Preface

The first version of this book was a set of lecture notes for a graduate course on *Data mining and applications in science and technology* organized by the Swedish *National Graduate School in Scientific Computing (NGSSC).* Since then the material has been used and further developed for an undergraduate course on *Numerical algorithms for data mining and IT* at Linköping University. This is a second course in scientific computing for computer science students.

The book is intended primarily for undergraduate students who have previously taken an introductory scientific computing/numerical analysis course. It may also be useful for early graduate students in various data mining and pattern recognition areas, who need an introduction to linear algebra techniques.

The purpose of the book is to demonstrate that there are several very powerful numerical linear algebra techniques for solving problems in different areas of data mining and pattern recognition. To achieve this goal, it is necessary to present material that goes beyond what is normally covered in a first course in scientific computing (numerical analysis) at a Swedish university. On the other hand, since the book is application-oriented, it is not possible to give a comprehensive treatment of the mathematical and numerical aspects of the linear algebra algorithms used.

The book has three parts. After a short introduction to a couple of areas of data mining and pattern recognition, linear algebra concepts and matrix decompositions are presented. I hope that this is enough for the student to use matrix decompositions in problem solving environments like MATLAB. Some mathematical proofs are given, but the emphasis is on the existence and properties of the matrix decompositions rather than how they are computed. In Part II the linear algebra techniques are applied to data mining problems. Naturally, the data mining and pattern recognition repertoire is quite limited: I have chosen problem areas that are well suited for linear algebra techniques. In order to use intelligently the powerful software for computing matrix decompositions available in MATLAB etc. it is necessary to have some understanding of the underlying algorithms. A very short introduction to eigenvalue and singular value algorithms is given in Part III.

I have not had the ambition to write a book of recipes: "given a certain problem, here is an algorithm for its solution". That would be difficult, as the area is far too diverse to give clear-cut, and simple solutions. Instead, my intention has been to give the student a set of tools that may be tried as they are, but more likely, will need to be modified to be useful for a particular application. Some of the methods in the book are described using MATLAB scripts. They should not be

considered as serious algorithms, but rather as pseudo-codes given for illustration purposes.

A collection of exercises and computer assignments are available at the web page of the book `http://www.mai.liu.se/~laeld/matrix-methods/`.

The support from NGSSC for producing the original lecture notes is gratefully acknowledged. The lecture notes have been used by a couple of colleagues. Thanks are due to Gene Golub and Saara Hyvönen for helpful comments. Several of my own students have helped me to improve the presentation by pointing out inconsistencies and asking questions. I am indebted to Berkant Savas for letting me use results from his Masters thesis in Chapter 10. Three anonymous referees have read earlier versions of the book and given suggestions for improvements. Finally, I would like to thank Nick Higham, series editor at SIAM, for carefully reading the manuscript. His thoughtful advice helped me improve the contents and the presentation considerably.

Linköping, October 2006

Lars Eldén