BACHELOR THESIS

Michal Outrata

Approaches to analysis of Krylov subspace methods

Department of Numerical Mathematics

Study programme: Mathematics
Study branch: General Mathematics

Prague 2016
I declare that I carried out this bachelor thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In ....... date ............ signature of the author
Title: Approaches to analysis of Krylov subspace methods

Author: Michal Outrata

Department: Department of Numerical Mathematics

Supervisor: prof. Ing. Zdeněk Strakoš, DrSc., Department of Numerical Mathematics

Abstract: The text deals with the understanding of the convergence behaviour of the GMRES method. The first part reviews results formulated for the linear algebraic finite-dimensional problem $Ax = b$. The second part revisits the question on whether the algebraic GMRES behaviour can be analyzed using bounded operators on an infinite-dimensional Hilbert spaces.

Keywords: Krylov Subspace Methods, GMRES, Convergence Behaviour, Linear Bounded Operators, Resolvent of Linear Bounded Operator
Dedication

I am most grateful to my supervisor for his time, patience and guidance. My gratitude also belongs to Jan Papež for his numerous consultations. Last but not least, I would like to thank to Iveta Hnětýnková for assistance and inspiration.
Contents

1 Krylov subspace methods 2
   1.1 Basic notation and the general framework ................. 2
   1.2 Projection methods ........................................ 3
   1.3 Krylov subspace methods .................................... 4

2 The GMRES method and its convergence behaviour 8
   2.1 GMRES basic properties and motivation .................... 8
   2.2 Matrices that generates the same Krylov residual subspaces . 10
   2.3 Convergence behaviour of GMRES and the eigenvalues ........ 13

3 Behaviour of algebraic methods and operators in infinite-dimensional Hilbert or Banach spaces 23

Conclusion 27

Bibliography 28
1. Krylov subspace methods

1.1 Basic notation and the general framework

This text deals with methods for solving a linear algebraic problem

\[ Ax = b, \]

with \( A \) being an \( N \times N \) nonsingular matrix (complex or real) and \( b \) is a (complex or real) vector of the length \( N \). It examines theoretical aspects of particular methods for solving the linear algebraic problem (1.1). Along with the notation, most of the statements in this chapter are adopted from [Liesen and Strakoš, 2015, Chapters 1 and 2]. The references are always stated explicitly.

Problems in form of (1.1) often arise as approximations of infinite-dimensional problems; see Figure 1.1 ([Liesen and Strakoš, 2015, Chapter 1, Figure 1.1]).

![Figure 1.1](image)

Though the computation of (1.1) is performed on the algebraic level, one should never forget that the ultimate goal is to solve the original problem. Therefore, each of the steps in Figure ?? should be treated with respect to the others. This requires, in particular, that one should always understand the circumstances, from which the given linear algebraic system arose. Otherwise, a useful information for handling the system (1.1) properly is lost. Readers interested in this "computation-modelling philosophy" can find more in, e.g., [Liesen and Strakoš, 2015, Chapter 1] or [Málek and Strakoš, 2014, Chapter 1]. These ideas combined with the Figure ?? result into the following major observations:

(a) One does not typically aim at the exact solution of (1.1). The computational error should be in balance with the error in the model, the discretization error and the algebraic error. In other words, to make the computation cost-effective, one intentionally aims at stopping the computation whenever the sufficient (problem-dependent) accuracy is reached.
(b) One should also not forget that some problems are uncomputable \textit{in principle} (such as determining the exact values of eigenvalues of a general matrix $A$). Moreover, numerical computations are performed in finite precision arithmetic, which can limit the maximal attainable accuracy of the computed approximations.

(c) The linear system is often sparse and gargantuan ($10^7 - 10^9$ unknowns is standard in large scale computations);

The property (c) can be viewed as a consequence of the first two steps of the solution process described in Figure 7. The sparsity as well as the large dimension of the problem arise from the mathematical formulation and discretization of the problem (using, e.g., the finite elements method (FEM); see [Gockenbach 2006, Chapters 1 to 5]). For example, studying the flow of a medium in some area, it is natural to decompose the area into a large amount of small areas and proceed almost separately on those smaller ones. That makes sense since one usually assumes that the interactions of the medium are only \textit{local}. The sparsity will then have a regular block structure. For more detailed reasoning see, e.g., [Gockenbach 2006, Section 4.1.2]. Collecting the thoughts presented above, one should aim for methods that

- will be \textit{safe} and \textit{robust} (will not uncontrollably increase the errors present in the data and will work for a substantially class of problems with similar performance);
- will make a good use of a possible sparsity;
- will not necessarily try to compute the exact solution of the system (1.1), but rather will provide a (sufficiently) good approximation;
- will avoid costly operations with the system matrix, which could lead to unfeasible memory and computational requirements.

Such thoughts are reflected in the development of the \textit{iterative methods}, such as \textit{SOR} ([Saad 2003, Chapter 4]) or the \textit{Chebyshev semi-iterative method} (for summary see, e.g., [Liesen and Strakoš 2015, Section 5.5.2]). The origins of the Chebyshev semi-iterative method can be linked with the works [Flanders and Shortley 1950], [Lanczos 1953] and [Young 1954], according to [Gergelits and Strakoš 2014]. Another important representative of the iterative methods are \textit{Krylov subspace methods} (for summary see, e.g., [Liesen and Strakoš 2015, Chapter 2]). Some interesting notes about the evolution and the history of the Krylov subspace methods can be found in [Liesen and Strakoš 2015, Chapter 2, Section 2.5.7].

1.2 Projection methods

As the name suggests, the idea behind the projection methods is to successively project the original system onto some, carefully chosen, subspaces of $\mathbb{C}^N$ of dimensions \textit{much} smaller than $N$. In this way, one obtains smaller systems, which
will yield the approximations of the solution. It follows that there are two essential ingredients for construction of a projection method - the subspaces and the instruction of how shall one project onto the subspaces.

**Definition 1.1.** Consider the problem $Ax = b$; see (1.1); and an initial guess $x_0$. Furthermore, take two nested sequences of subspaces

$$S_1 \subset S_2 \subset \ldots \subset \mathbb{C}^N$$

$$C_1 \subset C_2 \subset \ldots \subset \mathbb{C}^N$$

such that the dimensions of $S_k$ and $C_k$ are equal to $k$, i.e.,

$$\dim(S_k) = \dim(C_k) = k, \quad \text{ for } k = 1, 2, \ldots.$$ 

A method belongs to the family of iterative projection methods with the search spaces $S_1, S_2, \ldots$ and the constraint spaces $C_1, C_2, \ldots$ providing that the following holds

(i) at the $k$-th step the method computes an approximation $x_k$ and the corresponding residual $r_k := b - Ax_k$;

(ii) $x_k \in x_0 + S_k, \quad k = 1, 2, \ldots$;

(iii) $r_k \perp C_k, \quad k = 1, 2, \ldots$.

The definition above formalizes the thoughts of the first paragraph of this section. Indeed, condition (iii) fixes the $k$ degrees of freedom, which arises in the choice of the approximation $x_k$ according to condition (ii). To express the projection explicitly in the matrix form, take $S_k$ and $C_k$ as $N \times k$ matrices, with their columns forming orthonormal bases of $S_k$ and $C_k$ respectively. The condition (ii) of Definition [1.1] implies that there exists $t_k \in S_k$ such that $x_k = x_0 + S_k t_k$.

and from the condition (iii) one observes that

$$0 = C_k^* r_k = C_k^* (b - A(x_0 + S_k t_k)) = C_k^* r_0 - C_k^* AS_k t_k.$$ 

If $C_k^* AS_k$ is nonsingular, then the above equality implies that

$$t_k = (C_k^* AS_k)^{-1} C_k^* r_0$$

$$x_k = x_0 + S_k (C_k^* AS_k)^{-1} C_k^* r_0$$

$$r_k = r_0 - AS_k (C_k^* AS_k)^{-1} C_k^* r_0.$$ 

(1.2)

**Proposition 1.1** ([Liesen and Strakos 2015, Chapter 2, Theorem 2.1.2]). Consider the $k$-th step of a projection method described in Definition [1.1]. Take matrices $S_k$ and $C_k$ as the matrices considered in (??). Then $x_k$ is uniquely determined if and only if $C_k^* AS_k$ is nonsingular. This is equivalent to $C_k^+ \oplus AS_k = \mathbb{C}^N$.

The most natural way to ensure the latter condition, i.e.,

$$C_k^+ \oplus AS_k = \mathbb{C}^N.$$ 

4
is to consider $C_k = A S_k$. Such projection methods are called orthogonal since they perform the orthogonal decomposition of the initial residual; see [Liesen and Strakoš, 2015, Section 2, Definition 2.1.3]. Indeed, if $C_k^* A S_k$ is nonsingular and one considers the operator

$$P_k := A S_k (C_k^* A S_k)^{-1} C_k^*$$

then

$$P_k^2 = A S_k (C_k^* A S_k)^{-1} C_k^* A S_k (C_k^* A S_k)^{-1} C_k^* = A S_k (C_k^* A S_k)^{-1} C_k^* = P_k.$$  

Thus $P_k$ is a projector onto $A S_k$ implying that also the operator $I - P_k$ is a projector onto $A S_k$. One can see from (1.3) that $r_k$ is the projection of $r_0$ under $I - P_k$.

