Contents

Preface *v*

Acknowledgments *ix*

Part I Introduction

1 Pattern Analysis as Data Reduction *1*
1.1 Why Data Reduction? *2*
1.1.1 Digital Imaging Systems *3*
1.1.2 Numerical Simulations *4*
1.1.3 Experimental Apparatus *6*
1.2 Dimensionality Reducing Mappings *8*
1.2.1 Definitions of Dimension *9*
1.2.2 Mathematical Framework of Reduction *11*
1.2.3 Reduction Mapping Optimality Criteria *13*
1.2.4 Design of Reduction Mappings *14*
1.3 Patterns and Data *18*
1.3.1 Pattern Types *18*
1.3.2 Continuous and Discrete Data *20*
Problems *21*

2 Linear Spaces and Transformations 23
2.1 Linear Transformations 24
2.2 Change of basis 25
2.3 Operations on Subspaces 28
2.4 Important Subspaces 32
2.5 Projection Matrices 33
2.5.1 Invariant Subspaces 34
2.6 Orthogonal Projection Matrices 36
2.6.1 Best Approximation Theorem 37
2.6.2 Criterion for Orthogonal Projections 38
2.6.3 Orthogonalization 40
2.7 Application: The Novelty Filter 41
2.8 Eigenvalues and Eigenvectors 43
2.9 The Singular Value Decomposition 47
2.9.1 Reduction and Compression of Matrices 51
2.9.2 Applications of the SVD 53
2.9.3 Computation of the SVD 54
Problems 56

Part II Optimal Orthogonal Pattern Representations

3 The Karhunen-Loève (KL) Expansion 63
3.1 Introduction 64
3.2 What is an optimal basis? 65
3.3 On Lines of Best Fit 67
3.4 Construction of the Optimal Basis 71
3.4.1 The Simultaneous Derivation 72
3.4.2 The Sequential Derivation 74
3.5 General Properties of the KL Expansion 76
3.6 The Snapshot Method 84
3.6.1 The Rogues Gallery Problem 86
3.6.2 Reduction of Lip Motion 91
3.7 The Singular Value Decomposition and KL 92
3.7.1 Translationally Invariant Data 98
3.8 Summary 100
Problems 101

4 Additional Theory and Applications of the KL Expansion 107
4.1 The Continuous KL Transform 108
4.2 Vector Function KL Expansions 112
4.2.1 Continuous Vector Functions 112
4.2.2 Discrete Vector Functions 114
4.3 Optimal Symmetric Expansions 115
4.3.1 Optimal Symmetric Eigenfunctions 115
4.3.2 Optimal Symmetric Eigenvectors 121
4.4 Low-Dimensional Dynamical Equations 123
4.4.1 The Galerkin Projection 124
4.4.2 The Fourier-Galerkin Projection 125
4.4.3 KL-Galerkin 126
4.4.4 A Weighted Sobolev Norm 127
4.4.5 Further Discussion 129
4.5 Implementation with Missing Data 130
4.5.1 Estimating Missing Data 130
4.5.2 Estimating a KL Basis with Missing Data 132
4.6 Application to Noisy Data 134
4.7 The Local Karhunen-Loève Expansion 137
4.7.1 Global KL Procedure on Closed Curves 139
4.7.2 The Local Approach 141
4.7.3 Local Representation and the Taylor Series 146
4.7.4 Computation of the Tangent Space 146
4.7.5 Local Whitening Transformation 148
Problems 153

Part III Time, Frequency and Scale Analysis

5 Fourier Analysis 161
5.1 The Discrete Fourier Transform 162
5.1.1 Derivation of the Transform Pair 163
5.1.2 Properties of the DFT 165
5.1.3 The DFT as an Optimal Basis 169
5.1.4 The Convolution Theorem. 170
5.1.5 The 2-D DFT 173
5.1.6 The Fast Fourier Transform 176
5.2 Continuous-Time Fourier Transform 177

