**Introduction**

Some Useful Linear Algebra Functions

<table>
<thead>
<tr>
<th>Function</th>
<th>Input</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>UNS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UNR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UNQ</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Solving Linear Systems of Equations

<table>
<thead>
<tr>
<th>Function</th>
<th>Input</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>TNW</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNU</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNQ</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Graphics

<table>
<thead>
<tr>
<th>Function</th>
<th>Input</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNU</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SNT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SNR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SNQ</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Nonsingular Functions and Other Data Types

<table>
<thead>
<tr>
<th>Function</th>
<th>Input</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNY</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNV</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNQ</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Scalar Calculations

Overdetermined and Underdetermined Linear Systems

<table>
<thead>
<tr>
<th>Function</th>
<th>Input</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>QNY</td>
<td></td>
<td></td>
</tr>
<tr>
<td>QNX</td>
<td></td>
<td></td>
</tr>
<tr>
<td>QNW</td>
<td></td>
<td></td>
</tr>
<tr>
<td>QNU</td>
<td></td>
<td></td>
</tr>
<tr>
<td>QNT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>QNS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>QNR</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


directory of Mathematics

Index

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
</tr>
<tr>
<td>16</td>
</tr>
<tr>
<td>24</td>
</tr>
<tr>
<td>116</td>
</tr>
<tr>
<td>119</td>
</tr>
<tr>
<td>124</td>
</tr>
<tr>
<td>130</td>
</tr>
<tr>
<td>133</td>
</tr>
<tr>
<td>136</td>
</tr>
<tr>
<td>139</td>
</tr>
<tr>
<td>146</td>
</tr>
<tr>
<td>149</td>
</tr>
<tr>
<td>149</td>
</tr>
<tr>
<td>150</td>
</tr>
<tr>
<td>150</td>
</tr>
<tr>
<td>151</td>
</tr>
<tr>
<td>151</td>
</tr>
<tr>
<td>152</td>
</tr>
<tr>
<td>152</td>
</tr>
<tr>
<td>153</td>
</tr>
<tr>
<td>153</td>
</tr>
<tr>
<td>154</td>
</tr>
<tr>
<td>154</td>
</tr>
<tr>
<td>155</td>
</tr>
<tr>
<td>155</td>
</tr>
<tr>
<td>156</td>
</tr>
<tr>
<td>156</td>
</tr>
<tr>
<td>157</td>
</tr>
<tr>
<td>157</td>
</tr>
<tr>
<td>157</td>
</tr>
<tr>
<td>158</td>
</tr>
<tr>
<td>158</td>
</tr>
<tr>
<td>158</td>
</tr>
<tr>
<td>158</td>
</tr>
<tr>
<td>159</td>
</tr>
<tr>
<td>159</td>
</tr>
<tr>
<td>160</td>
</tr>
<tr>
<td>160</td>
</tr>
<tr>
<td>161</td>
</tr>
<tr>
<td>161</td>
</tr>
<tr>
<td>161</td>
</tr>
<tr>
<td>161</td>
</tr>
<tr>
<td>162</td>
</tr>
<tr>
<td>162</td>
</tr>
<tr>
<td>162</td>
</tr>
<tr>
<td>162</td>
</tr>
<tr>
<td>163</td>
</tr>
<tr>
<td>167</td>
</tr>
</tbody>
</table>

ASCII Table
Introduction

MATLAB is an interactive software package which was developed to perform numerical calculations on vectors and matrices. Initially, it was simply a MATrix LABoratory. However, today it is much more powerful:

- It can do quite sophisticated graphics in two and three dimensions.
- It contains a high-level programming language (a "baby C") which makes it quite easy to code complicated algorithms involving vectors and matrices.
- It can numerically solve nonlinear initial-value ordinary differential equations.
- It can numerically solve nonlinear boundary-value ordinary differential equations.
- It contains a wide variety of toolboxes which allow it to perform a wide range of applications from science and engineering. Since users can write their own toolboxes, the breadth of applications is quite amazing.

Mathematics is the basic building block of science and engineering, and MATLAB makes it easy to handle many of the computations involved. You should not think of MATLAB as another complication programming language, but as a powerful calculator that gives you fingertip access to exploring interesting problems in science, engineering, and mathematics. And this access is available by using only a small number of commands and functions because MATLAB's basic data element is a matrix (or an array).

This is a crucial feature of MATLAB — it was designed to group large amounts of data in arrays and to perform mathematical operations on this data as individual arrays rather than as groups of data. This makes it very easy to apply complicated operations to the data, and it make it very difficult to do it wrong. In high-level computer languages you would usually have to work on each piece of data separately and use loops to cycle over all the pieces. In MATLAB this can frequently do complicated "things" in one, or a few, statements (and no loops).

In addition, in a high-level language many mathematical operations require the use of sophisticated software packages, which you have to find and, much worse, to work with since the interfaces to these packages are frequently quite complicated and the documentation must be read and mastered. In MATLAB, on the other hand, these operations have simple and consistent interfaces which are quite easy to master. For an overview of the capabilities of MATLAB, type in the Help Navigator and click on MATLAB.

