## Contents

*Acknowledgments* xxiii

*List of Figures* xxv

*List of Algorithms* xxxi

*List of Boxes* xxxiii

1 **Introduction** 1

1.1 Motivation 1

1.2 Structured Probabilistic Models 2  
1.2.1 Probabilistic Graphical Models 3  
1.2.2 Representation, Inference, Learning 5

1.3 Overview and Roadmap 6  
1.3.1 Overview of Chapters 6  
1.3.2 Reader's Guide 9  
1.3.3 Connection to Other Disciplines 11

1.4 Historical Notes 12

2 **Foundations** 15

2.1 Probability Theory 15  
2.1.1 Probability Distributions 15  
2.1.2 Basic Concepts in Probability 18  
2.1.3 Random Variables and Joint Distributions 19  
2.1.4 Independence and Conditional Independence 23  
2.1.5 Querying a Distribution 25  
2.1.6 Continuous Spaces 27  
2.1.7 Expectation and Variance 31

2.2 Graphs 34  
2.2.1 Nodes and Edges 34  
2.2.2 Subgraphs 35  
2.2.3 Paths and Trails 36
7 Gaussian Network Models 247
7.1 Multivariate Gaussians 247
7.1.1 Basic Parameterization 247
7.1.2 Operations on Gaussians 249
7.1.3 Independencies in Gaussians 250
7.2 Gaussian Bayesian Networks 251
7.3 Gaussian Markov Random Fields 254
7.4 Summary 257
7.5 Relevant Literature 258
7.6 Exercises 258

8 The Exponential Family 261
8.1 Introduction 261
8.2 Exponential Families 261
8.2.1 Linear Exponential Families 263
8.3 Factored Exponential Families 266
8.3.1 Product Distributions 266
8.3.2 Bayesian Networks 267
8.4 Entropy and Relative Entropy 269
8.4.1 Entropy 269
8.4.2 Relative Entropy 272
8.5 Projections 273
8.5.1 Comparison 274
8.5.2 M-Projections 277
8.5.3 I-Projections 282
8.6 Summary 282
8.7 Relevant Literature 283
8.8 Exercises 283

II Inference 285

9 Variable Elimination 287
9.1 Analysis of Complexity 288
9.1.1 Analysis of Exact Inference 288
9.1.2 Analysis of Approximate Inference 290
9.2 Variable Elimination: The Basic Ideas 292
9.3 Variable Elimination 296
9.3.1 Basic Elimination 297
9.3.2 Dealing with Evidence 303
9.4 Complexity and Graph Structure: Variable Elimination 306
9.4.1 Simple Analysis 306
9.4.2 Graph-Theoretic Analysis 306
9.4.3 Finding Elimination Orderings ★ 310
9.5 Conditioning ★ 315
CONTENTS

9.5.1 The Conditioning Algorithm 315
9.5.2 Conditioning and Variable Elimination 318
9.5.3 Graph-Theoretic Analysis 322
9.5.4 Improved Conditioning 323

9.6 Inference with Structured CPDs 325
9.6.1 Independence of Causal Influence 325
9.6.2 Context-Specific Independence 329
9.6.3 Discussion 335

9.7 Summary and Discussion 336
9.8 Relevant Literature 337
9.9 Exercises 338

10 Clique Trees 345
10.1 Variable Elimination and Clique Trees 345
10.1.1 Cluster Graphs 346
10.1.2 Clique Trees 346
10.2 Message Passing: Sum Product 348
10.2.1 Variable Elimination in a Clique Tree 349
10.2.2 Clique Tree Calibration 355
10.2.3 A Calibrated Clique Tree as a Distribution 361
10.3 Message Passing: Belief Update 364
10.3.1 Message Passing with Division 364
10.3.2 Equivalence of Sum-Product and Belief Update Messages 368
10.3.3 Answering Queries 369
10.4 Constructing a Clique Tree 372
10.4.1 Clique Trees from Variable Elimination 372
10.4.2 Clique Trees from Chordal Graphs 374

