1}(k) \mid 0 \leq k \leq b_{i-1}\} \cup \{\mu_{i-1}(k) - 1 \mid 1 \leq k \leq b_{i-1}\} \tag{6.21}$$

for every $i \in \mathbb{N}$.

Remark 6.4 Theorem 6.3 implies that the restarted exact monomial look-ahead Lanczos method, using shifts polynomials as described in Theorem 6.3, generates after at most $N-1$ steps a tridiagonal matrix $H_{(N-1)}$ if the conditions of Theorem 6.3 are fulfilled in each step. The condition 'γ_i is not an eigenvalue of $H_{(i)}^{(k)}$' can be fulfilled by choosing $\gamma_i \in \mathbb{C}$ such that $|\gamma_i| > \|H_{(i)}^{(k)}\|$.

Remark 6.5 Theorem 6.3 implies that the ordinary LR algorithm without pivoting is equivalent to the restarted ordinary Lanczos method in the sense described in Theorem 6.3. This connection is known since the LR algorithm was developed by Rutishauser [63] in 1958.

In the proof of Theorem 6.3 and in the following Lemmas we use the permutation matrices X and Y defined as follows: Let A be a given Gragg matrix. Let $(\mu(k))_{0\le k\le b}$ denote the indices of A. Then we define the associated permutation matrices $X = \operatorname{diag}(X_1, X_2, \ldots, X_b)$ and $Y = \operatorname{diag}(Y_1, Y_2, \ldots, Y_b)$ by:

For $k = 1, 2, \ldots, b$

if $\mu(k) - \mu(k-1) \ge 2$ then set

$$X_k = \begin{bmatrix} 0 & 1 & & 0 \\ \vdots & \ddots & \ddots & \\ & & & \\ 0 & \ldots & 0 & 1 \\ 1 & 0 & \cdots & 0 \end{bmatrix} \in \mathrm{Mat}_{\mu(k)-\mu(k-1)}, \qquad Y_k = \begin{bmatrix} 0 & & 1 & 0 \\ & \iddots & 0 & \vdots \\ & & & \\ 1 & \iddots & \vdots & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix} \in \mathrm{Mat}_{\mu(k)-\mu(k-1)}$$

else put $X_k = Y_k = [1]$.

The following Lemmas are used in the proof of Theorem 6.3.

Lemma 6.7 *Let A be a Gragg matrix with indices $(\mu(k))_{0\le k\le b}$. Let $\hat{w}_1 := Xe_{\mu(1)}$ and $\hat{A} := XAX^T$. Then $\mathcal{K}(i, \hat{A}^T, \hat{w}_1) = Y\,\mathrm{span}\{e_1, e_2, \ldots, e_i\}$ for all $i \in \{1, 2, \ldots, N\}$.*

Inspection of the structure of $\hat{A}$ proves Lemma 6.7.

Lemma 6.8 *Let ψ be a non-constant polynomial. Let A be a Gragg matrix with indices $(\mu(k))_{0\le k\le b}$. Define $(w_1, A, v_1) := (e_{\mu(1)}, A, \psi(A)e_1)$. Suppose there is a LU decomposition of $X\psi(A)$. Then the degree indices $(\nu(k))_{k\in K}$ of the chain $(\mathcal{K}(j, A^T, w_1), \mathcal{K}(j, A, v_1))_{1\le j\le N}$ are:*

$$\{\nu(k) \mid k \in K\} = \Lambda := \{\mu(k) \mid 0 \le k \le b\} \cup \{\mu(k) - 1 \mid 1 \le k \le b\}$$

Proof. Define $(\hat{w}_1, \hat{A}, \hat{v}_1) := (Xw_1, XAX^T, Xv_1)$. Using Lemma 6.7 and the equation

$$\mathcal{K}(j, \hat{A}, \hat{v}_1) = X\psi(A)\mathrm{span}\{e_1, e_2, \ldots, e_j\} \qquad \text{for } j = 1, 2, \ldots, N$$

we obtain: If the matrix $\Gamma_l := (e_i^T X\psi(A)e_j)_{1\le i,j\le l}$ is invertible and $l \in \Lambda$, then the matrix $(w_1^T A^{i+j} v_1)_{0\le i,j\le l-1} = (\hat{w}_1^T \hat{A}^{i+j} \hat{v}_1)_{0\le i,j\le l-1}$ is invertible, too. By assumption the matrix Γ_l is invertible for $l = 1, 2, \ldots, N$. ∎

Lemma 6.9 *Let $A \in \mathrm{Mat}_N$ be a Gragg matrix with indices $(\mu(k))_{0\le k\le b}$. Let ψ be a polynomial of degree 1. Suppose there is an invertible lower triangular matrix L and an invertible upper triangular matrix R such that $X\psi(A) = LR$. Then the exact monomial look-ahead Lanczos method applied to the triple $(w_1, A, v_1) := (e_{\mu(1)}, A, \psi(A)e_1)$ terminates at step $s = N$ and generates a Krylov tuple (A, V, Π, H, N) with the following properties:*

a) There is an invertible diagonal matrix O such that $L = XVO$.

b) $O^{-1}HO = (X^TL)^{-1}A(X^TL)$.

c) The indices $(\tilde{\mu}_k)_{k\in L}$ of the Gragg matrix H are

$$\{\tilde{\mu}(k) : k \in L\} = \{\mu(k) \mid 0 \le k \le b\} \cup \{\mu(k) - 1 \mid 1 \le k \le b\}. \tag{6.22}$$

Proof. By Lemma 6.8 the exact monomial look-ahead Lanczos method applied to the triple $(w_1, A, v_1) := (e_{\mu(1)}, A, \psi(A)e_1)$ terminates at step $s = N$ and generates a Krylov tuple (A, V, Π, H, N) with the property c). The degree indices $(\nu(k)_{k\in K}$ of (w_1, A, v_1) are identical with the indices $(\tilde{\mu}(k)_{k\in L}$ of H. Define $\hat{w}_1 := Xw_1$, $\hat{v}_1 := Xv_1$ and $\hat{A} := XAX^T$. Consider the Krylov tuple $(\hat{A}, \hat{V}, \hat{\Pi}, \hat{H}, \hat{s})$ generated by the exact monomial look-ahead Lanczos method applied to the triple $(\hat{w}_1, \hat{A}, \hat{v}_1)$. By Lemma 5.8 and the uniqueness of the LU decomposition of $X\psi(A)$ it is sufficient to show that the matrix $\hat{V}$ is lower triangular. We denote the i-th column vector of $\hat{V}$ by $\hat{v}_i$. We recall that the regular vector $\hat{v}_{\nu(k)+1}$ is defined as $\hat{v}_{\nu(k)+1} := (I - \hat{\Pi}_{\nu(k)})\hat{A}\hat{v}_{\nu(k)}$ (up to a scalar factor). Here $\hat{\Pi}_{\nu(k)}$ is the projection on $\mathcal{K}(\nu(k), \hat{A}, \hat{v}_1)$ with kernel $\mathcal{K}(\nu(k), \hat{A}^T, \hat{w}_1)^\perp$. Using Lemma 6.7 we obtain the relation

$$\hat{v}_{\nu(k)+1} \in \mathrm{Im}\,(I - \hat{\Pi}_{\nu(k)}) \subseteq \mathrm{Ker}\,\hat{\Pi}_{\nu(k)} = \mathcal{K}(\nu(k), \hat{A}^T, \hat{w}_1)^\perp = \mathrm{span}\{e_1, e_2, \ldots, e_{\nu(k)}\}^\perp \tag{6.23}$$

for all $k \in K \setminus \{0\}$. First, we show that $\hat{v}_{\mu(k)} \in \mathrm{span}\{e_{\mu(k)}, e_{\mu(k+1)}\}$ and $\hat{v}_{\mu(k)+1} \in \mathrm{span}\{e_{\mu(k)+1}\}$. The upper Hessenberg form of A yields

$$\begin{aligned} \hat{v}_{\mu(k)} \in \mathcal{K}(\mu(k), \hat{A}, \hat{v}_1) &= X\psi(A)\mathcal{K}(\mu(k), A, e_1) \\ &= X\psi(A)\mathrm{span}\{e_1, e_2, \ldots, e_{\mu(k)}\} \\ &\subseteq X\mathrm{span}\{e_1, e_2, \ldots, e_{\mu(k)+1}\} \\ &= \mathrm{span}\{e_1, e_2, \ldots, e_{\mu(k)}\} + \mathrm{span}\{e_{\mu(k+1)}\}. \end{aligned} \tag{6.24}$$

Combining relations (6.23) and (6.24), we obtain $\hat{v}_{\mu(k)} \in \mathrm{span}\{e_{\mu(k)}, e_{\mu(k+1)}\}$. Hence we have

$$\hat{A}\hat{v}_{\mu(k)} \in \mathrm{span}\{\hat{A}e_{\mu(k)}, \hat{A}e_{\mu(k+1)}\} = \mathrm{span}\{e_{\mu(k-1)+1}, e_{\mu(k)+1}\}.$$

Hence relation (6.23) yields $\hat{v}_{\mu(k)+1} \in \mathrm{span}\{e_{\mu(k)+1}\}$. Finally, we conclude that

$$\hat{v}_{\mu(k)+j} \in \mathrm{span}\{\hat{A}^{j-1}\hat{v}_{\mu(k)+1}\} = \mathrm{span}\{e_{\mu(k)+j}\}$$

for $j = 1, 2, \ldots, \mu(k+1) - \mu(k) - 1$. Thus, the matrix $\hat{V}$ is lower triangular. ∎

Lemma 6.10 *Let $C \in \mathrm{Mat}_N$ be an arbitrary invertible matrix and let $A_i \in \mathrm{Mat}_N$ be a Gragg matrix. $(\mu(k))_{0\le k\le b}$ denotes the indices of A_i. Let $A_i^{(k)}$ denote the principal $\mu(k)\times\mu(k)$ submatrix of A_i (i.e. $A_i^{(k)} = (e_l^T A_i e_j)_{1\le l,j\le\mu(k)}$). Let $\psi_{i-1}(x) = x - \gamma_{i-1}$, $\gamma_{i-1} \in \mathbb{C}$. Suppose γ_{i-1} is not an eigenvalue of $A_i^{(k)}$ for all $k \in \{1,2,\ldots,b\}$. Let $H_{(i)}$ be the Gragg matrix generated by the exact monomial look-ahead Lanczos method applied to the triple $(w_1^{(i)}, \tilde{A}, v_1^{(i)}) := ((C^T)^{-1}e_{\mu(1)}, CA_iC^{-1}, C\psi_{i-1}(A_i)e_1)$. Then one step of the LR algorithm with exact pivoting applied to A_i generates a Gragg matrix A_{i+1} with the property:*

$$A_{i+1} = O_i^{-1} H_{(i)} O_i \tag{6.25}$$

for a suitably chosen invertible diagonal matrix O_i. The indices $(\tilde{\mu}(k)_{k\in L}$ of the Gragg matrix A_{i+1} are given by

$$\{\tilde{\mu}(k) \,|\, k \in L\} = \{\mu(k) \,|\, 0 \le k \le b\} \cup \{\mu(k) - 1 \,|\, 1 \le k \le b\}. \tag{6.26}$$

Proof. By Lemma 5.8, Lemma 6.9 and the structure of $\psi_{i-1}(A_i)$ it is sufficient to show that there is an invertible lower triangular matrix L and an invertible upper triangular matrix R such that $X\psi_{i-1}(A_i) = LR$. Define $\hat{A}_r := (e_l^T X\psi_{i-1}(A_i)e_j)_{1\le l,j\le r}$ for all $r \in \{1,2,\ldots,N\}$. The structure of $X\psi_{i-1}(A_i)$ yields for all $j \in \{2,3,\ldots,\mu(k+1)-\mu(k)\}$ the equation

$$\det(\hat{A}_{\mu(k)+j}) = \alpha_{k2}\alpha_{k3}\cdots\alpha_{kj}\det(\hat{A}_{\mu(k)})$$

where α_{kl} are the entries of the Gragg matrix A_i as described in Definition 6.3. Hence it remains to show that $\det(\hat{A}_{\mu(k)})$ does not vanish. The equation

$$\begin{aligned} |\det(\hat{A}_{\mu(k)})| &= |\det(\mathrm{diag}\,(X_1, X_2, \ldots, X_k)(\gamma_{i-1}I - A_i^{(k)}))| \\ &= |\det(\gamma_{i-1}I - A_i^{(k)})| \neq 0 \end{aligned}$$

yields the assertion. ∎

Now we are ready to prove Theorem 6.3.

Proof of Theorem 6.3.

Let $((A^T, W_{(i)}, \Pi_{(i)}^T, \hat{H}_{(i)}, m_i))_{i\in\mathbb{N}_0}$ and $((A, V_{(i)}, \Pi_{(i)}, H_{(i)}, m_i))_{i\in\mathbb{N}_0}$ be generated by the monomial look-ahead Lanczos method.
We prove Theorem 6.3 by induction on i.

Set $O_0 := I$. By assumption the equations $m_0 = N$ and $A_1 = O_0^{-1}H_{(0)}O_0$ hold.
Assume that the equation (6.20) holds for some $i \ge 0$.
Now we proceed to $i+1$:
Set $C := V_{(i)}O_i$. If the relations

$$(C^T)^{-1}e_{\mu(1)} = \alpha W_{(i+1)}e_1 \qquad \text{for some } \alpha \in \mathbb{C}\setminus\{0\} \tag{6.27}$$

$$A = CA_{i+1}C^{-1} \tag{6.28}$$

$$C\psi_i(A_{i+1})e_1 = \beta V_{(i+1)}e_1 \qquad \text{for some } \beta \in \mathbb{C}\setminus\{0\} \tag{6.29}$$

hold then Lemma 6.10 applies and yields the desired equations (6.20) and (6.21). Using $V_{(i+1)}e_1 = \psi_i(A)V_{(i)}e_1$, the induction hypothesis and the equation $A = V_{(i)}H_{(i)}V_{(i)}^{-1}$ we obtain the relations (6.28) and (6.29).

It remains to prove relation (6.27). By induction hypothesis and construction of the monomial look-ahead Lanczos vectors, the matrices $W_{(i)}$ and $V_{(i)}$ are invertible $N \times N$ matrices. Further, by Lemma 3.2, the matrix $D_{(i)} := W_{(i)}^T V_{(i)}$ is block diagonal

$$D_{(i)} = \operatorname{diag}(D_{(i),1}, D_{(i),2}, \ldots, D_{(i),b}) \tag{6.30}$$

and the diagonal blocks $D_{(i),j} \in \mathrm{Mat}_{\mu_i(j)-\mu_i(j-1)}$ have antitriangular structure

$$D_{(i),j} = \begin{bmatrix} & & & * \\ & 0 & \cdot & \\ & \cdot & & \\ & \cdot & * & \\ * & & & \end{bmatrix}.$$

Hence

$$D_{(i),j}^{-1} = \begin{bmatrix} & & & * \\ & * & \cdot & \\ & \cdot & & \\ & \cdot & 0 & \\ * & & & \end{bmatrix}. \tag{6.31}$$

Using $D_{(i)} = W_{(i)}^T V_{(i)}$, the relations (6.18), (6.30), (6.31) and the induction hypothesis, we obtain

$$(C^T)^{-1}e_{\mu(1)} = W_{(i)}D_{(i)}^{-1}e_{\mu(1)} = \alpha W_{(i)}e_1 = \alpha\hat{\psi}(A^T)W_{(i)}e_1 = \alpha W_{(i+1)}e_1$$

for some $\alpha \in \mathbb{C} \setminus \{0\}$. ∎

6.5 Convergence of restarted Krylov subspace projection methods

In this section we discuss the convergence properties of restarted Krylov subspace projection methods. With assistance of Theorem 6.1 we can carry over the convergence results for GR algorithms of Watkins and Elsner [76] to the situation of Krylov subspace projection methods.

The following convergence theorem has two parts. In the first part it is stated under which conditions the Krylov subspaces

$$\mathcal{S}_{i(k)} := \mathcal{K}(k, A, V_{(i)}e_1) = V_{(i)}\mathrm{span}\{e_1, \ldots, e_k\}$$

converge to an appropriate invariant subspace $\mathcal{H}$ of A. To estimate the convergence of subspaces we use the standard metric

$$d(\mathcal{X}, \mathcal{Y}) := \sup_{x \in \mathcal{X}, \|x\|=1} \inf_{y \in \mathcal{Y}} \|x - y\| \tag{6.32}$$

on the set of k-dimensional subspaces of $\mathbb{C}^N$.

In the second part it is stated that the convergence of the Krylov subspaces implies the convergence of the matrices $H_{(i)}$ to block triangular form if the norms of the projections $\Pi_{(i)}$ and the condition numbers of the matrices $V_{(i)}$ are bounded above. We recall that a Krylov tuple (A, V, Π, H, s) is called normalized if $\|Ve_j\| = 1$ for all $j \in \{1, 2, \ldots, s\}$.

Theorem 6.4 *Let $((A, V_{(i)}, \Pi_{(i)}, H_{(i)}, s_i, \psi_i))_{i \in \mathbb{N}_0}$ be a restarted normalized Krylov tuple. Let m be defined as $\dim \mathcal{K}(N, A, V_0 e_1)$. Assume there is a polynomial $\psi \in \mathrm{Pol}_N$ and a sequence of scalars $(\alpha_i)_{i \in \mathbb{N}_0} \subseteq \mathbb{C} \setminus \{0\}$ with the following properties a) and b):*

a) $\lim_{i \to \infty} \alpha_i \psi_i = \psi$ *(where* $\psi_i \in \mathrm{Pol}_N$ *for all* $i \in \mathbb{N}$ *)*

b) There is an integer k less than $\min_{i \in \mathbb{N}_0} s_i$ such that

$$|\psi(\lambda_k)| > |\psi(\lambda_{k+1})| \quad \text{and} \quad \psi_i(\lambda_j) \neq 0$$

for $j = 1, 2, \ldots, m$ and all $i \in \mathbb{N}_0$.
Here $\lambda_1, \lambda_2, \ldots, \lambda_m$ are the eigenvalues of the operator $A|_{\mathcal{K}(N, A, V_0 e_1)}$ ordered so that

$$|\psi(\lambda_1)| \geq |\psi(\lambda_2)| \geq \ldots \geq |\psi(\lambda_m)|.$$

Then for every ρ satisfying $|\psi(\lambda_{k+1})|/|\psi(\lambda_k)| < \rho < 1$ there is a constant L and an A-invariant subspace $\mathcal{H}$ of $\mathcal{K}(N, A, V_{(0)}e_1)$ such that

$$d(\mathcal{S}_{i(k)}, \mathcal{H}) \leq L\rho^i \qquad \text{for all} \quad i \in \mathbb{N}$$

where $\mathcal{S}_{i(k)} := \mathcal{K}(k, A, V_{(i)}e_1)$. The constant L depends on the matrix A and the starting vectors $V_{(i)}e_1$. It does not depend on the choice of the bases $(V_{(i)}e_j)_{1 \leq j \leq s_i}$ of the Krylov chain $(\mathcal{K}(j, A, V_{(i)}e_1))_{1 \leq j \leq s_i}$ nor on the choice of the projections $\Pi_{(i)}$.

If there are further constants $K_1 > 0$ and K_2 such that

$$\sigma_{\min}(V_{(i)}) \geq K_1 \qquad \text{and} \qquad \|\Pi_{(i)}\| \leq K_2 \qquad \text{for all } i \in \mathbb{N}_0 \tag{6.33}$$

then the sequence $(H_{(i)})_{i\in\mathbb{N}_0}$ *tends to block triangular form in the following sense: Consider the partition*

$$H_{(i)} = \begin{bmatrix} H_{11}^{(i)} & H_{12}^{(i)} \\ H_{21}^{(i)} & H_{22}^{(i)} \end{bmatrix}$$

where $H_{11}^{(i)} \in \mathrm{Mat}_k$ *and* $H_{22}^{(i)} \in \mathrm{Mat}_{s_i-k}$ *are unreduced upper Hessenberg matrices and the matrix* $H_{21}^{(i)} \in \mathrm{Mat}_{(s_i-k)\times k}$ *has at most one non-zero element in the upper right corner. Then*

$$\|H_{21}^{(i)}\| \le K_0 \rho^i \qquad \textit{for all} \quad i \in \mathbb{N}$$

where

$$K_0 := L\, 2\sqrt{2N}\, \|A\| \left(\frac{K_2}{K_1} + K_2 + 1\right).$$

In the proof of Theorem 6.4 we use Lemma 6.11 below.

Lemma 6.11 *Let* $(G_i)_{i\in\mathbb{N}_0}$ *and* $(A_i)_{i\in\mathbb{N}}$ *be the matrices generated by a GR algorithm. Let* $(\psi_i)_{i\in\mathbb{N}_0}$ *denote the shift polynomials used in the GR process. Let* $(\alpha_i)_{i\in\mathbb{N}_0} \subseteq \mathbb{C}\setminus\{0\}$ *be an arbitrary sequence. Then there is a GR algorithm that generates matrices* $(\check{G}_i)_{i\in\mathbb{N}_0}$, $(\check{A}_i)_{i\in\mathbb{N}}$ *and uses shift polynomials* $(\check{\psi}_i)_{i\in\mathbb{N}_0}$ *such that*

$$\check{G}_i = G_i, \qquad \check{A}_i = A_i \qquad \textit{and} \qquad \check{\psi}_i = \alpha_i \psi_i$$

for all i.

Proof. Using

$$\alpha_{i-1}\psi_{i-1}(A_i) = \alpha_{i-1} G_i R_i = G_i(\alpha_{i-1} R_i), \quad \text{and} \quad A_{i+1} = G_i^{-1} A_i G_i,$$

induction on i proves the lemma. ∎

Proof of Theorem 6.4. The assumption b) implies that ψ_i and the minimal polynomial of $A|_{\mathcal{K}(N,A,V_{(0)}e_1)}$ do not have a common divisor. Thus, by Lemma 6.4, the identity $m_i = m_0 =: m$ holds for all $i \in \mathbb{N}$. Recall that m_i denotes the dimension of $\mathcal{K}(N, A, V_{(i)}e_1)$. By Theorem 6.1 f) and Lemma 6.11 there is, for every $i \in \mathbb{N}$, a maximal normalized continuation $(A, \tilde{V}_{(i)}, \tilde{\Pi}_{(i)}, \tilde{H}_{(i)}, m)$ and a GR algorithm which generates the sequence $(\tilde{H}_{(i)})_{i\in\mathbb{N}}$, when applied to $\tilde{H}_{(0)}$ and the shift polynomials $\check{\psi}_i := \alpha_i\psi_i$ are used. Further by Theorem 6.1 e) the estimates (6.6) and (6.7) hold.
The convergence theorems for GR algorithms of Watkins and Elsner ([76], Theorem 6.2, Theorem 5.4), applied to this GR algorithm, yields the claim of Theorem 6.4 if the subspace condition

$$\mathrm{span}\{e_1, e_2, \ldots, e_k\} \cap \mathcal{U} = \{0\} \tag{6.34}$$

is fulfilled. Here $\mathcal{U}$ denotes the invariant subspace of $\tilde{H}_{(0)}$ associated with the eigenvalues

$$\lambda_{k+1}, \lambda_{k+2}, \ldots, \lambda_m.$$

Now we show that the subspace condition (6.34) is indeed satisfied.
The matrix $\tilde{H}_{(0)}$ is unreduced upper Hessenberg. Thus the equation

$$\mathcal{K}(k, \tilde{H}_{(0)}, e_1) = \operatorname{span}\{e_1, e_2, \ldots, e_k\}$$

holds.
Suppose there is an $x \in (\operatorname{span}\{e_1, e_2, \ldots, e_k\} \cap \mathcal{U}) \setminus \{0\}$.
The subspace $\mathcal{K}(N, A, x)$ is the smallest A-invariant subspace containing x. By assumption $\mathcal{U}$ is A-invariant and contains x. Consequently, the relation

$$\mathcal{K}(N, A, x) \subseteq \mathcal{U}$$

is valid. Using Lemma 5.14 we obtain the inequality

$$\dim \mathcal{U} \geq \dim \mathcal{K}(N, A, x) \geq m - k + 1.$$

This inequality contradicts the assumption $\dim \mathcal{U} = m - k$. Consequently the subspace condition (6.34) is fulfilled. ■

Example 6.1 (Restarted Arnoldi method) Let $((A, V_{(i)}, \Pi_{(i)}, H_{(i)}, s_i, \psi_i))_{i \in \mathbb{N}_0}$ be generated by the restarted Arnoldi method. Then the equations $||\Pi_{(i)}|| = 1$ and $\sigma_{min}(V_{(i)}) = 1$ hold for all $i \in \mathbb{N}_0$. Thus, the restarted Arnoldi method converges if the conditions a) and b) of Theorem 6.4 are satisfied.

Example 6.2 (power method) As described in Section 6.3, the power method is a special case of the restarted Arnoldi method. Thus the power method converges if the conditions a) and b) of Theorem 6.4 are fulfilled.
Set $\alpha_i := ||A v_1^{(i)}||$ and $\psi(x) := x$. Then the equations

$$\alpha_i \psi_i(x) = x = \psi(x) \qquad \text{for all } i \in \mathbb{N}_0$$

show that condition a) of Theorem 6.4 is fulfilled.

Example 6.3 (Restarted look-ahead Lanczos methods) A restarted practical look-ahead Lanczos method generates two restarted Krylov tuples

$$((A, V_{(i)}, \Pi_{(i)}, H_{(i)}, s_i, \psi_i))_{i \in \mathbb{N}_0} \qquad \text{and} \qquad ((A^+, W_{(i)}, \Pi^+_{(i)}, \hat{H}_{(i)}, s_i, \hat{\psi}_i))_{i \in \mathbb{N}_0}$$

such that for all $i \in \mathbb{N}_0$

$$\kappa_2(V_{(i)}) \leq K_0, \qquad \kappa_2(W_{(i)}) \leq K_0 \qquad \text{and} \qquad ||\Pi_{(i)}|| \leq K_0$$

where the constant K_0 can be chosen by the user (see Remark 5.7). Thus

- $(H_{(i)})_{i\in\mathbb{N}_0}$ tends to block triangular form if ψ_i and A satisfies the conditions a) and b) of Theorem 6.4.
- $(\hat{H}_{(i)})_{i\in\mathbb{N}_0}$ tends to block triangular form if $\hat{\psi}_i$ and A fulfil the conditions a) and b) of Theorem 6.4.

Choice of shifts: The polynomials ψ_i can be used to filter out the unwanted eigenvalues. If k eigenvalues (of a large matrix A) with specified properties are wanted then one can use the following strategy which was introduced by Sorensen [70] for the restarted Arnoldi method:

Proceed from step i to step $i+1$ as follows:

(1) Determine all eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_{s_i}$ of the matrix $H_{(i)} \in \mathrm{Mat}_{s_i}$.

(2) Choose $s_i - k$ unwanted eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_{s_i-k}$.
Set $\psi_i(x) := (x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_{s_i-k})$.

(3) Compute $V_{(i+1)}e_1 := V_{(i)}\psi_i(H_{(i)})e_1$.

In general the polynomials ψ_i constructed this way converge. In this case Theorem 6.4 yields linear convergence of the restarted Krylov subspace method if the conditions (6.33) are fulfilled. In practical applications the numbers s_i are chosen not much greater than k, for the following reasons:

a) *Storage requirements:* The matrix $V_{(i)} \in \mathrm{Mat}_{N\times s_i}$ and the matrix $H_{(i)} \in \mathrm{Mat}_{s_i\times s_i}$ are required for the computation of the next starting vector $V_{i+1}e_1$.

b) *Computational work:* The eigenvalues of the matrix $H_{(i)}$ have to be computed.

c) *Rounding Errors:* Rounding errors accumulate with increasing s_i.

If all eigenvalues of a (small) matrix A are wanted we use the following strategy:

a) Every Krylov tuple $(A, V_{(i)}, \Pi_{(i)}, H_{(i)}, m_i)$ is chosen maximal (i.e. $s_i = m_i := \dim \mathcal{K}(N, A, V_{(i)}e_1)$).

b) The polynomials ψ_i are chosen by the generalized Rayleigh-quotient shift strategy (known as a tool in the QR process): Choose $\psi_i = \gamma_i p_i$ where p_i is the characteristic polynomial of $H_{22}^{(i)}$, the trailing $(m_i - k) \times (m_i - k)$ submatrix of

$$H_{(i)} = \begin{bmatrix} H_{11}^{(i)} & H_{12}^{(i)} \\ H_{21}^{(i)} & H_{22}^{(i)} \end{bmatrix} \tag{6.35}$$

and $\gamma_i \in \mathbb{C}\setminus\{0\}$ can be chosen arbitrarily. For reasons of stability γ_i should be used to normalize the new starting vector $V_{(i+1)}e_1 = V_{(i)}\psi_i(H_{(i)})e_1$

Combining Theorem 6.1, Theorem 6.4 and the convergence theorems for GR algorithms of Watkins and Elsner ([76], Theorem 6.3, Theorem 6.5), we can establish quadratic and in some cases cubic convergence.

Theorem 6.5 (Quadratic and cubic convergence) *Let $A \in \mathrm{Mat}_N$ have distinct eigenvalues. Let $((A, V_{(i)}, \Pi_{(i)}, H_{(i)}, m_i, \psi_i))_{i\in\mathbb{N}_0}$ be generated by a restarted Krylov subspace projection method applied to A. Let the polynomials $\psi_i \in \mathrm{Pol}_{m_i-k}$ be chosen by the generalized Rayleigh-quotient shift strategy. Suppose there is a constant K_0 such that*

$$\kappa_2(V_{(i)}) \leq K_0 \qquad \text{for all } i \in \mathbb{N}. \tag{6.36}$$

If the sequence $(H_{(i)})_{i\in\mathbb{N}_0}$ tends to block triangular form, in the sense described in Theorem 6.4, then the convergence is quadratic. Furthermore, if there is a constant K_1 such that

$$\|H_{12}^{(i)}\| \leq K_1 \|H_{21}^{(i)}\| \qquad \text{for all} \quad i \in \mathbb{N} \tag{6.37}$$

then the convergence is cubic.

Remark 6.6 The proofs of the convergence theorems for GR algorithms ([76], Theorem 6.3, Theorem 6.5) are based on the convergence of the Krylov subspaces $\mathcal{S}_{i(k)} = \mathcal{K}(k, A, V_{(i)}e_1) = V_{(i)}\,\mathrm{span}\{e_1, \ldots, e_k\}$ to a suitable invariant subspace $\mathcal{H}$ of A and on versions of the Bauer-Fike theorem (see Lemma 5.13 and [76], Lemma 6.4).

Example 6.4 (Restarted Arnoldi method) Let $(V_{(i)})_{i\in\mathbb{N}_0}$ and $(H_{(i)})_{i\in\mathbb{N}_0}$ be the matrices generated by the restarted Arnoldi method using the Rayleigh-quotient shift strategy. Then $\kappa_2(V_{(i)}) = 1$ holds for all $i \in \mathbb{N}_0$. Thus, by Theorem 6.5, the restarted Arnoldi method converges quadratically if it converges. If further $A = A^*$ then the equation

$$\|H_{12}^{(i)}\| = \|H_{21}^{(i)}\|$$

holds for all $i \in \mathbb{N}$. Thus, by Theorem 6.5, the convergence is even cubic.

The following Theorem states that the convergence rates for the subspaces $\mathcal{S}_{i(k)} := \mathcal{K}(k, A, V_{(i)}e_1)$ and for the matrices $H_{21}^{(i)}$ are identical.

Theorem 6.6 *Let $A \in \mathrm{Mat}_N$ have distinct eigenvalues. Let the restarted Krylov tuple $((A, V_{(i)}, \Pi_{(i)}, H_{(i)}, m_i, \psi_i))_{i\in\mathbb{N}_0}$ be generated by a restarted Krylov subspace projection method using the Rayleigh-quotient shift strategy. Suppose there is a constant K_0 such that*

$$\kappa_2(V_{(i)}) \leq K_0 \qquad \text{for all } i \in \mathbb{N}.$$

a) Assume that the sequence of matrices $(H_{21}^{(i)})_{i\in\mathbb{N}}$ *tends to zero. Then there are constants* K_1 *and* K_2 *and an* A*-invariant subspace* $\mathcal{H}$ *such that*

$$\|H_{21}^{(i)}\| \le K_1\, d(\mathcal{S}_{i(k)}, \mathcal{H}) \le K_2 \|H_{21}^{(i)}\| \qquad \textit{for all} \quad i \in \mathbb{N}. \tag{6.38}$$

b) Assume that there is an A*-invariant subspace* $\mathcal{H}$ *such that*

$$d(\mathcal{S}_{i(k)}, \mathcal{H}) \to 0 \qquad \textit{for} \qquad i \to \infty$$

then the estimates (6.38) *hold.*

Theorem 6.6 can be proved in a similar manner as [76] Theorem 6.3.

6.6 Implicit GR algorithms and iterated Krylov methods

To save computational work, GR algorithms are usually implemented implicitly (see Watkins and Elsner [77]),

Implicit GR algorithm

(1) Choose an invertible matrix $G_0 \in \mathrm{Mat}_N$ such that $\tilde{A}_1 := G_0^{-1} A G_0$ is an upper Hessenberg matrix. Consider the partition

$$\tilde{A}_1 = \begin{bmatrix} A_{11}^{(1)} & & & \star \\ & A_{22}^{(1)} & & \\ & & \ddots & \\ 0 & & & A_{l_1 l_1}^{(1)} \end{bmatrix},$$

where $A_{jj}^{(1)}$ is unreduced upper Hessenberg for every $j \in \{1, 2, \ldots, l_1\}$.
Set $A_1 := A_{11}^{(1)}$.

(2) Construct two sequences of matrices $(A_i)_{i\in\mathbb{N}}$ and $(G_i)_{i\in\mathbb{N}}$ as follows:
Let A_i be an unreduced upper Hessenberg matrix.
Choose a (shift) polynomial ψ_{i-1}.
Compute an invertible matrix G_i such that $G_i e_1 = \psi_{i-1}(A_i) e_1$ and the matrix $\tilde{A}_{i+1} := G_i^{-1} A_i G_i$ is upper Hessenberg. Consider the decomposition

$$\tilde{A}_{i+1} = \begin{bmatrix} A_{11}^{(i+1)} & & & \star \\ & A_{22}^{(i+1)} & & \\ & & \ddots & \\ 0 & & & A_{l_{i+1} l_{i+1}}^{(i+1)} \end{bmatrix}$$

where $A_{jj}^{(i+1)}$ is unreduced upper Hessenberg for every $j \in \{1, 2, \ldots, l_{i+1}\}$.
Set $A_{i+1} := A_{11}^{(i+1)}$.

Remark 6.7 The eigenvalues of the matrices $A_{jj}^{(k)}$ are computed as described in (2) by setting $A_1 := A_{jj}^{(k)}$

To clarify the connection between restarted Krylov subspace projection methods and implicit GR algorithms we introduce 'iterated Krylov subspace projection methods' which are closely related to restarted Krylov subspace projection methods and implicit GR algorithms.

We recall that a Krylov tuple (A, V, Π, H, s) is called maximal if

$$s = m := \mathcal{K}(N, A, Ve_1).$$

Iterated Krylov subspace projection method

Let $A \in \mathrm{Mat}_N$ be given. Put $H_{(-1)} := A$.
Choose a starting vector $G_{(0)}e_1 \in \mathbb{C}^N \setminus \{0\}$.
Apply a Krylov subspace projection method to $H_{(-1)}$ and $G_{(0)}e_1$ to obtain a maximal Krylov tuple $(H_{(-1)}, G_{(0)}, \Pi_{(0)}, H_{(0)}, m_0)$.
Let a maximal Krylov tuple $(H_{(i-1)}, G_{(i)}, \Pi_{(i)}, H_{(i)}, m_i)$ be given.
Choose a shift polynomial ψ_i which approximates a divisor of the characteristic polynomial of $H_{(i)}$.
Next compute a new starting vector

$$G_{(i+1)}e_1 := \psi_i(H_{(i)})e_1.$$

Apply a Krylov subspace projection method to $H_{(i)}$ and $G_{(i+1)}e_1$ to obtain a new maximal Krylov tuple $(H_{(i)}, G_{(i+1)}, \Pi_{(i+1)}, H_{(i+1)}, m_{i+1})$.

For shortness of notation and clarity we introduce the notation of an iterated Krylov tuple.

Definition 6.4 (Iterated Krylov tuples) *Let $A \in \mathrm{Mat}_N$ and $G_{(0)}e_1 \in \mathbb{C}^N \setminus \{0\}$ be given. Put $H_{(-1)} := A$. A sequence of six-tuples $((H_{(i-1)}, G_{(i)}, \Pi_{(i)}, H_{(i)}, m_i, \psi_i))_{i \in \mathbb{N}_0}$ is called an iterated Krylov tuple if for every $i \in \mathbb{N}_0$ the five-tuple $(H_{(i-1)}, G_{(i)}, \Pi_{(i)}, H_{(i)}, m_i)$ is a maximal Krylov tuple, ψ_i is a polynomial and $G_{(i+1)}e_1 = \psi_i(H_{(i)})e_1$.*

Remark 6.8 Let $G_{(0)}e_1 \in \mathbb{C}^N \setminus \{0\}$ and $A \in \mathrm{Mat}_N$ be given. Then an iterated Krylov subspace projection method is a method which generates matrices $G_{(i)}$ and $H_{(i)}$ such that $((H_{(i-1)}, G_{(i)}, \Pi_{(i)}, H_{(i)}, m_i, \psi_i))_{i \in \mathbb{N}_0}$ is an iterated Krylov tuple for an appropriate choice of $\Pi_{(i)}$ and ψ_i.

The following theorem shows that iterated Krylov subspace projection methods and restarted Krylov subspace projection methods generate identical matrices $H_{(i)}$ if s_i is chosen maximal (i.e. $s_i = m_i$).

Theorem 6.7 (Iterated and restarted Krylov methods)

a) Let $((H_{(i-1)}, G_{(i)}, \hat{\Pi}_{(i)}, H_{(i)}, m_i, \psi_i))_{i\in\mathbb{N}_0}$ *be generated by an iterated Krylov subspace projection method applied to* A *and* $G_{(0)}e_1$.
For each $i \in \mathbb{N}_0$ *define* $V_{(i)} := G_{(0)}G_{(1)}\cdots G_{(i)}$ *and choose an arbitrary projection* $\Pi_{(i)}$ *on* $\mathcal{K}(N, A, G_{(0)}e_1)$. *Then there is a restarted Krylov subspace projection method which generates the tuple* $((A, V_{(i)}, \Pi_{(i)}, H_{(i)}, m_i, \psi_i))_{i\in\mathbb{N}_0}$ *when applied to* A *and* $V_{(0)}e_1 = G_{(0)}e_1$.

b) Conversely, let $((A, V_{(i)}, \Pi_{(i)}, H_{(i)}, m_i, \psi_i))_{i\in\mathbb{N}_0}$ *be generated by a restarted Krylov subspace projection method applied to* A *and* $V_{(0)}e_1$.
Set $H_{(-1)} := A$ *and* $m_{-1} := m_0$. *For each* $i \in \mathbb{N}_0$ *define* $G_{(i)} \in \mathrm{Mat}_{m_{i-1}\times m_i}$ *by the equation* $V_{(i)} = G_{(0)}G_{(1)}\cdots G_{(i)}$ *and choose an arbitrary projection* $\hat{\Pi}_{(i)}$ *on* $\mathcal{K}(N, H_{(i-1)}, G_{(i)}e_1)$. *Then there is an iterated Krylov subspace projection method which generates the tuple* $((H_{(i-1)}, G_{(i)}, \hat{\Pi}_{(i)}, H_{(i)}, m_i, \psi_i))_{i\in\mathbb{N}_0}$ *when applied to* A *and* $G_{(0)}e_1 = V_{(0)}e_1$.

Now we state the connections between iterated Krylov subspace projection methods and implicit GR algorithms.

Theorem 6.8 (Iterated Krylov methods and implicit GR algorithms)

a) Let $((H_{(i-1)}, G_{(i)}, \hat{\Pi}_{(i)}, H_{(i)}, m_i, \psi_i))_{i\in\mathbb{N}_0}$ *be generated by an iterated Krylov subspace projection method. Then there is an implicit GR algorithm which generates a sequence of invertible matrices* $(\tilde{G}_i)_{i\in\mathbb{N}_0}$ *and a sequence of upper Hessenberg matrices* $(\tilde{A}_i)_{i\in\mathbb{N}}$ *such that*

$$G_{(i+1)}e_1 = \psi_i(H_{(i)})e_1, \qquad \tilde{G}_i = [G_{(i)}, \star], \tag{6.39}$$

$$\tilde{A}_{i+1} = \tilde{G}_i^{-1}H_{(i-1)}\tilde{G}_i, \qquad \text{and} \qquad \tilde{A}_{i+1} = \begin{bmatrix} H_{(i)} & \star \\ 0 & \star \end{bmatrix}. \tag{6.40}$$

b) Let $(\tilde{G}_i)_{i\in\mathbb{N}}$ *and* $(\tilde{A}_i)_{i\in\mathbb{N}}$ *be the sequences of matrices generated by an implicit GR algorithm. Let* ψ_i *be the shift polynomials used in this implicit GR algorithm. Then there is an iterated Krylov subspace projection method which generates a tuple* $((H_{(i-1)}, G_{(i)}, \tilde{\Pi}_{(i)}, H_{(i)}, m_i, \psi_i))_{i\in\mathbb{N}_0}$ *such that* $\tilde{G}_i = [G_{(i)}, \star]$ *and* $A_{11}^{(i+1)} = H_{(i)}$.

Theorem 6.7 and Theorem 6.8 follow immediately from Theorem 6.1 and Lemma 6.5.

6.6.1 Convergence theorems

Using Theorem 6.7 and Theorem 6.8 the convergence results for restarted Krylov subspace projection methods (Theorem 6.4 and Theorem 6.5) can be easily carried over to the situation of implicit QR algorithms and iterated Krylov subspace projection methods, respectively. Theorem 6.9 and Theorem 6.10 below are formulated for iterated Krylov subspace projection methods. Using the equations (6.39) and (6.40) these convergence theorems can be immediately reformulated for implicit GR algorithms.

Theorem 6.9 *Let $((H_{(i-1)}, G_{(i)}, \Pi_{(i)}, H_{(i)}, m_i, \psi_i))_{i\in\mathbb{N}_0}$ be generated by an iterated Krylov subspace projection method applied to a matrix $A \in \mathrm{Mat}_N$. Let m denote the order of $H_{(0)}$ (i.e. $H_{(0)} \in \mathrm{Mat}_m$). Assume there is a polynomial $\psi \in \mathrm{Pol}_N$ and a sequence of scalars $(\alpha_i)_{i\in\mathbb{N}_0} \subseteq \mathbb{C} \setminus \{0\}$ with the following properties a) and b):*

a) $\lim_{i\to\infty} \alpha_i\psi_i = \psi$ *(where* $\psi_i \in \mathrm{Pol}_N$ *for all* $i \in \mathbb{N}$*)*

b) There is an integer k less than m such that

$$|\psi(\lambda_k)| > |\psi(\lambda_{k+1})| \quad \text{and} \quad \psi_i(\lambda_j) \neq 0$$

for $j = 1, 2, \ldots, m$ and all $i \in \mathbb{N}_0$.
Here $\lambda_1, \lambda_2, \ldots, \lambda_m$ are the eigenvalues of the operator $H_{(0)} = A|_{\mathcal{K}(N,A,G_{(0)}e_1)}$ ordered so that

$$|\psi(\lambda_1)| \geq |\psi(\lambda_2)| \geq \ldots \geq |\psi(\lambda_m)|.$$

Then for every ρ satisfying $|\psi(\lambda_{k+1})|/|\psi(\lambda_k)| < \rho < 1$ there is a constant L and an A-invariant subspace $\mathcal{H}$ of $\mathcal{K}(N, A, G_{(0)}e_1)$ such that

$$d(\mathcal{S}_{i(k)}, \mathcal{H}) \leq L\rho^i \qquad \text{for all} \quad i \in \mathbb{N}$$

where $\mathcal{S}_{i(k)} := \mathcal{K}(k, A, G_{(0)}e_1)$. The constant L depends on the matrix A and the starting vectors $G_{(i)}e_1$.

Furthermore, if there is a constant K_0 such that

$$\kappa_2(G_{(0)}G_{(1)}\cdots G_{(i)}) \leq K_0 \qquad \text{for all } i \in \mathbb{N}_0 \tag{6.41}$$

then the sequence $(H_{(i)})_{i\in\mathbb{N}_0}$ tends to block triangular form in the following sense:
Consider the partition

$$H_{(i)} = \begin{bmatrix} H_{11}^{(i)} & H_{12}^{(i)} \\ H_{21}^{(i)} & H_{22}^{(i)} \end{bmatrix}$$

where $H_{11}^{(i)} \in \mathrm{Mat}_k$ and $H_{22}^{(i)} \in \mathrm{Mat}_{m-k}$ are unreduced upper Hessenberg matrices and the matrix $H_{21}^{(i)} \in \mathrm{Mat}_{(m-k)\times k}$ has exactly one non-zero element in the upper right corner.