This tutorial is designed to be a concise introduction to many of the capabilities of MATLAB. It makes no attempt to cover either the range of topics or the depth of detail you can find in a reference manual, such as Mastering MATLAB by Duane Hanselman and Bruce Littlefield (which is over 950 pages long) or MATLAB Guide, 2nd edition by Desmond and Nicholas Higham (which is almost 400 pages long).

This tutorial was initially written to provide students with a free "basic" overview of functions which are useful in an undergraduate course on linear algebra.

In this tutorial MATLAB is first introduced as a calculator and then as a plotting package. Only afterwards are more technical topics discussed. We take this approach because most people are quite familiar with calculators, and it is only a small step to understand how to apply these same techniques to matrices rather than individual numbers or variables. In addition, by viewing MATLAB as a simple but powerful calculator, rather than as a complicated software package or computer language, you will be in the correct frame of mind to use MATLAB.

You should view MATLAB as a tool that you are "playing with" — trying ideas out and seeing how they work. If an idea works, fine; if it doesn't, investigate further and figure out why. Maybe you misunderstood some MATLAB command/function, or maybe your idea needs some refinement. "Play around" interactively and figure it out. There are no hard and fast rules for figuring it out — try things and see what happens. It is OK to make mistakes; MATLAB won't call you an idiot for making a mistake. When you first learned to ride a bicycle, you fell down a lot — and you looked pretty silly. But you kept at it until you didn't fall down. You didn't study Newton's laws of motion and try to analyze the motion of a bicycle; you didn't take classes in how to ride a bicycle; you didn't get videos from the library on how to ride a bicycle. You just kept at it, possibly with the assistance of someone who studied the bicycle and gave you a little push to get you started. This is how you should learn MATLAB.

However, this tutorial is not designed for "playing around". It is very ordered, because it has been designed as a brief introduction to all the basic topics that I consider important and then as a reference manual. It would be very useful for students to have a document which uses this "play around" approach so you would learn topics by using them in exploring some exercise. This is how workbooks should be written: present some exercise for students to investigate, and let them investigate it themselves. And these exercises should be interesting, having some connection to physical or mathematical models that the students — or at least a reasonable fraction thereof — have some knowledge of and some interest in. This tutorial is designed to be a reference manual that could be used alongside such a workbook — if only someone would write it.

Summary of Contents

We have tried to make this tutorial as linear as possible so that the building blocks necessary for a section are contained in preceding sections. This is not the best way to learn MATLAB, but it is a good way to document it. In addition, we try to separate these building blocks and put them in short subsections so that they are easy to find and to understand. Next, we collect all the commands/functions discussed in a subsection and put them in a box at the end along with a very brief discussion to make it easy to remember these commands. Finally, we collect them all and put them in the appendix, again boxed up by topic. MATLAB has a huge number of commands/functions and this is one way to collect them for easy reference.

Warning: Usually we do not discuss the complete behavior of these commands/functions, but only their most useful behavior. Typing gives you complete information about the command/function.

Notation: help <command/function> means to enter whatever command/function you desire (without the braces).

c demo command/function means to type these two words as written.

Section 1 of this tutorial discusses how to use MATLAB as a "scalar" calculator, and Section 2 how to use it as a "matrix" calculator. Following this, you will be able to set up and solve the matrix equation \( \mathbf{Ax} = \mathbf{b} \) where \( \mathbf{A} \) is a square nonsingular matrix.

The following sections delve more deeply into particular topics. Section 5 discusses how to find any and all solutions of \( \mathbf{Ax} = \mathbf{b} \) where \( \mathbf{A} \in \mathbb{C}^{m \times n} \) need not be a square matrix; there might be no solutions, one solution, or an infinite number of solutions to this linear system. When no solution exists, it discusses how to calculate a least-squares solution (i.e., the "best" approximation to a solution). In addition, it discusses how round-off errors can corrupt the solution, and how to determine if this is likely to occur.

Section 6 is quite brief and discusses advanced functions to input data into MATLAB and output it to a file. (The basic functions are discussed in Section 4.1.) This is useful if the data is being shared between various computer programs and/or software packages.

Section 7 discusses a number of functions which are useful in linear algebra and numerical linear algebra. Probably the most useful of these is calculating some or all of the eigenvalues of a square matrix.
Section 8 discusses MATLAB as a programming language — a very "baby C." Since the basic data element of MATLAB is a matrix, this programming language is very simple to learn and to use. Most of this discussion focuses on writing your own MATLAB functions, called function m-files (which are similar to functions in C and to functions, more generally subprograms, in Fortran). Using these functions, you can code a complicated sequence of statements such that all these statements as well as all the the variables used by these functions are hidden and will not affect the remainder of your MATLAB session. The only way to pass data into and out of these functions is through the argument list.

Section 9 discusses how to generate sparse matrices (i.e., matrices where most of the elements are zero). These matrices could have been discussed in Section 2, but we felt that it added too much complexity at too early a point in this tutorial. Unless the matrix is very large it is usually not worthwhile to generate sparse matrices — however, when it is worthwhile the time and storage saved can be enormous.

Section 10 discusses how to use MATLAB to numerically solve initial-value ordinary differential equations. This section is divided up into a "basic" part and an "advanced" part. It often requires very little effort to solve even complicated odes; when it does we discuss in detail what to do and provide a number of examples. Section 11 discusses how to use MATLAB to numerically solve boundary-value ordinary differential equations.