10.5 Summary 376
10.6 Relevant Literature 377
10.7 Exercises 378

11 Inference as Optimization 381
11.1 Introduction 381
11.1.1 Exact Inference Revisited 382
11.1.2 The Energy Functional 384
11.1.3 Optimizing the Energy Functional 386
11.2 Exact Inference as Optimization 386
11.2.1 Fixed-Point Characterization 388
11.2.2 Inference as Optimization 390
11.3 Propagation-Based Approximation 391
11.3.1 A Simple Example 391
11.3.2 Cluster-Graph Belief Propagation 396
11.3.3 Properties of Cluster-Graph Belief Propagation 399
11.3.4 Analyzing Convergence 401
11.3.5 Constructing Cluster Graphs 404
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Start Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.3.6</td>
<td>Variational Analysis</td>
<td>411</td>
</tr>
<tr>
<td>11.3.7</td>
<td>Other Entropy Approximations</td>
<td>414</td>
</tr>
<tr>
<td>11.3.8</td>
<td>Discussion</td>
<td>428</td>
</tr>
<tr>
<td>11.4</td>
<td>Propagation with Approximate Messages</td>
<td>430</td>
</tr>
<tr>
<td>11.4.1</td>
<td>Factorized Messages</td>
<td>431</td>
</tr>
<tr>
<td>11.4.2</td>
<td>Approximate Message Computation</td>
<td>433</td>
</tr>
<tr>
<td>11.4.3</td>
<td>Inference with Approximate Messages</td>
<td>436</td>
</tr>
<tr>
<td>11.4.4</td>
<td>Expectation Propagation</td>
<td>442</td>
</tr>
<tr>
<td>11.4.5</td>
<td>Variational Analysis</td>
<td>445</td>
</tr>
<tr>
<td>11.4.6</td>
<td>Discussion</td>
<td>448</td>
</tr>
<tr>
<td>11.5</td>
<td>Structured Variational Approximations</td>
<td>448</td>
</tr>
<tr>
<td>11.5.1</td>
<td>The Mean Field Approximation</td>
<td>449</td>
</tr>
<tr>
<td>11.5.2</td>
<td>Structured Approximations</td>
<td>456</td>
</tr>
<tr>
<td>11.5.3</td>
<td>Local Variational Methods</td>
<td>469</td>
</tr>
<tr>
<td>11.6</td>
<td>Summary and Discussion</td>
<td>473</td>
</tr>
<tr>
<td>11.7</td>
<td>Relevant Literature</td>
<td>475</td>
</tr>
<tr>
<td>11.8</td>
<td>Exercises</td>
<td>477</td>
</tr>
<tr>
<td>12</td>
<td>Particle-Based Approximate Inference</td>
<td>487</td>
</tr>
<tr>
<td>12.1</td>
<td>Forward Sampling</td>
<td>488</td>
</tr>
<tr>
<td>12.1.1</td>
<td>Sampling from a Bayesian Network</td>
<td>488</td>
</tr>
<tr>
<td>12.1.2</td>
<td>Analysis of Error</td>
<td>490</td>
</tr>
<tr>
<td>12.1.3</td>
<td>Conditional Probability Queries</td>
<td>491</td>
</tr>
<tr>
<td>12.2</td>
<td>Likelihood Weighting and Importance Sampling</td>
<td>492</td>
</tr>
<tr>
<td>12.2.1</td>
<td>Likelihood Weighting: Intuition</td>
<td>492</td>
</tr>
<tr>
<td>12.2.2</td>
<td>Importance Sampling</td>
<td>494</td>
</tr>
<tr>
<td>12.2.3</td>
<td>Importance Sampling for Bayesian Networks</td>
<td>498</td>
</tr>
<tr>
<td>12.2.4</td>
<td>Importance Sampling Revisited</td>
<td>504</td>
</tr>
<tr>
<td>12.3</td>
<td>Markov Chain Monte Carlo Methods</td>
<td>505</td>
</tr>
<tr>
<td>12.3.1</td>
<td>Gibbs Sampling Algorithm</td>
<td>505</td>
</tr>
<tr>
