Lecture Notes in Control and Information Sciences 434

For further volumes:
http://www.springer.com/series/642

S. Bhatnagar, H.L. Prasad, and L.A. Prashanth

Stochastic Recursive Algorithms for Optimization

Simultaneous Perturbation Methods

Springer

Authors
Prof. S. Bhatnagar
Department of Computer Science
and Automation
Indian Institute of Science
Bangalore
India

L.A. Prashanth
Department of Computer Science
and Automation
Indian Institute of Science
Bangalore
India

H.L. Prasad
Department of Computer Science
and Automation
Indian Institute of Science
Bangalore
India

ISSN 0170-8643
e-ISSN 1610-7411
ISBN 978-1-4471-4284-3
e-ISBN 978-1-4471-4285-0
DOI 10.1007/978-1-4471-4285-0
Springer London Heidelberg New York Dordrecht

Library of Congress Control Number: 2012941740

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To SB's parents Dr. G. K. Bhatnagar and Mrs. S.K. Bhatnagar, his wife Aarti and daughter Shriya

To HLP's parents Dr. H. R. Laxminarayana Bhatta and Mrs. G. S. Shreemathi, brother-in-law Giridhar N. Bhat and sister Vathsala G. Bhat

To LAP's daughter Samudyata

Preface

The area of *stochastic approximation* has its roots in a paper published by Robbins and Monro in 1951, where the basic stochastic approximation algorithm was introduced. Ever since, it has been applied in a variety of applications cutting across several disciplines such as control and communication engineering, signal processing, robotics and machine learning.

Kiefer and Wolfowitz, in a paper in 1952 (nearly six decades ago) published the first stochastic approximation algorithm for optimization. The algorithm proposed by them was a gradient search algorithm that aimed at finding the maximum of a regression function and incorporated finite difference gradient estimates. It was later found that whereas the Kiefer-Wolfowitz algorithm is efficient in scenarios involving scalar parameters, this is not necessarily the case with vector parameters, particularly those for which the parameter dimension is high. The problem that arises is that the number of function measurements needed at each update epoch grows linearly with the parameter dimension. Many times, it is also possible that the objective function is not observable as such and one needs to resort to simulation. In such scenarios, with vector parameters, one requires a corresponding (linear in the parameter-dimension) number of system simulations. In the case of large or complex systems, this can result in a significant computational overhead.

Subsequently, in a paper published in 1992, Spall proposed a stochastic approximation scheme for optimization that does a random search in the parameter space and only requires two system simulations regardless of the parameter dimension. This algorithm that came to be known as *simultaneous perturbation stochastic approximation* or *SPSA* for short, has become very popular because of its high efficiency, computational simplicity and ease of implementation. Amongst other impressive works, Katkovnik and Kulchitsky, in a paper published in 1972, also proposed a random search scheme (the *smoothed functional (SF) algorithm*) that only requires one system simulation regardless of the parameter dimension. Subsequent work showed that a two-simulation counterpart of this scheme performs well in practice. Both the Katkovnik-Kulchitsky as well as the Spall approaches involve perturbing the parameter randomly by generating certain *i.i.d.* random variables.

The difference between these schemes lies in the distributions these perturbation random variables can possess and the forms of the gradient estimators.

Stochastic approximation algorithms for optimization can be viewed as counterparts of deterministic search schemes with noise. Whereas, the SPSA and SF algorithms are gradient-based algorithms, during the last decade or so, there have been papers published on Newton-based search schemes for stochastic optimization. In a paper in 2000, Spall proposed the first Newton-based algorithm that estimated both the gradient and the Hessian using a simultaneous perturbation approach incorporating *SPSA-type* estimates. Subsequently, in papers published in 2005 and 2007, Bhatnagar proposed more Newton-based algorithms that develop and incorporate both SPSA and SF type estimates of the gradient and Hessian. In this text, we commonly refer to all approaches for stochastic optimization that are based on randomly perturbing parameters in order to estimate the gradient/Hessian of a given objective function as *simultaneous perturbation methods*. Bhatnagar and coauthors have also developed and applied such approaches for constrained stochastic optimization, discrete parameter stochastic optimization and reinforcement learning – an area that deals with the adaptive control of stochastic systems under real or simulated outcomes. The authors of this book have also studied engineering applications of the simultaneous perturbation approaches for problems of performance optimization in domains such as communication networks, vehicular traffic control and service systems.