Section 12 discusses how to numerically handle standard polynomial calculations such as evaluating polynomials, differentiating polynomials, and finding their zeroes. Polynomials and piecewise polynomials can also be used to interpolate data.

Section 13 discusses how to numerically calculate zeroes, extrema, and integrals of functions.

Section 14 discusses the discrete Fourier transform and shows how it arises from the continuous Fourier transform. We also provide an example which shows how to recover a simple signal which has been severely corrupted by noise.

Finally, Section 15 discusses how to apply mathematical functions to matrices.

There is one appendix which collects all the commands/functions discussed in this tutorial and boxes them up by topic. If a command/function has more than one use, it might appear in two or more boxes.

This tutorial closes with an index. It is designed to help in finding things that are "just on the tip of your tongue." All the MATLAB commands/functions discussed here are listed at the beginning of the index, followed by all the symbols, then followed by a list of all the script and function m-files which are in your tongue. The nbumbc commands/functions discussed here are listed at the beginning of the index.

1. Scalar Calculations

1.1. Simple Arithmetical Operations

MATLAB can be used as a scientific calculator. To begin a MATLAB session, click on a MATLAB icon or type `matlab` in a terminal and wait for the prompt, i.e., `>>`, to appear. (To exit MATLAB, click on Exit MATLAB in the File menu item or type `exit` or `quit`.) You are now in the MATLAB workspace.

You can calculate 3.17 * 5.7 + 17/3 by entering

```
3.17 * 5.7 + 17/3
```

and 20 by entering

```
20
```

And \( \sum_{j=1}^{10} \frac{1}{j} \) can be entered as

```
x = 1; x = x + 1/5 + 1/4 + 1/3 + 1/2 + 1 + 1/2 + 1/3 + 1/4 + 1/5
```

You can enter a number in scientific notation using the \( \times \) operator. For example, you can enter

```
2e-20
```

MATLAB, however, uses \( \times \) to represent \( \times 10^x \) so that MATLAB displays

```
2.0000e-20
```

The "standard" way to input \( 2 \times 10^{-20} \) is as \( 2e-20 \) or \( 2 \times 10^{-20} \) or \( 2 \times 10^{-20} \) (even \( 2.000000000000000000020 \) is acceptable).

Warning: \( 10^{20} \) cannot be input as \( e-20 \), but must be input as \( 1e-20 \) or \( 1E-20 \) or \( 1.0e-20 \) or \( 1.0E-20 \) or \( 1.000000000000000000020 \).

MATLAB can also handle complex numbers, where \( 1 \) or \( j \) represents \( \sqrt{-1} \). For example, \( 5j \) can be input as \( 51 \) or \( 5 \times 10^1 \), while \( 5 \times 10^8 \) cannot be input as \( 5e301 \) or \( 5e301 \) or \( 5 \times 10^{301} \). MATLAB considers to be \( 5 \times 10^{301} \). To calculate \( (2 + 3) \), enter

```
(2 + 3)
```

and MATLAB returns \( 5 \).

You can also save all of your input to MATLAB and most of the output (plots are not saved) by using the `diary` command. This archive of your work can be invaluable when you are solving homework problems. You can later use an editor to extract the part you want to turn in, while "burying" all the false starts and typing mistakes that occur. Conversely, if you are involved in a continuing project, this archive can be invaluable in keeping a record of your progress.

If you do not specify a file, this archive is saved to the file `diary` in the present directory. If the file already exists, this is appended — but if it is very long, this can get very time-consuming. Instead, you can use the `up-arrow` key, i.e., \( \uparrow \), to move backward, one statement at a time (or the down-arrow key, i.e., \( \downarrow \), to move forward). Then hit the enter (or return) key to execute the line.
Variables

Notation: We always use lowercase letters to denote scalar variables.

- Variables can be used to store numerical values. For example, you can store the value $2^{1/3}$ in the variable $x$ by entering:
  \[ x = 2^{1/3} \]
  This variable can then be used on the right-hand side of an equation such as:
  \[ fx = 3x^2 - 17xy + 79 \]
  There can also be more than one statement on a line. For example, if you type:
  \[ x = 2^{1/3}; fx = 3x^2 - 17xy + 79; g = 3/fx; \]
  then all three statements will be executed. Nothing will be printed out because semicolons follow each statement. If you want everything printed out then type:
  \[ x = 2^{1/3}, fx = 3x^2 - 17xy + 79, g = 3/fx; \]
  Thus, you can separate statements on a line by commas or semicolons. If semicolons are used, the results of the statement are not displayed, but if commas are used, the results appear on the computer screen.

Warning: A variable can be overwritten at will. For example, at present $x = 2^{1/3}$. If you now type:
  \[ x = x + 5 \]
  then $x$ becomes $2^{1/3} + 5$. No warning messages are printed if a variable is overwritten, just as in a programming language.

Although we do not discuss vectors and matrices until the next section, it is important to understand that MATLAB considers scalar variables to be vectors of length one or matrices of size $1 \times 1$. For example, if you type:
  \[ fx \]
  the number $57$ is returned. But you can also type:
  \[ fx(1) \]
  or
  \[ fx(1,1) \]
  and obtain the same result.

