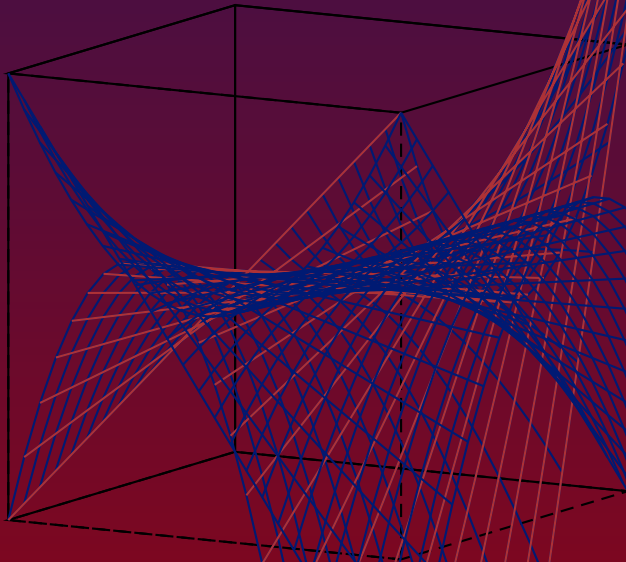


S. Chekanov

The Art of Scientific Computation

*Introduction to data analysis using
Python scripting, Jython and Java*



Copyright ©2009 by S.V. Chekanov. All Rights Reserved.

No part of this book shall be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written permission from the publisher. No patent liability is assumed with respect to the use of the information contained here. Although every precaution has been taken in the preparation of this book, the publisher and author assume no responsibility for errors or omissions. Neither is any liability assumed for damages resulting from the use of the information contained herein. The author and publisher shall not be liable in any event for incidental or consequential damage in connection with, or arising out of, the furnishing, performance, or use of the code examples given in this book.

ISBN 978-0-557-12010-9

Second Edition (2009). 517 pages.

Book preview. Not for redistribution

Contents

Introduction	1
1 Jython, Java and jHepWork	5
1.1 Introduction	5
1.1.1 Books you may read before	6
1.1.2 Yes, it is pure Java	7
1.1.3 Some warnings	8
1.1.4 Errors	9
1.2 Introduction to scientific computing	10
1.2.1 Book examples and the power of Jython	10
1.2.2 The history of jHepWork	11
1.2.3 Why Jython?	12
1.2.4 Differences with other data-analysis packages	13
1.2.5 How fast it is?	15
1.2.6 Jython and CPython versions	16
1.3 Installation	17
1.4 Introduction to the jHepWork IDE	18
1.4.1 Source Code Editor	19
1.4.2 jHepWork Java libraries and Python packages	19
1.4.3 Jython and Bean shell consoles	22
1.4.4 Accessing methods of instances	24
1.4.5 Editing Jython scripts	24
1.4.6 Running Jython scripts	25
1.4.7 Running a BeanShell scripts	26
1.4.8 Compiling and running Java code	26
1.4.9 Working with command-line scripts	26
1.4.10 jHepWork code assist	27

1.4.11	Other features	28
1.5	Third-party packages and the license	29
1.5.1	Contributions and third-party packages	29
1.5.2	Disclaimer of Warranty	31
1.5.3	jHepWork License	32
2	Introduction to Jython	33
2.1	Code structure and commentary	34
2.2	Quick introduction to Jython objects	34
2.2.1	Numbers as objects	37
2.2.2	Formatted output	39
2.2.3	Mathematical functions	40
2.2.4	Complex Numbers	41
2.3	Strings as objects	42
2.4	Import statements	43
2.4.1	Executing native applications	44
2.5	Comparison tests and loops	44
2.5.1	The 'if-else' statement	44
2.5.2	Loops. The "for" statement	46
2.5.3	The 'continue' and 'break' statements	47
2.5.4	Loops. The 'while' statement	47
2.6	Collections	48
2.6.1	Lists	48
2.6.2	List creation	49
2.6.3	Iteration over elements	51
2.6.4	Removal of duplicates	52
2.6.5	Tuples	54
2.6.6	Functional programming. Operations with lists	54
2.6.7	Dictionaries	57
2.7	Java collections in Jython	59
2.7.1	List. An ordered collection	59
2.7.2	Set. A collection without duplicate elements	62
2.7.3	SortedSet. Sorted unique elements	63
2.7.4	Map. Mapping keys to values	64
2.7.5	Java map with sorted elements	65
2.7.6	Real life example: sorting and removing duplicates	65
2.8	Random numbers	66
2.9	Time module	68
2.9.1	Benchmarking	69
2.10	Python functions and modules	70
2.11	Python classes	74