The main focus of this text is on simultaneous perturbation methods for stochastic optimization. This book is divided into six parts and contains a total of fourteen chapters and five appendices. Part I of the text essentially provides an introduction to optimization problems - both deterministic and stochastic, gives an overview of search algorithms and a basic treatment of the Robbins-Monro stochastic approximation algorithm as well as a general multi-timescale stochastic approximation scheme. Part II of the text deals with gradient search stochastic algorithms for optimization. In particular, the Kiefer-Wolfowitz, SPSA and SF algorithms are presented and discussed. Part III deals with Newton-based algorithms that are in particular presented for the long-run average cost objective. These algorithms are based on SPSA and SF based estimators for both the gradient and the Hessian. Part IV of the book deals with a few variations to the general scheme and applications of SPSA and SF based approaches there. In particular, we consider adaptations of simultaneous perturbation approaches for problems of discrete optimization, constrained optimization (under functional constraints) as well as reinforcement learning. The long-run average cost criterion will be considered here for the objective functions. Part V of the book deals with three important applications related to vehicular traffic control, service systems as well as communication networks. Finally, five short appendices at the end summarize some of the basic material as well as important results used in the text.

This book in many ways summarizes the various strands of research on simultaneous perturbation approaches that SB has been involved with during the course of the last fifteen years or so. Both HLP and LAP have also been working in this area for over five years now and have been actively involved in the various aspects of the research reported here. A large portion of this text (in particular, Parts III-V as well as portions of Part II) is based mainly on the authors' own contributions to this area. The text provides a compact coverage of the material in a way that both researchers and practitioners should find useful. The choice of topics is intended to cover a sufficient width while remaining tied to the common theme of simultaneous perturbation methods. While we have made attempts at conveying the main ideas behind the various schemes and algorithms as well as the convergence analyses, we have also included sufficient material on the engineering applications of these algorithms in order to highlight the usefulness of these methods in solving real-life engineering problems. As mentioned before, an entire part of the text, namely Part IV, comprising of three chapters is dedicated for this purpose. The text in a way provides a balanced coverage of material related to both theory and applications.

Acknowledgements

SB was first introduced to the area of stochastic approximation during his Ph.D work with Prof. Vivek Borkar and Prof. Vinod Sharma at the Indian Institute of Science. Subsequently, he began to look at simultaneous perturbation approaches while doing a post doctoral with Prof. Steve Marcus and Prof. Michael Fu at the Institute for Systems Research, University of Maryland, College Park. He has also benefitted significantly from reading the works of Prof. James Spall and through interactions with him. He would like to thank all his collaborators over the years. In particular, he would like to thank Prof. Vivek Borkar, Prof. Steve Marcus, Prof. Michael Fu, Prof. Richard Sutton, Prof. Csaba Szepesvari, Prof. Vinod Sharma, Prof. Karmeshu, Prof. M. Narasimha Murty, Prof. N. Hemachandra, Dr. Ambedkar Dukkipati and Dr. Mohammed Shahid Abdulla. He would like to thank Prof. Anurag Kumar and Prof. K. V. S. Hari for several helpful discussions on optimization approaches for certain problems in vehicular traffic control (during the course of a joint project), which is also the topic of Chapter 13 in this book. SB considers himself fortunate to have had the pleasure of guiding and teaching several bright students at IISc. He would like to acknowledge the work done by all the current and former students of the Stochastic Systems Laboratory. A large part of SB's research during the last ten years at IISc has been supported through projects from the Department of Science and Technology, Department of Information Technology, Texas Instruments, Satyam Computers, EMC and Wibhu Technologies. SB would also like to acknowledge the various institutions where he worked and visited during the last fifteen years where portions of the work reported here have been conducted: The Institute for Systems Research, University of Maryland, College Park; Vrije Universiteit, Amsterdam; Indian Institute of Technology, Delhi; The RLAI Laboratory, University of Alberta; and the Indian Institute of Science. A major part of the work reported here has been conducted at IISc itself. Finally, he would like to thank his parents Dr. G. K. Bhatnagar and Mrs. S. K. Bhatnagar for their support, help and guidance all through the years, his wife Aarti and daughter Shriya for their patience,

understanding and support, and his brother Dr. Shashank for his guidance and teaching during SB's formative years.

