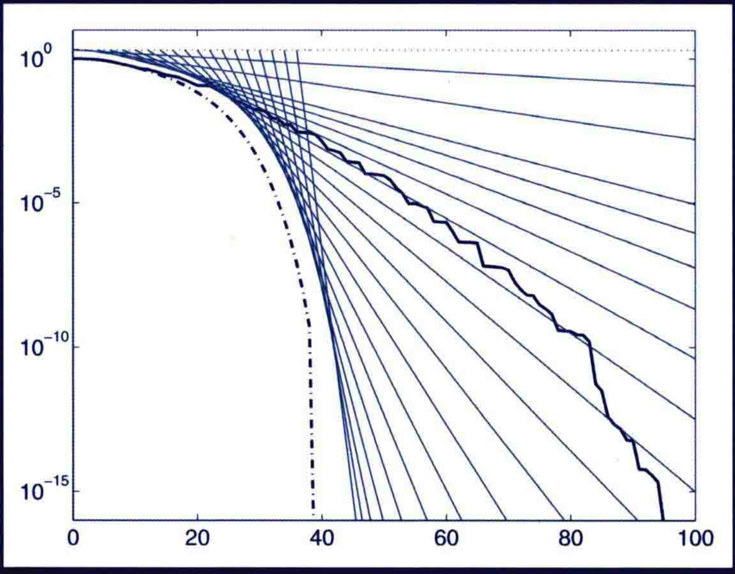
NUMERICAL MATHEMATICS
AND SCIENTIFIC COMPUTATION

Krylov Subspace Methods

Principles and Analysis

JÖRG LIESEN
ZDENĚK STRAKOŠ



OXFORD SCIENCE PUBLICATIONS

Krylov Subspace Methods

Principles and Analysis

JÖRG LIESEN

Technical University of Berlin

ZDENĚK STRAKOŠ

Charles University in Prague

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

First Edition published in 2013

Impression: 1

British Library Cataloguing in Publication Data

Data available

Library of Congress Cataloging in Publication Data

Library of Congress Control Number: 2012945094

ISBN 978-0-19-965541-0

Printed and bound by
CPI Group (UK) Ltd, Croydon, CR0 4YY

NUMERICAL MATHEMATICS AND SCIENTIFIC COMPUTATION

Series Editors

A. M. STUART   E. SÜLI

NUMERICAL MATHEMATICS AND SCIENTIFIC COMPUTATION

Books in the series

Monographs marked with an asterisk (*) appeared in the series 'Monographs in Numerical Analysis' which is continued by the current series.

For a full list of titles please visit
http://www.oup.co.uk/academic/science/maths/series/nmsc

* J. H. Wilkinson: *The Algebraic Eigenvalue Problem*
* I. Duff, A. Erisman, and J. Reid: *Direct Methods for Sparse Matrices*
* M. J. Baines: *Moving Finite Elements*
* J. D. Pryce: *Numerical Solution of Sturm–Liouville Problems*