Then

$$||H_{21}^{(i)}|| \leq K_1 \rho^i \qquad \textit{for all} \quad i \in \mathbb{N}$$

where

$$K_1 = 2\sqrt{2}\,||A||\,L\,K_0.$$

Theorem 6.10 (Quadratic and cubic convergence) *Let $A \in \mathrm{Mat}_N$ have distinct eigenvalues. Let $((H_{(i-1)}, G_{(i)}, \Pi_{(i)}, H_{(i)}, m_i, \psi_i))_{i\in\mathbb{N}_0}$ be generated by an iterated Krylov subspace projection method applied to A. Let the polynomials $\psi_i \in \mathrm{Pol}_{m_i-k}$ be chosen by the generalized Rayleigh-quotient shift strategy. Suppose there is a constant K such that*

$$\kappa_2(G_{(0)}G_{(1)}\cdots G_{(i)}) \leq K \qquad \textit{for all } i \in \mathbb{N}_0. \tag{6.42}$$

If the sequence $(H_{(i)})_{i\in\mathbb{N}_0}$ tends to block triangular form, in the sense described in Theorem 6.9, then the convergence is quadratic. Furthermore, if there is a constant K_1 such that

$$||H_{12}^{(i)}|| \leq K_1 ||H_{21}^{(i)}|| \qquad \textit{for all} \quad i \in \mathbb{N}, \tag{6.43}$$

then the convergence is cubic.

Remark 6.9 In Theorems 6.7, 6.8, 6.9 and 6.10 the maximality of the Krylov tuples, generated by the iterated Krylov subspace projection method, is necessary. Indeed, let $((H_{(i-1)}, G_{(i)}, \hat{\Pi}_{(i)}, H_{(i)}, s_i, \psi_i))_{i\in\mathbb{N}_0}$ be generated by an iterated Krylov subspace projection method applied to a matrix A. Suppose that there is an $i_0 \in \mathbb{N}_0$ such that $s_{i_0} < m_i$ and $s_i = m_i := \dim \mathcal{K}(N, H_{(i-1)}, G_{(i)}e_1)$ for $i \neq i_0$. Then the characteristic polynomial of $H_{(i)}$ is a divisor of the characteristic polynomial of $H_{(i_0)}$ for every $i \geq i_0$. In general the eigenvalues of $H_{(i_0)}$ are not eigenvalues of the starting matrix $H_{(-1)} = A$. This is demonstrated by the following example:

Choose

$$A = \begin{bmatrix} 2 & 3 \\ 3 & 0 \end{bmatrix}, \qquad G_{(0)}e_1 = e_1, \qquad i_0 = 1$$

and shift poynomials $\psi_i(x) = \alpha_i x$. Here α_i is chosen such that $||G_{(i+1)}e_1|| = 1$. The Arnoldi method applied to $H_{(-1)} := A$ and $G_{(0)}e_1 = e_1$ yields $(H_{(-1)}, G_{(0)}, \Pi_{(0)}, H_{(0)}, m_0)$ where

$$m_0 = 2, \qquad G_{(0)} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad H_{(0)} = A.$$

Set

$$G_{(1)}e_1 := \psi_0(H_{(0)})e_1 = \frac{1}{\sqrt{13}} \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

Then one step of the Arnoldi method applied to $H_{(0)}$ and $G_{(1)}e_1$ generates the Krylov tuple $(H_{(0)}, G_{(1)}, \Pi_{(1)}, H_{(1)}, s_1)$ where

$$s_1 = 1 \neq 2 = m_1, \qquad G_{(1)} = G_{(1)}e_1 \qquad \text{and} \qquad H_{(1)} = [\frac{44}{13}].$$

Continuing the iterated Arnoldi method we obtain for all $i \geq 2$ the Krylov tuple

$$(H_{(i-1)}, G_{(i)}, \Pi_{(i)}, H_{(i)}, m_i) = ([\frac{44}{13}], [1], [1], [\frac{44}{13}], 1).$$

The number $\frac{44}{13}$ is not an eigenvalue of A.

7 Restarted practical look-ahead Lanczos methods

In this section we discuss the convergence properties of restarted practical look-ahead Lanczos methods.

First we recall (see Example 6.3) that a restarted practical look-ahead Lanczos method generates two restarted Krylov tuples

$$((A, V_{(i)}, \Pi_{(i)}, H_{(i)}, s_i, \psi_i))_{i\in\mathbb{N}_0} \qquad \text{and} \qquad ((A^+, W_{(i)}, \Pi^+_{(i)}, \hat{H}_{(i)}, s_i, \hat{\psi}_i))_{i\in\mathbb{N}_0} \tag{7.1}$$

such that

$$\kappa_2(V_{(i)}) \le K_0, \qquad \kappa_2(W_{(i)}) \le K_0 \qquad \text{and} \qquad \|\Pi_{(i)}\| \le K_0 \qquad \text{for all } i \in \mathbb{N}_0 \tag{7.2}$$

where the constant K_0, the indices $s_i \in \{\nu_i(k) \mid k \in K_i\}$ and the shift polynomials ψ_i and $\hat{\psi}_i$ can be chosen by the user.

By Theorem 6.4 the following statements hold:

- If A and ψ_i fulfil the conditions a) and b) of Theorem 6.4, then the sequence $(H_{(i)})_{i\in\mathbb{N}_0}$ tends to block triangular form at least linearly.
- If A and $\hat{\psi}_i$ fulfil the conditions a) and b) of Theorem 6.4, then the sequence $(\hat{H}_{(i)})_{i\in\mathbb{N}_0}$ tends to block triangular form at least linearly.

We say that a restarted look-ahead Lanczos method converges if the sequence $(H_{(i)})_{i\in\mathbb{N}_0}$ tends to block triangular form. In the following we consider the case where the practical look-ahead Lanczos process is carried out up to termination. In other words, s_i is chosen such that

$$s_i = r_i := \max\{\nu(k) \mid k \in K_i\}$$

for all $i \in \mathbb{N}_0$ (see Remark 5.4).
We recall that the only possibility for a breakdown of a look-ahead Lanczos method is the 'incurable breakdown' (see Remark 5.4). If no such (very unlikely) breakdown occurs then the following equation holds:

$$r_i = \min\{m_i, \hat{m}_i\} \tag{7.3}$$

where

$$m_i := \dim \mathcal{K}(N, A, V_{(i)}e_1) \qquad \text{and} \qquad \hat{m}_i := \dim \mathcal{K}(N, A^+, W_{(i)}e_1).$$

We say that a restarted (practical) look-ahead Lanczos method does not break down if the equation

$$m_i = r_i \qquad \text{for all} \quad i \in \mathbb{N}_0 \tag{7.4}$$

or the equation

$$\hat{m}_i = r_i \qquad \text{for all} \quad i \in \mathbb{N}_0 \tag{7.5}$$

holds.

Applying Theorem 6.5 to the restarted practical look-ahead Lanczos methods we obtain the following convergence theorem.

Theorem 7.1 *Let $A \in \mathrm{Mat}_N$ have distinct eigenvalues. Suppose that the restarted practical look-ahead Lanczos method does not break down.*

a) Let the shift polynomials ψ_i be chosen by the generalized Rayleigh-quotient shift strategy. If $r_0 = m_0$ and the sequence $(H_{(i)})_{i\in\mathbb{N}_0}$ tends to block triangular form, in the sense described in Theorem 6.4, then convergence is quadratic. Furthermore, if there is a a constant K_1 such that

$$\|H_{12}^{(i)}\| \leq K_1 \, \|H_{21}^{(i)}\| \tag{7.6}$$

then convergence is cubic.

b) Let the shift polynomials $\hat{\psi}_i$ be chosen by the generalized Rayleigh-quotient shift strategy. If $r_0 = \hat{m}_0$ and the sequence $(\hat{H}_{(i)})_{i\in\mathbb{N}_0}$ tends to block triangular form then convergence is quadratic. Furthermore, if there is a constant $\hat{K}_1$ such that

$$\|\hat{H}_{12}^{(i)}\| \leq \hat{K}_1 \, \|\hat{H}_{21}^{(i)}\| \tag{7.7}$$

then the convergence is cubic.

Lemma 7.1 below states a sufficient condition for the restarted look-ahead Lanczos methods not to break down.

Lemma 7.1 *Let $\lambda_1, \lambda_2, \ldots, \lambda_N$ denote the eigenvalues of A. Assume that*

$$\psi_i(\lambda_j) \neq 0 \qquad \textit{and} \qquad \hat{\psi}_i(\lambda_j) \neq 0 \tag{7.8}$$

for all $i \in \mathbb{N}_0$ and all j.
Then the equations

$$m_i = m_0, \qquad \hat{m}_i = \hat{m}_0 \qquad \textit{and} \qquad r_i = r_0$$

hold for all $i \in \mathbb{N}_0$.
If further $r_0 = \min\{m_0, \hat{m}_0\}$, then the restarted look-ahead Lanczos method does not break down.

Proof. Using Lemma 6.1 a) and Lemma 6.4 we obtain the equations

$$\mathcal{K}(N, A, V_{(i)}e_1) = \mathcal{K}(N, A, V_{(0)}e_1) \quad \text{and} \quad \mathcal{K}(N, A^+, W_{(i)}e_1) = \mathcal{K}(N, A^+, W_{(0)}e_1). \tag{7.9}$$

Hence the equations $m_i = m_0$ and $\hat{m}_i = \hat{m}_0$ hold. Finally Theorem 5.4 yields the equation $r_i = r_0$. ■

Lemma 7.1 combined with Theorem 7.1 shows that restarted look-ahead Lanczos methods typically converge at least quadratic.

Example 7.1 (Quadratic convergence) By Theorem 6.3 the restarted monomial look-ahead Lanczos method with shift polynomials of the form

$$\begin{aligned} \psi_i(x) &= x - \gamma_i \qquad (\text{where } \gamma_i \in \mathbb{C}) \\ \hat{\psi}_i(x) &= 1 \end{aligned} \tag{7.10}$$

generates under mild conditions matrices $H_{(i)}$ which are diagonally similar to the matrices A_{i+1} generated by the LR algorithm with exact pivoting. The LR algorithm with exact pivoting, using the generalized Rayleigh-quotient shift strategy, converges typically quadratically and not cubically, if it converges at all (see [79]). Thus, the monomial look-ahead Lanczos method converges typically quadratically and not cubically if the shift polynomials ψ_i are chosen by the generalized Rayleigh quotient-shift strategy and the shift polynomials $\hat{\psi}_i$ are chosen as described in (7.10).

7.1 Cubic convergence

In the following we discuss how to choose the shift polynomials $\hat{\psi}_i$ and the first starting vectors $V_{(0)}e_1$ and $W_{(0)}e_1$ so that the inequalities (7.6) and (7.7) hold and the restarted look-ahead Lanczos methods do not break down. In this situation Theorem 7.1 applies and yields cubic convergence.

Definition 7.1 *A restarted look-ahead Lanczos method is called identically restarted if the shift polynomials ψ_i and $\hat{\psi}_i$ are chosen such that there is a scalar $\gamma_i \in \mathbb{C} \setminus \{0\}$ satisfying the equation*

$$\gamma_i \, \psi_i = \hat{\psi}_i$$

for all $i \in \mathbb{N}_0$.

Let $A \in \mathrm{Mat}_N$ have distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_N$ with eigenvectors $x_1, x_2, \ldots, x_N$. Let $\langle \cdot, \cdot \rangle_S$ be a non-degenerate bilinear form. Let $y_1, y_2, \ldots, y_N$ be the eigenvectors of A^+. Let two vectors $v \in \mathbb{C}^N$ and $w \in \mathbb{C}^N$ be given. Let α_i and β_i be the coefficients in the expansions

$$v = \sum_{i=1}^{N} \alpha_i x_i \qquad \text{and} \qquad w = \sum_{i=1}^{N} \beta_i y_i.$$

Definition 7.2 *The vectors v and w have 'dual directions' (with respect to A and $\langle\cdot,\cdot\rangle_S$) if the equivalence*

$$\alpha_i = 0 \Longleftrightarrow \beta_i = 0$$

holds for all $i \in \{1, 2, \ldots, N\}$.

Definition 7.3 *The vectors v and w have 'general directions' (with respect to A and $\langle\cdot,\cdot\rangle_S$) if $\alpha_i \neq 0$ and $\beta_i \neq 0$ for all $i \in \{1, 2, \ldots, N\}$.*

Notice that the pairs of vectors (w, v) having general directions (with respect to A and $\langle\cdot,\cdot\rangle_S$) form an open dense set in $\mathbb{C}^N \times \mathbb{C}^N$. Notice further that every pair of vectors having general directions also have dual directions.

Let $\mathcal{W}$ be a subspace of $\mathbb{C}^N$. Throughout this section we use the notation

$$\mathcal{K}_{(i)k} := \mathcal{K}(k, A, V_{(i)}e_1), \qquad \hat{\mathcal{K}}_{(i)k} := \mathcal{K}(k, A^+, W_{(i)}e_1)$$

and

$$\overline{W} := \{\overline{w} \,|\, w \in \mathcal{W}\}$$

where $\overline{w}$ is the component-wise complex conjugate of $w \in \mathbb{C}^N$. Notice that $\overline{W}$ is a subspace of $\mathbb{C}^N$.

For later use we recall the definition of the condition number of an eigenvalue (see e.g. [28]). Let $A \in \mathrm{Mat}_N$ have distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_N$ with eigenvectors $x_1, x_2, \ldots, x_N$. Let $y_1, y_2, \ldots, y_N$ be the eigenvectors of A^*. The number

$$cond(\lambda_i) := \frac{||x_i||\ ||y_i||}{|x_i^* y_i|} \tag{7.11}$$

is called the condition number of λ_i.

The following Theorem 7.2 states that every identically restarted practical look-ahead Lanczos method does not break down and that convergence is cubic, if it converges at all and if A has distinct eigenvalues, the first starting vectors $V_{(0)}e_1$ and $W_{(0)}e_1$ have dual directions and ψ_i is chosen by the generalized Rayleigh-quotient shift strategy.

Theorem 7.2 (Cubic convergence of the identically restarted look-ahead Lanczos method) *Let $A \in \mathrm{Mat}_N$ have distinct eigenvalues. Choose an arbitrary non-degenerate bilinear form $\langle x, y\rangle_S = x^T S y$ on $\mathbb{C}^N \times \mathbb{C}^N$. Choose the first starting vectors $V_{(0)}e_1$ and $W_{(0)}e_1$ such that they have dual directions (with respect to A and $\langle\cdot,\cdot\rangle_S$). Then every identically restarted practical look-ahead Lanczos method does not break down and generates two sequences of Krylov tuples*

$$((A, V_{(i)}, \Pi_{(i)}, H_{(i)}, m_i, \psi_i))_{i\in\mathbb{N}_0} \quad \textit{and} \quad ((A^+, W_{(i)}, \Pi^+_{(i)}, \hat{H}_{(i)}, m_i, \hat{\psi}_i))_{i\in\mathbb{N}_0}.$$

All eigenvalues of the matrix $H_{(i)}$ are eigenvalues of the matrix A. The matrices $H_{(i)}$ and $\hat{H}_{(i)}$ are similar.
Further the identity

$$m_i := \dim \mathcal{K}(N, A, V_{(i)}e_1) = \dim \mathcal{K}(N, A^+, W_{(i)}e_1)$$

holds for all $i \in \mathbb{N}$.
Suppose that the shift polynomials $\psi_i \in \mathrm{Pol}_{m_i-k}$ are chosen by the generalized Rayleigh-quotient shift strategy ($k \in \{1, \ldots, m_0 - 1\}$ can be chosen arbitrarily). Consider the partitions

$$H_{(i)} = \begin{bmatrix} H_{11}^{(i)} & H_{12}^{(i)} \\ H_{21}^{(i)} & H_{22}^{(i)} \end{bmatrix}, \qquad \hat{H}_{(i)} = \begin{bmatrix} \hat{H}_{11}^{(i)} & \hat{H}_{12}^{(i)} \\ \hat{H}_{21}^{(i)} & \hat{H}_{22}^{(i)} \end{bmatrix}$$

where $H_{11}^{(i)}$, $\hat{H}_{11}^{(i)} \in \mathrm{Mat}_k$ and $H_{22}^{(i)}$, $\hat{H}_{22}^{(i)} \in \mathrm{Mat}_{m_i-k}$. Suppose that the sequence $(H_{21}^{(i)})_{i\in\mathbb{N}_0}$ tends to zero. Then the following statements hold:

a) **(Convergence of the subspaces)**
There is a k-dimensional A-invariant subspace $\mathcal{H}$ and an k-dimensional A^+-invariant subspace $\hat{\mathcal{H}}$ such that

$$\begin{aligned} d(\mathcal{K}_{(i)k}, \mathcal{H}) &\to 0 \qquad \text{for} \quad i \to \infty \\ d(\hat{\mathcal{K}}_{(i)k}, \hat{\mathcal{H}}) &\leq C\, d(\mathcal{K}_{(i)k}, \mathcal{H}) \qquad \text{for all} \quad i \in \mathbb{N}_0 \end{aligned}$$

where the constant C depends on the matrix A, the first starting vectors $V_{(0)}e_1$ and $W_{(0)}e_1$ and on the form $\langle x, y\rangle_S = x^T S y$.

b) The sequence $(\hat{H}_{21}^{(i)})_{i\in\mathbb{N}_0}$ tends to zero.

c) **(Existence of regular vectors in the $k+1$-st step)**
There is an $i_0 \in \mathbb{N}$ such that for all $i \geq i_0$ the vectors $V_{(i)}e_{k+1}$ and $W_{(i)}e_{k+1}$ can be chosen as regular Lanczos vectors. In other words: The Lanczos projection $\Pi_{k,(i)}$ ($\mathrm{Im}\,\Pi_{k,(i)} = \mathcal{K}_{(i)k}$ and $\mathrm{Ker}\,\Pi_{k,(i)} = (\hat{\mathcal{K}}_{(i)k})^{\perp_S}$) exists for all $i \geq i_0$. For the norm of the projection $\Pi_{k,(i)}$ the estimates

$$||\Pi_{k,(i)}|| \leq \frac{1}{\sqrt{1 - d(\mathcal{K}_{(i)k}, \overline{S^T\hat{\mathcal{K}}_{(i)k}})^2}} \tag{7.12}$$

$$d(\mathcal{K}_{(i)k}, \overline{S^T\hat{\mathcal{K}}_{(i)k}}) \leq C_1\, d(\mathcal{K}_{(i)k}, \mathcal{H}) + d(\mathcal{H}, \overline{S^T\hat{\mathcal{H}}}) \tag{7.13}$$

hold for all $i \geq i_0$.
$\overline{S^T\hat{\mathcal{H}}}$ is a k-dimensional A^*-invariant subspace.

The number $d(\mathcal{H}, \overline{S^T\hat{\mathcal{H}}})$ *is less than one and depends on the condition numbers of the eigenvalues of* A.
The constant C_1 *depends on the starting vectors* $V_{(0)}e_1$ *and* $W_{(0)}e_1$, *the matrix* A *and the form* $\langle x, y\rangle_S = x^T S y$.

d) **(Quadratic convergence)**
The sequences $(H_{21}^{(i)})_{i\in\mathbb{N}_0}$, *and* $(\hat{H}_{21}^{(i)})_{i\in\mathbb{N}_0}$ *tend to zero at least quadratically.*

e) **(Cubic convergence)**
If the vectors $V_{(i)}e_{k+1}$ *and* $W_{(i)}e_{k+1}$ *(for all* $i \geq i_1$; $i_1 \in \mathbb{N}$ *arbitrary) are chosen as regular Lanczos vectors then the sequences* $(H_{21}^{(i)})_{i\in\mathbb{N}_0}$, $(H_{12}^{(i)})_{i\in\mathbb{N}_0}$, $(\hat{H}_{21}^{(i)})_{i\in\mathbb{N}_0}$ *and* $(\hat{H}_{12}^{(i)})_{i\in\mathbb{N}_0}$ *tend to zero cubically.*

In the proof of Theorem 7.2 we use the following Lemmas.

Lemma 7.2 *Let* $((A, V_{(i)}, \Pi_{(i)}, H_{(i)}, m_i, \psi_i))_{i\in\mathbb{N}_0}$ *and* $((A^+, W_{(i)}, \Pi_{(i)}^+, \hat{H}_{(i)}, m_i, \hat{\psi}_i))_{i\in\mathbb{N}_0}$ *be generated by an identically restarted look-ahead Lanczos method. Define*

$$\Phi_i := \psi_{i-1}\psi_{i-2}\cdots\psi_0 \qquad \text{for all} \quad i \in \mathbb{N}.$$

Then, for every $i \in \mathbb{N}$, *there is a scalar* $\hat{\gamma}_i \in \mathbb{C} \setminus \{0\}$ *such that the identities*

$$\begin{aligned} V_i e_1 &= \psi_{i-1}(A)V_{(i-1)}e_1 = \Phi_i(A)V_{(0)}e_1 \\ W_i e_1 &= \gamma_{i-1}\psi_{i-1}(A^+)W_{(i-1)}e_1 = \hat{\gamma}_{i-1}\Phi_i(A^+)W_{(0)}e_1 \end{aligned}$$

are valid.

Proof.
By assumption, we have $V_{(i)}e_1 = \psi_{i-1}(A)V_{(i-1)}e_1$ and $W_{(i)}e_1 = \gamma_{i-1}\psi_{i-1}(A^+)W_{(i-1)}e_1$. Induction on i yields the desired identities. ∎

Lemma 7.3 *Let* X, $Y \in \mathrm{Mat}_{N\times l}$ $(l \leq N)$ *with* $X^*X = I$ *and* $Y^*Y = I$ *be given. Then*

$$d(\operatorname{Im} X, \operatorname{Im} Y) = \sqrt{1 - \sigma_{min}(Y^*X)^2}.$$

Lemma 7.3 is proved in [28] pp. 77-78.

Lemma 7.4 *Let* $\mathcal{X}$ *and* $\mathcal{Y}$ *be subspaces of* $\mathbb{C}^N$ *of the same dimension. Then the following statements hold:*

a) There is a projection Π *with* $\operatorname{Im}\Pi = \mathcal{X}$ *and* $\operatorname{Ker}\Pi = \mathcal{Y}^{\perp_*}$ *if and only if* $d(\mathcal{X},\mathcal{Y}) < 1$. *In this case the inequality*

$$||\Pi|| \leq \frac{1}{\sqrt{1 - d(\mathcal{X},\mathcal{Y})^2}} \tag{7.14}$$

holds. Π *is an orthogonal projection if and only if* $d(\mathcal{X},\mathcal{Y}) = 0$.

b) Let $G, F \in \mathrm{Mat}_{N\times k}$ $(N \geq k)$ *be matrices with* $\operatorname{Im} G = \mathcal{Y}$ *and* $\operatorname{Im} F = \mathcal{X}$. *Then the following estimate holds:*

$$\sigma_{min}(G^*F) \geq \sigma_{min}(G)\,\sigma_{min}(F)\sqrt{1 - d(\mathcal{X},\mathcal{Y})^2}. \tag{7.15}$$

Proof. Let k denote the dimension of $\mathcal{X}$. Let $X, Y \in \mathrm{Mat}_{N\times k}$ be matrices with orthonormal columns such that $\operatorname{Im} X = \mathcal{X}$ and $\operatorname{Im} Y = \mathcal{Y}$. Applying Lemma 7.3 we have

$$d(\mathcal{X},\mathcal{Y}) = \sqrt{1 - \sigma_{min}(Y^*X)^2}. \tag{7.16}$$

Thus, the inequality $d(\mathcal{X},\mathcal{Y}) < 1$ holds if and only if the matrix Y^*X is invertible. By Lemma 2.3 there is a projection Π with $\operatorname{Im}\Pi = \mathcal{X}$ and $\operatorname{Ker}\Pi = \mathcal{Y}^{\perp_*}$ if and only if the matrix Y^*X is invertible. In this case Π has the representation $\Pi = X(Y^*X)^{-1}Y^*$. Using (7.16) and the inequality

$$||\Pi|| = ||X(Y^*X)^{-1}Y^*|| \leq ||X||\,||(Y^*X)^{-1}||\,||Y^*|| = ||(Y^*X)^{-1}|| = \frac{1}{\sigma_{min}(Y^*X)} \tag{7.17}$$

we find the desired inequality (7.14).
The projection Π is orthogonal if and only if $\mathcal{X} = \mathcal{Y}$. This is the case if and only if $d(\mathcal{X},\mathcal{Y}) = 0$.
Now we turn to the proof of inequality (7.15). There are invertible matrices $R_1, R_2 \in \mathrm{Mat}_k$ such that $G = YR_1$ and $F = XR_2$. Using the identities $G^*F = R_1^*(Y^*X)R_2$, $\sigma_{min}(G) = \sigma_{min}(R_1)$, $\sigma_{min}(F) = \sigma_{min}(R_2)$ and the estimates (7.16), (7.17) we obtain

$$\begin{aligned}\sigma_{min}(G^*F) &\geq \sigma_{min}(G)\,\sigma_{min}(Y^*X)\,\sigma_{min}(F)\\ &= \sigma_{min}(G)\sqrt{1 - d(\mathcal{X},\mathcal{Y})^2}\sigma_{min}(F).\end{aligned}$$

■

Lemma 7.5 *Let* (A, V, Π, H, s) *and* $(A^+, W, \Pi^+, \hat{H}, s)$ *be Krylov tuples generated by a look ahead Lanczos method. Let* $\langle x, y\rangle_S = x^TSy$ *be the dual form used in the look-ahead Lanczos process. Define* $D := W^TSV$. *Suppose that the vectors* Ve_{k+1} *and* We_{k+1} *are regular Lanczos vectors. Consider the partitions*

$$H = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}, \qquad \hat{H} = \begin{bmatrix} \hat{H}_{11} & \hat{H}_{12} \\ \hat{H}_{21} & \hat{H}_{22} \end{bmatrix}$$

where $H_{11},\ \hat{H}_{11} \in \mathrm{Mat}_k$ *and* $H_{22},\ \hat{H}_{22} \in \mathrm{Mat}_{m-k}$. *Then*

$$d(\mathcal{K}(k, A, Ve_1), \overline{S^T\mathcal{K}(k, A^+, We_1)}) < 1$$

and the estimates

$$||H_{12}|| \le \kappa_2(D)\,||\hat{H}_{21}|| \le \frac{\kappa_2(S)\,\kappa_2(W)\,\kappa_2(V)}{\sqrt{1 - d(\mathcal{K}(k, A, Ve_1), \overline{S^T\mathcal{K}(s, A^+, We_1)})^2}}\,||\hat{H}_{21}|| \tag{7.18}$$

hold.

Proof. By assumption Ve_{k+1} and We_{k+1} are regular vectors. Therefore the projection Π with $\mathrm{Im}\,\Pi = \mathcal{K}(k, A, Ve_1)$ and $\mathrm{Ker}\,\Pi = \mathcal{K}(k, A^+, We_1)^{\perp_S}$ exists. The equation

$$\mathcal{K}(s, A^+, We_1)^{\perp_S} = (S^*\overline{\mathcal{K}(s, A^+, We_1)})^{\perp_*}$$

together with Lemma 7.4 a) yields $d(\mathcal{K}(s, A, Ve_1), S^*\overline{\mathcal{K}(s, A^+, We_1)}) < 1$. By Lemma 5.5 the identity $H = D^{-1}\hat{H}^T D$ holds and there are matrices $D_1 \in \mathrm{Mat}_k$ and $D_2 \in \mathrm{Mat}_{s-k}$ such that $D = \mathrm{diag}\,(D_1, D_2)$. The block diagonal structure of D yields the equation

$$H_{12} = {D_1}^{-1}\hat{H}_{21}^T D_2$$

which implies

$$||H_{12}|| \le ||{D_1}^{-1}||\,||\hat{H}_{21}^T||\,||D_2|| \le \kappa_2(D)||\hat{H}_{21}^T||. \tag{7.19}$$

Lemma 7.4 b) applied to $G := \overline{S^TW}$ and $F := V$ yields

$$||D^{-1}|| = \frac{1}{\sigma_{min}(W^TSV)} = \frac{1}{\sigma_{min}((\overline{S^TW})^*V)} \tag{7.20}$$

$$\le \frac{1}{\sigma_{min}(\overline{S^TW})\,\sigma_{min}(V)\sqrt{1 - d(\mathcal{K}(s, A, Ve_1), S^*\overline{\mathcal{K}(s, A^+, We_1)})}}. \tag{7.21}$$

Combining the relations (7.19) and (7.21) and using

$$||D|| \le \sigma_{max}(W)\,\sigma_{max}(S)\,\sigma_{max}(V) \qquad \text{and} \qquad \sigma_{min}(\overline{S^TW}) \ge \sigma_{min}(S)\,\sigma_{min}(W)$$

we find the estimates (7.18). ∎

Lemma 7.6 *Let* $\mathcal{K}, \mathcal{H}$ *be two subspaces of* $\mathbb{C}^N$ *of the same dimension. Let* $A \in \mathrm{Mat}_N$. *Suppose there is a subspace* $\mathcal{W}$ *of* $\mathbb{C}^N$ *with the properties:*

a) $\mathcal{W} \supseteq \mathcal{K} + \mathcal{H}$.

b) $A\mathcal{W} = \mathcal{W}$.

Then $d(A\mathcal{K}, A\mathcal{H}) \leq \kappa_2(A|_{\mathcal{W}})\, d(\mathcal{K}, \mathcal{H})$

Lemma 7.6 is evident.

Lemma 7.7 *Let* $A \in \mathrm{Mat}_N$ *have distinct eigenvalues. Let* $\langle x, y\rangle = x^T S y$ *be an arbitrary non-degenerate bilinear form. Then there is an invertible matrix* Y *such that* $Y^T = Y$ *and* $A^T = YAY^{-1}$ *(see* Lemma 5.6*).*
Define $X := (S^{-1})^T Y$. *Then*

$$A^+ = XAX^{-1}.$$

Let $\mathcal{H}$ *be an arbitrary* A*-invariant subspace. Let* $\lambda_1, \lambda_2, \ldots, \lambda_k$ *be the eigenvalues of* $A|_{\mathcal{H}}$. *Then* $\hat{\mathcal{H}} := X\mathcal{H}$ *is an* A^+*-invariant subspace and* $\lambda_1, \lambda_2, \ldots, \lambda_k$ *are the eigenvalues of* $A^+|_{\hat{\mathcal{H}}}$.

Lemma 7.7 follows immediately from Lemma 5.6. ∎

Lemma 7.8 *Let* A *have distinct eigenvalues. Let* v *and* w *be vectors with dual directions (with respect to* A *and* $\langle x, y\rangle_S = x^T S y$*). Let* Φ *be an arbitrary polynomial. Let* $\hat{\gamma} \in \mathbb{C}\setminus\{0\}$. *Let* X *be defined as in Lemma 7.7. Then the following statements hold:*

a) The vectors $\Phi(A)v$ *and* $\hat{\gamma}\,\Phi(A^+)w$ *have dual directions (with respect to* A *and* $\langle x, y\rangle_S$*).*

b) $$X\mathcal{K}(N, A, \Phi(A)v) = \mathcal{K}(N, A^+, \Phi(A^+)w) = \mathcal{K}(N, A^+, \hat{\gamma}\,\Phi(A^+)w).$$

c) There is a polynomial φ *such that*

$$w = X\varphi(A)v \qquad \textit{and} \qquad \varphi(A)\mathcal{K}(N, A, v) = \mathcal{K}(N, A, v).$$

Proof. By assumption and Lemma 7.7 there are eigenvectors $x_1, x_2, \ldots, x_k$ of A and eigenvectors $y_1, y_2, \ldots, y_k$ of A^+ and non-vanishing coefficients α_i and β_i such that

$$v = \sum_{i=1}^{k} \alpha_i x_i, \qquad\qquad w = \sum_{i=1}^{k} \beta_i y_i$$

and $y_i = X x_i$ for all $i \in \{1, 2 \ldots, k\}$.
Let λ_i be the eigenvalue of A belonging to x_i. Choose a polynomial φ of degree $k-1$ such that

$$\varphi(\lambda_i) = \frac{\beta_i}{\alpha_i} \qquad \text{for } i = 1, 2, \ldots, k. \tag{7.22}$$

Then

$$\begin{aligned} \varphi(A)v &= \sum_{i=1}^{k} \alpha_i \varphi(A) x_i = \sum_{i=1}^{k} \alpha_i \varphi(\lambda_i) x_i \\ &= \sum_{i=1}^{k} \beta_i x_i = \sum_{i=1}^{k} \beta_i X^{-1} y_i \\ &= X^{-1} w \end{aligned}$$

and $\varphi(A)|_{\mathcal{K}(N,A,v)}$ is invertible. Hence c) is proved.
Using c) we obtain b):

$$\begin{aligned} X\mathcal{K}(N,A,\Phi(A)v) &= X\varphi(A)\mathcal{K}(N,A,\Phi(A)v) \\ &= X\mathcal{K}(N,A,\varphi(A)\Phi(A)v) = X\mathcal{K}(N,A,\Phi(A)\varphi(A)v) \\ &= \mathcal{K}(N,XAX^{-1},X\Phi(A)\varphi(A)v) \\ &= \mathcal{K}(N,XAX^{-1},\Phi(XAX^{-1})X\varphi(A)v) \\ &= \mathcal{K}(N,A^+,\Phi(A^+)w) \\ &= \mathcal{K}(N,A^+,\hat{\gamma}\,\Phi(A^+)w). \end{aligned}$$

Finally the identities

$$\begin{aligned} \Phi(A)v &= \sum_{i=1}^{k}\alpha_i\Phi(A)x_i = \sum_{i=1}^{k}\alpha_i\Phi(\lambda_i)x_i \\ \hat{\gamma}\Phi(A^+)w &= \sum_{i=1}^{k}\hat{\gamma}\beta_i\Phi(A^+)y_i = \sum_{i=1}^{k}\hat{\gamma}\beta_i\Phi(\lambda_i)y_i \end{aligned}$$

imply

$$\alpha_i\Phi(\lambda_i)=0 \quad\Leftrightarrow\quad \Phi(\lambda_i)=0 \quad\Leftrightarrow\quad \hat{\gamma}\beta_i\Phi(\lambda_i)=0$$

which proves a).

■

Lemma 7.9 *Let $A\in \mathrm{Mat}_N$ have distinct eigenvalues. Let $\mathcal{H}$ be an invariant subspace of A. Let $\hat{\mathcal{H}}$ be the invariant subspace of A^+ defined in Lemma 7.7. Then the following statements hold:*

a) $d(\mathcal{H},\overline{S^T\hat{\mathcal{H}}})<1.$

b) There is a projection Π with $\operatorname{Im}\Pi=\mathcal{H}$ and $\operatorname{Ker}\Pi=\hat{\mathcal{H}}^{\perp_S}=(\overline{S^T\hat{\mathcal{H}}})^{\perp_*}$.

Proof. Let the matrices X, Y and S be defined as in Lemma 7.7. Then $\overline{S^T\hat{\mathcal{H}}}$ and $\mathcal{H}^{\perp_S}$ have the representations

$$\overline{S^T\hat{\mathcal{H}}} = \overline{S^TX\mathcal{H}} = \overline{S^T(S^T)^{-1}Y\mathcal{H}} = \overline{Y\mathcal{H}} \tag{7.23}$$

$$\mathcal{H}^{\perp_S} = (S^T\hat{\mathcal{H}})^{\perp} = (\overline{S^T\hat{\mathcal{H}}})^{\perp_*} = \overline{Y\mathcal{H}}^{\perp_*}. \tag{7.24}$$

By definition $A^T=YAY^{-1}$. Hence

$$A^* = \overline{YAY}^{-1}.$$

Let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be the eigenvalues of $A|_{\mathcal{H}}$. Let x_i denote the eigenvector of A belonging to λ_i. Then $\mathcal{H} = \operatorname{span}\{x_1, x_2, \ldots, x_k\}$. Consider the equation

$$A^* \overline{Y x_i} = \overline{Y A Y^{-1}} \overline{Y x_i} = \overline{Y A x_i} = \overline{\lambda Y x_i}.$$

Consequently $z_i := \overline{Y x_i}$ is the eigenvector of A^* belonging to the eigenvalue $\overline{\lambda_i}$ and

$$\overline{Y\mathcal{H}} = \operatorname{span}\{z_1, z_2, \ldots, z_k\}.$$

Using that $z_i^* x_j = 0$ for $i \neq j$ and $z_i^* x_i \neq 0$ for all i we conclude that

$$D := [z_1, z_2, \ldots, z_k]^* [x_1, x_2, \ldots, x_k]$$

is invertible. Hence, by Lemma 2.3, the projection Π with $\operatorname{Im}\Pi = \mathcal{H}$ and $\operatorname{Ker}\Pi = \overline{Y\mathcal{H}}$ exists. Finally Lemma 7.4 a) yields

$$d(\mathcal{H}, \overline{Y\mathcal{H}}) < 1$$

which, combined with (7.23), proves the lemma. ∎

Lemma 7.10 *Let $A \in \operatorname{Mat}_N$ have distinct eigenvalues. Choose a non-degenerate bilinear form $\langle x, y\rangle_S = x^T S y$ on $\mathbb{C}^N \times \mathbb{C}^N$. Choose the starting vectors Ve_1 and We_1 such that they have dual directions (with respect to A and $\langle\cdot,\cdot\rangle_S$). Then every look-ahead Lanczos method does not break down and generates two Krylov tuples (A, V, Π, H, m) and $(A^+, W, \Pi^+, \hat{H}, m)$. Further the identity*

$$m := \dim \mathcal{K}(N, A, Ve_1) = \dim \mathcal{K}(N, A^+, We_1) \tag{7.25}$$

is valid.

Proof. By Lemma 7.8 b) the identity (7.25) holds. Lemma 7.8 b) and c) together with Lemma 7.9 b) applied to $\mathcal{H} := \mathcal{K}(N, A, Ve_1)$ yields the existence of the Lanczos projection Π with

$$\operatorname{Im}\Pi = \mathcal{K}(m, A, Ve_1) \qquad \text{and} \qquad \operatorname{Ker}\Pi = \mathcal{K}(m, A^+, We_1)^{\perp_S}. \tag{7.26}$$

Hence the look-ahead Lanczos process does not break down (see Remark 5.4).

∎

Lemma 7.11 below states that the convergence of the Krylov subspaces $\mathcal{K}_{(i)k}$ implies the convergence of the dual Krylov subspaces $\hat{\mathcal{K}}_{(i)k}$

Lemma 7.11 *Let $\mathcal{H}$ be an arbitrary A-invariant subspace. Let $\hat{\mathcal{H}}$ be the A^+-invariant subspace defined in Lemma 7.7. Assume the hypothesis of Theorem 7.2. Then the inequality*

$$d(\hat{\mathcal{K}}_{(i)k}, \hat{\mathcal{H}}) \leq C\, d(\mathcal{K}_{(i)k}, \mathcal{H}) \tag{7.27}$$

holds for all $i \in \mathbb{N}_0$. The constant C depends on the matrix A, the choice of the form $\langle x, y\rangle_S = x^T S y$ and the choice of the first starting vectors $W_{(0)}e_1$, $V_{(0)}e_1$.

Proof. Let X be defined as in Lemma 7.7. Let Φ_i be defined as in Lemma 7.2. Let k denote the dimension of $\mathcal{H}$.
By Lemma 7.8 there is a polynomial φ such that

$$W_{(0)}e_1 = X\varphi(A)V_{(0)}e_1 \qquad \text{and} \qquad \varphi(A)\mathcal{K}(N, A, V_{(0)}e_1) = \mathcal{K}(N, A, V_{(0)}e_1). \tag{7.28}$$

Using Lemma 7.2, Lemma 7.8 and equation (7.28), we obtain

$$\begin{aligned}
\hat{\mathcal{K}}_{(i)k} &= \mathcal{K}(k, A^+, W_{(i)}e_1) = \mathcal{K}(k, A^+, \Phi_i(A^+)W_{(0)}e_1) \\
&= \mathcal{K}(k, XAX^{-1}, \Phi_i(XAX^{-1})W_{(0)}e_1) = X\mathcal{K}(k, A, X^{-1}X\Phi_i(A)X^{-1}W_{(0)}e_1) \\
&= X\mathcal{K}(k, A, \Phi_i(A)X^{-1}X\varphi(A)V_{(0)}e_1) = X\mathcal{K}(k, A, \varphi(A)\Phi_i(A)V_{(0)}e_1) \\
&= X\varphi(A)\mathcal{K}(k, A, \Phi_i(A)V_{(0)}e_1) = X\varphi(A)\mathcal{K}(k, A, V_{(i)}e_1) \\
&= X\varphi(A)\mathcal{K}_{(i)k}.
\end{aligned} \tag{7.29}$$

Now we are ready to prove the inequality (7.27). Using the equation (7.29), the relation $\hat{\mathcal{H}} = X\mathcal{H}$, Lemma 7.6, the equations (7.28) and $A\mathcal{H} \subset \mathcal{H}$ we find

$$\begin{aligned}
d(\hat{\mathcal{K}}_{(i)k}, \hat{\mathcal{H}}) &= d(X\varphi(A)\mathcal{K}_{(i)k}, \hat{\mathcal{H}}) = d(X\varphi(A)\mathcal{K}_{(i)k}, X\mathcal{H}) \\
&\leq \kappa_2(X)\, d(\varphi(A)\mathcal{K}_{(i)k}, \mathcal{H}) = \kappa_2(X)\, d(\varphi(A)\mathcal{K}_{(i)k}, \varphi(A)\mathcal{H}) \\
&\leq \kappa_2(X)\, \kappa_2(\varphi(A|_{\mathcal{K}(N,A,V_0e_1)}))\, d(\mathcal{K}_{(i)k}, \mathcal{H}).
\end{aligned}$$

Defining $C := \kappa_2(X)\, \kappa_2(\varphi(A|_{\mathcal{K}(N,A,V_0e_1)}))$, the lemma is proved. ∎

Proof of Theorem 7.2

Corollary 5.7 yields that all eigenvalues of $H_{(i)}$ are eigenvalues of A and that the matrices $H_{(i)}$ and $\hat{H}_{(i)}$ are similar.
By Lemma 7.2 the starting vectors $V_{(i)}e_1$ and $W_{(i)}e_1$ have, for all $i \in \mathbb{N}$, the representation

$$V_{(i)}e_1 = \Phi_i(A)V_{(0)}e_1 \qquad \text{and} \qquad W_{(i)}e_1 = \hat{\gamma}_i\Phi_i(A^+)W_{(0)}e_1$$

where $\Phi_i := \psi_{i-1}\psi_{i-2}\cdots\psi_0$ and $\hat{\gamma}_i \in \mathbb{C} \setminus \{0\}$.
Lemma 7.8 applies and yields that for all $i \in \mathbb{N}$ the vectors $V_{(i)}e_1$ and $W_{(i)}e_1$ have dual directions. Thus, by Lemma 7.10, the restarted look-ahead Lanczos method does not break down and the identities

$$s_i = m_i := \dim \mathcal{K}(N, A, V_ie_1) = \dim \mathcal{K}(N, A^+, W_ie_1)$$

hold for all $i \in \mathbb{N}$.

By the construction of a practical restarted look-ahead Lanczos method there is a constant K_0 such that

$$\kappa_2(V_{(i)}) \leq K_0 \qquad \text{and} \qquad \kappa_2(W_{(i)}) \leq K_0 \qquad \text{for all} \quad i \in \mathbb{N}_0. \tag{7.30}$$

By assumption the sequence $(H_{21}^{(i)})_{i\in\mathbb{N}_0}$ tends to zero. Thus, by Theorem 6.6 there is a k-dimensional A-invariant subspace $\mathcal{H}$ and there are constants K_1 and K_2 such that

$$||H_{21}^{(i)}|| \le K_1\, d(\mathcal{K}_{(i)k}, \mathcal{H}) \le K_2||H_{21}^{(i)}|| \qquad \text{for all} \quad i \in \mathbb{N}. \tag{7.31}$$

Consequently

$$d(\mathcal{K}_{(i)k}, \mathcal{H}) \to 0 \qquad \text{for} \quad i \to \infty. \tag{7.32}$$

By Lemma 7.11 there is a k-dimensional A^+-invariant subspace $\hat{\mathcal{H}}$ and a constant C depending on A, $W_{(0)}e_1$, $V_{(0)}e_1$ and $\langle\cdot,\cdot\rangle_S$ such that

$$d(\hat{\mathcal{K}}_{(i)k}, \hat{\mathcal{H}}) \le C\, d(\mathcal{K}_{(i)k}, \mathcal{H}) \qquad \text{for all} \quad i \in \mathbb{N}. \tag{7.33}$$

The inequality (7.33) combined with (7.32) yields

$$d(\hat{\mathcal{K}}_{(i)k}, \hat{\mathcal{H}}) \to 0 \qquad \text{for} \quad i \to \infty.$$

Hence, by Theorem 6.6, there is a constant K_3 such that

$$||\hat{H}_{12}^{(i)}|| \le K_3 d(\hat{\mathcal{K}}_{(i)k}, \hat{\mathcal{H}}) \qquad \text{for all} \quad i \in \mathbb{N}. \tag{7.34}$$

Thus, statements a) and b) of Theorem 7.2 are proved.

