These proofs may contain colour figures. Those figures may print black and white in the final printed book if a colour print product has not been planned. The colour figures will appear in colour in all electronic versions of this book.

B978-0-323-95576-8.09993-1, 09993

Contents

Preface *xi*
Acknowledgment *xvii*
Advice on scripting for beginners *xix*

1. Data analysis with MATLAB(R) or Python **1**

Part A: MATLAB(R) 2
1A.1 Getting started 2
1A.2 Navigating folders 4
1A.3 Simple arithmetic and algebra 5
1A.4 Vectors and matrices 6
1A.5 Addition, subtraction and multiplication of column-vectors and matrices 7
1A.6 Element access 9
1A.7 Representing functions 10
1A.8 To loop or not to loop 12
1A.9 The matrix inverse 14
1A.10 Loading data from a file 15
1A.11 Plotting data 16
1A.12 Saving data to a file 19
1A.13 Some advice on writing scripts 20

Part B: Python 22
1B.1 Installation 22
1B.2 The first cell in the `edapy_01` Jupyter Notebook 22
1B.3 A very simple Python script 25
1B.4 Navigating folders 25
1B.5 Simple arithmetic and algebra 26
1B.6 List-like data 27
1B.7 Creating column-vectors and matrices 28
1B.8 Addition, subtraction and multiplication of column-vectors and matrices 30
1B.9 Element access 32
1B.10 Representing functions 33
1B.11 To loop or not to loop 36
1B.12 The matrix inverse 38
1B.13 Loading data from a file 38
1B.14 Plotting data 39
1B.15 Saving data to a file 42

Insert Registered Trademark Symbol in two places

Menke, 978-0-323-95576-8

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter SPi. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

These proofs may contain colour figures. Those figures may print black and white in the final printed book if a colour print product has not been planned. The colour figures will appear in colour in all electronic versions of this book.

B978-0-323-95576-8.09993-1, 09993

1B.16 Some advice on writing scripts 42
Problems 45

2. Systematic explorations of a new dataset 47

2.1 Case study: Black rock forest temperature data 47
2.2 More on graphics 59
2.3 Case study: Neuse River hydrograph used to explore rate data 66
2.4 Case study, Atlantic Rock dataset used to explore scatter plots 69
2.5 More on character strings 73
Problems 76

3. Modeling observational noise with random variables 77

3.1 Random variables 77
3.2 Mean, median, and mode 79
3.3 Variance 85
3.4 Two important probability density functions 87
3.5 Functions of a random variable 89
3.6 Joint probabilities 90
3.7 Bayesian inference 93
3.8 Joint probability density functions 94
3.9 Covariance 100
3.10 The multivariate normal p.d.f. 102
3.11 Linear functions of multivariate data 105
Problems 109
Reference 110

4. Linear models as the foundation of data analysis 111

4.1 Quantitative models, data, and model parameters 111
4.2 The simplest of quantitative models 113
4.3 Curve fitting 114
4.4 Mixtures 119
4.5 Weighted averages 121
4.6 Examining error 124
4.7 Least squares 129
4.8 Examples 131
4.9 Covariance and the behavior of error 136
Problems 140

5. Least squares with prior information 141

5.1 When least square fails 141
5.2 Prior information 143

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter SPi. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

These proofs may contain colour figures. Those figures may print black and white in the final printed book if a colour print product has not been planned. The colour figures will appear in colour in all electronic versions of this book.

B978-0-323-95576-8.09993-1, 09993

5.3 Bayesian inference 144
5.4 The product of Normal probability density functions 146
5.5 Generalized least squares 148
5.6 The role of the covariance of the data 150
5.7 Smoothness as prior information 152
5.8 Sparse matrices 156
5.9 Reorganizing grids of model parameters 160
Problems 169
References 170

6. Detecting periodicities with Fourier analysis 171
6.1 Describing sinusoidal oscillations 171
6.2 Models composed only of sinusoidal functions 173
6.3 Going complex 182
6.4 Lessons learned from the integral transform 187
6.5 Normal curve 188
6.6 Spikes 188
6.7 Area under a function 190
6.8 Time-delayed function 192
6.9 Derivative of a function 193
6.10 Integral of a function 195
6.11 Convolution 196
6.12 Nontransient signals 197
Problems 203
Further reading 203

7. Modeling time-dependent behavior with filters 205
7.1 Behavior sensitive to past conditions 205
7.2 Filtering as convolution 209
7.3 Solving problems with filters 211
7.4 Case study: Empirically-derived filter for Hudson River discharge 221
7.5 Predicting the future 223
7.6 A parallel between filters and polynomials 225
7.7 Filter cascades and inverse filters 227
7.8 Making use of what you know 231
Problems 237
Reference 237

8. Undirected data analysis using factors, empirical orthogonal functions, and clusters 239
8.1 Samples as mixtures 239

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter SPi. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

These proofs may contain colour figures. Those figures may print black and white in the final printed book if a colour print product has not been planned. The colour figures will appear in colour in all electronic versions of this book.

