# Contents

1 Getting Started ................................................................. 4
   1.1 Course Description .................................................. 4
   1.2 Install Python ....................................................... 4
   1.3 Install LibreOffice .................................................. 5
   1.4 Weekly Reports ........................................................ 5
      1.4.1 Graphics ......................................................... 6
      1.4.2 Fonts .............................................................. 6
      1.4.3 Using Styles .................................................... 6
   1.5 IPython Notebook ..................................................... 6
      1.5.1 Magics ............................................................. 8
      1.5.2 Markdown ........................................................ 9
      1.5.3 LaTeX ................................................................ 9

2 Programming Python ........................................................ 10
   2.1 Numbers .................................................................. 10
   2.2 Booleans ................................................................ 11
   2.3 Strings .................................................................. 11
   2.4 Type Conversions .................................................... 13
   2.5 Variable Names ....................................................... 14
   2.6 Modules .................................................................. 15
   2.7 Lists .................................................................. 17
   2.8 Tuples .................................................................. 19
   2.9 Sets .................................................................. 20
   2.10 Dictionaries ............................................................ 21
   2.11 Boolean Expressions ............................................... 22
   2.12 If Statements .......................................................... 23
   2.13 For Loops ............................................................... 25
   2.14 While Loops ............................................................. 27
   2.15 Break and Continue .................................................. 27
   2.16 Comprehensions ....................................................... 28
   2.17 Functions ............................................................... 28
   2.18 Reading and Writing Files ......................................... 31
   2.19 Comments ............................................................... 32
1 Getting Started

1.1 Course Description

This course covers the following areas:

- Programming using Python, the scientific computing package NumPy, and the plotting library Matplotlib.
- Scientific computing methods used in number theory, random number generation, initial value problems, dynamical systems, root-finding, linear regression and optimization.
- Using computers to explore topics in the mathematical and natural sciences.
- Presentation of experiments, observations and conclusions in the form of written reports.

1.2 Install Python

We will be using Python 2.7. It is recommended that you use the Anaconda distribution, which is available free on Windows, Mac and Linux and contains all the packages we need (NumPy, SciPy, IPython, Matplotlib).
Weekly reports must be written using LibreOffice Writer. LibreOffice is a free, cross-platform suite of applications including the word processor Writer, which is similar to Microsoft Word.

1.4 Weekly Reports

Reports will be submitted every week on UBlearns as a file “⟨lastname⟩⟨number⟩.odt” (e.g. “cunningham01.odt”). The final report will be a pdf compilation of all the weekly reports, including a title page and table of contents.

Reports usually need to include:

- An introduction. The topic should explained in a way that would be comprehensible to another member of the class.
- A clear statement of the specific question or task.
- A description of the approach used to tackle the question or task.
- Clearly presented results, including appropriate diagrams and plots.
- An interpretation of the results.
- An appropriate conclusion.
- If appropriate, a list of references to books, articles and websites consulted.
- A complete listing of the Python code used to generate all figures and data used in the report. This should be presented in a way that can be directly copied, pasted, and run by the instructor.

Extra credit will be assigned for extra or unusual work or insight.
CHAPTER 1. GETTING STARTED

1.5. IPYTHON NOTEBOOK

1.4.1 Graphics

Reports are graphics-intensive, and will contain a lot of embedded images. Plots made with Matplotlib can be saved as .png (Portable Network Graphics) files, then embedded in a LibreOffice document using Insert → Image → From File. Make sure that Insert as Link is not checked (this should be the default anyway). Reports are submitted on UBlearns as a single document, so they should contain no dependencies on other files.

Graphs should be labelled in a way that makes the content of the graph clear. They should include the following information (at a minimum):

- Labels for the x- and y-axes, specifying the units when displaying quantities.
- A legend below the graph. Although graphs can be given a title when the graph is created, this information then becomes part of the graph and can’t be changed later. Adding this information later using LibreOffice means that the font and style will be consistent across the report, and the description can be easily changed.
- A number for the legend, which can be used to refer to the graph in the report.

1.4.2 Fonts

Feel free to change the default fonts to suit your own tastes. However:

- Try not to mix up too many different fonts in one document. A good rule of thumb is to limit the number of different fonts to two or three.
- Titles, subtitles etc can be a different font to the body text.
- Use a fixed-width font such as Courier for computer code. Indenting matters in Python, and is an intrinsic part of the language. Using a fixed-width font ensures that code blocks that are supposed to line up actually do line up on the screen.

1.4.3 Using Styles

Any changes to fonts, font sizes etc should be done by modifying existing styles or creating new ones. This will keep a consistent look across the whole report, save the work of repeated font changes, and allow changes to a style to be made in just one place.

- Modify the built-in styles using Format → Styles and Formatting, then right click Modify on the style.
- Create a new style (e.g. for Python code) using Format → Styles and Formatting, then right click New on the style you want to inherit from.

1.5 IPython Notebook

The development environment we will be using is IPython notebook. This provides:

- An interactive environment for writing and running code.
- A way to integrate code, text and graphics in a single document.
A new IPython notebook opens in a web browser, and looks as follows:

The notebook consists of:

- A **Title** bar, containing the name of the notebook.
- A **Menu** bar, containing all the functions available in the notebook.
- A **Tool** bar, for easy access to the most commonly-used functions.
- A list of cells, containing code or text, and the results of executing the code.

The menu bar and tool bar contain the following functions.

- Open, save, and modify notebooks
- Edit, move, and select cells
- Toggle header and toolbar
- Insert new cells into notebook
- Run cells and set cell type
- Interrupt or restart Python
- Help with Python and libraries
- File
- Edit
- View
- Insert
- Cell
- Kernel
- Help
- Save book
- Open cell
- Cut cell
- Paste cell
- Insert cell above
- Insert cell below
- Run cell
- Halt cell
- Set celltype
Code and text are entered in cells, which come in three important types:

- **Code** cells, which contain Python code.
  - Click on a cell with the mouse to start entering code.
  - Enter adds a new line in the cell, without executing the code.
  - Shift-Enter (or clicking the “Play” button in the toolbar, or **Cell → Run** in the menubar) executes the code in the cell and moves the cursor to the next cell.
  - Tab brings up help for the function the cursor is currently in.

- **Markdown** cells contain text formatted using the Markdown language, and mathematical formulas defined using LaTeX math syntax.

- **Heading** cells contain headings to organize the notebook.

The type can be selected by either using the “Cell Type” pull-down menu in the toolbar, or **Cell → Cell Type** in the menubar.

### 1.5.1 Magics

Magics are instructions to IPython that perform specialized tasks. They are entered and executed in code cells, and prefaced by “%” for a line magic (which just applies to one line) or “%%” for a cell magic (which applies to the whole cell). The main ones we will be using are:

- `%pylab inline` imports numpy and matplotlib (making the functions and variables in these modules available to us), with plots drawn inline (in the notebook itself).

- `%pylab` imports numpy and matplotlib, with plots drawn in separate windows so they can be resized and saved.

- `%run ⟨file⟩` executes the Python commands in ⟨file⟩.

- `%timeit ⟨code⟩` records the time it takes to run a line of Python code.

- `%%timeit` records the time it takes to run all the Python code in a cell.

An example of timing code execution using `%timeit` is as follows,

<table>
<thead>
<tr>
<th>Python Code</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>In [1]: <code>%timeit x = range(100000)</code></td>
<td>The first line “x = range(10000)” is run once but not timed. The “max(x)” code is timed</td>
</tr>
<tr>
<td><code>max(x)</code></td>
<td>1000 loops, best of 3: 382 µs per loop</td>
</tr>
</tbody>
</table>
15.2. Markdown

Text can be added to IPython notebooks using Markdown cells. Markdown is a language that can be used to specify formatted text such as italic and bold text, lists, hyperlinks, tables and images. Some examples are shown below.

<table>
<thead>
<tr>
<th>Markdown</th>
<th>How it prints</th>
</tr>
</thead>
<tbody>
<tr>
<td>An h1 header</td>
<td>An h1 header</td>
</tr>
<tr>
<td>=========</td>
<td>=============</td>
</tr>
<tr>
<td>An h2 header</td>
<td>italic</td>
</tr>
<tr>
<td>=========</td>
<td>--------------</td>
</tr>
<tr>
<td><em>italic</em></td>
<td>bold</td>
</tr>
<tr>
<td><strong>bold</strong></td>
<td></td>
</tr>
<tr>
<td>This is a bullet list</td>
<td>This is a bullet list</td>
</tr>
<tr>
<td>* First item</td>
<td>• First item</td>
</tr>
<tr>
<td>* Second item</td>
<td>• Second item</td>
</tr>
<tr>
<td>This is an enumerated list</td>
<td>This is an enumerated list</td>
</tr>
<tr>
<td>1. First item</td>
<td>1. First item</td>
</tr>
<tr>
<td>2. Second item</td>
<td>2. Second item</td>
</tr>
<tr>
<td>A horizontal line</td>
<td>A horizontal line</td>
</tr>
<tr>
<td>***</td>
<td>***</td>
</tr>
</tbody>
</table>

15.3. LaTeX

Mathematical expressions in Markdown cells are specified using the typesetting language LaTeX. These expressions are identified using: $⟨formula⟩$ for an inline formula (one displayed within a line of text), or codebox$$⟨formula⟩$$ for a larger formula displayed on a separate line. Some of the most useful options are shown below.

<table>
<thead>
<tr>
<th>⟨formula⟩</th>
<th>How it prints</th>
</tr>
</thead>
<tbody>
<tr>
<td>(x^2)</td>
<td>(x^2)</td>
</tr>
<tr>
<td>(x_1)</td>
<td>(x_1)</td>
</tr>
<tr>
<td>(\frac{1}{2})</td>
<td>(\frac{1}{2})</td>
</tr>
<tr>
<td>(\alpha, \beta, \omega)</td>
<td>(\alpha, \beta, \omega)</td>
</tr>
<tr>
<td>(\sum_{i = 1}^n)</td>
<td>(\sum_{i = 1}^n)</td>
</tr>
<tr>
<td>(\int_a^b)</td>
<td>(\int_a^b)</td>
</tr>
<tr>
<td>(\sqrt{a + b})</td>
<td>(\sqrt{a + b})</td>
</tr>
<tr>
<td>(\bar{x})</td>
<td>(\bar{x})</td>
</tr>
<tr>
<td>((x + y)^2)</td>
<td>((x + y)^2)</td>
</tr>
<tr>
<td>{1, 2, \ldots, n}</td>
<td>{1, 2, \ldots, n}</td>
</tr>
</tbody>
</table>
2  Programming Python

Python is a flexible and powerful high-level language that is well suited to scientific and mathematical computing. It has been designed with a clear and expressive syntax with a focus on ensuring that code is readable.

2.1  Numbers

The basic numerical types used in Python are:

- Integers.
- Floats (reals).
- Complex numbers (pairs of floats).

Python will automatically convert numbers from one type to another when appropriate. For example, adding two integers yields an integer, but adding an integer and a float yields a float. The main arithmetic operations are +, −, *, /, **, and %, illustrated below.

