Parallel optimization: Theory, algorithms, and applications

Book · January 1997

Stavros A. Zenios
University of Cyprus

PARALLEL OPTIMIZATION
Theory, Algorithms and Applications

Series on Numerical Mathematics and Scientific Computation

PARALLEL OPTIMIZATION
Theory, Algorithms and Applications

Yair Censor
Department of Mathematics and Computer Science
University of Haifa

Stavros A. Zenios
Department of Public and Business Administration
University of Cyprus

New York *Oxford*
Oxford University Press
1997

Oxford University Press

Material to be provided by OUP.

To Erga, Aviv, Nitzan and Keren — Y.C.

To Christiana, Efy and Elena — S.A.Z.

Foreword

This book is a must for anyone interested in entering the fascinating new world of **parallel optimization** using parallel processors — computers capable of doing an enormous number of complex operations in a nanosecond.

The authors are among the pioneers of this fascinating new world and they tell us what new applications they explored, what algorithms appear to work best, how parallel processors differ in their design, and what the comparative results were using different types of algorithms on different types of parallel processors to solve them.

According to an old adage, the whole can sometimes be much more than the sum of its parts. I am thoroughly in agreement with the authors' belief in the added value of bringing together **Applications**, **Mathematical Algorithms** and **Parallel Computing techniques**. This is exactly what they found true in their own research and report on in the book.

Many years ago, I, too, experienced the thrill of combining three diverse disciplines: the **Application** (in my case Linear Programs), the **Solution Algorithm** (the Simplex Method), and the then **New Tool** (the Serial Computer). The union of the three made possible the optimization of many real-world problems. Parallel processors are the new generation and they have the power to tackle applications which require solution in real time, or have model parameters which are not known with certainty, or have a vast number of variables and constraints.

Image restoration tomography, radiation therapy, finance, industrial planning, transportation and economics are the sources for many of the interesting practical problems used by the authors to test the methodology.

George B. Dantzig
Stanford University, 1996

Preface

> As the sun eclipses the stars by his brilliancy so the one of knowledge will eclipse the fame of the assemblies of the people if he proposes algebraic problems, and still more if he solves them.
>
> **Brahmagupta**, 650 AD.

Problems of mathematical optimization are encountered in diverse areas of the exact sciences, the natural sciences, the social sciences and engineering. Many of them are rooted in real-world applications. Developments in the vast field of optimization are, to a great extent, motivated by these applications and have drawn, over the years, both from mathematics and from computer science. Mathematics creates the foundation for the design and analysis of optimization algorithms. Computer science provides the tools for the design of data-structures, and for the translation of the mathematical algorithms into numerical procedures that are implementable on a computer. The efficient and robust implementation of an optimization algorithm becomes crucial when one deals with the solution of large-scale, real-world applications.

Recent technological innovations, with the introduction of parallel computer architectures, are having a significant impact on every area of scientific computing where large-scale problems are attacked. In this book we give an introduction to methods of **parallel optimization**. We do so by introducing parallel computing ideas and techniques into both optimization theory and into numerical algorithms for large-scale optimization problems. We also examine significant broad areas of application where the problems are particularly suitable for solution on parallel machines, and where substantial progress has been made in recent years with the application of parallel optimization algorithms.

Some mathematical algorithms that are recognized today as being parallel algorithms date back to the 1920s and some efforts in using parallel computers to solve optimization problems were made in the late 1970s, with the introduction of the Illiac IV array processor at the University of Illinois. However, it was in the early 1980s that concentrated and systematic efforts started by several researchers in the field of parallel optimization. Several of the contributions that were made over the last two decades have matured to the point where a coherent theoretical framework has been developed, extensive numerical experiments have been carried out, and large-scale problems from diverse areas of application have been solved successfully.

This book gives a comprehensive account of these developments. The coverage is unavoidably not exhaustive, since parallel computing technology has influenced recent developments in all areas of optimization. A series of books could be written on parallel computing for linear programming, large-scale constrained optimization, unconstrained optimization, global optimization and combinatorial optimization; see Section 1.5 for references. This book focuses on parallel optimization methods for large-scale constrained optimization problems and structured linear programs. Hence, it provides a comprehensive chart of part of the vast intersection between parallel computing and optimization. We set out to describe a domain where parallel computing is having a great impact — precisely because of the large-scale nature of the applications — and where many of the recent research developments have occurred. Even within this domain we do not claim that the material about theory, parallel algorithms and applications presented here is exhaustive. However, related developments that are not treated in the book are discussed in extensive "Notes and References" sections at the end of each chapter.