<td>12.3.2</td>
<td>Markov Chains</td>
<td>507</td>
</tr>
<tr>
<td>12.3.3</td>
<td>Gibbs Sampling Revisited</td>
<td>512</td>
</tr>
<tr>
<td>12.3.4</td>
<td>A Broader Class of Markov Chains</td>
<td>515</td>
</tr>
<tr>
<td>12.3.5</td>
<td>Using a Markov Chain</td>
<td>518</td>
</tr>
<tr>
<td>12.4</td>
<td>Collapsed Particles</td>
<td>526</td>
</tr>
<tr>
<td>12.4.1</td>
<td>Collapsed Likelihood Weighting</td>
<td>527</td>
</tr>
<tr>
<td>12.4.2</td>
<td>Collapsed MCMC</td>
<td>531</td>
</tr>
<tr>
<td>12.5</td>
<td>Deterministic Search Methods</td>
<td>536</td>
</tr>
<tr>
<td>12.6</td>
<td>Summary</td>
<td>540</td>
</tr>
<tr>
<td>12.7</td>
<td>Relevant Literature</td>
<td>541</td>
</tr>
<tr>
<td>12.8</td>
<td>Exercises</td>
<td>544</td>
</tr>
<tr>
<td>13</td>
<td>MAP Inference</td>
<td>551</td>
</tr>
<tr>
<td>13.1</td>
<td>Overview</td>
<td>551</td>
</tr>
<tr>
<td>13.1.1</td>
<td>Computational Complexity</td>
<td>551</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>-----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>13.1.2</td>
<td>Overview of Solution Methods</td>
<td>552</td>
</tr>
<tr>
<td>13.2</td>
<td>Variable Elimination for (Marginal) MAP</td>
<td>554</td>
</tr>
<tr>
<td>13.2.1</td>
<td>Max-Product Variable Elimination</td>
<td>554</td>
</tr>
<tr>
<td>13.2.2</td>
<td>Finding the Most Probable Assignment</td>
<td>556</td>
</tr>
<tr>
<td>13.2.3</td>
<td>Variable Elimination for Marginal MAP</td>
<td>559</td>
</tr>
<tr>
<td>13.3</td>
<td>Max-Product in Clique Trees</td>
<td>562</td>
</tr>
<tr>
<td>13.3.1</td>
<td>Computing Max-Marginals</td>
<td>562</td>
</tr>
<tr>
<td>13.3.2</td>
<td>Message Passing as Reparameterization</td>
<td>564</td>
</tr>
<tr>
<td>13.3.3</td>
<td>Decoding Max-Marginals</td>
<td>565</td>
</tr>
<tr>
<td>13.4</td>
<td>Max-Product Belief Propagation in Loopy Cluster Graphs</td>
<td>567</td>
</tr>
<tr>
<td>13.4.1</td>
<td>Standard Max-Product Message Passing</td>
<td>567</td>
</tr>
<tr>
<td>13.4.2</td>
<td>Max-Product BP with Counting Numbers</td>
<td>572</td>
</tr>
<tr>
<td>13.4.3</td>
<td>Discussion</td>
<td>575</td>
</tr>
<tr>
<td>13.5</td>
<td>MAP as a Linear Optimization Problem</td>
<td>577</td>
</tr>
<tr>
<td>13.5.1</td>
<td>The Integer Program Formulation</td>
<td>577</td>
</tr>
<tr>
<td>13.5.2</td>
<td>Linear Programming Relaxation</td>
<td>579</td>
</tr>
<tr>
<td>13.5.3</td>
<td>Low-Temperature Limits</td>
<td>581</td>
</tr>
<tr>
<td>13.6</td>
<td>Using Graph Cuts for MAP</td>
<td>588</td>
</tr>
<tr>
<td>13.6.1</td>
<td>Inference Using Graph Cuts</td>
<td>588</td>
</tr>
<tr>
<td>13.6.2</td>
<td>Nonbinary Variables</td>
<td>592</td>
</tr>
<tr>
<td>13.7</td>
<td>Local Search Algorithms</td>
<td>595</td>
</tr>
<tr>
<td>13.8</td>
<td>Summary</td>
<td>597</td>
</tr>
<tr>
<td>13.9</td>
<td>Relevant Literature</td>
<td>598</td>
</tr>
<tr>
<td>13.10</td>
<td>Exercises</td>
<td>601</td>
</tr>
</tbody>
</table>