Now we show that there is a number i_0 such that for all $i \ge i_0$ the Lanczos projection $\Pi_{k,(i)}$ with

$$\operatorname{Im}\Pi_{k,(i)} = \mathcal{K}_{(i)k} \qquad \text{and} \qquad \operatorname{Ker}\Pi_{k,(i)} = (\hat{\mathcal{K}}_{(i)k})^{\perp_S}$$

exists. In view of Lemma 7.4 a) and the identity

$$(\hat{\mathcal{K}}_{(i)k})^{\perp_S} = (S^T\hat{\mathcal{K}}_{(i)k})^{\perp} = (\overline{S^T\hat{\mathcal{K}}_{(i)k}})^{\perp_*} \tag{7.35}$$

it is sufficient to prove that there is a number i_0 and a constant C_3 such that

$$d(\mathcal{K}_{(i)k}, \overline{S^T\hat{\mathcal{K}}_{(i)k}}) \le C_3 < 1 \qquad \text{for all} \quad i \ge i_0. \tag{7.36}$$

Using the triangle inequality, Lemma 7.6, the inequality (7.33) and the identity

$$d(\overline{S^T\hat{\mathcal{H}}}, \overline{S^T\hat{\mathcal{K}}_{(i)k}}) = d(S^T\hat{\mathcal{H}}, S^T\hat{\mathcal{K}}_{(i)k}),$$

we find

$$\begin{aligned}
d(\mathcal{K}_{(i)k}, \overline{S^T\hat{\mathcal{K}}_{i(k)}}) &\le d(\mathcal{K}_{(i)k}, \mathcal{H}) + d(\mathcal{H}, \overline{S^T\hat{\mathcal{H}}}) + d(\overline{S^T\hat{\mathcal{H}}}, \overline{S^T\hat{\mathcal{K}}_{(i)k}}) \\
&= d(\mathcal{K}_{(i)k}, \mathcal{H}) + d(\mathcal{H}, \overline{S^T\hat{\mathcal{H}}}) + d(S^T\hat{\mathcal{H}}, S^T\hat{\mathcal{K}}_{(i)k}) \\
&\le d(\mathcal{K}_{(i)k}, \mathcal{H}) + d(\mathcal{H}, \overline{S^T\hat{\mathcal{H}}}) + \kappa_2(S^T)\, d(\hat{\mathcal{H}}, \hat{\mathcal{K}}_{(i)k}) \\
&\le d(\mathcal{K}_{(i)k}, \mathcal{H}) + d(\mathcal{H}, \overline{S^T\hat{\mathcal{H}}}) + \kappa_2(S^T)\, C\, d(\mathcal{H}, \mathcal{K}_{(i)k}) \\
&= C_1 d(\mathcal{K}_{(i)k}, \mathcal{H}) + d(\mathcal{H}, \overline{S^T\hat{\mathcal{H}}})
\end{aligned} \tag{7.37}$$

for all $i \in \mathbb{N}$. Here $C_1 := 1 + \kappa_2(S^T)\, C$. Lemma 7.9 a) yields the estimate

$$d(\mathcal{H}, \overline{S^T \hat{\mathcal{H}}}) < 1. \tag{7.38}$$

Combining (7.32), (7.37) and (7.38), we find the desired inequality (7.36).
Further Lemma 7.4 combined with the identity (7.35) yields the estimates (7.12) and (7.13).

Finally we turn to the proof of the cubic convergence of the sequences

$$(H_{21}^{(i)})_{i\in\mathbb{N}_0}, \qquad (H_{12}^{(i)})_{i\in\mathbb{N}_0}, \qquad (\hat{H}_{21}^{(i)})_{i\in\mathbb{N}_0} \quad \text{and} \quad (\hat{H}_{12}^{(i)})_{i\in\mathbb{N}_0}. \tag{7.39}$$

By assumption the vectors $V_{(i)}e_{k+1}$ and $W_{(i)}e_{k+1}$ are regular Lanczos vectors for all $i \geq i_1$. Using Lemma 7.5, Theorem 6.6 and the inequalities (7.30), (7.33) and (7.36), we obtain the estimates

$$\begin{aligned}
||H_{12}^{(i)}|| &\leq \frac{\kappa_2(S)\,\kappa_2(W_{(i)})\kappa_2(V_{(i)})}{\sqrt{1 - d(\mathcal{K}_{(i)k}, \overline{S^T\hat{\mathcal{K}}_{(i)k}})^2}} ||\hat{H}_{21}^{(i)}|| \\
&\leq \frac{\kappa_2(S)\, K_0^2}{\sqrt{1 - C_3^2}} ||\hat{H}_{21}^{(i)}|| \\
&\leq \frac{\kappa_2(S)\, K_0^2\, K_3}{\sqrt{1 - C_3^2}} d(\hat{\mathcal{K}}_{(i)k}, \hat{\mathcal{H}}) \\
&\leq \frac{\kappa_2(S)\, K_0^2\, K_3\, C}{\sqrt{1 - C_3^2}} d(\mathcal{K}_{(i)k}, \mathcal{H}) \qquad (7.40) \\
&\leq \frac{\kappa_2(S)\, K_0^2\, K_3\, C\, K_2}{\sqrt{1 - C_3^2}} ||H_{21}^{(i)}|| \qquad (7.41)
\end{aligned}$$

for all $i \geq i_1$.
Finally, Theorem 7.1 yields the cubic convergence of the sequences (7.39). ∎

In the following we present and discuss a variant of the identically restarted practical look-ahead Lanczos method:

A variant of the identically restarted practical look-ahead Lanczos method

(1) Let $A \in \mathrm{Mat}_N$ be given.

(2) Choose two starting vectors $v_1, w_1 \in \mathbb{C}^N \setminus \{0\}$.

(3) Apply a practical look-ahead Lanczos method (using the form $\langle x, y\rangle = x^T y$) on A, v_1, w_1 to obtain two Krylov tuples

$$(A, V, \Pi, H, r) \qquad \text{and} \qquad (A^T, W, \Pi^T, \hat{H}, r).$$

(4) (Generalized Rayleigh-quotient shift): Choose a number $k \in \{1, 2, \ldots, r-1\}$. Consider the partition

$$H = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} \qquad \text{where} \quad H_{11} \in \mathrm{Mat}_k.$$

Compute the characteristic polynomial ψ of H_{22}.

(5) Compute $v_1^{(0)} := \psi(H)e_1$ and $w_1^{(0)} := \psi(H)D^T e_1$ where $D := W^T V$.

(6) Apply an identically restarted practical look-ahead Lanczos method on H, $v_1^{(0)}$, $w_1^{(0)}$. Determine the shift polynomials ψ_i (for use in the restarted look-ahead Lanczos process) by the generalized Rayleigh-quotient shift strategy and use the form $\langle x, y\rangle = x^T y$ in the look-ahead Lanczos process.

For shortness of notation we call this method 'VRL method'.

Corollary 7.1 below states that the VRL method converges cubically if it converges and if A has pairwise distinct eigenvalues. Notice that, in contrast to Theorem 7.2 (cubic convergence of the identically restarted look-ahead Lanczos method), Corollary 7.1 does not require the assumption that the first starting vectors have dual directions.

Corollary 7.1 (Cubic convergence of the VRL method) *Let $A \in \mathrm{Mat}_N$. Choose arbitrary starting vectors $v_1, w_1 \in \mathbb{C}^N \setminus \{0\}$. Then the VRL method generates matrices $V, W \in \mathrm{Mat}_{N\times r}$, $H \in \mathrm{Mat}_r$ and two restarted look-ahead Lanczos Krylov tuples*

$$((H, V_{(i)}, \Pi_{(i)}, H_{(i)}, m_i, \psi_i))_{i\in\mathbb{N}_0} \qquad \textit{and} \qquad ((H^T, W_{(i)}, \Pi_{(i)}^T, \hat{H}_{(i)}, m_i, \psi_i))_{i\in\mathbb{N}_0}.$$

The characteristic polynomial of the matrix H is a divisior of the minimal polynomial of the matrix A.
The characteristic polynomial of $H_{(i)}$ is a divisor of the characteristic polynomial of H.
The matrices $H_{(i)}$ and $\hat{H}_{(i)}$ are similar.
Further the relations

$$r \geq m_i := \dim \mathcal{K}(N, H, V_{(i)}e_1) = \dim \mathcal{K}(N, H^T, W_{(i)}e_1) \geq m_{i+1}$$

hold for all $i \in \mathbb{N}_0$.
Consider the partitions

$$H_{(i)} = \begin{bmatrix} H_{11}^{(i)} & H_{12}^{(i)} \\ H_{21}^{(i)} & H_{22}^{(i)} \end{bmatrix}, \qquad \hat{H}_{(i)} = \begin{bmatrix} \hat{H}_{11}^{(i)} & \hat{H}_{12}^{(i)} \\ \hat{H}_{21}^{(i)} & \hat{H}_{22}^{(i)} \end{bmatrix}$$

where $H_{11}^{(i)}$, $\hat{H}_{11}^{(i)} \in \mathrm{Mat}_k$ and $H_{22}^{(i)}$, $\hat{H}_{22}^{(i)} \in \mathrm{Mat}_{m_i-k}$.
Suppose that the sequence $(H_{21}^{(i)})_{i\in\mathbb{N}_0}$ tends to zero and that A has distinct eigenvalues.
Then the following statements hold:

a) *The sequence* $(\hat{H}_{21}^{(i)})_{i\in\mathbb{N}_0}$ *tends to zero.*

b) **(Existence of regular vectors)**
There is an $i_0 \in \mathbb{N}$ *such that for all* $i \geq i_0$ *the vectors* $V_{(i)}e_{k+1}$ *and* $W_{(i)}e_{k+1}$ *can be chosen as regular Lanczos vectors. In other words: The Lanczos projection* $\Pi_{k,(i)}$ *(with* $\operatorname{Im}\Pi_{k,(i)} = \mathcal{K}(k, H, V_{(i)}e_1)$ *and* $\operatorname{Ker}\Pi_{k,(i)} = \mathcal{K}(k, H^T, W_{(i)}e_1)^{\perp}$*) exists for all* $i \geq i_0$.
Further there is a constant C *such that*

$$||\Pi_{k,(i)}|| \leq C \tag{7.42}$$

for all $i \geq i_0$.
The constant C *depends on the starting vectors* v_1, w_1, *the conditon number of* V *and the matrix* A.

c) **(Quadratic convergence)**
The sequences $(H_{21}^{(i)})_{i\in\mathbb{N}_0}$ *and* $(\hat{H}_{21}^{(i)})_{i\in\mathbb{N}_0}$ *tend to zero at least quadratically.*

d) **(Cubic convergence)**
If the vectors $V_{(i)}e_{k+1}$ *and* $W_{(i)}e_{k+1}$ *(for all* $i \geq i_1$*;* $i_1 \in \mathbb{N}$ *arbitrary) are chosen as regular Lanczos vectors, then the sequences* $(H_{21}^{(i)})_{i\in\mathbb{N}_0}$, $(H_{12}^{(i)})_{i\in\mathbb{N}_0}$, $(\hat{H}_{21}^{(i)})_{i\in\mathbb{N}_0}$ *and* $(\hat{H}_{12}^{(i)})_{i\in\mathbb{N}_0}$ *tend to zero cubically.*

Remark 7.1 If the matrix A is not sparse then the VRL method requires less arithmetic operations than the identically restarted practical look-ahead Lanczos algorithm, since the matrix H is block tridiagonal and upper Hessenberg.

Proof. Corollary 5.7 yields that the characteristic polynomial of H is a divisor of the minimal polynomial of A and that the characteristic polynomial of $H_{(i)}$ is a divisor of the minimal polynomial of H. By Lemma 5.5 the matrices $H_{(i)}$ and $H_{(i)}$ are similar.

The matrices $H \in \mathrm{Mat}_r$ and $\hat{H} \in \mathrm{Mat}_r$ are unreduced upper Hessenberg matrices. Hence the equations

$$\mathcal{K}(r, H, e_1) = \mathbb{C}^r = \mathcal{K}(r, \hat{H}, e_1) \tag{7.43}$$

hold. Consequently, the minimal polynomial of H and the characteristic polynomial of H are identical.
Using $H^T = D^T\hat{H}(D^T)^{-1}$ (where $D := W^TV$) (see Lemma 5.5) we obtain

$$\mathcal{K}(r, H^T, D^Te_1) = \mathcal{K}(r, D^T\hat{H}(D^T)^{-1}, D^Te_1) = D^T\mathcal{K}(r, \hat{H}, (D^T)^{-1}D^Te_1) = \mathbb{C}^r \tag{7.44}$$

which implies together with (7.43) that the vectors e_1 and D^Te_1 have dual directions with respect to the matrix H and $\langle x, y\rangle = x^Ty$.
Consequently, by Lemma 7.8, the vectors $v_1^{(0)} = \psi(H)e_1$ and $w_1^{(0)} = \psi(H)D^Te_1$ have dual directions with respect to H and $\langle x, y\rangle = x^Ty$, too.
Finally, Theorem 7.2 applies and Corollary 7.1 follows. ∎

7.2 Numerical example

The following examples show that the identically restarted look-ahead Lanczos method and the VRL method are unstable because of instabilities in the look-ahead Lanczos process.

$$A = \begin{bmatrix} 0.00 & 0.06 & -0.28 & 0.41 & 0.55 & -0.66 & 0.64 & 0.32 \\ 0.16 & 0.95 & 0.14 & -0.32 & 0.12 & 0.59 & -0.17 & 0.74 \\ -0.60 & -0.31 & -0.56 & 0.61 & 0.33 & 0.66 & 0.93 & 0.49 \\ 0.96 & 0.30 & -0.25 & 0.57 & -0.45 & -0.05 & 0.99 & -0.22 \\ 0.28 & 0.29 & -1.00 & 0.04 & -0.81 & 0.02 & -0.45 & -0.70 \\ -0.88 & 0.41 & -0.64 & -0.81 & -0.09 & -0.71 & 0.00 & 0.49 \\ 0.17 & -0.46 & 0.99 & -0.24 & -0.98 & -0.85 & -0.09 & -0.63 \\ -0.59 & -0.02 & -0.45 & -0.50 & 0.40 & 0.29 & -0.17 & -0.43 \end{bmatrix}, \quad w_1^{(0)} = \begin{bmatrix} 0.85 \\ 0.66 \\ 0.56 \\ 0.04 \\ -0.58 \\ 0.59 \\ 0.02 \\ -0.27 \end{bmatrix}$$

$$v_1^{(0)} = \begin{bmatrix} 0.74 & -0.45 & -0.35 & -0.35 & -0.46 & -0.65 & 0.68 & -0.82 \end{bmatrix}^T$$

The entries of the matrix A and the starting vectors $w_1^{(0)}$ and $v_1^{(0)}$ are generated by the random number generator ($\approx$ uniform distribution in [-1,1]) of 'Turbo Pascal' [6].

The matrix A has the following eigenvalues (rounded):

Eigenvalues of A (rounded)		
Eigenvalue	Real part	Imaginary part
λ_1	$1.94768032815462 \cdot 10^{00}$	0
$\lambda_{2/3}$	$7.22771408213559 \cdot 10^{-01}$	$\pm 3.86823730013324 \cdot 10^{-01}$
$\lambda_{4/5}$	$-4.63268021120600 \cdot 10^{-01}$	$\pm 3.06680358938131 \cdot 10^{-01}$
λ_6	$-1.07573663811272 \cdot 10^{00}$	0
$\lambda_{7/8}$	$-1.23547523211391 \cdot 10^{00}$	$\pm 1.23396246460755 \cdot 10^{00}$

The eigenvalues of the matrix A are not ill-conditioned ($cond(lambda_j) \leq 2.5$ for all j, see Definition (7.11)); The vectors $v_1^{(0)}$ and $w_1^{(0)}$ have general directions with respect to A and $\langle x, y \rangle = x^T y$ (see Definition 7.3). Thus, by Theorem 7.2, every identically restarted look-ahead Lanczos method, using the generalized Rayleigh-quotient shift strategy, does not break down and converges cubically, if it converges at all. Indeed, we observe cubic convergence even of the identically restarted ordinary Lanczos method using the generalized

Rayleigh-quotient shift strategy of order two:

Eigenvalues of $H_{22}^{(i)}$ (ordinary Lanczos method, exact arithmetic)				
i	$\Re\,\lambda_{4/5}^{(i)}$	$\Im\,\lambda_{4/5}^{(i)}$	$B_6^{(i)}$	$C_6^{(i)}$
0	$-5.59229549963598 \cdot 10^{-01}$	$0.00000000000000 \cdot 10^{00}$	$3 \cdot 10^{-01}$	$-4 \cdot 10^{-01}$
	$-3.64371917679964 \cdot 10^{-01}$	$0.00000000000000 \cdot 10^{00}$		
1	$-4.61432540407506 \cdot 10^{-01}$	$\pm 3.06571398605869 \cdot 10^{-01}$	$6 \cdot 10^{-02}$	$-7 \cdot 10^{-02}$
2	$-4.63268035075506 \cdot 10^{-01}$	$\pm 3.06680350809448 \cdot 10^{-01}$	$2 \cdot 10^{-04}$	$5 \cdot 10^{-05}$
3	$-4.63268021120600 \cdot 10^{-01}$	$\pm 3.06680358938131 \cdot 10^{-01}$	$3 \cdot 10^{-12}$	$1 \cdot 10^{-12}$

Here

$$H_{(i)} = \begin{bmatrix} H_{11}^{(i)} & H_{12}^{(i)} \\ H_{21}^{(i)} & H_{22}^{(i)} \end{bmatrix} = \left[\begin{array}{cccccc|cc} E_1^{(i)} & C_1^{(i)} & 0 & 0 & 0 & 0 & 0 & 0 \\ B_1^{(i)} & E_2^{(i)} & C_2^{(i)} & 0 & 0 & 0 & 0 & 0 \\ 0 & B_2^{(i)} & E_3^{(i)} & C_3^{(i)} & 0 & 0 & 0 & 0 \\ 0 & 0 & B_3^{(i)} & E_4^{(i)} & C_4^{(i)} & 0 & 0 & 0 \\ 0 & 0 & 0 & B_4^{(i)} & E_5^{(i)} & C_5^{(i)} & 0 & 0 \\ 0 & 0 & 0 & 0 & B_5^{(i)} & E_6^{(i)} & C_6^{(i)} & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & B_6^{(i)} & E_7^{(i)} & C_7^{(i)} \\ 0 & 0 & 0 & 0 & 0 & 0 & B_7^{(i)} & E_8^{(i)} \end{array}\right].$$

If the restarted Lanczos process is carried out in finite precision arithmetic then the eigenvalues of $H_{22}^{(i)}$ do not tend to eigenvalues of A although the eigenvalues of $H_{11}^{(i)}$ converge to eigenvalues of A and the entries $B_6^{(i)}$ and $C_6^{(i)}$ of $H_{(i)}$ tend to zero. Tab. 7.2.1, Tab, 7.2.2 and Tab. 7.2.3 below show the results of the identically restarted Lanczos

process carried out in the IEEE-arithmetic 'extended'.

Eigenvalues of $H_{22}^{(i)}$ (ordinary Lanczos method, finite precision arithmetic)				
i	$\Re\,\lambda_{4/5}^{(i)}$	$\Im\,\lambda_{4/5}^{(i)}$	$B_6^{(i)}$	$C_6^{(i)}$
0	$-5.59229549963598 \cdot 10^{-01}$	$0.00000000000000 \cdot 10^{\ 00}$	$3 \cdot 10^{-01}$	$-4 \cdot 10^{-01}$
	$-3.64371917679964 \cdot 10^{-01}$	$0.00000000000000 \cdot 10^{\ 00}$		
1	$-\underline{4.6}1432540407506 \cdot 10^{-01}$	$\pm\underline{3.06}571398605869 \cdot 10^{-01}$	$6 \cdot 10^{-02}$	$-7 \cdot 10^{-02}$
2	$-\underline{4.63268}035075506 \cdot 10^{-01}$	$\pm\underline{3.06680350}809448 \cdot 10^{-01}$	$2 \cdot 10^{-04}$	$5 \cdot 10^{-05}$
3	$-\underline{4.6326}3869946471 \cdot 10^{-01}$	$\pm\underline{3.066}90297973582 \cdot 10^{-01}$	$3 \cdot 10^{-12}$	$1 \cdot 10^{-12}$
4	$1.97792532368542 \cdot 10^{\ 00}$	$0.00000000000000 \cdot 10^{\ 00}$	$2 \cdot 10^{-14}$	$-5 \cdot 10^{-14}$
	$-1.14995739998428 \cdot 10^{\ 01}$	$0.00000000000000 \cdot 10^{\ 00}$		

Tab.7.2.1

Eigenvalues of $H_{11}^{(3)}$ and $H_{11}^{(4)}$ (finite precision arithmetic)		
Eigenvalue	Real part	Imaginary part
$\lambda_1^{(3)} = \lambda_1^{(4)}$	$\underline{1.94768032815462} \cdot 10^{\ 00}$	$\underline{0.00000000000000} \cdot 10^{\ 00}$
$\lambda_2^{(3)} = \lambda_2^{(4)}$	$\underline{7.22771408213559} \cdot 10^{-01}$	$-\underline{3.86823730013324} \cdot 10^{-01}$
$\lambda_3^{(3)} = \lambda_3^{(4)}$	$\underline{7.22771408213559} \cdot 10^{-01}$	$\underline{3.86823730013324} \cdot 10^{-01}$
$\lambda_6^{(3)} = \lambda_6^{(4)}$	$-\underline{1.07573663811272} \cdot 10^{\ 00}$	$\underline{0.00000000000000} \cdot 10^{\ 00}$
$\lambda_7^{(3)} = \lambda_7^{(4)}$	$-\underline{1.23547523211391} \cdot 10^{\ 00}$	$-\underline{1.23396246460755} \cdot 10^{\ 00}$
$\lambda_8^{(3)} = \lambda_8^{(4)}$	$-\underline{1.23547523211391} \cdot 10^{\ 00}$	$\underline{1.23396246460755} \cdot 10^{\ 00}$

Tab.7.2.2

Ordinary Lanczos method, finite precision arithmetic					
i	0	1	2	3	4
$\sigma_{min}(D_{(i)})$	$1 \cdot 10^{-01}$	$2 \cdot 10^{-01}$	$2 \cdot 10^{-01}$	$2 \cdot 10^{-01}$	$1 \cdot 10^{-01}$
$\sigma_{max}(D_{(i)})$	$4 \cdot 10^{-01}$	$7 \cdot 10^{-01}$	$7 \cdot 10^{-01}$	$9 \cdot 10^{-01}$	$9 \cdot 10^{-01}$
$\kappa_2(W_{(i)})$	$4 \cdot 10^{01}$	$9 \cdot 10^{00}$	$2 \cdot 10^{01}$	$2 \cdot 10^{01}$	$1 \cdot 10^{03}$
$\kappa_2(V_{(i)})$	$4 \cdot 10^{01}$	$9 \cdot 10^{00}$	$2 \cdot 10^{01}$	$2 \cdot 10^{01}$	$6 \cdot 10^{03}$
$\max_{1\le k\le 8} \mu_w(k)$	$6 \cdot 10^{00}$	$2 \cdot 10^{00}$	$2 \cdot 10^{00}$	$4 \cdot 10^{00}$	$8 \cdot 10^{00}$
$\max_{1\le k\le 8} \mu_v(k)$	$4 \cdot 10^{00}$	$2 \cdot 10^{00}$	$2 \cdot 10^{00}$	$4 \cdot 10^{00}$	$8 \cdot 10^{00}$
$\min_{1\le k\le 7, k\neq 6}\{\lvert B_k^{(i)}\rvert, \lvert C_k^{(i)}\rvert\}$	$3 \cdot 10^{-01}$	$4 \cdot 10^{-01}$	$3 \cdot 10^{-01}$	$1 \cdot 10^{-01}$	$4 \cdot 10^{-02}$
$\max_{1\le k\le 7, k\neq 6}\{\lvert B_k^{(i)}\rvert, \lvert C_k^{(i)}\rvert\}$	$8 \cdot 10^{00}$	$2 \cdot 10^{00}$	$2 \cdot 10^{00}$	$6 \cdot 10^{00}$	$2 \cdot 10^{00}$
τ_6	$2 \cdot 10^{-17}$	$3 \cdot 10^{-18}$	$2 \cdot 10^{-17}$	$3 \cdot 10^{-16}$	$3 \cdot 10^{-16}$
τ_8	$2 \cdot 10^{-16}$	$4 \cdot 10^{-16}$	$2 \cdot 10^{-12}$	$3 \cdot 10^{-03}$	$9 \cdot 10^{-01}$

Tab.7.2.3

The quantities $\mu_v(k)$, $\mu_w(k)$ and τ_s are defined in Section 4. We recall that τ_s is a measure for the block biorthogonality of the basis $((W_{(i)}e_j, V_{(i)}e_j))_{1\le j\le s}$, generated by a (look-ahead) Lanczos method carried out in finite-precision arithmetic.

Now we try to explain the surprising observation that the eigenvalues of $H_{22}^{(i)}$ do not converge to eigenvalues of A although the eigenvalues of $H_{11}^{(i)}$ tend to eigenvalues of A. By Tab. 7.2.3, row τ_6, the biorthogonality of the basis $(W_{(i)}e_j, V_{(i)}e_j)_{1\le j\le 6}$ is nearly as good as if exact arithmetic is used in the Lanczos process. Thus, for every i the matrix $H_{11}^{(i)}$(finite-precision arithmetic) approximates $H_{11}^{(i)}$(exact arithmetic) very well. The numbers τ_8 (see Tab. 7.2.3) show that the biorthogonality of the basis $(W_{(i)}e_j, V_{(i)}e_j)_{1\le j\le 8}$ is nearly destroyed for $i \ge 3$. Consequently, the matrix $H_{22}^{(i)}$(finite precision arithmetic) is a poor approximation to $H_{22}^{(i)}$(exact arithmetic) for $i \ge 3$.

Why is the number τ_8 much greater than the number τ_6 for $i \ge 3$?
We recall that

$$B_k^{(i)} = \|(I - \Pi_{k,(i)})AV_{(i)}e_k\| \text{ and } C_k^{(i)} = \|(I - \Pi_{k,(i)}^T)A^T W_{(i)}e_k\| \frac{\delta_{k+1}^{(i)}}{\delta_k^{(i)}} \tag{7.45}$$

where $D_{(i)} = \operatorname{diag}(\delta_1^{(i)}, \delta_2^{(i)}, \ldots, \delta_8^{(i)}) := W_{(i)}^T V_{(i)}$ and $\Pi_{k,(i)}$ is the Lanczos projection with $\operatorname{Im}\Pi_{k,(i)} = \mathcal{K}_{i(k)}$.

If $B_k^{(i)}$ is very small then $AV_{(i)}e_k \approx \Pi_{k,(i)}AV_{(i)}e_k$. In this case the computation of

$$V_{(i)}e_{k+1} = \frac{1}{B_k^{(i)}}(AV_{(i)}e_k - \Pi_{k,(i)}AV_{(i)}e_k)$$

is ill-conditioned (see also Section 4). Similarly, we find that the computation of $W_{(i)}e_{k+1}$ is ill-conditioned if $|C_k^{(i)}|$ is very small.

The numbers $B_6^{(i)}$ and $|C_6^{(i)}|$ are very small for $i \geq 3$ (see Tab. 7.2.3). Consequently, the computation of $W_{(i)}e_7$ and $V_{(i)}e_7$ is ill-conditioned for $i \geq 3$. This implies the quick loss of biorthogonality (τ_8 is not small).

By Tab.7.2.3 the matrices $D_{(i)}$, $W_{(i)}$ and $V_{(i)}$ are well-conditioned and the Lanczos projection $\Pi_{k,(i)}$ is not 'very skew' (i.e. $||\Pi_{k,(i)}||$ is of moderate size) for all $k = 1, 2, \ldots, 8$ and all i. Further

$$\min_{1\leq k\leq 5}\{B_k^{(i)}, |C_k^{(i)}|\} \geq 0.04$$

for all i. Thus, the computation of $V_{(i)}e_{k+1}$ and $W_{(i)}e_{k+1}$ is well conditioned for $k = 0, 1, 2, 3, 4, 5$ and all i. This leads to very small numbers τ_6.

Is it possible to improve the eigenvalue approximation by using a practical look-ahead Lanczos method ?

The identically restarted practical look-ahead Lanczos method using orthogonal inner vectors (i.e. $W_{(i)}e_{\hat{\nu}(k)+j} \perp W_{(i)}e_{\hat{\nu}(k)+l}$ for $1 \leq l, j \leq \hat{\nu}(k+1) - \hat{\nu}(k)$, $l \neq j$) and the checks (4.10) and (4.11) with $tol = 0.01$, $\xi = 10$ and $\Psi = 1 \cdot 10^5$ (see Section 4) generates the following:

Degree indices generated in the look-ahead Lanczos process (finite precision arithmetic)	
i	$(\hat{\nu}(k))_{k\in\hat{K}} \subseteq (\nu(k))_{k\in K} = (k)_{1\leq k\leq 8}$
0	$\hat{\nu}(k) = k$ for $1 \leq k \leq 8$
1	$\hat{\nu}(k) = k$ for $1 \leq k \leq 8$
2	$\hat{\nu}(k) = k$ for $1 \leq k \leq 8$
3	$\hat{\nu}(k) = k$ for $1 \leq k \leq 5$, $\hat{\nu}(6) = 7$, $\hat{\nu}(7) = 8$
4	$\hat{\nu}(k) = k$ for $1 \leq k \leq 5$, $\hat{\nu}(6) = 7$, $\hat{\nu}(7) = 8$

Tab.7.2.4

Eigenvalues of $H_{22}^{(i)}$ (look-ahead Lanczos method, finite precision arithmetic)				
i	$\Re\,\lambda_{4/5}^{(i)}$	$\Im\,\lambda_{4/5}^{(i)}$	$\|\|H_{21}^{(i)}\|\|_\infty$	$\|\|H_{12}^{(i)}\|\|_\infty$
0	$-5.59229549963598 \cdot 10^{-01}$	$0.00000000000000 \cdot 10^{00}$	$3 \cdot 10^{-01}$	$-4 \cdot 10^{-01}$
	$-3.64371917679964 \cdot 10^{-01}$	$0.00000000000000 \cdot 10^{00}$		
1	$-\underline{4.6}1432540407506 \cdot 10^{-01}$	$\pm\underline{3.06}571398605869 \cdot 10^{-01}$	$6 \cdot 10^{-02}$	$-7 \cdot 10^{-02}$
2	$-\underline{4.63268}035075506 \cdot 10^{-01}$	$\pm\underline{3.06680350}809448 \cdot 10^{-01}$	$2 \cdot 10^{-04}$	$5 \cdot 10^{-05}$
3	$-\underline{4.63267}902841025 \cdot 10^{-01}$	$\pm\underline{3.06680}687178604 \cdot 10^{-01}$	$2 \cdot 10^{-12}$	$2 \cdot 10^{00}$
4	$2.13504696626625 \cdot 10^{00}$	$0.00000000000000 \cdot 10^{00}$	$3 \cdot 10^{-15}$	$-1 \cdot 10^{00}$
	$-5.39106851565365 \cdot 10^{-01}$	$0.00000000000000 \cdot 10^{00}$		

Tab.7.2.5

Look-ahead Lanczos method, finite precision arithmetic					
i	0	1	2	3	4
$\sigma_{min}(D_{(i)})$	$1 \cdot 10^{-01}$	$2 \cdot 10^{-01}$	$2 \cdot 10^{-01}$	$2 \cdot 10^{-01}$	$5 \cdot 10^{-01}$
$\sigma_{max}(D_{(i)})$	$4 \cdot 10^{-01}$	$7 \cdot 10^{-01}$	$7 \cdot 10^{-01}$	$9 \cdot 10^{-01}$	$9 \cdot 10^{-01}$
$\kappa_2(W_{(i)})$	$4 \cdot 10^{01}$	$9 \cdot 10^{00}$	$2 \cdot 10^{01}$	$2 \cdot 10^{01}$	$2 \cdot 10^{02}$
$\kappa_2(V_{(i)})$	$4 \cdot 10^{01}$	$9 \cdot 10^{00}$	$2 \cdot 10^{01}$	$2 \cdot 10^{01}$	$2 \cdot 10^{02}$
$\max_{1\le k\le 7}\mu_w(k)$	$6 \cdot 10^{00}$	$2 \cdot 10^{00}$	$2 \cdot 10^{00}$	$4 \cdot 10^{00}$	$3 \cdot 10^{00}$
$\max_{1\le k\le 7}\mu_v(k)$	$4 \cdot 10^{00}$	$2 \cdot 10^{00}$	$2 \cdot 10^{00}$	$4 \cdot 10^{00}$	$2 \cdot 10^{00}$
τ_5	$2 \cdot 10^{-17}$	$3 \cdot 10^{-18}$	$2 \cdot 10^{-17}$	$3 \cdot 10^{-18}$	$1 \cdot 10^{-17}$
τ_8	$2 \cdot 10^{-16}$	$4 \cdot 10^{-16}$	$2 \cdot 10^{-12}$	$2 \cdot 10^{-04}$	$8 \cdot 10^{-01}$

Tab.7.2.6

For $i \geq 3$

$$H_{(i)} = \begin{bmatrix} H_{11}^{(i)} & H_{12}^{(i)} \\ H_{21}^{(i)} & H_{22}^{(i)} \end{bmatrix} = \left[\begin{array}{cccccc|cc} E_1^{(i)} & C_1^{(i)} & 0 & 0 & 0 & 0 & 0 & 0 \\ B_1^{(i)} & E_2^{(i)} & C_2^{(i)} & 0 & 0 & 0 & 0 & 0 \\ 0 & B_2^{(i)} & E_3^{(i)} & C_3^{(i)} & 0 & 0 & 0 & 0 \\ 0 & 0 & B_3^{(i)} & E_4^{(i)} & C_4^{(i)} & 0 & 0 & 0 \\ 0 & 0 & 0 & B_4^{(i)} & E_5^{(i)} & C_{5,1}^{(i)} & C_{5,2}^{(i)} & 0 \\ 0 & 0 & 0 & 0 & B_5^{(i)} & E_{6,1}^{(i)} & E_{6,2}^{(i)} & C_{6,1}^{(i)} \\ \hline 0 & 0 & 0 & 0 & 0 & E_{6,3}^{(i)} & E_{6,4}^{(i)} & C_{6,2}^{(i)} \\ 0 & 0 & 0 & 0 & 0 & 0 & B_6^{(i)} & E_7^{(i)} \end{array}\right].$$

Comparing Tab.7.2.1, Tab.7.2.5 and Tab.7.2.3, Tab.7.2.6, we find that the eigenvalue approximation, generated by the restarted look-ahead Lanczos method, is not better than the eigenvalue approximation computed by the restarted ordinary Lanczos method. The fast loss of block biorthogonality in the look-ahead Lanzos process has the same reasons as in the ordinary Lanczos process.

In the following chapter we discuss how the instabilities in the look-ahead Lanczos process can be avoided.

8 Look-ahead Lanczos matrices by similarity transformations

In this chapter we derive a new algorithm which uses a sequence of transformation matrices to generate the look-ahead Lanczos matrices H in a stable way.

For this purpose we introduce the concept of block Hessenberg matrices of index $(\mu(k))_{0\leq k\leq b}$ and block tridiagonal matrices of index $(\mu(k))_{0\leq k\leq b}$.

Definition 8.1 *A matrix $H \in \mathrm{Mat}_N$ is called block Hessenberg of index*

$$0 = \mu(0) < \mu(1) < \ldots < \mu(b-1) < \mu(b) = N \tag{8.1}$$

if H has the form

$$H = \begin{bmatrix} E_1 & & & \star \\ B_1 & E_2 & & \\ & \ddots & \ddots & \\ 0 & & B_{b-1} & E_b \end{bmatrix}, \quad E_k \in \mathrm{Mat}_{\mu(k)-\mu(k-1)}, \quad B_k \in \mathrm{Mat}_{\mu(k+1)-\mu(k)\times\mu(k)-\mu(k-1)}$$

and each of the non-diagonal blocks B_k has rank ≤ 1.

Remark 8.1 Notice the difference between k-Hessenberg matrices (see Definiton 6.2) and block Hessenberg matrices of index $(\mu(k))_{0\leq k\leq b}$.

Definition 8.2 *A matrix $H \in \mathrm{Mat}_N$ is called block tridiagonal of index*

$$0 = \mu(0) < \mu(1) < \ldots < \mu(b-1) < \mu(b) = N \tag{8.2}$$

if H has the form

$$H = \begin{bmatrix} E_1 & C_1 & & 0 \\ B_1 & E_2 & \ddots & \\ & \ddots & \ddots & C_{b-1} \\ 0 & & B_{b-1} & E_b \end{bmatrix}, \quad E_k \in \mathrm{Mat}_{\mu(k)-\mu(k-1)}, \quad B_k \in \mathrm{Mat}_{\mu(k+1)-\mu(k)\times\mu(k)-\mu(k-1)}$$

and $C_k \in \mathrm{Mat}_{\mu(k)-\mu(k-1)\times\mu(k+1)-\mu(k)}$.

Let a matrix $A \in \mathrm{Mat}_N$ be given. First we show that if there are invertible matrices $V, W \in \mathrm{Mat}_N$ and a sequence of numbers (8.1) satisfying the following relations

$$H := V^{-1}AV \quad \text{is block Hessenberg of index } (\mu(k))_{0\le k\le b} \tag{8.3}$$

$$\hat{H} := W^{-1}A^TW \quad \text{is block Hessenberg of index } (\mu(k))_{0\le k\le b} \tag{8.4}$$

$$D := W^TV = \mathrm{diag}\,(\Theta_1, \Theta_2, \ldots, \Theta_b), \quad \Theta_j \in \mathrm{Mat}_{\mu(j)-\mu(j-1)} \tag{8.5}$$

then the matrices H and $\hat{H}$ are block tridiagonal of index $(\mu(k))_{0\le k\le b}$ with non-diagonal blocks of rank ≤ 1.

Lemma 8.1 *Assume* (8.3)-(8.5). *Then the equation* $H = D^{-1}\hat{H}^TD$ *holds.*

Proof. By assumption the matrices W and V are invertible. Thus D is invertible, too. The equation

$$H = V^{-1}AV = D^{-1}W^TA(W^T)^{-1}D = D^{-1}(W^{-1}A^TW)^TD = D^{-1}\hat{H}^TD$$

proves Lemma 8.1. ∎

In the following l_j is defined as $l_j := \mu(j) - \mu(j-1)$.

Lemma 8.2 *Assume* (8.3)-(8.5). *Then the matrices* H *and* $\hat{H}$ *are block tridiagonal with non-diagonal blocks of rank* ≤ 1. *In other words, there are vectors* $d_j^{(1)}, \hat{d}_j^{(1)} \in \mathbb{C}^{l_j+1}$, $d_j^{(2)}, \hat{d}_j^{(2)} \in \mathbb{C}^{l_j}$ *and matrices* $E_j, \hat{E}_j \in \mathrm{Mat}_{l_j}$, $C_j, \hat{C}_j \in \mathrm{Mat}_{l_j\times l_{j+1}}$ $B_j, \hat{B}_j \in \mathrm{Mat}_{l_{j+1}\times l_j}$ *such that*

$$H = \begin{bmatrix} E_1 & C_1 & & 0 \\ B_1 & E_2 & \ddots & \\ & \ddots & \ddots & C_{b-1} \\ 0 & & B_{b-1} & E_b \end{bmatrix}, \quad B_j = d_j^{(1)}\{d_j^{(2)}\}^T, \quad C_j = \{\Theta_j^{-1}\hat{d}_j^{(2)}\}\{\Theta_{j+1}^T\hat{d}_j^{(1)}\}^T \tag{8.6}$$

$$\hat{H} = \begin{bmatrix} \hat{E}_1 & \hat{C}_1 & & 0 \\ \hat{B}_1 & \hat{E}_2 & \ddots & \\ & \ddots & \ddots & \hat{C}_{b-1} \\ 0 & & \hat{B}_{b-1} & \hat{E}_b \end{bmatrix}, \quad \hat{B}_j = \hat{d}_j^{(1)}\{\hat{d}_j^{(2)}\}^T, \quad \hat{C}_j = \{(\Theta_j^{-1})^Td_j^{(2)}\}\{\Theta_{j+1}d_j^{(1)}\}^T \tag{8.7}$$

and

$$E_j = \Theta_j^{-1}\hat{E}_j^T\Theta_j. \tag{8.8}$$

Proof. Using Lemma 8.1 and the block Hessenberg form of $\hat{H}$ we obtain the identity

$$
\begin{aligned}
H = D^{-1}\hat{H}^T D &= \begin{bmatrix} \Theta_1^{-1} & & & 0 \\ & \Theta_2^{-1} & & \\ & & \ddots & \\ 0 & & & \Theta_b^{-1} \end{bmatrix} \begin{bmatrix} \hat{E}_1^T & \hat{B}_1^T & & 0 \\ & \hat{E}_2^T & \ddots & \\ & & \ddots & \hat{B}_{b-1}^T \\ \star & & & \hat{E}_b^T \end{bmatrix} \begin{bmatrix} \Theta_1 & & & 0 \\ & \Theta_2 & & \\ & & \ddots & \\ 0 & & & \Theta_b \end{bmatrix} \\
&= \begin{bmatrix} (\Theta_1^{-1}\hat{E}_1^T\Theta_1) & (\Theta_1^{-1}\hat{B}_1^T\Theta_2) & & 0 \\ & (\Theta_2^{-1}\hat{E}_2^T\Theta_2) & \ddots & \\ & & \ddots & (\Theta_{b-1}^{-1}\hat{B}_{b-1}^T\Theta_b) \\ \star & & & (\Theta_b^{-1}\hat{E}_b^T\Theta_b) \end{bmatrix}. \qquad (8.9)
\end{aligned}
$$

By assumption, B_k has rank 1 and $\hat{H}$ is block Hessenberg. Hence equation (8.9) yields the desired identities (8.6) and (8.8). In a similar manner we obtain the identity (8.7). ■

The following theorem clarifies the connection between matrices satisfying the relations (8.3)-(8.5) and the matrices generated by a (practical) look-ahead Lanczos method.

Theorem 8.1 *Let a matrix $A \in \mathrm{Mat}_N$ and two starting vectors $v_1, w_1 \in \mathbb{C}^N \setminus \{0\}$ be given. Suppose there are invertible matrices $V, W \in \mathrm{Mat}_N$ and a sequence of numbers $0 = \mu(0) < \mu(1) < \ldots < \mu(b-1) \leq \mu(b) := N$ such that*

$$
\begin{aligned}
Ve_1 &= v_1 && (8.10)\\
We_1 &= w_1 && (8.11)\\
H &:= V^{-1}AV \quad \textit{is unreduced } \mu(b-1)\textit{-Hessenberg} && (8.12)\\
\hat{H} &:= W^{-1}A^TW \quad \textit{is unreduced } \mu(b-1)\textit{-Hessenberg} && (8.13)\\
D &:= W^TV = \mathit{diag}\,(\Theta_1, \Theta_2, \ldots, \Theta_b), \quad \Theta_j \in \mathrm{Mat}_{\mu(j)-\mu(j-1)}. && (8.14)
\end{aligned}
$$

Let $H^{(1)}$ denote the leading $\mu(b-1)\times\mu(b-1)$ submatrix of H. Let $\hat{H}^{(1)}$ denote the leading $\mu(b-1)\times\mu(b-1)$ submatrix of $\hat{H}$. Consider the partitions

$$V = [V^{(1)}, V^{(2)}] \qquad \textit{and} \qquad W = [W^{(1)}, W^{(2)}]$$

where $V^{(1)}, W^{(1)} \in \mathrm{Mat}_{N\times\mu(b-1)}$. Then there is a (practical) look-ahead Lanczos method which can be carried out at least up to step $\mu(b-1)$ and which generates the Krylov tuples

$$(A, V^{(1)}, \Pi, H^{(1)}, \mu(b-1)) \qquad \textit{and} \qquad (A^T, W^{(1)}, \Pi^T, \hat{H}^{(1)}, \mu(b-1)), \qquad (8.15)$$

where Π is the projection with

$$\mathrm{Im}\,\Pi = \mathcal{K}(\mu(b-1), A, v_1) \quad \textit{and} \quad \mathrm{Im}\,\Pi^T = \mathcal{K}(\mu(b-1), A^T, w_1)^\perp.$$

Proof. The assumptions (8.10)-(8.13) imply that the column vectors of $V^{(1)}$ and $W^{(1)}$ form bases of the Krylov chains $(\mathcal{K}(i, A, v_1))_{1\le i\le \mu(b-1)}$ and $(\mathcal{K}(i, A^T, w_1))_{1\le i\le \mu(b-1)}$, respectively. By assumption (8.14) the sequence of vectors $((W^{(1)}e_i, V^{(1)}e_i))_{1\le i\le \mu(b-1)}$ forms a basis to $(\mu(k))_{1\le k\le b-1}$ (see Definition 4.1). Thus there is a (practical) look-ahead Lanczos method which generates the Krylov tuples (8.15) (see Section 4). ∎

Construction. Our next goal is to construct matrices H, $\hat{H}$ and V, with the properties (8.10)-(8.14), block by block in a stable way. First we choose transformation matrices $\tilde{P}_1 \in \mathrm{Mat}_N$ and $\tilde{O}_1 \in \mathrm{Mat}_N$ such that

$$\tilde{P}_1^{-1} v_1 \in \mathrm{span}\{e_1\}, \qquad \tilde{O}_1^{-1} w_1 \in \mathrm{span}\{e_1\} \tag{8.16}$$

$$\kappa_2(\tilde{O}_1) \le tol, \qquad \kappa_2(\tilde{P}_1) \le tol \tag{8.17}$$

and the matrices $\hat{H}_1 := \tilde{O}_1^{-1} A^T \tilde{O}_1$, $H_1 := \tilde{P}_1^{-1} A \tilde{P}_1$ and $D_1 := \tilde{O}_1^T \tilde{P}_1$ have the forms

$$\begin{aligned} H_1 &= \begin{bmatrix} E_1 & F_1 \\ G_1 & M_1 \end{bmatrix}, \quad G_1 = g_1 e_{l_1}^T, \quad g_1 \in \mathbb{C}^{N-\mu(1)} \setminus \{0\} \\ \hat{H}_1 &= \begin{bmatrix} \hat{E}_1 & \hat{F}_1 \\ \hat{G}_1 & \hat{M}_1 \end{bmatrix}, \quad \hat{G}_1 = \hat{g}_1 e_{l_1}^T, \quad \hat{g}_1 \in \mathbb{C}^{N-\mu(1)} \setminus \{0\} \\ D_1 &:= \mathrm{diag}\,(\Theta_1, \tilde{\Theta}_1), \qquad \Theta_1 \in \mathrm{Mat}_{l_1} \end{aligned}$$

where $\mu(1) := l_1 \in \{1, 2, \ldots, N\}$, $E_1, \hat{E}_1 \in \mathrm{Mat}_{l_1}$ are unreduced upper Hessenberg matrices and the matrices F_1 and $\hat{F}_1$ are of rank 1.
The constant *tol* is chosen by the user.

