Ivan Markovsky

Low Rank Approximation

Algorithms, Implementation, Applications

July 3, 2011

Springer

Preface

Mathematical models are obtained from first principles (natural laws, interconnection, *etc*.) and experimental data. Modeling from first principles is common in natural sciences, while modeling from data is common in engineering. In engineering, often experimental data is available and a simple approximate model is preferred to a complicated detailed one. Indeed, although optimal prediction and control of a complex (high-order, nonlinear, time-varying) system is currently difficult to achieve, robust analysis and design methods, based on a simple (low-order, linear, time-invariant) approximate model, may achieve sufficiently high performance. This book addresses the problem of *data approximation by low-complexity models*.

A unifying theme of the book is low rank approximation: a prototypical data modeling problem. The rank of a matrix constructed from the data corresponds to the complexity of a linear model that fits the data exactly. The data matrix being full rank implies that there is no exact low complexity linear model for that data. In this case, the aim is to find an approximate model. One approach for approximate modeling, considered in the book, is to find small (in some specified sense) modification of the data that renders the modified data exact. The exact model for the modified data is an optimal (in the specified sense) approximate model for the original data. The corresponding computational problem is low rank approximation. It allows the user to trade off accuracy vs complexity by varying the rank of the approximation.

The distance measure for the data modification is a user choice that specifies the desired approximation criterion or reflects prior knowledge about the accuracy of the data. In addition, the user may have prior knowledge about the system that generates the data. Such knowledge can be incorporated in the modeling problem by imposing constraints on the model. For example, if the model is known (or postulated) to be a linear time-invariant dynamical system, the data matrix has Hankel structure and the approximating matrix should have the same structure. This leads to a Hankel structured low rank approximation problem.

A tenet of the book is: the estimation accuracy of the basic low rank approximation method can be improved by exploiting prior knowledge, *i.e.*, by adding constraints that are known to hold for the data generating system. This path of development leads to weighted, structured, and other constrained low rank approximation problems. The theory and algorithms of these new classes of problems are interesting in their own right and being application driven are practically relevant.

Stochastic estimation and deterministic approximation are two complementary aspects of data modeling. The former aims to find from *noisy data*, generated by a low-complexity system, an estimate of that data generating system. The latter aims to find from *exact data*, generated by a high complexity system, a low-complexity approximation of the data generating system. In applications both the stochastic estimation and deterministic approximation aspects are likely to be present. The data is likely to be imprecise due to measurement errors and is likely to be generated by a complicated phenomenon that is not exactly representable by a model in the considered model class. The development of data modeling methods in system identification and signal processing, however, has been dominated by the stochastic estimation point of view. If considered, the approximation error is represented in the mainstream data modeling literature as a random process. This is not natural because the approximation error is by definition deterministic and even if considered as a random process, it is not likely to satisfy standard stochastic regularity conditions such as zero mean, stationarity, ergodicity, and Gaussianity.

An exception to the stochastic paradigm in data modeling is the behavioral approach, initiated by J. C. Willems in the mid 80's. Although the behavioral approach is motivated by the deterministic approximation aspect of data modeling, it does not exclude the stochastic estimation approach. In this book, we use the behavioral approach as a *language* for defining different modeling problems and presenting their solutions. We emphasize the importance of deterministic approximation in data modeling, however, we formulate and solve stochastic estimation problems as low rank approximation problems.

Many well known concepts and problems from systems and control, signal processing, and machine learning reduce to low rank approximation. Generic examples in system theory are model reduction and system identification. The principal component analysis method in machine learning is equivalent to low rank approximation, which suggests that related dimensionality reduction, classification, and information retrieval problems can be phrased as low rank approximation problems. Sylvester structured low-rank approximation has applications in computations with polynomials and is related to methods from computer algebra.

The developed ideas lead to algorithms, which are implemented in software. The algorithms clarify the ideas and the software implementation clarifies the algorithms. Indeed, the software is the ultimate unambiguous description of how the ideas are put to work. In addition, the provided software allows the reader to reproduce the examples in the book and to modify them. The exposition reflects the sequence

$$\text{theory} \mapsto \text{algorithms} \mapsto \text{implementation}.$$

Correspondingly, the text is interwoven with code that generates the numerical examples being discussed.

Prerequisites and practice problems

A common feature of the current research activity in all areas of science and engineering is the narrow specialization. In this book, we pick applications in the broad area of data modeling, posing and solving them as low rank approximation problems. This unifies seemingly unrelated applications and solution techniques by emphasising their common aspects (*e.g.*, complexity–accuracy trade-off) and abstracting from the application specific details, terminology, and implementation details. Despite of the fact that applications in systems and control, signal processing, machine learning, and computer vision are used as examples, the only real prerequisites for following the presentation is knowledge of linear algebra.

The book is intended to be used for self study by researchers in the area of data modeling and by advanced undergraduate / graduate level students as a complementary text for a course on system identification or machine learning. In either case, the expected knowledge is undergraduate level linear algebra. In addition, MATLAB code is used, so that familiarity with MATLAB's programming language is required.

Passive reading of the book gives a broad perspective on the subject. Deeper understanding, however, requires active involvement, such as supplying missing justification of statements and specific examples of the general concepts, application and modification of presented ideas, and solution of the provided exercises and practice problems. There are two types of practice problems: analytical, asking for a proof of a statement clarifying or expanding the material in the book, and computational, asking for experiments with real or simulated data of specific applications. Most of the problems are easy to medium difficulty. A few problems (marked with stars) can be used as small research projects.

The code in the book has been tested with MATLAB version 7.9.0, running under Linux, and uses the Optimization Toolbox version 4.3, Control System Toolbox version 8.4, and Symbolic Math Toolbox version 5.3. An Octave compatible version of the code is available from the book's web page.

Software tools used for typesetting the book

The book is typeset using LaTeX and a number of extension packages. Diagrams are created using the PSTricks and `xy` packages and the standalone `xfig` program. All editing is done in the GNU Emacs editor, using `org-mode` and `latex-mode`. MATLAB code, presented in a literate programming style is an integral part of the text. This is achieved by the Norman Ramsey's `noweb` system.

The process of compiling the book involves a number of steps automated by a GNU Make file. First the LaTeX source code and the MATLAB functions and scripts are extracted from `noweb`'s source files. The obtained scripts are then executed in MATLAB in order to generate the figures and numerical results. Finally, LaTeX, bibTeX, `dvips`, and `ps2pdf` are run on the files created in the previous steps to generate the final pdf file of the book. The described process guarantees that the code in the text is the actual code that has generated the results shown in the book.

A laptop and a desktop computers were used for working on the book. Keeping the files synchronized between the two computers was done using the Unison file synchronizer.

Acknowledgements

A number of individuals and the European Research Council contributed and supported me during the preparation of the book. Oliver Jackson—Springer's editor (engineering)—encouraged me to embark on the project. My colleagues in ESAT/SISTA, K.U.Leuven and ECS/ISIS, Southampton, U.K. created the right environment for developing the ideas in the book. In particular, I am in debt to Jan C. Willems (SISTA) for his personal guidance and example of critical thinking. The behavioral approach that Jan initiated in the early 1980's is present in this book.

Maarten De Vos (SISTA), Konstantin Usevich (Saint Petersburg state university), and Jan Willems (SISTA) proofread and suggested improvements for the first chapters of the book. I gratefully acknowledge funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement number 258581 "Structured low-rank approximation: Theory, algorithms, and applications".

Southampton, *Ivan Markovsky*
July 3, 2011

Contents

Part I Linear modeling problems

Part II Miscellaneous generalizations