<table>
<thead>
<tr>
<th>Python Code</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>In [1]: 3 + 2</td>
<td>Addition</td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
<tr>
<td>In [2]: 3 − 2</td>
<td>Subtraction</td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>In [3]: 3 * 2</td>
<td>Multiplication</td>
</tr>
<tr>
<td>6</td>
<td></td>
</tr>
<tr>
<td>In [4]: 3 / 2</td>
<td>Integer division (truncates - the remainder is discarded)</td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>In [5]: 3. / 2</td>
<td>Float division (a decimal point turns an integer into a float)</td>
</tr>
<tr>
<td>1.5</td>
<td></td>
</tr>
<tr>
<td>In [6]: 3**2</td>
<td>Exponentiation (not (3^2))</td>
</tr>
<tr>
<td>9</td>
<td></td>
</tr>
<tr>
<td>In [7]: 26 % 5</td>
<td>Remainder (modulo division)</td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>In [8]: abs(-88)</td>
<td>abs returns the absolute value</td>
</tr>
<tr>
<td>88</td>
<td></td>
</tr>
</tbody>
</table>
In [9]: 1 + 2j
   (1+2j)  Generate a complex number
           (j is used instead of i)

In [10]: complex(1, 2)
   (1+2j)  Another way to generate a
           complex number

In [11]: (1+2j).real
   1.0    real returns the real part of a
           complex number

In [12]: (1+2j).imag
   1.0    imag returns the imaginary
           part of a complex number

In [13]: abs(3+4j)
   5.0    abs returns the modulus when
           applied to a complex number

Operations are evaluated in standard order - Parentheses, Exponentiation, Multiplication,
Division, Addition, Subtraction. To avoid possible ambiguity, use parentheses to make
the order of evaluation clear.

2.2 Booleans

Python also has a Boolean type, which only takes the values True or False. These also
work like numbers, where True has the value 1 and False the value 0.

<table>
<thead>
<tr>
<th>Python Code</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>In [1]: True or False</td>
<td>Logical disjunction</td>
</tr>
<tr>
<td>True</td>
<td></td>
</tr>
<tr>
<td>In [2]: True and False</td>
<td>Logical conjunction</td>
</tr>
<tr>
<td>False</td>
<td></td>
</tr>
<tr>
<td>In [3]: not True</td>
<td>Logical negation</td>
</tr>
<tr>
<td>False</td>
<td></td>
</tr>
<tr>
<td>In [4]: True + 41</td>
<td>True has the numerical value 1.</td>
</tr>
<tr>
<td>42</td>
<td></td>
</tr>
<tr>
<td>In [5]: False * 41</td>
<td>False has the numerical value 0.</td>
</tr>
<tr>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

2.3 Strings

Strings are sequences of characters. They are identified by surrounding quote marks.

- To generate a string, enclose a sequence of characters in either single (') or double (") quotes (Python doesn’t care which).
- A single character in Python is just a one-element string.
Python strings are immutable - once defined, they can’t be changed. They can of course still be copied or operated on to create new strings.

Python Code | Explanation
--- | ---
In [1]: print "abc" | print outputs text to the screen (discarding the quotes)
abc
In [2]: "abc" + "def" | Adding two strings makes a new string by concatenation
"abcdef"
In [3]: "abc"*3 | Multiplying a string by an integer repeats the string
"abcabcabc"
In [4]: print "I love 'MTH 337'!" | Embedding quote marks within a string
I love 'MTH 337'!

A "\" within a string is used to specify special characters such as newlines and tabs.

Python Code | Explanation
--- | ---
In [1]: string1 = "abc \ndef" | The "\n" character specifies a newline
print string1
abc
def
In [2]: string2 = "abc \tdef" | The "\t" character specifies a tab
print string2
abc   def

Strings elements are accessed using square brackets, [].

- **Indexing** obtains characters from the string using a single integer to identify the position of the character.
- **Slicing** obtains a substring using `start:stop:step` to identify which characters to select.
- Indexing and slicing is *zero-based* - the first character is at position 0.
- Indexing and slicing is “up to but not including” the stop position.
- A “:" can be used to select all characters either before or after a given position.

Python Code | Explanation
--- | ---
In [1]: "abcde"[1] | Indexing returns the character at index 1 (indices start at 0, not 1)
"b"
In [2]: "abcde"[-1]  
   "e"  
   
In [3]: "abcde"[1:4]  
   "bcd"  
   
In [4]: "abcde"[2:]  
   "cde"  
   
In [5]: "abcde"[:2]  
   "ab"  
   
In [6]: "abcde"[::2]  
   "ace"  
   
In [7]: "abcdefg"[1:5:2]  
   "bd"  
   
In [8]: "abcde"[::-1]  
   "edcba"  

Strings can be formatted using the **format** function. This allows “replacement fields” surrounded by curly brackets `{}` in a string to be replaced by some other data.

<table>
<thead>
<tr>
<th>Python Code</th>
<th>Explanation</th>
</tr>
</thead>
</table>
| In [1]: "{} {}".format("a", "b")  
   "a b"  |
| "Replacement fields" `{}` are filled in order by **format** |
| In [2]: "1st: {0}, 2nd: {1}".format(3,4)  
   "1st: 3, 2nd: 4"  |
| The arguments to **format** can also be identified by position, starting at 0 |

### 2.4 Type Conversions

Objects can be explicitly converted from one type to another, as long as the conversion makes sense. This is called **type casting**:

- Ints can be cast to floats, and both ints and floats can be cast to complex numbers.
- Complex numbers can’t be converted to ints or floats.
- Strings can be cast to numerical types, but only if the string represents a valid number.

Casting is done using the functions **bool**, **int**, **float**, and **str**.
### Variable Names

Variable names can be used to refer to objects in Python. They:

- Must start with either a letter or an underscore.
- Are case sensitive. So value, VALUE, and Value all refer to different variables.
- Are assigned a value using “=”. The variable name goes to the left of the “=”, and the value to assign on the right.

<table>
<thead>
<tr>
<th>Python Code</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>In [1]: <code>bool(1)</code></td>
<td>Convert integer to boolean</td>
</tr>
<tr>
<td>True</td>
<td></td>
</tr>
<tr>
<td>In [2]: <code>bool(42)</code></td>
<td>Any nonzero value counts as True</td>
</tr>
<tr>
<td>True</td>
<td></td>
</tr>
<tr>
<td>In [3]: <code>bool(0)</code></td>
<td>Zero equates to False</td>
</tr>
<tr>
<td>False</td>
<td></td>
</tr>
<tr>
<td>In [4]: <code>bool(&quot;&quot;)</code></td>
<td>An empty string is also False</td>
</tr>
<tr>
<td>False</td>
<td></td>
</tr>
<tr>
<td>In [5]: <code>int(2.99)</code></td>
<td>Convert float to integer (the decimal part is discarded)</td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>In [6]: <code>int(&quot;22&quot;)</code></td>
<td>Convert string to int</td>
</tr>
<tr>
<td>22</td>
<td></td>
</tr>
<tr>
<td>In [7]: <code>float(&quot;4.567&quot;)</code></td>
<td>Convert string to float</td>
</tr>
<tr>
<td>4.567</td>
<td></td>
</tr>
<tr>
<td>In [8]: <code>complex(&quot;1+2j&quot;)</code></td>
<td>Convert string to complex</td>
</tr>
<tr>
<td>(1+2j)</td>
<td></td>
</tr>
<tr>
<td>In [9]: <code>float(10)</code></td>
<td>Convert integer to float</td>
</tr>
<tr>
<td>10.0</td>
<td></td>
</tr>
<tr>
<td>In [10]: <code>complex(10)</code></td>
<td>Convert integer to complex number</td>
</tr>
<tr>
<td>(10+0j)</td>
<td></td>
</tr>
<tr>
<td>In [11]: <code>str(True)</code></td>
<td>Convert boolean to string</td>
</tr>
<tr>
<td>&quot;True&quot;</td>
<td></td>
</tr>
<tr>
<td>In [12]: <code>str(1)</code></td>
<td>Convert integer 1 to string</td>
</tr>
<tr>
<td>&quot;1&quot;</td>
<td></td>
</tr>
<tr>
<td>In [13]: <code>str(1.234)</code></td>
<td>Convert float to string</td>
</tr>
<tr>
<td>&quot;1.234&quot;</td>
<td></td>
</tr>
</tbody>
</table>
### 2.6 Modules

A module is a file containing Python definitions and statements. These allow us to use code created by other developers, and greatly extend what we can do with Python. Since many different modules are available, it is possible that the same names are used by different developers. We therefore need a way to identify which module a particular

<table>
<thead>
<tr>
<th>Python Code</th>
<th>Explanation</th>
</tr>
</thead>
</table>
| In [1]: ```python
x = 5
print x
``` | Assign x the value 5 (note that “=” is used for assignment, not “==”) |
| In [2]: ```python
y = x + 3
print y
``` | Assign y the value of x + 3 |
| In [3]: ```python
course = "MTH 337"
p = print course
``` | course is a string (printed without quotes) |
| In [4]: ```python
a, b = 2, 3
print a, b
``` | Multiple variables can be assigned at the same time |
| In [5]: ```python
a, b = b, a
print a, b
``` | Values of a and b are swapped (the right hand side is evaluated before the assignment) |
| In [6]: ```python
z = 3
z += 2
print z
``` | Same as z = z + 2 |
| In [7]: ```python
z -= 1
print z
``` | Same as z = z - 1 |
| In [8]: ```python
z *= 3
print z
``` | Same as z = z * 3 |
| In [9]: ```python
z /= 2
print z
``` | Same as z = z / 2 |
| In [10]: ```python
z %= 5
print z
``` | Same as z = z % 5 |
variable or function came from. The `import` statement is used to make the variables and functions in a module available for use. We can either:

- Import everything from a module for immediate use.
- Import only certain named variables and functions from a module.
- Import everything from a module, but require that variable and function names be prefaced by either the module name or some alias.

<table>
<thead>
<tr>
<th>Python Code</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>In [1]: <code>from math import pi</code></td>
<td><code>print pi</code></td>
</tr>
<tr>
<td></td>
<td>3.14159265359</td>
</tr>
<tr>
<td>In [2]: <code>from math import *</code></td>
<td><code>print e</code></td>
</tr>
<tr>
<td></td>
<td>2.71828182846</td>
</tr>
<tr>
<td>In [3]: <code>import numpy</code></td>
<td><code>print numpy.arcsin(1)</code></td>
</tr>
<tr>
<td></td>
<td>1.57079632679</td>
</tr>
<tr>
<td>In [4]: <code>import numpy as np</code></td>
<td><code>print np.cos(0)</code></td>
</tr>
<tr>
<td></td>
<td>1.0</td>
</tr>
</tbody>
</table>

If we want to know what a module contains, we can use the `dir` function. This returns a list of all the variable and function names in the module.