Now we proceed by induction on k.
Assume that we have already constructed transformation matrices $\tilde{P}_1, \tilde{P}_2, \ldots, \tilde{P}_k \in \mathrm{Mat}_N$ and $\tilde{O}_1, \tilde{O}_2, \ldots, \tilde{O}_k \in \mathrm{Mat}_N$ and a sequence of numbers $0 =: \mu(0) < \mu(1) < \ldots < \mu(k)$ such that

$$\kappa_2(\tilde{O}_j) \le tol, \qquad \kappa_2(\tilde{P}_j) \le tol \qquad \text{for all} \quad j \in \{1, 2, \ldots, k\} \tag{8.18}$$

and the matrices

$$H_k := \tilde{P}_k^{-1} \tilde{P}_{k-1}^{-1} \cdots \tilde{P}_1^{-1} A \tilde{P}_1 \tilde{P}_2 \cdots \tilde{P}_k \tag{8.19}$$

$$\hat{H}_k := \tilde{O}_k^{-1} \tilde{O}_{k-1}^{-1} \cdots \tilde{O}_1^{-1} A^T \tilde{O}_1 \tilde{O}_2 \cdots \tilde{O}_k \tag{8.20}$$

$$D_k := \tilde{O}_k^T \tilde{O}_{k-1}^T \cdots \tilde{O}_1^T \tilde{P}_1 \tilde{P}_2 \cdots \tilde{P}_k \tag{8.21}$$

have the forms

$$H_k = \begin{bmatrix} E_1 & C_1 & & & 0 \\ B_1 & E_2 & \ddots & & \\ & \ddots & \ddots & C_{k-1} & \\ & & B_{k-1} & E_k & F_k \\ 0 & & & G_k & M_k \end{bmatrix}, \quad G_k = g_k e_{l_k}^T,\ g_k \in \mathbb{C}^{N-\mu(k)} \setminus \{0\} \tag{8.22}$$

$$\hat{H}_k = \begin{bmatrix} \hat{E}_1 & \hat{C}_1 & & & 0 \\ \hat{B}_1 & \hat{E}_2 & \ddots & & \\ & \ddots & \ddots & \hat{C}_{k-1} & \\ & & \hat{B}_{k-1} & \hat{E}_k & \hat{F}_k \\ 0 & & & \hat{G}_k & \hat{M}_k \end{bmatrix}, \quad \hat{G}_k = \hat{g}_k e_{l_k}^T,\ \hat{g}_k \in \mathbb{C}^{N-\mu(k)} \setminus \{0\} \tag{8.23}$$

$$D_k = \operatorname{diag}(\Theta_1, \Theta_2, \ldots, \Theta_k, \tilde{\Theta}_k) \tag{8.24}$$

where $l_k := \mu(k) - \mu(k-1)$, $E_i, \hat{E}_i \in \mathrm{Mat}_{l_i}$ are unreduced upper Hessenberg matrices, the matrices C_i, $\hat{C}_i$, F_k and $\hat{F}_k$ are of rank 1 and the matrices B_i, $\hat{B}_i$ have the form

$$B_i = d_i e_{l_i}^T, \qquad \hat{B}_i = \hat{d}_i e_{l_i}^T \qquad \text{where} \quad d_i, \hat{d}_i \in \operatorname{span}\{e_1\}.$$

Suppose further that $\tilde{O}_i$ and $\tilde{P}_i$ have the form

$$\tilde{O}_i = \begin{bmatrix} I_{\mu(i-1)} & 0 \\ 0 & O_i \end{bmatrix}, \qquad \tilde{P}_i = \begin{bmatrix} I_{\mu(i-1)} & 0 \\ 0 & P_i \end{bmatrix} \tag{8.25}$$

for $i = 1, 2, \ldots, k$.

In the next step $(k \to k+1)$ we construct transformation matrices $\tilde{O}_{k+1}$ and $\tilde{P}_{k+1}$ of the form (8.25) such that

$$\kappa_2(\tilde{O}_{k+1}) \le tol, \qquad \kappa_2(\tilde{P}_{k+1}) \le tol \tag{8.26}$$

and the matrices $\hat{H}_{k+1}$, H_{k+1} and D_{k+1} have, for an appropriate choice of the number $\mu(k+1)$, the form (8.22), (8.23) and (8.24), respectively.

By definition, the matrices $\hat{H}_{k+1}, H_{k+1}$ and D_{k+1} have the representations

$$H_{k+1} := \tilde{P}_{k+1}^{-1} H_k \tilde{P}_{k+1} = \begin{bmatrix} E_1 & C_1 & & & 0 \\ B_1 & E_2 & \ddots & & \\ & \ddots & \ddots & C_{k-1} & \\ & & B_{k-1} & E_k & F_k P_{k+1} \\ 0 & & & P_{k+1}^{-1} G_k & P_{k+1}^{-1} M_k P_{k+1} \end{bmatrix} \tag{8.27}$$

$$\hat{H}_{k+1} := \tilde{O}_{k+1}^{-1} \hat{H}_k \tilde{O}_{k+1} = \begin{bmatrix} \hat{E}_1 & \hat{C}_1 & & & 0 \\ \hat{B}_1 & \hat{E}_2 & \ddots & & \\ & \ddots & \ddots & \hat{C}_{k-1} & \\ & & \hat{B}_{k-1} & \hat{E}_k & \hat{F}_k O_{k+1} \\ 0 & & & O_{k+1}^{-1} \hat{G}_k & O_{k+1}^{-1} \hat{M}_k O_{k+1} \end{bmatrix} \tag{8.28}$$

$$D_{k+1} := \tilde{O}_{k+1}^T D_k \tilde{P}_{k+1} = \operatorname{diag}(\Theta_1, \Theta_2, \ldots, \Theta_k, O_{k+1}^T \tilde{\Theta}_k P_{k+1}). \tag{8.29}$$

Comparing the equations (8.22), (8.23), (8.24) with (8.27), (8.28), (8.29) we recognize that the equations (8.22), (8.23), (8.24) hold (for the index $k+1$) if the matrices O_{k+1} and P_{k+1} fulfil the conditions:

a)

$$P_{k+1}^{-1} g_k \in \operatorname{span}\{e_1\} \setminus \{0\}. \tag{8.30}$$

b)

$$O_{k+1}^{-1} \hat{g}_k \in \operatorname{span}\{e_1\} \setminus \{0\}. \tag{8.31}$$

c) There is a number $\mu(k+1) \in \{\mu(k)+1, \mu(k)+2, \ldots, N\}$ such that the matrices $P_{k+1}^{-1} M_k P_{k+1}$, $O_{k+1}^{-1} \hat{M}_k O_{k+1}$ and $O_{k+1}^T \tilde{\Theta}^k P_{k+1}$ can be partitioned as follows:

$$P_{k+1}^{-1} M_k P_{k+1} = \begin{bmatrix} E_{k+1} & F_{k+1} \\ G_{k+1} & M_{k+1} \end{bmatrix} \tag{8.32}$$

$$O_{k+1}^{-1} \hat{M}_k O_{k+1} = \begin{bmatrix} \hat{E}_{k+1} & \hat{F}_{k+1} \\ \hat{G}_{k+1} & \hat{M}_{k+1} \end{bmatrix} \tag{8.33}$$

$$O_{k+1}^T \tilde{\Theta}_k P_{k+1} = \operatorname{diag}(\Theta_{k+1}, \tilde{\Theta}_{k+1}) \qquad (\Theta_{k+1} \in \operatorname{Mat}_{l_{k+1}}). \tag{8.34}$$

Here $E_{k+1}, \hat{E}_{k+1} \in \operatorname{Mat}_{l_{k+1}}$ are unreduced upper Hessenberg matrices and the matrices G_{k+1}, $\hat{G}_{k+1}$ have the form

$$G_{k+1} = g_{k+1} e_{l_{k+1}}^T, \quad \hat{G}_{k+1} = \hat{g}_{k+1} e_{l_{k+1}}^T \qquad (g_{k+1}, \hat{g}_{k+1} \in \mathbb{C}^{N-\mu(k+1)}).$$

Notice that by Lemma 8.2 and Theorem 8.1 the conditions (8.30)-(8.34) imply that the matrices F_{k+1} and $\hat{F}_{k+1}$ have rank 1.

For the solution of the eigenvalue problem $Ax = \lambda x$ it is sufficient to know H_{k+1} and the matrices $P_0, P_1, P_2, \ldots, P_{k+1}$. It is not necessary to know $\hat{H}_{k+1}$ or $O_1, O_2, \ldots, O_{k+1}$. Hence it is useful to reduce the computational effort in (8.30)-(8.34) by replacing O_{k+1} by a suitable chosen matrix Q_{k+1}.

By setting

$$g_0 := v_1, \qquad \hat{g}_0 := w_1, \qquad \tilde{\Theta}_0 := I, \qquad Q_{k+1} := \tilde{\Theta}_k^T O_{k+1}$$

and using the identity $\hat{M}_k = (\tilde{\Theta}_k^{-1})^T M_k^T \tilde{\Theta}_k^T$ (see Lemma 8.2) we obtain the following conditions for the matrices P_{k+1} and Q_{k+1} (for $k \geq 0$):

a)

$$d_k := P_{k+1}^{-1} g_k \in \operatorname{span}\{e_1\} \setminus \{0\}. \tag{8.35}$$

b)

$$\hat{d}_k := Q_{k+1} \tilde{\Theta}_k^T \hat{g}_k \in \operatorname{span}\{e_1\} \setminus \{0\}. \tag{8.36}$$

c) There is a number $\mu(k+1) \in \{\mu(k)+1, \mu(k)+2, \ldots, N\}$ such that the matrices $P_{k+1}^{-1} M_k P_{k+1}$, $Q_{k+1}^{-1} M_k^T Q_{k+1}$ and $Q_{k+1}^T P_{k+1}$ can be partitioned as follows:

$$P_{k+1}^{-1} M_k P_{k+1} = \begin{bmatrix} E_{k+1} & F_{k+1} \\ G_{k+1} & M_{k+1} \end{bmatrix} \tag{8.37}$$

$$Q_{k+1}^{-1} M_k^T Q_{k+1} = \begin{bmatrix} \hat{E}_{k+1} & \hat{F}_{k+1} \\ \hat{G}_{k+1} & \hat{M}_{k+1} \end{bmatrix} \tag{8.38}$$

$$Q_{k+1}^T P_{k+1} = \operatorname{diag}(\Theta_{k+1}, \tilde{\Theta}_{k+1}) \qquad (\Theta_{k+1} \in \operatorname{Mat}_{l_{k+1}}). \tag{8.39}$$

Here $E_{k+1}, \hat{E}_{k+1} \in \operatorname{Mat}_{l_{k+1}}$ are unreduced upper Hessenberg matrices and the matrices G_{k+1}, $\hat{G}_{k+1}$ have the form

$$G_{k+1} = g_{k+1} e_{l_{k+1}}^T, \quad \hat{G}_{k+1} = \hat{g}_{k+1} e_{l_{k+1}}^T \qquad (g_{k+1}, \hat{g}_{k+1} \in \mathbb{C}^{N-\mu(k+1)}).$$

The conditions (8.35)-(8.39) for the matrices Q_{k+1} and P_{k+1} have, compared to the conditions (8.30)-(8.34) for O_{k+1} and P_{k+1}, the following advantages:

- The computation of Θ_{k+1} and $\tilde{\Theta}_{k+1}$ requires less arithmetic operations.
- The matrix $\hat{M}_{k+1}$ is not needed in a later step.

The procedure 'generate H, V and D' and Theorem 8.2 below summarize our observations about the construction of well-conditioned look-ahead Lanczos matrices satisfying the relations (8.10)-(8.14).

procedure generate H, V and D

input: $N \in \mathbb{N}$; $A \in \mathrm{Mat}_N$; $v_1, w_1 \in \mathbb{C}^N \setminus \{0\}$; $tol \geq 1$;
output: $(\mu(k))_{0 \leq k \leq b} \subseteq \mathbb{N}$; $d_k, \hat{d}_k \in \mathbb{C}^{l_k}$ ($0 \leq k \leq b-2$);
$E_k \in \mathrm{Mat}_{l_k}$; $P_k \in \mathrm{Mat}_{N-\mu(k-1)}$; $\Theta_k \in \mathrm{Mat}_{l_k}$ ($1 \leq k \leq b-1$);
$M_{b-1} \in \mathrm{Mat}_{l_b}$, g_{b-1}, $\hat{g}_{b-1} \in \mathbb{C}^{l_b}$;

$g_0 := v_1$; $\hat{g}_0 := w_1$; $M_0 := A$; $\tilde{\Theta}_0 := I$; $k := 0$; $\mu(0) := 0$;
repeat

Find the smallest number

$$\mu(k+1) \in \{\mu(k)+1, \mu(k)+2, \ldots, N\}$$

and matrices Q_{k+1} and P_{k+1} such that

$$\kappa_2(Q_{k+1}) \leq tol, \qquad \kappa_2(P_{k+1}) \leq tol \tag{8.40}$$

and the conditions (8.35)-(8.39) are satisfied; $l_{k+1} := \mu(k+1) - \mu(k)$;
Compute d_k, $\hat{d}_k$, E_{k+1}, M_{k+1}, g_{k+1}, $\hat{g}_{k+1}$ and Θ_{k+1} (see (8.35)-(8.39));
$k := k+1$;

until there is no solution of the relations (8.35)-(8.40);
$b := k+1$; $\mu(b) := N$.

In the following Theorem 8.2 we describe how to construct the look-ahead Lanczos matrices from the output of the procedure 'generate H, V and D'.

Theorem 8.2 *Let $A \in \mathrm{Mat}_N$ and $v_1, w_1 \in \mathbb{C}^N \setminus \{0\}$ be given. Let the sequences $(P_k)_{1 \leq k \leq b-1}$, $(E_k)_{1 \leq k \leq b-1}$, $(\Theta_k)_{1 \leq k \leq b-1}$, $(d_k)_{0 \leq k \leq b-2}$, $(\hat{d}_k)_{0 \leq k \leq b-2}$, $(\mu(k))_{0 \leq k \leq b}$, M_{b-1}, g_{b-1} and $\hat{g}_{b-1}$ be generated by the procedure 'generate H, V and D'.*
Define

$$C_k := (e_1^T \hat{d}_k)\{\Theta_k^{-1} e_{l_k}\}\{\Theta_{k+1}^T e_1\}^T, \quad B_k := d_k e_{l_k}^T, \quad D^{(1)} := diag\,(\Theta_1, \Theta_2, \ldots, \Theta_{b-1})$$

$$H^{(1)} := \begin{bmatrix} E_1 & C_1 & & 0 \\ B_1 & E_2 & \ddots & \\ & \ddots & \ddots & C_{b-2} \\ 0 & & B_{b-2} & E_{b-1} \end{bmatrix}, \qquad \tilde{P}_k := \begin{bmatrix} I_{\mu(k-1)} & 0 \\ 0 & P_k \end{bmatrix}$$

$$\hat{H}^{(1)} := (D^{(1)}H^{(1)}(D^{(1)})^{-1})^T = \begin{bmatrix} \hat{E}_1 & \hat{C}_1 & & 0 \\ \hat{B}_1 & \hat{E}_2 & \ddots & \\ & \ddots & \ddots & \hat{C}_{b-2} \\ 0 & & \hat{B}_{b-2} & \hat{E}_{b-1} \end{bmatrix}$$

and $V^{(1)} := \tilde{P}_1\tilde{P}_2\cdots\tilde{P}_{b-1}[e_1, e_2, \ldots, e_{\mu(b-1)}]$. *Further define* $W^{(1)} \in \mathrm{Mat}_{N\times\mu(b-1)}$ *by the equation* $(W^{(1)})^TV^{(1)} = D^{(1)}$. *Then there is a (practical) look-ahead Lanczos method which can be carried out at least up to step* $\mu(b-1)$ *and which generates the Krylov tuples*

$$(A, V^{(1)}, \Pi, H^{(1)}, \mu(b-1)) \qquad \textit{and} \qquad (A^T, W^{(1)}, \Pi^T, \hat{H}^{(1)}, \mu(b-1)). \tag{8.41}$$

Here Π *is the Lanczos projection (see Theorem 8.1).*

If $\mu(b-1) = N$ *then the equations (8.10)-(8.14) are satisfied by the matrices defined above.*
If $\mu(b-1) < N$ *then the following matrices satisfy the equations (8.10)-(8.14):*

$$V := \tilde{P}_1\tilde{P}_2\cdots\tilde{P}_{b-1}, \qquad D := diag\,(D^{(1)}, I_{N-\mu(b-1)}), \qquad W := (V^{-1})^TD^T$$

and

$$H := \begin{bmatrix} E_1 & C_1 & & & 0 \\ B_1 & E_2 & \ddots & & \\ & \ddots & \ddots & C_{b-2} & \\ & & B_{b-2} & E_{b-1} & F_{b-1} \\ 0 & & & G_{b-1} & M_{b-1} \end{bmatrix}, \quad \hat{H} := \begin{bmatrix} \hat{E}_1 & \hat{C}_1 & & & 0 \\ \hat{B}_1 & \hat{E}_2 & \ddots & & \\ & \ddots & \ddots & \hat{C}_{b-2} & \\ & & \hat{B}_{b-2} & \hat{E}_{b-1} & \hat{F}_{b-1} \\ 0 & & & \hat{G}_{b-1} & M_{b-1}^T \end{bmatrix} \tag{8.42}$$

where

$$G_{b-1} := g_{b-1}e_{l_{b-1}}^T, \qquad F_{b-1} := \{\Theta_{b-1}^{-1}\hat{g}_{b-1}\}e_1^T$$
$$\hat{F}_{b-1} := \{(\Theta_{b-1}^{-1})^Tg_{b-1}\}e_1^T, \qquad \hat{G}_{b-1} := \hat{g}_{b-1}e_{l_{b-1}}^T.$$

Theorem 8.2 follows immediately from Lemma 8.1 and Theorem 8.1.

Up to now we have reduced the problem of generating matrices H and V which satisfy the equations (8.10)-(8.14) (for an appropriate choice of W) to the solution of 'two-block problems' of the following form:

Let a matrix $A \in \mathrm{Mat}_N$ and vectors $g_0, \tilde{g}_0 \in \mathbb{C}^N \setminus \{0\}$ be given.

Choose a number $l \in \{1, 2, \ldots, N\}$ and well-conditioned matrices $P \in \mathrm{Mat}_N$, and $Q \in \mathrm{Mat}_N$ such that the following relations hold:

a)

$$d := P^{-1} g_0 \in \mathrm{span}\{e_1\}. \tag{8.43}$$

b)

$$\hat{d} := Q^{-1} \tilde{g}_0 \in \mathrm{span}\{e_1\}. \tag{8.44}$$

c) The matrix $P^{-1}AP$ has the form

$$P^{-1}AP = \begin{bmatrix} E & F \\ G & M_1 \end{bmatrix} \tag{8.45}$$

where $G = g e_l^T$, $g \in \mathbb{C}^{N-l}$ and $E \in \mathrm{Mat}_l$ is an unreduced upper Hessenberg matrix.

d) The matrix $Q^{-1}A^TQ$ has the form

$$Q^{-1}A^TQ = \begin{bmatrix} \hat{E} & \hat{F} \\ \hat{G} & \hat{M}_1 \end{bmatrix} \tag{8.46}$$

where $\hat{G} = \hat{g} e_l^T$, $\hat{g} \in \mathbb{C}^{N-l}$ and $\hat{E} \in \mathrm{Mat}_l$ is an unreduced upper Hessenberg matrix.

e) There are matrices $\Theta \in \mathrm{Mat}_l$ and $\tilde{\Theta} \in \mathrm{Mat}_{N-l}$ such that

$$Q^T P = \mathrm{diag}\,(\Theta, \tilde{\Theta}). \tag{8.47}$$

Notice that the relations (8.36) and (8.44) are connected by $\tilde{g}_0 = \tilde{\Theta}_k^T \hat{g}_k$. In the following $\tilde{g}$ is defined by $\tilde{g} := \tilde{\Theta}^T \hat{g}$.

Now we turn to the solution of the two-block problem (8.43)-(8.47).

We make the ansatz

$$Q = \tilde{\mathcal{Q}}_0 \tilde{\mathcal{Q}}_1 \cdots \tilde{\mathcal{Q}}_l, \qquad P = \tilde{\mathcal{P}}_0 \tilde{\mathcal{P}}_1 \cdots \tilde{\mathcal{P}}_l$$

where $\tilde{\mathcal{Q}}_j$, $\tilde{\mathcal{P}}_j \in \mathrm{Mat}_N$ (for $j = 1, 2, \ldots, l$) are invertible matrices of the form

$$\tilde{\mathcal{Q}}_j = \begin{bmatrix} I_j & \mathcal{Q}_{j1} \\ 0 & \mathcal{Q}_{j2} \end{bmatrix}, \qquad \mathcal{Q}_{j2} \in \mathrm{Mat}_{N-j} \tag{8.48}$$

$$\tilde{\mathcal{P}}_j = \begin{bmatrix} I_j & \mathcal{P}_{j1} \\ 0 & \mathcal{P}_{j2} \end{bmatrix}, \qquad \mathcal{P}_{j2} \in \mathrm{Mat}_{N-j} \tag{8.49}$$

and $\mathcal{P}_0 = \mathcal{P}_{02}$, $\mathcal{Q}_0 = \mathcal{Q}_{02}$. One easily verifies that the first j column vectors of Q and P are uniquely determined by the matrices $\tilde{\mathcal{Q}}_0, \tilde{\mathcal{Q}}_1, \ldots, \tilde{\mathcal{Q}}_{j-1}$ and $\tilde{\mathcal{P}}_0, \tilde{\mathcal{P}}_1, \ldots, \tilde{\mathcal{P}}_{j-1}$, respectively.

We start with the matrices $\tilde{\mathcal{Q}}_0$ and $\tilde{\mathcal{P}}_0$. The matrices $\tilde{\mathcal{Q}}_0$ and $\tilde{\mathcal{P}}_0$ are used to fulfil the conditions (8.43) and (8.44):

$$\begin{aligned} Q^{-1}\tilde{g}_0 \in \operatorname{span}\{e_1\} &\iff \tilde{g}_0 \in \operatorname{span}\{Qe_1\} \\ &\iff \tilde{g}_0 \in \operatorname{span}\{\tilde{\mathcal{Q}}_0 e_1\} \\ &\iff \tilde{\mathcal{Q}}_0^{-1}\tilde{g}_0 \in \operatorname{span}\{e_1\} \end{aligned} \tag{8.50}$$

and

$$P^{-1}g_0 \in \operatorname{span}\{e_1\} \Longleftrightarrow \tilde{\mathcal{P}}_0^{-1}g_0 \in \operatorname{span}\{e_1\}. \tag{8.51}$$

Suppose that we have already generated matrices $\tilde{\mathcal{Q}}_0, \tilde{\mathcal{Q}}_1, \ldots, \tilde{\mathcal{Q}}_{j-1}$ and $\tilde{\mathcal{P}}_0, \tilde{\mathcal{P}}_1, \ldots, \tilde{\mathcal{P}}_{j-1}$ $(j \geq 1)$ such that

$$\begin{aligned} \mathcal{H}_j &:= \tilde{\mathcal{P}}_{j-1}^{-1}, \tilde{\mathcal{P}}_{j-2}^{-1}, \ldots, \tilde{\mathcal{P}}_0^{-1}A\tilde{\mathcal{P}}_0, \tilde{\mathcal{P}}_0, \ldots, \tilde{\mathcal{P}}_{j-1} \\ \hat{\mathcal{H}}_j &:= \tilde{\mathcal{Q}}_{j-1}^{-1}, \tilde{\mathcal{Q}}_{j-2}^{-1}, \ldots, \tilde{\mathcal{Q}}_0^{-1}A^T\tilde{\mathcal{Q}}_0, \tilde{\mathcal{Q}}_0, \ldots, \tilde{\mathcal{Q}}_{j-1} \end{aligned}$$

have the form

$$\mathcal{H}_j = \begin{bmatrix} \mathcal{E}_j & \mathcal{F}_j \\ \mathcal{G}_j & \mathcal{M}_j \end{bmatrix}, \qquad \hat{\mathcal{H}}_j = \begin{bmatrix} \hat{\mathcal{E}}_j & \hat{\mathcal{F}}_j \\ \hat{\mathcal{G}}_j & \hat{\mathcal{M}}_j \end{bmatrix} \tag{8.52}$$

where $\mathcal{E}_j, \hat{\mathcal{E}}_j \in \mathrm{Mat}_j$ are unreduced upper Hessenberg matrices and $\mathcal{G}_j = g^{(j)}e_j^T$, $\hat{\mathcal{G}}_j = \hat{g}^{(j)}e_j^T$ for appropriate vectors $g^{(j)}, \hat{g}^{(j)} \in \mathbb{C}^{N-j}$.

Now we construct matrices $\tilde{\mathcal{Q}}_j$ and $\tilde{\mathcal{P}}_j$ such that $\mathcal{H}_{j+1}$ and $\hat{\mathcal{H}}_{j+1}$ have the form (8.52).

The matrix $\mathcal{H}_{j+1}$ has the representation

$$\mathcal{H}_{j+1} = \tilde{\mathcal{P}}_j^{-1}\mathcal{H}_j\tilde{\mathcal{P}}_j = \begin{bmatrix} I_j & -\mathcal{P}_{j1}\mathcal{P}_{j2}^{-1} \\ 0 & \mathcal{P}_{j2}^{-1} \end{bmatrix} \begin{bmatrix} \mathcal{E}_j & \mathcal{F}_j \\ \mathcal{G}_j & \mathcal{M}_j \end{bmatrix} \begin{bmatrix} I_j & \mathcal{P}_{j1} \\ 0 & \mathcal{P}_{j2} \end{bmatrix} \tag{8.53}$$

$$= \begin{bmatrix} (\mathcal{E}_j - \mathcal{P}_{j1}\mathcal{P}_{j2}^{-1}\mathcal{G}_j) & \star \\ (\mathcal{P}_{j2}^{-1}\mathcal{G}_j) & \star \end{bmatrix}. \tag{8.54}$$

Hence the first equation in (8.52) holds (for $j+1$) if and only if $\mathcal{P}_j$ satisfys the relations

(1) $\mathcal{E}_j - \mathcal{P}_{j1}\mathcal{P}_{j2}^{-1}\mathcal{G}_j$ is upper Hessenberg.

(2) $\mathcal{P}_{j2}^{-1}\mathcal{G}_j = \alpha_j e_1 e_j^T$ for an appropriate $\alpha_j \in \mathbb{C}$.

Relation (1) holds for every choice of matrices $\mathcal{P}_{j1}$ and $\mathcal{P}_{j2}$ because $\mathcal{E}_j$ is upper Hessenberg and

$$\mathcal{E}_j - \mathcal{P}_{j1}\mathcal{P}_{j2}^{-1}\mathcal{G}_j = \mathcal{E}_j - (\mathcal{P}_{j1}\mathcal{P}_{j2}^{-1}g^{(j)})e_j^T.$$

Relation (2) is equivalent to the relation

$$\mathcal{P}_{j2}^{-1} g^{(j)} \in \text{span}\{e_1\}.$$

In a similar manner we obtain the condition

$$\mathcal{Q}_{j2}^{-1} \hat{g}^{(j)} \in \text{span}\{e_1\}$$

for the matrix $\mathcal{Q}_j$.
Let

$$m := \min\{\dim \mathcal{K}(N, A, g_0), \dim \mathcal{K}(N, A^T, \tilde{g}_0)\}.$$

Then the conditions (8.43)-(8.46) can be fulfilled for every number $l \in \{1, 2, \ldots, m\}$ by choosing

$$\mathcal{P}_{j2}^{-1} g^{(j)} \in \text{span}\{e_1\} \tag{8.55}$$

$$\mathcal{Q}_{j2}^{-1} \hat{g}^{(j)} \in \text{span}\{e_1\} \tag{8.56}$$

for $j = 0, 1, \ldots, l-1$. Notice that the conditions (8.43)-(8.46) hold (if (8.55) and (8.56) are fulfilled) for every choice of matrices $\mathcal{P}_{j1}$, $\mathcal{Q}_{j1}$ $(j = 1, \ldots, l)$ and for every choice of invertible matrices $\mathcal{P}_{l2}$ and $\mathcal{Q}_{l2}$.
The matrices $\mathcal{P}_{j2}$ and $\mathcal{Q}_{j2}$ are not uniquely determined by the conditons (8.55) and (8.56). For reasons of stability and simplicity we set

$$\mathcal{P}_{j1} := 0, \qquad \mathcal{Q}_{j1} := 0 \qquad \text{for} \quad j = 1, 2, \ldots, l-1$$

and choose appropriate unitary Housholder matrices

$$\begin{aligned} \mathcal{P}_{j2} &= I - 2x_j x_j^*, \qquad x_j \in \mathbb{C}^{N-j}, \quad \|x_j\| = 1 \\ \mathcal{Q}_{j2} &= I - 2y_j y_j^*, \qquad y_j \in \mathbb{C}^{N-j}, \quad \|y_j\| = 1 \end{aligned}$$

to satisfy the conditions (8.55) and (8.56) for $j = 0, 1, \ldots, l-1$.

It remains to fulfil the condition (8.47).

Define

$$\tilde{D}_l := \tilde{\mathcal{Q}}_{l-1}^T \cdots \tilde{\mathcal{Q}}_0^T \tilde{\mathcal{P}}_0 \cdots \tilde{\mathcal{P}}_{l-1}.$$

Consider the equation

$$\begin{aligned} Q^T P &= \tilde{D}_{l+1} = \begin{bmatrix} D_{l+1,1} & D_{l+1,2} \\ D_{l+1,3} & D_{l+1,4} \end{bmatrix} = \tilde{\mathcal{Q}}_l^T \tilde{D}_l \tilde{\mathcal{P}}_l \\ &= \begin{bmatrix} I_l & 0 \\ \mathcal{Q}_{l1}^T & \mathcal{Q}_{l2}^T \end{bmatrix} \begin{bmatrix} D_{l,1} & D_{l,2} \\ D_{l,3} & D_{l,4} \end{bmatrix} \begin{bmatrix} I_l & \mathcal{P}_{l1} \\ 0 & \mathcal{P}_{l2} \end{bmatrix} \\ &= \begin{bmatrix} D_{l,1} & D_{l,1}\mathcal{P}_{l1} + D_{l,2}\mathcal{P}_{l,2} \\ \mathcal{Q}_{l,1}^T D_{l,1} + \mathcal{Q}_{l,2}^T D_{l,3} & D_{l+1,4} \end{bmatrix} \end{aligned} \tag{8.57}$$

where

$$D_{l+1,4} = \mathcal{Q}_{l,1}^T(D_{l,1}\mathcal{P}_{l,1} + D_{l,2}\mathcal{P}_{l,2}) + \mathcal{Q}_{l,2}^T(D_{l,3}\mathcal{P}_{l,1} + D_{l,4}\mathcal{P}_{l,2}),$$

$D_{l,1} = D_{l+1,1} \in \mathrm{Mat}_l$ and $D_{l,4}, D_{l+1,4} \in \mathrm{Mat}_{N-l}$.
If $D_{l,1}$ is invertible then the equation (8.47) can be fulfilled by setting

$$\mathcal{P}_{l2} := I_{N-l}, \qquad \mathcal{Q}_{l2} := I_{N-l} \tag{8.58}$$

and

$$\mathcal{P}_{l1} := -D_{l,1}^{-1}D_{l,2}, \qquad \mathcal{Q}_{l1} := -D_{l,3}D_{l,1}^{-1}. \tag{8.59}$$

For numerical stability and efficiency we choose the smallest number l such that

$$\sigma_{\min}(D_{l,1}) \geq tol_1 \tag{8.60}$$

where $tol_1 \in (0,1]$ is chosen by the user. Notice that $\kappa_2(D_{l,1}) \leq (\sigma_{\min}(D_{l,1}))^{-1}$ because D_l is, by construction, the product of unitary matrices.
An easy calculation yields the estimates

$$\kappa_2(P) \leq \left(1 + \frac{1}{\sigma_{min}(D_{l,1})}\right)^2 \leq \left(1 + \frac{1}{tol_1}\right)^2 \tag{8.61}$$

$$\kappa_2(Q) \leq \left(1 + \frac{1}{\sigma_{min}(D_{l,1})}\right)^2 \leq \left(1 + \frac{1}{tol_1}\right)^2. \tag{8.62}$$

Putting these observations together, we obtain an algorithm (see procedure 'generate one block' below) for the computation of matrices P, Q E, M_1, Θ and vectors $g, \tilde{g}, d, \hat{d}$ satisfying the relations (8.43)-(8.47).

procedure generate one block

input: $N \in \mathbb{N}$; $A \in \mathrm{Mat}_N$; $g_0, \tilde{g}_0 \in \mathbb{C}^N$; $tol_1 \in (0,1]$;
output: $l \in \{1, 2, \ldots, N\}$; P, $Q \in \mathrm{Mat}_N$; $E \in \mathrm{Mat}_l$; $M_1 \in \mathrm{Mat}_{N-l}$; $\Theta \in \mathrm{Mat}_l$; $g, \tilde{g} \in \mathbb{C}^{N-l}$; $d, \hat{d} \in \mathrm{span}\{e_1\}$;

$\mathcal{H}_0 := A$; $\hat{\mathcal{H}}_0 := A^T$; $\tilde{D}_0 := I$; $g^{(0)} := g_0$; $\hat{g}^{(0)} := \tilde{g}_0$ $j := 0$;
repeat

if $j \geq 1$ **then** $\mathcal{P}_{j1} := 0$; $\mathcal{Q}_{j1} := 0$;
Construct unitary Housholder matrices $\mathcal{P}_{j2}$ and $\mathcal{Q}_{j2}$ satisfying the relations (8.55) and (8.56);
Compute $\mathcal{H}_{j+1} = \tilde{\mathcal{P}}_j^{-1}\mathcal{H}_j\tilde{\mathcal{P}}_j$; $\hat{\mathcal{H}}_{j+1} = \tilde{\mathcal{Q}}_j^{-1}\hat{\mathcal{H}}_j\tilde{\mathcal{Q}}_j$; $\tilde{D}_{j+1} := \tilde{\mathcal{Q}}_j^T\tilde{D}_j\tilde{\mathcal{P}}_j$;
if $j = 0$ **then** set $d := \tilde{\mathcal{P}}_0^{-1}g_0$ and $\hat{d} := \tilde{\mathcal{Q}}_0^{-1}\tilde{g}_0$;
$j := j + 1$

until $\sigma_{\min}(D_{j,1}) \geq tol_1$ **or** $j = N$ **or** $g^{(j)} = 0$ **or** $\hat{g}^{(j)} = 0$ (8.63)
$l := j$; $g := g^{(l)}$;
if $\sigma_{\min}(D_{l,1}) \geq tol_1$ **and** $l < N$ **then**
begin

Choose $\mathcal{P}_{l1}$, $\mathcal{P}_{l2}$, $\mathcal{Q}_{l1}$ and $\mathcal{Q}_{l2}$ as described in (8.58) and (8.59);
Compute $\mathcal{E}_{l+1}$, $\mathcal{M}_{l+1}$, $g^{(l+1)}$, $\hat{g}^{(l+1)}$ and $\tilde{D}_{l+1}$ (see (8.52))
Set $E := \mathcal{E}_{l+1}$, $M_1 := \mathcal{M}_{l+1}$, $\Theta := D_{l,1}$, $\tilde{\Theta} := D_{l+1,4}$, $g := g^{(l+1)}$, $\tilde{g} := \tilde{\Theta}^T \hat{g}^{(l+1)}$

end
else Set $E := \mathcal{E}_l$, $\Theta := D_{l,1}$.

Remark 8.2 The matrix $\tilde{D}_l$ is the product of unitary matrices. Hence $\kappa_2(\tilde{D}_l) = 1$. Notice that $\tilde{D}_l = D_{l,1}$ if $l = N$.

To obtain a numerically stable algorithm we replace the relation (8.63) by

until $\sigma_{\min}(D_{j,1}) \geq tol_1$ **or** $j = N$. (8.64)

We call this algorithm **'procedure compute one block'**.

Remark 8.3 If $\sigma_{\min}(D_{j,1}) \geq tol$ for some $j \geq \min\{\mathcal{K}(N, A, g_0), \mathcal{K}(N, A^T, \tilde{g}_0)\}$ then the procedure 'generate one block' and the procedure 'compute one block' generate identical solutions of the relations (8.43)-(8.47).
In the other case the matrices P, Q, E, M_1, Θ, $g, \tilde{g}, d, \hat{d}$, computed by the procedure 'compute one block', do not satisfy the relations (8.43)-(8.47). However, these matrices satisfy (8.43), (8.44), (8.47) and the relations c') and d') below.

c') The matrix $P^{-1}AP$ has the form

$$P^{-1}AP = \begin{bmatrix} E & F \\ G & M_1 \end{bmatrix} \tag{8.65}$$

where $G = ge_l^T$, $g \in \mathbb{C}^{N-l}$ and $E \in \mathrm{Mat}_l$ is an upper Hessenberg matrix.

d') The matrix $Q^{-1}A^TQ$ has the form

$$Q^{-1}A^TQ = \begin{bmatrix} \hat{E} & \hat{F} \\ \hat{G} & \hat{M}_1 \end{bmatrix} \tag{8.66}$$

where $\hat{G} = \hat{g}e_l^T$, $\hat{g} \in \mathbb{C}^{N-l}$ and $\hat{E} \in \mathrm{Mat}_l$ is an upper Hessenberg matrix.

Implementation details

In the following we give some implementation details for the efficient computation of $\mathcal{E}_{l+1}$, $\mathcal{M}_{l+1}$, g and $\tilde{g}$. Assume that $\sigma_{\min}(D_{l,1}) \geq tol_1$ and $l < N$. By construction the equations

$$\begin{bmatrix} \mathcal{E}_{l+1} & \mathcal{F}_{l+1} \\ \mathcal{G}_{l+1} & \mathcal{M}_{l+1} \end{bmatrix} = \begin{bmatrix} I_l & \mathcal{P}_{l1} \\ 0 & I_{N-l} \end{bmatrix}^{-1} \begin{bmatrix} \mathcal{E}_j & \mathcal{F}_j \\ \mathcal{G}_j & \mathcal{M}_j \end{bmatrix} \begin{bmatrix} I_l & \mathcal{P}_{l1} \\ 0 & I_{N-l} \end{bmatrix} \tag{8.67}$$

$$= \begin{bmatrix} (\mathcal{E}_l - \mathcal{P}_{l1}\mathcal{G}_l) & \star \\ \mathcal{G}_l & (G_l\mathcal{P}_{l1} + \mathcal{M}_l) \end{bmatrix}, \tag{8.68}$$

$$G_l = g^{(l)} e_l^T \qquad \text{and} \qquad \mathcal{P}_{l1} = -D_{l,1}^{-1} D_{l,2} \tag{8.69}$$

hold. Using (8.57)-(8.59) and (8.67)-(8.69), we obtain the representations

$$\mathcal{E}_{l+1} = \mathcal{E}_l + \{D_{l,1}^{-1} D_{l,2} g^{(l)}\} e_l^T = [\mathcal{E}_l e_1, \mathcal{E}_l e_2, \ldots, \mathcal{E}_l e_{l-1}, \mathcal{E}_l e_l + D_{l,1}^{-1} D_{l,2} g^{(l)}]$$

$$\mathcal{M}_{l+1} = \mathcal{M}_l - g^{(l)} \{D_{l,2}^T (D_{l,1}^{-1})^T e_l\}^T$$

$$g = g^{(l+1)} = g^{(l)}$$

$$\hat{g}^{(l+1)} = \hat{g}^{(l)}$$

and

$$\begin{aligned} \tilde{g} &= \tilde{\Theta}^T \hat{g}^{(l+1)} = \tilde{\Theta}^T \hat{g}^{(l)} = \tilde{D}_{l+1,4}^T \hat{g}^{(l)} \\ &= \{\mathcal{Q}_{l1}^T (D_{l,1}\mathcal{P}_{l1} + D_{l,2}\mathcal{P}_{l2}) + \mathcal{Q}_{l2}^T (D_{l,3}\mathcal{P}_{l1} + D_{l,4}\mathcal{P}_{l2})\}^T \hat{g}^{(l)} \\ &= (D_{l,3}\mathcal{P}_{l1} + D_{l,4})^T \hat{g}^{(l)} \\ &= D_{l,4}^T \hat{g}^{(l)} - D_{l,2}^T (D_{l,1}^{-1})^T D_{l,3}^T \hat{g}^{(l)}. \end{aligned}$$

For the computation of the singular values of the matrices $D_{1,1}, D_{2,1}, \ldots, D_{l-1,1}$ and the generation of the singular value decomposition of the matrix $D_{l,1}$ we use the SVD-algorithm of Golub and Reinsch [26], [28].
Notice, that the matrix $D_{l+1,4}$ need not be computed.

8.1 Reducing the computational effort

Considering equation (8.70) below we find that the procedure 'compute one block' and the procedure 'generate one block' in general compute one dimensional solutions (i.e. $l = 1$) of the two-block problem (8.43)-(8.47), (8.60) if the constant tol_1 is not chosen too large.

$$\sigma_{\min}(D_{1,1}) = |D_{1,1}| = |e_1^T Q^T P e_1| = |(Qe_1)^T (Pe_1)| = \frac{|\tilde{g}_0^T g_0|}{||\tilde{g}_0||\ ||g_0||} \tag{8.70}$$

In this situation the matrix P, generated by the procedure 'compute one block', is the product of the two elementary matrices $\mathcal{P}_0$ and $\mathcal{P}_1$. Now we show how P can be chosen as a single elementary matrix. For this purpose we solve the following problem:

Let the vectors g_0 and $\tilde{g}_0$ be given. Suppose that

$$g_0^T \tilde{g}_0 \neq 0. \tag{8.71}$$

Find invertible elementary matrices $P = I - \gamma x y^T$ and Q satisfying the relations (8.43)-(8.47) for $l = 1$.

The relations (8.45) and (8.46) are satisfied for every choice of invertible matrices P and Q. The equation (8.47) is equivalent to the identities

$$\frac{1}{\beta}\tilde{g}_0 = Qe_1 = (P^T)^{-1}(P^T Q)e_1 = \Theta(P^T)^{-1}e_1 \tag{8.72}$$

$$\frac{1}{\alpha}g_0 = Pe_1 = (Q^T)^{-1}(Q^T P)e_1 = \Theta(Q^T)^{-1}e_1 \tag{8.73}$$

where $\alpha, \beta \in \mathbb{C} \setminus \{0\}$. Hence the conditions (8.43)-(8.46) are satisfied if and only if the relations

$$P^{-1}g_0 \in \text{span}\{e_1\} \tag{8.74}$$
$$P^T\tilde{g}_0 \in \text{span}\{e_1\} \tag{8.75}$$
$$Q^{-1}\tilde{g}_0 \in \text{span}\{e_1\} \tag{8.76}$$
$$Q^T g_0 \in \text{span}\{e_1\} \tag{8.77}$$

hold. Now we proceed as follows: We construct an invertible elementary matrix $P = I - \gamma x y^T$ satisfying the relations (8.74) and (8.75) and set $Q = (P^{-1})^T$. Then the equations (8.43)-(8.47), are satisfied, too.
We recall that a matrix of the form $I - \gamma x y^T$ is invertible if and only if $\gamma(y^T x) \neq 1$. In this case the matrix $I - \hat{\gamma} x y^T$ with

$$\hat{\gamma} = \frac{\gamma}{\gamma(y^T x) - 1} \tag{8.78}$$

is the inverse of $I - \gamma x y^T$. The relations (8.74) and (8.75) are equivalent to the equations (8.78)-(8.80).

$$\alpha e_1 = g_0 - \hat{\gamma} x (y^T g_0) \tag{8.79}$$
$$\beta e_1 = \tilde{g}_0 - \gamma y (x^T \tilde{g}_0). \tag{8.80}$$

Here $\alpha, \beta, \gamma, \hat{\gamma} \in \mathbb{C} \setminus \{0\}$ and $x, y \in \mathbb{C}^N$.
(8.79) and (8.80) imply the identity

$$\alpha\beta = \alpha e_1^T e_1 \beta = g_0^T (P^{-1})^T P^T \tilde{g}_0 = g_0^T \tilde{g}_0. \tag{8.81}$$

One easily verifies that the equations (8.79) and (8.80) hold if the identities (8.81)-(8.85) are satisfied.

$$\begin{aligned} x &= g_0 - \alpha e_1 && (8.82)\\ 1 &= \hat{\gamma}(y^T g_0) && (8.83)\\ y &= \tilde{g}_0 - \beta e_1 && (8.84)\\ 1 &= \gamma(x^T \tilde{g}_0). && (8.85) \end{aligned}$$

Equation (8.86) below shows that the equations (8.81)-(8.85) imply the identity (8.78).

$$\begin{aligned} \gamma\hat{\gamma}(x^T y) &= \gamma\hat{\gamma}(g_0^T g_0 - \beta e_1^T g_0 - \alpha e_1^T \tilde{g}_0 + \alpha\beta)\\ &= \gamma\hat{\gamma}(g_0^T g_0 - \beta e_1^T g_0 - \alpha e_1^T \tilde{g}_0 + g_0^T \tilde{g}_0)\\ &= \gamma\hat{\gamma}(y^T g_0 + x^T \tilde{g}_0)\\ &= \gamma\hat{\gamma}(\frac{1}{\gamma} + \frac{1}{\hat{\gamma}}) = \gamma + \hat{\gamma}. && (8.86) \end{aligned}$$

Consequently, every solution $\alpha, \beta, \gamma, \hat{\gamma} \in \mathbb{C} \setminus \{0\}$, $x, y \in \mathbb{C}^N$ of (8.81)-(8.85) is a solution of (8.78)-(8.80).

