

CONTENTS

Preface	xvii
 PART I INTRODUCTION	 1
1 What is Chemometrics?	3
1.1 Subject of Chemometrics, 3	
1.2 Historical Digression, 5	
 2 What the Book Is About?	 8
2.1 Useful Hints, 8	
2.2 Book Syllabus, 9	
2.3 Notations, 10	
 3 Installation of Chemometrics Add-In	 11
3.1 Installation, 11	
3.2 General Information, 14	
 4 Further Reading on Chemometrics	 15
4.1 Books, 15	
4.1.1 The Basics, 15	
4.1.2 Chemometrics, 16	
4.1.3 Supplements, 16	
4.2 The Internet, 17	
4.2.1 Tutorials, 17	
4.3 Journals, 17	

- 4.3.1 Chemometrics, 17
- 4.3.2 Analytical, 18
- 4.3.3 Mathematical, 18
- 4.4 Software, 18
 - 4.4.1 Specialized Packages, 18
 - 4.4.2 General Statistic Packages, 19
 - 4.4.3 Free Ware, 19

PART II THE BASICS **21**

5 Matrices and Vectors **23**

- 5.1 The Basics, 23
 - 5.1.1 Matrix, 23
 - 5.1.2 Simple Matrix Operations, 24
 - 5.1.3 Matrices Multiplication, 25
 - 5.1.4 Square Matrix, 26
 - 5.1.5 Trace and Determinant, 27
 - 5.1.6 Vectors, 28
 - 5.1.7 Simple Vector Operations, 29
 - 5.1.8 Vector Products, 29
 - 5.1.9 Vector Norm, 30
 - 5.1.10 Angle Between Vectors, 30
 - 5.1.11 Vector Representation of a Matrix, 30
 - 5.1.12 Linearly Dependent Vectors, 31
 - 5.1.13 Matrix Rank, 31
 - 5.1.14 Inverse Matrix, 31
 - 5.1.15 Pseudoinverse, 32
 - 5.1.16 Matrix–Vector Product, 33
- 5.2 Advanced Information, 33
 - 5.2.1 Systems of Linear Equations, 33
 - 5.2.2 Bilinear and Quadratic Forms, 34
 - 5.2.3 Positive Definite Matrix, 34
 - 5.2.4 Cholesky Decomposition, 34
 - 5.2.5 Polar Decomposition, 34
 - 5.2.6 Eigenvalues and Eigenvectors, 35
 - 5.2.7 Eigenvalues, 35
 - 5.2.8 Eigenvectors, 35
 - 5.2.9 Equivalence and Similarity, 36
 - 5.2.10 Diagonalization, 37
 - 5.2.11 Singular Value Decomposition (SVD), 37
 - 5.2.12 Vector Space, 38
 - 5.2.13 Space Basis, 39
 - 5.2.14 Geometric Interpretation, 39
 - 5.2.15 Nonuniqueness of Basis, 39
 - 5.2.16 Subspace, 40
 - 5.2.17 Projection, 40

6 Statistics**42**

- 6.1 The Basics, 42
 - 6.1.1 Probability, 42
 - 6.1.2 Random Value, 43
 - 6.1.3 Distribution Function, 43
 - 6.1.4 Mathematical Expectation, 44
 - 6.1.5 Variance and Standard Deviation, 44
 - 6.1.6 Moments, 44
 - 6.1.7 Quantiles, 45
 - 6.1.8 Multivariate Distributions, 45
 - 6.1.9 Covariance and Correlation, 45
 - 6.1.10 Function, 46
 - 6.1.11 Standardization, 46
- 6.2 Main Distributions, 46
 - 6.2.1 Binomial Distribution, 46
 - 6.2.2 Uniform Distribution, 47
 - 6.2.3 Normal Distribution, 48
 - 6.2.4 Chi-Squared Distribution, 50
 - 6.2.5 Student's Distribution, 52
 - 6.2.6 F -Distribution, 53
 - 6.2.7 Multivariate Normal Distribution, 54
 - 6.2.8 Pseudorandom Numbers, 55
- 6.3 Parameter Estimation, 56
 - 6.3.1 Sample, 56
 - 6.3.2 Outliers and Extremes, 56
 - 6.3.3 Statistical Population, 56
 - 6.3.4 Statistics, 57
 - 6.3.5 Sample Mean and Variance, 57
 - 6.3.6 Sample Covariance and Correlation, 58
 - 6.3.7 Order Statistics, 59
 - 6.3.8 Empirical Distribution and Histogram, 60
 - 6.3.9 Method of Moments, 61
 - 6.3.10 The Maximum Likelihood Method, 62
- 6.4 Properties of the Estimators, 62
 - 6.4.1 Consistency, 62
 - 6.4.2 Bias, 63
 - 6.4.3 Effectiveness, 63
 - 6.4.4 Robustness, 63
 - 6.4.5 Normal Sample, 64
- 6.5 Confidence Estimation, 64
 - 6.5.1 Confidence Region, 64
 - 6.5.2 Confidence Interval, 65
 - 6.5.3 Example of a Confidence Interval, 65
 - 6.5.4 Confidence Intervals for the Normal Distribution, 65
- 6.6 Hypothesis Testing, 66
 - 6.6.1 Hypothesis, 66
 - 6.6.2 Hypothesis Testing, 66