<table>
<thead>
<tr>
<th>Python Code</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>In [1]: <code>import math</code></td>
<td><code>print dir(math)</code></td>
</tr>
</tbody>
</table>
|                                       | `['__doc__', '__name__', '__package__',
|                                       | 'acos', 'acosh', 'asin', 'asinh',
|                                       | 'atan', 'atan2', 'atanh', 'ceil',
|                                       | 'copysign', 'cos', 'cosh',
|                                       | 'degrees', 'e', 'erf', 'erfc',
|                                       | 'exp', 'expm1', 'fabs', 'factorial',
|                                       | 'floor', 'fmod', 'frexp', 'fsum',
|                                       | 'gamma', 'hypot', 'isinf', 'isnan',
|                                       | 'ldexp', 'lgamma', 'log', 'log10',
|                                       | 'log1p', 'modf', 'pi', 'pow',
|                                       | 'radians', 'sin', 'sinh', 'sqrt',
|                                       | 'tan', 'tanh', 'trunc']              |

The `math` module contains all the standard mathematical functions.
2.7 Lists

List are a type of container - they contain a number of other objects. A list is an ordered sequence of objects, identified by surrounding square brackets, `[]`.

- To generate a list, enclose a sequence of objects (separated by commas) in square brackets.
- List elements can be of any type, and can be of different types within the same list.
- Lists are mutable - once created, elements can be added, replaced or deleted.

List elements can be of any type, and can be of different types within the same list.

Python Code | Explanation
--- | ---
In [1]: mylist = [1, "a", 6.58] print mylist | Use square brackets to create a list
[1, "a", 6.58]
In [2]: len(mylist) | `len` returns the number of elements in a list
3
In [3]: list1 = [1, 2, 3] list2 = [4, 5, 6] list1 + list2 | Adding two lists makes a new list by concatenation
[1, 2, 3, 4, 5, 6]
In [4]: list1 * 3 | Multiplying a list by an integer repeats the list
[1, 2, 3, 1, 2, 3, 1, 2, 3]
In [5]: list3 = [] print list3 | `list3` is an empty list
[]
In [6]: list4 = list() print list4 | Another way to create an empty list
[]

Lists can be indexed and sliced in the same way as strings, using square brackets.

Python Code | Explanation
--- | ---
In [1]: primes = [2, 3, 5, 7, 11, 13, 17] primes[1] | Access the element at index 1 (indexing starts with 0)
3
In [2]: primes[3:] | List slicing, start at position 3, through to the end
[7, 11, 13, 17]
In [3]: primes[:3] | List slicing, start at the beginning, end at position 2
[2, 3, 5]
CHAPTER 2. PROGRAMMING PYTHON

2.7. LISTS

The **range** function generates a list of integers in a given range. The start, stop and step parameters to **range** are similar to those used to slice lists and strings.

<table>
<thead>
<tr>
<th>Python Code</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>In [1]: numbers = range(5) print numbers</td>
<td>range(n) creates a list of (n) consecutive integers, starting at 0</td>
</tr>
<tr>
<td>In [2]: teens = range(13, 20) print teens</td>
<td>The third step argument to range specifies the increment from one integer to the next</td>
</tr>
<tr>
<td>In [3]: evens = range(0, 9, 2) print evens</td>
<td>range can also count backwards using a negative step size</td>
</tr>
<tr>
<td>In [4]: range(10, 0, -1)</td>
<td></td>
</tr>
</tbody>
</table>

List elements can be changed, added, and deleted, modifying an existing list.

<table>
<thead>
<tr>
<th>Python Code</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>In [1]: mylist = [&quot;a&quot;, &quot;b&quot;] mylist.append(&quot;c&quot;) print mylist</td>
<td>append adds an element to the end of a list</td>
</tr>
<tr>
<td>In [2]: del(mylist[1]) print mylist</td>
<td>del deletes an element from a list</td>
</tr>
<tr>
<td>In [3]: mylist.insert(1, &quot;d&quot;) print mylist</td>
<td>insert inserts an element at a given position</td>
</tr>
<tr>
<td>In [4]: mylist[1] = &quot;e&quot; print mylist</td>
<td>List elements can be changed by assigning a new element at a given index</td>
</tr>
</tbody>
</table>

In [4]: primes[2:5] [5, 7, 11] List slicing, start at position 2, end at position 4

In [5]: primes[::-1] [17, 13, 11, 7, 5, 3, 2] One way to reverse a list

In [4]: primes[2:5] List slicing, start at position 2, end at position 4
[5, 7, 11]
In [5]: primes[::-1] One way to reverse a list
[17, 13, 11, 7, 5, 3, 2]
Lists can be sorted and reversed.

<table>
<thead>
<tr>
<th>Python Code</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>In [1]: <code>letters = [&quot;a&quot;, &quot;b&quot;, &quot;c&quot;]</code></td>
<td><code>reverse</code> changing an existing list, reversing the order of elements</td>
</tr>
<tr>
<td><code>letters.reverse()</code></td>
<td></td>
</tr>
<tr>
<td><code>print mylist</code></td>
<td></td>
</tr>
<tr>
<td><code>[&quot;c&quot;, &quot;b&quot;, &quot;a&quot;]</code></td>
<td></td>
</tr>
<tr>
<td>In [2]: <code>numbers = [2, 10, 3, 26, 5]</code></td>
<td><code>sorted</code> returns a sorted list, but does not modify the existing list</td>
</tr>
<tr>
<td><code>print sorted(numbers)</code></td>
<td></td>
</tr>
<tr>
<td><code>[2, 3, 5, 10, 26]</code></td>
<td></td>
</tr>
<tr>
<td>In [3]: <code>numbers.sort()</code></td>
<td><code>sort</code> sorts a list in place, modifying the existing list</td>
</tr>
<tr>
<td><code>print numbers</code></td>
<td></td>
</tr>
<tr>
<td><code>[2, 3, 5, 10, 26]</code></td>
<td></td>
</tr>
<tr>
<td>In [4]: <code>sorted(numbers, reverse=True)</code></td>
<td>The <code>reverse</code> keyword is used to sort in descending order</td>
</tr>
<tr>
<td><code>print 26, 10, 5, 3, 2</code></td>
<td></td>
</tr>
</tbody>
</table>

The `min` and `max` functions find the smallest and largest items in a list.

<table>
<thead>
<tr>
<th>Python Code</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>In [1]: <code>numbers = [2, 10, 3, 26, 5]</code></td>
<td><code>min</code> and <code>max</code> find the smallest and largest items</td>
</tr>
<tr>
<td><code>print min(numbers), max(numbers)</code></td>
<td></td>
</tr>
<tr>
<td><code>2 26</code></td>
<td></td>
</tr>
</tbody>
</table>

2.8 Tuples

Tuples are containers like lists, with the difference being that they are *immutable* - once defined, elements cannot be changed or added. Tuples are identified by surrounding standard parentheses, ().

- To generate a tuple, enclose a sequence of objects (separated by commas) in standard parentheses.
- Tuple indexing and slicing works in the same way as for lists and strings.
- It is an error to try to change a tuple element once the tuple has been created.

Tuples are simpler and more efficient than lists in terms of memory use and performance, and are often preferred for “temporary” variables that will not need to be modified.

<table>
<thead>
<tr>
<th>Python Code</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>In [1]: <code>tuple1 = (&quot;a&quot;, &quot;b&quot;, &quot;c&quot;)</code></td>
<td>Create a tuple using standard parentheses</td>
</tr>
<tr>
<td><code>print tuple1</code></td>
<td></td>
</tr>
<tr>
<td>(&quot;a&quot;, &quot;b&quot;, &quot;c&quot;)</td>
<td></td>
</tr>
</tbody>
</table>
In [2]: tuple1[2]  
"c"  

Elements can be indexed just like lists or strings

In [3]: tuple1[1:]  
("b", "c")  

Slicing works the same way as for lists or strings

Any comma-separated sequence of values defines a tuple, which can be used to assign values to multiple variables at a time.

<table>
<thead>
<tr>
<th>Python Code</th>
<th>Explanation</th>
</tr>
</thead>
</table>
| In [1]: tuple2 = 1, 2, 3  
print tuple2  
(1, 2, 3) | A comma-separated sequence of values defines a tuple |
| In [2]: (x, y) = (10, 20)  
print "x =", x  
print "y =", y  
x = 10  
y = 20 | The variables on the left-hand side are assigned to the values on the right |
| In [3]: a, b = (2, 4)  
print a, b  
2 4 | The parentheses are not strictly necessary, and can be discarded |

2.9 Sets

Sets are containers with the same meaning they do in mathematics - unordered collections of items with no duplicates. Sets are identified by surrounding curly brackets, `{}`.

- To generate a set, enclose a sequence of objects (separated by commas) in curly brackets.
- Duplicates will be removed when creating a set or operating on existing sets.
- Sets can be used instead of lists when we know that each element is unique and immutable (unchanging).

<table>
<thead>
<tr>
<th>Python Code</th>
<th>Explanation</th>
</tr>
</thead>
</table>
| In [1]: myset = {1, 2, 3}  
print myset  
set([1, 2, 3]) | Sets are created using curly brackets |
| In [2]: myset = set([1, 2, 3, 2])  
print myset  
set([1, 2, 3]) | Creating a set from a list (note that duplicates are removed) |
### 2.10. Dictionaries

Dictionaries are containers where items are accessed by a key. This makes them different from sequence type objects such as strings, lists, and tuples, where items are accessed by position.

- To generate a dictionary, enclose a sequence of `key:value` pairs (separated by commas) in curly brackets.
- The key can be any immutable object - a number, string, or tuple.
- New dictionary elements can be added, and existing ones can be changed, by using an assignment statement.
- It is an error to attempt to access a dictionary using a key that does not exist. This can be avoided by using the `get` method, which returns a default value if the key is not found.
- Order is *not* preserved in a dictionary, so printing a dictionary will not necessarily print items in the same order that they were added.

---

<table>
<thead>
<tr>
<th>Python Code</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>In [1]: set1 = {1, 2, 3} set2 = {3, 4, 5}</td>
<td>Create 2 sets</td>
</tr>
<tr>
<td>In [2]: 1 in set1</td>
<td><code>in</code> tests for set membership</td>
</tr>
<tr>
<td></td>
<td>True</td>
</tr>
<tr>
<td>In [3]: set1</td>
<td>Union (the <code>union</code> operator can also be used)</td>
</tr>
<tr>
<td></td>
<td>{1, 2, 3, 4, 5}</td>
</tr>
<tr>
<td>In [4]: set1</td>
<td>Set difference (can also use the <code>difference</code> operator)</td>
</tr>
<tr>
<td></td>
<td>{1, 2}</td>
</tr>
<tr>
<td>In [5]: set1</td>
<td>Set intersection (can also use the <code>intersection</code> operator)</td>
</tr>
<tr>
<td></td>
<td>{3}</td>
</tr>
<tr>
<td>In [6]: set1</td>
<td>Symmetric difference (can also use the <code>symmetric_difference</code> operator)</td>
</tr>
<tr>
<td></td>
<td>{1, 2, 4, 5}</td>
</tr>
<tr>
<td>In [7]: set1 &lt;= set2</td>
<td>Test if one set is a subset of another (can also use the <code>issubset</code> operator)</td>
</tr>
<tr>
<td></td>
<td>False</td>
</tr>
</tbody>
</table>
### 2.11 Boolean Expressions

Boolean expressions are statements that either evaluate to True or False. An important use of these expressions is for tests in conditional code that only executes if some condition is met. Examples of Boolean expressions include the standard comparison operators below.