Combining the equations (8.81)-(8.85) and using the assumption (8.71) we obtain the equivalent system of equations:

$$\begin{aligned} \alpha\beta &= g_0^T \tilde{g}_0 && (8.87)\\ \hat{\gamma} &= (g_0^T \tilde{g}_0)^{-1}(1 - \frac{e_1^T g_0}{\alpha})^{-1} && (8.88)\\ \gamma &= (g_0^T \tilde{g}_0 - \alpha e_1^T \tilde{g}_0)^{-1} && (8.89)\\ x &= g_0 - \alpha e_1 && (8.90)\\ y &= \tilde{g}_0 - \beta e_1. && (8.91) \end{aligned}$$

If $\alpha \in \mathbb{C} \setminus \{0\}$ is chosen such that

$$g_0^T \tilde{g}_0 \neq \alpha e_1^T \tilde{g}_0 \qquad \text{and} \qquad \alpha \neq e_1^T g_0 \tag{8.92}$$

then there are uniquely determined numbers $\gamma, \hat{\gamma}, \beta$ and vectors x and y satisfying the equations (8.87)-(8.91).
Our goal is to choose the scalar α so that the matrix $P = I - \gamma x y^T$ is well-conditioned. For this purpose we first derive a lower bound for the condition number of every invertible matrix $P \in \mathrm{Mat}_N$ satisfying the relations (8.74) and (8.75). By (8.81) there are scalars α and β such that $\alpha\beta = g_0^T \tilde{g}_0 \neq 0$,

$$||P|| \geq ||Pe_1|| = ||\frac{g_0}{\alpha}|| = \frac{||g_0||}{|\alpha|}, \tag{8.93}$$

and

$$||P^{-1}|| = ||(P^T)^{-1}|| \geq ||(P^T)^{-1}e_1|| = ||\frac{\tilde{g}_0}{\beta}|| = \frac{||\tilde{g}_0||}{|g_0^T\tilde{g}_0|}\,|\alpha|. \tag{8.94}$$

Combining the equations (8.93) and (8.94) we obtain the estimate

$$\kappa_2(P) = ||P||\;||P^{-1}|| \geq \frac{||g_0||}{|\alpha|}\frac{||\tilde{g}_0||}{|g_0^T\tilde{g}_0|}\,|\alpha| = \frac{||g_0||\;||\tilde{g}_0||}{|g_0^T\tilde{g}_0|}. \tag{8.95}$$

An easy calculation yields for invertible elementary matrices $P = I - \gamma xy^T$ the identity

$$\kappa_2(P) = a + \sqrt{a^2 - 1} \tag{8.96}$$

where

$$a = \frac{1}{2}|\gamma\hat{\gamma}|\;||x||^2\;||y||^2 - \Re\left(\frac{\gamma\hat{\gamma}}{|\gamma\hat{\gamma}|}\right) \tag{8.97}$$

and $\hat{\gamma}$ is defined in (8.78) above. Using the relations (8.96) and (8.97) we obtain in the case $\Re(\gamma\hat{\gamma}) \geq 0$ the estimates

$$\frac{|\gamma\hat{\gamma}|\;||x||^2\;||y||^2}{2} - 1 \leq \kappa_2(P) \leq |\gamma\hat{\gamma}|\;||x||^2\;||y||^2. \tag{8.98}$$

In the other case ($\Re(\gamma\hat{\gamma}) < 0$) the inequalities

$$\frac{|\gamma\hat{\gamma}|\;||x||^2\;||y||^2}{2} \leq \kappa_2(P) \leq |\gamma\hat{\gamma}|\;||x||^2\;||y||^2 + 2 \tag{8.99}$$

hold. Considering the equations (8.95), (8.98) and (8.99) we expect that the expression $|\gamma\hat{\gamma}|\;||x||^2\;||y||^2$ can be bounded by a polynomial of low degree in $||g_0||\;||\tilde{g}_0||/|g_0^T\tilde{g}_0|$ if the number α is appropriate chosen. Indeed, by choosing

$$\alpha = -2\;||g_0||\;e^{i(\varphi_1-\varphi_2)} \tag{8.100}$$

where $g_0^T\tilde{g}_0 = |g_0^T\tilde{g}_0|e^{i\varphi_1}$ and $e_1^T\tilde{g}_0 = |e_1^T\tilde{g}_0|e^{i\varphi_2}$ (if $e_1^T\tilde{g}_0 = 0$ then we take $\varphi_2 = 0$) we obtain the inequalities

$$\begin{aligned} |\gamma|^{-1} &= |g_0^T\tilde{g}_0 - \alpha e_1^T\tilde{g}_0| && (8.101)\\ &= |\;|g_0^T\tilde{g}_0|e^{i\varphi_1} + 2\;||g_0||\;e^{i(\varphi_1-\varphi_2)}|e_1^T\tilde{g}_0|e^{i\varphi_2}\;\;| \\ &= |g_0^T\tilde{g}_0| + 2\;||g_0||\;|e_1^T\tilde{g}_0| \\ &\geq |g_0^T\tilde{g}_0| && (8.102) \end{aligned}$$

$$\begin{aligned} |\hat{\gamma}|^{-1} &= |g_0^T\tilde{g}_0|\;|1 + \frac{e_1^Tg_0}{2\;||g_0||\;e^{i(\varphi_1-\varphi_2)}}| \\ &\geq |g_0^T\tilde{g}_0|\;|1 - \frac{|e_1^Tg_0|}{2\;||g_0||}| \\ &\geq \frac{|g_0^T\tilde{g}_0|}{2} && (8.103) \end{aligned}$$

which leads to the estimate

$$\begin{aligned}
\kappa_2(P) &\leq |\gamma|\,|\hat{\gamma}|\,||x||^2\,||y||^2 + 2 \leq \frac{2\,||x||^2\,||y||^2}{|g_0^T\tilde{g}_0|^2} + 2 \\
&\leq \frac{2\,(||g_0|| + |\alpha|)^2\,(||\tilde{g}_0|| + |\beta|)^2}{|g_0^T\tilde{g}_0|^2} + 2 \\
&\leq 2\frac{(||g_0|| + 2||g_0||)^2\,(||\tilde{g}_0|| + \frac{|g_0^T\tilde{g}_0|}{2||g_0||})^2}{|g_0^T\tilde{g}_0|^2} + 2 \\
&\leq 18\left(\frac{||g_0||\,||\tilde{g}_0||}{|g_0^T\tilde{g}_0|} + \frac{1}{2}\right)^2 + 2. \qquad (8.104)
\end{aligned}$$

Remark 8.4 If $g_0, \tilde{g}_0 \in \mathbb{R}^N$ then the relation (8.100) reads as follows:

$$\alpha = -2\,||g_0||\,e^{i(\varphi_1-\varphi_2)} = -2\,||g_0|| \cdot \text{signum}(\,(e_1^T\tilde{g}_0)(g_0^T\tilde{g}_0)\,). \qquad (8.105)$$

In this case ($g_0, \tilde{g}_0 \in \mathbb{R}^N$) the matrix $P = I - \gamma xy^T$ has real entries and the inequality $\gamma\hat{\gamma} > 0$ holds. This leads to the estimate

$$\kappa_2(P) \leq 18\left(\frac{||g_0||\,||\tilde{g}_0||}{|g_0^T\tilde{g}_0|} + \frac{1}{2}\right)^2 - 2.$$

The procedure 'compute one block of dimension 1' below summarizes how to compute in a stable and efficient way matrices P, Q, E, Θ and vectors $g, \tilde{g}, d, \hat{d}$ satisfying the relations (8.43)-(8.47) under the assumption

$$\frac{|g_0^T\tilde{g}_0|}{||g_0||\,||\tilde{g}_0||} \geq tol_2$$

where the constant $tol_2 \in (0, 1]$ is specified by the user.

procedure compute one block of dimension 1

input: $N \in \mathbb{N}$; $A \in \text{Mat}_N$; g_0, $\tilde{g}_0 \in \mathbb{C}^N \setminus \{0\}$; $g_0^T\tilde{g}_0$; $||g_0||$;

output: P, $Q \in \text{Mat}_N$; $E \in \text{Mat}_1$; $M_1 \in \text{Mat}_{N-1}$; $\Theta \in \text{Mat}_1$;
g, $\tilde{g} \in \mathbb{C}^{N-1}$; d, $\hat{d} \in \text{span}\{e_1\}$;

begin

$\alpha := -2\,||g_0||\,e^{i(\varphi_1-\varphi_2)}$; {where $g_0^T\tilde{g}_0 = |g_0^T\tilde{g}_0|e^{i\varphi_1}$ and $e_1^T\tilde{g}_0 = |e_1^T\tilde{g}_0|e^{i\varphi_2}$}
$\beta := g_0^T\tilde{g}_0/\alpha$;

$d := \alpha e_1$;
$\hat{d} := \beta e_1$; {Generation of $P = I - \gamma x y^T$ and $P^{-1} = I - \hat{\gamma} x y^T$}
$x := g_0$;
$e_1^T x := e_1^T g_0 - \alpha$;
$y := \tilde{g}_0$;
$e_1^T y := e_1^T \tilde{g}_0 - \beta$;
$\gamma := -1/(\alpha \cdot e_1^T y)$;
$\hat{\gamma} := -\alpha/(g_0^T \tilde{g}_0 \cdot e_1^T x)$;
$\Theta := 1$;
compute $P^{-1}AP$ and define E, M_1, g, and $\tilde{g}$ by

$$P^{-1}AP =: \begin{bmatrix} E & \tilde{g}^T \\ g & M_1 \end{bmatrix};$$

end .

Remark 8.5 (Stability) Considering the procedure 'compute one block of dimension 1' we see immediately that the computation of the numbers α, β, d, $\hat{d}$, Θ, γ and $\hat{\gamma}$ is well-conditioned.
The relations

$$|e_1^T g_0 - \alpha| \geq |\, 2||g_0|| - |e_1^T g_0|\, | \geq ||g_0|| \geq \frac{1}{3}(|e_1^T g_0| + 2\,||g_0||) = \frac{1}{3}(|e_1^T g_0| + |\alpha|)$$

and

$$\begin{aligned} |e_1^T \tilde{g}_0 - \beta| &= \left| e_1^T \tilde{g}_0 + \frac{g_0^T \tilde{g}_0}{2\,||g_0|| e^{i(\varphi_1 - \varphi_2)}} \right| = \left| |e_1^T \tilde{g}_0| e^{i\varphi_2} + \frac{|g_0^T \tilde{g}_0|}{2\,||g_0||} e^{i\varphi_2} \right| \\ &= |e_1^T \tilde{g}_0| + \frac{|g_0^T \tilde{g}_0|}{2\,||g_0||} = |e_1^T \tilde{g}_0| + |\beta| \end{aligned}$$

proves that the subtractions used in the computation of the vectors x and y are well-conditioned. Finally, the estimate (8.104) for the condition number of the matrix P shows that the computation of $P^{-1}AP$ is well-conditioned if the user choose the constant tol_2 not to small.

Now we are set to formulate the complete algorithm for the stable and efficient computation of the look-ahead Lanczos matrices H and D.

procedure compute H and D

input: $N \in \mathbb{N}$; $A \in \mathrm{Mat}_N$; v_1,$w_1 \in \mathbb{C}^N \setminus \{0\}$; $tol \in (0, 1]$;
output: $(\mu(k))_{0 \leq k \leq b} \subseteq \mathbb{N}$; H, $D \in \mathrm{Mat}_N$;

$g_0 := v_1$; $\tilde{g}_0 := w_1$; $M_0 := A$; $\tilde{\Theta}_0 := I$; $k := 0$; $\mu(0) := 0$;
repeat

if $(|g_k^T \tilde{g}_k|/(||g_k||\ ||\tilde{g}_k||)) \geq tol$ **then**
begin

compute one block of dimension 1 **input**: $N - \mu(k), M_k, g_k, \tilde{g}_k, g_k^T \tilde{g}_k, ||g_k||$;
output: P_{k+1}, Q_{k+1}, E_{k+1}, M_{k+1}, Θ_{k+1}, g_{k+1}, $\tilde{g}_{k+1}$, d_{k+1}, $\hat{d}_{k+1}$;

$l_{k+1} := 1$;
end
else
begin

compute one block **input**: $N - \mu(k)$, M_k, g_k, $\tilde{g}_k$, tol;
output: l_{k+1}, P_{k+1}, Q_{k+1}, E_{k+1}, M_{k+1}, Θ_{k+1}; g_{k+1}, $\tilde{g}_{k+1}$, d_{k+1}, $\hat{d}_{k+1}$;

end
$\mu(k+1) := \mu(k) + l_{k+1}$;

until $\mu(k+1) = N$;
$b := k+1$; $E_b := M_b$
for $k := 1$ **to** $b-1$ set
$C_k := (e_1^T \hat{d}_k)\{\Theta_k^{-1} e_{l_k}\}\{\Theta_{k+1}^T e_1\}^T$;
$B_k := d_k e_{l_k}^T$;
Set

$$H := \begin{bmatrix} E_1 & C_1 & & 0 \\ B_1 & E_2 & \ddots & \\ & \ddots & \ddots & C_{b-1} \\ 0 & & B_{b-1} & E_b \end{bmatrix}; \qquad D := \operatorname{diag}(\Theta_1, \Theta_2, \ldots, \Theta_b). \tag{8.106}$$

Theorem 8.3 below states the properties of the output of the procedure 'compute H and D'.

Theorem 8.3 *Let H, D, $(P_k)_{1 \leq k \leq b}$ and $(Q_k)_{1 \leq k \leq b}$ be generated by the procedure 'compute H and D'. For $k = 1, 2, \ldots, b$ define*

$$\tilde{P}_k := \begin{bmatrix} I_{\mu(k-1)} & 0 \\ 0 & P_k \end{bmatrix}, \qquad V := \tilde{P}_1 \tilde{P}_2 \cdots \tilde{P}_b$$

and $W := (V^{-1})^T D^T$. *Then the relations* (8.107)-(8.116) *below hold.*

$$
\begin{aligned}
Ve_1 &= v_1 && (8.107)\\
We_1 &= w_1 && (8.108)\\
H &= V^{-1}AV \quad \textit{is an upper Hessenberg matrix} && (8.109)\\
\hat{H} &= W^{-1}A^TW \quad \textit{is an upper Hessenberg matrix} && (8.110)\\
D &= W^TV = \mathit{diag}\,(\Theta_1, \Theta_2, \ldots, \Theta_b), \quad \Theta_j \in \mathrm{Mat}_{\mu(j)-\mu(j-1)}. && (8.111)\\
\kappa_2(V) &\leq b\left(18\left(\frac{1}{tol}+\frac{1}{2}\right)^2+2\right) && (8.112)\\
\kappa_2(Q_k) &\leq 18\left(\frac{1}{tol}+\frac{1}{2}\right)^2+2 \qquad (k=1,2,\ldots,b) && (8.113)\\
\sigma_{min}(D) &\geq tol && (8.114)\\
\sigma_{max}(D) &\leq 1 && (8.115)\\
\kappa_2(W) &\leq \frac{\kappa_2(V)}{tol}. && (8.116)
\end{aligned}
$$

Let $\tilde{b}$ *be the largest number such that* H *and* $\hat{H}$ *are unreduced* $\mu(\tilde{b})$*-Hessenberg matrices. Let* $H^{(1)}$ *denote the leading* $\mu(\tilde{b}) \times \mu(\tilde{b})$ *submatrix of* H. *Let* $\hat{H}^{(1)}$ *denote the leading* $\mu(\tilde{b}) \times \mu(\tilde{b})$ *submatrix of* $\hat{H}$. *Consider the partitions*

$$V = [V^{(1)}, V^{(2)}] \qquad \textit{and} \qquad W = [W^{(1)}, W^{(2)}]$$

where $V^{(1)}, W^{(1)} \in \mathrm{Mat}_{N\times\mu(\tilde{b})}$. *Then there is a (practical) look-ahead Lanczos method which generates the Krylov tuples*

$$(A, V^{(1)}, \Pi, H^{(1)}, \mu(\tilde{b})) \qquad \textit{and} \qquad (A^T, W^{(1)}, \Pi^T, \hat{H}^{(1)}, \mu(\tilde{b})) \tag{8.117}$$

where Π *is the projection with*

$$\mathrm{Im}\,\Pi = \mathcal{K}(\mu(\tilde{b}), A, v_1) \quad \textit{and} \quad \mathrm{Im}\,\Pi^T = \mathcal{K}(\mu(\tilde{b}), A^T, w_1)^{\perp}.$$

Theorem 8.3 follows directly from Theorem 8.1, Remark 8.2 and the estimates (8.61), (8.62) and (8.104).

In the following we discuss the output of the procedure 'compute H and D' in greater detail. Let the matrix H be generated by the procedure 'compute H and D'. Choose the smallest number $m \in \{1, 2, \ldots, N\}$ such that

$$H = \begin{bmatrix} H^{(1)} & F \\ G & H^{(2)} \end{bmatrix}$$

where

$$F = 0 \qquad \text{or} \qquad G = 0 \tag{8.118}$$

and $H^{(1)} \in \mathrm{Mat}_m$ is unreduced upper Hessenberg.
Using Lemma 8.2, Theorem 8.2, Theorem 5.14 and Corollary 5.5 we find immediately the following cases:

a) If $N = m$ then there is a practical look-ahead Lanczos method which generates the matrix $H^{(1)}$ when applied to v_1, w_1 and A. Further, if A has pairwise distinct eigenvalues, then the vectors v_1 and w_1 have 'general directions' with respect to A and $\langle x, y\rangle = x^T y$ (see Definition 7.3).

b) If $\mu(\tilde{b}) = m$ then there is a practical look-ahead Lanczos method which generates the matrix $H^{(1)}$ when applied to v_1, w_1 and A.

c) If $\mu(\tilde{b}) < m$ then in general there is no look-ahead Lanczos method, starting from v_1, w_1 and A, which generates the matrix $H^{(1)}$.

8.2 Flop count

There are at least two different definitions of 'flop' used in the mathematical literature:

a) Old flop: One flop is an arithmetic operation of the form $a + b \cdot c$, i.e., a floating point add and a floating point multiply (see e.g. [27]).

b) New flop: One flop is a single floating point operation (see e.g. [28]).

Notice that an 'old flop' involves two 'new flops'. In this thesis the definition b) is used.

For every starting matrix $A \in \mathrm{Mat}_N(\mathbb{R})$ and for every choice of v_1, $w_1 \in \mathbb{R}^N \setminus \{0\}$ and $tol \in (0, 1]$ the procedure 'compute H and D' requires less than

$$\begin{aligned} \frac{8}{3}N^3 \;&+\; \left(8\left(15 - \frac{7}{l_{max}}\right)s + 7\right)N^2 + \left(8s^2 - 9s + \frac{37}{3}\right)N \\ &+\; 40s^3 + 9s^2 + \left(\frac{2}{3}l_{max}^3 + O(l_{max}^2)\right)s \end{aligned} \tag{8.119}$$

flops for the computation of the look-ahead Lanczos matrices $H \in \mathrm{Mat}_N(\mathbb{R})$ and $D \in \mathrm{Mat}_N(\mathbb{R})$. Here and below the number l_{max} is the size of the largest diagonal blocks E_j of the matrix H and the number s is the sum of the sizes of all diagonal blocks of H which are not 1×1 matrices. In other words

$$l_{max} = \max_{1 \le k \le b} \underbrace{(\mu(k) - \mu(k-1))}_{=l_k} \tag{8.120}$$

and

$$s = \sum_{k=1, l_k \neq 1}^{b} l_k. \tag{8.121}$$

Notice that the identities $s = 0$ and $l_{max} = 1$ hold if the matrix H is tridiagonal.

For later use we consider the case that the starting matrix $A \in \mathrm{Mat}_N(\mathbb{R})$ and the starting vectors v_1, $w_1 \in \mathbb{R}^N \setminus \{0\}$ can be partitioned as follows:

$$A = \begin{bmatrix} \tilde{E}_1 & \tilde{C}_1 & & 0 \\ \tilde{B}_1 & \tilde{E}_2 & \ddots & \\ & \ddots & \ddots & \tilde{C}_{\tilde{b}-1} \\ 0 & & B_{\tilde{b}-1} & \tilde{E}_{\tilde{b}} \end{bmatrix}, \qquad v_1 = \begin{bmatrix} v_1^{(1)} \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \qquad w_1 = \begin{bmatrix} w_1^{(1)} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{8.122}$$

where $\tilde{E}_j \in \mathrm{Mat}_{\tilde{l}_j}$ are upper Hessenberg matrices for $j = 2, \ldots, \tilde{b}$, the matrix E_1 is an arbitrary $\tilde{l}_1 \times \tilde{l}_1$ matrix, the matrices $\tilde{B}_1, \tilde{B}_2, \ldots \tilde{B}_{\tilde{b}-1}$ have exactly one non-zero element in the upper right corner and $v_1^{(1)}, w_1^{(1)} \in \mathbb{R}^{\tilde{l}_1}$. In this situation the procedure 'compute H and D' requires only

$$\begin{aligned} N(8\tilde{s}^2 + 4\tilde{s}\tilde{l}_{max} + 7\tilde{s} + 10) \quad &+ \quad s(19\tilde{s}^2 + ((8\tilde{l}_{max} - 4)l_{max} - 8\,\tilde{l}_{max} + 18)\tilde{s}) \\ &+ \quad s\left(\frac{8}{3}\tilde{l}_{max}^2 l_{max}^2 + 3\tilde{l}_{max}^2 + \frac{2}{3}l_{max}^3 + O(\tilde{l}_{max} l_{max}^2)\right) \end{aligned} \tag{8.123}$$

flops. The numbers s and l_{max} are defined in (8.120) and (8.121). The number $\tilde{s}$ is the sum of the sizes of all diagonal blocks $\tilde{E}_j$ of A which are not 1×1 matrices. The number $\tilde{l}_{max}$ is the size of the largest diagonal blocks $\tilde{E}_2, \tilde{E}_3 \ldots \tilde{E}_{\tilde{b}}$ of the matrix A. In other words

$$\tilde{l}_{max} = \max_{2 \leq k \leq \tilde{b}} \tilde{l}_k \tag{8.124}$$

and

$$\tilde{s} = \sum_{k=1, \tilde{l}_k \neq 1}^{\tilde{b}} \tilde{l}_k. \tag{8.125}$$

In our numerical experiments (see Subsection 11.2) the look-ahead Lanczos matrices H are near to tridiagonal matrices (i.e. l_{max} and s are small numbers) if the number tol is chosen small enough (e.g. for $N \leq 100$ and $tol = 10^{-4}$ we have observed $l_{max} \leq 6$ and $s \leq 8$). Combining these observations with the expressions (8.119) and (8.123) we find the following:

- For nearly tridiagonal starting matrices $A \in \mathrm{Mat}_N(\mathbb{R})$ and starting vectors v_1, $w_1 \in \mathbb{R}^N$ having the form (8.122) the procedure 'compute H and D' requires typically

$$O(N) \tag{8.126}$$

flops.

- For full matrices $A \in \mathrm{Mat}_N(\mathbb{R})$ the procedure 'compute H and D' requires typically

$$\frac{8}{3}N^3 + O(N^2) \tag{8.127}$$

  flops.

In comparison we note the flop counts of two well known methods used for the reduction of full matrices $A \in \mathrm{Mat}_N(\mathbb{R})$ to upper Hessenberg form:

- The reduction to upper Hessenberg form via Housholder transformations (see e.g. [28]) requires $10N^3/3 + O(N^2)$ flops.
- The reduction to upper Hessenberg form via Gauss transformations (see e.g. [28]) requires $4N^3/3 + O(N^2)$ flops.

9 The restarted similarity look-ahead Lanczos algorithm

We recall (see Subsection 7.2) that the restarted look-ahead Lanczos method, discussed in Section 7, is not a good choice for the computation of eigenvalues because of instabilities. In the following algorithm these instabilities are avoided by using the procedure 'compute H and D' (see Section 8) for the computation of the look-ahead Lanczos matrices.

Restarted similarity look-ahead Lanczos algorithm

(1) Let $A \in \mathrm{Mat}_N$ be given. Choose two starting vectors $v_1^{(0)}$, $w_1^{(0)} \in \mathbb{C}^N \setminus \{0\}$ and a number $tol \in (0,1]$. Set $i := 0$, $k := 0$, $A_k := A$, $N_i := N$, $n(k) := 0$.

(2) Apply the procedure 'compute H and D' on N_i, A_k, $v_1^{(i)}$, $w_1^{(i)}$, tol to obtain a block tridiagonal upper Hessenberg matrix $H_{(i+1)} \in \mathrm{Mat}_{N_i}$ and an invertible block diagonal matrix $D_{(i)} \in \mathrm{Mat}_{N_i}$.
Choose the smallest number $N_{i+1} \in \{1, 2, \ldots, N_i\}$ such that

$$H_{(i+1)} = \begin{bmatrix} H_{(i+1)}^{(1)} & F_{(i+1)} \\ G_{(i+1)} & H_{(i+1)}^{(2)} \end{bmatrix}$$

where

$$F_{(i+1)} = 0 \qquad \text{or} \qquad G_{(i+1)} = 0 \tag{9.1}$$

and $H_{(i+1)}^{(1)} \in \mathrm{Mat}_{N_{i+1}}$ is unreduced upper Hessenberg.
If $N_{i+1} \leq 2$ then goto (5).
Compute the characteristic polynomial ψ_i of

$$[H_{(i+1)}^{(1)}(j,l)]_{N_{i+1}-1 \leq j,l \leq N_{i+1}},$$

the trailing 2×2 submatrix of $H_{(i+1)}^{(1)}$.
If $N_{i+1} = N_i$ and $i \geq 1$ then goto (3) else goto (4).

(3) Compute $v_1^{(i+1)} := \psi_i(A_k) v_1^{(i)}$.
Compute $w_1^{(i+1)} := \psi_i(A_k^T) w_1^{(i)}$.
Set $i := i + 1$.
Goto (2).

(4) Set $A_{k+1} := H_{(i+1)}^{(1)}$.
Compute $v_1^{(i+1)} := \psi_i(A_{k+1}) e_1$.
Compute $w_1^{(i+1)} := \psi_i(A_{k+1}^T)[I_{N_{i+1}}, 0] D_{(i)}^T e_1$.

Set $k := k + 1$.
Set $i := i + 1$.
Set $n(k) := i$.
Goto (2).

(5) Set $A_k := H^{(1)}_{(i+1)}$.
end.

Remark 9.1 (Reduction to upper Hessenberg block tridiagonal form) To reduce the computational effort the starting matrix A is reduced in step $i = 0$ to upper Hessenberg block tridiagonal form ($A \to A_1$). In the following steps ($i \geq 1$) only matrices with this structure occur.

The restarted similarity look-ahead Lanczos algorithm (described above) generates sequences $(n(k))_{0\leq k\leq k_0}$, $(A_k)_{0\leq k\leq k_0}$, $(H_{(i+1)})_{i\in X}$ and $(D_{(i)})_{i\in X}$ where $k_0 < N$ and $X = \{0, 1, 2, \ldots\} \subseteq \mathbb{N}_0$. The numbers $n(0) = 0 < n(1) = 1 < n(2) < n(3) \cdots < n(k_0)$ are those iteration indices i at which a block decouples (see (9.1) and (4)).
If X is a finite set then the matrix A_{k_0} is a 1×1 or a 2×2 matrix. In this very unlikely case we have computed after a finite number $n(k_0)$ of steps at least one exact eigenvalue of the matrix A. In the other case ($X = \mathbb{N}_0$) there is, by Theorem 9.1 below, an identically restarted practical look-ahead Lanczos method, using the generalized Rayleigh-quotient shift strategy, which generates the same matrices $(H_{(i+1)})_{i\in\mathbb{N}_0\setminus\{0,1,\ldots,n(k_0)-1\}}$ and $(D_{(i)})_{i\in\mathbb{N}_0\setminus\{0,1,\ldots,n(k_0)-1\}}$ when applied to $v_1^{(n(k_0))}$, $w_1^{(n(k_0))}$ and A_{k_0}.

Theorem 9.1 *Let $A \in \mathrm{Mat}_N$. Choose starting vectors $v_1^{(0)}, w_1^{(0)} \in \mathbb{C} \setminus \{0\}$. Choose tol $\in (0, 1]$. Let $(n(k))_{0\leq k\leq k_0}$, $(A_k)_{0\leq k\leq k_0}$, $(H_{(i+1)})_{i\in X}$ and $(D_{(i)})_{i\in X}$ be the matrices computed by the restarted similarity look-ahead Lanczos algorithm applied to $v_1^{(0)}$, $w_1^{(0)}$, A. Let $V_{(i)}$ be the transformation matrix generated by the procedure 'compute H and D' in the i-th step of the restarted similarity look-ahead Lanczos process. Then the characteristic polynomials of the matrices A_k, $H_{(i+1)}$, $H^{(1)}_{(i+1)}$ and $H^{(2)}_{(i+1)}$ are divisors of the characteristic polynomial of the matrix A.*
Suppose that the restarted similarity look-ahead Lanczos method does not stop in finitely many steps (i.e $X = \mathbb{N}_0$). Define the matrix $W_{(i)} \in \mathrm{Mat}_{N_{i+1}\times N_i}$ by the equation $D_{(i)} = W_{(i)}^T V_{(i)}$ and $\hat{H}_{(i+1)} := (D_{(i)} H_{(i+1)} (D_{(i)})^{-1})^T$.
Then the identities

$$N_{n(k_0)} = N_i = m_i := \dim \mathcal{K}(N, A_{k_0}, V_{(i)}e_1) = \dim \mathcal{K}(N, A_{k_0}^T, W_{(i)}e_1) \tag{9.2}$$

hold for all $i \in \mathbb{N}_0 \setminus \{0, 1, \ldots, n(k_0) - 1\}$.
Furthermore there is an identically restarted practical look-ahead Lanczos method, using

the generalized Rayleigh-quotient shift strategy, which generates, when applied to A_{k_0}, $v_1^{(n(k_0))}$, $w_1^{(n(k_0))}$, *the restarted Krylov tuples*

$$((A_{k_0}, V_{(i)}, \Pi_{(i)}, H_{(i+1)}, m_i, \psi_i))_{i \in \mathbb{N}_0 \setminus \{0,1,\ldots,n(k_0)-1\}}$$

and

$$((A_{k_0}^T, W_{(i)}, \Pi_{(i)}^T, \hat{H}_{(i+1)}, m_i, \psi_i))_{i \in \mathbb{N}_0 \setminus \{0,1,\ldots,n(k_0)-1\}} \tag{9.3}$$

where $\Pi_{(i)} = I$. *Further the following estimates hold for all* $i \in \mathbb{N}_0$:

$$\kappa_2(D_{(i)}) \leq \frac{1}{tol} \tag{9.4}$$

$$\kappa_2(V_{(i)}) \leq N \cdot \left(18 \left(\frac{1}{tol} + \frac{1}{2}\right)^2 + 2\right) \tag{9.5}$$

$$\kappa_2(W_{(i)}) \leq \frac{\kappa_2(V_{(i)})}{tol}. \tag{9.6}$$

Finally, if the matrix A *is diagonalizable then the starting vectors* $v_1^{(n(k_0))}$ *and* $w_1^{(n(k_0))}$ *have dual directions with respect to the matrix* A_{k_0} *and the form* $\langle x, y \rangle = x^T y$.

Theorem 9.1 follows immediately from Theorem 8.3.

Combining Theorem 6.4, Theorem 7.2 and Theorem 9.1 we find the following convergence properties of the restarted similarity look-ahead Lanczos method:

Theorem 9.2 (Convergence) *Let* $(H_{(i)})_{i \in \mathbb{N}}$ *be the matrices generated by the restarted look-ahead Lanczos method. Consider the partition*

$$H_{(i)} = \begin{bmatrix} H_{11}^{(i)} & H_{12}^{(i)} \\ H_{21}^{(i)} & H_{22}^{(i)} \end{bmatrix}, \qquad H_{22}^{(i)} \in \mathrm{Mat}_2.$$

Then the following statements hold:

a) **(Convergence)**
If the shift polynomials ψ_i *fulfil the conitions a) and b) of Theorem 6.4 then the sequence* $(H_{21}^{(i)})_{i \in \mathbb{N}}$ *tends to zero.*

b) **(Quadratic convergence)**
Suppose that the sequence $(H_{21}^{(i)})_{i \in \mathbb{N}}$ *tends to zero. Suppose further that the matrix* A *has pairwise distinct eigenvalues. Then the sequence* $(H_{21}^{(i)})_{i \in \mathbb{N}}$ *tends to zero at least quadratically.*

c) **(Existence of regular vectors)**
Suppose that the sequence $(H_{21}^{(i)})_{i\in\mathbb{N}}$ *tends to zero. Suppose further that the matrix* A *has pairwise distinct eigenvalues. If the number* $tol \in (0,1]$ *is chosen small enough then there is an* $i_1 \in \mathbb{N}$ *such that the block diagonal matrices* $(D_{(i)}^{(1)})_{i\geq i_1}$ *can be partitioned as follows:*

$$D_{(i)}^{(1)} = diag\,(\Theta_{(i)}^{(1)}, \Theta_{(i)}^{(2)}) \qquad where \quad \Theta_{(i)}^{(2)} \in \mathrm{Mat}_2. \tag{9.7}$$

d) **(Cubic convergence)**
Suppose that the sequence $(H_{21}^{(i)})_{i\in\mathbb{N}}$ *tends to zero. Suppose further that the matrix* A *has pairwise distinct eigenvalues. If the number* $tol \in (0,1]$ *is chosen small enough (i.e. such that* (9.7) *holds) then the sequences* $(H_{21}^{(i)})_{i\in\mathbb{N}}$ *and* $(H_{12}^{(i)})_{i\in\mathbb{N}}$ *tend to zero cubically.*

In all our numerical experiments (see Subsection 11.2) we have observed that for the choice $tol = 10^{-4}$ the matrices $D_{(i)}^{(1)}$ can be partitioned as described in (9.7).

Remark 9.2 (Convergence of the QR algorithm) We recall that the QR algorithm (with Rayleigh-quotient shifts) converges typically quadratically and not cubically when applied to non-normal matrices. Notice further that there are starting matrices A for which the QR algorithm with Rayleigh-quotient shifts does not converge (see e.g. [80]).

Next we present an implementation of the restarted similarity look-ahead Lanczos method for the computation of all eigenvalues of an arbitrary $N \times N$ matrix A. For shortness of notation we call this algorithm RSL method

This algorithm computes a sequence $(\tilde{H}_{(i)})_{i\in X}$, $(X \subseteq \mathbb{N}_0)$ of $N \times N$ upper Hessenberg matrices such that there are invertible matrices $G_i \in \mathrm{Mat}_N$ and matrices $R_i \in \mathrm{Mat}_N$ satisfying the relations

$$A + R_i = G_i \tilde{H}_{(i+1)} G_i^{-1} \qquad \text{and} \qquad \|R_{(i)}\| \leq K \cdot tol_3 \tag{9.8}$$

for all $i \in X$. The numbers K and $\kappa_2(G_i)$ depend on the number tol (see below). The numbers $tol, tol_3 \in (0,1]$ are chosen by the user.
We say that the RSL algorithm converges if it terminates after finitely many steps (i.e. $X = \{0, 1, \ldots, i_0\}$). In this case the computed upper Hessenberg matrix $\tilde{H}_{(i_0+1)}$ has the form

$$\tilde{H}_{(i_0+1)} = \begin{bmatrix} \tilde{H}^{(1)} & \star & & \star \\ 0 & \tilde{H}^{(2)} & \ddots & \\ & \ddots & \ddots & \star \\ 0 & & 0 & \tilde{H}^{(q)} \end{bmatrix}, \qquad \tilde{H}^{(p)} = \begin{bmatrix} E_{p,1} & 0 & & 0 \\ B_{p,1} & E_{p,2} & \ddots & \\ & \ddots & \ddots & 0 \\ 0 & & B_{p,j_p-1} & E_{p,j_p} \end{bmatrix} \tag{9.9}$$

where the diagonal blocks $E_{p,i}$ are 1×1 or 2×2 matrices for all $p \in \{1, 2, \ldots, q\}$ and $i \in \{1, 2, \ldots, j_p\}$.
The eigenvalues of the matrix $\tilde{H}_{(i_0+1)}$ (i.e. the eigenvalues of the matrices $E_{p,i}$) are used as approximations to eigenvalues of A.

An implementation of the restarted similarity look-ahead Lanczos method (*RSL* algorithm)

(0) Let $A \in \mathrm{Mat}_N$ be given. Choose numbers tol, $tol_3 \in (0, 1]$.

(1) Choose two starting vectors $v_1^{(0)}$, $w_1^{(0)} \in \mathbb{C}^N \setminus \{0\}$.

(2) Set $i := 0$, $k := 0$, $A_k := A$, $\tilde{H}_{(i)} := A$, $N_i := N$, $j_i := 0$, $n(k) := 0$.

(3) Set $v_1^{(i)} := v_1^{(i)}/\|v_1^{(i)}\|$ and $w_1^{(i)} := w_1^{(i)}/\|w_1^{(i)}\|$.

(4) Apply the procedure 'compute H and D' on N_i, A_k, $v_1^{(i)}$, $w_1^{(i)}$, tol to obtain an invertible block diagonal matrix $D_{(i)} \in \mathrm{Mat}_{N_i}$ and a block tridiagonal upper Hessenberg matrix

$$H_{(i+1)} = \begin{bmatrix} E_1^{(i+1)} & C_1^{(i+1)} & & 0 \\ B_1^{(i+1)} & E_2^{(i+1)} & \ddots & \\ & \ddots & \ddots & C_{b_{i+1}-1}^{(i+1)} \\ 0 & & B_{b_{i+1}-1}^{(i+1)} & E_{b_{i+1}}^{(i+1)} \end{bmatrix} \in \mathrm{Mat}_{N_i} \tag{9.10}$$

of index $(\mu_{i+1}(l))_{0 \le l \le b_{i+1}}$.

(5) Set to zero all subdiagonal elements $H_{(i+1)}(j+1, j)$ that satisfy

$$|H_{(i+1)}(j+1, j)| \le tol_3 \tag{9.11}$$

and all submatrices $C_j^{(i+1)}$ that satisfy

$$\|C_j^{(i+1)}\| \le tol_3. \tag{9.12}$$

(6) Set $\tilde{H}_{(i+1)} := \tilde{H}_{(i)}$ and update

$$\tilde{H}_{(i+1)} = \begin{bmatrix} \tilde{H}_{(i)}^{(1)} & \star & \star \\ \star & H_{(i)} & \star \\ 0 & \star & \tilde{H}_{(i)}^{(3)} \end{bmatrix} \in \mathrm{Mat}_N \qquad (\text{where} \quad \tilde{H}_{(i)}^{(1)} \in \mathrm{Mat}_{j_i})$$

by replacing $H_{(i)}$ with $H_{(i+1)}$.

(7) Find the smallest number $j_{i+1} \in \{j_i, j_i+1, \ldots, N-3\}$ and the smallest number $N_{i+1} \geq 3$ such that $\tilde{H}_{(i+1)}$ has one of the following forms:

$$\tilde{H}_{(i+1)} = \left[\begin{array}{c|cc} \tilde{H}^{(1)}_{(i+1)} & \star & \star \\ \hline 0 & \tilde{H}^{(2)}_{(i+1)} & 0 \\ 0 & \star & \tilde{H}^{(3)}_{(i+1)} \end{array}\right], \qquad \tilde{H}_{(i+1)} = \left[\begin{array}{c|cc} \tilde{H}^{(1)}_{(i+1)} & 0 & 0 \\ \hline \star & \tilde{H}^{(2)}_{(i+1)} & 0 \\ 0 & \star & \tilde{H}^{(3)}_{(i+1)} \end{array}\right]$$

$$\tilde{H}_{(i+1)} = \left[\begin{array}{cc|c} \tilde{H}^{(1)}_{(i+1)} & \star & \star \\ 0 & \tilde{H}^{(2)}_{(i+1)} & \star \\ \hline 0 & 0 & \tilde{H}^{(3)}_{(i+1)} \end{array}\right], \qquad \tilde{H}_{(i+1)} = \left[\begin{array}{cc|c} \tilde{H}^{(1)}_{(i+1)} & 0 & \star \\ \star & \tilde{H}^{(2)}_{(i+1)} & \star \\ \hline 0 & 0 & \tilde{H}^{(3)}_{(i+1)} \end{array}\right].$$

Here $\tilde{H}^{(2)}_{(i+1)}$ is an unreduced $N_{i+1} \times N_{i+1}$ upper Hessenberg matrix and $\tilde{H}^{(1)}_{(i+1)}$ is a $j_{i+1} \times j_{i+1}$ upper Hessenberg matrix.

(8) If there are no numbers j_{i+1} and N_{i+1} satisfying the relations (7) then goto (12).
Compute the characteristic polynomial ψ_i of

$$[\tilde{H}^{(2)}_{(i+1)}(j,l)]_{N_{i+1}-1 \leq j,l \leq N_{i+1}}, \tag{9.13}$$

the trailing 2×2 submatrix of $\tilde{H}^{(2)}_{(i+1)}$.
If $j_i = j_{i+1}$, $N_{i+1} = N_i$ and $k \geq 1$ then goto (9).
Set $A_{k+1} := \tilde{H}^{(2)}_{(i+1)}$.
Set $n(k+1) := i+1$.
If $l_i = l_{i+1}$ then goto (10) else goto (11).

(9) Compute $v_1^{(i+1)} := \psi_i(A_k)v_1^{(i)}$.
Compute $w_1^{(i+1)} := \psi_i(A_k^T)w_1^{(i)}$.
Set $i := i+1$.
Goto (3).

(10) Compute $v_1^{(i+1)} := \psi_i(A_{k+1})e_1$.
If $\mu_{l+1}(1) \leq N_{i+1}$ then compute $w_1^{(i+1)} := \psi_i(A_{k+1}^T)[I_{N_{i+1}}, 0]D_{(i)}^T e_1$
else compute $w_1^{(i+1)} := \psi_i(A_{k+1}^T)e_1$.
Set $k := k+1$, $i := i+1$.
Goto (3).

(11) Compute $v_1^{(i+1)} := \psi_i(A_{k+1})e_1$.
Compute $w_1^{(i+1)} := \psi_i(A_{k+1}^T)e_1$.
Set $k := k+1$, $i := i+1$.
Goto (3).

(12) Set $k_0 := k + 1$.
Set $n(k_0) := i + 1$.
Set $i_0 := i$.
end.

Remark 9.3 (Tridiagonal starting matrices) If the starting matrix A is tridiagonal we set $H_{(1)} := A$, $D_{(0)} := I$ and do not execute (1), (3) and (4) in step $i = 0$.

Remark 9.4 (Real matrices) If the matrix A has real entries and the starting vectors in (1) are chosen with real entries then the complete algorithm can be carried out in real arithmetic.