C. Schwab: *p- and hp- Finite Element Methods: Theory and Applications in Solid and Fluid Mechanics*
J. W. Jerome: *Modelling and Computation for Applications in Mathematics, Science, and Engineering*
A. Quarteroni and A. Valli: *Domain Decomposition Methods for Partial Differential Equations*
G. Em Karniadakis and S. J. Sherwin: *Spectral/hp Element Methods for CFD*
I. Babuška and T. Strouboulis: *The Finite Element Method and its Reliability*
B. Mohammadi and O. Pironneau: *Applied Shape Optimization for Fluids*
S. Succi: *The Lattice Boltzmann Equation for Fluid Dynamics and Beyond*
P. Monk: *Finite Element Methods for Maxwell's Equations*
A. Bellen and M. Zennaro: *Numerical Methods for Delay Differential Equations*
J. Modersitzki: *Numerical Methods for Image Registration*
M. Feistauer, J. Felcman, and I. Straškraba: *Mathematical and Computational Methods for Compressible Flow*
W. Gautschi: *Orthogonal Polynomials: Computation and Approximation*
M. K. Ng: *Iterative Methods for Toeplitz Systems*
M. Metcalf, J. Reid, and M. Cohen: *Fortran 95/2003 Explained*
G. Em Karniadakis and S. Sherwin: *Spectral/hp Element Methods for Computational Fluid Dynamics, Second Edition*
D. A. Bini, G. Latouche, and B. Meini: *Numerical Methods for Structured Markov Chains*
H. Elman, D. Silvester, and A. Wathen: *Finite Elements and Fast Iterative Solvers: with Applications in Incompressible Fluid Dynamics*
M. Chu and G. Golub: *Inverse Eigenvalue Problems: Theory, Algorithms, and Applications*
J.-F. Gerbeau, C. Le Bris, and T. Lelièvre: *Mathematical Methods for the Magnetohydrodynamics of Liquid Metals*
G. Allaire and A. Craig: *Numerical Analysis and Optimization: An Introduction to Mathematical Modelling and Numerical Simulation*
K. Urban: *Wavelet Methods for Elliptic Partial Differential Equations*
B. Mohammadi and O. Pironneau: *Applied Shape Optimization for Fluids, Second Edition*
K. Boehmer: *Numerical Methods for Nonlinear Elliptic Differential Equations: A Synopsis*
M. Metcalf, J. Reid, and M. Cohen: *Modern Fortran Explained*
J. Liesen and Z. Strakoš: *Krylov Subspace Methods: Principles and Analysis*

To Anja and to Iva

This seems to be a very elementary problem without
deeper meaning. However, one meets this task again
and again in the electric industry and in all kinds
of oscillation problems.
A short while ago, I found a rather elegant solution.
The reason why I am strongly drawn to such
approximation mathematics problems is not
the practical applicability of the solution,
but rather the fact that a very "economical"
solution is possible only when it is very "adequate".
To obtain a solution in very few steps
means nearly always that one has found a way
that does justice to the inner nature of the problem.

CORNELIUS LANCZOS in a letter to Albert Einstein on March 9, 1947.

Your remark on the importance of
adapted approximation methods makes very
good sense to me, and I am convinced
that this is a fruitful mathematical aspect,
and not just a utilitarian one.

EINSTEIN'S reply to Lanczos on March 18, 1947[1].

1. We thank the Albert Einstein Archives, The Hebrew University of Jerusalem, Israel, which holds the copyrights on Einstein's letters, for giving us access to the correspondence between Lanczos and Einstein. The two corresponded in German, and the above are our translations of the following original quotes:

"Das scheint ja ein reichlich elementares Problem ohne tiefere Bedeutung zu sein. Doch stösst man in der elektrischen Industrie und bei allen möglichen Schwingungsproblemen immer wieder auf diese Aufgabe. Vor kurzem ist mir eine recht elegante Lösung geglückt. Der Grund weshalb mich solche Probleme der Approximationsmathik [spelling error by Lanczos; most likely he meant to write "Approximationsmathematik"] stark anziehen, ist nicht die praktische Verwendbarkeit der Lösung, sondern vielmehr der Umstand, dass eine stark "ökonomische" Lösung nur möglich ist, wenn sie auch stark "adequat" ist. Mit wenig Schritten zum Resultat zu kommen bedeutet fast immer, dass man einen Weg gefunden hat, der dem inneren Wesen des Problems gerecht wird." (Lanczos to Einstein, Albert Einstein Archives document number 15-313.)

"Ihre Bemerkung über die Bedeutung angepasster Approximations-Methoden leuchtet mir sehr ein, und ich bin überzeugt, dass dies ein fruchtbarer mathematischer Gesichtspunkt ist, nicht nur ein utilitaristischer." (Einstein to Lanczos, Albert Einstein Archives document number 15-315.)