**14 Inference in Hybrid Networks**  

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>14.1</td>
<td>Introduction</td>
<td>605</td>
</tr>
<tr>
<td>14.1.1</td>
<td>Challenges</td>
<td>605</td>
</tr>
<tr>
<td>14.1.2</td>
<td>Discretization</td>
<td>606</td>
</tr>
<tr>
<td>14.1.3</td>
<td>Overview</td>
<td>607</td>
</tr>
<tr>
<td>14.2</td>
<td>Variable Elimination in Gaussian Networks</td>
<td>608</td>
</tr>
<tr>
<td>14.2.1</td>
<td>Canonical Forms</td>
<td>609</td>
</tr>
<tr>
<td>14.2.2</td>
<td>Sum-Product Algorithms</td>
<td>611</td>
</tr>
<tr>
<td>14.2.3</td>
<td>Gaussian Belief Propagation</td>
<td>612</td>
</tr>
<tr>
<td>14.3</td>
<td>Hybrid Networks</td>
<td>615</td>
</tr>
<tr>
<td>14.3.1</td>
<td>The Difficulties</td>
<td>615</td>
</tr>
<tr>
<td>14.3.2</td>
<td>Factor Operations for Hybrid Gaussian Networks</td>
<td>618</td>
</tr>
<tr>
<td>14.3.3</td>
<td>EP for CLG Networks</td>
<td>621</td>
</tr>
<tr>
<td>14.3.4</td>
<td>An “Exact” CLG Algorithm</td>
<td>626</td>
</tr>
<tr>
<td>14.4</td>
<td>Nonlinear Dependencies</td>
<td>630</td>
</tr>
<tr>
<td>14.4.1</td>
<td>Linearization</td>
<td>631</td>
</tr>
<tr>
<td>14.4.2</td>
<td>Expectation Propagation with Gaussian Approximation</td>
<td>637</td>
</tr>
<tr>
<td>14.5</td>
<td>Particle-Based Approximation Methods</td>
<td>642</td>
</tr>
<tr>
<td>14.5.1</td>
<td>Sampling in Continuous Spaces</td>
<td>642</td>
</tr>
<tr>
<td>14.5.2</td>
<td>Forward Sampling in Bayesian Networks</td>
<td>643</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>17.1.1</td>
<td>The Thumbtack Example</td>
<td>717</td>
</tr>
<tr>
<td>17.1.2</td>
<td>The Maximum Likelihood Principle</td>
<td>720</td>
</tr>
<tr>
<td>17.2</td>
<td>MLE for Bayesian Networks</td>
<td>722</td>
</tr>
<tr>
<td>17.2.1</td>
<td>A Simple Example</td>
<td>723</td>
</tr>
<tr>
<td>17.2.2</td>
<td>Global Likelihood Decomposition</td>
<td>724</td>
</tr>
<tr>
<td>17.2.3</td>
<td>Table-CPDs</td>
<td>725</td>
</tr>
<tr>
<td>17.2.4</td>
<td>Gaussian Bayesian Networks</td>
<td>728</td>
</tr>
<tr>
<td>17.2.5</td>
<td>Maximum Likelihood Estimation as M-Projection</td>
<td>731</td>
</tr>
<tr>
<td>17.3</td>
<td>Bayesian Parameter Estimation</td>
<td>733</td>
</tr>
<tr>
<td>17.3.1</td>
<td>The Thumbtack Example Revisited</td>
<td>733</td>
</tr>
<tr>
<td>17.3.2</td>
<td>Priors and Posteriors</td>
<td>737</td>
</tr>
<tr>
<td>17.4</td>
<td>Bayesian Parameter Estimation in Bayesian Networks</td>
<td>741</td>
</tr>
<tr>
<td>17.4.1</td>
<td>Parameter Independence and Global Decomposition</td>
<td>742</td>
</tr>
<tr>
<td>17.4.2</td>
<td>Local Decomposition</td>
<td>746</td>
</tr>
<tr>
<td>17.4.3</td>
<td>Priors for Bayesian Network Learning</td>
<td>748</td>
</tr>
<tr>
<td>17.4.4</td>
<td>MAP Estimation</td>
<td>751</td>
</tr>
<tr>
<td>17.5</td>
<td>Learning Models with Shared Parameters</td>
<td>754</td>
</tr>
<tr>
<td>17.5.1</td>
<td>Global Parameter Sharing</td>
<td>755</td>
</tr>
<tr>
<td>17.5.2</td>
<td>Local Parameter Sharing</td>
<td>760</td>
</tr>
<tr>
<td>17.5.3</td>
<td>Bayesian Inference with Shared Parameters</td>
<td>762</td>
</tr>
<tr>
<td>17.5.4</td>
<td>Hierarchical Priors</td>
<td>763</td>
</tr>
<tr>
<td>17.6</td>
<td>Generalization Analysis</td>
<td>769</td>
</tr>
<tr>
<td>17.6.1</td>
<td>Asymptotic Analysis</td>
<td>769</td>
</tr>
<tr>
<td>17.6.2</td>
<td>PAC-Bounds</td>
<td>770</td>
</tr>
<tr>
<td>17.7</td>
<td>Summary</td>
<td>776</td>
</tr>
<tr>
<td>17.8</td>
<td>Relevant Literature</td>
<td>777</td>
</tr>
<tr>
<td>17.9</td>
<td>Exercises</td>
<td>778</td>
</tr>
</tbody>
</table>