Now we turn to the proof of the relations (9.8).
Let $A \in \mathrm{Mat}_N$. Let the starting vectors (see (1)) and the numbers $tol, tol_3 \in (0, 1]$ be chosen arbitrarily. Let $(\tilde{H}_{(i+1)})_{i \in X}$ be the computed upper Hessenberg matrices. Let $V_{(i)}$ be the transformation matrix generated by the procedure 'compute H and D' in the i-th step of the RSL algorithm. Define for all $i \in X$

$$\check{V}_i := \begin{bmatrix} I_{j_i} & 0 & 0 \\ 0 & V_{(i)} & 0 \\ 0 & & I_{N-N_i-j_i} \end{bmatrix}.$$

Then, by construction, there is a matrix $\check{R}_{n(k)} \in \mathrm{Mat}_N$ such that

$$\check{V}_{n(k)-1}^{-1} \tilde{H}_{(n(k-1))} \check{V}_{n(k)-1} = \tilde{H}_{(n(k))} + \check{R}_{n(k)} \qquad \text{and} \qquad ||\check{R}_{n(k)}|| \leq tol_3 \tag{9.14}$$

for all $k \in \{1, 2, \ldots, k_0\}$. Furthermore the equation

$$\check{V}_i^{-1} \tilde{H}_{(n(k))} \check{V}_i = \tilde{H}_{(i+1)} \tag{9.15}$$

holds for every $i \in X \setminus \{n(1) - 1, n(2) - 1, \ldots, n(k_0) - 1\}$. Here and in the following $k \in \{0, 1, 2, \ldots, k_0\}$ is the greatest number such that $n(k) \leq i$. Define for every $i \in X$ the matrices

$$G_0 := \check{V}_0, \qquad G_i := \check{V}_{n(1)-1} \check{V}_{n(2)-1} \ldots \check{V}_{n(k)-1} \check{V}_i.$$

Using the relations (9.14) and (9.15) we find by induction on i the representation

$$A + R_i = G_i \tilde{H}_{(i+1)} G_i^{-1} \tag{9.16}$$

where

$$R_i = -\sum_{l=1}^{k} G_{n(l)-1} \check{R}_{n(l-1)} G_{n(l)-1}^{-1}.$$

One easily verifies the estimate

$$||R_i|| \leq \sum_{l=1}^{k} \kappa_2(G_{n(l)-1})||\check{R}_{n(l-1)}||. \tag{9.17}$$

Using the inequalities (9.17), $k \leq k_0 \leq N$, $||\check{R}_{n(l-1)}|| \leq tol_3$,

$$\kappa_2(G_i) \leq \kappa_2(\check{V}_{n(1)-1})\kappa_2(\check{V}_{n(2)-1}) \cdots \kappa_2(\check{V}_{n(k)-1})\kappa_2(\check{V}_i),$$

$$\kappa_2(G_{n(l)-1}) \leq \kappa_2(\check{V}_{n(1)-1})\kappa_2(\check{V}_{n(2)-1}) \cdots \kappa_2(\check{V}_{n(l-1)-1})\kappa_2(\check{V}_{n(l)-1})$$

and the estimate (8.112) we obtain the inequalities

$$\begin{aligned} ||R_i|| &\leq tol_3 \cdot \sum_{l=1}^{k} \alpha^l = \left(\frac{1-\alpha^{k+1}}{1-\alpha} - 1\right) \cdot tol_3 \\ &\leq \left(\frac{1-\alpha^{N+1}}{1-\alpha} - 1\right) \cdot tol_3 \end{aligned} \tag{9.18}$$

and

$$\kappa_2(G_i) \leq \alpha^{k+1} \leq \alpha^{k_0+1} \leq \alpha^{N+1} \tag{9.19}$$

where

$$\alpha = N\left(18\left(\frac{1}{tol} + \frac{1}{2}\right)^2 + 2\right). \tag{9.20}$$

Putting the relations (9.16), (9.18), (9.19) and (9.20) together we obtain the desired representation (9.8).

Using the representation (9.8), the inequality (9.18) and the perturbation theorems of Golub and Van Loan [28] Theorem 7.2.2, Theorem 7.2.3 we find Theorem 9.3 below.

Theorem 9.3 *Let $A \in \mathrm{Mat}_N$. Choose starting vectors $v_1^{(0)}, w_1^{(0)} \in \mathbb{C} \setminus \{0\}$. Choose $tol, tol_3 \in (0,1]$. Let $(\tilde{H}_{(i+1)})_{i \in X}$ be the sequence of upper Hessenberg matrices computed by the RSL algorithm. Let the number α be defined by* (9.20).

a) Let $Q^*AQ = diag(\lambda_1, \lambda_2, \ldots, \lambda_N) + L$ be a Schur decomposition of the matrix A. Let $\beta^{(i+1)}$ be an eigenvalue of $\tilde{H}_{(i+1)}$. Let p be the smallest positive integer such that $|L|^p = 0$. Then there is an eigenvalue λ_j of A such that

$$|\beta^{(i+1)} - \lambda_j| \leq \max\{\theta, \theta^{1/p}\}$$

where

$$\theta = \left(\frac{1-\alpha^{N+1}}{1-\alpha} - 1\right) \cdot tol_3 \cdot \sum_{k=0}^{p-1} ||L||^k.$$

b) Suppose there is a matrix $Y \in \mathrm{Mat}_N$ *such that* $Y^{-1}AY = diag\,(\lambda_1, \lambda_2, \ldots, \lambda_N)$. *Let* $\beta^{(i+1)}$ *be an eigenvalue of* $H_{(i+1)}$. *Then there is an eigenvalue* λ_j *of* A *such that*

$$|\beta^{(i+1)} - \lambda_j| \leq \kappa_2(Y) \cdot \left(\frac{1 - \alpha^{N+1}}{1 - \alpha} - 1 \right) \cdot tol_3.$$

Implementation details

By construction (see procedure 'compute one block') the matrices $C_j^{(i+1)}$ have the form

$$C_j^{(i+1)} = c_{j,1}^{(i+1)} (c_{j,2}^{(i+1)})^T$$

where $c_{j,1}^{(i+1)}$ and $c_{j,2}^{(i+1)}$ are suitable vectors. Thus the numbers $\|C_j^{(i+1)}\|$ (see (9.12)) can be computed with less effort by $\|C_j^{(i+1)}\| = \|c_{j,1}^{(i+1)}\| \, \|c_{j,2}^{(i+1)}\|$.

In (9) the new starting vector $v_1^{(i+1)}$ is computed as follows:

Let $\psi_i(x) = x^2 + b_i x + c_i$ be the characteristic polynomial of the matrix (9.13).
Compute $y := A_k v_1^{(i)}$.
Compute $y := A_k y + b_i y$.
Compute $v_1^{(i+1)} := y + c_i v_1^{(i)}$.

The new starting vectors $v_1^{(i+1)}$ and $w_1^{(i+1)}$ in (10) and (11) are computed in a similar manner.

10 The look-ahead Lanczos GR algorithm

In this section we present a new and stable implicit GR algorithm which is mathematically equivalent to the identically restarted practical look-ahead Lanczos algorithm discussed in Subsection 7.1.

Look-ahead Lanczos GR algorithm

(1) Let $A \in \mathrm{Mat}_N$ be given. Choose two starting vectors $v_1^{(0)}$, $w_1^{(0)} \in \mathbb{C}^N \setminus \{0\}$ and a number $tol \in (0,1]$. Set $i := 0$, $H_{(0)}^{(1)} := A$, $N_i := N$.

(2) Apply the procedure 'compute H and D' on N_i, $H_{(i)}^{(1)}$, $v_1^{(i)}$, $w_1^{(i)}$, tol to obtain a block tridiagonal upper Hessenberg matrix $H_{(i+1)} \in \mathrm{Mat}_{N_i}$ and an invertible block diagonal matrix $D_{(i)} \in \mathrm{Mat}_{N_i}$.
Choose the smallest number $N_{i+1} \in \{1, 2, \ldots, N_i\}$ such that

$$H_{(i+1)} = \begin{bmatrix} H_{(i+1)}^{(1)} & F_{(i+1)} \\ G_{(i+1)} & H_{(i+1)}^{(2)} \end{bmatrix}$$

where

$$F_{(i+1)} = 0 \qquad \text{or} \qquad G_{(i+1)} = 0 \tag{10.1}$$

and $H_{(i+1)}^{(1)} \in \mathrm{Mat}_{N_{i+1}}$ is unreduced upper Hessenberg.
If $N_{i+1} \leq 2$ then goto (4).
Compute the characteristic polynomial ψ_i of

$$[H_{(i+1)}^{(1)}(j,l)]_{N_{i+1}-1 \leq j,l \leq N_{i+1}},$$

the trailing 2×2 submatrix of $H_{(i+1)}^{(1)}$.

(3) Compute $v_1^{(i+1)} := \psi_i(H_{(i+1)}^{(1)})e_1$.
Compute $w_1^{(i+1)} := \psi_i((H_{(i+1)}^{(1)})^T)[I_{N_{i+1}}, 0]D_{(i)}^T e_1$.
Set $i := i + 1$.
Goto (2).

(4) Set $i_0 := i + 1$.
end.

Let $(H_{(i+1)}^{(1)})_{i \in X}$ and $(H_{(i+1)})_{i \in X}$ be the sequences generated by the look-ahead Lanczos GR method described above.
The set $X = \{0, 1, 2, \ldots\} \subseteq \mathbb{N}_0$ is finite, say $X = \{0, 1, 2, \ldots, i_0\}$, if and only if $H_{(i_0)}^{(1)}$ is

a 1×1 or a 2×2 matrix. In this very unlikely case this algorithm has computed after a finite number i_0 of steps at least one exact eigenvalue of the matrix A.
In the other case ($X = \mathbb{N}_0$) there is a number $i_1 \in \mathbb{N}_0$ and an identically restarted practical look-ahead Lanczos method, using the generalized Rayleigh-quotient shift strategy, which generates the same matrices $(H_{(i+1)})_{i \in \mathbb{N}_0 \setminus \{0,1,\ldots,i_1-1\}}$ when applied to $H_{(i_1)}$, $v_1^{(i_1)}$ and $w_1^{(i_1)}$. This is the content of Theorem 10.1 below.

Theorem 10.1 *Let $A \in \mathrm{Mat}_N$. Choose starting vectors $v_1^{(0)}, w_1^{(0)} \in \mathbb{C} \setminus \{0\}$. Choose $tol \in (0,1]$. Let $(H_{(i+1)})_{i \in X}$ and $(D_{(i)})_{i \in X}$ be the matrices computed by the look-ahead Lanczos GR algorithm applied to $v_1^{(0)}$, $w_1^{(0)}$, A. Let $V_{(i)}$ be the transformation matrix generated by the procedure 'compute H and D' in the i-th step of the look-ahead Lanczos GR process. Then the characteristic polynomials of the matrices $H_{(i+1)}$, $H_{(i+1)}^{(1)}$ and $H_{(i+1)}^{(2)}$ are divisors of the characteristic polynomial of the matrix A.*
Suppose that the look-ahead Lanczos GR algorithm does not stop in finitely many steps (i.e $X = \mathbb{N}_0$). Then there is a number i_1 such that $N_i = N_{i+1}$ for all $i \in \mathbb{N}_0 \setminus \{0,1,\ldots,i_1-1\}$. Define the matrix $W_{(i)} \in \mathrm{Mat}_{N_i}$ by the equation $D_{(i)} = W_{(i)}^T V_{(i)}$.
Define further

$$\tilde{V}_{(i_1-1)} := I, \qquad \tilde{V}_{(i)} := V_{(i_1)} V_{(i_1+1)} \cdots V_{(i)}, \qquad \tilde{W}_{(i)} := (\tilde{V}_{(i-1)}^{-1})^T W_{(i)} \tag{10.2}$$

and

$$\hat{H}_{(i+1)} := (D_{(i)} H_{(i+1)} (D_{(i)})^{-1})^T \tag{10.3}$$

for all $i \in \mathbb{N}_0 \setminus \{0,1,\ldots,i_1-1\}$.
Then the identities

$$N_{i_1} = N_i = m_i := \dim \mathcal{K}(N, H_{(i_1)}, \tilde{V}_{(i)} e_1) = \dim \mathcal{K}(N, H_{(i_1)}^T, \tilde{W}_{(i)} e_1) \tag{10.4}$$

hold for all $i \in \mathbb{N}_0 \setminus \{0,1,\ldots,i_1-1\}$.
Furthermore, there is an identically restarted look-ahead Lanczos method, using the generalized Rayleigh-quotient shift strategy, which generates, when applied to $H_{(i_1)}$, $v_1^{(i_1)}$, $w_1^{(i_1)}$, the restarted Krylov tuples

$$((H_{(i_1)}, \tilde{V}_{(i)}, \Pi_{(i)}, H_{(i+1)}, m_i, \psi_i))_{i \in \mathbb{N}_0 \setminus \{0,1,\ldots,i_1-1\}} \tag{10.5}$$

and

$$((H_{(i_1)}^T, \tilde{W}_{(i)}, \Pi_{(i)}^T, \hat{H}_{(i+1)}, m_i, \psi_i))_{i \in \mathbb{N}_0 \setminus \{0,1,\ldots,i_1-1\}} \tag{10.6}$$

where $\Pi_{(i)} = I$. The inequalities

$$\kappa_2(D_{(i)}) \leq \frac{1}{tol} \tag{10.7}$$

$$\kappa_2(V_{(i)}) \leq N \cdot \left(18 \left(\frac{1}{tol} + \frac{1}{2}\right)^2 + 2\right) \tag{10.8}$$

$$\kappa_2(W_{(i)}) \leq \frac{\kappa_2(V_{(i)})}{tol} \tag{10.9}$$

are valid for all $i \in \mathbb{N}_0$.
The following estimates hold for all $i \in \mathbb{N}_0 \setminus \{0, 1, \ldots, i_1 - 1\}$*:*

$$\kappa_2(\tilde{V}_{(i)}) \leq \left(N \cdot \left(18 \left(\frac{1}{tol} + \frac{1}{2} \right)^2 + 2 \right) \right)^{(i-i_1+1)} \tag{10.10}$$

$$\kappa_2(\tilde{W}_{(i)}) \leq \frac{\kappa_2(\tilde{V}_{(i)})}{tol}. \tag{10.11}$$

Finally, if the matrix A *is diagonalizable, then the starting vectors* $v_1^{(i_1)}$ *and* $w_1^{(i_1)}$ *have dual directions with respect to the matrix* $H_{(i_1)}$ *and the form* $\langle x, y \rangle = x^T y$.

Proof. To prove that there is an identically restarted look-ahead Lanczos method, which generates (10.5) and (10.6), it is sufficient to show a)-f) below for all $i \in \mathbb{N}_0 \setminus \{0, 1 \ldots, i_1 - 1\}$.

a) The matrices $D_{(i)}$ are invertible and equation (10.4) holds.

b) $H_{(i+1)}$ and $\hat{H}_{(i+1)}$ are unreduced upper Hessenberg matrices.

c) $\tilde{V}_{(i)} H_{(i+1)} = \Pi_{(i)} H_{(i_1)} \tilde{V}_{(i)}$.

d) $\tilde{V}_{(i+1)} e_1 = \psi_i(H_{(i_1)}) \tilde{V}_{(i)} e_1$.

e) $\tilde{W}_{(i)} \hat{H}_{(i+1)} = \Pi_{(i)}^T H_{(i_1)}^T \tilde{W}_{(i)}$.

f) $\tilde{W}_{(i+1)} e_1 = \psi_i(H_{(i_1)}^T) \tilde{W}_{(i)} e_1$.

By Theorem 8.3 the matrices $D_{(i)}$ (for all $i \in \mathbb{N}_0$) are invertible, the inequalities (10.7)-(10.9) hold (for all $i \in \mathbb{N}_0$) and the matrices $H_{(i+1)}$ and $\hat{H}_{(i+1)}$ are unreduced upper Hessenberg (for $i \geq i_1$) and satisfy the equations

$$H_{(i+1)} = V_{(i)}^{-1} H_{(i)} V_{(i)} \qquad \text{and} \qquad \hat{H}_{(i+1)} = W_{(i)}^{-1} H_{(i)}^T W_{(i)}. \tag{10.12}$$

One easily verifies the equation

$$\tilde{W}_{(i)}^T \tilde{V}_{(i)} = W_{(i)} \tilde{V}_{(i-1)}^{-1} \tilde{V}_{(i)} = W_{(i)}^T V_{(i)} = D_{(i)}. \tag{10.13}$$

With assistance of the equations (10.3), (10.12) and (10.13) we find inductively the representations

$$H_{(i+1)} = \tilde{V}_{(i)}^{-1} H_{(i_1)} \tilde{V}_{(i)} \qquad \text{and} \qquad \hat{H}_{(i+1)} = \tilde{W}_{(i)}^{-1} H_{(i_1)}^T \tilde{W}_{(i)}. \tag{10.14}$$

Using the equations (10.3), (10.12), (10.13), (10.14), the unreduced upper Hessenberg structure of the matrices $H_{(i+1)}$ and $\hat{H}_{(i+1)}$ and the invertibility of the matrices $\tilde{V}_{(i)}$ and $\tilde{W}_{(i)}$ we obtain the identities

$$\begin{aligned} \mathcal{K}(N, H_{(i_1)}, \tilde{V}_{(i)}e_1) &= \mathcal{K}(N, \tilde{V}_{(i)}H_{(i+1)}\tilde{V}_{(i)}^{-1}, \tilde{V}_{(i)}e_1) \\ &= \tilde{V}_{(i)}\mathcal{K}(N, H_{(i+1)}, \tilde{V}_{(i)}^{-1}\tilde{V}_{(i)}e_1) = \mathbb{C}^{N_{i_1}} \end{aligned} \tag{10.15}$$

and

$$\begin{aligned} \mathcal{K}(N, H_{(i_1)}^T, \tilde{W}_{(i)}e_1) &= \mathcal{K}(N, \tilde{W}_{(i)}\hat{H}_{(i+1)}\tilde{W}_{(i)}^{-1}, \tilde{W}_{(i)}e_1) \\ &= \tilde{W}_{(i)}\mathcal{K}(N, \hat{H}_{(i+1)}, \tilde{W}_{(i)}^{-1}\tilde{W}_{(i)}e_1) = \mathbb{C}^{N_{i_1}}. \end{aligned} \tag{10.16}$$

The identities (10.15) and (10.16) imply the equations (10.4).
The Lanczos projection $\Pi_{(i)}$ is defined by

$$\operatorname{Im}\Pi_{(i)} = \mathcal{K}(N, H_{i_1}, \tilde{V}_{(i)}e_1) \quad \text{and} \quad \operatorname{Ker}\Pi_{(i)}^T = \mathcal{K}(N, H_{i_1}^T, \tilde{W}_{(i)}e_1)^\perp.$$

In view of the equations (10.15) and (10.16) the identity $\Pi_{(i)} = I$ holds. Putting these observations together we have proved a), b), c) and e). The equations

$$\begin{aligned} \psi_i(H_{(i_1)})\tilde{V}_{(i)}e_1 &= \psi_i(\tilde{V}_{(i)}H_{(i+1)}\tilde{V}_{(i)}^{-1})\tilde{V}_{(i)}e_1 = \tilde{V}_{(i)}\psi_i(H_{(i+1)})\tilde{V}_{(i)}^{-1}\tilde{V}_{(i)}e_1 = \tilde{V}_{(i)}V_{(i+1)}e_1 \\ &= \tilde{V}_{(i+1)}e_1 \end{aligned}$$

and

$$\begin{aligned} \psi_i(H_{(i_1)}^T)\tilde{W}_{(i)}e_1 &= \psi_i(H_{(i_1)}^T)(\tilde{V}_{(i)}^T)^{-1}D_{(i)}^T e_1 = (\tilde{V}_{(i)}^T)^{-1}\psi_i(\tilde{V}_{(i)}^T H_{(i_1)}^T(\tilde{V}_{(i)}^T)^{-1})D_{(i)}^T e_1 \\ &= (\tilde{V}_{(i)}^T)^{-1}\psi_i(H_{(i+1)}^T)D_{(i)}^T e_1 = (\tilde{V}_{(i)}^T)^{-1}W_{(i+1)}e_1 \\ &= \tilde{W}_{(i+1)}e_1 \end{aligned}$$

prove d) and f). The equations (10.13) together with the estimates (10.7), (10.8) and

$$\kappa_2(\tilde{V}_{(i)}) \leq \kappa_2(V_{(i_1)})\kappa_2(V_{(i_1+1)})\cdots\kappa_2(V_{(i)}) \tag{10.17}$$

yield the estimates (10.10) and (10.11). Finally the equations (10.15) and (10.16) imply that the starting vectors $v_1^{(i_1)}$ and $w_1^{(i_1)}$ have dual directions with respect to the matrix $H_{(i_1)}$ and the form $\langle x, y\rangle = x^T y$. ∎

We discuss common properties and an important difference between the restarted similarity look-ahead Lanczos method and the look-ahead Lanczos GR algorithm: By Theorem 9.1 and Theorem 10.1 both methods generate sequences of upper Hessenberg matrices such that there are identically restarted look-ahead Lanczos methods which generate the identical sequences of matrices. Further the generated starting vectors have dual directions. Another common property is that the condition numbers of all transformation

matrices used in both algorithms are bounded above by a number which is specified by the user. Now we turn to the most important difference between these methods. In the case of the restarted similarity look-ahead Lanczos method the condition numbers of the transformation matrices $V_{(i)}$ and $W_{(i)}$ of the accompanying identically restarted look-ahead Lanczos method are bounded above by a number which is chosen by the user (see Theorem 9.1). In case of the look-ahead Lanczos GR algorithm the condition numbers of the transformation matrices $\tilde{V}_{(i)}$ and $\tilde{W}_{(i)}$ of the accompanying identically restarted look-ahead Lanczos method may increase with i (see Theorem 10.1). Notice that the condition numbers of the matrices $\tilde{V}_{(i)}$ and $\tilde{W}_{(i)}$ are not relevant for the stability of the look-ahead Lanczos GR algorithm because these matrices are not used in the look-ahead Lanczos GR process. However, for the convergence properties of the look-ahead Lanczos GR algorithm the condition numbers of these matrices are important. Indeed, Theorem 6.4 and Theorem 7.2 yield the following convergence results for the look-ahead Lanczos GR algorithm:

Theorem 10.2 (Convergence) *Let $A \in \mathrm{Mat}_N$ be given. Choose two arbitrary starting vectors $v_1^{(0)}$, $w_1^{(0)} \in \mathbb{C} \setminus \{0\}$ and a number tol $\in (0,1]$. Let $(H_{(i+1)})_{i \in X}$ be the upper Hessenberg matrices computed by the look-ahead Lanczos GR algorithm applied to $v_1^{(0)}$, $w_1^{(0)}$, A.*
We recall that the look-ahead Lanczos GR algorithm terminates after a finite number i_0 of steps (i.e $X = \{0, 1, \ldots, i_0\}$) only if the algorithm has found at least one exact eigenvalue of the matrix A.
Now we turn to the other case ($X = \mathbb{N}_0$).
Let the number i_1 and let the matrices $\tilde{V}_{(i)}$ be defined as in Theorem 10.1.
Suppose that there is a constant K such that

$$\kappa_2(\tilde{V}_{(i)}) \leq K \tag{10.18}$$

for all $i \in \mathbb{N}_0 \setminus \{0, 1, \ldots, i_1 - 1\}$.
Consider the partition

$$H_{(i)} = \begin{bmatrix} H_{11}^{(i)} & H_{12}^{(i)} \\ H_{21}^{(i)} & H_{22}^{(i)} \end{bmatrix}, \qquad H_{22}^{(i)} \in \mathrm{Mat}_2.$$

Then the following statements hold:

a) **(Convergence)**
If the shift polynomials ψ_i fulfil the conitions a) and b) of Theorem 6.4 then the sequence $(H_{21}^{(i)})_{i \in \mathbb{N}}$ tends to zero.

b) **(Quadratic convergence)**
Suppose that the sequence $(H_{21}^{(i)})_{i \in \mathbb{N}}$ tends to zero. Suppose further that the ma-

trix A has distinct eigenvalues. Then the sequence $(H_{21}^{(i)})_{i\in\mathbb{N}}$ tends to zero at least quadratically.

c) **(Existence of regular vectors)**
Suppose that the sequence $(H_{21}^{(i)})_{i\in\mathbb{N}}$ tends to zero. Suppose further that the matrix A has distinct eigenvalues. If the number $tol \in (0,1]$ is chosen small enough then there is an $i_2 \in \mathbb{N}_0$ such that the block diagonal matrices $(D_{(i)}^{(1)})_{i\geq i_2}$ can be partitioned as follows:

$$D_{(i)}^{(1)} = diag\,(\Theta_{(i)}^{(1)}, \Theta_{(i)}^{(2)}) \qquad where \quad \Theta_{(i)}^{(2)} \in \mathrm{Mat}_2. \tag{10.19}$$

d) **(Cubic convergence)**
Suppose that the sequence $(H_{21}^{(i)})_{i\in\mathbb{N}}$ tends to zero. Suppose further that the matrix A has distinct eigenvalues. If the number $tol \in (0,1]$ is chosen small enough (i.e. such that (10.19) holds) then the sequences $(H_{21}^{(i)})_{i\in\mathbb{N}}$ and $(H_{12}^{(i)})_{i\in\mathbb{N}}$ tend to zero cubically.

In all our numerical experiments (see Subsection 11.2) we have observed cubic convergence of the look-ahead Lanczos GR algorithm for the choice $tol = 10^{-4}$.

By small changes in the RSL algorithm we obtain the following implementation of the look-ahead Lanczos GR algorithm.

An implementation of the look-ahead Lanczos GR method (LGR method)

(0) Let $A \in \mathrm{Mat}_N$ be given. Choose numbers tol, $tol_3 \in (0,1]$.

(1) Choose two starting vectors $v_1^{(0)}$, $w_1^{(0)} \in \mathbb{C}^N \setminus \{0\}$.

(2) Set $i := 0$, $\tilde{H}_{(i)}^{(2)} := A$, $\tilde{H}_{(i)} := A$, $N_i := N$, $j_i := 0$.

(3) Set $v_1^{(i)} := v_1^{(i)}/\|v_1^{(i)}\|$ and $w_1^{(i)} := w_1^{(i)}/\|w_1^{(i)}\|$.

(4) Apply the procedure 'compute H and D' on N_i, $\tilde{H}_{(i)}^{(2)}$, $v_1^{(i)}$, $w_1^{(i)}$, tol to obtain an invertible block diagonal matrix $D_{(i)} \in \mathrm{Mat}_{N_i}$ and a block tridiagonal upper Hessenberg matrix

$$H_{(i+1)} = \begin{bmatrix} E_1^{(i+1)} & C_1^{(i+1)} & & 0 \\ B_1^{(i+1)} & E_2^{(i+1)} & \ddots & \\ & \ddots & \ddots & C_{b_{i+1}-1}^{(i+1)} \\ 0 & & B_{b_{i+1}-1}^{(i+1)} & E_{b_{i+1}}^{(i+1)} \end{bmatrix} \in \mathrm{Mat}_{N_i} \tag{10.20}$$

of index $(\mu_{i+1}(l))_{0\leq l\leq b_{i+1}}$.

(5) Set to zero all subdiagonal elements $H_{(i+1)}(j+1,j)$ that satisfy

$$|H_{(i+1)}(j+1,j)| \leq tol_3 \tag{10.21}$$

and all submatrices $C_j^{(i+1)}$ that satisfy

$$\|C_j^{(i+1)}\| \leq tol_3. \tag{10.22}$$

(6) Set $\tilde{H}_{(i+1)} := \tilde{H}_{(i)}$ and update

$$\tilde{H}_{(i+1)} = \begin{bmatrix} \tilde{H}_{(i)}^{(1)} & \star & \star \\ \star & H_{(i)} & \star \\ 0 & \star & \tilde{H}_{(i)}^{(3)} \end{bmatrix} \in \mathrm{Mat}_N \qquad (\text{where} \quad \tilde{H}_{(i)}^{(1)} \in \mathrm{Mat}_{j_i})$$

by replacing $H_{(i)}$ with $H_{(i+1)}$.

(7) Find the smallest number $j_{i+1} \in \{j_i, j_i+1, \ldots, N-3\}$ and the smallest number $N_{i+1} \geq 3$ such that $\tilde{H}_{(i+1)}$ has one of the following forms:

$$\tilde{H}_{(i+1)} = \left[\begin{array}{c|cc} \tilde{H}_{(i+1)}^{(1)} & \star & \star \\ \hline 0 & \tilde{H}_{(i+1)}^{(2)} & 0 \\ 0 & \star & \tilde{H}_{(i+1)}^{(3)} \end{array}\right], \qquad \tilde{H}_{(i+1)} = \left[\begin{array}{c|cc} \tilde{H}_{(i+1)}^{(1)} & 0 & 0 \\ \hline \star & \tilde{H}_{(i+1)}^{(2)} & 0 \\ 0 & \star & \tilde{H}_{(i+1)}^{(3)} \end{array}\right]$$

$$\tilde{H}_{(i+1)} = \left[\begin{array}{cc|c} \tilde{H}_{(i+1)}^{(1)} & \star & \star \\ 0 & \tilde{H}_{(i+1)}^{(2)} & \star \\ \hline 0 & 0 & \tilde{H}_{(i+1)}^{(3)} \end{array}\right], \qquad \tilde{H}_{(i+1)} = \left[\begin{array}{cc|c} \tilde{H}_{(i+1)}^{(1)} & 0 & \star \\ \star & \tilde{H}_{(i+1)}^{(2)} & \star \\ \hline 0 & 0 & \tilde{H}_{(i+1)}^{(3)} \end{array}\right].$$

Here $\tilde{H}_{(i+1)}^{(2)}$ is an unreduced $N_{i+1} \times N_{i+1}$ upper Hessenberg matrix and $\tilde{H}_{(i+1)}^{(1)}$ is a $j_{i+1} \times j_{i+1}$ upper Hessenberg matrix.

(8) If there are no numbers j_{i+1} and N_{i+1} satisfying the relations (7) then goto (11). Compute the characteristic polynomial ψ_i of

$$[\tilde{H}_{(i+1)}^{(2)}(j,l)]_{N_{i+1}-1 \leq j,l \leq N_{i+1}},$$

the trailing 2×2 submatrix of $\tilde{H}_{(i+1)}^{(2)}$.
If $l_i = l_{i+1}$ then goto (9) else goto (10).

(9) Compute $v_1^{(i+1)} := \psi_i(\tilde{H}^{(2)}_{(i+1)})e_1$.
If $\mu_{l+1}(1) \leq N_{i+1}$ then compute $w_1^{(i+1)} := \psi_i((\tilde{H}^{(2)}_{(i+1)})^T)[I_{N_{i+1}}, 0]D^T_{(i)}e_1$
else compute $w_1^{(i+1)} := \psi_i(\tilde{H}^{(2)}_{(i+1)})e_1$.
$i := i + 1$.
Goto (3).

(10) Compute $v_1^{(i+1)} := \psi_i(\tilde{H}^{(2)}_{(i+1)})e_1$.
Compute $w_1^{(i+1)} := \psi_i((\tilde{H}^{(2)}_{(i+1)})^T)e_1$.
Set $i := i + 1$.
Goto (3).

(11) Set $i_0 := i$
end.

Remark 10.1 (Tridiagonal starting matrices) If the starting matrix A is tridiagonal we set $H_{(1)} := A$, $D_{(0)} := I$ and do not execute (1), (3) and (4) in step $i = 0$.

Remark 10.2 (Real matrices) If the matrix A has real entries and the starting vectors in (1) are chosen with real entries then the complete LGR algorithm does not require complex arithmetic.

For completeness we discuss in the following the properties of the $N \times N$ upper Hessenberg matrices $(\tilde{H}_{(i+1)})_{i \in X}$ $(X \subseteq \mathbb{N}_0)$ computed by the LGR algorithm.
Let $V_{(i)}$ be the transformation matrices generated by the procedure 'compute H and D' in the i-th step of the LGR algorithm. Define

$$\check{V}_i := \begin{bmatrix} I_{j_i} & 0 & 0 \\ 0 & V_{(i)} & 0 \\ 0 & & I_{N-N_i-j_i} \end{bmatrix} \in \mathrm{Mat}_N.$$

and

$$G_i := \check{V}_0 \check{V}_1 \cdots \check{V}_i.$$

Then we find, with the arguments used in the proof of (9.8), that there are matrices $R_i \in \mathrm{Mat}_N$ such that

$$A + R_i = G_i \tilde{H}_{(i+1)} G_i^{-1} \qquad \text{and} \qquad \|R_{(i)}\| \leq K_2 \cdot tol_3 \tag{10.23}$$

for all $i \in X$. Here

$$K_2 = N^2 \cdot \left(18 \left(\frac{1}{tol} + \frac{1}{2}\right)^2 + 2\right)^{i_1} \tag{10.24}$$

and i_1 is the minimal number such that $N_i = N_{i+1}$ for all $i \geq i_1$.
One easily verifies that the matrix $\tilde{H}_{(i_0+1)}$, generated by the LGR method, has the form (9.9).
Finally Theorem 10.3 below states estimates for the eigenvalue approximations.

Theorem 10.3 *Let $A \in \mathrm{Mat}_N$. Choose starting vectors $v_1^{(0)}, w_1^{(0)} \in \mathbb{C} \setminus \{0\}$. Choose $tol, tol_3 \in (0,1]$. Let $(\tilde{H}_{(i+1)})_{i \in X}$ be the sequence of upper Hessenberg matrices computed by the LGR algorithm. Let the number K_2 be defined as in* (10.24). *Then for all $i \in X$ the following statements hold.*

a) Let $Q^*AQ = diag\,(\lambda_1, \lambda_2, \ldots, \lambda_N) + L$ be a Schur decompostion of the matrix A. Let $\beta^{(i+1)}$ be an eigenvalue of $\tilde{H}_{(i+1)}$. Let p be the smallest positive integer such that $|L|^p = 0$. Then there is an eigenvalue λ_j of A such that

$$|\beta^{(i+1)} - \lambda_j| \leq \max\{\theta, \theta^{1/p}\}$$

where

$$\theta = K_2 \cdot tol_3 \cdot \sum_{k=0}^{p-1} ||L||^k.$$

b) Suppose there is a matrix $Y \in \mathrm{Mat}_N$ such that $Y^{-1}AY = diag\,(\lambda_1, \lambda_2, \ldots, \lambda_N)$. Let $\beta^{(i+1)}$ be an eigenvalue of $\tilde{H}_{(i+1)}$. Then there is an eigenvalue λ_j of A such that

$$|\beta^{(i+1)} - \lambda_j| \leq \kappa_2(Y) \cdot K_2 \cdot tol_3.$$

Theorem 10.3 follows immediately from (10.23) and the perturbation theorems of Golub and Van Loan [28] Theorem 7.2.2, Theorem 7.2.3.

Remark 10.3 For implementation details see Section 9, 'Implementation details for the RSL algorithm'.

11 Comparing the RSL, LGR and QR algorithms

11.1 Flop counts

First we consider the case that the matrix A is a full $N \times N$ matrix with real entries and the starting vectors $v_1^{(0)}$, $w_1^{(0)} \in \mathbb{R}^N \setminus \{0\}$ are chosen arbitrarily.

In step zero (i.e. $i = 0$) the matrix A is reduced in the RSL and the LGR algorithms to upper Hessenberg block tridiagonal form ($A \to A_1$) and the next starting vectors $v_1^{(1)}$ and $w_1^{(1)}$ are computed. Using estimate (8.119) we find that step zero requires less than

$$\begin{aligned} \frac{8}{3}N^3 &+ \left(8\left(15 - \frac{7}{l_{max}^{(1)}}\right)s_1 + 7\right)N^2 + \left(8s_1^2 - 9s_1 + \frac{55}{3}\right)N \\ &+ 40s_1^3 + 9s_1^2 + \left(\frac{2}{3}(l_{max}^{(1)})^3 + O((l_{max}^{(1)})^2)\right)s_1 \\ &+ 16(l_{max}^{(1)})^2 + l_{max}^{(1)} + 12 \end{aligned} \tag{11.1}$$

flops. Here $l_{max}^{(1)}$ is the size of the largest diagonal blocks $E_j^{(1)}$ of the computed matrix $H_{(1)}$ (see (8.120)). The number s_1 is the sum of the sizes of all diagonal blocks of $H_{(1)}$ which are not 1×1 matrices (see (8.121)). One easily verifies that the new starting vectors have the forms

$$v_1^{(1)} = \begin{bmatrix} v_{1,1}^{(1)} \\ 0 \end{bmatrix} \qquad \text{and} \qquad w_1^{(1)} = \begin{bmatrix} w_{1,1}^{(1)} \\ 0 \end{bmatrix} \tag{11.2}$$

where $v_{1,1}^{(1)} \in \mathbb{R}^3$ and $w_{1,1}^{(1)} \in \mathbb{R}^{\mu_1(3)}$.

Now we are going to derive an estimate for the amount of flops required in the i-th step ($i \geq 1$) of the RSL algorithm:
The i-th step of the RSL method starts with a block tridiagonal upper Hessenberg matrix A_k ($k \geq 1$) and vectors $v_1^{(i)}$, $w_1^{(i)}$ of the form

$$v_1^{(i)} = \begin{bmatrix} v_{1,1}^{(i)} \\ 0 \end{bmatrix} \qquad \text{and} \qquad w_1^{(i)} = \begin{bmatrix} w_{1,1}^{(i)} \\ 0 \end{bmatrix} \tag{11.3}$$

where $v_{1,1}^{(i)} \in \mathbb{R}^{1+2\varepsilon}$ and $w_{1,1}^{(1)} \in \mathbb{R}^{\mu_{n(k)}(1+2\varepsilon)}$.
Here and in the following ε denotes the number of steps carried out with the starting matrix A_k. In other words,

$$\varepsilon = i - n(k) + 1 \tag{11.4}$$

where k is chosen such that $n(k) \leq i < n(k+1)$.

Using the flop count (8.123) we find immediately that the i-th step of the RSL algorithm involves

$$\begin{aligned} N_i(8\tilde{s}_{i+1}^2 + 4\tilde{s}_{i+1} l_{max}^{(n(k))} + 7\tilde{s}_{i+1} + 10) &+ s_{i+1} 19\tilde{s}_{i+1}^2 \\ &+ s_{i+1}(8 l_{max}^{(n(k))} - 4) l_{max}^{(i+1)} - 8\, l_{max}^{(n(k))} + 18)\tilde{s}_{i+1} \\ &+ s_{i+1}\left(\frac{8}{3}(l_{max}^{(n(k))})^2 (l_{max}^{(i+1)})^2 + 3(l_{max}^{(n(k))})^2\right) \\ &+ s_{i+1}\left(\frac{2}{3}(l_{max}^{(i+1)})^3 + O(l_{max}^{(n(k))}(l_{max}^{(i+1)})^2)\right) \\ &+ f_{i+1} \end{aligned} \tag{11.5}$$

flops. Here $l_{max}^{(i+1)}$ is the size of the largest diagonal blocks $E_j^{(i+1)}$ of the computed matrix $H_{(i+1)}$ and $l_{max}^{(n(k))}$ is the size of the largest diagonal blocks of the matrix A_k. The number s_{i+1} is the sum of the sizes of all diagonal blocks of $H_{(i+1)}$ which are not 1×1 matrices and

$$\tilde{s}_{i+1} = \mu_{n(k)}(1 + 2\varepsilon) + \sum_{j=2(\varepsilon+1), l_j^{(n(k))} \neq 1}^{b_{n(k)}} l_j^{(n(k))}$$

where $l_j^{(n(k))}$ is the size of the j-th diagonal block of A_k and $b_{n(k)}$ is defined in (9.10). Further, if $i + 1 < n(k+1)$ then f_{i+1} is defined as

$$\begin{aligned} f_{i+1} &= 4(\mu_{n(k)}(1 + 2\varepsilon))^2 + \mu_{n(k)}(1 + 2\varepsilon)(12\, l_{max}^{(n(k))} + 8) \\ &+ 2(1 + 2\varepsilon)^2 + 13(1 + 2\varepsilon) + 8(l_{max}^{(n(k))})^2 - l_{max}^{(n(k))} + 1. \end{aligned} \tag{11.6}$$

In the other case $(i + 1 = n(k+1))$ the number f_{i+1} is defined by

$$f_{i+1} = 3\,\mu_{n(k)}(1 + 2\varepsilon) + 3\,(1 + 2\varepsilon) + 16(l_{max}^{(i+1)})^2 + l_{max}^{(i+1)} + 12. \tag{11.7}$$

To simplify the expression (11.5) we introduce the numbers $l_{max,i+1}$, $\delta^{(i+1)}$ and δ_{i+1} defined below.

$$l_{max,i+1} = \max\{l_{max}^{(n(k))}, l_{max}^{(i+1)}\}$$

$\delta^{(i+1)}$ denotes the number of diagonal blocks of the matrix $H_{(i+1)}$ which are not 1×1 matrices. Finally δ_{i+1} is defined by

$$\delta_{i+1} = \max\{\delta^{(n(k))}, \delta^{(i+1)}\}.$$

Using the inequalities

$$s_{i+1} \leq \delta_{i+1} l_{max,i+1} \qquad \text{and} \qquad \tilde{s}_{i+1} \leq (1 + 2\varepsilon + \delta_{i+1}) l_{max,i+1} \tag{11.8}$$

we obtain the estimate

$$\begin{aligned} &N_i((32\varepsilon^2 + 54\varepsilon + 32i\delta_{i+1} + 8\delta_{i+1}^2 + 27\delta_{i+1} + 19) l_{max,i+1}^2 + 10) \\ &+ \frac{8}{3}\delta_{i+1} l_{max,i+1}^5 + O(\varepsilon^2 \delta_{i+1}^2 l_{max,i+1}^4) + O(\varepsilon^2 l_{max,i+1}^2) \end{aligned} \tag{11.9}$$

for the number of flops required in the i-th step of the RSL algorithm. Notice that $\delta_{i+1} = 0$ and $l_{max,i+1} = 1$ if the matrices A_k and $H_{(i+1)}$ are tridiagonal.

Next we derive an estimate for the number of flops required in the i-th step ($i \geq 1$) of the LGR algorithm:
Considering (9) and (10) in the LGR method we recognize immediately that, in contrast to the RSL method (see (11.3), the starting vectors have for all $i \geq 1$ the forms

$$v_1^{(i)} = \begin{bmatrix} v_{1,1}^{(i)} \\ 0 \end{bmatrix} \qquad \text{and} \qquad w_1^{(i)} = \begin{bmatrix} w_{1,1}^{(i)} \\ 0 \end{bmatrix} \tag{11.10}$$

where $v_{1,1}^{(i)} \in \mathbb{R}^3$ and $w_{1,1}^{(i)} \in \mathbb{R}^{\mu_{i+1}(3)}$.
Using the flop count (8.123) and the partitions (11.10) we find that the i-th step (for $i \geq 1$) of the LGR algorithm involves

$$\begin{aligned} N_i(8\tilde{s}_{i+1}^2 + 4\tilde{s}_{i+1}l_{max}^{(i)} + 7\tilde{s}_{i+1} + 10) \quad &+ \quad s_{i+1}(19\tilde{s}_{i+1}^2 + ((8l_{max}^{(i)} - 4)l_{max}^{(i+1)} - 8\,l_{max}^{(i)} + 18)\tilde{s}_{i+1}) \\ &+ \quad s_{i+1}\left(\frac{8}{3}(l_{max}^{(i)})^2(l_{max}^{(i+1)})^2 + 3(l_{max}^{(i)})^2\right) \\ &+ \quad s_{i+1}\left(\frac{2}{3}(l_{max}^{(i+1)})^3 + O(l_{max}^{(i)}(l_{max}^{(i+1)})^2)\right) \\ &+ \quad 16(l_{max}^{(i+1)})^2 + l_{max}^{(i+1)} + 9l_{max}^{(i)} + 21 \end{aligned} \tag{11.11}$$

flops. Here $l_{max}^{(i+1)}$ is the size of the largest diagonal blocks $E_j^{(i+1)}$ of the computed matrix $H_{(i+1)}$. The number s_{i+1} is the sum of the sizes of all diagonal blocks of $H_{(i+1)}$ which are not 1×1 matrices and

$$\tilde{s}_{i+1} = \mu_i(3) + \sum_{j=4, l_j^{(i)} \neq 1}^{b_i} l_j^{(i)}$$

where $l_j^{(i)}$ is the size of the j-th diagonal block of $H_{(i)}$ and the number b_i is defined in (10.20).
To simplify the expression (11.11) we use, as in the case of the RSL method, the numbers $l_{max,i+1}$, $\delta^{(i+1)}$ and δ_{i+1} defined below:

$$l_{max,i+1} = \max\{l_{max}^{(i)}, l_{max}^{(i+1)}\}$$

$\delta^{(i+1)}$ denotes the number of diagonal blocks of the matrix $H_{(i+1)}$ which are not 1×1 matrices. δ_{i+1} is defined by

$$\delta_{i+1} = \max\{\delta^{(i)}, \delta^{(i+1)}\}.$$

Using the estimates

$$s_{i+1} \leq \delta_{i+1} l_{max,i+1} \qquad \text{and} \qquad \tilde{s}_{i+1} \leq (3 + \delta_{i+1}) l_{max,i+1} \tag{11.12}$$

the flop count (11.11) turns into

$$N_i(84\,l_{max,i+1}^2 + 21\,l_{max,i+1} + 8\,\delta_{i+1}^2 l_{max,i+1}^2 + 52\,\delta_{i+1} l_{max,i+1}^2 + 7\,\delta_{i+1} l_{max,i+1} + 10)$$
$$+\frac{8}{3}\delta_{i+1} l_{max,i+1}^5 + O(\delta_{i+1}^2 l_{max,i+1}^4) + O(l_{max,i+1}^2). \tag{11.13}$$

We recall that $\delta_{i+1} = 0$ and $l_{max,i+1} = 1$ if the generated matrices are tridiagonal.

In our numerical experiments (see Subsection 9.4) we have observed that all matrices generated by the RSL and LGR algorithms are typically nearly tridiagonal (i.e. $l_{max,i+1}$ and δ_{i+1} are small numbers) if the tolerance tol is chosen small enough (e.g. for $N \leq 100$ and $tol = 10^{-4}$ we have observed $l_{max,i+1} \leq 6$ and $\delta_{i+1} \leq 3$). Further we have observed that in both methods on the average 2.8 iterations are required to separate a 1×1 or a 2×2 block if we choose $tol_3 = 10^{-8}$ in the RSL method, $tol_3 = 10^{-16}$ in the LGR method and $tol = 10^{-4}$ in both methods (see Subsection 11.3). Combining these observations with the flop counts (11.9) and (11.13) we find that in both methods typically $O(N)$ flops are required for every step $i \geq 1$. Putting these results together we obtain the following:

- Let $A \in \mathrm{Mat}_N(\mathbb{R})$ be a nearly tridiagonal upper Hessenberg matrix with distinct eigenvalues. If the starting vectors $v_1,\ w_1 \in \mathbb{R}^N \setminus \{0\}$ have the form (8.122) then the RSL algorithm and the LGR algorithm require typically

  $$O(N^2) \tag{11.14}$$

  flops for the computation of all eigenvalues of the matrix A.