PREFACE

Quite frequently in life, the most elegant solution to a problem also turns out to be the most efficient one for practical purposes. This heuristical observation certainly applies to Krylov subspace methods. The algorithms devised by Magnus Hestenes, Eduard Stiefel, Cornelius Lanczos, and others in the early 1950s for iteratively solving large and sparse linear algebraic systems and eigenvalue problems can hardly be more elegant and aesthetically pleasing. Yet, these algorithms and their numerous later variants and extensions are nowadays used widely and successfully throughout science and engineering. Because of their overwhelming success in applications, Krylov subspace methods are counted among the 'Top 10 Algorithms' of the 20th century [116, 140].

Not surprisingly, several first-rate books describing Krylov subspace methods are available. They have been written by excellent communicators and leading researchers in the field, including Bernd Fischer [183] (republished by SIAM as a 'Classics in Applied Mathematics' title in 2011), Anne Greenbaum [272], Gérard Meurant [452, 453], Yousef Saad [543], and Henk van der Vorst [636]. Last but not least we mention the closely related book by Gene Golub and Gérard Meurant [246]. These books, as well as the other books and most survey papers on the subject we are aware of, reflect the current state-of-the-art, which is the outcome of explosive algorithmic developments over the last few decades. Such developments were necessary because of tremendous challenges raised by an ever growing variety of application problems. For many years, investigations on *how* to solve problems computationally and the derivation of new methods and their algorithmic realisations have dominated the analysis of existing approaches.

Our aim with this book is to complement the existing literature by focusing on *mathematical fundamentals* of Krylov subspace methods rather than their algorithmic details, and on *addressing the why* more than the how. In the quote given above, Lanczos announces to Einstein that he has found a new elegant algorithm (he refers to his nowadays classical method for computing eigenvalue approximations) and, more importantly, he explains his main motivation for working on iterative methods. He is attracted to them because they can only be made to work efficiently when they uncover a deeper truth, namely the 'inner nature of the problem'. The utilitarian viewpoint of practical applicability is secondary to him, and apparently to Einstein as well, who in his answer to Lanczos points out the 'fruitful mathematical aspect' involved in Lanczos' thinking. In our book we try to explore precisely this aspect, which in the past has been overshadowed by the algorithmic view.

In the process of writing we went back to the early papers by Krylov (1931), Gantmacher (1934), Lanczos (1950 and 1952), Hestenes and Stiefel (1952), and others from that period. These authors presented many close relationships of their methods with mathematical concepts beyond the realm of what is known today as 'matrix computations'. Examples include quadrature methods, orthogonal polynomials, continued fractions, moments, projections, and invariant subspaces. Reading these original works was a fascinating experience.

Our feeling is nicely expressed by the following quote from the German mathematician Eduard Study (1862–1930): 'Mathematics is neither the art of calculation nor the art of avoiding calculations. Mathematics, however, entails the art of avoiding superfluous calculations and conducting the necessary ones skilfully. In this respect one could have learned from the older authors.' [601, p. 4] (our translation). The fact that reading older authors is not just a matter of studying history has also been stressed by the English writer Joseph Rudyard Kipling (1865–1936), author of *The Jungle Book* and Nobel Prize winner for literature in 1907, who wrote in his essay 'The Uses of Reading' (1928) that 'it is only when one reads what men wrote long ago that one realises how absolutely modern the best of the old things are.' Lanczos formulated a similar point of view even more strongly in his essay 'Why Mathematics?' from 1967: 'But to hail our times as the originator of an entirely new science, which need not bother with the past and has the right to construct everything from scratch, betrays a dangerous short-sightedness which can lead to a dissolution of mathematical research into an empty play with words.'

Returning back to the original sources led us to examine a wide range of areas and their interconnections, and to study many results developed in the 19th and the early 20th centuries. Examples are such classics as continued fractions with their relationship to orthogonal polynomials, quadrature and minimal partial realisation in the works of Gauss (1814), Jacobi (1826), Christoffel (1857, 1858), Chebyshev (1855, 1859), Markov (1884), Stieltjes (1884, 1894), and many others. Further classical results we looked at and include in this book are Cauchy's interlacing theorem (1829), Jacobi's reduction to tridiagonal form (1848), Jordan's canonical form (1870), Stieltjes' moment problem (1894), and the Riemann–Stieltjes integral representation of operators by von Neumann (1927, 1932) going back to Hilbert (1906, 1912), who praised the work of Stieltjes.

