

## Contents

**Preface** xv

**About the Companion Website** xviii

### **Part I Mathematical Review** 1

#### **1 Methods of Proof and Some Notation** 3

1.1 Methods of Proof 3

1.2 Notation 5

Exercises 5

#### **2 Vector Spaces and Matrices** 7

2.1 Vector and Matrix 7

2.2 Rank of a Matrix 11

2.3 Linear Equations 16

2.4 Inner Products and Norms 18

Exercises 20

#### **3 Transformations** 23

3.1 Linear Transformations 23

3.2 Eigenvalues and Eigenvectors 24

3.3 Orthogonal Projections 26

3.4 Quadratic Forms 27

3.5 Matrix Norms 32

Exercises 35

#### **4 Concepts from Geometry** 39

4.1 Line Segments 39

4.2 Hyperplanes and Linear Varieties 39

4.3 Convex Sets 41

4.4 Neighborhoods 43

4.5 Polytopes and Polyhedra 44

Exercises 45

<b>5</b>	<b>Elements of Calculus</b>	<b>47</b>
5.1	Sequences and Limits	47
5.2	Differentiability	52
5.3	The Derivative Matrix	54
5.4	Differentiation Rules	57
5.5	Level Sets and Gradients	58
5.6	Taylor Series	61
	Exercises	65
	<b>Part II Unconstrained Optimization</b>	<b>67</b>
<b>6</b>	<b>Basics of Set-Constrained and Unconstrained Optimization</b>	<b>69</b>
6.1	Introduction	69
6.2	Conditions for Local Minimizers	70
	Exercises	78
<b>7</b>	<b>One-Dimensional Search Methods</b>	<b>87</b>
7.1	Introduction	87
7.2	Golden Section Search	87
7.3	Fibonacci Method	91
7.4	Bisection Method	97
7.5	Newton's Method	98
7.6	Secant Method	101
7.7	Bracketing	103
7.8	Line Search in Multidimensional Optimization	103
	Exercises	105
<b>8</b>	<b>Gradient Methods</b>	<b>109</b>
8.1	Introduction	109
8.2	Steepest Descent Method	110
8.3	Analysis of Gradient Methods	117
	Exercises	126
<b>9</b>	<b>Newton's Method</b>	<b>133</b>
9.1	Introduction	133
9.2	Analysis of Newton's Method	135
9.3	Levenberg–Marquardt Modification	138
9.4	Newton's Method for Nonlinear Least Squares	139
	Exercises	142
<b>10</b>	<b>Conjugate Direction Methods</b>	<b>145</b>
10.1	Introduction	145
10.2	Conjugate Direction Algorithm	146
10.2.1	Basic Conjugate Direction Algorithm	146
10.3	Conjugate Gradient Algorithm	151
10.4	Conjugate Gradient Algorithm for Nonquadratic Problems	154
	Exercises	156

<b>11</b>	<b>Quasi-Newton Methods</b>	159
11.1	Introduction	159
11.2	Approximating the Inverse Hessian	160
11.3	Rank One Correction Formula	162
11.4	DFP Algorithm	166
11.5	BFGS Algorithm	170
	Exercises	173
<b>12</b>	<b>Solving Linear Equations</b>	179
12.1	Least-Squares Analysis	179
12.2	Recursive Least-Squares Algorithm	187
12.3	Solution to a Linear Equation with Minimum Norm	190
12.4	Kaczmarz's Algorithm	191
12.5	Solving Linear Equations in General	194
	Exercises	201
<b>13</b>	<b>Unconstrained Optimization and Neural Networks</b>	209
13.1	Introduction	209
13.2	Single-Neuron Training	211
13.3	Backpropagation Algorithm	213
	Exercises	222
<b>14</b>	<b>Global Search Algorithms</b>	225
14.1	Introduction	225
14.2	Nelder–Mead Simplex Algorithm	225
14.3	Simulated Annealing	229
14.3.1	Randomized Search	229
14.3.2	Simulated Annealing Algorithm	229
14.4	Particle Swarm Optimization	231
14.4.1	Basic PSO Algorithm	232
14.4.2	Variations	233
14.5	Genetic Algorithms	233
14.5.1	Basic Description	233
14.5.1.1	Chromosomes and Representation Schemes	234
14.5.1.2	Selection and Evolution	234
14.5.2	Analysis of Genetic Algorithms	238
14.5.3	Real-Number Genetic Algorithms	243
	Exercises	244
	<b>Part III Linear Programming</b>	247
<b>15</b>	<b>Introduction to Linear Programming</b>	249
15.1	Brief History of Linear Programming	249
15.2	Simple Examples of Linear Programs	250
15.3	Two-Dimensional Linear Programs	256
15.4	Convex Polyhedra and Linear Programming	258
15.5	Standard Form Linear Programs	260

- 15.6 Basic Solutions 264
- 15.7 Properties of Basic Solutions 267
- 15.8 Geometric View of Linear Programs 269
- Exercises 273
  
