

# Contents

<b>Preface to the Second Edition</b>	<b>vii</b>
<b>Preface to the First Edition</b>	<b>ix</b>
<b>1 Elementary Optimization</b>	<b>1</b>
1.1 Introduction . . . . .	1
1.2 Univariate Optimization . . . . .	1
1.3 Multivariate Optimization . . . . .	7
1.4 Constrained Optimization . . . . .	10
1.5 Problems . . . . .	17
<b>2 The Seven C's of Analysis</b>	<b>23</b>
2.1 Introduction . . . . .	23
2.2 Vector and Matrix Norms . . . . .	23
2.3 Convergence and Completeness . . . . .	26
2.4 The Topology of $\mathbb{R}^n$ . . . . .	30
2.5 Continuous Functions . . . . .	34
2.6 Semicontinuity . . . . .	42
2.7 Connectedness . . . . .	44
2.8 Uniform Convergence . . . . .	46
2.9 Problems . . . . .	47
<b>3 The Gauge Integral</b>	<b>53</b>
3.1 Introduction . . . . .	53
3.2 Gauge Functions and $\delta$ -Fine Partitions . . . . .	54

3.3	Definition and Basic Properties of the Integral . . . . .	57	8.8	Geometric and Signomial Programs . . . . .	194
3.4	The Fundamental Theorem of Calculus . . . . .	62	8.9	Poisson Processes . . . . .	197
3.5	More Advanced Topics in Integration . . . . .	66	8.10	Transmission Tomography . . . . .	198
3.6	Problems . . . . .	71	8.11	Poisson Multigraphs . . . . .	202
<b>4</b>	<b>Differentiation</b>	<b>75</b>	8.12	Problems . . . . .	204
4.1	Introduction . . . . .	75	<b>9</b>	<b>The EM Algorithm</b>	<b>221</b>
4.2	Univariate Derivatives . . . . .	75	9.1	Introduction . . . . .	221
4.3	Partial Derivatives . . . . .	79	9.2	Definition of the EM Algorithm . . . . .	222
4.4	Differentials . . . . .	81	9.3	Missing Data in the Ordinary Sense . . . . .	224
4.5	Multivariate Mean Value Theorem . . . . .	88	9.4	Allele Frequency Estimation . . . . .	225
4.6	Inverse and Implicit Function Theorems . . . . .	89	9.5	Clustering by EM . . . . .	226
4.7	Differentials of Matrix-Valued Functions . . . . .	93	9.6	Transmission Tomography . . . . .	228
4.8	Problems . . . . .	98	9.7	Factor Analysis . . . . .	230
<b>5</b>	<b>Karush-Kuhn-Tucker Theory</b>	<b>107</b>	9.8	Hidden Markov Chains . . . . .	234
5.1	Introduction . . . . .	107	9.9	Problems . . . . .	236
5.2	The Multiplier Rule . . . . .	108	<b>10</b>	<b>Newton's Method and Scoring</b>	<b>245</b>
5.3	Constraint Qualification . . . . .	114	10.1	Introduction . . . . .	245
5.4	Taylor-Made Higher-Order Differentials . . . . .	117	10.2	Newton's Method and Root Finding . . . . .	246
5.5	Applications of Second Differentials . . . . .	123	10.3	Newton's Method and Optimization . . . . .	248
5.6	Problems . . . . .	128	10.4	MM Gradient Algorithm . . . . .	250
<b>6</b>	<b>Convexity</b>	<b>137</b>	10.5	Ad Hoc Approximations of $d^2f(\theta)$ . . . . .	252
6.1	Introduction . . . . .	137	10.6	Scoring and Exponential Families . . . . .	254
6.2	Convex Sets . . . . .	138	10.7	The Gauss-Newton Algorithm . . . . .	257
6.3	Convex Functions . . . . .	142	10.8	Generalized Linear Models . . . . .	258
6.4	Continuity, Differentiability, and Integrability . . . . .	149	10.9	Accelerated MM . . . . .	259
6.5	Minimization of Convex Functions . . . . .	152	10.10	Problems . . . . .	262
6.6	Moment Inequalities . . . . .	159	<b>11</b>	<b>Conjugate Gradient and Quasi-Newton</b>	<b>273</b>
6.7	Problems . . . . .	162	11.1	Introduction . . . . .	273
<b>7</b>	<b>Block Relaxation</b>	<b>171</b>	11.2	Centers of Spheres and Centers of Ellipsoids . . . . .	274
7.1	Introduction . . . . .	171	11.3	The Conjugate Gradient Algorithm . . . . .	275
7.2	Examples of Block Relaxation . . . . .	172	11.4	Line Search Methods . . . . .	278
7.3	Problems . . . . .	180	11.5	Stopping Criteria . . . . .	280
<b>8</b>	<b>The MM Algorithm</b>	<b>185</b>	11.6	Quasi-Newton Methods . . . . .	281
8.1	Introduction . . . . .	185	11.7	Trust Regions . . . . .	285
8.2	Philosophy of the MM Algorithm . . . . .	186	11.8	Problems . . . . .	286
8.3	Majorization and Minorization . . . . .	187	<b>12</b>	<b>Analysis of Convergence</b>	<b>291</b>
8.4	Allele Frequency Estimation . . . . .	189	12.1	Introduction . . . . .	291
8.5	Linear Regression . . . . .	191	12.2	Local Convergence . . . . .	292
8.6	Bradley-Terry Model of Ranking . . . . .	193	12.3	Coercive Functions . . . . .	297
8.7	Linear Logistic Regression . . . . .	194	12.4	Global Convergence of the MM Algorithm . . . . .	299
			12.5	Global Convergence of Block Relaxation . . . . .	302