### 18 Structure Learning in Bayesian Networks | 783

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>18.1</td>
<td>Introduction</td>
<td>783</td>
</tr>
<tr>
<td>18.1.1</td>
<td>Problem Definition</td>
<td>783</td>
</tr>
<tr>
<td>18.1.2</td>
<td>Overview of Methods</td>
<td>785</td>
</tr>
<tr>
<td>18.2</td>
<td>Constraint-Based Approaches</td>
<td>786</td>
</tr>
<tr>
<td>18.2.1</td>
<td>General Framework</td>
<td>786</td>
</tr>
<tr>
<td>18.2.2</td>
<td>Independence Tests</td>
<td>787</td>
</tr>
<tr>
<td>18.3</td>
<td>Structure Scores</td>
<td>790</td>
</tr>
<tr>
<td>18.3.1</td>
<td>Likelihood Scores</td>
<td>791</td>
</tr>
<tr>
<td>18.3.2</td>
<td>Bayesian Score</td>
<td>794</td>
</tr>
<tr>
<td>18.3.3</td>
<td>Marginal Likelihood for a Single Variable</td>
<td>797</td>
</tr>
<tr>
<td>18.3.4</td>
<td>Bayesian Score for Bayesian Networks</td>
<td>799</td>
</tr>
<tr>
<td>18.3.5</td>
<td>Understanding the Bayesian Score</td>
<td>801</td>
</tr>
<tr>
<td>18.3.6</td>
<td>Priors</td>
<td>804</td>
</tr>
<tr>
<td>18.3.7</td>
<td>Score Equivalence</td>
<td>807</td>
</tr>
<tr>
<td>18.4</td>
<td>Structure Search</td>
<td>807</td>
</tr>
<tr>
<td>18.4.1</td>
<td>Learning Tree-Structured Networks</td>
<td>808</td>
</tr>
</tbody>
</table>
CONTENTS

18.4.2 Known Order 809
18.4.3 General Graphs 811
18.4.4 Learning with Equivalence Classes * 821
18.5 Bayesian Model Averaging * 824
18.5.1 Basic Theory 824
18.5.2 Model Averaging Given an Order 826
18.5.3 The General Case 828
18.6 Learning Models with Additional Structure 832
18.6.1 Learning with Local Structure 833
18.6.2 Learning Template Models 837
18.7 Summary and Discussion 838
18.8 Relevant Literature 840
18.9 Exercises 843

