Half Title Page

Title Page

LOC Page

Vince: to Riggins
Geraint: also, to Riggins

Contents

Foreword

This is the foreword

Preface

This is the preface.

Contributors

Michaél Aftosmis
NASA Ames Research Center
Moffett Field, California

Pratul K. Agarwal
Oak Ridge National Laboratory
Oak Ridge, Tennessee

Sadaf R. Alam
Oak Ridge National Laboratory
Oak Ridge, Tennessee

Gabrielle Allen
Louisiana State University
Baton Rouge, Louisiana

Martin Sandve Alnæs
Simula Research Laboratory and University of Oslo, Norway
Norway

Steven F. Ashby
Lawrence Livermore National Laboratory
Livermore, California

David A. Bader
Georgia Institute of Technology
Atlanta, Georgia

Benjamin Bergen
Los Alamos National Laboratory
Los Alamos, New Mexico

Jonathan W. Berry
Sandia National Laboratories
Albuquerque, New Mexico

Martin Berzins
University of Utah
Salt Lake City, Utah

Abhinav Bhatele
University of Illinois
Urbana-Champaign, Illinois

Christian Bischof
RWTH Aachen University
Germany

Rupak Biswas
NASA Ames Research Center
Moffett Field, California

Eric Bohm
University of Illinois
Urbana-Champaign, Illinois

James Bordner
University of California, San Diego
San Diego, California

Geörge Bosilca
University of Tennessee
Knoxville, Tennessee

Grèg L. Bryan
Columbia University
New York, New York

Marian Bubak
AGH University of Science and Technology
Kraków, Poland

Andrew Canning
Lawrence Berkeley National Laboratory
Berkeley, California

Jonathan Carter
Lawrence Berkeley National Laboratory
Berkeley, California

Zizhong Chen
Jacksonville State University
Jacksonville, Alabama

Joseph R. Crobak
Rutgers, The State University of New Jersey
Piscataway, New Jersey

Roxana E. Diaconescu
Yahoo! Inc.
Burbank, California

Roxana E. Diaconescu
Yahoo! Inc.
Burbank, California

I

Getting Started

CHAPTER 1

Introduction

THANK you for starting to read this book. This book aims to bring together two fascinating topics:

- Problems that can be solved using mathematics;
- Software that is free to use and change.

What we mean by both of those things will become clear through reading this chapter and the rest of the book.

1.1 WHO IS THIS BOOK FOR?

Anyone who is interested in using mathematics and computers to solve problems will hopefully find this book helpful.

If you are a student of a mathematical discipline, a graduate student of a subject like operational research, a hobbyist who enjoys solving the travelling salesman problem or even if you get paid to do this stuff: this book is for you. We will introduce you to the world of open source software that allows you to do all these things freely.

If you are a student learning to write code, a graduate student using databases for their research, an enthusiast who programmes applications to help coordinate the neighbourhood watch, or even if you get paid to write software: this book is for you. We will introduce you to a world of problems that can be solved using your skill sets.

It would be helpful for the reader of this book to:

- Have access to a computer and be able to connect to the internet (at least once) to be able to download the relevant software.
- Be prepared to read some mathematics. Technically you do not need to understand the specific mathematics to be able to use the tools in this book. The topics covered use some algebra, calculus and probability.

1.2 WHAT DO WE MEAN BY APPLIED MATHEMATICS?

We consider this book to be a book on applied mathematics. This is not however a universal term, for some applied mathematics is the study of mechanics and involves modelling projectiles being fired out of canons. We will use the term a bit more freely here and mean any type of real world problem that can be tackled using mathematical tools. This is sometimes referred to as operational research, operations research, mathematical modelling or indeed just mathematics.

One of the authors, Vince, used mathematics to plan the sitting plan at his wedding. Using a particular area of mathematics call graph theory he was able to ensure that everyone sat next to someone they liked and/or knew.

The other author, Geraint, used mathematics to find the best team of Pokemon. Using an area of mathematics call linear programming which is based on linear algebra he was able to find the best makeup of pokemon.

Here, applied mathematics is the type of mathematics that helps us answer questions that the real world asks.

1.3 WHAT IS OPEN SOURCE SOFTWARE

Strictly speaking open source software is software with source code that anyone can read, modify and improve. In practice this means that you do not need to pay to use it which is often one of the first attractions. This financial aspect can also be one of the reasons that someone will not use a particular piece of software due to a confusion between cost and value: if something is free is it really going to be any good?