Character strings can also be stored in variables. For example, to store the string “And now for something completely different” in a variable, enter:
  \[ str = ’And now for something completely different’ \]
  (We discuss text variables in more detail in Section 3.)

Note: To put a single quote mark into the string, use two single quote marks.

You can change a variable from a scalar to a vector or a matrix whenever you desire — or whenever you forget that the variable has already been defined. Unlike C, for example, variables do not need to be declared (or typed). A variable springs into existence the first time it is assigned a value, and its type depends on its context.

At start-up time, MATLAB also contains some predefined variables. Many of these are contained in the table below. Probably the most useful of these is pi.

Warning: Be careful since you can redefine these predefined variables. For example, if you type:
  \[ pi = 2 \]
  then you have redefined $\pi$ — and no error messages will be printed out!

Another very useful predefined variable is ans, which contains the last calculated value which was not stored in a variable. For example, it sometimes happens that you forget to put a value into a variable. Then MATLAB sets the expression equal to the variable ans. For example, if you type:
  \[ (3.2 + 17.5 - 5/3)^2 \]
  but then realize that you wanted to save this value, simply enter:
  \[ x = ans \]
  and $x$ now contains $(3.2 + 17.5 - 5/3)^2$.

We will discuss character strings in detail in Section 3.3. For now, put the text “Silly Walks” into the variable $x$:
  \[ x = ’Silly Walks’ \]
  In MATLAB it is trivial to display a variable: simply type it. For example, if $x$ has the value $-23.6$ then:
  \[ x \]
  returns:
  \[ x = -23.6000 \]
  It is sometimes useful to display the value of a variable or an expression or a character string without displaying the name of the variable or ans. This is done by using disp. For example:
  \[ \text{disp}(x) \]
  \[ \text{disp}(pi/3) \]
  \[ \text{disp}(’And now for something completely different’) \]
  \[ \text{disp}(-23.6000) \]
  \[ 31.0063 \]
  And now for something completely different
  
  (The function fprintf, which will be discussed in Section 6, allows much finer formatting of variables.)

Note: When disp displays a variable or an array or an expression, it follows with a blank line. However, when it displays a string or a string variable, it does not.

Incidentally, a valid name for a MATLAB variable is a character string containing letters (upper or lower case), digits, and underscores where the first character must be a letter. The maximum length of a name is too long to worry about. However, there are a few names which are reserved because they have special meanings. The reserved words, called keywords, are:

<table>
<thead>
<tr>
<th>Keyword</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>break</td>
<td>continue</td>
</tr>
<tr>
<td>case</td>
<td>continue</td>
</tr>
<tr>
<td>catch</td>
<td>elseif</td>
</tr>
<tr>
<td>classif</td>
<td>end</td>
</tr>
<tr>
<td></td>
<td>exact</td>
</tr>
<tr>
<td></td>
<td>global</td>
</tr>
<tr>
<td></td>
<td>implements</td>
</tr>
<tr>
<td></td>
<td>import</td>
</tr>
<tr>
<td></td>
<td>interface</td>
</tr>
<tr>
<td></td>
<td>module</td>
</tr>
<tr>
<td></td>
<td>namespace</td>
</tr>
<tr>
<td></td>
<td>operator</td>
</tr>
<tr>
<td></td>
<td>package</td>
</tr>
<tr>
<td></td>
<td>procedure</td>
</tr>
<tr>
<td></td>
<td>property</td>
</tr>
<tr>
<td></td>
<td>public</td>
</tr>
<tr>
<td></td>
<td>private</td>
</tr>
<tr>
<td></td>
<td>protected</td>
</tr>
<tr>
<td></td>
<td>static</td>
</tr>
<tr>
<td></td>
<td>struct</td>
</tr>
<tr>
<td></td>
<td>typedef</td>
</tr>
<tr>
<td></td>
<td>type</td>
</tr>
<tr>
<td></td>
<td>union</td>
</tr>
<tr>
<td></td>
<td>void</td>
</tr>
<tr>
<td></td>
<td>while</td>
</tr>
</tbody>
</table>

(Of course, you can still use end or END — but you probably shouldn’t.)

Variables can also be deleted by using clear. For example, to delete $x$ type:
  \[ clear \]

Warning: This is a very dangerous command because it is so easy to lose a great deal of work.

If you mean to type:
  \[ clear \]
  \[ but instead you type \]
  \[ clear x \]
  \[ you will delete all the variables you have created in the workspace!\]
The most important principle for you to understand about computers is the following:

**Predefined Variables**

- `ans`: The default variable name when one has not been specified.
- `pi`: \( \pi \)
- `eps`: Approximately the smallest positive real number on the computer such that \( 1 + \text{eps} \neq 1 \).
- `Inf`: \( \infty \) (as in \( 1/0 \)). You can also type `inf`.
- `NaN`: Not-a-Number (as in \( 0/0 \)). You can also type `nan`.
- `i`: \( \sqrt{-1} \) (the same as \( z \) because engineers often use these interchangeably).
- `realmin`: The smallest “usable” positive real number on the computer. This is “approximately” the smallest positive real number that can be represented on the computer (on some computer `realmin`/2 returns 0).
- `realmax`: The largest “usable” positive real number on the computer. This is “approximately” the largest positive real number that can be represented on the computer (on most computer `2*realmax` returns `Inf`).