<table>
<thead>
<tr>
<th>Python Code</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>5 == 5</code></td>
<td>Check for equality</td>
</tr>
<tr>
<td><code>5 != 5</code></td>
<td>Check for inequality</td>
</tr>
<tr>
<td><code>3 &lt; 2</code></td>
<td>Less than</td>
</tr>
</tbody>
</table>
Chapter 2. Programming Python

2.12. If Statements

Python if statements provide a way to execute a block of code only if some condition is True. The syntax is:

```
if <condition>:
    <code to execute when condition True>
<following code>
```

Note that:

- `<condition>` is a Boolean expression, which must evaluate to True or False.
- `<condition>` must be followed by a colon, :.
- The block of code to execute if `<condition>` is True starts on the next line, and must be indented. The convention in Python is that code blocks are indented with 4 spaces.
- This block of code is finished by de-indenting back to the previous level.

### Python Code

```
from math import *
if pi > e:
    print "Pi is bigger than e!"
```

### Explanation

The block of code following the if statement only executes if the condition is met.

Pi is bigger than e!

An else statement can be added after an if statement is complete. This will be followed by the code to execute if the condition is False. The syntax is:

```
if <condition>:
    <code to execute when condition True>
else:
    <code to execute when condition False>
<following code>
```

The following example illustrates an if-else statement.

```
if pi > e:
    print "Pi is bigger than e!"
else:
    print "Pi is smaller than e!"
```

Note that any empty container evaluates to False in a Boolean expression. Examples include empty strings ('"'), lists ([]), and dictionaries ({}).
CHAPTER 2. PROGRAMMING PYTHON

2.12. IF STATEMENTS

Python Code

| In [1]: | x, y = 2**3, 3**2  
        | if x < y:  
        |     print "x < y"  
        | else:  
        |     print "x >= y"  
 |
| --- | --- |

Explanation

The block of code following the else statement executes if the condition is not met.

Multiple elif statements (short for else-if) can be added to create a series of conditions that are tested in turn until one succeeds. Each elif must also be followed by a condition and a colon. The syntax is:

```
if <condition 1>:
    <code to execute when condition 1 True>
elif <condition 2>:
    <code to execute when condition 2 True>
else:
    <code to execute if neither condition is True>
```

Python Code

| In [1]: | score = 88  
        | if score >= 90:  
        |     print "A"  
        | elif score >= 80:  
        |     print "B"  
        | elif score >= 70:  
        |     print "C"  
        | elif score >= 60:  
        |     print "D"  
        | else:  
        |     print "F"  
 |
| --- | --- |

Explanation

Only the first two conditions are tested - the rest are skipped since the second condition is True.

It often happens that we want to assign a variable some value if a condition is True, and another value if a condition is False. Using an if statement, we would have:

```
if <condition>:
    x = <true_value>
else:
    x = <false_value>
```

Python provides an elegant way to do this in a single line using a conditional expression. This has the following syntax.
An example is given below.

```python
x = <true_value> if <condition> else <false_value>
```

Note that \( x \mod 2 \) returns the remainder when \( x \) is divided by 2. Any nonzero value evaluates as True.

### 2.13 For Loops

Python **for** loops provide a way to iterate (loop) over the items in a list, string, tuple, or any other iterable object, executing a block of code on each pass through the loop. The syntax is:

```python
for <iteration variable(s)> in <iterable>:
    <code to execute each time>
    <following code>
```

Note that:

- The **for** statement must be followed by a colon, :.
- One or more iteration variables are bound to the values in \( \langle \text{iterable} \rangle \) on successive passes through the loop.
- The block of code to execute each time through the loop starts on the next line, and must be indented.
- This block of code is finished by de-indenting back to the previous level.

Sequence objects, such as strings, lists and tuples, can be iterated over as follows.

<table>
<thead>
<tr>
<th>Python Code</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>In [1]:</td>
<td></td>
</tr>
<tr>
<td>for i in [2, 4, 6]: print(i)</td>
<td>Iterate over the elements of a list. The iteration variable ( i ) gets bound to each item in turn</td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
</tr>
<tr>
<td>In [2]:</td>
<td></td>
</tr>
<tr>
<td>for char in &quot;abc&quot;: print char</td>
<td>Iterate over the characters in a string. The iteration variable ( char ) gets bound to each character in turn</td>
</tr>
<tr>
<td>a</td>
<td></td>
</tr>
<tr>
<td>b</td>
<td></td>
</tr>
<tr>
<td>c</td>
<td></td>
</tr>
<tr>
<td>Python Code</td>
<td>Explanation</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>In [3]: <code>for i, char in enumerate(&quot;abc&quot;):</code>&lt;br&gt;print i, char</td>
<td><code>enumerate</code> allows an iteration variable to be bound to the index of each item, as well as to the item itself</td>
</tr>
<tr>
<td>0 a&lt;br&gt;1 b&lt;br&gt;2 c</td>
<td>Sum the numbers from 1 to 5. <code>xrange</code> works like <code>range</code> in <code>for</code> loops, but doesn’t generate a list first</td>
</tr>
</tbody>
</table>
| In [4]: `total = 0<br>for i in xrange(1, 6):  
    total += i<br>print total` | Dictionaries elements consist of key:value pairs. When iterated over, variables can be bound to the key, the value, or both. |
| 15 | |
| In [1]: `mydict = {"x" : 1, "y" : 2, "z" : 3}
for key in mydict:
    print key` | Iteration over a dictionary binds to the key (note that order is not preserved in a dictionary) |
| y<br>x<br>z | Use `itervalues` to iterate over the dictionary values rather than the keys |
| In [2]: `for value in mydict.itervalues():
    print value` | |
| 2<br>1<br>3 | Use `iteritems` to iterate over the dictionary keys and values together |
| In [3]: `for key, val in mydict.iteritems():
    print key, val` | |
| y 2<br>x 1<br>z 3 | |
| In [1]: `courses = [141, 142, 337]<br>ranks = ["good", "better", "best!"]
zipped = zip(courses, ranks)<br>print zipped` | The function `zip` can be used to iterate in parallel over multiple lists of equal length. This returns a list of tuples, with one element of each tuple drawn from each list. |
| `[(141, "good"), (142, "better"), (337, "best")]
zip(courses, ranks)` creates a list of tuples. Each tuple contains one course and one rank, with the tuples in the same order as the list elements |
In [2]:

```python
for c, r in zipped:
    print c, r
```

141 good
142 better
337 best!

Multiple iteration variables can be bound at each iteration of a loop

2.14 While Loops

Python `while` loops execute a block of code repeatedly as long as some condition is met. The syntax is:

```python
while <condition>:
    <code to execute repeatedly>
    <following code>
```

Note that for the loop to terminate, the code must change some part of the condition so that it eventually returns `False`.

<table>
<thead>
<tr>
<th>Python Code</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>i = 3</code>&lt;br&gt;<code>while i &gt; 0:</code>&lt;br&gt;<code>    print i</code>&lt;br&gt;<code>    i -= 1</code></td>
<td><code>i</code> is printed while it remains greater than zero. The code must change the value of <code>i</code> to ensure that the loop eventually terminates</td>
</tr>
<tr>
<td>3&lt;br&gt;2&lt;br&gt;1</td>
<td></td>
</tr>
</tbody>
</table>

2.15 Break and Continue

Sometimes we need to end a loop early, either by ending just the current iteration, or by quitting the whole loop. The statements `break` and `continue` provide a way to do this.

- To end the loop completely and jump to the following code, use the `break` statement.
- To end the current iteration and skip to the next item in the loop, use the `continue` statement. This can often help to avoid nested if-else statements.

<table>
<thead>
<tr>
<th>Python Code</th>
<th>Explanation</th>
</tr>
</thead>
</table>
| `vowels = "aeiou"
for char in "bewgfiagf":`<br>`    if char in vowels:`<br>`        print "First vowel is", char`<br>`        break` | The `for` loop is terminated by `break` once the first vowel is found |
| | First vowel is `e` |
CHAPTER 2. PROGRAMMING PYTHON

2.16. COMPREHENSIONS

Often we want to create a container by modifying and filtering the elements of some other container. Comprehensions provide an elegant way to do this, similar to mathematical set-builder notation. For lists, the syntax is:

\[
\left\{ \text{expression} \mid \text{for variables in container if condition} \right\}
\]

The code in \( \langle \text{expression} \rangle \) is evaluated for each item in the \( \langle \text{container} \rangle \), and the result becomes an element of the new list. The \( \langle \text{condition} \rangle \) does not have to be present but, if it is, only elements which satisfy the condition become incorporated into the new list.

### Python Code Explanation

**In [1]:**

\[
[\text{i}**2 \text{ for i in range}(5)]
\]

\[0, 1, 4, 9, 16\]

\( \text{i}**2 \) is evaluated for every item \( \text{i} \) in the list

**In [2]:**

\[
[\text{d for d in range}(1,7) \text{ if } 6 \% d == 0]
\]

\[1, 2, 3, 6\]

Divisors of 6 - only elements passing the test \( 6 \% d == 0 \) are included

We can also use a dictionary comprehension to create a dictionary without needing to repeatedly add key:value pairs.

**In [1]:**

\[
\{\text{i}:\text{i}**2 \text{ for i in range}(4)\}
\]

\{0:0, 1:1, 2:4, 3:9\}

Create a dictionary from a list. Note the key:value pairs and surrounding curly brackets

### 2.17 Functions

Functions provide a way to reuse a block of code by giving it a name. The code can then be executed just by calling the function name, with the option of passing in additional data to be used inside the function. The variables used to identify this additional data are the function **parameters**, and the particular values passed in when the function is called are the function **arguments**.
• Functions take a list of required arguments, identified by position.

• Functions can take keyword arguments, identified by name. These can also be assigned default values in the function definition to use if no values are passed in.

• Functions can return one or more values using the return statement. Note that functions do not have to return a value - they could just perform some action instead. A function stops executing as soon as a return statement is encountered.

• An optional documentation string can be added at the start of the function (before the code) to describe what the function does. This string is usually enclosed in triple quotes.

Functions are defined in Python using the def statement, with the syntax:

\[
\text{def <name> (<parameters>):} \\
\text{<documentation string>} \\
\text{<code>}
\]

The arguments to a function can be specified by position, keyword, or some combination of both. Some examples using just positional arguments are as follows.

<table>
<thead>
<tr>
<th>Python Code</th>
<th>Explanation</th>
</tr>
</thead>
</table>
| In [1]: def square(x):
  return x**2
  print square(3)
>
| The function exits as soon as the return statement is called |
|-------------|-------------|
| 9 |
| In [2]: def multiply(x, y):
  """Return the product xy""
  return x*y
  print multiply(3, 2)
>
| Parameters are bound to input data in the order given. The documentation string is placed after the colon and before the code |
|-------------|-------------|
| 6 |
| In [3]: def minmax(data):
  return min(data), max(data)
  print minmax([1, 3, 7, 2, 10])
>
| Multiple values are returned as a tuple |
|-------------|-------------|
| (1, 10) |

Using keyword arguments allows default values to be assigned. This is particularly useful when a function can be called with many different options, and avoids having to call functions with a long list of arguments.

• Keyword arguments are specified using key=default in place of a positional argument.

• Using keyword instead of positional arguments means we don’t need to remember the order of arguments, and allows the defaults to be used most of the time.
Positional and keyword arguments can be used in the same function, as long as the positional arguments come first.

