

Contents

| | |
|--|------------|
| Preface | vii |
| 1 Elementary Optimization | 1 |
| 1.1 Introduction | 1 |
| 1.2 Univariate Optimization | 1 |
| 1.3 Multivariate Optimization | 7 |
| 1.4 Constrained Optimization | 10 |
| 1.5 Problems | 15 |
| 2 The Seven C's of Analysis | 19 |
| 2.1 Introduction | 19 |
| 2.2 Vector and Matrix Norms | 19 |
| 2.3 Convergence and Completeness | 22 |
| 2.4 The Topology of \mathbb{R}^n | 26 |
| 2.5 Continuous Functions | 30 |
| 2.6 Connectedness | 35 |
| 2.7 Uniform Convergence | 37 |
| 2.8 Problems | 38 |
| 3 Differentiation | 43 |
| 3.1 Introduction | 43 |
| 3.2 Univariate Derivatives | 43 |
| 3.3 The Gauge Integral | 46 |
| 3.4 Higher-Order Derivatives | 47 |

| | | |
|----------|--|------------|
| 3.5 | Partial Derivatives | 50 |
| 3.6 | Differentials | 52 |
| 3.7 | Multivariate Mean Value Theorem | 56 |
| 3.8 | Inverse and Implicit Function Theorems | 58 |
| 3.9 | Differentials of Matrix-Valued Functions | 61 |
| 3.10 | Problems | 65 |
| 4 | Karush-Kuhn-Tucker Theory | 69 |
| 4.1 | Introduction | 69 |
| 4.2 | The Multiplier Rule | 70 |
| 4.3 | Constraint Qualification | 75 |
| 4.4 | Taylor-Made Second Differentials | 78 |
| 4.5 | A Sufficient Condition for a Minimum | 85 |
| 4.6 | Problems | 88 |
| 5 | Convexity | 93 |
| 5.1 | Introduction | 93 |
| 5.2 | Convex Sets | 94 |
| 5.3 | Convex Functions | 95 |
| 5.4 | Minimization of Convex Functions | 104 |
| 5.5 | Moment Inequalities | 109 |
| 5.6 | Problems | 112 |
| 6 | The MM Algorithm | 119 |
| 6.1 | Introduction | 119 |
| 6.2 | Philosophy of the MM Algorithm | 120 |
| 6.3 | Majorization and Minorization | 121 |
| 6.4 | Allele Frequency Estimation | 123 |
| 6.5 | Linear Regression | 125 |
| 6.6 | Bradley-Terry Model of Ranking | 127 |
| 6.7 | Linear Logistic Regression | 127 |
| 6.8 | Poisson Processes | 128 |
| 6.9 | Transmission Tomography | 129 |
| 6.10 | Problems | 133 |
| 7 | The EM Algorithm | 137 |
| 7.1 | Introduction | 137 |
| 7.2 | Definition of the EM Algorithm | 138 |
| 7.3 | Allele Frequency Estimation | 140 |
| 7.4 | Transmission Tomography | 141 |
| 7.5 | Factor Analysis | 143 |
| 7.6 | Hidden Markov Chains | 147 |
| 7.7 | Problems | 149 |

| | | |
|------------------|--|------------|
| 8 | Newton's Method | 155 |
| 8.1 | Introduction | 155 |
| 8.2 | Newton's Method and Root Finding | 156 |
| 8.3 | Newton's Method and Optimization | 158 |
| 8.4 | MM Gradient Algorithm | 160 |
| 8.5 | Ad Hoc Approximations of $d^2 f(\theta)$ | 161 |
| 8.6 | Scoring | 164 |
| 8.7 | Problems | 168 |
| 9 | Conjugate Gradient and Quasi-Newton | 175 |
| 9.1 | Introduction | 175 |
| 9.2 | Centers of Spheres and Centers of Ellipsoids | 176 |
| 9.3 | The Conjugate Gradient Algorithm | 177 |
| 9.4 | Line Search Methods | 180 |
| 9.5 | Quasi-Newton Methods | 182 |
| 9.6 | Trust Regions | 186 |
| 9.7 | Problems | 187 |
| 10 | Analysis of Convergence | 191 |
| 10.1 | Introduction | 191 |
| 10.2 | Local Convergence | 192 |
| 10.3 | Global Convergence of the MM Algorithm | 196 |
| 10.4 | Global Convergence of Gradient Algorithms | 200 |
| 10.5 | Problems | 202 |
| 11 | Convex Programming | 207 |
| 11.1 | Introduction | 207 |
| 11.2 | Adaptive Barrier Methods | 208 |
| 11.3 | Dykstra's Algorithm | 216 |
| 11.4 | Dual Programs | 219 |
| 11.5 | Linear Classification | 223 |
| 11.6 | Support Vector Machines | 226 |
| 11.7 | Problems | 228 |
| Appendix: | The Normal Distribution | 233 |
| A.1 | Univariate Normal Random Variables | 233 |
| A.2 | Multivariate Normal Random Vectors | 235 |
| | References | 237 |
| | Index | 247 |