**Variables**

- `disp(X)`: Displays a variable (including an array) or a string without printing the variable name or `ans`.
- `;`: Separates multiple statements on the same line. The results appear on the screen.
- `:`: When this ends a MATLAB statement, the result is not printed on the screen. This can also separate multiple statements on the same line.

**About Variables**

- **This is a very dangerous command — use it at your own risk.**

### 1.3. Round-off Errors

The most important principle for you to understand about computers is the following. **Principle 1.1. Computers cannot add, subtract, multiply, or divide correctly!**

Computers do integer arithmetic correctly (as long as the numbers are not too large to be stored in the computer). However, computers cannot store most floating-point numbers (i.e., real numbers) correctly. For example, the fraction \( 1/3 \) is equal to the real number 0.3333... Since a computer cannot store this infinite sequence of threes, the number has to be truncated. `eps` is “close to” the difference between the exact number \( 1/3 \) and the approximation to \( 1/3 \) used in MATLAB. `1 + eps` is the smallest floating-point number after 1 which can be stored precisely in the computer. For example, in MATLAB `1 + 0.1` is clearly greater than 1; however, on our computer `1 + 1e-40` is not. To see this, when we enter `1 + 1e-40 - 1` we obtain 0.1000 as expected.

**Note:** MATLAB guarantees that the expression in parentheses is evaluated first, and then 1 is subtracted from the result. However, when we enter `1 + 1e-40 - 1` MATLAB returns `0` rather than `1.6e-40`. The smallest positive integer `n` for which `(1 + 10^n*eps) - 1` returns `0` is computer dependent. (On our computer it is 16.) What is not computer dependent is that this leads to errors in numerical calculations. For example, when we enter `n = 5; (n^((1/3)))^3 = n` MATLAB returns `-1.77635683989344` rather than the correct result of 0. If you obtain 0, try some different values of `n`. You should be able to rerun the last statement executed without having to retype it by using the up-arrow key. Alternatively, on a Mac or a PC use the `copy` command in the menu; in Unix enter `cp`.

**Note:** It might not seem important that MATLAB does not do arithmetical operations precisely. However, you will see in Section 5.2 that there are simple examples where this can lead to very incorrect results.

One function which is occasionally useful when you are just “playing around” is the `input` function, which displays a prompt on the screen and waits for you to enter some input from the keyboard. For example, if you want to try some different values of \( n \) in experimenting with the expression \((n^{1/3})^3 - n\), enter `n = input('n = '); (n^((1/3)))^3 = n` the argument to the function `input` is the string which prompts you for input, and the input is stored in the variable `n`; the semicolon `;` keeps the result of this function from being printed out. You can easily rerun this line for different values of `n` (as we described above) and explore how round-off errors can affect simple expressions.

**Note:** You can input a character string in two ways:

- `str = input('input string: ')`;
- and then enter, for example, ‘Upper class twit of the year’ or `str = input('input string: ', 's');`
- and then enter `Upper class twit of the year`

The first must enter a string while the second simply enters text until the line ends.

**Warning:** `eps` and `realmin` are very different numbers. `realmin` is approximately the smallest positive number that can be represented on the computer, whereas `eps` is approximately the smallest positive number on the computer such that `1 + eps ≠ 1`. (`eps/realmin` is larger than the total number of atoms in the known universe.)

### 1.4. Formatting Printing

The reason that \((n^{1/3})^3 - n\) can be nonzero numerically is that MATLAB only stores real numbers to a certain number of digits of accuracy. **Type**

```
log10(1/eps)
```