- 6.6.3 Type I and Type II Errors, 67
- 6.6.4 Example, 67
- 6.6.5 Pearson's Chi-Squared Test, 67
- 6.6.6 *F*-Test, 69
- 6.7 Regression, 70
 - 6.7.1 Simple Regression, 70
 - 6.7.2 The Least Squares Method, 71
 - 6.7.3 Multiple Regression, 72
- Conclusion, 73

7 Matrix Calculations in Excel

74

- 7.1 Basic Information, 74
 - 7.1.1 Region and Language, 74
 - 7.1.2 Workbook, Worksheet, and Cell, 76
 - 7.1.3 Addressing, 77
 - 7.1.4 Range, 78
 - 7.1.5 Simple Calculations, 78
 - 7.1.6 Functions, 78
 - 7.1.7 Important Functions, 81
 - 7.1.8 Errors in Formulas, 85
 - 7.1.9 Formula Dragging, 86
 - 7.1.10 Create a Chart, 87
- 7.2 Matrix Operations, 88
 - 7.2.1 Array Formulas, 88
 - 7.2.2 Creating and Editing an Array Formula, 90
 - 7.2.3 Simplest Matrix Operations, 91
 - 7.2.4 Access to the Part of a Matrix, 91
 - 7.2.5 Unary Operations, 93
 - 7.2.6 Binary Operations, 95
 - 7.2.7 Regression, 95
 - 7.2.8 Critical Bug in Excel 2003, 99
 - 7.2.9 Virtual Array, 99
- 7.3 Extension of Excel Possibilities, 100
 - 7.3.1 VBA Programming, 100
 - 7.3.2 Example, 101
 - 7.3.3 Macro Example, 103
 - 7.3.4 User-Defined Function Example, 104
 - 7.3.5 Add-Ins, 105
 - 7.3.6 Add-In Installation, 106
- Conclusion, 107

8 Projection Methods in Excel

108

- 8.1 Projection Methods, 108
 - 8.1.1 Concept and Notation, 108
 - 8.1.2 PCA, 109
 - 8.1.3 PLS, 110
 - 8.1.4 Data Preprocessing, 111

8.1.5	Didactic Example, 112
8.2	Application of Chemometrics Add-In, 113
8.2.1	Installation, 113
8.2.2	General, 113
8.3	PCA, 114
8.3.1	ScoresPCA, 114
8.3.2	LoadingsPCA, 114
8.4	PLS, 116
8.4.1	ScoresPLS, 116
8.4.2	UScoresPLS, 117
8.4.3	LoadingsPLS, 118
8.4.4	WLoadingsPLS, 119
8.4.5	QLoadingsPLS, 120
8.5	PLS2, 121
8.5.1	ScoresPLS2, 121
8.5.2	UScoresPLS2, 122
8.5.3	LoadingsPLS2, 124
8.5.4	WLoadingsPLS2, 125
8.5.5	QLoadingsPLS2, 126
8.6	Additional Functions, 127
8.6.1	MIdent, 127
8.6.2	MIdentD2, 127
8.6.3	MCutRows, 129
8.6.4	MTrace, 129
	Conclusion, 130