In order to make the book as self-contained as possible, we have included complete proofs of many stated results. In addition, we have tried to the best of our abilities to give references to the original sources. Throughout the book we have used framed boxes to highlight points in the development that we consider particularly important. In these boxes we usually skip some technical details in order to focus on the main message of the corresponding mathematical results or questions.

To keep the project manageable for ourselves, we have focused on methods for solving linear algebraic systems, and we left aside the equally interesting area of

Krylov subspace based eigenvalue solvers. However, many results presented in this book are relevant for such solvers as well, since they are usually based on the same principles as the methods for linear algebraic systems (with the Lanczos and Arnoldi algorithms as the basic building blocks).

As indicated above, we strongly believe that a mathematical theory can be better understood when it is viewed in its historical context. We therefore discuss the original developments in extensive historical notes. In our opinion, the knowledge of early developments can also help in understanding very recent computational developments. The outcome of the historically motivated approach therefore is practical and readily applicable. It shows what can and what cannot be expected from Krylov subspace methods *today*. Moreover, it challenges some common 'modern' views that have been articulated in the justification of 'practical', though mathematically questionable, approaches.

When solving real-world and large-scale problems, Krylov subspace methods must always be combined with acceleration techniques. The goal is to improve the behaviour of Krylov subspace methods, and the techniques are (somewhat imprecisely) called 'preconditioning'. Construction of preconditioners is usually based on some specific properties of the real-world problem or on empirical (which does not mean simple!) observations and heuristics. Successful application of preconditioning often requires an extensive and deep theoretical knowledge from many areas combined with skilful implementation of graph-theoretical ideas. Therefore it is sometimes viewed as 'a combination of art and science'. In this book we do not explicitly consider preconditioning techniques. (They are studied, for example, in [272, Part II], [452, Chapters 8–10], and [543, Chapters 9–14]; see also the survey [54].) Nevertheless, we believe that our book also contributes to the area of preconditioning. Since most of the presented analysis is applicable to preconditioned systems, it applies, assuming exact arithmetic, to preconditioned Krylov subspace methods. Moreover, many results can be modified in order to describe finite precision computations with preconditioning. Most important of all, we believe that a better understanding of the fundamentals of Krylov subspace methods is a prerequisite for establishing an analytic base on which a theory of preconditioning can be developed.

OVERVIEW OF THE BOOK

The book contains five chapters. The first chapter, *Introduction*, introduces the general setting of the book and the context of solving a real-world problem via the stages of modelling, discretisation, and computation. We also recall the richness of ideas related to the above mentioned original works of Hestenes, Stiefel, and Lanczos. Many of the mathematical topics addressed in the related works are closely examined in Chapters 2–4, which form the theoretical core of our book. In these chapters we consider Krylov subspace methods from different points of view, and we make links between these viewpoints.

Chapter 2, *Krylov Subspace Methods*, focuses on the idea of *projections*. The so-called 'finite termination property' then naturally leads to the introduction of the

Krylov subspaces. We characterise the major methods CG, MINRES, GMRES, and SYMMLQ in terms of their projection properties. Using these properties and the standard approaches for generating Krylov subspace bases (namely the Lanczos and Arnoldi algorithms), we derive algorithmic descriptions of some Krylov subspace methods.

In Chapter 3, *Matching Moments and Model Reduction View,* we consider the ideas of *moments* and *model reduction,* starting from (a simplified version of) Stieltjes' classical moment problem. We discuss important related concepts, ranging from the Gauss–Christoffel quadrature and orthogonal polynomials to continued fractions. Through Jacobi matrices we find the matrix computations analogies in Krylov subspace methods, and we characterise the methods in terms of their moment matching properties.