**and remember the integer part of this number.** This is approximately the maximum number of digits of accuracy of any calculation performed in MATLAB. For example, if you type `1/3` in MATLAB the result is only accurate to approximately this number of digits. You do not see the decimal representation of `1/3` to this number of digits because on start-up MATLAB only prints the result to four decimal digits — or five significant digits if scientific notation is used (e.g., the calculation `1/30000` is displayed in scientific notation). To change how the results are printed out, use the `format` command in MATLAB. Use each of these six `format` functions and then type in `1/3` to see how the result is printed out.

```
```

```
the term within parentheses is used in the calculation of important numbers. In fact, Mathematica has introduced two functions to reduce roundoff error: 

\[
\text{exp}(x) - 1 \approx \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots 
\]

for all angles \( x \). Mathematica tries to calculate the correct result, but it cannot quite do it. It is up to you to interpret what Mathematica is trying to tell you.

Note: If you obtained zero for the above calculation, try

\[
\text{inc}(\text{deg} + \pi/180); x = (\cos(\text{deg})^2 + \sin(\text{deg})^2)^{1/2} 
\]

for various values. Some of these calculations should be nonzero.

There are a number of issues in this overview where we reiterate that Mathematica cannot usually calculate results exactly. Sometimes these errors are small and unimportant—other times they are very important. In fact, Mathematica has introduced two functions to reduce round-off errors. The relative error in the calculation of \( e^x - 1 \), i.e., \( (\exp(x) - 1) / \exp(x) \) can be very large when \( x < 1 \) since

\[
e^x - 1 \approx \left(1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots \right) - 1 - \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots
\]

the term within parentheses is very close to 1 and so subtracting by 1 causes a loss of many digits in the result. For example,

\[
\exp(1.0e-8) - 1 = 9.99999993825290e-09
\]

\[
\exp(1.0e-8) + 1 = 1.00000000000000e+08
\]

\[
\exp(1.0e-20) - 1 = 0
\]

\[
\exp(1.0e-20) = 9.99999999999999e-21
\]

Similarly,

\[
\log_2 \left( \frac{z-1}{1} \right) = \frac{(z-1)^1}{2} + \frac{(z-1)^2}{3} + \frac{(z-1)^3}{4} + \cdots
\]

so if \( z = 1 \) then accuracy is lost in the calculation of \( z-1 \). This can be avoided by entering \( x \) directly in

\[
\log(1+x) = \frac{x}{1} - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots
\]

which is evaluated by \( \log \text{int}(x) \).

Warning: There is one technical detail about functions that will trip you up occasionally: how does Mathematica determine whether a word you enter is a variable or a function? The answer is that Mathematica first checks if the word is a variable and, only if it fails, does it check if the word is a function. For example, suppose you enter

\[
\text{sin} = 20
\]

by mistake (possibly you meant \( \text{sin} = 20 \) but were thinking about something else). If you now type

\[
\text{sin}(3)
\]

MATLAB will reply

\[
??? \text{Index exceeds matrix dimensions.}
\]

because it recognizes that \( \text{sin} \) is a variable. Since Mathematica considers a variable to be a vector of length one, its complaint is that you are asking for the value of the third element of the vector \( \sin \) (which only has one element). Similarly, if you enter

\[
\text{sin}(25*pi)
\]

MATLAB will reply

\[
??? \text{Index exceeds matrix dimensions.}
\]

because it thinks you are asking for the 25\( \pi \)-th element of the vector \( \sin \). The way to undo your mistake is by typing

\[
\text{clear sin}
\]
1.6. Complex Numbers

MATLAB can work with complex numbers as easily as with real numbers. For example, to find the roots of the quadratic polynomial \( x^2 + 2x + 5 \) enter

```
>> a = 1; b = 2; c = 5;
>> x1 = (-b + sqrt(b^2 - 4*a*c)) / (2*a)
>> x2 = (-b - sqrt(b^2 - 4*a*c)) / (2*a)
```

The output is

\[-1.0000 + 2.0000i\]

and

\[-1.0000 - 2.0000i\]

As another example, to calculate \( e^{i\pi/2} \) enter

```
>> exp(i*pi/2)
```

and obtain

\[0.0000 + 1.00000i\]

There are standard functions for obtaining the real part, the imaginary part, and the complex conjugate\(^\dagger\) of a complex number or variable. For example,