A subspace $S \subset \mathbb{C}^N$ is called $A$-invariant, provided $A S = S$. The following proposition shows that $A$-invariant subspaces play an important role in the projection methods.

**Proposition 1.2** ([Liesen and Strakoš, 2015, Chapter 2, Lemma 2.1.7]). A projection method finds the exact solution, i.e., $r_k = 0$ if and only if $r_0 = P_k r_0$. That is equivalent to $r_0 \in A S_k$. Consider the initial search space $S_1 = \text{span}\{r_0\}$ and build up the nested sequence of the search and the constraint spaces as in Definition 1.1. If the search space $S_k$ is $A$-invariant for some $k$, then the solution is found at the step $k$, i.e., $x = x_k$.

Considerations in Proposition 1.2 naturally lead to using the Krylov subspaces in the projection methods for deriving the search and the constraint spaces, as is shown in the following section. For a more detailed description of the framework of projection methods see, e.g., [Liesen and Strakoš, 2015, Chapter 2], [Eiermann and Ernst, 2001, Sections 2 and 3] or [Saad, 2003, Chapter 5].

### 1.3 Krylov subspace methods

We will now consider projection methods with the search and the constraint spaces derived from the Krylov subspaces.

**Definition 1.2.** Consider an $N \times N$ matrix $A$ and a nonzero vector $v$ of the length $N$. Then, given $k \leq N$, define the $k$-th Krylov subspace associated with the matrix $A$ and the vector $v$ as the space $K_k(A,v)$ generated by the first $k$ powers of $A$ with respect to $v$,

$$K_k(A,v) := \text{span}\{v, Av, A^2 v, \ldots, A^{k-1} v\}.$$  

Furthermore, define the degree of the vector $v$ with respect to the matrix $A$ as the smallest integer $d = d(A,v)$ such that the set $\{v, Av, \ldots, A^{d-1} v\}$ is linearly independent and the set $\{v, Av, \ldots, A^d v\}$ is not.

The following theorem gives a mathematical description of some Krylov subspace methods in terms of the search and the constraint spaces, supplemented by characterizations in terms of the approximations properties.
Theorem 1.1 \cite{Liesen and Strakos 2015, Theorem 2.3.1}. Consider the problem $Ax = b$ (see (1.1)) and an initial guess $x_0$. Taking $r_0 = b - Ax_0$ and $d = d(A, r_0)$, the following hold

1. If $A$ is Hermitian and positive definite and one takes $S_k = C_k = K_k(A, r_0)$ for $k = 1, 2, \ldots$, then the projection method generated by these spaces is well-defined and terminates exactly at step $d$. Moreover, at the $k$-th step of the method, one has the following characterizations for $x_k$:

- $x - x_k \perp_A K_k(A, r_0)$,
- $\|x - x_k\|_A = \min_{z \in x_0 + K_k(A, r_0)} \|x - z\|_A$.

(Mathematical characterization of the Conjugate Gradient method (CG); see \cite{Hestenes and Stiefel 1952}.)

2. If $A$ is Hermitian and nonsingular and one takes $S_k = A K_k(A, r_0)$ and $C_k = K_k(A, r_0)$, $k = 1, 2, \ldots$, then the projection method generated by these spaces is well-defined and terminates exactly at step $d$. Moreover, at the $k$-th step of the method, one has the following characterizations for $x_k$:

- $x - x_k \perp_A K_k(A, r_0)$,
- $\|x - x_k\| = \min_{z \in x_0 + A K_k(A, r_0)} \|x - z\|_A$.

(Mathematical characterization of the SYMMLQ method; see \cite{Paige and Saunders 1975, Section 2, Case (a)}.)

3. If $A$ is nonsingular and one takes $S_k = K_k(A, r_0)$, $C_k = A K_k(A, r_0)$, then the projection method generated by these spaces is well-defined and terminates exactly at step $d$. Moreover, at the $k$-th step of the method, one has the following characterizations for $x_k$:

- $x - x_k \perp_{A^* A} K_k(A, r_0)$,
- $\|x - x_k\|_{A^* A} = \min_{z \in x_0 + K_k(A, r_0)} \|z - x\|_{A^* A}$,

or, equivalently, in terms of residual

- $r_k \perp A K_k(A, r_0)$,
- $\|r_k\| = \min_{z \in x_0 + K_k(A, r_0)} \|b - A z\|$.

(Mathematical characterization of the GMRES method \cite{Saad and Schultz 1986}. Provided $A$ is symmetric, the above characterize the MINRES method \cite{Paige and Saunders 1975, Section 2, Case (b)}.)

It is important to realize that the use of Krylov subspaces for deriving the search and the constraint spaces is not a voluntary choice. In fact it is a very natural one as stated bellow Proposition 1.2. If one wishes to approximate the problem, the aim should be on approximating the dominant features first and, as the iterations go on, refine the outputs. That is precisely the case with the Krylov subspaces. At the beginning, only the basic information about the pair $A, r_0$ lies
within the Krylov subspace, whereas with the growing \( k \), \( K_k(A,r_0) \) approximates the desired \( A \)-invariant subspace more closely. For further reading on this topic, one may look into [Liesen and Strakoš, 2015, Sections 2.2 to 2.6 and Figure 2.3] or [Saad, 2003, Sections 6.1 and 6.2].
2. The GMRES method and its convergence behaviour

The following chapter focuses on the GMRES method and summarizes the triplet of papers Greenbaum and Strakoš [1994], Greenbaum et al. [1996] and Arioli et al. [1998], which study the convergence behaviour of GMRES. For that purpose, it is necessary to start with basic observations and propositions and then present the line of development of these articles with the associated arguments.

2.1 GMRES basic properties and motivation

The section starts with a basic but important observation about the optimality characterization of the GMRES method given in Theorem 1.1.

**Proposition 2.1.** Consider the problem $Ax = b$ (see (1.1)) and an initial guess $x_0$. The GMRES method can be also characterized via the following optimality properties

$$
||r_k|| = \min_{z \in r_0 + AK_k(A,r_0)} ||z||, \quad k = 1, 2, \ldots,
$$

or, equivalently,

$$
||r_k|| = \min_{p \in \Pi_k} ||p(A)r_0||, \quad k = 1, 2, \ldots,
$$

with $r_0 = b - Ax_0$ and $\Pi_k$ being the set of polynomials of a degree less or equal to $k$ and with constant term equal to one. The sequence of the residual norms $\{|r_0|, |r_1|, \ldots\}$ will be denoted by $GMRES(A,r_0)$.

**Proof.** One can write

$$
\min_{z \in r_0 + AK_k(A,r_0)} ||z|| = \min_{z \in x_0 + K_k(A,r_0)} ||b - Az||.
$$

Observing that vectors in $r_0 + AK_k(A,r_0)$ are of form $p(A)r_0$ for some $k$-th degree polynomial with constant term equal to one, the second reformulation follows. □

**Remark 2.1 (Fischer [1996]).** The fact that the $k$-th residual (or $k$-th error) can be expressed as a certain polynomial of degree $k$ in the system matrix applied to the initial residual (initial error), is a common feature for the Krylov subspace methods. Hence, the Krylov subspace methods are sometimes called polynomial methods. This will be further discussed in the last chapter.

**Assumption 2.1.** For the rest of the work we will assume that $d(A,r_0) = N$, i.e., there is no early termination for GMRES applied to (1.1). The general case can be discussed in the same manner.

A bound on the residual (error) norms of an iterative method (e.g. GMRES) is called sharp, provided at each step of the method there are initial data (dependent on the particular step) such that the bound is reached at the considered step.

With the GMRES method, the “goodness” of the $k$-th approximation is measured with the norm of the $k$-th residual. Deriving a sharp (or nearly sharp)
upper bound on the convergence behaviour in terms of the convergence rate, i.e., the ratio $||r_k||/||r_0||$, is hence an important task in the convergence behaviour analysis.

Consider the problem $Ax = b$; see (1.1). Assume that the system matrix $A$ has a complete set of eigenvectors. Take $Z$ as a matrix with columns formed by those vectors and take $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_N)$ where $\lambda_1, \ldots, \lambda_N$ stands for the spectrum of $A$. A well-known bound on the GMRES residual norms in terms of the eigenvalues of the matrix and the condition number of the eigenvector matrix can be derived as follows (see, e.g., [Saad, 2003, Chapter 6, Proposition 6.32]).

$$
||r_k|| = \min_{p \in \Pi_k} ||p(A)r_0|| = \min_{p \in \Pi_k} ||Zp(\Lambda)Z^{-1}r_0|| \leq \kappa(Z) \min_{p \in \Pi_k} ||p(\Lambda)|| \cdot ||r_0|| = \kappa(Z) ||r_0|| \min_{p \in \Pi_k, \lambda \in \text{sp}(A)} ||p(\lambda)||,
$$

(2.1)

where $\kappa(Z) = ||Z|| \cdot ||Z^{-1}||$ is the condition number of the matrix $Z$.