In practice open source software is used all of the world and powers some of the most important infrastructure around. For example, one should never use any cryptographic software that is not open source: if you cannot open up and read things than you should not trust it (this is indeed why most cryptographic systems used are open source).

Today, open source software is a lot more than a licensing agreement: it is a community of practice. Bugs are fixed faster, research is implemented immediately and knowledge is spread more widely thanks to open source software. Bugs are fixed faster because anyone can read and inspect the source code. Most open source software projects also have a clear mechanisms for communicating with the developers and even reviewing and accepting code contributions from the general public. Research is implemented immediately because when new algorithms are discovered they are often added directly to the software by the researchers who found them. This all contributes to the spread of knowledge: open source software is the modern should of giants that we all stand on.

Open source software is software that, like scientific knowledge is not restricted in its use.

1.4 HOW TO GET THE MOST OUT OF THIS BOOK

The book itself is open source. You can find the source files for this book online at `github.com/drvinceknight/ampwoss`. There will will also find a number of *Jupyter notebooks* and *R markdown files* that include code snippets that let you follow along.

We feel that you can choose to read the book from cover to cover, writing out the code examples as you go; or it could also be used as a reference text when faced with particular problem and wanting to know where to start.

The book is made up of 10 chapters that are paired in two 4 parts. Each part corresponds to a particular area of mathematics, for example “Emergent Behaviour”. Two chapters are paired together for each chapter, usually these two chapters correspond to the same area of mathematics but from a slightly different scale that correspond to different ways of tackling the problem.

Every chapter has the following structure:

1. Introduction - a brief overview of a given problem type. Here we will describe the problem at hand in general terms.

2. An Example problem. This will provide a tangible example problem that offers the reader some intuition for the rest of the discussion.

3. Solving with Python. We will describe the mathematical tools available to us in a programming language called Python to solve the problem.

4. Solving with R. Here we will do the same with the R programming language.

5. Brief theoretic background with pointers to reference texts. Some readers might like to delve in to the mathematics of the problem a bit further, we will include those details here.

6. Examples of research using these methods. Finally, some readers might even be interested in finding out a bit more of what mathematicians are doing on these problems. Often this will include some descriptions of the problem considered but perhaps at a much larger scale than the one presented in the example.

For a given reader, not all sections of a chapter will be of interest. Perhaps a reader is only interested in R and finding out more about the research. Please do take from the book what you find useful.

CHAPTER 2

Software

THIS book will involve using software, the particular interface to software we will use is to write code. There are numerous reasons why this is the correct way to do things but one of them is reproducibility.

This chapter will go over the basics of getting your computer set up to use the software: the programming languages R and Python. It will also briefly discuss using the command line: a particular interface to your whole compute and finally it will give a brief introduction to R and Python.

What this Chapter (and indeed this whole book) is not is a place to learn R and Python completely. We will cover very specific tasks and how to carry them out but we will not cover the entire intricacies of each language. There are numerous places (books, websites, courses) that are available to do that. A lot of these places would argue that you should not learn multiple programming books from one book and instead concentrate on a single skill at a time. We agree and the single skill to concentrate on with this book is the use of software to solve applied mathematical problems. The software itself is not the most important component.

2.1 SOFTWARE INSTALLATION

There are a number of different places from which you can buy your vegetables, you can grow them yourself, you can go to a market and pick fresh fruit from specific stalls, you can go to a supermarket and buy a bag of a collection of vegetables and in some places you can even get a box of vegetables regularly posted to you. Software is similar, there are a variety of places from which you can get it and a number of different forms in which it can be obtained.

IF you're comfortable with using R and Python then you probably do not need to read this section and you might even use different so called "distributions" of each piece of software but for the purpose of this book here is where we will be getting what we need:

- Python: we will use the Anaconda distribution: `https://www.anaconda.com/distribution/`
- R: we will be getting this directly from the Comprehensive R Archive Network (commonly referred to as CRAN): `https://cran.r-project.org`. We will also use another piece of software called Rstudio: `https://rstudio.com`.

2.1.1 Installing Python

Installing Python and all the software we need around it is done by downloading and running the installer for the Anaconda distribution.

1. Go to this webpage: `https://www.anaconda.com/download/`.
2. Identify and download the version of Python 3 for your operating system (Windows, Mac OSX, Linux). Run the installer.

2.1.2 Installing R

There are actually two pieces of software we need to install to use R for the purposes of this book, first the R language itself and second an application with which we will write R code.

1. Go to this webpage: `https://cran.r-project.org`.
2. Identify and download the latest version of R for your operating system (Windows, Mac OSX, Linux). Run the installer.
3. Go to this webpage: `https://rstudio.com`.
4. Identify and download the latest version of Rstudio for your operating system (Windows, Mac OSX, Linux). Run the installer.