```
>> x = 3 - 5i
>> real(x)
>> imag(x)
>> conj(x)
```

returns \(3 - 5i\), and \(0.0000 + 5.00000i\) respectively.

Note that many of the common mathematical functions can take complex arguments. Above, MATLAB has calculated \( e^{i\pi/2} \), which is evaluated using the formula

\[e^{i\theta} = \cos \theta + i \sin \theta\]

Similarly,

\[\cos z = \frac{e^{iz} + e^{-iz}}{2}\]

\[\sin z = \frac{e^{iz} - e^{-iz}}{2i}\]

\(^\dagger\)This function is in the symbolic math toolbox. If it is not on your computer, the code is shown on page 103.

### Some Common Real Mathematical Functions

<table>
<thead>
<tr>
<th>Function (b)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>abs(x)</td>
<td>The absolute value of x.</td>
</tr>
<tr>
<td>acos(x)</td>
<td>arccos x.</td>
</tr>
<tr>
<td>acosh(x)</td>
<td>arccosh x.</td>
</tr>
<tr>
<td>acot(x)</td>
<td>arccot x.</td>
</tr>
<tr>
<td>acoth(x)</td>
<td>arccoth x.</td>
</tr>
<tr>
<td>asin(x)</td>
<td>arcsin x.</td>
</tr>
<tr>
<td>asind(x)</td>
<td>arcsind x.</td>
</tr>
<tr>
<td>asinh(x)</td>
<td>arsinh x.</td>
</tr>
<tr>
<td>atan(x)</td>
<td>arctan x.</td>
</tr>
<tr>
<td>atand(x)</td>
<td>arctand x.</td>
</tr>
<tr>
<td>atanh(x)</td>
<td>arctanh x.</td>
</tr>
<tr>
<td>cos(x)</td>
<td>cos x.</td>
</tr>
<tr>
<td>cosd(x)</td>
<td>cos x where x is in degrees.</td>
</tr>
<tr>
<td>cosh(x)</td>
<td>cosh x.</td>
</tr>
<tr>
<td>cot(x)</td>
<td>cot x where x is in degrees.</td>
</tr>
<tr>
<td>cotd(x)</td>
<td>cot x where x is in degrees.</td>
</tr>
<tr>
<td>csc(x)</td>
<td>csc x.</td>
</tr>
<tr>
<td>cscd(x)</td>
<td>csc x where x is in degrees.</td>
</tr>
<tr>
<td>csche(x)</td>
<td>csche x.</td>
</tr>
<tr>
<td>exp(x)</td>
<td>(e^x).</td>
</tr>
<tr>
<td>expm1(x)</td>
<td>(e^x - 1).</td>
</tr>
<tr>
<td>factorial(n)</td>
<td>n! for n a non-negative integer.</td>
</tr>
<tr>
<td>fix(x)</td>
<td>If (x \geq 0) this is the largest integer which is (\leq x).</td>
</tr>
<tr>
<td>floor(x)</td>
<td>This is the largest integer which is (&lt; x).</td>
</tr>
<tr>
<td>heaviside(x)(^\dagger)</td>
<td>If (x &gt; 0) this returns 1, if (x &lt; 0) return 0, and if (x = 0) this returns (1/2).</td>
</tr>
<tr>
<td>log(x)</td>
<td>The natural log of x, i.e., (\ln x).</td>
</tr>
<tr>
<td>log2(x)</td>
<td>The common log of x, i.e., (\log_2 x).</td>
</tr>
<tr>
<td>log10(x)</td>
<td>The modulus after division.</td>
</tr>
<tr>
<td>logit(x)</td>
<td>(\log(x+1)).</td>
</tr>
<tr>
<td>mod(x,y)</td>
<td>(x \mod y) where (n = \text{floor}(x/y)).</td>
</tr>
<tr>
<td>rem(x,y)</td>
<td>The remainder of (x/y).</td>
</tr>
<tr>
<td>round(x)</td>
<td>The integer which is closest to (x).</td>
</tr>
<tr>
<td>sec(x)</td>
<td>(1/\cos x).</td>
</tr>
<tr>
<td>secd(x)</td>
<td>(1/\cos x) where x is in degrees.</td>
</tr>
<tr>
<td>sign(x)</td>
<td>(x/</td>
</tr>
<tr>
<td>sinh(x)</td>
<td>(\sinh x).</td>
</tr>
<tr>
<td>sinh(x)</td>
<td>(\sinh x).</td>
</tr>
<tr>
<td>sqrt(x)</td>
<td>(\sqrt{x}).</td>
</tr>
<tr>
<td>tand(x)</td>
<td>(\tan x).</td>
</tr>
<tr>
<td>tanh(x)</td>
<td>(\tanh x).</td>
</tr>
<tr>
<td>angle(x)</td>
<td>The angle of (x).</td>
</tr>
<tr>
<td>conj(z)</td>
<td>(x^\ast + y^\ast i).</td>
</tr>
<tr>
<td>imag(z)</td>
<td>The imaginary part of (z), i.e., (y).</td>
</tr>
<tr>
<td>real(z)</td>
<td>The real part of (z), i.e., (x).</td>
</tr>
</tbody>
</table>

### Some Common Complex Mathematical Functions

<table>
<thead>
<tr>
<th>Function (b)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>abs(z)</td>
<td>The absolute value of (z = x + iy).</td>
</tr>
<tr>
<td>angle(z)</td>
<td>The angle of (z).</td>
</tr>
<tr>
<td>conj(z)</td>
<td>(x^\ast + y^\ast i).</td>
</tr>
<tr>
<td>imag(z)</td>
<td>The imaginary part of (z), i.e., (y).</td>
</tr>
<tr>
<td>real(z)</td>
<td>The real part of (z), i.e., (x).</td>
</tr>
</tbody>
</table>

1.7. Script M-files

So far we have always entered MATLAB statements directly into the text window so that they are executed immediately. However, if we want to repeatedly execute a number of statements we have to put them all on one line and recutecte the whole line. This line can get very long on the screen! The solution is to type the sequence of statements in a separate file named \(<\text{file name}>.m\). It is easy to edit this file to remove any errors, and the sequence can be executed whenever desired by typing

```
>> use <file name>
```

The MATLAB statements themselves are not printed out, but the result of each statement is, unless a semicolon ends it. This type of file is called a script M-file: when MATLAB executes the statement \(<\text{file name}>\) the contents of the file \("<\text{file name}>.m\)" are executed just as if you had typed them into the text window. We will not emphasize script M-files further, but you will find many occasions where they are very helpful.

You can easily work on a script M-file by clicking on the menu item File and then on Run to create a new m-file and finally, if it asks, click on Script. Or click on File and then Open to open an already existing one (if you want to modify it). You can also create a new m-file or open an already existing one by

