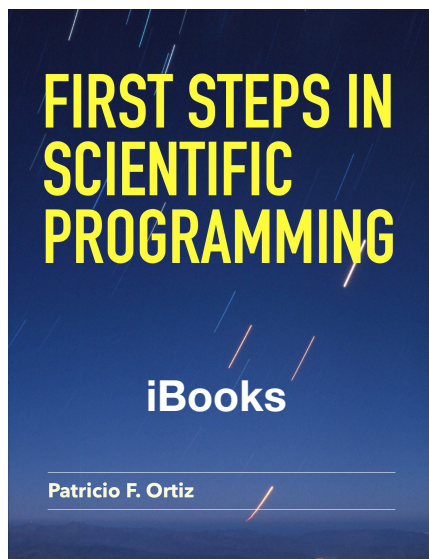


# First steps in Scientific Programming



This book was conceived as a quick yet rich guide for anyone starting to program in (physical) sciences. Through the years, the author has found out that undergrads, graduate students and post-docs sooner or later face the task of having to do some programming to help their research. Science and research are at the forefront of knowledge, which means that on many occasions there are no computational tools to deal with the problems faced. Hence, the only path is to learn how to do it, do it quickly and do it efficiently. In a high percentage of the cases, young scientists have exposure to one computer language only, barely touching practical subjects. Today's reality is that people in a research environment have to learn to code for parallel super-machines or even programming for resources on the Cloud. Most of those systems are base on Unix-derived operating systems (mostly Linux).

This book covers issues of programming in a generic way, not tied to any particular language because once people learn the principles of programming, then what language is used becomes a secondary issue. Concepts which deal with how computers work are covered, like byte order, and internal representation. And, of course, what tools are at the user's disposal to tackle from simple to highly complex tasks. Issues particularly useful in the scientific context are covered, like testing code, performance enhancers, code scalability, debugging, working with remote machines, working with time, coordinate systems, and tips on data processing.

## Patricio F. Ortiz



Patricio Ortiz holds a PhD. in astronomy from the University of Toronto, Canada. He has a keen interest in programming as a mean to create tools to help his research when no tools were available. He has taught at the graduate and undergraduate levels in subjects about astronomy, instrumentation, and applied programming. Throughout his career, Patricio has interacted with students at any level as well as post-graduates, helping him identify the most critical subjects needed by young scientists in the physical sciences and usually not covered by current literature. He has worked on projects involving automated super-nova detection systems; detection of fast moving solar system bodies, including Near-Earth objects and he was involved in the Gaia project (European Space Agency) for nearly ten years. Patricio also developed an ontology system used since its conception by the astronomical community to identify equivalent quantities. He also worked on an Earth Observation project, which gave him the opportunity to work extensively with high-performance computers, leading to his development of an automated task submission system which significantly decreases the execution time of data reduction of extended missions.

Patricio now works as a Research Software Engineer at the [Department of Automatic Control and Systems Engineering at the University of Sheffield](#). He uses C, Fortran, Python, Java and Perl as his main toolkits, and as a pragmatic person, he uses the language which suits a problem best. Amongst his interests are: scientific data visualisation as a discovery tool, photography and (human) languages.

# Contents:

## **1 [The modern computer](#)**

- [1.1 CPU](#)
- [1.2 Memory](#)
- [1.3 GPU](#)
- [1.4 Permanent storage](#)
- [1.5 The operating system \(OS\)](#)

## **2 [How computers store information](#)**

- [2.1 How to organise the disks: paths](#)
- [2.2 Internal representation](#)
- [2.3 Data-type conversion](#)
- [2.4 Input/Output](#)
- [2.5 The price of I/O](#)
- [2.6 File permissions](#)
- [2.7 Number of files in a directory](#)
- [2.8 Pipe operations](#)

## **3 [Which language should I use?](#)**

- [3.1 Languages to consider](#)
- [3.2 Combining languages](#)
- [3.3 Languages recap](#)

## **4 [What software can and can't do](#)**

- [4.1 What software can do](#)
- [4.2 What software cannot do](#)

## **5 [How to write \(scientific\) software](#)**

- [5.1 Planning before coding](#)
- [5.2 From typed code to executables](#)
- [5.3 Build automation tools or make et al.](#)
- [5.4 Protecting your code](#)

## **6 [Main software elements/tools](#)**

- [6.1 Assignment](#)
- [6.2 Flow control: Conditionals](#)
- [6.3 Beware of the = sign](#)
- [6.4 Flow control: Loops](#)
- [6.5 Exception handling](#)
- [6.6 Methods / subprograms](#)
- [6.7 Memory contents and usage](#)
- [6.8 Memory management](#)
- [6.9 Commenting your code](#)
- [6.10 Code structures and Scope](#)
- [6.11 String handling](#)
- [6.12 Basics of flowcharts](#)
- [6.13 UML: a serious design tool](#)

## **7 [Interfaces](#)**

- [7.1 Software Interface](#)
- [7.2 Data interface](#)
- [7.3 Naming files is important](#)
- [7.4 How to store/distribute files on disk](#)

## **8 [Demo code vs production code](#)**

## **9 [Altering someone else's code](#)**

## **10 [Finding problems in the code](#)**

- [10.1 Debugging](#)
- [10.2 Keep historical records](#)

## **11 [Testing code](#)**

- [11.1 Performance test](#)
- [11.2 Test for memory usage and leaks](#)
- [11.3 Assume nothing](#)
- [11.4 Individual variable testing](#)
- [11.5 Visual testing](#)

## **12 [Performance Enhancers](#)**

- [12.1 Look-Up tables \(LUT\)](#)
- [12.2 Avoid repetition](#)
- [12.3 Avoid expensive operations:](#)
- [12.4 Looking for minimum/maximum](#)
- [12.5 Create new elements only when necessary](#)
- [12.6 Make your searches smart operations](#)
- [12.7 Numerical compression](#)
- [12.8 Compression by packing files](#)
- [12.9 Be aware of overheads](#)

## **13 [Code scalability](#)**

## **14 [Working in parallel environments](#)**

- [14.1 Generating unique file names](#)

## **15 [Working with remote computers](#)**

- [15.1 ftp, sftp, scp, rsync, wget](#)
- [15.2 Connecting to a remote machine](#)

## **16 [Unix basics](#)**

- [16.1 Shells](#)
- [16.2 Some \(very\) useful shell features](#)
- [16.3 The terminal](#)
- [16.4 Various command line tools](#)
- [16.5 Unix command examples](#)
- [16.6 Links: Symbolic links and Hard links](#)

## **17 [Automated execution](#)**

## **18 [Random numbers](#)**

## **19 [Working with time](#)**

- [19.1 Is there only one way to measure time?](#)
- [19.2 Different flavours of time](#)
- [19.3 Leap years](#)
- [19.4 Leap seconds](#)
- [19.5 Time differences and usage of Julian Day](#)
- [19.6 Keeping time in your data variables](#)
- [19.7 Displaying time in graphs](#)

## **20 [Coordinate systems](#)**

- [20.1 Cartesian coordinates](#)
- [20.2 Spherical coordinates](#)
- [20.3 Projections from the Sphere to a plane: maps](#)
- [20.4 Angular variables as a function of time](#)
- [20.5 Angular variables: Spherical Trigonometry or vectorial algebra?](#)

## **21 [Data Processing](#)**

## **22 [Databases: the basics](#)**

- [22.1 What is a database?](#)
- [22.2 When is a database necessary?](#)
- [22.3 The choice of a database management system](#)