HLP's interest in the area of control engineering and decision making, which was essentially sown in him by interactions with Prof. U. R. Prasad at I.I.Sc., Dr. K. N. Shubhanga and Jora M. Gonda at NITK, led him to the area of operations research followed by that of stochastic approximation. He derives inspirations from the works of Prof. Vivek Borkar, Prof. James Spall, Prof. Richard Sutton, Prof. Shalabh Bhatnagar and several eminent personalities in the field of stochastic approximation. He thanks Dr. Nirmit V. Desai at IBM Research, India, collaboration with whom, brought up several new stochastic approximation algorithms with practical applications to the area of service systems. He thanks Prof. Manjunath Krishnapur and Prof. P. S. Sastry for equipping him with mathematical rigour needed for stochastic approximation. He thanks I. R. Rao at NITK who has been constant source of motivation. He thanks his father Dr. H. R. Laxminarayana Bhatta, mother Mrs. G. S. Shreemathi, brother-in-law Giridhar N. Bhat and sister Vathsala G. Bhat, for their constant support and understanding.

LAP would like to thank his supervising professor SB for introducing stochastic optimization during his PhD work. The extensive interactions with SB served to sharpen the understanding of the subject. LAP would also like to thank HLP, collaboration with whom has been most valuable. LAP's project associateship with Department of Information Technology as well as his internship at IBM Research presented many opportunities for developing as well as applying simultaneous pertubation methods in various practical contexts and LAP would like to thank Prof. Anurag Kumar, Prof. K.V.S. Hari of ECE department, IISc and Nirmit Desai and Gargi Dasgupta of IBM Research for several useful interactions on the subject matter. Finally, LAP would like to thank his family members - particularly his parents, wife and his daughter, for their support in this endeavour.

Bangalore,
May 2012

S. Bhatnagar
H.L. Prasad
L.A. Prashanth

Contents

Part I Introduction to Stochastic Recursive Algorithms

1 Introduction 3
1.1 Introduction 3
1.2 Overview of the Remaining Chapters 7
1.3 Concluding Remarks 11
References 11

2 Deterministic Algorithms for Local Search 13
2.1 Introduction 13
2.2 Deterministic Algorithms for Local Search 14
References 15

3 Stochastic Approximation Algorithms 17
3.1 Introduction 17
3.2 The Robbins-Monro Algorithm 18
3.2.1 Convergence of the Robbins-Monro Algorithm 19
3.3 Multi-timescale Stochastic Approximation 23
3.3.1 Convergence of the Multi-timescale Algorithm 24
3.4 Concluding Remarks 26
References 27

Part II Gradient Estimation Schemes

4 Kiefer-Wolfowitz Algorithm 31
4.1 Introduction 31
4.2 The Basic Algorithm 31
4.2.1 Extension to Multi-dimensional Parameter 35
4.3 Variants of the Kiefer-Wolfowitz Algorithm 36
4.3.1 Fixed Perturbation Parameter 36
4.3.2 One-Sided Variants 37

4.4 Concluding Remarks 38
References 38

5 Gradient Schemes with Simultaneous Perturbation Stochastic Approximation 41
5.1 Introduction 41
5.2 The Basic SPSA Algorithm 41
5.2.1 Gradient Estimate Using Simultaneous Perturbation 42
5.2.2 The Algorithm 43
5.2.3 Convergence Analysis 44
5.3 Variants of the Basic SPSA Algorithm 47
5.3.1 One-Measurement SPSA Algorithm 47
5.3.2 One-Sided SPSA Algorithm 49
5.3.3 Fixed Perturbation Parameter 49
5.4 General Remarks on SPSA Algorithms 51
5.5 SPSA Algorithms with Deterministic Perturbations 52
5.5.1 Properties of Deterministic Perturbation Sequences 52
5.5.2 Hadamard Matrix Based Construction 54
5.5.3 Two-Sided SPSA with Hadamard Matrix Perturbations 56
5.5.4 One-Sided SPSA with Hadamard Matrix Perturbations 62
5.5.5 One-Measurement SPSA with Hadamard Matrix Perturbations 63
5.6 SPSA Algorithms for Long-Run Average Cost Objective 65
5.6.1 The Framework 65
5.6.2 The Two-Simulation SPSA Algorithm 65
5.6.3 Assumptions 66
5.6.4 Convergence Analysis 68
5.6.5 Projected SPSA Algorithm 73
5.7 Concluding Remarks 75
References 75