If the matrix $A$ is moreover normal, i.e., $A^*A = AA^*$, then $Z$ is unitary, i.e., $||Z|| = ||Z||^{-1} = 1$ and instead of (2.1) one gets the following upper bound on the convergence rate of GMRES

$$
\frac{||r_k||}{||r_0||} \leq \min_{p \in \Pi_k, \lambda \in \text{sp}(A)} ||p(\lambda)||.
$$

(2.2)

As shown in [Greenbaum and Trefethen, 1994, Section 4] the bound (2.2) is sharp (for normal matrices). Thereby, the convergence behaviour of GMRES applied to a linear algebraic problem with a normal matrix is in this sense governed by the eigenvalues of the matrix. It should be understood, however, that (2.2) is valid for any initial residual, i.e., it represents the so called worst case bound. For particular initial residual the bound (2.2) can be a very poor indicator of the real convergence behaviour.

Moreover, in many cases of interest the matrix is not normal and then the bound (2.1) may be a large overestimate. Looking for a “sharper” bound, in [Greenbaum and Strakoš, 1994, Section 1, (1.11)] the authors refer to an upper bound on the convergence rate based on the $\varepsilon$-pseudo-eigenvalues.

**Definition 2.1** ([Trefethen and Embree, 1999, Introduction, Definition 1 and 2]).

Let $A$ be an $N \times N$ matrix and $\varepsilon$ a nonnegative scalar. One says that $\lambda$ is an $\varepsilon$-pseudo-eigenvalue of $A$, provided there exists an $N \times N$ perturbation matrix $E$ such that $||E|| \leq \varepsilon$ and $\lambda$ is an eigenvalue of the matrix $A + E$.

However, the example [Greenbaum and Strakoš, 1994, Section 1, Example (1.12)] proves that those bounds can be large overestimates. Summarizing with a quote from [Greenbaum and Strakoš, 1994, Section 1]:

“The problem of deriving a sharp bound on the convergence rate of the GMRES method, in terms of the eigenvalues or the pseudo-eigenvalues or other simple properties of the matrix, appears to be a difficult one.”

The authors approach this problem in the following way. Consider the problem $Ax = b$ (see (1.1)) and assume that the system matrix $A$ is not normal. Suppose
there exists a normal matrix $B$ such that the problem (1.1) with $B$ behaves the same as it does with $A$ (with respect to the GMRES method with the fixed initial residual). Then one has the sharp upper bound (2.2) in terms of the eigenvalues of $B$. If one could relate the spectrum of $B$ to some properties of $A$, then one would obtain the sharp bound in terms of these properties of $A$.

The immediate question arises, whether for any given problem one can find such matrix $B$. The answer is affirmative, as will be shown in the following section. However, relating the spectrum of the normal matrix to some properties of the former system matrix is a difficult problem and is, up to our knowledge, not solved.

### 2.2 Matrices that generates the same Krylov residual subspaces

Let us point out that we assume that there is no early termination in the GMRES method (see Assumption 2.1).

Consider the problem $Ax = b$ (see (1.1)) with two different matrices $A$ and $B$ and the same right-hand side vector $b$ and an initial guess $x_0$. From Proposition 2.1 is clear that if $A\mathsf{K}_k(A, r_0) = B\mathsf{K}_k(B, r_0)$, $k = 1, 2, \ldots, N$, then the GMRES method applied to $B$ with the initial residual $r_0$ will generate the same residual vectors as if applied to $A$ with initial residual $r_0$, since the minimisation is done over the same subspaces. Hence, the important task follows.

Consider the problem $Ax = b$ (see (1.1)) and an initial guess $x_0$. Find all matrices $B$ such that $A\mathsf{K}_k(A, r_0) = B\mathsf{K}_k(B, r_0)$ for all $k = 1, \ldots, N$.

The characterization of such matrices from Greenbaum and Strakoš [1994, Section 2] is given below. We also adopt the notation from this article, as follows.

Using the Arnoldi process (see Liesen and Strakoš [2015, Section 2.4]) on the matrix $A$ and the initial residual $r_0$, one gets orthonormal vectors $w_1, \ldots, w_N$ such that for any $k \leq N$ the vectors $w_1, \ldots, w_k$ form an orthonormal basis of the $k$-th Krylov residual subspace associated with $A$ and $r_0$, i.e.,

$$\text{span}\{w_1, \ldots, w_k\} = \text{span}\{A r_0, \ldots, A^k r_0\}, \quad k = 1, \ldots, N. \quad (2.3)$$

Taking the unitary matrix $W = [w_1 | w_2 | \cdots | w_N]$, the final step of the Arnoldi process applied to $A$ and $r_0$ gives the Arnoldi decomposition of the matrix $A$

$$AW = WH, \quad (2.4)$$

with $H$ being an $N \times N$ unreduced upper Hessenberg matrix. This means that all entries bellow the first subdiagonal of $H$ are zeros and the entries on the first subdiagonal are nonzero. The following theorem gives the characterization of the matrices that generate the same Krylov residual subspaces for a given initial residual $r_0$. 

10
Theorem 2.1 (Greenbaum and Strakoš 1994 Theorem 2.2]). Consider the problem \( Ax = b \) (see (1.1)) and an initial guess \( x_0 \). Take \( W \) as the matrix from (2.4). Let \( B \) be an \( N \times N \) matrix of the form
\[
B = W \hat{R} \hat{H} W^*,
\]
where \( \hat{R} \) is a nonsingular upper triangular matrix and \( \hat{H} \) is taken as
\[
\hat{H} = \begin{bmatrix}
0 & \cdots & 0 & 1/ \langle r_0, w_N \rangle \\
1 & \cdots & 0 & - \langle r_0, w_1 \rangle / \langle r_0, w_N \rangle \\
\vdots & \ddots & \vdots & \vdots \\
0 & \cdots & 1 & - \langle r_0, w_{N-1} \rangle / \langle r_0, w_N \rangle 
\end{bmatrix}.
\]
Then
\[
A K_k(A, r_0) = B K_k(B, r_0), \quad k = 1, \ldots, N.
\]
(2.7)
Conversely, any matrix \( B \) that satisfies (2.7) is of the form (2.5, 2.6).

Proof. To prove the statement, it is useful to consider the following observation.

Condition (2.7) is met if and only if there exists a nonsingular upper triangular matrix \( R_1 \) such that
\[
\begin{bmatrix} A r_0 & A^2 r_0 & \cdots & A^N r_0 \end{bmatrix} R_1 = B \begin{bmatrix} r_0 & w_1 & \cdots & w_{N-1} \end{bmatrix}.
\]
(2.8)

To verify this, observe that (2.8) holds if and only if the following assertions hold
- \( Br_0 \) is a nonzero multiple of \( A r_0 \) (and hence of \( w_1 \); see (2.7))
- \( Bw_1 \) is a linear combination of \( A^2 r_0 \) and \( A r_0 \), with the coefficient of \( A^2 r_0 \) being nonzero. Consequently, the same goes for \( B^2 r_0 \).
- \( Bw_2 \) is a linear combination of \( A^3 r_0, A^2 r_0 \) and \( A r_0, \ldots \)
- Proceeding in this way up to the final step yields the observation.

Thanks to (2.3) there exists a matrix \( R_2 \) such that
\[
\begin{bmatrix} A r_0 & A^2 r_0 & \cdots & A^N r_0 \end{bmatrix} = W R_2.
\]
(2.9)

Another simple observation follows as
\[
\begin{bmatrix} r_0 & w_1 & \cdots & w_{N-1} \end{bmatrix} = \begin{bmatrix} \langle r_0, w_1 \rangle & 1 & 0 & \cdots & 0 \\
\langle r_0, w_2 \rangle & 0 & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\langle r_0, w_{N-1} \rangle & 0 & 0 & \cdots & 1 \\
\langle r_0, w_N \rangle & 0 & 0 & \cdots & 0 
\end{bmatrix}.
\]
(2.10)

Denoting the last matrix from (2.10) by \( M \), observe that \( M^{-1} \) exists. That is the case because \( \langle r_0, w_n \rangle \neq 0 \) (thanks to the Assumption 2.1). Computing the inverse of \( M \) imply that \( M^{-1} \) is equal to \( \hat{H} \) from the statement; see (2.6). Multiplying (2.10) with the matrix \( B \), one obtains
\[
B W M = \begin{bmatrix} A r_0 & A^2 r_0 & \cdots & A^N r_0 \end{bmatrix} R_1 = W R_2 R_1,
\]
where the first equality is due to the observation in the beginning and the second one follows from (2.9). It follows that \( B = W R_2 R_1 \hat{H} W^* \), which completes the proof with \( \hat{R} = R_2 \hat{R}_1 \).
Consider the set of all matrices, for which GMRES provides the same residual norms with a fixed right-hand side vector \( b \) and an initial guess \( x_0 \). Denoting this set \( \text{GMRES}_A(\cdot, r_0) \), the aim is at finding a normal matrix in this set, as suggested at the end of the previous section. Two particular cases of normal matrices contained in this set are given below.

**Unitary matrices** Consider the same setting as before, i.e., the problem \( Ax = b \); see (1.1); and an initial guess \( x_0 \). The Arnoldi process applied to \( A \) and \( r_0 \) terminates after \( N \) steps with the Arnoldi decomposition \( AW = WH \) (see (2.4)). Performing the RQ-decomposition of the matrix \( H \), one gets matrices \( Q \) and \( R \) such that \( H = RQ \). Considering the matrix \( WR^{-1}HW^* = WQW^* \), Theorem 2.1 implies that it is contained in the set \( \text{GMRES}_A(\cdot, r_0) \). But surely, the matrix \( WQW^* \) is unitary. Hence, any convergence behaviour that can be seen with GMRES can be also seen with GMRES applied to the problem with a unitary matrix.