Python Code

```
In [1]: def close_enough(x, y, tolerance=.1)
    ...:     return abs(x - y) <= tolerance
```

Explanation

The `tolerance` argument is 0.1 by default

```
In [2]: close_enough(1, 1.05)
```

True

The default tolerance of 0.1 is used in this case

```
In [3]: close_enough(1, 1.05, tolerance=.01)
```

False

The default tolerance is overridden by the value of 0.01

If the number of arguments is not known in advance, functions can be defined to take a variable number of positional arguments and/or a variable number of keyword arguments. We are unlikely to be using these options ourselves, although they occur frequently in the documentation for Matplotlib.

- The positional arguments are usually specified as `*args` and are available as a tuple. Individual positional arguments can then be accessed by indexing into the tuple by position.
- The keyword arguments are usually specified as `**kwargs` and are available as a dictionary. Individual keyword arguments can then be accessed by indexing into this dictionary by key.

We may on occasion need to use a simple function in a single place, and not want to have to define and name a separate function for this purpose. In this case we can define an anonymous or `lambda` function just in the place where it is needed. The syntax for a lambda function is:

```
lambda <arguments> : <code>
```

The `lambda` statement returns an unnamed function which takes the arguments given before the colon, and returns the result of executing the code after the colon. Typical uses for lambda functions are where one function needs to be passed in as an argument to a different function.

Python Code

```
In [1]: ages = [21, 19, 150]
   ...: names = ['Bruce', 'Sheila', 'Adam']
   ...: data = zip(ages, names)
   ...: sorted(data, key=lambda x : x[0])
```

Explanation

The `key` argument to `sorted` lets us sort the data based on the first list. Using a lambda function means not having to define a separate function for this simple task

```
[(19, 'Sheila'), (21, 'Bruce'),
 (150, 'Adam')]
```
2.18 Reading and Writing Files

Several reports for this class will involve reading and analyzing data that has been stored in a file. This typically involves three steps:

- Open the file using the `open` function. This returns a file handle - an object we then use to access the text that the file contains.
- Process the file, either line-by-line, or as a single text string.
- Close the file. This is done using the `close` function.

It is possible to read in the entire contents of a file in one go using the functions `read` and `readlines`. However, we may not read to read the entire contents into memory if we are dealing with a large file and just want to extract some information from the text. In this case, it is preferable to iterate over the lines of text that the file contains.

The following examples assume that a file named "filename.txt" has been created in the directory that Python was started in. This file contains the three lines:

```
Leonard
Penny
Sheldon
```

<table>
<thead>
<tr>
<th>Python Code</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>In [1]: <code>f = open(&quot;filename.txt&quot;)</code></td>
<td>Open &quot;filename.txt&quot; for reading using <code>open</code></td>
</tr>
</tbody>
</table>
| In [2]: `for line in f:
  print line,`                                                                 | The lines of an open file can be iterated over in a `for` loop. Note the use of a "," after `print line`, since each line already ends with a new line |
|   Leonard
   Penny
   Sheldon                                                                 |             |
| In [3]: `f.close()`                                                         | Close "filename.txt" using `close` |
| In [4]: `f = open("filename.txt")
  first_names = f.read()
  f.close()
  first_names`                                                             | `read` reads in the whole file as a single string. The newlines at the end of each line are shown as "\n" characters |
| "Leonard\nPenny\nSheldon\n"                                                 |             |
| In [5]: `print first_names`                                                | Printing a string causes the newline characters "\n" to be outputted as new lines |
| Leonard
Penny
Sheldon                                                        |             |
In [6]:
   f = open("filename.txt")
   data = f.readlines()
   f.close()
   data
      
      ["Leonard\n", "Penny\n", "Sheldon\n"]

Files can also be opened for writing using the "w" option to open.

Python Code               Explanation

In [1]:
   data = ["Hofstadter", "?", "Cooper"]
   output_file = open("names.txt", "w")
   for name in data:
      output_file.write(name + "\n")
   output_file.close()

Write each string in the data list to a separate line in the file. Note that new lines are not automatically included, so they need to be added.

In [2]:
   f = open("names.txt")
   last_names = f.read()
   f.close()
   print last_names

Hofstadter
?
Cooper

2.19 Comments

Comments are text that is included in the code but not executed. They are used to document and explain what the code is doing. Python allows two forms of comment.

- A hash symbol # means that the rest of the line is a comment, and is not to be executed.
- A documentation string is surrounded by triple quotes """. Everything inside the quotes is ignored.

Python Code               Explanation

In [1]:  # This is a single-line comment

In [2]:  """This is a documentation string, which can span multiple lines"""

32
NumPy (Numerical Python) is the fundamental package for scientific computing with Python. It defines a new kind of container - the ndarray (usually just referred to as an array) - that supports fast and efficient computation. NumPy also defines the basic routines for accessing and manipulating these arrays.

Arrays have the following properties (among others):

- A *shape*, which is a tuple of integers. The number of integers is the number of dimensions in the array, and the integers specify the size of each dimension.
- A *dtype* (data-type), which specifies the type of the objects stored in the array.

In NumPy, the dimensions of an array are referred to as axes. An example of an array with dtype int and shape \((4, 5)\) is shown below. The first axis has four elements, each of which is an array with 5 elements.

\[
\begin{bmatrix}
 0 & 1 & 2 & 3 & 4 \\
 5 & 6 & 7 & 8 & 9 \\
10 & 11 & 12 & 13 & 14 \\
15 & 16 & 17 & 18 & 19 \\
\end{bmatrix}
\]

The main differences between NumPy arrays and Python lists are:

- The objects in a NumPy array must all be of the same type - booleans, integers, floats, complex numbers or strings.
- The size of an array is fixed at creation, and can’t be changed later.
- Arrays can be multi-dimensional.
- Mathematical operations can be applied directly to arrays. When this is done they are applied elementwise to the array, generating another array as output. This is *much* faster than iterating over a list.
- Indexing for arrays is more powerful than that for lists, and includes indexing using integer and boolean arrays.
- Slicing an array produces a view of the original array, not a copy. Modifying this view will change the original array.

NumPy is well documented online, with a [standard tutorial](#) and [good introductory tutorial](#) available.
3.1 Array Creation

NumPy arrays can be created:

- From a list. The elements of the list need to all be of the same type, or of a kind that can all be cast to the same type. For example, a list consisting of both integers and floats will generate an array of floats, since the integers can all be converted to floats.

- According to a given shape. The array will be initialized differently depending on the function used.

- From another array. The new array will be of the same shape as the existing array, and could either be a copy, or initialized with some other values.

- As a result of an operation on other arrays. The standard mathematical operators can all be applied directly to arrays. The result is an array of the same shape where the operation has been performed separately on corresponding elements.

The following functions are the main ones you need to know.

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>array</td>
<td>Create an array from a list.</td>
</tr>
<tr>
<td>linspace</td>
<td>Return an array of evenly spaced numbers over a specified interval.</td>
</tr>
<tr>
<td>arange</td>
<td>Return an array of evenly spaced integers within a given interval.</td>
</tr>
<tr>
<td>empty</td>
<td>Return an a new array of a given shape and type, without initializing entries.</td>
</tr>
<tr>
<td>zeros</td>
<td>Return an a new array of a given shape and type, filled with zeros.</td>
</tr>
<tr>
<td>ones</td>
<td>Return an a new array of a given shape and type, filled with ones.</td>
</tr>
<tr>
<td>empty_like</td>
<td>Return a new array with the same shape and type as a given array.</td>
</tr>
<tr>
<td>zeros_like</td>
<td>Return an array of zeros with the same shape and type as a given array.</td>
</tr>
<tr>
<td>ones_like</td>
<td>Return an array of ones with the same shape and type as a given array.</td>
</tr>
<tr>
<td>copy</td>
<td>Return an array copy of the given object.</td>
</tr>
<tr>
<td>meshgrid</td>
<td>Returns a pair of 2D x and y grid arrays from 1D x and y coordinate arrays.</td>
</tr>
</tbody>
</table>

These array creation functions are illustrated below.

```
In [1]: from numpy import *
   x = array([1, 2, 3])
   print x
[1 2 3]
   
array(object) creates an array from a list - note that arrays are printed without commas
```

```
In [2]: x = array([1, 2, 3], dtype=float)
   print x
[1. 2. 3.]
   
array(object, dtype) creates an array of type dtype - the integers are now cast to floats
```
In [3]:
   x = linspace(0, 1, 6)
print x
[0. 0.2 0.4 0.6 0.8 1. ]

In [4]:
   x = arange(5)
print x
[0 1 2 3 4]

The functions `empty`, `zeros` and `ones` all take a `shape` argument and create an array of that shape, initialized as appropriate.

<table>
<thead>
<tr>
<th>Python Code</th>
<th>Explanation</th>
</tr>
</thead>
</table>
| In [1]:
   x = empty((3, 2))
   print x
   [[ 6.93946206e-310 6.93946206e-310]
    [ 6.36598737e-314 6.36598737e-314]
    [ 6.36598737e-314 0.00000000e+000]] | `empty(shape)` returns an array of shape `shape`, initially filled with garbage |
| In [2]:
   x = zeros((2, 3))
   print x
   [[ 0. 0. 0.]
    [ 0. 0. 0.]] | `zeros(shape)` returns an array of shape `shape` filled with zeros - note the default type is `float` |
| In [3]:
   x = ones((2, 3), dtype=int)
   print x
   [[ 1 1 1]
    [ 1 1 1]] | `ones(shape, dtype)` returns an array of shape `shape` filled with ones - using `dtype=int` casts the elements to type `int` |

Arrays can be created directly from other arrays using `empty_like`, `zeros_like`, `ones_like` and `copy`.

<table>
<thead>
<tr>
<th>Python Code</th>
<th>Explanation</th>
</tr>
</thead>
</table>
| In [1]:
   x = arange(3, dtype=float)
   print x
   [0. 1. 2.] | Create an array of floats using `arange` |
| In [2]:
   y = empty_like(x)
   print y
   [ 0.00000000e+000 6.51913678e+091 6.95022185e+310] | `y` has the same shape as `x`, but is initially filled with garbage |
| In [3]:
   y = zeros_like(x)
   print y
   [ 0. 0. 0.] | `y` has the same shape as `x`, but is initialized with zeros |
In [4]: y = ones_like(x)
print y
[ 1.  1.  1.]
y has the same shape as x, but is initialized with ones

In [5]: y = copy(x)
print y
[0.  1.  2.]
y is a copy of x - changing y will not change x

The function \texttt{meshgrid}(x, y) creates two-dimensional arrays from one-dimensional x- and y-coordinate axes. One array contains the x-coordinates of all the points in the xy-plane defined by these axes, and the other contains the y-coordinates.

<table>
<thead>
<tr>
<th>Python Code</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>In [1]: from numpy import * x = arange(4) y = arange(3) X, Y = meshgrid(x, y)</td>
<td>\texttt{meshgrid} creates 2D x- and y-coordinate arrays from 1D x- and y-coordinate arrays</td>
</tr>
<tr>
<td>In [2]: print X</td>
<td>X is a 2D array containing just the x-coordinates of points in the xy plane</td>
</tr>
<tr>
<td>[[0 1 2 3] [0 1 2 3] [0 1 2 3]]</td>
<td></td>
</tr>
<tr>
<td>In [3]: print Y</td>
<td>Y is a 2D array containing just the y-coordinates of points in the xy plane</td>
</tr>
<tr>
<td>[[0 0 0 0] [1 1 1 1] [2 2 2 2]]</td>
<td></td>
</tr>
</tbody>
</table>

### 3.2. Array Properties

NumPy provides a set of functions for accessing the properties of an array.