12.6	Global Convergence of Gradient Algorithms . . . . .	303	17.3	Normed Vector Spaces . . . . .	448
12.7	Problems . . . . .	306	17.4	Linear Operators and Functionals . . . . .	451
<b>13</b>	<b>Penalty and Barrier Methods</b>	<b>313</b>	17.5	Differentials . . . . .	453
13.1	Introduction . . . . .	313	17.6	The Euler-Lagrange Equation . . . . .	456
13.2	Rudiments of Barrier and Penalty Methods . . . . .	314	17.7	Applications of the Euler-Lagrange Equation . . . . .	459
13.3	An Adaptive Barrier Method . . . . .	318	17.8	Lagrange's Lacuna . . . . .	462
13.4	Imposition of a Prior in EM Clustering . . . . .	325	17.9	Variational Problems with Constraints . . . . .	464
13.5	Model Selection and the Lasso . . . . .	327	17.10	Natural Cubic Splines . . . . .	466
13.6	Lasso Penalized $\ell_1$ Regression . . . . .	329	17.11	Problems . . . . .	467
13.7	Lasso Penalized $\ell_2$ Regression . . . . .	330	<b>Appendix: Mathematical Notes</b>	<b>473</b>	
13.8	Penalized Discriminant Analysis . . . . .	333	A.1	Univariate Normal Random Variables . . . . .	473
13.9	Problems . . . . .	334	A.2	Multivariate Normal Random Vectors . . . . .	475
<b>14</b>	<b>Convex Calculus</b>	<b>341</b>	A.3	Polyhedral Sets . . . . .	477
14.1	Introduction . . . . .	341	A.4	Birkhoff's Theorem and Fan's Inequality . . . . .	480
14.2	Notation . . . . .	342	A.5	Singular Value Decomposition . . . . .	485
14.3	Fenchel Conjugates . . . . .	342	A.6	Hadamard Semidifferentials . . . . .	487
14.4	Subdifferentials . . . . .	351	A.7	Problems . . . . .	497
14.5	The Rules of Convex Differentiation . . . . .	358	<b>References</b>	<b>499</b>	
14.6	Spectral Functions . . . . .	365	<b>Index</b>	<b>519</b>	
14.7	A Convex Lagrange Multiplier Rule . . . . .	372			
14.8	Problems . . . . .	375			
<b>15</b>	<b>Feasibility and Duality</b>	<b>383</b>			
15.1	Introduction . . . . .	383			
15.2	Dykstra's Algorithm . . . . .	384			
15.3	Contractive Maps . . . . .	389			
15.4	Dual Functions . . . . .	393			
15.5	Examples of Dual Programs . . . . .	396			
15.6	Practical Applications of Duality . . . . .	402			
15.7	Problems . . . . .	406			
<b>16</b>	<b>Convex Minimization Algorithms</b>	<b>415</b>			
16.1	Introduction . . . . .	415			
16.2	Projected Gradient Algorithm . . . . .	416			
16.3	Exact Penalties and Lagrangians . . . . .	421			
16.4	Mechanics of Path Following . . . . .	426			
16.5	Bregman Iteration . . . . .	432			
16.6	Split Bregman Iteration . . . . .	436			
16.7	Convergence of Bregman Iteration . . . . .	439			
16.8	Problems . . . . .	440			
<b>17</b>	<b>The Calculus of Variations</b>	<b>445</b>			
17.1	Introduction . . . . .	445			
17.2	Classical Problems . . . . .	446			