Contents

Series Foreword xi

Preface xiii

1 Introduction: Optimization and Machine Learning
S. Sra, S. Nowozin, and S. J. Wright 1
1.1 Support Vector Machines 2
1.2 Regularized Optimization 7
1.3 Summary of the Chapters 11
1.4 References 15

2 Convex Optimization with Sparsity-Inducing Norms
F. Bach, R. Jenatton, J. Mairal, and G. Obozinski 19
2.1 Introduction 19
2.2 Generic Methods 26
2.3 Proximal Methods 27
2.4 (Block) Coordinate Descent Algorithms 32
2.5 Reweighted-ℓ_2 Algorithms 34
2.6 Working-Set Methods 36
2.7 Quantitative Evaluation 40
2.8 Extensions 47
2.9 Conclusion 48
2.10 References 49

3 Interior-Point Methods for Large-Scale Cone Programming
M. Andersen, J. Dahl, Z. Liu, and L. Vandenberghe 55
3.1 Introduction 56
3.2 Primal-Dual Interior-Point Methods 60
3.3 Linear and Quadratic Programming 64
3.4 Second-Order Cone Programming 71
3.5 Semidefinite Programming 74
3.6 Conclusion 79

3.7 References 79

4 Incremental Gradient, Subgradient, and Proximal Methods for Convex Optimization: A Survey
D. P. Bertsekas **85**
4.1 Introduction 86
4.2 Incremental Subgradient-Proximal Methods 98
4.3 Convergence for Methods with Cyclic Order 102
4.4 Convergence for Methods with Randomized Order 108
4.5 Some Applications 111
4.6 Conclusions 114
4.7 References 115

5 First-Order Methods for Nonsmooth Convex Large-Scale Optimization, I: General Purpose Methods
A. Juditsky and A. Nemirovski **121**
5.1 Introduction 121
5.2 Mirror Descent Algorithm: Minimizing over a Simple Set . . . 126
5.3 Problems with Functional Constraints 130
5.4 Minimizing Strongly Convex Functions 131
5.5 Mirror Descent Stochastic Approximation 134
5.6 Mirror Descent for Convex-Concave Saddle-Point Problems . 135
5.7 Setting up a Mirror Descent Method 139
5.8 Notes and Remarks 145
5.9 References 146

6 First-Order Methods for Nonsmooth Convex Large-Scale Optimization, II: Utilizing Problem's Structure
A. Juditsky and A. Nemirovski **149**
6.1 Introduction 149
6.2 Saddle-Point Reformulations of Convex Minimization Problems 151
6.3 Mirror-Prox Algorithm 154
6.4 Accelerating the Mirror-Prox Algorithm 160
6.5 Accelerating First-Order Methods by Randomization 171
6.6 Notes and Remarks 179
6.7 References 181

7 Cutting-Plane Methods in Machine Learning
V. Franc, S. Sonnenburg, and T. Werner **185**
7.1 Introduction to Cutting-plane Methods 187
7.2 Regularized Risk Minimization 191
7.3 Multiple Kernel Learning 197

7.4 MAP Inference in Graphical Models 203
7.5 References 214

8 Introduction to Dual Decomposition for Inference
D. Sontag, A. Globerson, and T. Jaakkola **219**
8.1 Introduction 220
8.2 Motivating Applications 222
8.3 Dual Decomposition and Lagrangian Relaxation 224
8.4 Subgradient Algorithms 229
8.5 Block Coordinate Descent Algorithms 232
8.6 Relations to Linear Programming Relaxations 240
8.7 Decoding: Finding the MAP Assignment 242
8.8 Discussion 245
8.10 References 252

9 Augmented Lagrangian Methods for Learning, Selecting, and Combining Features
R. Tomioka, T. Suzuki, and M. Sugiyama **255**
9.1 Introduction 256
9.2 Background 258
9.3 Proximal Minimization Algorithm 263
9.4 Dual Augmented Lagrangian (DAL) Algorithm 265
9.5 Connections 272
9.6 Application 276
9.7 Summary 280
9.9 References 282

10 The Convex Optimization Approach to Regret Minimization
E. Hazan **287**
10.1 Introduction 287
10.2 The RFTL Algorithm and Its Analysis 291
10.3 The "Primal-Dual" Approach 294
10.4 Convexity of Loss Functions 298
10.5 Recent Applications 300
10.6 References 302

11 Projected Newton-type Methods in Machine Learning
M. Schmidt, D. Kim, and S. Sra **305**
11.1 Introduction 305
11.2 Projected Newton-type Methods 306
11.3 Two-Metric Projection Methods 312

11.4 Inexact Projection Methods 316
11.5 Toward Nonsmooth Objectives 320
11.6 Summary and Discussion 326
11.7 References 327

12 Interior-Point Methods in Machine Learning
J. Gondzio **331**
12.1 Introduction 331
12.2 Interior-Point Methods: Background 333
12.3 Polynomial Complexity Result 337
12.4 Interior-Point Methods for Machine Learning 338
12.5 Accelerating Interior-Point Methods 344
12.6 Conclusions 347
12.7 References 347

13 The Tradeoffs of Large-Scale Learning
L. Bottou and O. Bousquet **351**
13.1 Introduction 351
13.2 Approximate Optimization 352
13.3 Asymptotic Analysis 355
13.4 Experiments 363
13.5 Conclusion 366
13.6 References 367

14 Robust Optimization in Machine Learning
C. Caramanis, S. Mannor, and H. Xu **369**
14.1 Introduction 370
14.2 Background on Robust Optimization 371
14.3 Robust Optimization and Adversary Resistant Learning . . . 373
14.4 Robust Optimization and Regularization 377
14.5 Robustness and Consistency 390
14.6 Robustness and Generalization 394
14.7 Conclusion 399
14.8 References 399

15 Improving First and Second-Order Methods by Modeling Uncertainty
N. Le Roux, Y. Bengio, and A. Fitzgibbon **403**
15.1 Introduction 403
15.2 Optimization Versus Learning 404
15.3 Building a Model of the Gradients 406
15.4 The Relative Roles of the Covariance and the Hessian 409

15.5 A Second-Order Model of the Gradients 412
15.6 An Efficient Implementation of Online Consensus Gradient: TONGA 414
15.7 Experiments 419
15.8 Conclusion 427
15.9 References 429

16 Bandit View on Noisy Optimization
J.-Y. Audibert, S. Bubeck, and R. Munos **431**
16.1 Introduction 431
16.2 Concentration Inequalities 433
16.3 Discrete Optimization 434
16.4 Online Optimization 443
16.5 References 452

17 Optimization Methods for Sparse Inverse Covariance Selection
K. Scheinberg and S. Ma **455**
17.1 Introduction 455
17.2 Block Coordinate Descent Methods 461
17.3 Alternating Linearization Method 469
17.4 Remarks on Numerical Performance 475
17.5 References 476

18 A Pathwise Algorithm for Covariance Selection
V. Krishnamurthy, S. D. Ahipaşaoğlu, and A. d'Aspremont **479**
18.1 Introduction 479
18.2 Covariance Selection 481
18.3 Algorithm 482
18.4 Numerical Results 487
18.5 Online Covariance Selection 491
18.6 References 494