<table>
<thead>
<tr>
<th>Python Code</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>In [1]: x = arange(6) type(x)</td>
<td>x is of \texttt{type} numpy.ndarray</td>
</tr>
<tr>
<td>numpy.ndarray</td>
<td></td>
</tr>
<tr>
<td>In [2]: x.dtype</td>
<td>\texttt{dtype} returns the element type - a 64-bit integer</td>
</tr>
<tr>
<td>\texttt{dtype(“int64“)}</td>
<td></td>
</tr>
<tr>
<td>In [3]: x.shape</td>
<td>x is a 1-dimensional array with 6 elements in the first axis</td>
</tr>
<tr>
<td>(6,)</td>
<td></td>
</tr>
</tbody>
</table>

It is possible to create an array from the elements of an existing array, but with the properties changed. The number of dimensions and size of each dimension can be changed.
using `reshape` (as long as the total number of elements is the same), and the dtype can be changed using `astype`.

<table>
<thead>
<tr>
<th>Python Code</th>
<th>Explanation</th>
</tr>
</thead>
</table>
| In [1]: x = arange(6).reshape((2, 3))
print x                  | reshape creates a view of an array with the same number of elements, but a different shape |
| [[0 1 2]
  [3 4 5]]                  |                                                    |
| In [2]: y = x.astype(float)
print y                  | astype casts the integers in x to floats in y. This creates a new array - modifying it will not alter the original |
| [[0.  1.  2.]
  [3.  4.  5.]]                  |                                                    |

### 3.3 Array Operations

Array arithmetic is done on an elementwise basis.

<table>
<thead>
<tr>
<th>Python Code</th>
<th>Explanation</th>
</tr>
</thead>
</table>
| In [1]: from numpy import *
x = arange(4)
print x
[0 1 2 3]                  | Create an array of consecutive integers using `arange` |
| In [2]: print x + 1
[1 2 3 4]                  | 1 is added to every element of the array x |
| In [3]: print x * 2
[0 2 4 6]                  | Every element of the array x is multiplied by 2 |
| In [4]: print x ** 2
[0 1 4 9]                  | Every element of the array x is squared |
| In [5]: y = array([3, 2, 5, 1])                  | Create a second array |
| In [6]: print x
print y
print x + y
[0 1 2 3]
[3 2 5 1]
[3 3 7 4]                  | The elements of x are added to the corresponding elements of y on an element-by-element basis |
| In [7]: print x**y
[ 0  1 32  3]                  | Exponentiation is done using corresponding elements |
Comparison operators and other Boolean expressions are also applied on an element-by-element basis. The result is an array of booleans.

| In [1]: | x = arange(5)  
|        | print x  
|        | print x % 2 == 0  
|        | [0 1 2 3 4]  
|        | [ True False True False True]  
|        | The Boolean expression is evaluated for each element separately, resulting in an array of booleans  

| In [2]: | x = arange(4)  
|        | y = array([3, 2, 5, 1])  
|        | print x  
|        | print y  
|        | print x < y  
|        | [0 1 2 3]  
|        | [3 2 5 1]  
|        | [ True True True False]  
|        | The comparison is done on an elementwise basis, resulting in an array of booleans  

NumPy contains vectorized versions of all the basic mathematical functions. Note that they need to be imported before we can use them. Some examples are given below.

| In [1]: | x = arange(3)  
|        | print x  
|        | [0 1 2]  
|        | Create an array  

| In [2]: | print sin(x)  
|        | [ 0. 0.84147098 0.90929743]  
|        | sin is applied to each element, to create a new array  

| In [3]: | print exp(x)  
|        | [ 1. 2.71828183 7.3890561]  
|        | exp is the exponential operator  

| In [4]: | x = random.randint(5, size=(2,3))  
|        | print x  
|        | [[1 4 3]  
|        | [2 2 3]]  
|        | random.randint returns an array of a given size filled with randomly selected integers from a given range  

| In [5]: | print x.min(), x.max()  
|        | 1 4  
|        | min and max calculate the minimum and maximum values across the entire array  

| In [6]: | print x.min(axis=0)  
|        | print x.min(axis=1)  
|        | [1 2 3]  
|        | [1 2]  
|        | The axis argument finds each minimum along a given axis. The resulting array is the shape of the original array, but with the given axis removed  

38
### 3.4 Accessing Arrays

Arrays can be indexed and sliced using square brackets in the same way as lists. An index or `start:end` filter needs to be provided for each axis, separated by commas. Indexing is zero-based, as it is with lists.

<table>
<thead>
<tr>
<th>Python Code</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>from numpy import *</code>&lt;br&gt;<code>x = arange(20).reshape((4, 5))</code>&lt;br&gt;<code>print x</code>&lt;br&gt;``[[ 0  1  2  3  4]&lt;br&gt; [ 5  6  7  8  9]&lt;br&gt; [10 11 12 13 14]&lt;br&gt; [15 16 17 18 19]]`</td>
<td><strong>reshape</strong> provides a fast way to create a 2D array</td>
</tr>
<tr>
<td><code>print x[1,2]</code>&lt;br&gt;7</td>
<td>Indexing is done into each axis in order - row 1, column 2</td>
</tr>
<tr>
<td><code>print x[1,:]</code>&lt;br&gt;``[5  6  7  8  9]`</td>
<td>Slicing selects every element of the first axis</td>
</tr>
<tr>
<td><code>print x[:,1]</code>&lt;br&gt;``[1  6 11 16]`</td>
<td>Slicing selects the first element of every axis</td>
</tr>
<tr>
<td><code>print x[1:3, 1:4]</code>&lt;br&gt;``[[ 6  7  8]&lt;br&gt; [11 12 13]]`</td>
<td>Slice rows 1 and 2 using 1:3, then slice columns 1, 2 and 3 using 1:4</td>
</tr>
</tbody>
</table>

NumPy also offers some fancy indexing tricks. The first is to index using an array of integers. An array with axes the same size as the indexing array is returned, with elements selected from the array according to the integers in the indexing array.

<table>
<thead>
<tr>
<th>Python Code</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>x = arange(9)**2</code>&lt;br&gt;<code>print x</code>&lt;br&gt;``[ 0  1  4  9 16 25 36 49 64]`</td>
<td>First create the array using <code>arange</code>, then square each element</td>
</tr>
</tbody>
</table>
CHAPTER 3. NUMPY

3.4. ACCESSING ARRAYS

In [2]:
   index = array([1, 3])
   print x[index]

[1 9]

An array is returned containing elements from the first array, selected according to the
integers in the second array

In [3]:
   index = array([[1, 3], [7, 2]])
   print index
   print x[index]

[[1 3]
 [7 2]]
[[1 9]
 [49 4]]

When indexing using an integer array, the returned array has the same shape as the index-
ing array

We can also index using arrays of booleans. In this case, only the elements corresponding
to True values in the indexing array are returned. A common use of this technique is
to filter out the elements of an array that satisfy some condition. This can be done by
first applying the condition to the array to generate an array of booleans, then using the
resulting array as an index. The result is that only elements for which the condition holds
true are selected.

Python Code          Explanation
---------------------  ---------------------------------------------------------------
In [1]:
   x = arange(20)
   index3 = (x % 3 == 0)
   index 5 = (x % 5 == 0)

index3 and index5 are boolean arrays containing True elements for the integers that are divisible by 3
and 5 respectively.

In [2]:
   print x[index3]
[ 0 3 6 9 12 15 18]

Select the elements of x that are divisible by 3

In [3]:
   print x[index5]
[ 0 5 10 15]

Select the elements of x that are divisible by 5

In [4]:
   print x[logical_or(index3, index5)]
[ 0 3 5 6 9 10 12 15 18]

The function logical_or performs an elementwise “or”. The result is the integers divisible by either 3 or 5
Matplotlib is a 2D plotting library for Python. It can be used to generate graphs, histograms, bar charts, contour plots, scatter plots, and many other kinds of mathematical graphics. The “pyplot” interface provides a MATLAB-like interface for simple plotting, and is the main one we will be using in class. The online reference provides a full description of the available functions. A good tutorial is also available online.

The following commands are the main ones used for creating and formatting graphs.

- **plot**: Plot lines and/or markers.
- **show**: Display a figure.
- **title**: Set a title for the graph.
- **xlabel/ylabel**: Set labels for the x and y axes.
- **xlim/ylim**: Get or set the range of x and y values to be displayed.
- **xticks/yticks**: Get or set the locations and labels for the tick marks on the x and y axes.
- **subplot**: Plot multiple graphs in one figure.
- **figure**: Create a new figure.
- **fill_between**: Fill the area between two curves.
- **legend**: Put a legend on the graph.

Colors, line styles, and marker styles can all be set to create customized graphs. These are usually specified as strings, with the most frequently used options as follows.

<table>
<thead>
<tr>
<th>Style options</th>
<th>Colors</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;-&quot;</td>
<td>&quot;b&quot;</td>
</tr>
<tr>
<td>&quot;- -&quot;</td>
<td>&quot;g&quot;</td>
</tr>
<tr>
<td></td>
<td>&quot;r&quot;</td>
</tr>
<tr>
<td>&quot;o&quot;</td>
<td>&quot;c&quot;</td>
</tr>
<tr>
<td>&quot;s&quot;</td>
<td>&quot;m&quot;</td>
</tr>
<tr>
<td>&quot;+&quot;</td>
<td>&quot;y&quot;</td>
</tr>
<tr>
<td>&quot;x&quot;</td>
<td>&quot;k&quot;</td>
</tr>
<tr>
<td>&quot;D&quot;</td>
<td>&quot;w&quot;</td>
</tr>
</tbody>
</table>

41
4.1 Basic Plotting

Functions can be graphed using a call to `plot(x, y)`, followed by a call to `show`. Note that:

- The `x` parameter contains the x-coordinates of the points to plot, and the `y` parameter contains the y-coordinates.
- We need to import the required NumPy and Pylab functions for array manipulation and plotting. If using IPython notebook, this can also be done using the IPython magic `%pylab`. In this case, the call to `show` is done for us automatically when the cell is executed.
- The default plotting behavior is to connect the points with a blue line.

The following example plots the exponential function in the range [0, 5].

```python
from numpy import *
import pylab as pl
x = linspace(0, 5) # Create array of equally spaced values
pl.plot(x, exp(x)) # Plot the exponential function
pl.show() # Finally, show the figure
```
A range of options are available for customizing plots. These are illustrated in the example below, which plots a sine and cosine curve on the same graph. Note that:

- The third argument to `plot` can be used to set colors, line types and marker types.
- `Plot` can be called multiple times, followed by a single call to `show`.

```python
from numpy import *
import matplotlib as mpl

x = linspace(0, 2*pi, 50)
mpl.figure(figsize=(10,7))
mpl.plot(x, sin(x), label='sine')
mpl.plot(x, cos(x), 'ro', label='cosine')
mpl.xlabel('theta')
mpl.xlim(0, 2*pi)
ticks = [i*pi/2 for i in range(5)]
labels = [r'$0$', r'$\pi/2$', r'$\pi$', r'$3\pi/2$', r'$2\pi$']
mpl.xticks(ticks, labels, size='large')
mpl.title('Sine and Cosine')
mpl.legend()
mpl.show()
```

# Imports linspace, sin, cos
# Plot 50 points on the x-axis
# Set the size of the figure
# Default style is a blue line
# Use 'ro' for red circles
# Label the x-axis
# Limit x-axis to this range
# Locations of ticks on x-axis
# Labels for the x-axis ticks
# These are LaTeX strings
# Add the x-ticks and labels
# Add a title
# Legend uses the plot labels
# Finally, show the figure
4.3 Bar Plots

The function `bar` is used to create bar plots.