6 Smoothed Functional Gradient Schemes 77
6.1 Introduction 77
6.2 Gaussian Based SF Algorithm 79
6.2.1 Gradient Estimation via Smoothing 79
6.2.2 The Basic Gaussian SF Algorithm 81
6.2.3 Convergence Analysis of Gaussian SF Algorithm 82
6.2.4 Two-Measurement Gaussian SF Algorithm 88
6.3 General Conditions for a Candidate Smoothing Function 91
6.4 Cauchy Variant of the SF Algorithm 92
6.4.1 Gradient Estimate 92
6.4.2 Cauchy SF Algorithm 94
6.5 SF Algorithms for the Long-Run Average Cost Objective 94
6.5.1 The G-SF1 Algorithm 95
6.5.2 The G-SF2 Algorithm 99

6.5.3 Projected SF Algorithms . . . 100
6.6 Concluding Remarks . . . 101
References . . . 101

Part III Hessian Estimation Schemes

7 Newton-Based Simultaneous Perturbation Stochastic Approximation 105
7.1 Introduction . . . 105
7.2 The Framework . . . 106
7.3 Newton SPSA Algorithms . . . 106
7.3.1 Four-Simulation Newton SPSA (N-SPSA4) . . . 107
7.3.2 Three-Simulation Newton SPSA (N-SPSA3) . . . 109
7.3.3 Two-Simulation Newton SPSA (N-SPSA2) . . . 110
7.3.4 One-Simulation Newton SPSA (N-SPSA1) . . . 111
7.4 Woodbury's Identity Based Newton SPSA Algorithms . . . 113
7.5 Convergence Analysis . . . 114
7.5.1 Assumptions . . . 114
7.5.2 Convergence Analysis of N-SPSA4 . . . 117
7.5.3 Convergence Analysis of N-SPSA3 . . . 125
7.5.4 Convergence Analysis of N-SPSA2 . . . 126
7.5.5 Convergence Analysis of N-SPSA1 . . . 128
7.5.6 Convergence Analysis of W-SPSA Algorithms . . . 130
7.6 Concluding Remarks . . . 130
References . . . 131

8 Newton-Based Smoothed Functional Algorithms . . . 133
8.1 Introduction . . . 133
8.2 The Hessian Estimates . . . 134
8.2.1 One-Simulation Hessian SF Estimate . . . 134
8.2.2 Two-Simulation Hessian SF Estimate . . . 136
8.3 The Newton SF Algorithms . . . 137
8.3.1 The One-Simulation Newton SF Algorithm (N-SF1) . . . 137
8.3.2 The Two-Simulation Newton SF Algorithm (N-SF2) . . . 138
8.4 Convergence Analysis of Newton SF Algorithms . . . 139
8.4.1 Convergence of N-SF1 . . . 139
8.4.2 Convergence of N-SF2 . . . 146
8.5 Concluding Remarks . . . 147
References . . . 147

Part IV Variations to the Basic Scheme

9 Discrete Parameter Optimization . . . 151
9.1 Introduction . . . 151
9.2 The Framework . . . 152
9.2.1 The Deterministic Projection Operator . . . 153
9.2.2 The Random Projection Operator . . . 154

9.2.3 A Generalized Projection Operator . . . 155
9.2.4 Regular Projection Operator to $\bar{C}$ . . . 157
9.2.5 Basic Results for the Generalized Projection Operator Case . 157
9.3 The Algorithms . . . 160
9.3.1 The SPSA Algorithm . . . 160
9.3.2 The SFA Algorithm . . . 161
9.3.3 Convergence Analysis . . . 162
9.4 Concluding Remarks . . . 164
References . . . 166

10 Algorithms for Constrained Optimization . . . 167
10.1 Introduction . . . 167
10.2 The Framework . . . 168
10.3 Algorithms . . . 171
10.3.1 Constrained Gradient-Based SPSA Algorithm (CG-SPSA) . 172
10.3.2 Constrained Newton-Based SPSA Algorithm (CN-SPSA) . . 173
10.3.3 Constrained Gradient-Based SF Algorithm (CG-SF) . . . 175
10.3.4 Constrained Newton-Based SF Algorithm (CN-SF) . . . 176
10.4 A Sketch of the Convergence . . . 178
10.5 Concluding Remarks . . . 185
References . . . 186