**Hermitian positive definite matrices** Consider the same setting as before, i.e., the problem \( Ax = b \); see (1.1); and an initial guess \( x_0 \). The Arnoldi process applied to \( A \) and \( r_0 \) terminates after \( N \) steps with the Arnoldi decomposition \( AW = WH \) (see (2.4)). If zero is outside the field of values of \( A \), i.e., for any nonzero \( N \) dimensional vector \( v \) one has that \( \langle v, Av \rangle \neq 0 \), then zero is also outside the field of values of \( H \). Performing the UL-decomposition of the matrix \( H \), one gets a nonsingular matrix \( L \) (lower triangular) and a nonsingular matrix \( U \) (upper triangular) such that \( UL = H \). This relation can be rewritten, using the fact that \( L \) is nonsingular, as

\[
A = WHW^* = WU (L^*)^{-1} L^* LW^*.
\]

Notice that the matrix \( (L^*)^{-1} \) is upper triangular. Choosing \( R = (U (L^*)^{-1})^{-1} \), Theorem 2.1 implies that the Hermitian positive definite matrix

\[
WR^{-1}HW^* = WL^* LW^*
\]

is also contained in \( \text{GMRES}_A(\cdot, r_0) \). Hence, any convergence behaviour that can be seen with GMRES for a problem with system matrix with zero outside its field of values can be also seen with GMRES applied to the problem with a system matrix that is **Hermitian and positive definite**.

An examples and also a further discussion of these particular cases of normal matrices contained in the set \( \text{GMRES}_A(\cdot, r_0) \) can be found in [Greenbaum and Strakos, 1994, Section 3].

But so far, for neither of the cases shown above have been established any connections between the spectrum of the normal matrices (unitary and Hermitian, positive definite respectively) and some properties of the matrix \( A \). Nonetheless, a different and interesting question arises from the results above.

Consider an \( N \times N \) nonsingular matrix \( A \) and an initial residual \( r_0 \). Taking \( N \) nonzero complex numbers \( \lambda_1, \ldots, \lambda_N \), is there always a matrix

\[ B \in \text{GMRES}_A(\cdot, r_0) \] such that \( \text{sp}(B) = \{ \lambda_i \}_{i=1}^N \)?

12
The answer is given in [Greenbaum and Strakoš, 1994, Section 4] and is affirmative. Such matrix can be constructed in the following way.

Consider the Arnoldi decomposition $AW = WH$ of the system matrix (see [2.4]). Take the monic polynomial

$$q(z) = \prod_{i=1}^{N} (z - \lambda_i)$$

and let $\alpha_i$ denote its coefficients, i.e.,

$$q(z) = z^N + \sum_{i=0}^{N-1} \alpha_i z^i.$$  \hspace{1cm} (2.11)

Take the companion matrix $C$ of the polynomial $q(z)$ and the upper triangular matrix $R$ as follows

$$C = \begin{bmatrix} 0 & 0 & \cdots & 0 & \alpha_0 \\ 1 & 0 & \cdots & 0 & \alpha_1 \\ 0 & 1 & \cdots & 0 & \alpha_2 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & \alpha_{N-1} \end{bmatrix}$$ \hspace{1cm} (2.12)

$$R = \begin{bmatrix} 1 & 0 & \cdots & 0 & \alpha_1 + \alpha_0 \langle r_0, w_1 \rangle \\ 0 & 1 & \cdots & 0 & \alpha_2 + \alpha_0 \langle r_0, w_2 \rangle \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & \alpha_{N-1} + \alpha_0 \langle r_0, w_{N-1} \rangle \\ 0 & 0 & \cdots & 0 & \alpha_0 \langle r_0, w_N \rangle \end{bmatrix}$$

Computing the product $\hat{H}R$, where $\hat{H}$ is taken of the form (2.6), it follows that $\hat{H}R = C$. Note that $R$ is nonsingular since

$$\alpha_0 = \prod_{i=1}^{N} \lambda_i \neq 0 \quad \text{and} \quad \langle r_0, w_n \rangle \neq 0.$$ 

The spectrum of $\hat{H}R$ is the same as the spectrum of $R\hat{H}RR^{-1} = \hat{H}$. Summarizing, the matrix $W^*\hat{H}RW$ is an element of $\text{GMRES}_A(\cdot, r_0)$ and at the same time has the desired spectrum.

As pointed out in [Greenbaum and Strakoš, 1994, Section 4], this result shows, in particular, that:

“In general, eigenvalue information alone cannot ensure fast convergence of the GMRES algorithm. Any behaviour that can be seen with the method can be seen with the method applied to a matrix having any nonzero eigenvalues”

This result opens up for the following questions.

Is any non-increasing convergence behaviour possible with the GMRES method? Is there a characterization for the pairs $\{A, r_0\}$ that produce a given non-increasing convergence curve? If $A$ has a fixed spectrum, are there always some initial data such that GMRES($A, r_0$) corresponds to some given non-increasing convergence curve? Is there any freedom left, if those two properties (spectrum and convergence curve) are fixed?

The following section will be dealing with those questions.
2.3 Convergence behaviour of GMRES and the eigenvalues

This section summarizes the papers [Greenbaum et al. 1996] and [Arioli et al. 1998], which answer the questions posed at the end of the previous section.

Let us point out that we assume that there is no early termination in the GMRES method (see Assumption 2.1).

**Assumption 2.2.** For the rest of this work it will be assumed without the loss of generality that $x_0 = 0$ (and hence $r_0 = b$).

Consider the problem $Ax = b$; see (1.1). Recalling Proposition 2.1, the $k$-th residual $r_k$ is chosen from the $k$-th Krylov residual subspace in such way that

$$||r_k|| = \min_{u \in A\mathcal{K}_k(A,b)} ||b - u||.$$

Hence, considering the orthonormal base $w_1, \ldots, w_k$ of the $k$-th Krylov residual subspace as in (2.3), it follows that $r_k$ is chosen as the orthogonal projection of $b$ onto the subspace $A\mathcal{K}_k(A,b)$, i.e.,

$$r_k = b - \sum_{i=1}^{k} \langle b, w_i \rangle w_i, \ k = 1, \ldots, N.$$

The last equality implies that

$$||r_{k-1}||^2 - ||r_k||^2 = \left< b - \sum_{i=1}^{k-1} \langle b, w_i \rangle w_i, b - \sum_{i=1}^{k-1} \langle b, w_i \rangle w_i \right> - \left< b - \sum_{i=1}^{k} \langle b, w_i \rangle w_i, b - \sum_{i=1}^{k} \langle b, w_i \rangle w_i \right> = \langle w_k, b \rangle^2.$$

As the immediate consequence one has

$$|\langle b, w_k \rangle| = \sqrt{||r_{k-1}||^2 - ||r_k||^2}, \ k = 1, \ldots, N - 1$$

$$|\langle b, w_N \rangle| = ||r_{N-1}||.$$

Note that the square roots above are well-defined, since the residual norms form a non-increasing sequence (see Theorem 1.1). The previous relation can also be expressed as

$$\Gamma W^* b = \left( \begin{array}{c} \sqrt{||b||^2 - ||r_1||^2} \\ \sqrt{||r_1||^2 - ||r_2||^2} \\ \vdots \\ \sqrt{||r_{N-2}||^2 - ||r_{N-1}||^2} \\ ||r_{N-1}|| \end{array} \right), \quad (2.13)$$

where $\Gamma = \text{diag}(\gamma_1, \ldots, \gamma_N)$ is the uniquely determined unitary diagonal matrix such that $\gamma_k \langle r_0, w_k \rangle = \sqrt{||r_{k-1}||^2 - ||r_k||^2}$, $k = 1, \ldots, N - 1$.

The above formulae yield the relation between the initial residual (in our case corresponding to the right-hand side vector $b$; see Assumption (2.2)) and
the convergence curve of the GMRES method (earlier denoted as GMRES(A, b)). Consider an arbitrary non-increasing sequence

\[ f(0) \geq f(1) \geq \ldots \geq f(N-1) > 0. \]  

(2.14)

The previous considerations imply that (2.13) corresponds to GMRES(A, b) if and only if

\[ \Gamma W^* b = \begin{pmatrix} \sqrt{f(0)^2 - f(1)^2} \\ \sqrt{f(1)^2 - f(2)^2} \\ \vdots \\ \sqrt{f(N-2)^2 - f(N-1)^2} \\ f(N-1) \end{pmatrix}. \]  

(2.15)

This can be shortened, for future use, to

\[ \Gamma W^* b = \begin{pmatrix} g(0)^2 \\ g(1)^2 \\ \vdots \\ g(N-2) \\ g(N-1) \end{pmatrix}, \]

with \( g(i) := \sqrt{f(i)^2 - f(i+1)^2} \) for \( i = 0, 1, \ldots, N-2 \) and \( g(N-1) = f(N-1) \). Note that computing the second power of the norm of both sides of (2.14) gives \( f(0) = ||b|| \).