- Bars are described by their height, width, and position of the left and bottom edges.
- The `width` argument can be used to make bars thinner or thicker.
- The face color and edge color of the bars can be specified independently.

The following example shows a bar plot with the face color set to "c" (cyan) and edge color set to "b" (blue). Labels are positioned at the centers of the bars.

```python
import pylab as pl

grades = ['A', 'B', 'C', 'D', 'F']
freqs = [30, 35, 20, 10, 5]
width = 0.8
ticks = [width/2 + i for i in range(5)]
pl.bar(range(5), freqs, fc='c', ec='b')
pl.xticks(ticks, grades)
pl.ylim(0, 40)
pl.title('Grade distribution')
pl.xlabel('Grade')
pl.ylabel('Frequency (%)')
pl.show()
```

```
import pylab as pl

grades = ['A', 'B', 'C', 'D', 'F']
freqs = [30, 35, 20, 10, 5]
width = 0.8
ticks = [width/2 + i for i in range(5)]
pl.bar(range(5), freqs, fc='c', ec='b')
pl.xticks(ticks, grades)
pl.ylim(0, 40)
pl.title('Grade distribution')
pl.xlabel('Grade')
pl.ylabel('Frequency (%)')
pl.show()
```
4.4 Histograms

The function `hist` is used to plot histograms. These group numerical data into “bins”, usually of equal width, in order to show how the data is distributed.

- Each bin covers a range of values, with the height of each bin indicating the number of points falling in that range.
- The first argument is an array or sequence of arrays.
- The `bins` argument specifies the number of bins to use.
- The `range` argument specifies the range of values to include.

The following example plots a histogram of 1000 samples drawn from a uniform probability distribution over $[0, 1)$.

```python
from numpy import 
import pylab as pl

x = random.rand(1000)  # 1000 random values in [0, 1)
x.hist(x, bins=20, range=(0,1), fc='g')  # Create histogram with 20 bins
pl.title('Uniform distribution')  # Add a title
pl.xlabel('Value')  # Add a label for the x-axis
pl.ylabel('Frequency')  # Add a label for the y-axis
pl.show()  # Finally, show the figure
```

![Histogram example](image)
4.5 Contour Plots

The functions `contour` and `contourf` are used for contour plots and filled contour plots respectively. These are projections of a graph surface onto a plane, with the contours showing the level curves of the graph.

- The first two arguments are one dimensional arrays representing the x- and y-coordinates of the points to plot.
- The third coordinate is a two dimensional array representing the z-coordinates.
- Contour levels are automatically set, although they can be customized.
- A colorbar can be added to display the level curves.

The following examples are of a filled and unfilled contour plot of the two-dimensional Gaussian function, \( f(x, y) = e^{-(x^2+y^2)} \).

```python
import numpy as np
from numpy import *

x = linspace(-2, 2)
y = linspace(-2, 2)
XX, YY = meshgrid(x, y)
z = exp(- (XX**2 + YY**2))

plt.figure(figsize=(14, 5))
plt.subplot('121')
plt.contour(x, y, z)
plt.title('Contour plot')
plt.colorbar()
plt.show()

plt.subplot('122')
plt.contourf(x, y, z)
plt.title('Filled contour plot')
plt.colorbar()
plt.show()
```

![Contour plot](image1.png) ![Filled contour plot](image2.png)
CHAPTER 4. MATPLOTLIB

4.6. MULTIPLE PLOTS

The function **subplot** is used to plot multiple graphs on a single figure. This divides a figure into a grid of rows and columns, with plotting done in the currently active subplot.

- Calls to subplot specify the number of rows, number of columns, and subplot number.
- Subplots are numbered from left to right, row by row, starting with 1 in the top left.
- All plotting is done in the location specified in the most recent call to subplot.
- If there are less than 10 rows, columns and subplots, subplot can be called with a string argument. For example, subplot(2, 3, 4) is the same as subplot(“234”).

The example below uses 2 rows and 3 columns. The “subplot” calls displayed on the figure show which call corresponds to each grid location.

```python
import matplotlib.pyplot as plt  
fig = plt.figure(figsize=(8, 5))  
nrows = 2  
ncols = 3  
for i in range(nrows * ncols):
    plt.subplot(nrows, ncols, i + 1)  
```

The example above uses 2 rows and 3 columns. The “subplot” calls displayed on the figure show which call corresponds to each grid location.
4.7 Formatting Mathematical Expressions

LaTeX provides a way to format mathematical expressions in Matplotlib graphs in a similar way to IPython notebook Markdown cells.

- Expressions are identified using \( r"\langle formula \rangle $\).
- The syntax for \( \langle formula \rangle \) is the same as that described in section 1.5.3.
- These expressions can be used anywhere a string is used, such as titles, axis and ticks labels, and legends.
5 Additional Topics

5.1 Loading Numerical Files

We often need to load files containing numerical data into a NumPy array for further processing and display. Such data files typically consist of:

- Header information. This describes what the data represents and how it is formatted.
- A set of rows of numerical data. Each row contains the same number of values, separated by some string such as a comma or tab.

The NumPy function `numpy.loadtxt` can be used to load such data. This returns a NumPy array, where each row corresponds to a line in the data file. The first argument to this function is the data file name. Some of the optional keyword arguments are shown below.

- `dtype`. This is the data type of values in the array, which are floats by default.
- `delimiter`. This is the string used to separate values in each row. By default, any whitespace such as spaces or tabs are considered delimiters.
- `skiprows`. This is the number of rows to ignore at the start of the file before reading in data. It is usually used to skip over the header information, and defaults to 0.

The example shown below uses a file called “weather.dat”, which contains the following:

<table>
<thead>
<tr>
<th>Day</th>
<th>High-Temp</th>
<th>Low-Temp</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>77</td>
<td>56</td>
</tr>
<tr>
<td>2</td>
<td>79</td>
<td>62</td>
</tr>
</tbody>
</table>

This data is loaded as follows.

<table>
<thead>
<tr>
<th>Python Code</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>from numpy import loadtxt</code></td>
<td>Import the <code>loadtxt</code> function</td>
</tr>
<tr>
<td><code>data = loadtxt(&quot;weather.dat&quot;, dtype=int, skiprows=1)</code></td>
<td>Load the &quot;weather.dat&quot; file, skipping the first row, and creating an array of integers</td>
</tr>
<tr>
<td><code>print data</code></td>
<td>The data array is a $2 \times 3$ integer array - for floats, the <code>dtype</code> argument would not be used</td>
</tr>
</tbody>
</table>
5.2. Animation

An animation consists of a sequence of frames which are displayed one after the other. Animation using Matplotlib essentially involves updating the data associated with some drawn object or objects (such as points or lines), and redrawing these objects. Producing an animation therefore involves the following steps:

- Set up the variables and data structures relating to the animation.
- Draw the first frame.
- Repeatedly update the frame with new data.

Animations are generated using `FuncAnimation` from the `matplotlib.animation` module. This takes the following required arguments:

- `fig`. This is the figure in which the animation is to be drawn. It can be obtained using either of the Matplotlib figure or subplots functions.

- `func`. This specifies the function to call to perform a single step of the animation. It should take a single argument which is the frame number (an integer). The frame number is used to update the values of drawn objects such as points or lines. If the `blit` keyword argument is True, this function should return a tuple of the modified objects that need to be redrawn.

`FuncAnimation` also takes the following keyword arguments.

- `frames`. An integer specifying the number of frames to generate.

- `init_func`. This specifies the function which is called once at the start to draw the background that is common to all frames. If the `blit` keyword argument is True, this function should also return a tuple of the modified objects that need to be redrawn.

- `interval`. This argument specifies the time (in ms) to wait between drawing successive frames.

- `blit`. If True, the animation only redraws the parts of the plot which have changed. This can help ensure that successive frames are displayed quickly.

- `repeat`. If True (the default), the animation will repeat from the beginning once it is finished.

The following example for IPython Notebook animates a point circling the origin with constant angular velocity. The `animate` function is defined to update the position of the point based on the frame number.
from matplotlib import animation

omega = .02
fig, ax = subplots(figsize=(4,4))
ax.set_aspect('equal')
point, = plot([], [], 'ro', ms=10)
xlim(-1.5, 1.5)
ylim(-1.5, 1.5)

# Initialization function. This is called once to plot the background.
def init():
    point.set_data([], [])
    return point,

# Animation function. This is called once per animation step.
# The integer i is the frame number.
def animate(i):
    x = cos(i*omega)
    y = sin(i*omega)
    point.set_data(x, y)
    return point,

# Start the animator with a call to "FuncAnimation"
animation.FuncAnimation(fig, animate, init_func=init, frames=100,
                      interval=20, blit=True)

Some frames from this animation are shown below.

Note that in IPython Notebook the IPython magic we need to use is %pylab rather than %pylab inline. Inline graphs in IPython Notebook are static, meaning that once drawn, they cannot be updated. Using %pylab generates graphs in a separate window, where the updated data can be displayed.
5.3 Images

Matplotlib provides functions for saving, reading, and displaying images. These images are either 2- or 3-dimensional NumPy arrays. In both cases, the first two axes of the array correspond to the rows and columns of the image. The third axis corresponds to the color of the pixel at each (column, row) coordinate.

- For a 2D array, the array values are floats in the range [0, 1]. These represent the luminance (brightness) of a grayscale image from black (0) to white (1).
- For a 3D array, the third axis can have either 3 or 4 elements. In both cases, the first three elements correspond to the red, green, and blue components of the pixel color. These can be either floats in the range [0, 1], or 8-bit integers of type 'uint8'. A fourth element corresponds to an “alpha” value representing transparency.

The main functions we use are:
- `imread` Read an image file into an array.
- `imsave` Save an image to file.
- `imshow` Display an image array.

<table>
<thead>
<tr>
<th>Python Code</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>In [1]: <code>img = zeros((100, 100, 3))</code></td>
<td>Create an image array of 100 rows and columns, set to zero</td>
</tr>
<tr>
<td>In [2]: <code>img[:60,:60,0] = 1.</code></td>
<td>Set the top-left corner to red</td>
</tr>
<tr>
<td>In [3]: <code>img[40:,:40,:1] = 1.</code></td>
<td>Set the lower-right corner to green</td>
</tr>
<tr>
<td>In [4]: <code>img[20:80,20:80,2] = 1.</code></td>
<td>Set the center square to blue</td>
</tr>
<tr>
<td>In [5]: <code>imsave(&quot;squares.png&quot;, img)</code> <code>img2 = imread(&quot;squares.png&quot;)</code> <code>imshow(img2)</code></td>
<td>Save the img array to the &quot;squares.png&quot; file, read the file back to the img2 array, and display the image</td>
</tr>
</tbody>
</table>
5.4 Random Number Generation

NumPy provides a library of functions for random number generation in the `random` module. These return either a sample, or an array of samples of a given size, drawn from a given probability distribution. The main functions we use are:

- `random.rand` Samples are drawn from a uniform distribution over $[0, 1)$.
- `random.randint` Samples are integers drawn from a given range.
- `random.randn` Samples are drawn from the “standard normal” distribution.
- `random.normal` Samples are drawn from a normal (Gaussian) distribution.