2.11.1	Initializing a class	76
2.11.2	Classes inherited from other classes	76
2.11.3	Java classes in Jython	77
2.11.4	Not covered topics	78
2.12	Used memory	78
2.13	Parallel computing and threads	78
2.14	Arrays in Jython	80
2.14.1	Array conversion and transformations	81
2.14.2	Performance issues	82
2.15	Exceptions in Python	83
2.16	Input and output	84
2.16.1	User interaction	84
2.16.2	Reading and writing files	84
2.16.3	Input and output for arrays	86
2.16.4	Working with CSV Python module	87
2.16.5	Saving objects in a serialized file	89
2.16.6	Storing multiple objects	90
2.16.7	Using Java for I/O	91
2.16.8	Reading data from the network	92
2.17	Real-life example. Collecting data files	93
2.18	Using Java for GUI programming	95
2.19	Concluding remarks	97
3	Mathematical functions	99
3.1	Jython functions	99
3.2	1D functions in jHepWork	101
3.2.1	Details of Java implementation	104
3.2.2	Integration and differentiation	105
3.3	Plotting 1D functions	106
3.3.1	Building a graphical canvas	106
3.3.2	Drawing 1D functions	110
3.3.3	Plotting 1D functions on different pads	113
3.3.4	Short summary of HPlot methods	113
3.3.5	Example	115
3.4	2D functions	116
3.4.1	Functions in two dimensions	116
3.4.2	Displaying 2D functions on a lego plot	117
3.4.3	Using a contour plot	120
3.5	3D functions	120
3.5.1	Functions in three dimensions	120
3.6	Functions in many dimensions	122

3.6.1	FND functions	122
3.6.2	Drawing FND functions	123
3.7	Custom functions defined by Jython scripts	124
3.7.1	Custom functions and their methods	124
3.7.2	Using external libraries	127
3.7.3	Plotting custom functions	129
3.8	Parametric surfaces in 3D	130
3.8.1	FPR functions	130
3.8.2	3D mathematical objects	134
3.9	Symbolic calculations	134
3.10	File input and output	137
4	One-dimensional data	139
4.1	One dimensional arrays	139
4.2	P0D data container	140
4.2.1	P0D transformations	143
4.2.2	Analyzing P0D and summary statistics	145
4.2.3	Displaying P0D data	147
4.3	Reading and writing P0D files	149
4.3.1	Serialization	150
4.3.2	XML format	151
4.3.3	Dealing with object collections	152
5	Two-dimensional data	155
5.1	Two dimensional data structures	155
5.2	Two dimensional data with errors	157
5.2.1	Viewing P1D data	161
5.2.2	Plotting P1D data	163
5.2.3	Contour plots	165
5.3	Manipulations with P1D data	166
5.3.1	Advanced P1D operations	168
5.3.2	Weighted average and systematical uncertainties	170
5.4	Reading and writing P1D data	173
5.4.1	Dealing with a single P1D container	173
5.4.2	Reading and writing collections	175
5.5	Real life example I: Henon attractor	177
5.6	Real-life example II. Weighted average	178