Recalling the end of the previous section one can conclude that any non-increasing convergence behaviour is possible with GMRES and, moreover, the system matrix of such problem can have any nonzero eigenvalues. Such problem can be constructed in the following way.

Take \( N \) nonzero complex numbers corresponding to the desired spectrum of the system matrix. Let \( v_1, \ldots, v_N \) be an orthonormal base of \( \mathbb{C}^N \) and take any non-increasing curve in form of (2.13). Define the right-hand side vector \( b \) as the unique solution of the linear algebraic problem

\[ [v_1 \mid \cdots \mid v_N]^* b = (g(0), \ldots, g(N-1))^T. \]  

(2.16)

Thanks to Assumption 2.1 one gets \( ||b, v_N|| = f(N-1) > 0 \). This implies that \( \mathcal{B} := \{b, v_1, \ldots, v_{N-1}\} \) is a base of \( \mathbb{C}^N \) and any linear operator on \( \mathbb{C}^N \) is fully determined by values on this set. Define the operator \( \mathcal{A} \) as

\[ \mathcal{A} b = v_1 \]
\[ \mathcal{A} v_1 = v_2 \]
\[ \vdots \]
\[ \mathcal{A} v_{N-2} = v_{N-1} \]
\[ \mathcal{A} v_{N-1} = \alpha_0 b + \alpha_1 v_1 + \cdots + \alpha_{N-1} v_{N-1}, \]
with \( \{ \alpha_i \}_{i=1}^{N-1} \) taken as in (2.11), i.e., as the coefficients of the monic polynomial with roots corresponding to the desired spectrum of the system matrix. The construction implies that the operator \( \mathcal{A} \) with respect to the base \( \mathcal{B} \) is the companion matrix of the polynomial (2.11), i.e., \( \mathcal{A}_{\mathcal{B}} = C \) with \( C \) being of the form (2.12). Therefore, \( \mathcal{A} \) can be expressed in the standard Euclidean base \( \mathcal{E} = \{ e_1, \ldots, e_N \} \) as

\[
A := \mathcal{A}_{\mathcal{E}} = BCB^{-1},
\]

where \( B \) is the \( N \times N \) matrix with columns formed by the vectors of the base \( \mathcal{B} \). The construction further implies that

\[
\mathcal{A} \mathcal{K}_k(A,b) = \text{span}\{Ab, \ldots, A^k b\} = \text{span}\{v_1, \ldots, v_k\}, \quad k = 1, \ldots, N
\]

and at the same time (??) holds with \( ||r_k|| = f(k) \), i.e., \( b \) is the solution of the system (2.15). Therefore, the GMRES method indeed produce the considered non-increasing convergence curve (2.13).

Summarizing, it was shown that any convergence curve is attainable with GMRES, even with fixed spectrum of the system matrix. Moreover, there is still some freedom, i.e., the problem (system matrix) is not unique. This fact is captured in the choice of the base \( v_1, \ldots, v_N \). An interesting problem is the quantification of the freedom left, i.e., finding all such system matrices. Though one may object that the freedom is contained exactly in the choice of the orthonormal base, our focus is on finding a more explicit quantification, e.g., finding a characterization of these system matrices in a similar way as in Theorem 2.1. One would like to have a decomposition of these system matrices that would explicitly state, which entries of the decomposition are free parameters and which are forced by the assumptions on the convergence behaviour of GMRES and the spectrum of the matrix.

This issue is challenged and solved in [Arioli et al., 1998] and also later in this chapter. First, let us collect the observations that appeared in the construction part in Theorem 2.2 below.

**Theorem 2.2** ([Greenbaum et al., 1996, Theorem 3.2]). Consider the problem \( Ax = b \) (see (1.1)) and \( x_0 = 0 \). Take a non-increasing sequence in form of (2.13), i.e., \( f(0) \geq f(1) \geq \ldots \geq f(N-1) > 0 \). Then the following assertions are equivalent.

1. The GMRES residual vectors \( r_k \) satisfy \( ||r_k|| = f(k) \) for \( k = 0, 1, \ldots, N-1 \);

2. The system matrix \( A \) is of the form (2.5, 2.6) and equality (2.14) holds, i.e.,

\[
A = WR\hat{H}W^* \quad \text{and} \quad W^*b = (g(0), \ldots, g(N-1))^T,
\]

where \( R \) is any nonsingular upper triangular matrix, \( W = [w_1, \ldots, w_N] \) is a unitary matrix and \( \hat{H} \) is of the form

\[
\hat{H} = \begin{bmatrix}
0 & \cdots & 0 & 1/\langle b, w_N \rangle \\
1 & \cdots & 0 & -\langle b, w_1 \rangle / \langle b, w_N \rangle \\
\vdots & \ddots & \vdots & \ddots \\
0 & \cdots & 1 & -\langle b, w_{N-1} \rangle / \langle b, w_N \rangle
\end{bmatrix}.
\]
Proof. The main ingredient is to generalize Theorem [2.1] in the following way (see [Greenbaum et al., 1996, Theorem 3.1.]).

Let \( E_1 \subset E_2 \subset \ldots \subset E_N \) be a nested sequence of subspaces of \( \mathbb{C}^N \), where \( \dim(E_k)=k \) for any \( k = 1, \ldots, N \). Take \( Q = \{q_1, \ldots, q_N\} \) an orthonormal basis of \( E_N \) such that for any \( k \)

\[
\text{span}\{q_1, \ldots, q_k\} = E_k
\]

and take an \( N \times N \) unitary matrix \( Q \) with the columns formed by the vectors \( q_1, \ldots, q_N \). Take an \( N \times N \) nonsingular upper triangular matrix \( R \) and let \( \hat{H} \) be the \( N \times N \) matrix from (2.6) with \( r_0 = b \), i.e.,

\[
\hat{H} = \begin{bmatrix}
0 & \cdots & 0 & 1/\langle b, q_N \rangle \\
1 & \cdots & 0 & -\langle b, q_1 \rangle / \langle b, q_N \rangle \\
\vdots & \ddots & \vdots & \vdots \\
0 & \cdots & 1 & -\langle b, q_{N-1} \rangle / \langle b, q_N \rangle 
\end{bmatrix}.
\]

Then

\[
A K_k(A, b) = E_k, \quad k = 1, \ldots, N \quad (2.17)
\]

if and only if \( \langle b, q_N \rangle \neq 0 \) and at the same time one has

\[
Q^* A Q = R \hat{H}. \quad (2.18)
\]

In other words, the considered nested sequence of subspaces \( E_1 \subset \ldots \subset E_N \) coincides with the Krylov residual subspaces \( AK_1(A, b) \subset \ldots \subset AK_N(A, b) \) at each step if and only if the matrix \( A \) expressed in the base \( Q \) is of the form (2.17) and the vector \( b \) is not contained in the first \( N-1 \) subspaces, i.e., \( \langle b, q_N \rangle \neq 0 \).

Note that this observation can be proved in the same way as Theorem [2.1]. Indeed, analogously to (2.8), condition (2.16) is equivalent to \( \langle b, q_N \rangle \neq 0 \) and the existence of the \( N \times N \) nonsingular upper triangular matrix \( R_1 \) such that

\[
[Ab, \ldots, A^N b] \quad R_1 = Q,
\]

which is equivalent to the existence of the \( N \times N \) nonsingular upper triangular matrix \( R_2 \) such that

\[
A [b, q_1, \ldots, q_{N-1}] \quad R_1 R_2 = Q.
\]

Denoting \( \hat{R} = R_1 R_2 \) and using the identity (2.10), i.e.,

\[
[b, q_1, \ldots, q_{N-1}] = Q \hat{H}^{-1},
\]

one can conclude that

\[
Q^* A Q = R \hat{H},
\]

with \( R = \hat{R}^{-1} \).

It is also useful to realize that for arbitrary matrix \( C \) and any unitary matrix \( Q \), the convergence curves \( \text{GMRES}(QCQ^*, b) \) and \( \text{GMRES}(C, Q^* b) \) coincide. This can be easily seen from Proposition [2.1] via the polynomial point of view. Indeed, for any polynomial \( p(z) \) one has
\[ p(QCQ^*)b = Qp(C)Q^*b, \]

which verifies the claim since \( Q \) is unitary.

Recalling \((2.14)\), GMRES\((A, b)\) is equal to \( \{f(0), \ldots, f(N - 1)\} \) if and only if

\[ \Gamma W^*b = (g(0), \ldots, g(N - 1))^T, \]

where \( g(i) := \sqrt{f(i)^2 - f(i + 1)^2} \) for \( i = 0, 1, \ldots, N - 2 \) and \( g(N - 1) = f(N - 1) \), the matrix \( \Gamma = \text{diag}(\gamma_1, \ldots, \gamma_N) \) is the uniquely determined unitary diagonal matrix such that \( \gamma_k \langle b, w_k \rangle = g(k - 1) \) for any \( k = 1, \ldots, N \) and \( W \) is the unitary matrix, whose columns are taken from the Arnoldi process applied to \( A \) and \( b \); see \((2.3)\) with \( r_0 = b \). Employing the observation \((2.16, 2.17)\) from the beginning of this proof yields the statement. \( \square \)

The above theorem nicely connects the task of constructing a problem with a specific convergence behaviour and the task of finding matrices that behave the same with the GMRES method. A characterization similar to the one suggested above Theorem \((2.2)\), i.e., a characterization that would reveal what is the structure of the problems with a fixed convergence behaviour (with GMRES) and with fixed nonzero eigenvalues of the system matrix, is given bellow.