2.2 USING THE COMMAND LINE

There are various interfaces to using a computer, the most common one is to use a mouse and keyboard and click on programmes we want to use. Another approach is to use what is called a command line interface this is where we do not interact visually with a computer but we type in specific commands.

We can use our command line to navigate the various directories on our computer.

There are two types of operating systems that we consider here:

- Windows
- Nix: this includes OSX (the Mac operating system) and Linux

Not all commands are the same on each type of operating system.

So let us start by opening our command line interface:

- Windows: after having installed Anaconda look to open the AnacondaPrompt. There are a number of other command line interfaces available but this is the one we recommend for the purposes of this book.
- Nix: look to open the Terminal.

This should open something that looks like and somewhat resembles a black box with some text in it. This is where we will write our commands to the computer.

For example to list the contents of the directory we are currently in:

On nix:

Cli input

```
ls
```

On Windows

Cli input

```
dir
```

It is also possible to get the name of the directory we are currently in:

On nix:

Cli input

```
pwd
```

On Windows

Cli input

```
cd
```

Finally we can also use the command line to move to another directory. The command for this are the same on Nix and on Windows.

Cli input

```
cd <name_of_subdirectory>
```

The command line is an important tool to learn to use when doing tasks:

- If we want to scale the tasks, a commonly heard phrase is that 'mouse clicks do not scale' highlighting that to repeat a task many times when using a graphical interface is inefficient.
- If we want someone else to be able to repeat the tasks, we can use screenshots of graphical interfaces but there will always be a level of ambiguity whereas the commands use in the command line are precise.

We can use our two programming languages right within the command line interface (we will actually be using a different tool that we will describe shortly).

To use Python, simply type the following and press Enter:

Cli input

```
python
```

This should make something like the following appear:

Cli output

```
Python 3.7.1 | packaged by conda-forge | (default, Nov 13 2018, 10:30:07)
[Clang 4.0.1 (tags/RELEASE_401/final)] :: Anaconda, Inc. on darwin
Type "help", "copyright", "credits" or "license" for more information.
>>>
```

The >>> is a prompt ready to accept a Python command. Let us start with the following:

Python input

```
>>> 2 + 2
```

When you press Enter, this will give:

Python output

```
4
```

This particular way of using Python is called a REPL which stands for: 'Read eval print loop' which indicates that it takes a command, evaluates it and waits for the next one.

To quit Python's REPL type the following (note that `()`, more about that later):

Python input

```
>>> quit()
```

We can do the same for R. To start R's REPL, in your command line type the following and press Enter:

Cli input

```
R
```

This should make something like the following appear:

Cli output

```
R version 3.5.1 (2018-07-02) -- "Feather Spray"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin13.4.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

>
```

The > is a prompt ready to accept an R command. Let us start with the following:

R input

```
> 2 + 2
```

When you press Enter, this will give:

R output

```
4
```

To quit R's REPL type the following:

R input

```
> q()
```

This will bring up a further prompt asking you to save some information about what you just did. You can type `n` for now:

R input

```
> Save workspace image? [y/n/c]: n
```

These two REPLs are not unique and also not the most efficient way of using the languages however they can at times be useful if you just want to type a very short command or perhaps check something quickly.

Another approach is to save a collection of commands in a plain text file and pass it to the interpreter at the command line.

For example, if we had a number of Python commands in `main.py` we could run this at the command line using:

Cli input

```
python main.py
```

Similarly for a file with a number of R commands `main.R`:

Cli input

```
Rscript main.R
```

These are just a few of many ways to use Python and R, an important notion to understand is that Python and R are not the particular tools that we use to interface to them. On a day to day basis the authors of this book will use both of the above approaches as well as the next ones, we recommend readers take time to experiment and understand the particular use cases for which each tool works best for them.

The two tools we recommend to use in this book are:

- For Python: the Jupyter notebook, a tool that behaves similarly to a REPL, runs in the web browser and is very popular in research.
- For R: RStudio, an integrated development environment with a lot of helpful features.

The best way to start the Jupyter notebook is to type the following in your command line:

Cli input

```
jupyter notebook
```

This will create a *notebook server* that runs on your computer and should open a page that looks like Note that despite running in a web browser this does not need the internet to run.

We can create a new notebook and write and run code in the *cells*.