PART III CHEMOMETRICS **131**

9 Principal Component Analysis (PCA) **133**

9.1	The Basics, 133
9.1.1	Data, 133
9.1.2	Intuitive Approach, 134
9.1.3	Dimensionality Reduction, 136
9.2	Principal Component Analysis, 136
9.2.1	Formal Specifications, 136
9.2.2	Algorithm, 137
9.2.3	PCA and SVD, 137
9.2.4	Scores, 138
9.2.5	Loadings, 139
9.2.6	Data of Special Kind, 140
9.2.7	Errors, 140
9.2.8	Validation, 143
9.2.9	Decomposition “Quality”, 143
9.2.10	Number of Principal Components, 144
9.2.11	The Ambiguity of PCA, 145
9.2.12	Data Preprocessing, 146
9.2.13	Leverage and Deviation, 146

- 9.3 People and Countries, 146
 - 9.3.1 Example, 146
 - 9.3.2 Data, 147
 - 9.3.3 Data Exploration, 147
 - 9.3.4 Data Pretreatment, 148
 - 9.3.5 Scores and Loadings Calculation, 149
 - 9.3.6 Scores Plots, 151
 - 9.3.7 Loadings Plot, 152
 - 9.3.8 Analysis of Residuals, 153
- Conclusion, 153

10 Calibration

156

- 10.1 The Basics, 156
 - 10.1.1 Problem Statement, 156
 - 10.1.2 Linear and Nonlinear Calibration, 157
 - 10.1.3 Calibration and Validation, 158
 - 10.1.4 Calibration “Quality”, 160
 - 10.1.5 Uncertainty, Precision, and Accuracy, 162
 - 10.1.6 Underfitting and Overfitting, 163
 - 10.1.7 Multicollinearity, 164
 - 10.1.8 Data Preprocessing, 166
- 10.2 Simulated Data, 166
 - 10.2.1 The Principle of Linearity, 166
 - 10.2.2 “Pure” Spectra, 166
 - 10.2.3 “Standard” Samples, 166
 - 10.2.4 X Data Creation, 167
 - 10.2.5 Data Centering, 168
 - 10.2.6 Data Overview, 168
- 10.3 Classic Calibration, 169
 - 10.3.1 Univariate (Single Channel) Calibration, 169
 - 10.3.2 The Vierordt Method, 172
 - 10.3.3 Indirect Calibration, 174
- 10.4 Inverse Calibration, 176
 - 10.4.1 Multiple Linear Calibration, 177
 - 10.4.2 Stepwise Calibration, 178
- 10.5 Latent Variables Calibration, 180
 - 10.5.1 Projection Methods, 180
 - 10.5.2 Latent Variables Regression, 184
 - 10.5.3 Implementation of Latent Variable Calibration, 185
 - 10.5.4 Principal Component Regression (PCR), 186
 - 10.5.5 Projection on the Latent Structures-1 (PLS1), 188
 - 10.5.6 Projection on the Latent Structures-2 (PLS2), 191
- 10.6 Methods Comparison, 193
 - Conclusion, 197

11 Classification

198

- 11.1 The Basics, 198

- 11.1.1 Problem Statement, 198
- 11.1.2 Types of Classes, 199
- 11.1.3 Hypothesis Testing, 199
- 11.1.4 Errors in Classification, 200
- 11.1.5 One-Class Classification, 200
- 11.1.6 Training and Validation, 201
- 11.1.7 Supervised and Unsupervised Training, 201
- 11.1.8 The Curse of Dimensionality, 201
- 11.1.9 Data Preprocessing, 201
- 11.2 Data, 202
 - 11.2.1 Example, 202
 - 11.2.2 Data Subsets, 203
 - 11.2.3 Workbook Iris.xls, 204
 - 11.2.4 Principal Component Analysis, 205
- 11.3 Supervised Classification, 205
 - 11.3.1 Linear Discriminant Analysis (LDA), 205
 - 11.3.2 Quadratic Discriminant Analysis (QDA), 210
 - 11.3.3 PLS Discriminant Analysis (PLSDA), 214
 - 11.3.4 SIMCA, 217
 - 11.3.5 k -Nearest Neighbors (kNN), 223
- 11.4 Unsupervised Classification, 225
 - 11.4.1 PCA Again (Revisited), 225
 - 11.4.2 Clustering by K -Means, 225
- Conclusion, 229