**Theorem 2.3** (Arioli et al., 1998, Theorem 2.1.) Consider the problem \( Ax = b \) (see \((1.1)\)) and \( x_0 = e \). Take a non-increasing sequence in form of \((2.13)\), i.e., \( f(0) \geq f(1) \geq \ldots \geq f(N - 1) > 0 \) and \( N \) nonzero complex numbers \( \lambda_1, \ldots, \lambda_N \). Then the following assertions are equivalent

1. \( \text{sp}(A) = \{\lambda_1, \ldots, \lambda_N\} \) and at the same time the GMRES residual vectors \( r_k \) satisfy \( ||r_k|| = f(k) \) for all \( k = 0, 1, \ldots, N - 1 \);

2. The system matrix \( A \) is of the form

\[ A = WRCR^{-1}W^* \quad \text{and} \quad b = W^*h, \]

with the notation and further specification

- \( h = (g(0), \ldots, g(N - 1))^T \) and \( g(i) := \sqrt{f(i)^2 - f(i + 1)^2} \); see \((2.14)\);

- \( W \) is a unitary matrix;

- \( C \) is the companion matrix of the monic polynomial \( q(z) \) with roots \( \lambda_1, \ldots, \lambda_N \); see \((2.11)\);

- consider the polynomial with roots \( \lambda_1, \ldots, \lambda_N \) and the constant term equal to one, i.e., the polynomial

\[ p(z) = \left(1 - \frac{z}{\lambda_1}\right) \left(1 - \frac{z}{\lambda_2}\right) \ldots \left(1 - \frac{z}{\lambda_N}\right). \]

Further consider the vector \( s = (\xi_1, \ldots, \xi_N)^T \) such that \( \xi_k \) is equal to the coefficient of \( z^k \) in \( p(z) \). Then \( R \) is any nonsingular, upper triangular \( N \times N \) matrix such that \( Rs = h \).
Proof. (1) ⇒ (2) Consider the matrix \( B = [Ab | \cdots | A^N b] \). It is nonsingular thanks to Assumption 2.1; see Proposition 1.2. It will be useful to prepare the following matrix equalities.

Denoting \( q(z) = z^N - q_{N-1}(z) \), the Cayley-Hamilton Theorem gives \( A^N = q_{N-1}(A) \). References concerning this theorem can be found, e.g., in [Liesen and Strakoš, 2015, Section 4.2.3]. Consequently

\[
AB = [A^2 b | \cdots | A^N b | A q_{N-1}(A) b] = BC. \tag{2.19}
\]

Writing \( p(z) = p_{N-1} + 1 \), the identity matrix \( I \) can be expressed from the Cayley-Hamilton Theorem as \( I = p_N(A) \). Hence

\[
b = \sum_{i=1}^{N} \xi_i A^i b = Bs. \tag{2.20}
\]

Performing the QR-decomposition of the matrix \( B \) yields matrices \( \tilde{W}, \tilde{R} \) such that \( \tilde{R} \) is nonsingular upper triangular, \( \tilde{W} \) is unitary and \( B = \tilde{W} \tilde{R} \). Surely \( A K_k(A, b) = \text{span} \{ \tilde{w}_1, \ldots, \tilde{w}_k \} \) for any \( k = 1, 2, \ldots, N \). Similarly as in (??), one has that \( g(i - 1) = |\langle \tilde{w}_i, b \rangle| \), which implies that there are complex scalars \( \gamma_1, \ldots, \gamma_N \) such that \( |\gamma_i| = 1 \) for all \( i \) and

\[
b = \sum_{i=1}^{N} (\tilde{w}_i, b) \tilde{w}_i = \sum_{i=1}^{N} \gamma_i g(i - 1) \tilde{w}_i.
\]

This, combined with (2.19), yields

\[
Bs = b = \tilde{W} \Gamma \mathbf{h},
\]

with \( \Gamma = \text{diag}(\gamma_1, \ldots, \gamma_N) \). Taking \( W = \tilde{W} \Gamma \) and \( R = \Gamma^* \tilde{R} \), one can verify the conditions (2) as

\[
AWR = A \tilde{W} \tilde{R} = AB = BC = WRC
\]

\[
Wh = b = Bs = WRs, \quad \text{i.e., } Rs = \mathbf{h}.
\]

(2) ⇒ (1) Since \( A \) is similar to \( C \), one has \( \text{sp}(C) = \{\lambda_1, \ldots, \lambda_N\} \), which implies that the spectral condition is clearly met. Therefore, it is sufficient to show

\[
\text{span} \{ w_1, \ldots, w_k \} = A K_k(A, b), \quad k = 1, 2, \ldots, N, \tag{2.21}
\]

where the vectors \( w_1, w_2, \ldots, w_N \) are the first \( k \) columns of \( W \). Indeed, if one has the above equality, then Theorem 2.2 yields the rest.

The proof of (2.20) will be done by induction along \( k \). Denote \( b_i \) as the \( i - \text{th} \) column of matrix \( B \), i.e., \( b_i = A^i b \). First, focus on proving \( \text{span} \{ w_1 \} = \text{span} \{ b_1 \} \).

Using the assumption (2), one may write

\[
b_1 = Ab = W R C R^{-1} W^* W R s = W R C s = W R e_1,
\]

19
which yields the result. The last equality can be justified for example as

\[
CS = \begin{bmatrix}
0 & 0 & \cdots & 0 & \alpha_0 \\
1 & 0 & \cdots & 0 & \alpha_1 \\
0 & 1 & \cdots & 0 & \alpha_2 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & 1 & \alpha_{N-1}
\end{bmatrix}
\begin{bmatrix}
\xi_1 \\
\xi_2 \\
\xi_3 \\
\vdots \\
\xi_N
\end{bmatrix} = \begin{bmatrix}
\xi_N \alpha_0 \\
\xi_1 + \alpha_1 \xi_N \\
\xi_2 + \alpha_2 \xi_N \\
\vdots \\
\xi_{N-1} + \alpha_{N-1} \xi_N
\end{bmatrix} = e_1,
\]

since

\[
\alpha_i = \sum_{k_1, \ldots, k_i \leq N} \left( \prod_{j \in \{k_1, \ldots, k_i\}} \lambda_j \right), \quad \xi_i = \sum_{k_1, \ldots, k_i \leq N} \left( \prod_{j \in \{k_1, \ldots, k_i\}} \frac{-1}{\lambda_j} \right).
\]

Taking any \( k < N \), the induction assumption ensures that \( b_{k-1} = WRe_{k-1} \) and also that \( \text{span}\{w_1, \ldots, w_{k-1}\} = AK_{k-1}(A, b) \). Using the first part, one can write

\[
b_k = Ab_{k-1} = WRCR^{-1}W^*b_{k-1} = WRCe_{k-1} = WRe_k,
\]

which yields the proof thanks to the second part of the induction assumption. \( \Box \)

The last theorem does not yet yield the desired characterization but it is already very close. The only thing that is left to do is to reformulate the assumption \( Rs = h \) into the matrix structure. This will be done later. First, let us step aside and collect the results shown so far. For that purpose, we will adopt the notation from Arioli et al. [1998] as follows.

Having any non-increasing sequence \( f = \{f(0) \geq f(1) \geq \ldots \geq f(N-1) > 0\} \) and set of \( N \) nonzero complex numbers \( \Lambda = \{\lambda_1, \ldots, \lambda_N\} \), denote \( \mathcal{S}_1(f, \Lambda) \) as the set of all pairs \( A, b \) described by Theorem 2.3, i.e., the set of pairs for which GMRES produces the residual vectors \( r_k \) with norms \( ||r_k|| = f(k), k = 1, 2, \ldots, N \) and the spectrum of \( A \) coincides with the set \( \Lambda \).

The set of all pairs \( A, b \) for which GMRES produces the residual vectors \( r_k \) with norms \( ||r_k|| = f(k), k = 1, 2, \ldots, N_1 \), i.e., the set of pairs described in Theorem 2.2 will be denoted by \( \mathcal{S}_2(f) \).

The immediate observation is that \( \mathcal{S}_1(f, \Lambda) \subset \mathcal{S}_2(f) \). However, the interesting question arises, whether there is an easy link between those two sets, i.e., how the characterization of the form of Theorem 2.2 can be linked to the characterization of the form of Theorem 2.3 for a given \( \Lambda \) and vice versa.