- For full matrices $A \in \mathrm{Mat}_N(\mathbb{R})$ with distinct eigenvalues and arbitrary starting vectors $v_1,\ w_1 \in \mathbb{R}^N \setminus \{0\}$ the RSL algorithm and the LGR algorithm typically require

  $$\frac{8}{3}N^3 + O(N^2) \tag{11.15}$$

  flops for the computation of all eigenvalues of the matrix A.

In the following we consider the case where $A \in \mathrm{Mat}_N(\mathbb{R})$ is a full matrix with distinct eigenvalues and the computed matrices $H_{(i+1)}$ are tridiagonal. In this important special case we can give more detailed estimates for the flops involved in the RSL and LGR methods. By (11.1) there are

$$\frac{8}{3}N^3 + 7N^2 + \frac{55}{3}N + 29 \tag{11.16}$$

flops involved in step $i = 0$, the reduction process of the matrix A to tridiagonal form. Using (11.5), (11.6), (11.7), (11.11) and the inequalities (11.8), (11.12) we obtain the following estimates for the i-th step ($i \geq 1$):

- If $i+1 < n(k)$ then the RSL algorithm requires

$$N_i\,(32\varepsilon^2 + 54\varepsilon + 29) + 24\varepsilon^2 + 90\varepsilon + 47 \tag{11.17}$$

flops in the i-th step.

- If $i+1 = n(k)$ then

$$N_i\,(32\varepsilon^2 + 54\varepsilon + 29) + 12\varepsilon + 35 \tag{11.18}$$

flops are necessary in the i-th step of the RSL algorithm.

- In the i-th step of the LGR method

$$115\,N_i + 47 \tag{11.19}$$

flops are involved.

In the following we assume that there are $\varepsilon_{max}+1$ steps necessary before the first 1×1 or 2×2 block decouples. Further we suppose that the decoupling of every following 1×1 or 2×2 block requires not more than ε_{max} steps. Under these assumptions we obtain the results summarized below:

- If the starting matrix $A \in \mathrm{Mat}_N(\mathbb{R})$ and all generated matrices $H_{(i+1)}$ are tridiagonal then the RSL algorithm requires not more than

$$\begin{aligned} & N^2\left(\frac{16}{3}\,\varepsilon_{max}^3 + \frac{43}{2}\,\varepsilon_{max}^2 + \frac{92}{3}\,\varepsilon_{max}\right) \\ + \;& N\left(\frac{40}{3}\,\varepsilon_{max}^3 + \frac{173}{2}\,\varepsilon_{max}^2 + \frac{500}{3}\,\varepsilon_{max} + 103\right) \\ - \;& 48\,\varepsilon_{max}^3 - 171\,\varepsilon_{max}^2 - 130\,\varepsilon_{max} + 112 \end{aligned} \tag{11.20}$$

flops for the computation of all eigenvalues of the matrix A.
The LGR algorithm needs not more than

$$\frac{115}{2}\,\varepsilon_{max}N^2 + \left(\frac{209}{2}\,\varepsilon_{max} + 115\right)N - 439\,\varepsilon_{max} + 76 \tag{11.21}$$

flops for the computation of all eigenvalues of the matrix A.

- For full matrices $A \in \mathrm{Mat}_N(\mathbb{R})$ and arbitrary starting vectors $v_1,\ w_1 \in \mathbb{R}^N \setminus \{0\}$ the RSL algorithm requires not more than

$$\begin{aligned} \frac{8}{3}N^3 \;+\; & N^2\left(\frac{16}{3}\,\varepsilon_{max}^3 + \frac{43}{2}\,\varepsilon_{max}^2 + \frac{92}{3}\,\varepsilon_{max} + 7\right) \\ + \;& N\left(\frac{40}{3}\,\varepsilon_{max}^3 + \frac{173}{2}\,\varepsilon_{max}^2 + \frac{500}{3}\,\varepsilon_{max} + \frac{364}{3}\right) \\ - \;& 48\,\varepsilon_{max}^3 - 171\,\varepsilon_{max}^2 - 130\,\varepsilon_{max} + 112 \end{aligned} \tag{11.22}$$

flops for the computation of all eigenvalues of the matrix A if all generated matrices $H_{(i+1)}$ are tridiagonal.
In this situation the LGR algorithm needs not more than

$$\frac{8}{3}N^3 + (\frac{115}{2}\,\varepsilon_{max} + 7)N^2 + \left(\frac{209}{2}\,\varepsilon_{max} + \frac{400}{3}\right) N - 439\,\varepsilon_{max} + 76 \tag{11.23}$$

flops for the computation of all eigenvalues of the matrix A.

Comparison with the QR algorithm

In the zero step of the QR method the matrix $A \in \mathrm{Mat}_N(\mathbb{R})$ is typically reduced to upper Hessenberg form. This is done to save work in the subsequent iterative phase of the QR method (see Golub and Van Loan [28], EISPACK [69]). The reduction of a full matrix to upper Hessenberg form requires

$$\frac{10}{3}N^3 - \frac{11}{2}N^2 + \frac{1}{6}N - 5$$

flops if Housholder transforms are used.
If Gauss transforms are used

$$\frac{4}{3}N^3 - \frac{5}{2}N^2 - \frac{11}{6}N + 3$$

flops are necessary (EISPACK routine 'elmhes'). Every step in the iterative phase of the QR method (EISPACK routine 'hqr') requires

$$10N_i^2 + 32N_i - 31 \tag{11.24}$$

flops where N_i is the current size of the iterated matrix.
In the following we assume that the QR method requires $\tilde{\varepsilon}_{max} + 1$ steps for decoupling the first 1×1 or 2×2 block. Further we suppose that $\tilde{\varepsilon}_{max}$ steps are necessary before the next 1×1 or 2×2 block decouples. Under these assumptions we obtain the following results for the complete QR algorithm:

- The QR algorithm does not preserve the structure of non-symmetric tridiagonal starting matrices $A \in \mathrm{Mat}_N(\mathbb{R})$. For non-symmetric tridiagonal matrices and upper Hessenberg matrices $A \in \mathrm{Mat}_N(\mathbb{R})$ the QR algorithm requires

$$\frac{10}{3}\,\tilde{\varepsilon}_{max}\,N^3 + (21\,\tilde{\varepsilon}_{max} + 10)N^2 + (32 - \frac{40}{3}\tilde{\varepsilon}_{max})N - 84\,\tilde{\varepsilon}_{max} - 31 \tag{11.25}$$

flops for the computation of all eigenvalues of the matrix A.

- For full matrices $A \in \mathrm{Mat}_N(\mathbb{R})$ the QR algorithm needs

$$\frac{10}{3}(\tilde{\varepsilon}_{max}+1)\,N^3+(21\,\tilde{\varepsilon}_{max}+\frac{9}{2})N^2+(\frac{193}{6}-\frac{40}{3}\tilde{\varepsilon}_{max})N-84\,\tilde{\varepsilon}_{max}-36 \quad (11.26)$$

  flops for the computation of all eigenvalues of the matrix A if Housholder transforms are used for the reduction of A to upper Hessenberg form.

- For full matrices $A \in \mathrm{Mat}_N(\mathbb{R})$ the QR algorithm requires

$$(\frac{10}{3}\,\tilde{\varepsilon}_{max}+\frac{4}{3})N^3+(21\,\tilde{\varepsilon}_{max}+\frac{15}{2})N^2+(\frac{181}{6}-\frac{40}{3}\tilde{\varepsilon}_{max})N-84\,\tilde{\varepsilon}_{max}-28 \quad (11.27)$$

  flops for the computation of all eigenvalues of the matrix A if Gauss transforms are used for the reduction of A to upper Hessenberg form.

In our numerical experiments (see Subsection 11.2) the RSL algorithm and the LGR algorithm require on the average 2.8 steps and the QR method 3.9 steps to separate a 1×1 or a 2×2 block. To compare the flop counts we suppose that all matrices $H_{(i+1)}$, generated by the RSL and the LGR algorithms, are tridiagonal and that $\varepsilon_{max} = 2.8$ and $\tilde{\varepsilon}_{max} = 3.9$. Considering the expressions (11.20), (11.21), (11.22), (11.23), (11.25), (11.26) and (11.27) we find the following estimates:

Type of the starting matrix $A \in \mathrm{Mat}_N$	The RSL algorithm requires less flops than the Gauss/QR algorithm for all	The RSL algorithm requires less flops than the Housholder/QR algorithm for all
Tridiagonal matrix	$N \geq 26$	$N \geq 26$
Hessenberg matrix	$N \geq 33$	$N \geq 33$
Full matrix	$N \geq 30$	$N \geq 26$

Type of the starting matrix $A \in \mathrm{Mat}_N$	The LGR algorithm requires less flops than the Gauss/QR algorithm for all	The LGR algorithm requires less flops than the Housholder/QR algorithm for all
Tridiagonal matrix	$N \geq 9$	$N \geq 9$
Hessenberg matrix	$N \geq 11$	$N \geq 11$
Full matrix	$N \geq 10$	$N \geq 9$

11.2 Numerical examples

In this subsection we report the output of the RSL, LGR and the QR algorithms applied to a wide range of test matrices. Throughout this subsection we use the EISPACK [69] implementation of the Gauss/QR method (routines hqr and elmhes). If one of these eigenvalue algorithms requires more than 50 steps to separate one block (i.e. $\varepsilon > 50$) then the algorithm stops and reports failure. We symbolize this case in the following tables by the sign '$***$'.
All calculations are carried out in the IEEE arithmetic double.

Example 11.1 (Random matrix (Continuation of Example 7.2)) In Example 7.2 we have discussed that the identically restarted look-ahead Lanczos method applied to the random matrix A and the random starting vectors $w_1^{(0)}$ and $v_1^{(0)}$

$$A = \begin{bmatrix} 0.00 & 0.06 & -0.28 & 0.41 & 0.55 & -0.66 & 0.64 & 0.32 \\ 0.16 & 0.95 & 0.14 & -0.32 & 0.12 & 0.59 & -0.17 & 0.74 \\ -0.60 & -0.31 & -0.56 & 0.61 & 0.33 & 0.66 & 0.93 & 0.49 \\ 0.96 & 0.30 & -0.25 & 0.57 & -0.45 & -0.05 & 0.99 & -0.22 \\ 0.28 & 0.29 & -1.00 & 0.04 & -0.81 & 0.02 & -0.45 & -0.70 \\ -0.88 & 0.41 & -0.64 & -0.81 & -0.09 & -0.71 & 0.00 & 0.49 \\ 0.17 & -0.46 & 0.99 & -0.24 & -0.98 & -0.85 & -0.09 & -0.63 \\ -0.59 & -0.02 & -0.45 & -0.50 & 0.40 & 0.29 & -0.17 & -0.43 \end{bmatrix}, \quad w_1^{(0)} = \begin{bmatrix} 0.85 \\ 0.66 \\ 0.56 \\ 0.04 \\ -0.58 \\ 0.59 \\ 0.02 \\ -0.27 \end{bmatrix}$$

$$v_1^{(0)} = \begin{bmatrix} 0.74 & -0.45 & -0.35 & -0.35 & -0.46 & -0.65 & 0.68 & -0.82 \end{bmatrix}^T$$

does not generate satisfactory approximations to the eigenvalues of A although all eigenvalues of A are not ill-conditoned ($cond(\lambda_j) \leq 2.5$ for all j).

Eigenvalues of A (rounded)		
Eigenvalue	Real part	Imaginary part
λ_1	$1.94768032815462 \cdot 10^{\ 00}$	0
$\lambda_{2/3}$	$7.22771408213559 \cdot 10^{-01}$	$\pm 3.86823730013324 \cdot 10^{-01}$
$\lambda_{4/5}$	$-4.63268021120600 \cdot 10^{-01}$	$\pm 3.06680358938131 \cdot 10^{-01}$
λ_6	$-1.07573663811272 \cdot 10^{\ 00}$	0
$\lambda_{7/8}$	$-1.23547523211391 \cdot 10^{\ 00}$	$\pm 1.23396246460755 \cdot 10^{\ 00}$

The QR algorithm applied to A terminates after 19 steps and yields a sequence ($i = 1, 2, \ldots, 20$) of upper Hessenberg matrices

$$A_i = \begin{bmatrix} H_1^{(i)} & & & \star \\ G_1^{(i)} & H_2^{(i)} & & \\ & \ddots & \ddots & \\ 0 & & G_4^{(i)} & H_5^{(i)} \end{bmatrix} = \left[\begin{array}{c|ccccccc} A_{1,1}^{(i)} & A_{1,2}^{(i)} & A_{1,3}^{(i)} & A_{1,4}^{(i)} & A_{1,5}^{(i)} & A_{1,6}^{(i)} & A_{1,7}^{(i)} & A_{1,8}^{(i)} \\ \hline A_{2,1}^{(i)} & A_{2,2}^{(i)} & A_{2,3}^{(i)} & A_{2,4}^{(i)} & A_{2,5}^{(i)} & A_{2,6}^{(i)} & A_{2,7}^{(i)} & A_{2,8}^{(i)} \\ 0 & A_{3,2}^{(i)} & A_{3,3}^{(i)} & A_{3,4}^{(i)} & A_{3,5}^{(i)} & A_{3,6}^{(i)} & A_{3,7}^{(i)} & A_{3,8}^{(i)} \\ \hline 0 & 0 & A_{4,3}^{(i)} & A_{4,4}^{(i)} & A_{4,5}^{(i)} & A_{4,6}^{(i)} & A_{4,7}^{(i)} & A_{4,8}^{(i)} \\ 0 & 0 & 0 & A_{5,4}^{(i)} & A_{5,5}^{(i)} & A_{5,6}^{(i)} & A_{5,7}^{(i)} & A_{5,8}^{(i)} \\ \hline 0 & 0 & 0 & 0 & A_{6,5}^{(i)} & A_{6,6}^{(i)} & A_{6,7}^{(i)} & A_{6,8}^{(i)} \\ 0 & 0 & 0 & 0 & 0 & A_{7,6}^{(i)} & A_{7,7}^{(i)} & A_{7,8}^{(i)} \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & A_{8,7}^{(i)} & A_{8,8}^{(i)} \end{array}\right].$$

After 8 steps of the QR algorithm, the diagonal element $H_5^{(i)} = A_{8,8}^{(i)}$ decouples. After two further steps the diagonal block $H_4^{(i)}$ is separated. The diagonal block $H_3^{(i)}$ decouples after 7 further steps. Finally, the diagonal blocks $H_1^{(i)}$ and $H_2^{(i)}$ are separated after 2 final steps.

Eigenvalue of $H_5^{(i+1)}$ (QR algorithm)			
i	ε	$\lambda_6^{(i+1)}$	$A_{8,7}^{(i+1)}$
0	0	$-7.82414021449142 \cdot 10^{-01}$	$-6 \cdot 10^{-01}$
1	1	$-8.85907577038759 \cdot 10^{-01}$	$4 \cdot 10^{-01}$
2	2	$-7.92135792046678 \cdot 10^{-01}$	$-1 \cdot 10^{-01}$
3	3	$-1.17006050745228 \cdot 10^{00}$	$-2 \cdot 10^{-01}$
4	4	$-1.07713461805488 \cdot 10^{00}$	$2 \cdot 10^{-02}$
5	5	$-1.07551001687239 \cdot 10^{00}$	$-3 \cdot 10^{-04}$
6	6	$-1.07573657360134 \cdot 10^{00}$	$-1 \cdot 10^{-07}$
7	7	$-1.07573663811272 \cdot 10^{00}$	$-1 \cdot 10^{-14}$
8	8	$-1.07573663811272 \cdot 10^{00}$	$2 \cdot 10^{-28}$

Eigenvalue of $H_4^{(i+1)}$ (QR algorithm)				
i	ε	$\Re\lambda_{4/5}^{(i+1)}$	$\Im\lambda_{4/5}^{(i+1)}$	$A_{6,5}^{(i+1)}$
9	1	$-4.63268021120600 \cdot 10^{-01}$	$\pm 3.06680358938131 \cdot 10^{-01}$	$1 \cdot 10^{-16}$
10	2	$-4.63268021120600 \cdot 10^{-01}$	$\pm 3.06680358938131 \cdot 10^{-01}$	$3 \cdot 10^{-32}$

Eigenvalues of $H_3^{(i+1)}$ (QR algorithm)				
i	ε	$\Re\lambda_{7/8}^{(i+1)}$	$\Im\lambda_{7/8}^{(i+1)}$	$A_{4,3}^{(i+1)}$
11	1	$-7.33385958807190 \cdot 10^{-01}$	$\pm 1.50892979473283 \cdot 10^{00}$	$-1 \cdot 10^{00}$
12	2	$-1.04514365968782 \cdot 10^{00}$	$\pm 1.29141778406387 \cdot 10^{00}$	$-4 \cdot 10^{-01}$
13	3	$-1.22625165471259 \cdot 10^{00}$	$\pm 1.22432557804061 \cdot 10^{00}$	$-4 \cdot 10^{-02}$
14	4	$-1.23550411247673 \cdot 10^{00}$	$\pm 1.23398607940271 \cdot 10^{00}$	$2 \cdot 10^{-04}$
15	5	$-1.23547523240123 \cdot 10^{00}$	$\pm 1.23396246537612 \cdot 10^{00}$	$3 \cdot 10^{-09}$
16	6	$-1.23547523211391 \cdot 10^{00}$	$\pm 1.23396246460755 \cdot 10^{00}$	$1 \cdot 10^{-18}$
17	7	$-1.23547523211391 \cdot 10^{00}$	$\pm 1.23396246460755 \cdot 10^{00}$	$4 \cdot 10^{-34}$

Eigenvalues of $H_2^{(i+1)}$ and $H_1^{(i+1)}$ (QR algorithm)					
i	ε		Real part	Imaginary part	$A_{2,1}^{(i+1)}$
18	1	$\lambda_{2/3}^{(i+1)}$	$7.22771408213560 \cdot 10^{-01}$	$\pm 3.86823730013325 \cdot 10^{-01}$	$1 \cdot 10^{-15}$
		$\lambda_1^{(i+1)}$	$1.94768032815462 \cdot 10^{00}$	$0.00000000000000 \cdot 10^{00}$	
19	2	$\lambda_{2/3}^{(i+1)}$	$7.22771408213560 \cdot 10^{-01}$	$\pm 3.86823730013323 \cdot 10^{-01}$	$4 \cdot 10^{-34}$
		$\lambda_1^{(i+1)}$	$1.94768032815462 \cdot 10^{00}$	$0.00000000000000 \cdot 10^{00}$	

We choose $tol_3 = 10^{-12}$ and $tol \leq 10^{-3}$. Then the RSL and the LGR algorithms, applied to the matrix A and the starting vectors $v_1^{(0)}$ and $w_1^{(0)}$, generate two sequences of

tridiagonal matrices

$$\tilde{H}_{(i)} = \begin{bmatrix} H_1^{(i)} & F_1^{(i)} & & 0 \\ G_1^{(i)} & H_2^{(i)} & \ddots & \\ & \ddots & \ddots & F_4^{(i)} \\ 0 & & G_4^{(i)} & H_5^{(i)} \end{bmatrix} = \left[\begin{array}{c|cc|cc|cc|c} E_1^{(i)} & C_1^{(i)} & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline B_1^{(i)} & E_2^{(i)} & C_2^{(i)} & 0 & 0 & 0 & 0 & 0 \\ 0 & B_2^{(i)} & E_3^{(i)} & C_3^{(i)} & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & B_3^{(i)} & E_4^{(i)} & C_4^{(i)} & 0 & 0 & 0 \\ 0 & 0 & 0 & B_4^{(i)} & E_5^{(i)} & C_5^{(i)} & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & B_5^{(i)} & E_6^{(i)} & C_6^{(i)} & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & B_6^{(i)} & E_7^{(i)} & C_7^{(i)} \\ 0 & 0 & 0 & 0 & 0 & 0 & B_7^{(i)} & E_8^{(i)} \end{array}\right].$$

The *RSL* method terminates after 10 steps. The *LGR* method terminates after 11 steps. The *LGR* method requires 4 steps to separate the diagonal block $H_5^{(i)}$. The *RSL* method requires 3 steps to separate the diagonal block $H_5^{(i)}$. For decoupling the remaining diagonal blocks $H_4^{(i)}$, $H_3^{(i)}$, $H_2^{(i)}$ and $H_1^{(i)}$ both methods require the same number of steps (see below).

Eigenvalues of $H_5^{(i+1)}$ (*LGR* algorithm)					
i	ε	$\Re\lambda_{4/5}^{(i+1)}$	$\Im\lambda_{4/5}^{(i+1)}$	$B_6^{(i+1)}$	$C_6^{(i+1)}$
0	0	$-5.59229549963591 \cdot 10^{-01}$	$0.00000000000000 \cdot 10^{00}$	$-4 \cdot 10^{-02}$	$3 \cdot 10^{00}$
		$3.64371917679969 \cdot 10^{-01}$	$0.00000000000000 \cdot 10^{00}$		
1	1	$-4.61432540407505 \cdot 10^{-01}$	$\pm 3.06571398605870 \cdot 10^{-01}$	$7 \cdot 10^{-03}$	$-5 \cdot 10^{-01}$
2	2	$-4.63268035075507 \cdot 10^{-01}$	$\pm 3.06680350809450 \cdot 10^{-01}$	$-4 \cdot 10^{-05}$	$-3 \cdot 10^{-04}$
3	3	$-4.63268021120601 \cdot 10^{-01}$	$\pm 3.06680358938134 \cdot 10^{-01}$	$2 \cdot 10^{-12}$	$3 \cdot 10^{-12}$
4	4	$-4.63268021120601 \cdot 10^{-01}$	$\pm 3.06680358938141 \cdot 10^{-01}$	$-1 \cdot 10^{-27}$	$2 \cdot 10^{-26}$

Eigenvalue of $H_4^{(i+1)}$ (*LGR* algorithm)				
i	ε	$\lambda_6^{(i+1)}$	$B_5^{(i+1)}$	$C_5^{(i+1)}$
5	1	$-1.07573663808503 \cdot 10^{00}$	$-1 \cdot 10^{-06}$	$-4 \cdot 10^{-05}$
6	2	$-1.07573663811272 \cdot 10^{00}$	$-9 \cdot 10^{-19}$	$3 \cdot 10^{-16}$

Eigenvalues of $H_3^{(i+1)}$ (LGR algorithm)					
i	ε	$\Re\lambda_{2/3}^{(i+1)}$	$\Im\lambda_{2/3}^{(i+1)}$	$B_3^{(i+1)}$	$C_3^{(i+1)}$
7	1	$7.22771408213563 \cdot 10^{-01}$	$\pm 3.86823730013325 \cdot 10^{-01}$	$2 \cdot 10^{-08}$	$-1 \cdot 10^{-06}$
8	2	$7.22771408213559 \cdot 10^{-01}$	$\pm 3.86823730013328 \cdot 10^{-01}$	$3 \cdot 10^{-24}$	$-3 \cdot 10^{-22}$

Eigenvalues of $H_2^{(i+1)}$ and $H_1^{(i+1)}$ (LGR algorithm)						
i	ε		Real part	Imaginary part	$B_1^{(i+1)}$	$C_1^{(i+1)}$
9	1	$\lambda_{7/8}^{(i+1)}$	$-1.23546988830601 \cdot 10^{00}$	$\pm 1.23398847819365 \cdot 10^{00}$	$-5 \cdot 10^{-02}$	$-2 \cdot 10^{-03}$
		$\lambda_1^{(i+1)}$	$1.94766964053881 \cdot 10^{00}$	$0.00000000000000 \cdot 10^{00}$		
10	2	$\lambda_{7/8}^{(i+1)}$	$-1.23547523211391 \cdot 10^{00}$	$\pm 1.23396246460756 \cdot 10^{00}$	$-4 \cdot 10^{-07}$	$-1 \cdot 10^{-08}$
		$\lambda_1^{(i+1)}$	$1.94768032815461 \cdot 10^{00}$	$0.00000000000000 \cdot 10^{00}$		
11	3	$\lambda_{7/8}^{(i+1)}$	$-1.23547523211391 \cdot 10^{00}$	$\pm 1.23396246460755 \cdot 10^{00}$	$-1 \cdot 10^{-22}$	$-2 \cdot 10^{-24}$
		$\lambda_1^{(i+1)}$	$1.94768032815461 \cdot 10^{00}$	$0.00000000000000 \cdot 10^{00}$		

Eigenvalues of $H_5^{(i+1)}$ (RSL algorithm)					
i	ε	$\Re\lambda_{4/5}^{(i+1)}$	$\Im\lambda_{4/5}^{(i+1)}$	$B_6^{(i+1)}$	$C_6^{(i+1)}$
0	0	$-5.59229549963591 \cdot 10^{-01}$	$0.00000000000000 \cdot 10^{00}$	$-4 \cdot 10^{-02}$	$3 \cdot 10^{00}$
		$3.64371917679969 \cdot 10^{-01}$	$0.00000000000000 \cdot 10^{00}$		
1	1	$-4.61432540407505 \cdot 10^{-01}$	$\pm 3.06571398605870 \cdot 10^{-01}$	$7 \cdot 10^{-03}$	$-5 \cdot 10^{-01}$
2	2	$-4.63268035075508 \cdot 10^{-01}$	$\pm 3.06680350809446 \cdot 10^{-01}$	$1 \cdot 10^{-04}$	$-1 \cdot 10^{-04}$
3	3	$-4.63268021120601 \cdot 10^{-01}$	$\pm 3.06680358938131 \cdot 10^{-01}$	$1 \cdot 10^{-12}$	$5 \cdot 10^{-12}$

Eigenvalue of $H_4^{(i+1)}$ (RSL algorithm)				
i	ε	$\lambda_6^{(i+1)}$	$B_5^{(i+1)}$	$C_5^{(i+1)}$
4	1	$-1.07573669391980 \cdot 10^{00}$	$-2 \cdot 10^{-05}$	$4 \cdot 10^{-03}$
5	2	$-1.07573663811271 \cdot 10^{00}$	$8 \cdot 10^{-14}$	$6 \cdot 10^{-11}$

Eigenvalues of $H_3^{(i+1)}$ (RSL algorithm)					
i	ε	$\Re\lambda_{2/3}^{(i+1)}$	$\Im\lambda_{2/3}^{(i+1)}$	$B_3^{(i+1)}$	$C_3^{(i+1)}$
6	1	$7.22771408103133 \cdot 10^{-01}$	$\pm 3.86823729973599 \cdot 10^{-01}$	$-1 \cdot 10^{-05}$	$-5 \cdot 10^{-05}$
7	2	$7.22771408213558 \cdot 10^{-01}$	$\pm 3.86823730013334 \cdot 10^{-01}$	$-2 \cdot 10^{-16}$	$-3 \cdot 10^{-16}$

Eigenvalues of $H_2^{(i+1)}$ and $H_1^{(i+1)}$ (RSL algorithm)						
i	ε		Real part	Imaginary part	$B_1^{(i+1)}$	$C_1^{(i+1)}$
8	1	$\lambda_{7/8}^{(i+1)}$	$-1.23547810529154 \cdot 10^{00}$	$\pm 1.23395917404653 \cdot 10^{00}$	$-3 \cdot 10^{-02}$	$9 \cdot 10^{-04}$
		$\lambda_1^{(i+1)}$	$1.94768607450986 \cdot 10^{00}$	$0.00000000000000 \cdot 10^{00}$		
9	2	$\lambda_{7/8}^{(i+1)}$	$-1.23547523211391 \cdot 10^{00}$	$\pm 1.23396246460754 \cdot 10^{00}$	$4 \cdot 10^{-08}$	$-4 \cdot 10^{-10}$
		$\lambda_1^{(i+1)}$	$1.94768032815461 \cdot 10^{00}$	$0.00000000000000 \cdot 10^{00}$		
10	3	$\lambda_{7/8}^{(i+1)}$	$-1.23547523211391 \cdot 10^{00}$	$\pm 1.23396246460720 \cdot 10^{00}$	$4 \cdot 10^{-16}$	$-7 \cdot 10^{-19}$
		$\lambda_1^{(i+1)}$	$1.94768032815461 \cdot 10^{00}$	$0.00000000000000 \cdot 10^{00}$		

We observe that the eigenvalues, generated by the QR algorithm, differ from the exact eigenvalues of the matrix A by not more than 10^{-15}. Further the LGR method yields eigenvalues approximations with a maximal error of 10^{-14}. The RSL method computes 6 eigenvalues approximations with errors less than 10^{-14} and 2 eigenvalue approximations with an error of $4 \cdot 10^{-13}$.

Counting the flops we find that the QR algorithm requires $\approx$ 12000 flops, the LGR algorithm uses $\approx$ 7600 flops and the RSL method requires $\approx$ 8400 flops.

Next we answer the question if it is possible to improve the accuracy of the eigenvalue approximations by choosing the tolerance tol_3 small enough.
We have experimented with tolerances tol_3 downto 10^{-16} ($\approx$ machine constant) and have observed the following:

- **LGR method:** For every choice of the number $tol_3 \in [10^{-12}, 10^{-16}]$ the LGR algorithm terminates after 11 steps and yields the same output as for the choice $tol_3 = 10^{-12}$, discussed above.

- **RSL method:** Choosing $tol_3 = 10^{-13}$ the RSL method terminates after 41 steps ($\approx$ 69000 flops) and generates 6 eigenvalues approximations with errors less than

$2 \cdot 10^{-14}$ and 2 eigenvalues approximations with an error of $4 \cdot 10^{-12}$.
If we choose $tol_3 = 10^{-14}$ then the RSL method terminates after 43 steps (≈ 73000 flops) and computes eigenvalues which differ from the exact eigenvalues of the matrix A by not more than $2 \cdot 10^{-14}$.
Choosing $tol_3 \in [10^{-16}, 10^{-15}]$ the RSL method stops with a failure report because it requires more than 50 steps to separate one block (i.e. $\varepsilon > 50$).

In all following examples we choose in the LGR process the tolerance $tol_3 = 10^{-16}$. Further we choose the tolerance $tol = 10^{-4}$ in the RSL and in the LGR process and use random starting vectors $v_1^{(0)}, w_1^{(0)}$ (generated by the random number generator of Turbo Pascal [6]).

Throughout the rest of the section λ_i denotes an exact eigenvalue of the matrix A whereas β_i is an approximation of λ_i, generated by the LGR, QR and RSL algorithm, respectively.

Example 11.2 (Cyclic matrices)

$$A = \begin{bmatrix} 0 & \dots & 0 & 1 \\ 1 & \ddots & & 0 \\ & \ddots & \ddots & \vdots \\ 0 & & 1 & 0 \end{bmatrix} \in \mathrm{Mat}_N.$$

The eigenvalues of the cyclic matrix A are the N-th complex roots of unity. All these eigenvalues are well-conditioned ($cond(\lambda_i) = 1$ for all i).
Applying the LGR, the QR and the RSL algorithm to A we obtain the results:

$N \times N$ Cyclic matrices, $tol = 10^{-4}$						
N	Algorithm	tol_3 (chosen)	Maximal rel. error $\max_i \frac{\lvert\lambda_i - \beta_i\rvert}{\lvert\lambda_i\rvert}$	Iteration steps required	Flops required	Steps to separate one block (mean value)
8	LGR	10^{-16}	$9 \cdot 10^{-13}$	17	$1.1 \cdot 10^4$	4.3
	QR		$5 \cdot 10^{-16}$	26	$1.8 \cdot 10^4$	6.5
	RSL	10^{-12}	$4 \cdot 10^{-14}$	17	$1.8 \cdot 10^4$	4.3

$N \times N$ Cyclic matrices, $tol = 10^{-4}$						
N	Algorithm	tol_3 (chosen)	Maximal rel. error $\max_i \frac{\lvert\lambda_i - \beta_i\rvert}{\lvert\lambda_i\rvert}$	Iteration steps required	Flops required	Steps to separate one block (mean value)
20	LGR	10^{-16}	$1 \cdot 10^{-11}$	29	$6.1 \cdot 10^4$	2.9
	QR		$8 \cdot 10^{-16}$	43	$1.3 \cdot 10^5$	4.3
	RSL	10^{-11}	$1 \cdot 10^{-12}$	25	$8.6 \cdot 10^4$	2.5
50	LGR	10^{-16}	$1 \cdot 10^{-09}$	69	$5.8 \cdot 10^5$	2.8
	QR		$2 \cdot 10^{-15}$	83	$1.1 \cdot 10^6$	3.3
	RSL	$\leq 10^{-08}$	$***$	$***$	$***$	$***$
100	LGR	10^{-16}	$1 \cdot 10^{-09}$	117	$3.5 \cdot 10^6$	2.3
	QR		$2 \cdot 10^{-15}$	147	$6.6 \cdot 10^6$	2.9
	RSL	$\leq 10^{-08}$	$***$	$***$	$***$	$***$

For $N = 8$, $N = 20$ and $N = 50$ all matrices generated by the LGR and the RSL method are tridiagonal. For $N = 100$ the LGR method computes in 113 iteration steps tridiagonal matrices and in 4 iteration steps block tridiagonal matrices with exactly one diagonal block of order 2 and $N - 2$ diagonal blocks of order 1.

Example 11.3 (Clement matrices) The Clement matrices

$$A = \begin{bmatrix} 0 & N-1 & 0 & & 0 \\ 1 & 0 & N-2 & & \\ 0 & 2 & \ddots & & \\ & & \ddots & \ddots & 1 \\ 0 & & \ddots & N-1 & 0 \end{bmatrix} \in \mathrm{Mat}_N$$

have the eigenvalues $\pm 1, \pm 3, \cdots, \pm(N-1)$ for N even and $\pm 0, \pm 2, \cdots, \pm(N-1)$ for odd N (see e.g. [38], [23], [74]). These eigenvalues are ill-conditioned. It is well known that the QR method is not able to compute accurate approximations of the inner eigenvalues of the Clement matrices for $N \geq 100$ (see e.g. [23], [74], [7]). In contrast to that the

LGR method generates even for $N = 250$ accurate approximations to all eigenvalues of the Clement matrix (see below).

$N \times N$ Clement matrices, $tol = 10^{-4}$						
N	Algorithm	tol_3 (chosen)	Maximal rel. error $\max_i \frac{\lvert\lambda_i-\beta_i\rvert}{\lvert\lambda_i\rvert}$	Iteration steps required	Flops required	Steps to separate one block (mean value)
8	*LGR*	10^{-16}	$4 \cdot 10^{-16}$	10	$6.1 \cdot 10^{3}$	3.3
	QR		$1 \cdot 10^{-15}$	14	$8.9 \cdot 10^{3}$	4.7
	RSL	10^{-13}	$7 \cdot 10^{-16}$	9	$9.5 \cdot 10^{3}$	3.0
20	*LGR*	10^{-16}	$3 \cdot 10^{-15}$	24	$3.5 \cdot 10^{4}$	2.7
	QR		$1 \cdot 10^{-14}$	32	$7.8 \cdot 10^{4}$	3.6
	RSL	10^{-10}	$2 \cdot 10^{-14}$	22	$8.3 \cdot 10^{4}$	2.4
50	*LGR*	10^{-16}	$1 \cdot 10^{-14}$	53	$1.7 \cdot 10^{5}$	2.2
	QR		$7 \cdot 10^{-11}$	73	$8.2 \cdot 10^{5}$	3.0
	RSL	10^{-08}	$7 \cdot 10^{-10}$	50	$2.9 \cdot 10^{5}$	2.1
100	*LGR*	10^{-16}	$5 \cdot 10^{-14}$	107	$6.7 \cdot 10^{5}$	2.2
	QR		$2 \cdot 10^{-03}$	142	$6.7 \cdot 10^{6}$	2.9
	RSL	$\leq 10^{-08}$	* * *	* * *	* * *	* * *
150	*LGR*	10^{-16}	$6 \cdot 10^{-14}$	158	$1.5 \cdot 10^{6}$	2.1
	QR		$1 \cdot 10^{+01}$	255	$2.0 \cdot 10^{7}$	2.7
	RSL	$\leq 10^{-08}$	* * *	* * *	* * *	* * *
200	*LGR*	10^{-16}	$1 \cdot 10^{-13}$	208	$2.5 \cdot 10^{6}$	2.1
	QR		$2 \cdot 10^{+01}$	391	$5.0 \cdot 10^{7}$	2.8
	RSL	$\leq 10^{-08}$	* * *	* * *	* * *	* * *
250	*LGR*	10^{-16}	$3 \cdot 10^{-14}$	259	$3.9 \cdot 10^{6}$	2.1
	QR		$2 \cdot 10^{+01}$	471	$8.6 \cdot 10^{7}$	3.0
	RSL	$\leq 10^{-08}$	* * *	* * *	* * *	* * *

Here all matrices generated by the LGR and the RSL algorithm are tridiagonal.
The amazing result that the LGR method computes accurate eigenvalue approximations and the QR algorithm fails has the following reason: The LGR method preserves in every step the tridiagonal structure of the starting matrices. The QR algorithm destroys the tridiagonal structure and creates in the first step a full upper Hessenberg matrix A_2. The small rounding errors which occur in the computation of the off-tridiagonal elements cause in double precision arithmetic the huge eigenvalue error observed above.

Example 11.4 (Hyperbolic matrices) Discretizing the model parabolic equation

$$u_t = u_{xx} \tag{11.28}$$

in an interval $[0, l]$ with zero Dirichlet boundary conditions by an explicit difference scheme with constant step sizes h and k we obtain the well known difference equation

$$u^{(i+1)} = \left(I + \frac{k}{h^2} A \right) u^{(i)} \tag{11.29}$$

where $u = (u_1, u_2, \ldots, u_N)^T$ and $A = tridiag(1, -2, 1)$ is the 'parabolic matrix'. The characteristic polynomials p_N of the parabolic matrices $A \in \mathrm{Mat}_N$ satisfy a three term recursion. Furthermore, the Tchebysheff polynomials U_N of the second kind and the polynomials p_N satisfy the equation $p_N(2x-2) = U_N(x)$ for all $x \in \mathbb{C}$ and all $N \in \mathbb{N}$.

Hadeler discussed in [36] a hyperbolic analogue of (11.28), (11.29) and introduced and studied in this context the 'hyperbolic matrices'

$$A = \begin{bmatrix} B & D \\ D & B^T \end{bmatrix} \in \mathrm{Mat}_{2n} \tag{11.30}$$

where

$$B = \begin{bmatrix} \alpha & & & 0 \\ \gamma & \ddots & & \\ & \ddots & \ddots & \\ 0 & & \gamma & \alpha \end{bmatrix} \in \mathrm{Mat}_n \qquad \text{and} \qquad D = \tilde{\mu} I \in \mathrm{Mat}_n.$$

Indeed, consider the model hyperbolic system (Goldstein-Kac system [44], [78]) for a correlated random walk on the real line

$$\begin{aligned} u_t^+ + \gamma u_x^+ &= \mu(u^- - u^+) \\ u_t^- - \gamma u_x^- &= \mu(u^+ - u^-) \end{aligned} \tag{11.31}$$

($\mu > 0$ and $\gamma > 0$) on an interval $[0, l]$ with zero Dirichlet boundary conditions. Discretizing the hyperbolic system (11.31) with an explicit difference scheme with step sizes h and k we obtain the difference equation

$$u^{(i+1)} = \left(I + \frac{k}{h} A \right) u^{(i)} \tag{11.32}$$

where $u = (u_1^+, u_2^+, \ldots, u_n^+, u_1^-, u_2^-, \ldots, u_n^-)^T$ and A is the hyperbolic matrix (11.30) with $\tilde{\mu} = h \cdot \mu$ and $\alpha = -\gamma - \tilde{\mu}$. Hadeler [36] derived the three term vector recursions
$Q_0(\lambda) = 0$, $P_0(\lambda) = 1$

$$\begin{aligned} Q_n(\lambda) &= \tilde{\mu} P_{n-1}(\lambda) - \gamma^2 Q_{n-1}(\lambda) \\ P_n(\lambda) &= \tilde{\mu} Q_n(\lambda) - (\lambda + \gamma + \tilde{\mu})^2 P_{n-1}(\lambda) \end{aligned} \tag{11.33}$$

for the characteristic polynomials P_n of the hyperbolic matrices A. Notice that P_n has degree $2n$ and Q_n has degree $2n - 2$.
Define polynomials $S_n(z)$, $R_n(z)$ of degree n by $R_0(z) = 0$, $S_0(z) = 1$,

$$\begin{aligned} R_n(\lambda) &= \tilde{\mu} S_{n-1}(z) - \gamma^2 R_{n-1}(z) \\ S_n(z) &= \tilde{\mu} R_n(z) - z S_{n-1}(z). \end{aligned} \tag{11.34}$$

Then

$$P_n(\lambda) = S_n((\lambda + \gamma + \tilde{\mu})^2). \tag{11.35}$$

Equation (11.35) implies that the spectrum of the hyperbolic matrix A is symmetric with respect to the line $\Re\lambda = -\tilde{\mu} - \gamma$. This situation is similar to the parabolic case. It is well known that the spectrum of the parabolic matrix $A = tridiag(1, -2, 1)$ is symmetric with respect to the line $\Re\lambda = -2$. We remark that the hyperbolic matrices have, in contrast to parabolic matrices, in general complex eigenvalues.