19 Partially Observed Data 849
19.1 Foundations 849
19.1.1 Likelihood of Data and Observation Models 849
19.1.2 Decoupling of Observation Mechanism 853
19.1.3 The Likelihood Function 856
19.1.4 Identifiability 860
19.2 Parameter Estimation 862
19.2.1 Gradient Ascent 863
19.2.2 Expectation Maximization (EM) 868
19.2.3 Comparison: Gradient Ascent versus EM 887
19.2.4 Approximate Inference * 893
19.3 Bayesian Learning with Incomplete Data * 897
19.3.1 Overview 897
19.3.2 MCMC Sampling 899
19.3.3 Variational Bayesian Learning 904
19.4 Structure Learning 908
19.4.1 Scoring Structures 909
19.4.2 Structure Search 917
19.4.3 Structural EM 920
19.5 Learning Models with Hidden Variables 925
19.5.1 Information Content of Hidden Variables 926
19.5.2 Determining the Cardinality 928
19.5.3 Introducing Hidden Variables 930
19.6 Summary 933
19.7 Relevant Literature 934
19.8 Exercises 935

20 Learning Undirected Models 943
20.1 Overview 943
20.2 The Likelihood Function 944
20.2.1 An Example 944
## CONTENTS

20.2.2 Form of the Likelihood Function 946
20.2.3 Properties of the Likelihood Function 947

20.3 Maximum (Conditional) Likelihood Parameter Estimation 949
20.3.1 Maximum Likelihood Estimation 949
20.3.2 Conditionally Trained Models 950
20.3.3 Learning with Missing Data 954
20.3.4 Maximum Entropy and Maximum Likelihood 956

20.4 Parameter Priors and Regularization 958
20.4.1 Local Priors 958
20.4.2 Global Priors 961

20.5 Learning with Approximate Inference 961
20.5.1 Belief Propagation 962
20.5.2 MAP-Based Learning 967

20.6 Alternative Objectives 969
20.6.1 Pseudolikelihood and Its Generalizations 970
20.6.2 Contrastive Optimization Criteria 974

20.7 Structure Learning 978
20.7.1 Structure Learning Using Independence Tests 979
20.7.2 Score-Based Learning: Hypothesis Spaces 981
20.7.3 Objective Functions 982
20.7.4 Optimization Task 985
20.7.5 Evaluating Changes to the Model 992

20.8 Summary 996
20.9 Relevant Literature 998
20.10 Exercises 1001

## IV Actions and Decisions 1007

21 Causality 1009

21.1 Motivation and Overview 1009
21.1.1 Conditioning and Intervention 1009
21.1.2 Correlation and Causation 1012

21.2 Causal Models 1014

21.3 Structural Causal Identifiability 1017
21.3.1 Query Simplification Rules 1017
21.3.2 Iterated Query Simplification 1020

21.4 Mechanisms and Response Variables 1026
21.5 Partial Identifiability in Functional Causal Models 1031
21.6 Counterfactual Queries 1034
21.6.1 Twinned Networks 1034
21.6.2 Bounds on Counterfactual Queries 1037

21.7 Learning Causal Models 1039
21.7.1 Learning Causal Models without Confounding Factors 1040
21.7.2 Learning from Interventional Data 1043
## CONTENTS

23.7 Value of Information 1119
   23.7.1 Single Observations 1120
   23.7.2 Multiple Observations 1122

23.8 Summary 1124

23.9 Relevant Literature 1125

23.10 Exercises 1128

### 24 Epilogue 1131

## A Background Material 1135

A.1 Information Theory 1135
   A.1.1 Compression and Entropy 1135
   A.1.2 Conditional Entropy and Information 1137
   A.1.3 Relative Entropy and Distances Between Distributions 1138

A.2 Convergence Bounds 1141
   A.2.1 Central Limit Theorem 1142
   A.2.2 Convergence Bounds 1143

A.3 Algorithms and Algorithmic Complexity 1144
   A.3.1 Basic Graph Algorithms 1144
   A.3.2 Analysis of Algorithmic Complexity 1145
   A.3.3 Dynamic Programming 1147
   A.3.4 Complexity Theory 1148

A.4 Combinatorial Optimization and Search 1152
   A.4.1 Optimization Problems 1152
   A.4.2 Local Search 1152
   A.4.3 Branch and Bound Search 1158

A.5 Continuous Optimization 1159
   A.5.1 Characterizing Optima of a Continuous Function 1159
   A.5.2 Gradient Ascent Methods 1161
   A.5.3 Constrained Optimization 1165
   A.5.4 Convex Duality 1169

### Bibliography 1171

### Notation Index 1209

### Subject Index 1213