**Theorem 2.3 \( \rightarrow \) Theorem 2.2** Consider a non-increasing sequence \( f = \{f(0) \geq f(1) \geq \ldots \geq f(N-1) > 0\} \) and set of \( N \) nonzero complex numbers \( \Lambda = \{\lambda_1, \ldots, \lambda_N\} \). Take a pair \( A, b \in \mathcal{S}_1(f, \Lambda) \), i.e., \( sp(A) = \Lambda \) and GMRES\((A, b)\) corresponds to \( f \). Take vectors \( s, h \) and matrices \( W, R, C \) corresponding to Theorem 2.3, i.e.,

- \( h \) is the \( N \) dimensional vector formed by the square root of the differences of the consecutive residual norms of GMRES;
- \( s \) is the \( N \) dimensional vector formed by the coefficients of the polynomial with the constant term equal to one and with the roots corresponding to the spectrum \( \Lambda \);
• $W$ is a unitary matrix such that $W^*b = h$;
• $R$ is an upper triangular nonsingular matrix such that $Rs = h$;
• $C$ is the companion matrix of the monic polynomial with roots $\Lambda$

Defining $Y = RC^{-1}$, it follows

$$Y = R \begin{bmatrix} -\alpha_1 & [I_{N-1}] \\ \vdots & \vdots \\ -\alpha_{N-1} & [I_{N-1}] \\ \frac{1}{\alpha_0} & 0 & \ldots & 0 \end{bmatrix} = R \begin{bmatrix} \xi_1 & [I_{N-1}] \\ \vdots & \vdots \\ \xi_{N-1} & 0 & \ldots & 0 \end{bmatrix} = 
$$

$$h \begin{bmatrix} R_{N-1} \\ 0 & \ldots & 0 \end{bmatrix} = h \begin{bmatrix} I_{N-1} \\ 0 & \ldots & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & R_{N-1} \end{bmatrix},$$

where $I_{N-1}$ is the identity matrix of the dimension $N - 1$ and $R_{N-1}$ stands for the $(N - 1)$-th left principal submatrix of the matrix $R$ (hence it is also a nonsingular upper triangular matrix but of order $N - 1$). With this notation, one can simply write the following link between the considered characterizations from Theorem 2.3 and Theorem 2.2:

$$RCR^{-1} = R (RC^{-1})^{-1} = \tilde{R} \tilde{H},$$

where $\tilde{H}$ is of the form (2.6) from Theorem 2.2 and Theorem 2.1 respectively and $\tilde{R}$ is given as

$$\tilde{R} = R \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & [R_{N-1}]^{-1} \end{bmatrix},$$

which corresponds to Theorem 2.2.

**Theorem 2.2 → Theorem 2.3** Consider a pair $A, b \in \mathcal{S}_2(f)$ for a fixed $f$, i.e., a pair which for which GMRES$(A, b) = f$. Denote $\Lambda = \text{sp}(A)$ and $C$ the companion matrix of the monic polynomial with roots corresponding to $\Lambda$. Then there exists a unique nonsingular upper triangular matrix $R_A$, which fits the second assumption of the Theorem 2.3 for the matrix $A$; see [Arioli et al., 1998, Section 2]. This gives the decomposition of the form of Theorem 2.3.
Recall the task below Theorem 2.3, i.e., fixing the system matrix (nonzero) eigenvalues, can one parametrize the set of all problems that produce a given non-increasing convergence curve with the GMRES method? The above insight into the structures of the set \( S(f, \Lambda) \) and gives a big hint to the solution.

Consider a pair \( \{A, b\} \in S(\Lambda, \Lambda) \) for a fixed \( f \) and \( \Lambda \). Take any nonsingular upper triangular matrix \( R \) from Theorem 2.3, i.e., such that \( Rs = h \) where \( h \) is the \( N \)-dimensional vector formed by the square root of the differences of the consecutive residual norms of GMRES and \( s \) is the \( N \)-dimensional vector formed by the coefficients of the polynomial with the constant term equal to one and with the roots corresponding to the spectrum \( \Lambda \). Denoting the last column of \( R \) by \( R_N \), the condition \( Rs = h \) gives

\[
\xi_N R_N = h - \begin{pmatrix}
R_{N-1} (\xi_1, \ldots, \xi_{N-1})^T \\
0
\end{pmatrix}.
\]

In other words, condition \( Rs = h \) fixes exactly \( N \) parameters and hence the rest of non-zero entries of \( R \), i.e., \( R_{N-1} \), are free parameters. Taking \( Y = RC^{-1} \) (see (2.21)) one gets the following theorem, which gives the desired parametrization.

**Theorem 2.4.** Consider the problem \( Ax = b; \) see (1.1); and a non-increasing sequence \( f(0) \geq f(1) \geq \ldots \geq f(N-1) > 0 \) and \( N \) nonzero complex numbers \( \{\lambda_1, \ldots, \lambda_N\} \). Then the following assertions are equivalent:

1. \( \text{sp}(A) = \{\lambda_1, \ldots, \lambda_N\} \) and at the same time the GMRES residual vectors \( r_k \) satisfy \( ||r_k|| = f(k) \) for all \( k = 0, 1, \ldots, N-1; \)

2. The matrix \( A \) is of the form

\[
A = WYCY^{-1}W^* \quad \text{and} \quad b = W^*h,
\]

with notation:

- \( h = (g(0), \ldots, g(N-1))^T \) and \( g(i) := \sqrt{f(i)^2 - f(i+1)^2} \); see (2.14);
- \( W \) is any unitary matrix;
- \( C \) is the companion matrix of the monic polynomial \( q(z) \) with roots \( \lambda_1, \ldots, \lambda_N \); see (2.11);

\[
Y = \begin{bmatrix}
\overrightarrow{h} & [R_{N-1}] \\
0 & \ldots & 0
\end{bmatrix}, \text{ where } R_{N-1} \text{ is any } (N-1) \times (N-1) 
\]

noningular upper triangular matrix.

This theorem encloses the second chapter and completes the pursuit of the second half of this chapter. It quantifies the freedom left in the construction of the linear algebraic problems \( Ax = b \) that have the given convergence behaviour with GMRES and the matrix \( A \) has any desired nonzero eigenvalues. This result was further extended in, e.g., Tebbens and Meurant [2012].
3. Behaviour of algebraic methods and operators in infinite-dimensional Hilbert or Banach spaces

In this chapter, we will approach the study of the convergence behaviour of the Krylov subspace methods applied to the problem \( Ax = b \); see (1.1); from a different perspective. Under a Banach space we will understand a complete normed vector space (complex or real) and under a Hilbert space we will understand a complete normed vector space (complex or real) with the norm induced by some inner product.

One can concentrate on the infinite-dimensional problem behind the linear algebraic system \( Ax = b \) (see Figure ??) and then study the applicability of the results for the finite-dimensional case. Questions related to the Krylov subspace methods on infinite-dimensional Hilbert or Banach spaces have been investigated by many authors; see Karush [1952] or Hayes [1954]. It has been proved, e.g., that the convergence rate of the CG method on an infinite-dimensional Hilbert space for an identity + compact operator must be \textit{superlinear}, i.e., the ratio of the norms of the \((k + 1)\)-th and \(k\)-th errors tends to zero as \( k \) goes to infinity (see Hayes [1954, Chapter 3, Section 13, Theorems 13.1, 13.2 and 13.3] and also Gautschi [2011, Section 4.2, Definition 4.2.2]). Applications of the infinite-dimensional considerations to the finite-dimensional problems have, however, also drawbacks and limitations. This chapter is devoted to examining one widely referenced approach that seems to be based on a misunderstanding.

Consider a Banach space \( X \). Take a linear bounded operator \( L \) on \( X \) and a vector \( f \in X \) and consider the linear problem

\[
Lu = f. \tag{3.1}
\]

Recalling the first chapter, the solution process of (3.1) usually first performs a discretization, followed by solving the discretized problem. In other words, a suitable discretization process approximate (3.1) with a problem

\[
L_h u_h = f_h, \tag{3.2}
\]

where \( f_h \in L_h(X) \) and \( L_h \) is \textit{finite-dimensional}, i.e., \( L_h \) is linear, bounded and \( \dim(L_h(X)) = N < \infty \). The problem (3.2) is then approached with, e.g., some Krylov subspace method as discussed in the first chapter.

If one wants to approximatively solve (3.1) in this way, it is very important to understand the relation of \( L \) and \( L_h \). As will be shown later, the fact that \( L_h \) is finite-dimensional naturally limits the sense, in which \( L_h \) approximates \( L \). In order to further investigate this topic, it is necessary to recall some definitions and facts from the functional analysis. Though most of the statements in this chapter are well-known, the references are always stated explicitly.

Consider a Banach space \( X \). The set of all linear bounded operators will be denoted by \( \mathcal{L}(X) \) and \( I \) will stand for the identity operator on \( X \). The set \( \mathcal{L}(X) \)
is equipped with the supremum norm induced by the norm of $X$, i.e., for any $L \in \mathcal{L}(X)$,
\[
||L|| = \sup_{x \in \mathbb{B}} ||Lx||,
\]
with $\mathbb{B}$ being the closed unit ball in $X$.

**Remark 3.1.** For the sake of simplicity, we choose to work with the induced norm (3.3) on $\mathcal{L}(X)$. Most of the statements, however, hold true also for an arbitrary norm.

Consider a sequence of linear bounded operators $\{L_m\}_{m=1}^{\infty}$ and a linear bounded operator $L$. One says that the operator $L$ is the **uniform limit of** $\{L_m\}_{m=1}^{\infty}$, provided
\[
\lim_{m \to \infty} ||L - L_m|| = 0.
\]
If for any $x \in X$ one has that
\[
\lim_{m \to \infty} ||Lx - L_m x|| = 0,
\]
then one says that $L$ is the **strong or pointwise limit of** $\{L_m\}_{m=1}^{\infty}$.