11 Reinforcement Learning . . . 187
11.1 Introduction . . . 187
11.2 Markov Decision Processes . . . 188
11.3 Numerical Procedures for MDPs . . . 191
11.3.1 Numerical Procedures for Discounted Cost MDPs . . . 192
11.3.2 Numerical Procedures for Long-Run Average Cost MDPs . . 193
11.4 Reinforcement Learning Algorithms for Look-up Table Case . . . 194
11.4.1 An Actor-Critic Algorithm for Infinite Horizon Discounted Cost MDPs . . . 195
11.4.2 The Q-Learning Algorithm and a Simultaneous Perturbation Variant for Infinite Horizon Discounted Cost MDPs . . . 198
11.4.3 Actor-Critic Algorithms for Long-Run Average Cost MDPs 202
11.5 Reinforcement Learning Algorithms with Function Approximation . 206
11.5.1 Temporal Difference (TD) Learning with Discounted Cost . . 206
11.5.2 An Actor-Critic Algorithm with a Temporal Difference Critic for Discounted Cost MDPs . . . 210
11.5.3 Function Approximation Based Q-learning Algorithm and a Simultaneous Perturbation Variant for Infinite Horizon Discounted Cost MDPs . . . 213
11.6 Concluding Remarks . . . 218
References . . . 218

Part V Applications

12 Service Systems 225
12.1 Introduction 225
12.2 Service System Framework 226
12.3 Problem Formulation 228
12.4 Solution Methodology 232
12.5 First Order Methods 234
12.5.1 SASOC-SPSA 234
12.5.2 SASOC-SF-N 235
12.5.3 SASOC-SF-C 236
12.6 Second Order Methods 236
12.6.1 SASOC-H 237
12.6.2 SASOC-W 237
12.7 Notes on Convergence 238
12.8 Summary of Experiments 239
12.9 Concluding Remarks 240
References 240

13 Road Traffic Control 243
13.1 Introduction 243
13.2 Q-Learning for Traffic Light Control 245
13.2.1 Traffic Control Problem as an MDP 245
13.2.2 The TLC Algorithm 246
13.2.3 Summary of Experimental Results 248
13.3 Threshold Tuning Using SPSA 248
13.3.1 The Threshold Tuning Algorithm 249
13.3.2 Traffic Light Control with Threshold Tuning 250
13.3.3 Summary of Experimental Results 254
13.4 Concluding Remarks 255
References 255

14 Communication Networks 257
14.1 Introduction 257
14.2 The Random Early Detection (RED) Scheme for the Internet 258
14.2.1 Introduction to RED Flow Control 258
14.2.2 The Framework 259
14.2.3 The B-RED and P-RED Stochastic Approximation Algorithms 263
14.2.4 Summary of Experimental Results 266
14.3 Optimal Policies for the Retransmission Probabilities in Slotted Aloha 267
14.3.1 Introduction to the Slotted Aloha Multiaccess Communication Protocol 267
14.3.2 The SDE Framework 268

14.3.3 The Algorithm 270
14.3.4 Summary of Experimental Results 272
14.4 Dynamic Multi-layered Pricing Schemes for the Internet 272
14.4.1 Introduction to Dynamic Pricing Schemes 273
14.4.2 The Pricing Framework 273
14.4.3 The Price Feed-Back Policies and the Algorithms 275
14.4.4 Summary of Experimental Results 277
14.5 Concluding Remarks 278
References 279

Part VI Appendix

A Convergence Notions for a Sequence of Random Vectors 283
Reference 285

B Martingales 287
References 289

C Ordinary Differential Equations 291
References 294

D The Borkar-Meyn Theorem for Stability and Convergence of Stochastic Approximation 295
References 296

E The Kushner-Clark Theorem for Convergence of Projected Stochastic Approximation 297
References 300

Index 301

Acronyms

Lists of abbreviations used in the ensuing text, are as follows.

SPSA	Simultaneous Perturbation Stochastic Approximation
SF	Smoothed Functional
SFA	Smoothed Functional Approximation
MDP	Markov Decision Process
SDP	Stationary Deterministic Policy
SRP	Stationary Randomized Policy
RL	Reinforcement Learning
AC	Actor-Critic
SLA	Service Level Agreement
R-M	Robbins-Monro
K-W	Kiefer-Wolfowitz
FDSA	Finite Difference Stochastic Approximation
IPA	Infinitesimal Perturbation Analysis
p.d.f.	Probability Density Function
i.i.d.	Independent and Identically Distributed
a.s.	Almost Surely
w.p.1	With Probability One
ODE	Ordinary Differential Equation
SDE	Stochastic Differential Equation
OCBA	Optimal computing budget allocation
R&S	Ranking and Selection
MCP	Multiple Comparison Procedures
TLC	Traffic Light Control
QoS	Quality of Service
RED	Random Early Detection
CSMA	Carrier Sense Multiple Access
CSMA/CD	CSMA with Collision Detection
TP	Tirupati Pricing
PMP	Paris Metro Pricing