- 16 Simplex Method 277**
- 16.1 Solving Linear Equations Using Row Operations 277
- 16.2 The Canonical Augmented Matrix 283
- 16.3 Updating the Augmented Matrix 284
- 16.4 The Simplex Algorithm 285
- 16.5 Matrix Form of the Simplex Method 291
- 16.6 Two-Phase Simplex Method 294
- 16.7 Revised Simplex Method 297
- Exercises 301
  
- 17 Duality 309**
- 17.1 Dual Linear Programs 309
- 17.2 Properties of Dual Problems 316
- 17.3 Matrix Games 321
- Exercises 324
  
- 18 Nonsimplex Methods 331**
- 18.1 Introduction 331
- 18.2 Khachiyan's Method 332
- 18.3 Affine Scaling Method 334
- 18.3.1 Basic Algorithm 334
- 18.3.2 Two-Phase Method 337
- 18.4 Karmarkar's Method 339
- 18.4.1 Basic Ideas 339
- 18.4.2 Karmarkar's Canonical Form 339
- 18.4.3 Karmarkar's Restricted Problem 341
- 18.4.4 From General Form to Karmarkar's Canonical Form 342
- 18.4.5 The Algorithm 345
- Exercises 349
  
- 19 Integer Linear Programming 351**
- 19.1 Introduction 351
- 19.2 Unimodular Matrices 351
- 19.3 The Gomory Cutting-Plane Method 358
- Exercises 366
  
- Part IV Nonlinear Constrained Optimization 369**
  
- 20 Problems with Equality Constraints 371**
- 20.1 Introduction 371
- 20.2 Problem Formulation 373

20.3	Tangent and Normal Spaces	374
20.4	Lagrange Condition	379
20.5	Second-Order Conditions	387
20.6	Minimizing Quadratics Subject to Linear Constraints	390
	Exercises	394
<b>21</b>	<b>Problems with Inequality Constraints</b>	<b>399</b>
21.1	Karush–Kuhn–Tucker Condition	399
21.2	Second-Order Conditions	406
	Exercises	410
<b>22</b>	<b>Convex Optimization Problems</b>	<b>417</b>
22.1	Introduction	417
22.2	Convex Functions	419
22.3	Convex Optimization Problems	426
22.4	Semidefinite Programming	431
22.4.1	Linear Matrix Inequalities and Their Properties	431
22.4.2	LMI Solvers	435
22.4.2.1	Finding a Feasible Solution Under LMI Constraints	436
22.4.2.2	Minimizing a Linear Objective Under LMI Constraints	438
22.4.2.3	Minimizing a Generalized Eigenvalue Under LMI Constraints	440
	Exercises	442
<b>23</b>	<b>Lagrangian Duality</b>	<b>449</b>
23.1	Overview	449
23.2	Notation	449
23.3	Primal–Dual Pair	450
23.4	General Duality Properties	451
23.4.1	Convexity of Dual Problem	451
23.4.2	Primal Objective in Terms of Lagrangian	451
23.4.3	Minimax Inequality Chain	452
23.4.4	Optimality of Saddle Point	452
23.4.5	Weak Duality	453
23.4.6	Duality Gap	453
23.5	Strong Duality	454
23.5.1	Strong Duality $\Leftrightarrow$ Minimax Equals Maximin	454
23.5.2	Strong Duality $\Rightarrow$ Primal Unconstrained Minimization	455
23.5.3	Strong Duality $\Rightarrow$ Optimality	455
23.5.4	Strong Duality $\Rightarrow$ KKT (Including Complementary Slackness)	455
23.5.5	Strong Duality $\Rightarrow$ Saddle Point	456
23.6	Convex Case	456
23.6.1	Convex Case: KKT $\Rightarrow$ Strong Duality	456
23.6.2	Convex Case: Regular Optimal Primal $\Rightarrow$ Strong Duality	457
23.6.3	Convex Case: Slater’s Condition $\Rightarrow$ Strong Duality	457
23.7	Summary of Key Results	457
	Exercises	458