What, then, has determined our choice of theory, algorithms and applications that were included in the book? We have focused on methods where substantial computational experiences have been accumulated over the years, and where, we feel, substantial integration has been achieved between the theory, the algorithms and the applications.

Quite often the implementation of an algorithm changes one's perspective of what are the important features of the algorithm, and such accumulated experience, that we have acquired through our own work in the field, is reflected in our treatment. The intricacies of exploiting the problem structure are also fully revealed only during an implementation. Finally, it is only with computational experiments that we can have full confidence in the efficiency and robustness of an algorithm. The material presented in this book leads to **implementable** parallel algorithms that have undergone the scrutiny of implementation on a variety of parallel architectures. In addition, our choice of topics is broad enough so that readers can get a comprehensive view of the landscape of parallel optimization methods.

While not all currently known parallel algorithms are discussed, the book introduces algorithms from three broad families of algorithms for constrained optimization. Those are defined later in the book as (i) **iterative projection algorithms**, (ii) **model decomposition algorithms**, and (iii) **interior point algorithms**. When viewed from the proper perspective these algorithms satisfy the design characteristics of "good" parallel algorithms.

The book starts with a basic introduction to parallel computers: what they are, how to assess their performance, how to design and implement parallel algorithms. This core knowledge on parallel computers is then linked with the theoretical algorithms. The combined mathematical algorithms and parallel computing techniques are brought together to bear on the solution of several important applications: image reconstruction from projections, matrix balancing, network optimization, nonlinear programming for planning under uncertainty, and financial planning. We also address implementation issues and study results from recent numerical works that highlight the efficiency of the developed algorithms, when implemented on suitable parallel computer platforms.

We believe that the value of bringing together applications, mathematical algorithms and parallel computing techniques, extends beyond the successful solution of the specific problems at hand. Namely, it introduces the reader to the complete process from the modeling of a problem through the design of solution algorithms, and to the art and science of parallel computations. It is not possible to study these three disciplinary efforts — modeling, mathematics of algorithms and parallel computing — in isolation from each other. The successful solution of real-world problems in scientific computing is the result of coordinated efforts across all three fronts. We hope that this book will help the reader to develop such a broad perspective and, thus, follow Brahmagupta's admonishment.

To keep the size of the book reasonable we had to make some decisions on what topics to exclude and about the prerequisites that are assumed by the reader. Many important topics related to the subject matter of the book have been left out or are only mentioned casually. These include questions of rate of convergence, computational complexity, stopping criteria, behavior of the algorithms in inconsistent cases and so on.

Regarding prerequisites we assume that the reader has been systematically exposed to differential and integral calculus, linear algebra, convex analysis and optimization theory. Sections 10.2, 10.3 and Chapter 13 assume familiarity with notions from probability theory.

Finally, inspite of the large bibliography included at the end of the book we might have missed relevant references or erred in crediting work done by others. We will be grateful to readers who bring to our attention such ommissions so that we can correct them in the future.

Organization of the Book

The material of this book is organized in three parts. First, Chapter 1 introduces the fundamental topics on parallel computing.

Part I: Theory, develops the theory of generalized distances and generalized projections (Chapter 2) and the theory for their use in solving linear programming problems via proximal minimization (Chapter 3). The theory of penalty and barrier methods and augmented Lagrangians is developed in (Chapter 4). This material provides the theoretical foundation upon which the algorithms in Chapters 5 to 8 are developed.

Part II: Algorithms, develops iterative projection algorithms, model decomposition algorithms, and interior point algorithms. Chapter 5 discusses iterative algorithms for the solution of convex feasibility problems, using the theory of generalized projections. Similarly, Chapter6 uses the theory of generalized distances and generalized projections to develop algorithms for linearly constrained optimization problems. Chapter 7 develops model decomposition algorithms, based on the theory of penalty methods and augmented Lagrangians. Chapter 8 introduces interior point algorithms for

Fig. 0.1 Organization of the book.

linear and quadratic programming, and explains ways in which the structure of some large-scale optimization problems can be exploited by these algorithms for parallel computations.