To start Rstudio, locate the application on your computer and double click on it. This will open an application that looks like

Rstudio includes its on REPL, so we can type and run single commands there but we can also write in a file that we can run

In the next sections we will cover some basics of Python and R.

2.3 BASIC PYTHON

This section gives a very brief overview of some introductory aspects of Python, there are excellent resources available for learning Python and we recommend the reader goes there if they feel they need an in depth understanding of the language

In the previous section, we saw how to get Python to perform a single calculation:

Python input

```python
3 + 5
```

which will give:

Python output

```
8
```

We can also assign values to a variable:

Python input

```python
a = 3
b = 5
c = a + b
c
```

This makes a point at 3 etc...

which will give:

Python output

```
8
```

There are a number of different types of variables in Python, here is a very brief list of some of them:

- Integers – `int` – for example `2`, `4`, `-459060`.
- Floats – `float` – for example `2.0`, `3.4`, `-3.459060`.
- Strings – `str` – for example `"two"`, `"hello world"`, `"3450"`.
- Booleans – `bool` – for example `True` or `False`.

Based on the values of a variable it is possible to construct Booleans:

Python input

```
is_a_larger_than_b = a > b
```

The variable `is_a_larger_than_b` will be the boolean variable `False`.

This is an important concept as boolean variable allow us to use conditional statements that let us write code that does specific things based on the value of variables. For example the following code will add 5 to the smallest variable:

Python input

```
a = 3
b = 5
if a < b:
    a = a + 3
elif a > b:
    b = b + 5
else:
    a = a + 3
    b = b + 3
a, b
```

which gives:

Python output

```
6, 5
```

If you are experimenting by typing the code as you go change the value of `a` or `b` to see how the behaviour changes. What happens if they are equal?

It is also possible to use these conditional statements to repeat code. For example the following code will repeatedly add 1 to the smallest variable until it becomes equal to the largest one:

Python input

```python
a = 3
b = 5
while a != b:
    if a < b:
        a = a + 1
    else:
        b = b + 1
```

It is important to be able to reuse code, this is done using a programming concept called a *function*, which acts similarly to a mathematical function.

The following code, creates a function that takes two variables as input and outputs the largest number and the smallest increased by 3.

Python input

```python
def add_3_to_smallest(a, b):
    """
    This function adds 3 to the smallest of a or b.
    """
    if a < b:
        return a + 3, b
    return a, b + 3
```

Once we have defined the function, the following is how we use it:

Python input

```python
add_3_to_smallest(a=5, b=-42)
```

which gives:

Python output

```
5, -39
```

Python has a type of variable that is in fact a collection of pointers to other variables. This is called a list. Here for example is a collection of strings:

Python input

```
tennis_players = ["Federer", "S. Williams", "V. Williams", "King"]
```

There are a number of things that can be done with lists but one particular aspect is that they are a sub type of something called an iterable in Python which means we can iterate over them. We do this in Python using a `for` loop. For example, the following code will iterate over the list and print all the values:

Python input

```
for name in tennis_players:
    print(name)
```

which gives:

Python output

```
"Federer"
"S. Williams"
"V. Williams"
"King"
```

We will often want to iterate over a set of integers, Python has a `range` command that can create such a set with ease. The following code will print every 3 integers from 30 to 50:

Python input

```
for integer in range(30, 50, 3):
    print(integer)
```

which will give:

Python output

```
30
33
36
39
42
45
48
```

A final important aspect of Python is that of libraries. The code examples above are from the so called ‘standard library’ but Python has numerous libraries specific to given problems. A lot of these libraries came bundled with the anaconda distribution but if you want to download one that is not you can always do so as long as you have an internet connection.

For example, to download a library for studying queueing systems `ciw` open your command line interface and type the following:

```
pip install ciw
```

Once you restart your python interpreter, for exampl if you are using a Jupyter notebook then restart the Kernel and then you can run the following to make `ciw` available to you:

Python input

```python
import ciw
```

2.4 BASIC R

This section gives a very brief overview of some introductory aspects of R, there are excellent resources available for learning R [1] and we recommend the reader goes there if they feel they need an in depth understanding of the language

In the previous section, we saw how to get R to perform a single calculation:

R input

```r
3 + 5
```

which will give:

R output

```
8
```

We can also assign values to a variable:

R input

```
a <- 3
b <- 5
c <- a + b
c
```

which will give:

Python output

```
8
```

An important difference between R and Python is that in R the base structure is in fact a vector, even if it only contains a single variable. We can use the `c` command to *concatenate* these base structures together:

R input

```
c(a, 4)
```

giving:

R output

```
3 8
```

There are a number of different types of variables in R, here is a very brief list of some of them:

- Integers – `integer` – for example `2`, `4`, `-459060`.
- Floats – `double` – for example `2.0`, `3.4`, `-3.459060`.
- Strings – `character` – for example `"two"`, `"hello world"`, `"3450"`.