<b>24</b>	<b>Algorithms for Constrained Optimization</b>	<b>459</b>
24.1	Introduction	459
24.2	Projections	459
24.3	Projected Gradient Methods with Linear Constraints	462
24.4	Convergence of Projected Gradient Algorithms	465
24.4.1	Fixed Points and First-Order Necessary Conditions	466
24.4.2	Convergence with Fixed Step Size	468
24.4.3	Some Properties of Projections	469
24.4.4	Armijo Condition	470
24.4.5	Accumulation Points	471
24.4.6	Projections in the Convex Case	472
24.4.7	Armijo Condition in the Convex Case	474
24.4.8	Convergence in the Convex Case	480
24.4.9	Convergence Rate with Line-Search Step Size	481
24.5	Lagrangian Algorithms	483
24.5.1	Lagrangian Algorithm for Equality Constraints	484
24.5.2	Lagrangian Algorithm for Inequality Constraints	486
24.6	Penalty Methods	489
	Exercises	495
<b>25</b>	<b>Multiobjective Optimization</b>	<b>499</b>
25.1	Introduction	499
25.2	Pareto Solutions	499
25.3	Computing the Pareto Front	501
25.4	From Multiobjective to Single-Objective Optimization	505
25.5	Uncertain Linear Programming Problems	508
25.5.1	Uncertain Constraints	508
25.5.2	Uncertain Objective Function Coefficients	511
25.5.3	Uncertain Constraint Coefficients	513
25.5.4	General Uncertainties	513
	Exercises	513
	<b>Part V Optimization in Machine Learning</b>	<b>517</b>
<b>26</b>	<b>Machine Learning Problems and Feature Engineering</b>	<b>519</b>
26.1	Machine Learning Problems	519
26.1.1	Data with Labels and Supervised Learning	519
26.1.2	Data Without Labels and Unsupervised Learning	521
26.2	Data Normalization	522
26.3	Histogram of Oriented Gradients	524
26.4	Principal Component Analysis and Linear Autoencoder	526
26.4.1	Singular Value Decomposition	526
26.4.2	Principal Axes and Principal Components of a Data Set	527
26.4.3	Linear Autoencoder	529
	Exercises	530

<b>27</b>	<b>Stochastic Gradient Descent Algorithms</b>	537
27.1	Stochastic Gradient Descent Algorithm	537
27.2	Stochastic Variance Reduced Gradient Algorithm	540
27.3	Distributed Stochastic Variance Reduced Gradient	542
27.3.1	Distributed Learning Environment	542
27.3.2	SVRG in Distributed Optimization	543
27.3.3	Communication Versus Computation	545
27.3.4	Data Security	545
	Exercises	546
<b>28</b>	<b>Linear Regression and Its Variants</b>	553
28.1	Least-Squares Linear Regression	553
28.1.1	A Linear Model for Prediction	553
28.1.2	Training the Model	554
28.1.3	Computing Optimal $\hat{w}$	554
28.1.4	Optimal Predictor and Performance Evaluation	555
28.1.5	Least-Squares Linear Regression for Data Sets with Vector Labels	556
28.2	Model Selection by Cross-Validation	559
28.3	Model Selection by Regularization	562
	Exercises	564
<b>29</b>	<b>Logistic Regression for Classification</b>	569
29.1	Logistic Regression for Binary Classification	569
29.1.1	Least-Squares Linear Regression for Binary Classification	569
29.1.2	Logistic Regression for Binary Classification	570
29.1.3	Interpreting Logistic Regression by Log Error	572
29.1.4	Confusion Matrix for Binary Classification	573
29.2	Nonlinear Decision Boundary via Linear Regression	575
29.2.1	Least-Squares Linear Regression with Nonlinear Transformation	576
29.2.2	Logistic Regression with Nonlinear Transformation	578
29.3	Multicategory Classification	580
29.3.1	One-Versus-All Multicategory Classification	580
29.3.2	Softmax Regression for Multicategory Classification	581
	Exercises	584
<b>30</b>	<b>Support Vector Machines</b>	589
30.1	Hinge-Loss Functions	589
30.1.1	Geometric Interpretation of the Linear Model	589
30.1.2	Hinge Loss for Binary Data Sets	590
30.1.3	Hinge Loss for Multicategory Data Sets	592
30.2	Classification by Minimizing Hinge Loss	593
30.2.1	Binary Classification by Minimizing Average Hinge Loss	593
30.2.2	Multicategory Classification by Minimizing $E_{\text{hww}}$ or $E_{\text{hcs}}$	594
30.3	Support Vector Machines for Binary Classification	596
30.3.1	Hard-Margin Support Vector Machines	596
30.3.2	Support Vectors	598
30.3.3	Soft-Margin Support Vector Machines	599

30.3.4	Connection to Hinge-Loss Minimization	602
30.4	Support Vector Machines for Multicategory Classification	602
30.5	Kernel Trick	603
30.5.1	Kernels	603
30.5.2	Kernel Trick	604
30.5.3	Learning with Kernels	605
30.5.3.1	Regularized Logistic Regression with Nonlinear Transformation for Binary Classification	605
30.5.3.2	Regularized Hinge-Loss Minimization for Binary Classification	606
	Exercises	607
<b>31</b>	<b><i>K</i>-Means Clustering</b>	<b>611</b>
31.1	<i>K</i> -Means Clustering	611
31.2	<i>K</i> -Means++ for Center Initialization	615
31.3	Variants of <i>K</i> -Means Clustering	617
31.3.1	<i>K</i> -Means Clustering Based on 1-Norm Regularization	617
31.3.2	PCA-Guided <i>K</i> -Means Clustering	619
31.4	Image Compression by Vector Quantization and <i>K</i> -Means Clustering	622
	Exercises	623
	<b>References</b>	<b>627</b>
	<b>Index</b>	<b>635</b>