Part III: Applications, discusses applications from several diverse real-world domains where the parallel algorithms are applicable. Each chapter contains a description of the real-world application, it develops one or more mathematical models for each problem, and discusses solution algorithms from one or more of the algorithm classes of Part II. Chapter 9 discusses problems of matrix estimation. Chapter 10 discusses problems of image reconstruction from projections. Chapter 11 discusses the problem of radiation therapy treatment planning. Chapter 12 discusses problems in transportation and the multicommodity flow problem. Chapter 13 discusses problems of planning under uncertainty using stochastic programming and robust optimization models.

Finally, two chapters are devoted to the parallel implementation and testing of the algorithms. Implementations are discussed in Chapter 14. The issue of implementations is an important one, but it is bound to be linked closely with the computer architecture. To the extent possible our discussion is linked to a whole class of machines and not just to a specific hardware model. Chapter15 summarizes numerical experiences with several of the algorithms that demonstrate their effectiveness for solving large-scale problems when implemented on parallel machines.

Figure 0.1 illustrates the interdependencies among the chapters. The sequence of chapters indicated in this diagram must be followed in order to appreciate fully a line of development from its theoretical foundations, to the algorithms and their applications. While we emphasize the importance of studying the continuum of theory-algorithms-applications, the chapters of the book are written in such a way that they can be used as a reference, without the need to study them sequentially. Readers who are interested only in the applications and the mathematical models may read the relevant chapters from Part III, without reading first the chapters on algorithms from Part II. Of course, in order to fully appreciate the solution algorithms for the models one has to read the earlier chapters as well. But even then, a solution algorithm can be understood without referring to the relevant chapters on theory from Part I, unless the reader wishes to understand the proof of convergence as well. The book can, therefore, be used either as a textbook or as a reference book.

Suggested course outlines

The book is organized in a way that allows it to be used as a text for graduate courses in large-scale optimization, parallel computing or large-scale mathematical modeling. There are three different avenues that an instructor may follow in teaching this material, especially bearing in mind that the whole book can not be covered in the usual time frame of a one-semester course. No matter which avenue an instructor may decide to follow, Chapter 1 gives a general introduction to the material and should be covered first.

One approach is to teach a course on theory and algorithms for constrained optimization and feasibility problems. Such a course will cover the theory part of the book, Chapters2, 3 and 4, followed by the algorithms in Chapters 5, 6, 7 and 8. Any one of the Chapters 3, 7 or 8 could be omitted without loss of continuity, but a balanced treatment of different families of optimization algorithms should include Chapters 6, 7 and 8. This course focuses on the theoretical aspects of parallel optimization, and references to parallel computations can be cursory.

A second approach is to teach a course on numerical methods for large-scale structured optimization problems. Such a course will focus on the algorithms part, without prior introduction to the theory that is essential for establishing convergence. The emphasis is on developing the students' understanding of the structure of an algorithm, taking for granted the theory that provides the foundation for its correctness. Such a course will cover the material from Chapters 6, 7 and 8. These chapters present general algorithms. Selected sections from Chapters 9, 10, 12 and 13 illustrate the development of algorithms for specific problems, starting from the general algorithms. In this course the exploitation of special structure is a key issue, and references to parallel computing become crucial. This course could also discuss the implementation of algorithms on parallel machines, with coverage of the material in Chapter 14.

Yet a third approach is to focus on applications of optimization, and present — in a cookbook fashion — implementable algorithms for the solution of real-world instances of large-scale problems. Such a course will teach material from the chapters on applications, and refer to the corresponding chapters in Part II where specific algorithms have been developed for the applications at hand. This course will start with the iterative optimization algorithms of Chapter 6 and move on to the applications Chapters 9, 10, 12 and 13. As in the previous course, the exploitation of special structure is a key issue and references to parallel computing become crucial. This course could also discuss the implementation of algorithms on parallel machines as given in Chapter 14.