5.2.1 Connection With the Fourier Series 179
5.2.2 Properties of the CTFT 181
5.2.3 The Uncertainty Principle 185
5.2.4 The Delta Function 188
5.2.5 Windowing in the Continuous-Time Domain 190
5.3 Short-Time Fourier Transform 193
Problems 195

6 Wavelet Expansions 199
6.1 The Continuous Wavelet Transform 200
6.1.1 Interpretting the CWT 202
6.1.2 Discretization of the CWT 208
6.2 The Discrete Wavelet Transform 210
6.2.1 The Multiresolution Property 213
6.2.2 The Multiresolution Analysis 215
6.2.3 The Pyramidal Decomposition 217
6.2.4 The Pyramidal Reconstruction 218
6.2.5 The Multiresolution Theorem 222
6.3 The Dilation Equation 224
6.3.1 Iteration of the Dilation Equation 224
6.3.2 Constraints on the coefficients $\{h_k\}$ 225
6.3.3 Solutions of the Scaling Equation 227
Problems 228

Part IV Adaptive Nonlinear Mappings

7 Radial Basis Functions 233
7.1 Function Approximation 234
7.1.1 The Interpolation Problem 234
7.2 The General RBF Expansion 237
7.2.1 Computing the RBF Weights 238
7.3 Direct Methods for Computing the Weights 240
7.3.1 The Normal Equations 240
7.3.2 The SVD Solution 241
7.3.3 Gram-Schmidt (QR) 242
7.4 Descent Methods for Computing the Weights 242
7.4.1 The Least Mean Square (LMS) Algorithm 243

7.4.2 Exact Line Search For Steepest Descent 247
7.4.3 The Conjugate Gradient (CG) Algorithm 249
7.4.4 The On-Line LMS Algorithm 250
7.5 Computing the RBF Centers 251
7.5.1 Nonlinear Optimization 252
7.6 Center Selection via Data Clustering 253
7.6.1 Competitive Learning 254
7.6.2 The LBG Algorithm 259
7.6.3 Topology Preserving Mappings 260
7.7 Additional Training Strategies 263
7.7.1 Orthogonal Least squares 263
7.7.2 Resource Allocating Networks 265
7.8 Summary of RBFs 265
Problems 267

8 Neural Networks 271
8.1 The Neural Network Model 273
8.2 The Simple Perceptron 276
8.3 Multi-Layer Perceptron 283
8.3.1 Function Approximation 283
8.3.2 Backpropagation 287
8.4 Optimization Techniques 291
8.4.1 Learning with Momentum 292
8.4.2 Conjugate Gradient 292
8.4.3 m-BFGS 292
8.4.4 Method of Lines 292
8.4.5 Training Strategies 293
8.5 Learning with Constraints 295
8.5.1 Regularization 296
8.5.2 Circular Nodes 298
8.6 Matrix Formulation of Backpropogation 299
8.6.1 Computation of the gradients 301
8.6.2 The Gradient Matrix 311
8.7 Neural Computation of K-L Eigenvectors 313
8.8 The Hopfield Network 317
8.9 Epilogue 320

9 Nonlinear Reduction Architectures 325

9.1 Data Parameterization *326*
9.2 The Whitney Reduction Network *328*
9.2.1 Network Architecture *328*
9.2.2 Theoretical Background *332*
9.2.3 Implementation *334*
9.2.4 Properties of the Whitney Network *337*
9.2.5 Applications *340*
9.3 Bottleneck Mappings *350*
9.3.1 Bottleneck Mappings with Constraints *354*
9.4 Neural Charts *355*
9.5 Epilogue *359*

Appendix A Mathemtical Preliminaries *363*
A.1 Sets, Classes *364*

Appendix B Linear Algebra *367*
B.1 Vector Spaces *367*
B.2 Mappings *369*
B.3 Inner Product Spaces *370*
B.4 Vector and Matrix Norms *371*
B.5 Metric Spaces *373*

Appendix C Point Set Topology *375*

Appendix D Calculus *377*

References *379*

Index *386*