Consider a Banach space $X$ and a nonempty set $S \subset X$. $S$ is called **totally bounded**, provided for any $\delta > 0$ there are points $x_1, \ldots, x_l$ in $X$ (finitely many) such that the union of the open $\delta$-balls with the centres at the points $x_1, \ldots, x_l$ covers the subset. If $S$ is totally bounded and closed it is called **compact**.

A linear bounded operator is called **invertible**, provided its inverse is also linear and bounded. A linear bounded operator on a Banach space is called **compact**, provided it maps any bounded set onto a totally bounded set. The set of all compact operators on $X$ will be denoted by $\mathcal{K}(X)$.

The set of all linear bounded functionals from $X$ to $\mathbb{C}$ (or $\mathbb{R}$, if $X$ is considered over $\mathbb{R}$), i.e., the dual space of $X$, is denoted by $X^*$. Consider a sequence $\{x_k\}_{k=1}^{\infty} \subset X$ and a vector $x \in X$. One says that $\{x_k\}_{k=1}^{\infty}$ **converges weakly to** $x$, provided for any $\Lambda \in X^*$ one has
\[
\Lambda(x_k) \to \Lambda(x), \quad \text{as } k \to \infty.
\]

**Remark 3.2** ([Rudin 1973, Introduction, Paragraph 1.9 (c)]). In many works, the term of a compact set is introduced more generally. A possible equivalent definition goes as follows.

If from any open cover of the set one can choose only a finite subcover, then one says that the set is compact. In any complete metric space, this definition is equivalent to the one above.

**Remark 3.3.** In literature, one may encounter substantially different equivalent definitions of a compact operator. For example, in [Friedman 2010, Chapter 5, Theorem 5.1.1], a linear bounded operator is called **compact** (or completely continuous), provided the image of any weakly converging sequence is a strongly converging sequence.

The first proposition shows the relation of the finite-dimensional and the compact operators on a Banach space.
Proposition 3.1 ([DeVito, 1990, Section 3.1, Exercise 1.3]). Any finite-dimensional operator on a Banach space is compact.

Proof. Consider a finite-dimensional operator $T$ on a Banach space $X$ and a bounded set $B \subset X$. $T$ is bounded and hence $T(B)$ is bounded as well. Since $\dim(T(X)) < \infty$, the compact sets in $T(X)$ coincide with sets that are bounded and closed. Hence $T(B)$ is a compact set, which yields the statement.

The following proposition gives a basic result regarding the set of compact operators on Banach spaces.

Proposition 3.2 ([Friedman, 2010, Section 5, Theorem 5.1.3]). Consider a Banach space $X$. Any uniform limit of compact operators is also a compact operator, i.e., $\mathcal{K}(X)$ is a closed subset of $\mathcal{L}(X)$.

Proof. Consider a sequence of compact linear bounded operators $\{T_n\}_{n=1}^{\infty}$ that uniformly converges to a linear bounded operator $T$. Take any bounded set $B \subset X$. Consider $\varepsilon > 0$ and find $n \in \mathbb{N}$ such that $||T - T_n|| < \varepsilon/3$. Since $T_n$ is compact the set $T_n(B)$ is totally bounded. Therefore, there are points $Tx_1, \ldots, Tx_l \in B$ such that for any $Tx \in T(B)$ one of the points $Tx_1, \ldots, Tx_l$ is closer to $Tx$ than $\varepsilon/3$. Take any $Tx \in T(B)$ and denote this point $Tx_i$. Considering the inequality

$$||Tx - Tx_i|| \leq ||Tx - T_nx|| + ||T_nx - T_nx_i|| + ||T_nx_i - Tx_i||,$$

implies

$$||Tx - Tx_i|| < \varepsilon.$$

In other words $T(B)$ is totally bounded and the statement follows.

Combining this result with Proposition 3.1 implies that any uniform limit of finite-dimensional operators on a Banach space must be a compact operator. If one considers a Hilbert space instead of a Banach space, the result of Proposition 3.2 can be extended in the following way.

Remark 3.4 ([DeVito, 1990, Section 5.5, Theorem 2]). Consider a Hilbert space. Any compact operator can be written as a uniform limit of finite-dimensional operators.

In other words, the set of all compact operators on a Hilbert space is the closure of the set of all finite-dimensional operators on that Hilbert space.

But even Proposition 3.2 implies that a general linear bounded operator on a Banach space cannot be approximated in norm by finite-dimensional operators to an arbitrary accuracy. Therefore, if the operator $L$ in (3.1) is not compact, then it makes no sense to consider the approximating operators $L_h$ in the uniform sense. The next proposition shows that if $L$ is invertible, i.e., (3.1) has a unique solution for any right-hand side vector $f$, then $L$ is not compact.

Proposition 3.3. Consider an infinite-dimensional Banach space $X$ and $L \in \mathcal{L}(X)$. If $L$ is invertible, then $L$ is not compact.
Proof. Recalling the fact that the closed unit ball in $X$ is compact if and only if $X$ is finite-dimensional, one has that the identity operator is not compact on an infinite-dimensional Banach space. It will be also useful to realize that if $L$ is a linear bounded operator and $T$ is a compact operator, then $TL$ is also a compact operator. Indeed, if one takes a bounded set $B \subset X$, then $L(B)$ is also bounded and hence $T(L(B))$ is totally bounded, which gives the observation.

Considering
\[ I = LL^{-1} \]
and the facts that $L^{-1}$ is bounded and $I$ is not compact implies that $L$ cannot be compact, according to the observation. \qed

Summarizing, if $L$ is invertible, i.e., for any right-hand side vector $f$ there is a unique solution to the problem (3.1), then it makes no sense to consider arbitrary close uniform approximations of $L$ by finite-dimensional operators $L_h$. Instead, one can consider the approximations in the strong (pointwise) sense, as shown in the following proposition.

**Proposition 3.4** ([Vorobyev, 1965, Section 4, Theorem II]). Let $L$ be a linear bounded operator on a Hilbert space. Then there exists a sequence of finite-dimensional operators converging to $L$ in a strong (pointwise) sense.

Moreover, approximating the operator in a strong sense is adequate to the problem (3.1), since one intentionally aims at approximating the solution not the operator.

In [Nevanlinna, 1993, Section 1.8], the author considers Krylov subspace methods and aims at analysing their convergence behaviour. In Krylov subspace methods, the residual (error) at the $k$-th iteration can be expressed as the $k$-th degree polynomial in the operator or matrix, applied to the initial residual (error); see Remark 2.1. Therefore, Krylov subspace methods are also called polynomial methods; see, e.g., Fischer [1996].

Considering
\[ \|p_k(A)r_0\| \leq \|p_k(A)\| \cdot |r_0|, \]
a (considerably) simplified analysis of the convergence behaviour can focus on the norm of the polynomial $\|p_k(A)\|$ instead of $\|p_k(A)r_0\|$.

In [Nevanlinna, 1993, Section 1.8], it is argued that this simplified analysis can be based on the infinite-dimensional operator $L$ from (3.1) with implications to the description of the behaviour of the Krylov subspace methods applied to the discretized finite-dimensional problem (3.2). The argument requires that $\|p_k(L) - p_k(L_h)\|$ is small, whenever $\|L - L_h\|$ is small. In particular, this requires
\[ \|L - L_{h}\| = O(h), \quad \text{as } h \to 0, \] (3.4)
which means that there is a constant $K > 0$ such that $\|L - L_h\| \leq Kh$ as $h \to 0$; see [Nevanlinna, 1993, Section 1.8, (1.8.1)]. However, this implies that the infinite-dimensional operator $L$ must be compact (see Propositions 3.1 and 3.2) and hence non-invertible (see Proposition 3.3). This shows that the assumption (3.4) makes the results based on it restricted onto the class of compact operators.
Conclusion

The iterative projection methods, widely used approach to a linear algebraic problem $Ax = b$, are introduced in the first chapter. The basic principles of the general framework of the projection methods are briefly explained. Some basic Krylov subspace methods (CG, SYMMLQ, MINRES and GMRES) are introduced and set into the framework of the projection methods. In the second chapter, we further focus on the GMRES method. In particular, the triplet of the articles Greenbaum and Strakoš [1994], Greenbaum et al. [1996] and Arioli et al. [1998] is summarized. The focus of these articles is on the convergence behaviour of the GMRES method and its connection with the spectrum of the system matrix. In particular, the matrices that generate the same Krylov residual spaces with a fixed starting vector are characterized. This characterization is then used to show that any convergence behaviour with GMRES is possible with the system matrix having any nonzero eigenvalues. In the end, it is given a characterization of the problems $Ax = b$ such that $A$ has any nonzero desired eigenvalues and at the same time the GMRES method applied to the problem $Ax = b$ with zero initial guess produce a given non-increasing convergence curve.

In the last chapter it is considered a different approach to the study of the convergence behaviour of the Krylov subspace methods. Often there is an infinite-dimensional problem behind the linear algebraic problem $Ax = b$. Concentrating on the infinite-dimensional problem behind $Ax = b$, one can aim at analysing the behaviour of the methods using the infinite-dimensional problem and then apply the results on the linear algebraic finite-dimensional problem. However, this approach of analysis has also its limitations. One particular example of misunderstanding using this approach is shown at the end of the last chapter.
Bibliography