12 Multivariate Curve Resolution

230

- 12.1 The Basics, 230
 - 12.1.1 Problem Statement, 230
 - 12.1.2 Solution Ambiguity, 232
 - 12.1.3 Solvability Conditions, 234
 - 12.1.4 Two Types of Data, 235
 - 12.1.5 Known Spectrum or Profile, 236
 - 12.1.6 Principal Component Analysis (PCA), 236
 - 12.1.7 PCA and MCR, 237
- 12.2 Simulated Data, 237
 - 12.2.1 Example, 237
 - 12.2.2 Data, 238
 - 12.2.3 PCA, 238
 - 12.2.4 The HELP Plot, 240
- 12.3 Factor Analysis, 241
 - 12.3.1 Procrustes Analysis, 241
 - 12.3.2 Evolving Factor Analysis (EFA), 244
 - 12.3.3 Windows Factor Analysis (WFA), 246
- 12.4 Iterative Methods, 249
 - 12.4.1 Iterative Target Transform Factor Analysis (ITTFA), 249
 - 12.4.2 Alternating Least Squares (ALS), 250
- Conclusion, 252

PART IV SUPPLEMENTS	255
13 Extension Of Chemometrics Add-In	257
13.1 Using Virtual Arrays, 257	
13.1.1 Simulated Data, 257	
13.1.2 Virtual Array, 259	
13.1.3 Data Preprocessing, 259	
13.1.4 Decomposition, 260	
13.1.5 Residuals Calculation, 260	
13.1.6 Eigenvalues Calculation, 262	
13.1.7 Orthogonal Distances Calculation, 263	
13.1.8 Leverages Calculation, 264	
13.2 Using VBA Programming, 265	
13.2.1 VBA Advantages, 265	
13.2.2 Virtualization of Real Arrays, 265	
13.2.3 Data Preprocessing, 266	
13.2.4 Residuals Calculation, 267	
13.2.5 Eigenvalues Calculation, 268	
13.2.6 Orthogonal Distances Calculation, 269	
13.2.7 Leverages Calculation, 270	
Conclusion, 271	
14 Kinetic Modeling of Spectral Data	272
14.1 The “Grey” Modeling Method, 272	
14.1.1 Problem Statement, 272	
14.1.2 Example, 274	
14.1.3 Data, 274	
14.1.4 Soft Method of Alternating Least Squares (Soft-ALS), 275	
14.1.5 Hard Method of Alternating Least Squares (Hard-ALS), 277	
14.1.6 Using Solver Add-In, 279	
Conclusions, 282	
15 MATLAB®: Beginner’s Guide	283
15.1 The Basics, 283	
15.1.1 Workspace, 283	
15.1.2 Basic Calculations, 285	
15.1.3 Echo, 285	
15.1.4 Workspace Saving: MAT-Files, 286	
15.1.5 Diary, 286	
15.1.6 Help, 287	
15.2 Matrices, 287	
15.2.1 Scalars, Vectors, and Matrices, 287	
15.2.2 Accessing Matrix Elements, 289	
15.2.3 Basic Matrix Operations, 289	
15.2.4 Special Matrices, 290	
15.2.5 Matrix Calculations, 292	

- 15.3 Integrating Excel and MATLAB®, 294
 - 15.3.1 Configuring Excel, 294
 - 15.3.2 Data Exchange, 294
- 15.4 Programming, 295
 - 15.4.1 M-Files, 295
 - 15.4.2 Script File, 296
 - 15.4.3 Function File, 297
 - 15.4.4 Plotting, 298
 - 15.4.5 Plot Printing, 300
- 15.5 Sample Programs, 301
 - 15.5.1 Centering and Scaling, 301
 - 15.5.2 SVD/PCA, 301
 - 15.5.3 PCA/NIPALS, 302
 - 15.5.4 PLS1, 303
 - 15.5.5 PLS2, 304
 - Conclusion, 306

Afterword. The Fourth Paradigm	307
Index	311