6	Multi-dimensional data	183
6.1	P2D data container	183
6.1.1	Drawing P2D and HPlot3D canvas	183
6.2	P3D data container	186
6.3	PND data container	189
6.3.1	Operations with PND data	191
6.4	Input and output	192
7	Arrays, matrices and linear algebra	193
7.1	Jaida data containers	193
7.1.1	Jaida clouds	195
7.2	jMathTools arrays and operations	197
7.2.1	1D arrays and operations	197
7.2.2	2D arrays	198
7.3	Colt data containers	200
7.4	Statistical analysis using Jython	202
7.5	Matrix packages	205
7.5.1	Basic matrix arithmetic	207
7.5.2	Elements of linear algebra	208
7.5.3	Jampack matrix computations and complex matrices ..	209
7.5.4	Jython vector and matrix operations	210
7.5.5	Matrix operations in SymPy	212
7.6	Lorentz vector and particle representations	213
7.6.1	Three-vector and Lorentz vector	213
7.6.2	Classes representing particles	215
8	Histograms	217
8.1	One-dimensional histogram	218
8.1.1	Probability distribution and probability density	222
8.1.2	Histogram characteristics	223
8.1.3	Histogram initialization and filling methods	224
8.1.4	Accessing histogram values	226
8.1.5	Integration	228
8.1.6	Histogram operations	229
8.1.7	Accessing low-level Jaida classes	230
8.1.8	Graphical attributes	230
8.2	Histogram in 2D	232
8.2.1	Histogram operations	234
8.2.2	Graphical representation	235
8.3	Histograms in Jaida	238
8.4	Histogram in 3D	240

8.5	Profile histograms	241
8.5.1	Profile histograms in 1D	241
8.5.2	Profile histograms in 2D	242
8.6	Histogram input and output	243
8.6.1	External programs for histograms	245
8.7	Real-life example. Analyzing histograms from multiple files	246
9	Random numbers and statistical samples	249
9.1	Random numbers in Jython	250
9.2	Random numbers in Java	251
9.3	Random numbers from the Colt package	252
9.4	Random numbers from the jhplot.math package	254
9.4.1	Apache common math package	255
9.5	Random sampling	256
9.5.1	Methods for 1D arrays from jhplot.math	257
9.5.2	Methods for 2D arrays from jhplot.math	259
9.6	Sampling using the Colt package	260
10	Graphical canvases	263
10.1	HPlot canvas	264
10.2	Working with the HPlot canvas	266
10.2.1	Saving plots in XML files	269
10.2.2	Reading data	269
10.2.3	Clear the HPlot canvas from graphics	270
10.2.4	Axes	270
10.2.5	Summary of the HPlot methods	271
10.2.6	Saving drawings in an image file	271
10.3	Labels and keys	273
10.3.1	Simple text labels	273
10.3.2	Interactive labels	274
10.3.3	Interactive text labels with keys	275
10.4	Geometrical primitives	277
10.5	Text strings and symbols	279
10.6	SHPlot class. HPlot canvas as a singleton	279
10.7	Visualizing interconnected objects	282
10.8	Showing charts	284
10.9	SPlot class. A simple canvas	285
10.9.1	Henon attractor again	288
10.10	Canvas for interactive drawing	289
10.10.1	Drawing diagrams	290
10.10.2	SHPlotJa class	292

10.11	HPlot2D canvas	292
10.12	3D canvas	296
10.13	HPlot3D canvas	296
10.13.1	HPlot3DP canvas	296
10.13.2	3D Geometry package	300
10.14	Combining graphs with Java Swing GUI components	301
10.15	Showing streams of data in real time	303
11	Input and output	307
11.1	Non-persistent data. Memory-based data	308
11.2	Serialization of objects	309
11.3	Storing data persistently	310
11.3.1	Sequential input and output	310
11.3.2	GUI browser for serialized objects	313
11.3.3	Saving event records persistently	314
11.3.4	Buffer size for I/O intensive operations	315
11.3.5	Input and output to XML files	316
11.3.6	Non-sequential input and output	317
11.4	Compressed PFile format	319
11.4.1	Browser dialog for PFile files	322
11.5	Reading ROOT and AIDA files	323
11.5.1	Reading ROOT histograms	323
11.5.2	Reading ROOT trees	324
11.5.3	Plotting ROOT or AIDA objects using jHepWork IDE	326
11.6	Working with relational SQL databases	327
11.6.1	Creating a SQL database	328
11.6.2	Working with a database	330
11.6.3	Creating a read-only compact database	330
11.7	Reading and writing CSV files	331
11.7.1	Reading CSV files	331
11.7.2	Writing CSV file	333
11.8	Google's Protocol Buffer format	334
11.8.1	Prototyping data records	335
11.8.2	Dealing with data using Java	337
11.8.3	Switching to Jython	339
11.8.4	Adding new data records	340
11.8.5	Using C++ for I/O in the Protocol Buffers format	341
11.8.6	Some remarks	343
11.9	EFile data output	344
11.10	Reading DIF files	347
11.11	Summary	347