The following examples illustrate the use of these functions.

<table>
<thead>
<tr>
<th>Python Code</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>In [1]: from numpy import * print random.rand()</td>
<td>Return a single random number uniformly drawn from the interval $[0, 1)$</td>
</tr>
<tr>
<td>In [2]: print random.rand(3)</td>
<td>Return an array of 3 random numbers drawn from $[0, 1)$</td>
</tr>
<tr>
<td>In [3]: print random.rand(2, 3)</td>
<td>Return an array of size $(2, 3)$ of random numbers drawn from $[0, 1)$</td>
</tr>
<tr>
<td>In [4]: print random.randint(7)</td>
<td>Return an integer drawn from ${0, 1, 2, 3, 4, 5, 6}$. Note that 7 is not included</td>
</tr>
<tr>
<td>In [5]: print random.randint(5, 9, size=(2,4))</td>
<td>Return an array of size $(2, 4)$ of integers drawn from ${5, 6, 7, 8}$</td>
</tr>
<tr>
<td>In [6]: print random.randn(3)</td>
<td>Return array of samples drawn from “standard” normal distribution ($\mu = 0$, $\sigma = 1$)</td>
</tr>
<tr>
<td>In [7]: print random.normal(100, 15)</td>
<td>Return a sample drawn from a normal distribution with $\mu = 100$, $\sigma = 15$</td>
</tr>
</tbody>
</table>
CHAPTER 5. ADDITIONAL TOPICS 5.5. SOUND FILES

5.5 Sound Files

Sound is a vibration that propagates through a medium such as air as a wave of pressure and displacement. Recording devices such as microphones convert this wave to an electrical signal. This signal is then sampled at regular intervals and converted to a sequence of numbers, which correspond to the wave amplitude at given times.

The WAVE (or WAV) file format is a standard for storing such audio data without compression. WAV files contain two main pieces of information:

- The rate at which the wave has been sampled, usually 44,100 times per second.
- The audio data, usually with 16 bits used per sample. This allows $2^{16} = 65,536$ different amplitude levels to be represented.

The module `scipy.io.wavfile` provides functions to read and write such files.

`scipy.io.wavfile.read` Read a WAV file, returning the sample rate and the data.

`scipy.io.wavfile.write` Write a NumPy array as a WAV file.

The following example creates and saves a WAV file with a single frequency at middle C, then plots the first 1000 samples of the data.

```python
from numpy import linspace
from scipy.io import wavfile
from pylab import plot, show

rate = 44100 # Number of samples/second
end = 10 # The file is 10 seconds long
freq = 261.625565 # Frequency of "middle C"

time = linspace(0, end, rate*end+1) # Time intervals are 1/rate
data = sin(2*pi*freq*time) # Generate the sine wave
wavfile.write("middleC.wav", rate, data) # Write the array to a wav file
plot(time[:1000], data[:1000]) # Plot the first 1000 samples
show() # Finally, show the figure
```

![Plot of sine wave](image)
5.6 Linear Programming

Linear programming problems are a special class of optimization problem. They involve finding the maximum (or minimum) of some linear objective function \( f(x) \) of a vector of variables \( x = (x_1, x_2, \ldots, x_n) \), subject to a set of linear equality and inequality constraints.

Since the objective function and constraints are linear, we can represent the problem as:

Maximize \( c^T x \), where the vector \( c \) contains the coefficients of the objective function, subject to \( A_{ub} \cdot x \leq b_{ub} \), where \( A_{ub} \) is a matrix and \( b_{ub} \) a vector, and \( A_{eq} \cdot x = b_{eq} \), where \( A_{eq} \) is a matrix and \( b_{eq} \) a vector.

An example of such a problem would be: \( x = \{x_1, x_2\} \). Maximize \( f(x) = 2x_1 + 3x_2 \) subject to the inequality constraints (i) \( 0 \leq x_1 \leq 80 \), (ii) \( x_2 \geq 0 \), (iii) \( x_1 + x_2 \leq 100 \), and (iv) \( x_1 + 2x_2 \leq 160 \). This example is graphed below, showing the level curves of \( f(x) \).

The function `scipy.optimize.linprog` implements the “simplex algorithm” we discuss in class to solve this problem. The arguments to this function are the values \( c, A_{ub}, b_{ub}, A_{eq} \) and \( b_{eq} \) given above. An optional `bounds` argument represents the range of permissible values that the variables can take, with `None` used to indicate no limit.

Applying `linprog` to this problem is done as shown below. Note that `linprog` finds the minimum of \( f(x) \), so we use negative values for the \( c \) coefficients to find a maximum.

```python
import numpy as np

c = np.array([-2, -3])  # Negative coefficients of f(x)
A_ub = np.array([[1, 1], [1, 2]])  # Matrix of the inequality coefficients
b_ub = np.array([100, 160])  # Vector of the inequality upper bounds
bounds = [(0, 80), (0, None)]  # Each tuple is a (lower, upper) bound
result = scipy.optimize.linprog(c, A_ub=A_ub, b_ub=b_ub, bounds=bounds)  # The "x" field holds the solution
print(result.x)
```

This yields the correct solution for \( x_1 \) and \( x_2 \), as seen in the graph above:

```
[ 40.  60.]
```
6 Programming Style

This chapter contains some tips on how to make programs easier to read and understand. Programs are written first and foremost to be understood by human beings, not by computers. Ideally, it should be possible a year from now for you to pick up the code that you’re writing today and still understand what you were doing and why. (It should also be possible for the instructor to understand it a week from now...)

6.1 Choosing Good Variable Names

Good variable names make reading and debugging a program much easier. Well chosen names are easy to decipher, and make the intent clear without additional comments.

- A variable name should fully and accurately describe the data it represents. As an example, date may be ambiguous whereas current_date is not. A good technique is to state in words what the variable represents, and use that for the name.

- Names that are too short don’t convey enough meaning. For example, using d for date or cd for current date is almost meaningless. Research shows that programs with variable names that are about 9 to 15 characters long are easiest to understand and debug.

- Variable names should be problem-oriented, referring to the problem domain, not how the problem is being solved. For example, planet_velocity refers to the problem, but vector_3d refers to how this information is being represented.

- Loop indices are often given short, simple names such as i, j and k. This is okay here, since these variables are just used in the loop, then thrown away.

- If loops are nested, longer index names such as row and column can help avoid confusion.

- Boolean variables should have names that imply either True or False. For example, prime_found implies that either a prime has been found, or it hasn’t.

- Boolean variables should be positive. For example, use prime_found rather than prime_not_found, since negative names are difficult to read (particularly if they are negated).

- Named constants should be in uppercase and refer to what the constant represents rather than the value it has. For example, if you want to use the same color blue for the font in every title, then define the color in one place as TITLE_FONT_COLOR rather than FONT_BLUE. If you later decide to have red rather than blue titles, just redefine TITLE_FONT_COLOR and it keeps the same meaning.
6.2 Choosing Good Function Names

The **recommended style** for naming functions in Python is to use all lowercase letters, separated by underscores as necessary. As with variable names, good function names can help make the intent of the code much easier to decipher.

- For procedures (functions that do something and don’t return a value), use a verb followed by an object. An example would be `plot_prime_distribution`.
- For functions that return values, use a description of what the returned value represents. An example would be `miles_to_kilometers`.
- Don’t use generic names such as `calculate_stuff` or numbered functions such as `function1`. These don’t tell you what the function does, and make the code difficult to follow.
- Describe everything that the function does, and make the function name as long as is necessary to do so. If the function name is too long, it may be a sign that the function itself is trying to do too much. In this case, the solution is to use shorter functions which perform just one task.

6.3 No “Magic Numbers”

Magic numbers are numbers such as 168 or 9.81 that appear in a program without explanation. The problem with such numbers is that the meaning is unclear from just reading the number itself.

- Numbers should be replaced with named constants which are defined in one place, close to the start of your code file.
- Named constants make code more readable. It’s a lot easier to understand what `HOURS_PER_WEEK` is referring to than the number 168.
- If a number needs to change, named constants allow this change to be done in one place easily and reliably.

6.4 Comments

It’s not necessary to comment every line of code, and “obvious” comments which just repeat what the code does should be avoided. For example, the endline comment in the following code is redundant and does nothing to explain what the code is for.

```
x += 1 # Add 1 to x
```

Good comments serve two main purposes:

- “Intent” comments explain the purpose of the code. They operate at the level of the problem (why the code was written) - rather than at the programming-language level (how the code operates). Intent is often one of the hardest things to understand when reading code written by another programmer.
“Summary” comments distill several lines of code into one or two sentences. These can be scanned faster than the code itself to quickly understand what the code is doing. For example, suppose you are creating several different graphs for a report. A summary comment before each plot and its associated set of formatting commands can describe which figure in the report the code is producing.

Endline comments are those at the end of a line, after the code. They are best avoided for a number of reasons.

- Endline comments are short by necessity as they need to fit into the space remaining on a line. This means that they tend to be cryptic and uninformative.
- Endline comments are difficult to keep aligned (particularly as the code changes), and if they’re not aligned they become messy and interfere with the visual structure of the code.

A final note is to get in the habit of documenting code files. At the top of every file, include a block comment describing the contents of the file, the author, and the date the file was created. An example would be:

```
# MTH 337: Intro to Scientific and Mathematical Computing, Fall 2015
# Report 1: Primitive Pythagorean Triples
# Created by Adam Cunningham 8/31/2015
```

6.5 Errors and Debugging

The following suggestions may help to reduce errors.

- Test each function completely as you go.
- In the initial stages of learning Python, test each few lines of code before moving on to the next.
- Add “print” statements inside a function to print out the intermediate values of a calculation. This can be used to check that a function is working as required, and can always be commented out afterwards.

In the event of an error being generated, IPython will typically give as much information as possible about the error. If this information is not sufficient, the %debug magic will start the IPython debugger. This lets the current values of variables inside a function be examined, and allows code to be stepped through one line at a time.
7 Further Reading

The following books may prove useful for further study or reference.

  A fairly short introduction to using NumPy and Matplotlib in IPython Notebooks. This is not a Python tutorial, although there is a brief review of Python in the appendix.

  A general guide to code writing and software construction, this book focuses on questions of software design rather than any specific language. More useful to an intermediate-level programmer who wants to improve their skills. No references to Python.

  A tutorial for Python, NumPy and Matplotlib that also covers many of the same scientific and mathematical topics as this class.

  A good introduction to Python, NumPy, Matplotlib and three-dimensional graphics. Extensive treatment of numerical solutions to ordinary, stochastic, and partial differential equations.