The material of the book can also be used for a course on parallel computing. Several of the algorithms are simple enough so that students with little background in optimization can readily understand them, and such algorithms can be the focus for implementation exercises. Furthermore, the structures of the underlying models vary from the very simple dense matrix (e.g., the dense transportation problems of Sections 12.2.1 and 12.4.1, to sparse and structured matrices and graph problems (e.g., the multicommodity transportation problems and the stochastic networks of Sections 12.2.2, 12.4.3, and 13.8, respectively). Hence, the material can be used to introduce students to the art of implementing algorithms on parallel machines. Such a course will provide some motivation by introducing applications from Chapters 9, 10, 12 and 13. For each application an algorithm can be introduced (specific implementable algorithms are found in the applications chapters), and references made to the implementation techniques of Chapter 14. The course should follow the sequence of topics as discussed in Chapter 14, but before each section of this chapter is presented in class the material from the corresponding application chapter should be introduced first. Finally, Chapter 15 can be used as a reference for students who wish to test the efficiency of their implementation, or the performance of their parallel machine.

Acknowledgments

Part of the material presented in this book is based on our own published work. We express our appreciation to our past and present collaborators from whom we learned a great deal. Particular thanks for many useful discussions and support, and for reading various parts of the book, are extended to Marty Altschuler, Dimitri Bertsekas, Dan Butnariu, Charlie Byrne, Alvaro De Pierro, Jitka Dupačova, Jonathan Eckstein, Tommy Elfving, Gabor Herman, Dan Gordon, Alfredo Iusem, Elizabeth Jessup, Arnold Lent, Robert Lewitt, Olvi Mangasarian, Jill Mesirov, Bob Meyer, John Mulvey, Soren Nielsen, Mustafa Pinar, Simeon Reich, Uri Rothblum, Michael Schneider, Jay Udupa, Paul Tseng and Dafeng Yang.

A draft version of the book was read by Dimitri Bertsekas, Jonathan Eckstein, Tommy Elfving, Michael Ferris and Gabor Herman. We thank them for their constructive comments. Any remaining errors or imperfections are, of course, our sole responsibility.

Part of the work of Yair Censor in this area was done in collaboration with the Medical Image Processing Group (**MIPG**) at the Department of Radiology, University of Pennsylvania. The support and encouragement of Gabor Herman for this continued collaboration is gratefully acknowledged. The work of Stavros A. Zenios was done while he was on the faculty at the Department of Operations and Information Management, The Wharton School, University of Pennsylvania, and while on leave at the Operations Research Center of the Sloan School, Massachussetts Institute of Technology, and with Thinking Machines Corporation, Cambridge, MA. Substantial parts of Stavros Zenios' work on this book were done during his visits to the University of Bergamo, Italy, and the support of Marida Bertocchi in making these visits possible is gratefully acknowledged. We express our appreciation to these organizations as well as to our current institutions, the University of Haifa, Haifa, Israel and the University of Cyprus, Nicosia, Cyprus, for creating an environment where international collaborations can be fostered and where long-term undertakings, such as the writing of this book, are encouraged and supported.

This text grew out of lecture notes that we prepared and delivered at the **19th Brazilian Mathematical Colloquium**, that took place in Rio de Janeiro in the summer of 1993. We thank the organizers of that Colloquium and Professor Jacob Palis, the Director of the Instituto de Mathemática Pura e Aplicada (IMPA) where the Colloquium took place, for giving us the opportunity to present our work at this forum, and thereby giving us the impetus to complete the project of writing this book.

This work was supported by grants from the National Institutes of Health (HL–28438), the Air-Force Office of Scientific Research (AFOSR-91–0168) and the National Science Foundation (CCR–9104042 and SES–9100216) in the USA, and the National Research Council (CNR), Italy. Computing resources were made available through the North-East Parallel Architectures Center (NPAC) at Syracuse University, the Army High-Performance Computing Research Center (AHPCRC) at the University of Minnesota, Thinking Machines Corporation, Cray Research Inc., and Argonne National Laboratory. We also thank Ms. Danielle Friedlander and Ms. Giuseppina La Mantia for their work in processing parts of the manuscript in LaTeX. Finally, we express our appreciation to Senior Editor Bill Zobrist and Editorial Assistant Krysia Bebick from the Oxford University Press office in New York for their help.

Haifa, Nicosia and Philadelphia
December 1996

Yair Censor
Stavros A. Zenios

Contents

PARALLEL OPTIMIZATION
Theory, Algorithms and Applications