The central concept of Chapter 4, *Short Recurrences for Generating Orthogonal Krylov Subspace Bases,* is the *invariant subspace.* We discuss how the length of a Krylov sequence is related to the Jordan canonical form of the given matrix. The main goal of the chapter is to explain when a Krylov sequence can be orthogonalised with a short recurrence. The main result in this context motivates the general distinction in the area of Krylov subspace methods between methods for Hermitian and non-Hermitian matrices.

Some of the general ideas and relationships presented in the first four chapters are illustrated in Chapter 5, *Cost of Computations Using Krylov Subspace Methods.* The chapter starts with a general discussion of the concept of computational cost and related issues, including the difference between direct and iterative methods, particular computations and complexity, and the concept of convergence in general. We then focus on the major methods CG and GMRES, and analyse their exact and finite precision behaviour. We summarise and present many results published previously (scattered throughout many papers), while making no claim for completeness of coverage. Some results and views presented here were not previously published, or were just briefly mentioned without an extensive treatment. Moreover, we feel that there is a need to pose and to investigate new questions in relation to application areas. For example, the questions of measuring the error and evaluating the cost when solving practical problems with Krylov subspace methods cannot be resolved, in our opinion, within the field of matrix computations alone. The fact that they need a much wider context is one of the challenges that is formulated in Chapter 5. The chapter ends with a discussion of some open questions, omitted topics, and an outlook.

Summing up, our goal is neither to give a classification of all existing Krylov subspace methods, nor to review or reference all existing approaches. Rather, we want to identify the major ideas and thoroughly analyse the resulting major methods, with algorithmic details presented only when they are relevant for the exposition. (For additional algorithmic descriptions we give appropriate references.) We thus attempt to be *analytic rather than algorithmic* and *focused rather than encyclopedic.* We hope that the readers of this book might find this approach stimulating for further analytic investigations of Krylov subspace methods as well as for using these methods more effectively in the future.

ACKNOWLEDGEMENTS

The book has been written out of our experience in research and teaching, and in discussing results with professionals in the field as well as in application areas. Many friends, students, and colleagues have helped us over the years with comments and suggestions for improvements of all kinds. It is our pleasure to thank, in particular, Michele Benzi, Hanka Bílková, Jurjen Duintjer-Tebbens, Martin Gander, André Gaul, Tomáš Gergelits, Anne Greenbaum, Lek Heng Lim, Robert Luce, Gerard Meurant, Chris Paige, Jan Papež, Miroslav Rozložník, Olivier Sète, Daniel Szyld, Petr Tichý and Miroslav Tůma. In both our personal and scientific lives, the late Gene Golub played an important role. His views have deeply influenced our understanding of the subject and thus the presentation in this book.

We gratefully acknowledge the Deutsche Forschungsgemeinschaft and the Grant Agencies of the Czech Republic and of the Academy of Science of the Czech Republic for support of the research that has led us to the publication of this book. Jörg Liesen was supported by the Emmy Noether and Heisenberg programmes of DFG, and Zdeněk Strakoš was supported by the projects IAA100300802, M100300901, and GAČR 201/09/0917. Furthermore, we thank the editors and staff at Oxford University Press for their interest in the topic and for their encouragement to get the book finished. Endre Süli, Clare Charles, Viki Mortimer, and Keith Mansfield were extremely helpful in this process.

In the writing we greatly benefited from the MATLAB computing language and interactive environment, which was used in all numerical experiments in this book, the LaTeX document preparation system, and the MathSciNet database of the American Mathematical Society.

Most of all, we would like to thank our families and our wives Anja and Ivanka for their lasting support. Without their love, patience, care, and encouragement this work would not have been done.

While working on this book we often felt, perhaps even more strongly than at other times, what is written in 1 Cor 4:7:

What do you possess that you have not received?

Jörg Liesen and Zdeněk Strakoš
August 2012

CONTENTS