In the following we discuss for which choice of time step sizes k the difference scheme (11.32) with parameters $\tilde{\mu} = 1$ and $\gamma = 2$ is stable, i.e. the eigenvalues $\hat{\lambda}_i$ of the iteration matrix

$$I + \frac{k}{h} A \tag{11.36}$$

satisfy the inequality

$$|\hat{\lambda}_i| < 1 \tag{11.37}$$

for all i.
One easily verifies that the inequality (11.37) is equivalent to

$$\frac{k}{h} < -2 \frac{\Re\lambda_i}{|\lambda_i|^2} \tag{11.38}$$

where λ_i are the eigenvalues of the hyperbolic matrix A. Computing the eigenvalues λ_i and using (11.38) we find that the difference scheme (11.32) with parameters $\tilde{\mu} = 1$ and $\gamma = 2$ is stable if the time step size k is chosen as described below:

N	8	20	50	100
$\frac{k}{h} \leq$	0.372	0.343	0.335	0.333

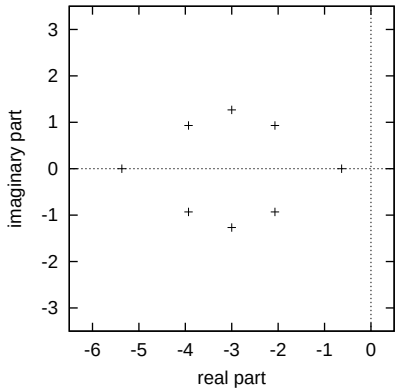


Fig.1: Eigenvalues of the 8×8 hyperbolic matrix

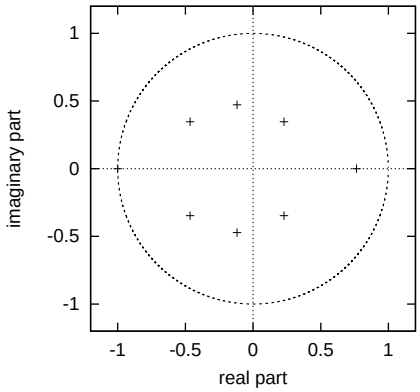


Fig.2: Eigenvalues of the 8×8 iteration matrix (11.36) with $k/h = 0.372$

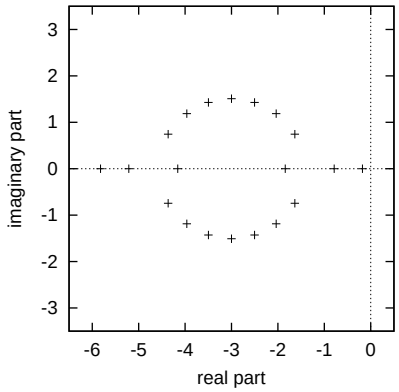


Fig.3: Eigenvalues of the 20×20 hyperbolic matrix

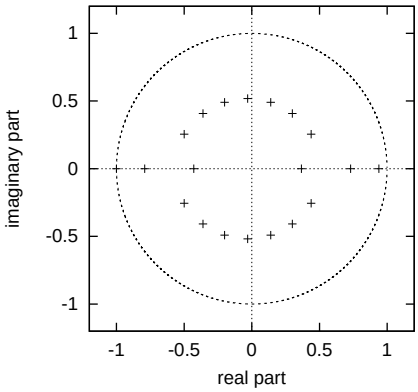


Fig.4: Eigenvalues of the 20×20 iteration matrix (11.36) with $k/h = 0.343$

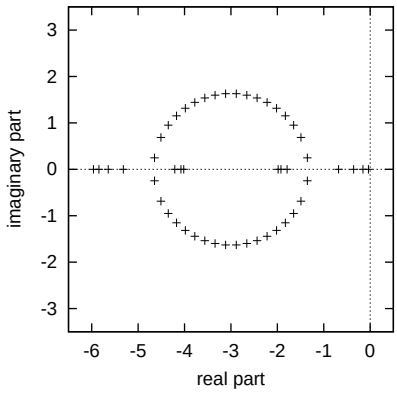


Fig.5: Eigenvalues of the 50×50 hyperbolic matrix

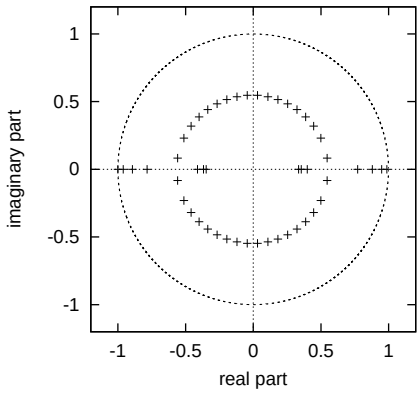


Fig.6: Eigenvalues of the 50×50 iteration matrix (11.36) with $k/h = 0.335$

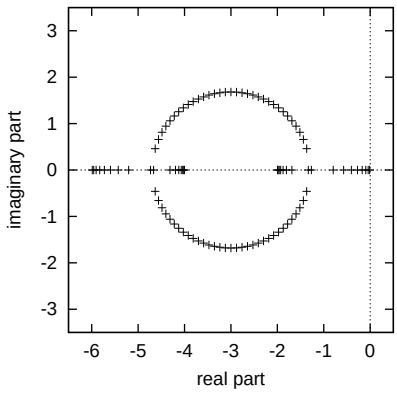


Fig.7: Eigenvalues of the 100×100 hyperbolic matrix

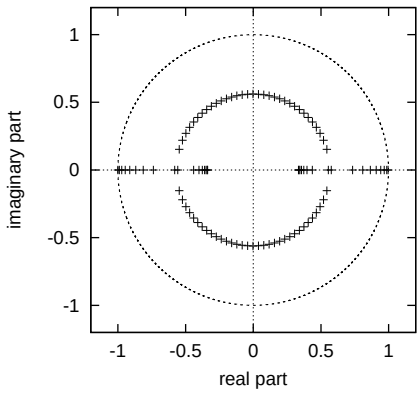


Fig.8: Eigenvalues of the 100×100 iteration matrix (11.36) with $k/h = 0.333$

$N \times N$ Hyperbolic matrices ($\gamma = 2$, $\tilde{\mu} = 1$), $tol = 10^{-4}$							
N	$\max_i$ $cond(\lambda_i)$	Algorithm	tol_3 (chosen)	Maximal rel. error $\max_i \frac{\lvert\lambda_i - \beta_i\rvert}{\lvert\lambda_i\rvert}$	Iteration steps required	Flops required	Steps to separate one block (mean value)
8	1.2	LGR	10^{-16}	$3 \cdot 10^{-10}$	12	$9.6 \cdot 10^3$	4.0
		QR		$2 \cdot 10^{-15}$	25	$1.5 \cdot 10^4$	8.3
		RSL	10^{-09}	$3 \cdot 10^{-10}$	11	$2.3 \cdot 10^4$	3.7
20	1.7	LGR	10^{-16}	$3 \cdot 10^{-11}$	27	$6.2 \cdot 10^4$	3.0
		QR		$3 \cdot 10^{-15}$	45	$1.3 \cdot 10^5$	5.0
		RSL	10^{-09}	$2 \cdot 10^{-10}$	23	$9.0 \cdot 10^4$	2.6
50	3.9	LGR	10^{-16}	$9 \cdot 10^{-08}$	66	$6.0 \cdot 10^5$	2.6
		QR		$2 \cdot 10^{-13}$	107	$1.5 \cdot 10^6$	4.5
		RSL	10^{-08}	$4 \cdot 10^{-10}$	52	$7.1 \cdot 10^5$	2.0
100	28.9	LGR	10^{-16}	$2 \cdot 10^{-04}$	123	$3.6 \cdot 10^6$	2.4
		QR		$2 \cdot 10^{-13}$	174	$8.5 \cdot 10^6$	3.4
		RSL	$\leq 10^{-08}$	$* * *$	$* * *$	$* * *$	$* * *$

$N \times N$ Hyperbolic matrices ($\gamma = 2$, $\tilde{\mu} = 1$)								
Block tridiagonal structure of the generated matrices								
N	Algorithm	Number of iteration steps	Number of diagonal blocks of order					
			1	2	3	4	5	6
8	*LGR*	12	8	0	0	0	0	0
	RSL	6	8	0	0	0	0	0
		5	6	1	0	0	0	0
20	*LGR*	26	20	0	0	0	0	0
		1	18	1	0	0	0	0
	RSL	23	20	0	0	0	0	0
50	*LGR*	57	50	0	0	0	0	0
		3	48	1	0	0	0	0
		3	46	2	0	0	0	0
		2	47	0	1	0	0	0
		1	45	1	1	0	0	0
	RSL	48	50	0	0	0	0	0
		4	48	1	0	0	0	0
100	*LGR*	104	100	0	0	0	0	0
		13	98	1	0	0	0	0
		4	96	2	0	0	0	0
		1	95	0	0	0	1	0
		1	92	1	0	0	0	1

The following test matrices are from the collection of non-hermitian eigenvalue problems of Bai et al. [2]

Example 11.5 (Brusselator wave model in chemical reaction)

The system

$$\begin{aligned} \frac{\partial x}{\partial t} &= \frac{\delta_1}{L^2}\frac{\partial^2 x}{\partial^2 z} + \alpha - (\beta+1)x + x^2 y, \\ \frac{\partial y}{\partial t} &= \frac{\delta_2}{L^2}\frac{\partial^2 y}{\partial^2 z} + \beta x - x^2 y \end{aligned} \tag{11.39}$$

with the initial conditions $x(0,z) = x_0(z), \quad y(0,z) = y_0(z)$ and the Dirichlet boundary conditions $x(t,0) = x(t,1) = x^*, \quad y(t,0) = y(t,1) = y^*$ is the so-called Brusselator wave model (see [60],[67]). This system models the concentration waves for reaction and transport interaction of chemical solutions in a tubular reactor. The Brusselator wave model has the stationary solution $x^* = \alpha$, $y^* = \beta/\alpha$. One is primarily interested in the existence of periodic solutions This can be studied by examining the spectrum of the Jacobian of the right-hand-side of the Brusselator system (11.39) at the stationary solution $x^* = \alpha$, $y^* = \beta/\alpha$. If the rightmost eigenvalues of the Jacobian are purely imaginary then there is a periodic solution of the Brusselator wave model. To discuss this problem numerically, the Jacobian is discretized in the interval $[0,1]$ with the mesh size $h = 1/(m+1)$ and the eigenvalues of the discretized Jacobian

$$A = \begin{bmatrix} \tau_1 T + (\beta-1)I & \alpha^2 I \\ -\beta I & \tau_2 T - \alpha^2 I \end{bmatrix} \in \mathrm{Mat}_{2m}$$

are computed. Here $T = \mathrm{tridiag}(1,-2,1) \in \mathrm{Mat}_m$, $\tau_1 = \frac{1}{h^2}\frac{\delta_1}{L^2}$ and $\tau_2 = \frac{1}{h^2}\frac{\delta_2}{L^2}$.

The exact eigenvalues of the Brusselator matrix A are known since there is a quadratic relation between the eigenvalues of the matrix A and those of the parabolic matrix $T = tridiag(1,-2,1)$ (see [2]).

In the following we consider the Brusselator system (11.39) with the parameters

$$\delta_1 = 0.008, \quad \delta_2 = 0.004, \quad \alpha = 2, \quad \beta = 5.45. \tag{11.40}$$

For small L the Jacobian has only eigenvalues with negative real parts. Following Saad [67] the rightmost pair of eigenvalues are purely imaginary at $L \approx 0.51302$. To verify this we compute the eigenvalues of the Brusselator matrices A with the parameters (11.40) and $L = 0.51302$.

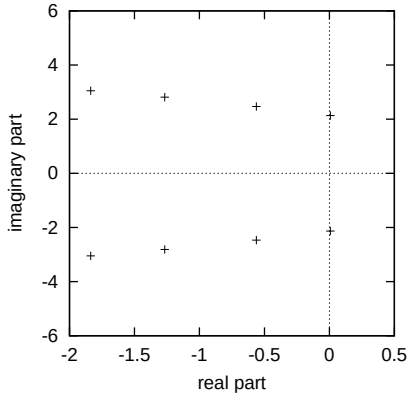


Fig.1: Eigenvalues of the 8×8 Brusselator matrix

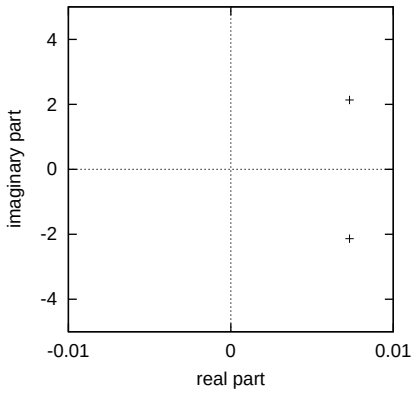


Fig.2: Rightmost eigenvalues of the 8×8 Brusselator matrix

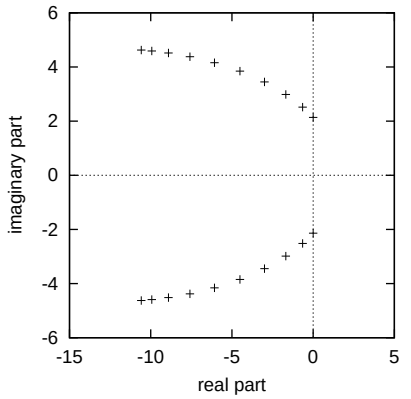


Fig.3: Eigenvalues of the 20×20 Brusselator matrix

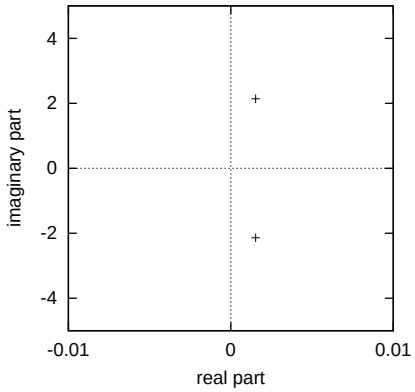


Fig.4: Rightmost eigenvalues of the 20×20 Brusselator matrix

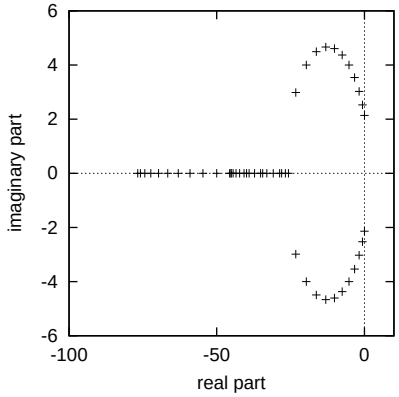


Fig.5: Eigenvalues of the 50×50 Brusselator matrix

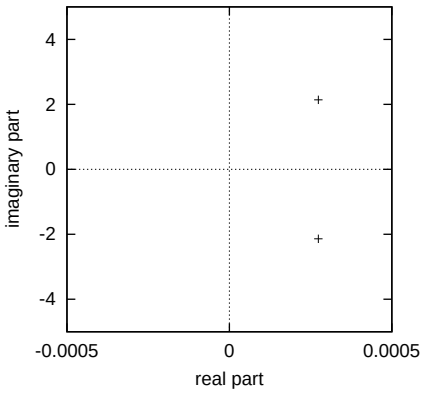


Fig.6: Rightmost eigenvalues of the 50×50 Brusselator matrix

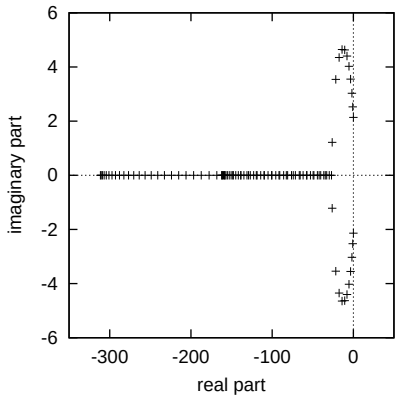


Fig.7: Eigenvalues of the 100×100 Brusselator matrix

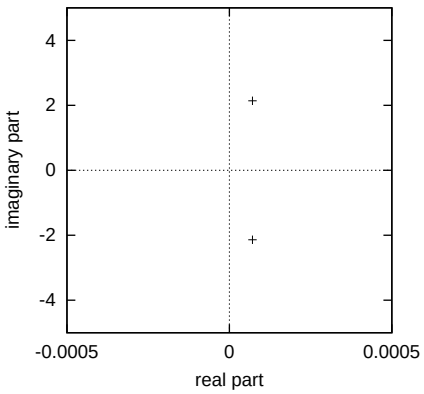


Fig.8: Rightmost eigenvalues of the 100×100 Brusselator matrix

$N \times N$ Brusselator matrices, $tol = 10^{-4}$						
N	Algorithm	tol_3 (chosen)	Maximal rel. error $\max_i \frac{\lvert\lambda_i - \beta_i\rvert}{\lvert\lambda_i\rvert}$	Iteration steps required	Flops required	Steps to separate one block (mean value)
8	LGR	10^{-16}	$1 \cdot 10^{-13}$	14	$9.1 \cdot 10^3$	4.7
	QR		$1 \cdot 10^{-15}$	16	$1.1 \cdot 10^4$	5.3
	RSL	10^{-12}	$9 \cdot 10^{-14}$	13	$1.2 \cdot 10^4$	4.3
20	LGR	10^{-16}	$5 \cdot 10^{-10}$	28	$6.5 \cdot 10^4$	3.1
	QR		$4 \cdot 10^{-15}$	40	$1.2 \cdot 10^5$	4.4
	RSL	10^{-10}	$1 \cdot 10^{-11}$	26	$1.6 \cdot 10^5$	2.8
50	LGR	10^{-16}	$3 \cdot 10^{-10}$	65	$5.6 \cdot 10^5$	2.5
	QR		$7 \cdot 10^{-15}$	79	$1.1 \cdot 10^6$	2.8
	RSL	$\leq 10^{-08}$	$***$	$***$	$***$	$***$
100	LGR	10^{-16}	$3 \cdot 10^{-09}$	133	$3.6 \cdot 10^6$	2.5
	QR		$4 \cdot 10^{-14}$	144	$7.6 \cdot 10^6$	2.6
	RSL	$\leq 10^{-08}$	$***$	$***$	$***$	$***$

$N \times N$ Brusselator matrices				
Block tridiagonal structure of the generated matrices				
N	Algorithm	Number of iteration steps	Number of diagonal blocks of order 1	Number of diagonal blocks of order 2
8	LGR	14	8	0
	RSL	13	8	0
20	LGR	27	20	0
		1	18	1
	RSL	26	20	0
50	LGR	65	50	0
100	LGR	131	100	0
		2	96	2

Example 11.6 (Ising model for ferromagnetic materials) The Ising matrices

$$A = KL \in \mathrm{Mat}_{2m}$$

where

$$K = \begin{bmatrix} E & & & \\ & E & & \\ & & \ddots & \\ & & & E \\ & & & & E \end{bmatrix}, \qquad L = \begin{bmatrix} \cos\beta & & & & -\sin\beta \\ & F & & & \\ & & \ddots & & \\ & & & F & \\ \sin\beta & & & & \cos\beta \end{bmatrix} \in \mathrm{Mat}_{2m}$$

$$E = \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix}, \qquad F = \begin{bmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{bmatrix}$$

arise in the analysis of the Ising model for ferromagnetic materials (see e.g. [46], [22], [15]).

It is well known that the eigenvalues of A are the $2m$ numbers that are obtained by computing the eigenvalues of the 2×2 matrices

$$B_k = \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} \cos\beta & -\theta^k \sin\beta \\ \theta^{m-k}\sin\beta & \cos\beta \end{bmatrix}$$

for $k = 1, 2, \ldots, m$, where $\theta = e^{i2\pi/m}$. One easily verifies that the matrices B_k and B_{m-k} have identical eigenvalues. Thus the Ising matrix $A \in \mathrm{Mat}_{2m}$ has maximal $m+1$ distinct eigenvalues.

In the following we compute the eigenvalues of the Ising matrices with $\alpha = \pi/4$ and $\beta = \pi/4$.

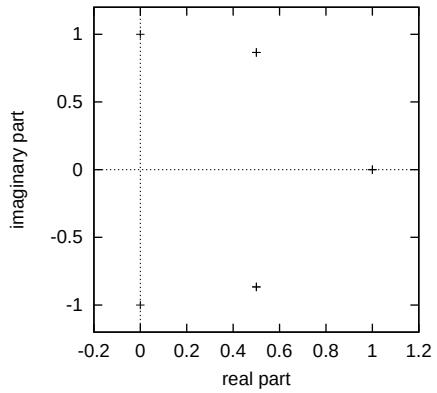


Fig.1: Eigenvalues of the 8×8 Ising matrix

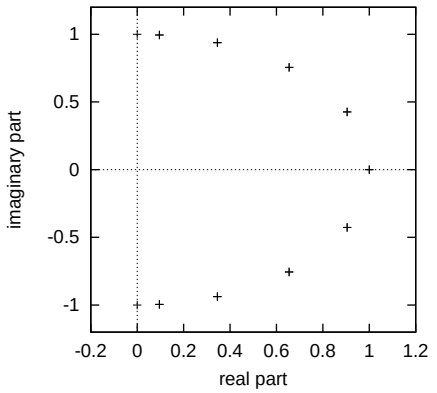


Fig.2: Eigenvalues of the 20×20 Ising matrix

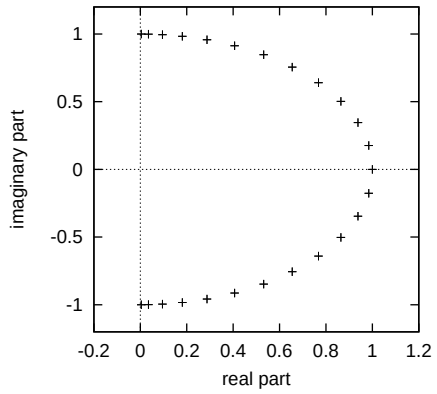


Fig.3: Eigenvalues of the 50×50 Ising matrix

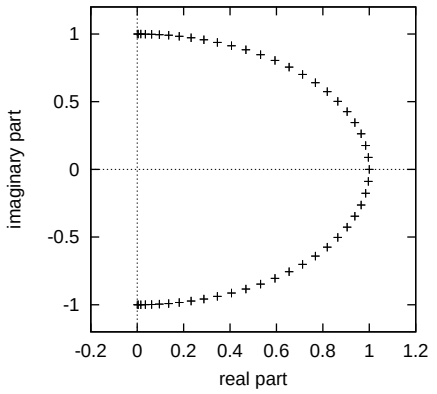


Fig.4: Eigenvalues of the 100×100 Ising matrix

$N \times N$ Ising matrices ($\alpha = \pi/4$, $\beta = \pi/4$), $tol = 10^{-4}$						
N	Algorithm	tol_3 (chosen)	Maximal rel. error $\max_i \frac{\lvert\lambda_i - \beta_i\rvert}{\lvert\lambda_i\rvert}$	Iteration steps required	Flops required	Steps to separate one block (mean value)
8	LGR	10^{-16}	$2 \cdot 10^{-11}$	8	$6.6 \cdot 10^3$	2.0
	QR		$5 \cdot 10^{-16}$	21	$4.2 \cdot 10^3$	7.0
	RSL	10^{-12}	$6 \cdot 10^{-14}$	12	$5.5 \cdot 10^3$	4.0
20	LGR	10^{-16}	$2 \cdot 10^{-12}$	17	$5.2 \cdot 10^4$	3.0
	QR		$2 \cdot 10^{-15}$	24	$8.3 \cdot 10^4$	2.7
	RSL	10^{-12}	$3 \cdot 10^{-12}$	25	$5.4 \cdot 10^4$	2.8
50	LGR	10^{-16}	$1 \cdot 10^{-08}$	48	$5.3 \cdot 10^5$	2.4
	QR		$3 \cdot 10^{-15}$	58	$9.2 \cdot 10^5$	2.4
	RSL	10^{-09}	$2 \cdot 10^{-11}$	55	$5.4 \cdot 10^5$	2.3
100	LGR	10^{-16}	$2 \cdot 10^{-07}$	105	$3.6 \cdot 10^6$	2.4
	QR		$7 \cdot 10^{-15}$	85	$6.2 \cdot 10^6$	1.7
	RSL	10^{-08}	$7 \cdot 10^{-09}$	116	$3.8 \cdot 10^6$	2.4

$N \times N$ Ising matrices ($\alpha = \pi/4$, $\beta = \pi/4$)								
Block tridiagonal structure of the generated matrices								
N	Algorithm	Number of iteration steps	Number of diagonal blocks of order					
			1	2	3	4	5	6
8	*LGR*	8	8	0	0	0	0	0
	RSL	12	8	0	0	0	0	0
20	*LGR*	16	20	0	0	0	0	0
		1	18	1	0	0	0	0
	RSL	25	20	0	0	0	0	0
50	*LGR*	31	50	0	0	0	0	0
		16	48	1	0	0	0	0
		1	45	1	1	0	0	0
	RSL	55	50	0	0	0	0	0
100	*LGR*	91	100	0	0	0	0	0
		10	98	1	0	0	0	0
		1	96	2	0	0	0	0
		1	94	3	0	0	0	0
		1	95	1	1	0	0	0
		1	92	1	0	0	0	1
	RSL	113	100	0	0	0	0	0
		2	98	1	0	0	0	0
		1	96	2	0	0	0	0

Example 11.7 (Model 2-D convection diffusion operator) Discretizing the convection diffusion Operator

$$Lu = -\Delta u + 2p_1 u_x + 2p_2 u_y - p_3 u \tag{11.41}$$

on the domain $\Omega = [0,1]^2$ with $u = 0$ on $\partial\Omega$ by a finite difference scheme with a 5-point stencil on a uniform $m \times m$ grid we obtain the matrix

$$A = \begin{bmatrix} T & (\beta-1)I & & & \\ -(\beta+1)I & T & (\beta-1)I & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & (\beta-1)I \\ & & & -(\beta+1)I & T \end{bmatrix} \in \mathrm{Mat}_{m^2}$$

with

$$T = \begin{bmatrix} 4-\sigma & \gamma-1 & & & \\ -\gamma-1 & 4-\sigma & \gamma-1 & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \gamma-1 \\ & & & -\gamma-1 & 4-\sigma \end{bmatrix} \in \mathrm{Mat}_m,$$

where $\beta = p_1 h, \gamma = p_2 h, \sigma = p_3 h^2$ and $h = 1/(m+1)$.
The exact eigenvalues of the convection diffusion matrix A are

$$\lambda_{k,l} = 4 - \sigma + 2\sqrt{1-\beta^2}\cos(kh\pi) + 2\sqrt{1-\gamma^2}\cos(lh\pi)$$

for $k, l = 1, 2, \ldots, m$ (see e.g. [2]).
The maximal condition number of the eigenvalues of the convection diffusion matrix A is

$$\max_{k,l} cond(\lambda_{k,l}) = \left(\frac{(1+\beta)(1+\gamma)}{(1-\beta)(1-\gamma)}\right)^{\frac{m-1}{2}}.$$

In the following we compute the eigenvalues of the convection diffusion matrix A with parameters $p_1 = 1$, $p_2 = 2$ and $p_3 = 30$. These parameters are also used in Bai et al. [2]. One easily verifies that for this choice of parameters the spectrum of the convection diffusion matrix A is real.

$N \times N$ Convection diffusion matrices ($p_1 = 1$, $p_2 = 2$, $p_3 = 30$), $tol = 10^{-4}$							
N	$\max_{k,l}$ $cond(\lambda_{k,l})$	Algorithm	tol_3 (chosen)	Maximal rel. error $\max_{k,l} \frac{\lvert\lambda_{k,l}-\beta_{k,l}\rvert}{\lvert\lambda_{k,l}\rvert}$	Iteration steps required	Flops required	Steps to separate one block (mean value)
9	5.0	LGR	10^{-16}	$4 \cdot 10^{-11}$	10	$9.3 \cdot 10^3$	2.5
		QR		$3 \cdot 10^{-15}$	16	$1.3 \cdot 10^4$	4.0
		RSL	10^{-10}	$5 \cdot 10^{-13}$	8	$9.6 \cdot 10^3$	2.0
25	7.8	LGR	10^{-16}	$2 \cdot 10^{-11}$	33	$9.9 \cdot 10^4$	2.2
		QR		$4 \cdot 10^{-14}$	36	$1.4 \cdot 10^5$	3.0
		RSL	$\leq 10^{-08}$	$***$	$***$	$***$	$***$
49	9.8	LGR	10^{-16}	$2 \cdot 10^{-09}$	61	$5.2 \cdot 10^5$	2.3
		QR		$1 \cdot 10^{-13}$	74	$9.6 \cdot 10^5$	3.0
		RSL	$\leq 10^{-08}$	$***$	$***$	$***$	$***$
100	11.9	LGR	10^{-16}	$2 \cdot 10^{-09}$	136	$3.8 \cdot 10^6$	2.3
		QR		$3 \cdot 10^{-14}$	149	$7.3 \cdot 10^6$	3.0
		RSL	$\leq 10^{-08}$	$***$	$***$	$***$	$***$

$N \times N$ Convection diffusion matrices					
Block tridiagonal structure of the generated matrices					
N	Algorithm	Number of iteration steps	Number of diagonal blocks of order		
			1	2	3
9	LGR	10	9	0	0
	RSL	8	9	0	0
25	LGR	33	25	0	0
49	LGR	61	49	0	0
100	LGR	132	100	0	0
		3	98	1	0
		1	97	0	1

11.3 Concluding remarks

In all our numerical experiments we have observed cubic convergence of the LGR algorithm (for the choice $tol = 10^{-4}$ and $tol_3 = 10^{-16}$) and quadratic convergence of the QR algorithm. Further we have observed that the RSL algorithm (with the choice $tol = 10^{-4}$ and $tol_3 \leq 10^{-8}$)

- does not converge in 12 out of 28 examples
- converges cubically in 12 out of 13 numerical examples for $N \leq 25$
- converges cubically in 3 out of 6 numerical examples for $N \in \{49, 50\}$
- converges cubically in 1 out of 9 numerical examples for $N \in \{100, 150, 200, 250\}$.

Steps:
The LGR method requires, on the average, 2.8 steps to separate a block (for $N \leq 100$).
The QR method requires, on the average, 3.9 steps to separate a block (for $N \leq 100$).
If the RSL method converges it needs about 2.8 steps to separate a block.

Flops:
The LGR algorithm requires, on the average, for all types of matrices (full, upper Hessenberg, tridiagonal) less flops than the QR algorithm.
For full and upper Hessenberg matrices of size $10 \leq N \leq 100$ the LGR algorithm needs about half the number of flops as compared to the QR method.
For tridiagonal matrices (and nearly tridiagonal matrices) we expect that the LGR method requires about
$1/2$ the number of flops of the QR method for $N = 20$,
$1/5$ the number of flops of the QR method for $N = 50$,
$1/10$ the number of flops of the QR method for $N = 100$,
$1/20$ the number of flops of the QR method for $N = 200$
(see Example 11.3).

Eigenvalue Error:
If the eigenvalues of a matrix $A \in \mathrm{Mat}_N$ are not ill-conditioned then the QR algorithm yields (in our examples) for $N \leq 100$ eigenvalue approximations with an average maximal relative error of $3 \cdot 10^{-14}$ (for details see the table below). If there are ill-conditioned eigenvalues then the QR method can fail (see Example 11.3).

The LGR algorithm generates (in our examples) for $N \leq 100$ eigenvalue approximations with an average maximal relative error of $2 \cdot 10^{-8}$ (for details see the table below). In 1 out of 28 numerical examples (see Example 11.4, $N = 100$) the LGR method yields the maximal relative eigenvalue error of $2 \cdot 10^{-04}$.

If the RSL method converges then it yields (in our examples) an average maximal relative eigenvalue error of $2 \cdot 10^{-9}$.

Mean values						
N	Algorithm	Number of steps to separate one block	Full matrix A $\frac{\text{flops required}}{\text{flops of } QR}$	Hessenberg matrix A $\frac{\text{flops required}}{\text{flops of } QR}$	Tridiagonal matrix A $\frac{\text{flops required}}{\text{flops of } QR}$	$\max_i \frac{\lvert\lambda_i - \beta_i\rvert}{\lvert\lambda_i\rvert}$
	LGR	3.4	0.76	0.72	0.69	$6 \cdot 10^{-11}$
8	QR	5.8	1	1	1	$1 \cdot 10^{-15}$
(9)	RSL	3.4	1.06	1.00	1.07	$5 \cdot 10^{-11}$
	LGR	2.8	0.59	0.47	0.45	$9 \cdot 10^{-11}$
20	QR	3.8	1	1	1	$1 \cdot 10^{-14}$
(25)	RSL	3.3	0.86	0.66	1.06	$4 \cdot 10^{-11}$
	LGR	2.5	0.49	0.52	0.21	$2 \cdot 10^{-08}$
50	QR	3.2	1	1	1	$(6 \cdot 10^{-14})_1$
(49)	RSL	2.1	0.56	$***$	0.35	$4 \cdot 10^{-10}$
	LGR	2.4	0.49	0.53	0.1	$(4 \cdot 10^{-08})_1$
100	QR	2.8	1	1	1	$(6 \cdot 10^{-14})_1$
	RSL	2.4	0.51	$***$	$***$	$7 \cdot 10^{-09}$

In the last column the numbers in brackets $(\cdot)_1$ are averages of only 5 out of 6 examples. In these cases an extreme outlier has been removed.

Summarizing we find that the LGR algorithm is an efficient method to compute eigenvalue approximations of non-Hermitian matrices. On the average, it requires less flops than the QR algorithm but it is in general not as accurate as the QR algorithm.

Outlook: In future work the LGR method will be compared with the DQR algorithm [74] and the BR algorithm [23].

References

[1] W. E. ARNOLDI, The principle of minimized iterations in the solution of the matrix eigenvalue problem, Quart. Appl. Math. 9 (1951), pp. 17-29.

[2] Z. BAI, D. DAY, J. DEMMEL, AND J. DONGARRA, A Test matrix Collection for Non-Hermitian Eigenvalue Problems, Tech. Rep., University of Kentucky, Lexington, KY, (1996). Available from ftp.ms.uky.edu/pub/misc/bai.

[3] F. L. BAUER, Sequential reduction to tridiagonal form, J. SIAM 7 (1959), pp. 107-113.

[4] P. BENNER AND H. FASSBENDER, An implicitly restarted symplectic Lanczos method for the Hamiltonian eigenvalue problem, Lin. Alg. Appl. 263 (1997), pp. 75-111.

[5] D.L. BOLEY, G.H. Golub, The nonsymmetric Lanczos algorithm and controllability, Syst. Control. Lett. 16 (1991), pp. 97-105.

[6] Turbo Pascal 7.0 Compiler of Borland International, 1992.

[7] A. BUNSE-GERSTNER, Der HR-Algorithmus zur Bestimmung der Eigenwerte einer Matrix, Ph.D. Thesis, Bielefeld, 148, 1978.

[8] A. BUNSE-GERSTNER, An analysis of the HR algorithm for computing the eigenvalues of a matrix, Linear Algebra Appl. 35 (1981), pp. 155-178.

[9] A. BUNSE-GERSTNER, On the similarity transformation to tridiagonal form, Lin. Multilin. Alg. 12 (1982), pp. 51-56.

[10] W. BUNSE AND A. BUNSE-GERSTNER, Numerische lineare Algebra, Teubner, Stuttgart, 1985.

[11] A. BUNSE-GERSTNER AND V. MEHRMANN, A symplectic QR-like algorithm for the solution of the real algebraic Riccati equation, IEEE Trans. Automat. Control AC-31 (1986), pp. 155-178.

[12] A. BUNSE-GERSTNER, V. MEHRMANN AND D.WATKINS, An SR algorithm for Hamiltonian matrices based on Gaussian elimination, Methods Oper. Res. 58 (1989), pp. 339-358.

[13] H. J. BUUREMA, A geometric proof of convergence for the QR method, Doctoral Dissertation, Univ. of Groningen, 1970.

[14] F. CHATELIN AND D. HO, Arnoldi-Tchebychev procedure for large scale nonsymmetric matricies, Math. Model. Numer. Anal. 24 (1990), pp. 53-65.

[15] B. A. CIPRA, An Introduction to the Ising Model, American Mathematical Monthly 94 (1987), pp. 937-959.

[16] J. DELLA-DORA, Numerical linear algorithms and group theory, Linear Algebra Appl. 10 (1975), 267-283.

[17] G. DE SAMBLANX AND A. BULTHEEL, Nested Lanczos: implicitly restarting an unsymmetric Lanczos algorithm, Numerical Algorithms 18 (1998), pp. 31-50.

[18] A. DRAUX, Polynômes Orthogonaux Formels - Applications, Lect. Notes in Math. 974, Springer Verlag, Berlin, 1983.

[19] J. G. F. FRANCIS, The QR transformation: A unitary analogue to the LR transformation, Part I and II, Comp. J. 4 (1961), pp. 265-72, pp. 332-45.

[20] R. W. FREUND, M. H. GUTKNECHT AND N. M. NACHTIGAL, An implementation of the look-ahead Lanczos algorithm for non-Hermitian matrices, SIAM J. Sci. Comput. 14 (1993), pp. 137-158.

[21] R. W. FREUND AND H. Zha, Simplifications of the nonsymmetric Lanczos process and a new algorithm for solving indefinite symmetric linear systems, Technical Report, RIACS, NASA Ames Research Center, Moffet Field, 1991.

[22] B. FRIEDMAN, Eigenvalues of composite matrices, Proc. Cambridge Philos. Soc. 57 (1961), pp. 37-49.

[23] G. A. GEIST, G. W. HOWELL AND D. S. WATKINS, The *BR* eigenvalue algorithm, SIAM J. Matrix Anal. Appl. 20 (1999), pp. 1083-1098.

[24] A. GEORGE, K. IKRAMOV, A. N. KRIVOSHAPOVA, AND W.-P. TANG, A finite procedure for the tridiagonalization of a general matrix, SIAM J. Matrix Anal. Appl. 16 (1995), pp. 377-387.

[25] W. GIVENS, Numerical computation of the characteristic values of a real symmetric matrix, Report 1574, Oak Ridge National Laboratory, 1954.

[26] G. H. GOLUB AND C. REINSCH, Singular value decomposition and least squares solution, Numer. Math. 14 (1970), pp. 403-420.

[27] G. H. GOLUB AND C. F. VAN LOAN, Matrix Computations, 1nd Edition, The John Hopkins University Press, Baltimore, 1983.

[28] G. H. GOLUB AND C. F. VAN LOAN, Matrix Computations, 2nd Edition, The John Hopkins University Press, Baltimore, 1989.

[29] G. H. GOLUB, B. KÅGSTRÖM AND P. VAN DOOREN, Direct block tridiagonalization of single-input single-output systems, Systems & Control Letters 18 (1992), pp. 109-120.

[30] W. B. GRAGG, Matrix interpretations and applications of the continued fraction algorithm, Rocky Mountain J. Math. 4 (1974), pp. 213-225.

[31] W. B. GRAGG AND A. LINDQUIST, On the partial realization problem, Linear Alg. Appl. 50 (1983), pp. 277-319.

[32] E. GRIMME, D. SORENSEN AND P. VAN DOOREN, Model reduction of state space systems via an implicitly restarted Lanczos method, Numer. Algorithms 12 (1996), pp. 1-31.

[33] M. H. GUTKNECHT, A completed theory of the unsymmetric Lanczos process and related algorithms, Part 1, SIAM J. Matrix Anal. Appl. 13 (1992), pp. 594-639.

[34] M. H. GUTKNECHT, A completed theory of the unsymmetric Lanczos process and related algorithms, Part 2, SIAM J. Matrix Anal. Appl. 15 (1994), pp. 15-58.

[35] M. H. GUTKNECHT, Lanczos-type solvers for nonsymmetric linear systems of equations, Acta Numerica 6, Cambridge University Press (1997), pp. 271-397.

[36] K.P. HADELER, Reaction transport systems in biological modelling. In: V. Capasso, O. Diekmann ed., Mathematics inspired by biology, Lect. Notes in Math. 1714, Springer, Berlin, 1997.

[37] P. HENRICI, The quotient-difference algorithm, Nat. Bur. Standards Appl. Math. Series 19 (1958), pp. 23-46.

[38] N. J. HIGHAM, The Test Matrix Toolbox for Matlab, Tech. Rep. 276, University of Manchester, 1995.

[39] M. HOCHBRUCK, The Padé Table and its relation to certain numerical algorithms, Habilitationsschrift, Universität Tübingen, 1996.

[40] M. HOCHBRUCK AND C. LUBICH, Error analysis of Krylov methods in a nutshell, SIAM J. Sci. Comput. 19 (1998), pp. 695-701.

[41] A. S. HOUSEHOLDER AND F. L. BAUER, On certain methods for expanding the characteristic polynomial, Num. Math. 1 (1959), pp. 29-37.

[42] A. S. HOUSEHOLDER, The theory of matrices in numerical analysis, Blaisdell, New York, 1964.

[43] Z. JIA, The convergence of generalized Lanczos methods for large unsymmetric eigenproblems, SIAM J. Matrix Anal. Appl. 16 (1995), pp. 843-862.

[44] M. KAC, A stochastic model related to the telegrapher's equation, Rocky Mountain J. Math. 4 (1956), pp.497-509. (1956).

[45] R.E. KALMAN, Canonical structure of linear dynamical systems, Proc. Nat. Acad. Sci. USA, 48 (1962), pp. 596-600.

[46] B. KAUFMANN, Crystal statistics II, Phys. Rev. 76 (1949), pp. 1232-1949.

[47] T. KOCH AND J. LIESEN, The conformal bratwurst maps and associated Faber polynomials, Numer. Math. 86 (2000), pp. 173-191.

[48] L. KRONECKER, Bemerkungen zur Determinanten-Theorie, J. Reine Angew. Math. 72 (1870), pp. 152-175.

[49] V.N. KUBLANOVSKAYA, On some algorithms for the solution of the complete eigenvalue problem, Zh. Vychisl. Mat. 1 (1961), pp. 555-570.

[50] C. D. La Budde, The reduction of an arbitrary real square matrix to tridiagonal form using similarity transformations, Math. Comp. 17 (1963), pp. 443-447.

[51] C. LANCZOS, An iteration method for the solution of the eigenvalue problem of linear differential and integral operators, J. Res. Nat. Bur. Standards 45 (1950), pp. 255-282.

[52] R.B. LEHOUCQ, Analysis and Implementation of an Implicitly Restarted Iteration, PhD thesis, Rice University, Houston, Texas, May 1995.

[53] J. LIESEN, Computable convergence bounds for GMRES, SIAM Matrix. Anal. Appl. 21 (2000), pp. 882-903.

[54] M. MARCUS AND H. MINC, A survey of matrix theory and matrix inequalities, Dover edition, New York, 1992.

[55] R. B. MORGAN, On restarting the Arnoldi method for large nonsymmetric eigenvalue problems, Math. Comp. 65 (1996), pp. 1213-1230.

[56] B. N. PARLETT, The development and use of methods of LR type, SIAM, Rev. 6 (1964), pp. 275-295.

[57] B. N. PARLETT, Global convergence of the basic QR algorithm on Hessenberg matrices, Math. Comp. 22 (1968), pp. 803-817.

[58] B. N. PARLETT, Reduction to tridiagonal form and minimal realizations, SIAM J. Matrix Anal. Appl. 13 (1992), pp. 567-593.

[59] B. N. PARLETT, D. R. TAYLOR AND Z. A. LIU, A look-ahead Lanczos algorithm for unsymmetric matrices, Math. Comp. 44 (1985), pp. 1-22.

[60] P. RASCHMANN, M. KUBICEK AND M. MAROS, Waves in distributed chemical systems: experiments and computations, P. J. Holmes ed., New Approaches to Nonlinear Problems in Dynamics - Proceedings of the Asilomar Conference Ground, Pacific Grove, California 1979. The Engineering Foundation, SIAM (1980), pp.271-288.

[61] H. RUTISHAUSER, Beiträge zur Kenntnis des Biorthogonalisierungs-Algorithmus von Lanczos, Z. Angew. Math. Physik 4 (1953), pp. 35-56.

[62] H. RUTISHAUSER, Der Quotienten-Differenzen-Algorithmus, Z. Angew. Math. Physik 5 (1954), pp. 233-251.

[63] H. RUTISHAUSER, Solution of eigenvalue problems with the LR-transformation, Nat. Bur. Standards Appl. Math. Ser. 49 (1958), pp. 47-81.

[64] Y. SAAD, Variation on Arnoldi's method for computing eigenelements of large unsymmetric matrices, Lin. Alg. Appl. 34 (1980), pp. 269-295.

[65] Y. SAAD, The Lanczos biorthogonalization algorithm and other oblique projection methods for solving large unsymmetric systems, SIAM J. Numer. Anal. 19 (1982), pp. 485-506.

[66] Y. SAAD, Chebyshev acceleration techniques for solving nonsymmetric eigenvalue problems, Math. Comp. 42 (1984), pp. 567-588.

[67] Y. SAAD, Numerical methods for large eigenvalue problems, Manchester University Press, Manchester, 1992.

[68] E. SCHMIDT, Über die Auflösung linearer Gleichungen mit unendlich vielen Unbekannten, Rend. Circ. Mat. Palermo 25 (1908), pp. 59-77

[69] B. T. SMITH, J. M. BOYLE, J.J. DONGARRA, B. S. GARBOW, Y. IKEBE, V. C. KLEMA, AND C.B. MOLER, Matrix Eigensystem Routines-EISPACK Guide, 2nd ed., Springer-Verlag, New York, 1976.

[70] D. C. SORENSEN, Implicit application of polynomial filters in a k-step Arnoldi method, SIAM J. Matrix Anal. Appl. 13 (1992), pp. 357-385.

[71] C. STRACHEY AND J. G. F. FRANCIS, The reduction of a matrix to codiagonal form by eliminations. Computer Journal 4 (1961), pp. 168-176.

[72] O. TAUSSKY AND H. ZASSENHAUS, On the similarity transformation between a matrix and its transpose, Pacific J. Math. 9 (1959), pp. 893-896.

[73] D. R. TAYLOR, Analysis of the look-ahead Lanczos algorithm, Ph.D. thesis, Dept. of Mathematics, University of California, Berkeley, CA, 1982.

[74] F. UHLIG, The DQR algorithm, basic theory, convergence and conditional stability, Numerische Mathematik 76 (1997), pp. 515-553.

[75] H. A. VAN DER VORST AND G. H. GOLUB, 150 years old and still alive: Eigenproblems. Duff, I. S. (ed.) et al., The state of the art in numerical analysis. Based on the proceedings of a conference organized by the Institute of Mathematics and its Applications (IMA), University of York, York, GB, April 1-4, 1996. Oxford: Clarendon Press. Inst. Math. Appl. Conf. Ser., New Ser. 63 (1997), pp. 93-119.

[76] D. S. WATKINS AND L. ELSNER, Convergence of algorithms of decomposition type for the eigenvalue problem, Lin. Alg. Appl. 143 (1991), pp. 19-47.

[77] D. S. WATKINS AND L. ELSNER, Chasing algorithms for the eigenvalue problem, SIAM J. Matrix Anal. Appl. 12 (1991), pp. 374-384.

[78] G. WEBB, Existence and asymptotic behavior for a strongly damped nonlinear wave equation, Canad. J. Math. 32 (1980), pp. 631-643.

[79] J. H. WILKINSON, The Algebraic Eigenvalue Problem, Oxford University Press, London, 1965.

[80] J. H. WILKINSON AND C. REINSCH, Handbook for Automatic Computation, Vol. II, Linear Algebra, Springer-Verlag, New York, 1971.

[81] M. ZIEGLER, Generalized biorthogonal bases and tridiagonalisation of matrices, Numer. Math. 77 (1997), pp. 407-421.

Lebenslauf

Markus Ziegler

18.10.1968	geboren in Rosenfeld
1975-1979	Grundschule in Rosenfeld
1979-1985	Progymnasium in Rosenfeld
1985-1988	Technisches Gymnasium in Balingen
1988	Allgemeine Hochschulreife
1988-1990	Zivildienst
1990-1996	Studium der Mathematik und Physik an der Universität Tübingen
1995	Wissenschaftliche Arbeit (für das Lehramt an Gymnasien) über die Tridiagonalisierung von Matrizen
1996	1. Staatsexamen für das Lehramt an Gymnasien
seit 1996	Promotion an der Universität Tübingen
1996-1997	Wissenschaftlicher Angestellter der Universität Tübingen im SFB 382 *Verfahren und Algorithmen zur Simulation* *physikalischer Prozesse auf Höchstleistungsrechnern*
1997-1999	Studienreferendar für das Lehramt an Gymnasien (Seminar Rottweil)
1999	2. Staatsexamen für das Lehramt an Gymnasien
seit 1999	verheiratet mit Sibylle Conzelmann
1999-2000	Studienassessor, tätig am Gymnasium Spaichingen
seit 2000	Studienrat, tätig am Gymnasium Spaichingen

Meine akademischen Lehrer waren

in Mathematik:	K.P. Hadeler, C. Hering, G. Huisken, W. Kaup, R. Nagel, H. Yserentant
in Physik:	M. Baumann, W.D. Hartmann, R. Hübener, O. Lutz, H. Müther, A. Schwenk, G. Staudt, G. Wagner.