B978-0-323-95576-8.09993-1, 09993

8.2 Determining the minimum number of factors 242
8.3 Application to the Atlantic Rocks dataset 246
8.4 Spiky factors 249
8.5 Weighting of elements 254
8.6 Q-mode factor analysis 255
8.7 Factors and factor loadings with natural orderings 256
8.8 Case study: EOF's of the equatorial Pacific Ocean Sea surface temperature dataset 258
8.9 Clusters of data 262
8.10 K-mean cluster analysis 267
8.11 Clustering the Atlantic Rocks dataset 270
Problems 274
References 275

9. Detecting and understanding correlations among data 277

9.1 Correlation is covariance 277
9.2 Computing autocorrelation by hand 283
9.3 Relationship to convolution and power spectral density 284
9.4 Cross-correlation 285
9.5 Using the cross-correlation to align time series 287
9.6 Least squares estimation of filters 291
9.7 The effect of smoothing on time series 296
9.8 Band-pass filters 300
9.9 Case study: Coherence of the Reynolds Channel water quality data 305
9.10 Windowing before computing Fourier transforms 310
9.11 Optimal window functions 313
Problems 317
References 317

10. Interpolation, Gaussian process regression, and kriging 319

10.1 Interpolation requires prior information 319
10.2 Linear interpolation 320
10.3 Cubic interpolation 323
10.4 Gaussian process regression and kriging 325
10.5 Example of Gaussian process regression 329
10.6 Tuning of parameters in Gaussian process regression 333
10.7 Example of tuning 335
10.8 Interpolation in two-dimensions 335
10.9 Fourier transforms in two dimensions 340
10.10 Using Fourier transforms to fill in missing data 342
Problems 347
References 348

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter SPi. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

These proofs may contain colour figures. Those figures may print black and white in the final printed book if a colour print product has not been planned. The colour figures will appear in colour in all electronic versions of this book.

B978-0-323-95576-8.09993-1, 09993

11. Approximate methods, including linearization and artificial neural networks 349

11.1 The value of simplicity 349
11.2 Polynomial approximations and Taylor series 350
11.3 Small number approximations 351
11.4 Small number approximation applied to distance on a sphere 352
11.5 Small number approximation applied to variance 353
11.6 Taylor series in multiple dimensions 354
11.7 Small number approximation applied to covariance 355
11.8 Solving nonlinear problems with iterative least squares 356
11.9 Fitting a sinusoid of unknown frequency 357
11.10 The gradient descent method 359
11.11 Precomputation of a function and table lookups 361
11.12 Artificial neural networks 363
11.13 Information flow in a neural net 365
11.14 Training a neural net 372
11.15 Neural net for a nonlinear filter 375
Problems 382
References 383

12. Assessing the significance of results 385

12.1 Rejecting the Null Hypothesis 385
12.2 The distribution of the total error 386
12.3 Four important probability density functions 389
12.4 Case study with common hypothesis testing scenarios 391
12.5 Chi-squared test for generalized least squares 398
12.6 Testing improvement in fit 400
12.7 Testing the significance of a spectral peak 402
12.8 Significance of a spectral peak for a correlated time series 407
12.9 Bootstrap confidence intervals 410
Problems 418

13. Notes 419

13.1 Note 1.1 On the persistence of variables 419
13.2 Note 2.1 On time 420
13.3 Note 2.2 On reading complicated text files 421
13.4 Note 3.1 On the rule for error propagation 423
13.5 Note 3.2 On the `eda_draw()` function 424
13.6 Note 4.1 On complex least squares 424
13.7 Note 5.1 On the derivation of generalized least squares 427
13.8 Note 5.2 MATLAB and Python functions

(R)
^

insert
Registered Trademark
symbol

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter SPi. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

These proofs may contain colour figures. Those figures may print black and white in the final printed book if a colour print product has not been planned. The colour figures will appear in colour in all electronic versions of this book.

B978-0-323-95576-8.09993-1, 09993

13.9 Note 5.3 On reorganizing matrices 429
13.10 Note 6.1 On the `atan2()` function 430
13.11 Note 6.2 On the orthonormality of the discrete Fourier data kernel 431
13.12 Note 6.3 On the expansion of a function in an orthonormal basis 432
13.13 Note 8.1 On singular value decomposition 433
13.14 Note 9.1 On coherence 434
13.15 Note 9.2 On Lagrange multipliers 434
13.16 Note 10.1 Covariance matrix corresponding to prior information of smoothness 435
13.17 Note 10.2 Issues encountered tuning Gaussian Process Regression 437
13.18 Note 11.1 On the chain rule for partial derivatives 438
References 439

Index *441*

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter SPi. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.