11.11.1	Dealing with single objects	348
11.11.2	Dealing with collections of objects	348
12	Miscellaneous analysis issues using jHepWork	351
12.1	Accessing third-party libraries	351
12.1.1	Extracting data points from a figure	351
12.1.2	Cellular automaton	352
12.2	Downloading files from the Web	353
12.3	Macro files for jHepWork editor	354
12.4	Data output to tables and spreadsheets	355
12.4.1	Showing data in a sortable table	355
12.4.2	Spreadsheet support	357
12.5	Accessing external Java and Jython libraries	357
12.6	Working with a jHepWork project	358
12.6.1	Pure Jython project	358
12.6.2	Pure Java project	359
12.6.3	Mixing Jython with Java	359
12.7	Working with images	360
12.7.1	Saving plots in external image file	360
12.7.2	View an image. IView class.	360
12.7.3	Analyzing and editing images	361
12.8	Complex numbers	361
12.9	Building list of files	362
12.10	Reading configuration files	363
12.10.1	Configuration files using Jython	363
12.10.2	Reading configuration files using Java	365
12.11	Jython scripting with jHepWork	366
12.11.1	Jython operations with data holders	369
12.12	Unwrapping Jython code. Back to Java	369
12.13	Embedding jHepWork in applets	372
13	Data clustering	377
13.1	Data clustering. Real-life example	377
13.1.1	Preparing a data sample	378
13.1.2	Clustering analysis	381
13.1.3	Interactive clustering with JMinHEP	385
14	Linear regression and curve fitting	387
14.1	Linear regression	387
14.1.1	Data set	388
14.1.2	Analyzing the data set	388

14.2	Curve fitting of data	391
14.2.1	Preparing a fit	392
14.2.2	Creating a fit function	394
14.3	Displaying a fit function	399
14.3.1	Making a fit	399
14.4	Real-life example. Signal plus background	401
14.4.1	Preparing a data sample	402
14.4.2	Performing curve fitting	403
14.4.3	Fitting multiple peaks	405
14.4.4	Fitting histograms in 3D	407
14.5	Interactive fit	409
15	Neural networks	411
15.1	Introduction	411
15.1.1	Generating a data sample	412
15.1.2	Data preparation	413
15.1.3	Building a neural net	416
15.1.4	Training and verifying	418
15.2	Bayesian networks	421
15.3	Self-organizing map	421
15.3.1	Non-interactive BSOM	423
15.4	Neural network using Python libraries	424
16	Steps in data analysis	429
16.1	Major analysis steps	429
16.2	Real life example. Analyzing a gene catalog	431
16.2.1	Data transformation	432
16.2.2	Data skimming	433
16.2.3	Data slimming	434
16.2.4	Data sorting	434
16.2.5	Removing duplicate records	436
16.2.6	Sorting and removing duplicate records using Java	437
16.3	Using metadata for data mining	439
16.3.1	Analyzing data using built-in metadata file	439
16.3.2	Using an external metadata file	443
16.4	Multiprocessor programming	444
16.4.1	Reading data in parallel	444
16.4.2	Processing a single input file in parallel	447
16.4.3	Numerical computations using multiple cores	450
16.5	Data consistency and security. MD5 class	451
16.5.1	MD5 fingerprint at runtime	451

16.5.2 Fingerprinting files 453

17 Real-life examples 455

17.1 Measuring single-particle densities 455

 17.1.1 Preparing a data sample 456

 17.1.2 Analyzing data 457

17.2 Many-particle densities, fluctuations and correlations..... 459

 17.2.1 Building a data sample for analysis 459

 17.2.2 Analyzing the data 464

 17.2.3 Reading the data and plotting multiplicities 465

17.3 Analyzing macro data: nearby galaxies 470

17.4 Analyzing micro data: elementary particles 472

17.5 A Monte Carlo simulation of particle decays 475

17.6 Measuring the speed of light using the Internet 477

 17.6.1 Getting host names in each continent..... 479

 17.6.2 Checking response from servers 480

 17.6.3 Creating histograms with the response time 481

 17.6.4 Final measurements 481

References 485

Index of examples 490

Index 491

Book preview. Not for redistribution