- Booleans – `logical` – for example **TRUE** or **FALSE**.

Based on the values of a variable it is possible to construct Booleans:

R input

```
is_a_larger_than_b <- a > b
```

The variable `is_a_larger_than_b` will be the boolean variable **FALSE**.

This is an important concept as boolean variable allow us to use conditional statements that let us write code that does specific things based on the value of variables. For example the following code will add 5 to the smallest variable:

R input

```
a <- 3
b <- 5
if (a < b) {
  a <- a + 3
} else if (a > b) {
  b <- b + 3
} else {
  a <- a + 3
  b <- b + 3
}
c(a, b)
```

which gives:

R output

```
6 5
```

If you are experimenting by typing the code as you go change the value of `a` or `b` to see how the behaviour changes. What happens if they are equal?

R is a so called “vectorized” language which means that there is often a more appropriate approach to doing things repeatedly using vectors. This applies to the `if` statement in that there exists a `ifelse` statement that applies to vectors of booleans. For example:

R input

```
booleans <- c(FALSE, TRUE, FALSE, FALSE)
ifelse(booleans, "cat", "dog")
```

which gives:

R output

```
"dog" "cat" "dog" "dog"
```

It is also possible to use conditional statements to repeat code. For example the following code will repeatedly add 1 to the smallest variable until it becomes equal to the largest one:

R input

```
a <- 3
b <- 5
while (a != b) {
  if (a < b) {
    a <- a + 1
  }
  else {
    b <- b + 1
  }
}
```

It is important to be able to reuse code, this is done using a programming concept called a *function*, which acts similarly to a mathematical function.

The following code, creates a function that takes two variables as input and outputs the largest number and the smallest increased by 3.

R input

```r
add_3_to_smallest <- function(a, b) {
  # This function adds 3 to the smallest of a or b.
  if (a < b) {
    return(c(a + 3, b))
  }
  else {
    return(c(a, b + 3))
  }
}
```

Note that R will implicitly return the last computed expression without the need for a `return` statement. So the above can also be written as:

R input

```r
add_3_to_smallest <- function(a, b) {
  # This function adds 3 to the smallest of a or b.
  if (a < b) {
    c(a + 3, b)
  }
  else {
    c(a, b + 3)
  }
}
```

Once we have defined the function, the following is how we use it:

R input

```r
add_3_to_smallest(a=5, b=-42)
```

which gives:

R output

```
5, -39
```

It is possible to iterate over elements inside R vectors:

R input

```r
tennis_players <- c("Federer", "S. Williams", "V. Williams", "King")
```

The following will print all the names contained in the vector:

R input

```r
for (name in tennis_players) {
    print(name)
}
```

which gives:

R output

```
"Federer"
"S. Williams"
"V. Williams"
"King"
```

We will often want to iterate over a vector of integers, R has a `seq` command that can create such a vector with ease. The following code will print every 3 integers from 30 to 50:

R input

```r
for (i in seq(30, 50, 3)) {
  print(i)
}
```

which will give:

R output

```
30
33
36
39
42
45
48
```

A final important aspect of R is that of packages. The code examples above are from the so called ‘base R’ but R has numerous packages specific to given problems. If you want to download and use one you can always do so as long as you have an internet connection.

For example, to download a very common collection of data science tools called `tidyverse` we use the following line of code inside of an R session:

R input

```
install.packages{"tidyverse")
```

Once this package is installed you load it using

R input

```
library(tidyverse)
```

2.5 A NOTE ON HOW CODE IS DISPLAYED IN THIS BOOK

FURTHER READING

Becskei, A. and Serrano, L. (2000). Engineering stability in gene networks by autoregulation. *Nature,* 405: 590–593.

Rosenfeld, N., Elowitz, M.B., and Alon, U. (2002). Negative auto-regulation speeds the response time of transcription networks. *J. Mol. Biol.*, 323: 785–793.

Savageau, M.A. (1976). *Biochemical Systems Analysis: A study of Function and Design in Molecular Biology.* Addison-Wesley. Chap. 16.

Savageau, M.A. (1974). Comparison of classical and auto-genous systems of regulation in inducible operons. *Nature*, 252: 546–549.

Bibliography

[1] Hadley Wickham. *Advanced r*. Chapman and Hall/CRC, 2014.