```
>> edit <file name>.m
```

Warning: The name of the file includes the extension \(".m\), i.e., \("<\text{file name}>.m\)\), but you execute it in MATLAB by typing \(<\text{file name}>\), i.e., without the extension.

Warning: The file name can consist of (almost any number of) letters (lowercase and/or uppercase), numbers, and underscores, i.e., \("\_\)\.

The first character must be a letter.

Spaces are not allowed.

Warning: There is one point we cannot overemphasize. Make sure your file name is not the same.

\(^\dagger\)If \(a\) is a complex number, then its complex conjugate, denoted by \(a^\ast\) is obtained by changing the sign of \(i\) whenever it appears in the expression for \(a\). For example, if \(a = 3 + 17i\), then \(a^\ast = 3 - 17i\); if \(a = e^{i\pi/4}\), then \(a^\ast = e^{-i\pi/4}\); if \(a = (2 + 3i)(1 + 3i)/(3 - \sqrt{5})\), then \(a^\ast = (2 - 3i)(1 - 3i)/(3 + \sqrt{5})\).
as one of MATLAB's commands/functions. If it is, your file might not execute — MATLAB's might! Or you might run a MATLAB function which tries to call a function which has the same name as yours! (The m-file it executes depends on the order in which directories are searched for m-files — see path for more details.) To check this, you can enter

- type <file name>

before you save your m-file. This will type out the entire file if it is written in MATLAB or type out

<f ile name> is a built-in function
if it is written in C or Fortran and so cannot be viewed directly. If the file name doesn't exist, MATLAB returns

??? Undefined function or variable '<f ile name>'
If MATLAB returns your m-file, it means you have already saved it. In this case enter

- doc <f ile name>

(which is discussed in the next section), which returns useful information about a MATLAB function, i.e., not one of yours. If it cannot find this particular function, you are safe.

A long expression can be continued to a new line by typing three periods followed by the "enter (or 'return')" key. For example, \[ \sum_{j=1}^{\infty} 1/j \] can be entered as

\[
1 + 1/2 + 1/3 + 1/4 + 1/5 + 1/6 + 1/7 + 1/8 + 1/9 + 1/10 + 1/11 + 1/12 + \ldots
\]

although there are much better ways to obtain this same expression with many fewer keystrokes (as you will see in Section 2.8). Lines can also be continued in the MATLAB workspace by using three periods, but it is much more common to use continuation in an m-file.

If your m-file is very long, it is often valuable to include comments to explain what you are doing. Each line of comments must begin with the percent character, i.e., "%". Comments can appear alone on a line or they can follow a statement that you have entered.

Options and Exit

edit Create a new m-file or edit an already existing one.
type Displays the actual MATLAB code for a command/function.
... Continue an expression onto the next line.
% Begin a comment

1.8. Help!

Before discussing how to obtain help in MATLAB, here is a good place to discuss a very frustrating situation where you desperately need help: how do you abort a MATLAB statement which is presently executing. The answer is simply to type -c (that is, hold down the control key and type "c").

The on-line help facility in MATLAB is quite extensive. If you type

- help

you will get a list of all the topics that you can peruse further by typing help followed by the name of the topic. If you want help on a specific command/function, simply type help followed by the name of the command, i.e.,

- help <command/function>

For example, if you forget the exact form of the format command, just type

- help format

and you will see all the various ways that the output can be formatted.

Note: Typing

- help ?
gives you lots of information about arithmetical and relational and logical operators and special characters.

There is a more general command that can help you determine which commands might be of use. The command lookfor searches through the first line of all MATLAB help entries for a particular string. It is case insensitive so capital letters need not be used. For example,

- lookfor plot

returns all the MATLAB functions that have something to do with plots. (There are over one hundred!) This function may be useful — or it may not be. However, it is worth a try if you cannot remember the name of the one you want to use.

Warning: All of the thousands of MATLAB commands/functions have to be checked, so this might run slowly.

Note: The string need not be a complete word. For example, the string compl is contained in the words "complement", "complex", "complete", "completion", and "incomplete" — and in the capitals of all these words.

If you want to find out more about a specific command/function, enter

- type <command/function>

or

- edit <command/function>

If the command is written in MATLAB's programming language (as discussed in Section 8), it will be typed out for the former and appear in a new window in the latter. However, these cannot return anything useful on internal MATLAB commands/functions, called built-in functions, which are coded in C.) MATLAB also has an entire reference manual on-line which can be accessed by entering

- doc

or

- helpbrowser

This hypertext documentation is displayed using your Web browser. It generally gives much more information than the help command, and is a more easily understood format.

After working for a while, you may well forget what variables you have defined in the workspace. Simply type who or whos to get a list of all your variables (but not their values). who also returns the names of the variables you have defined, while whos also returns the size and type of each variable. To see what a variable contains, simply type the name of the variable on a line.

By the way, the demonstrations available by running demo show many of the capabilities of MATLAB and include the actual code used. This is always a good place to look if you are not sure how do something.

Two functions that don’t quite fit in any category are save and load. However, since these are occasionally very helpful, this is a good place to discuss them. Occasionally, you might need to save one or more MATLAB variables; it might have taken you some time to generate these variables and you might have to quit your MATLAB session without finishing your work — or you just might be afraid that you will overwrite some of them by mistake. The save command saves the contents of all your variables to the file "matlab.mat". Use doc to learn how to save all the variables to a file of your own choice and how to save just some of the variables. The load command loads all the saved variables back into your MATLAB session.1 (As we discuss in Section 4.1, the load command can also be used to input our own data into MATLAB.)

1These variables are saved in binary format; when loaded back in using load the variables will be exactly the same as before. The contents of this file can be viewed by the user with an editor — but the contents will appear to be gibberish. The contents can only be interpreted by the load command.
1.9. Be Able To Do

After reading this section you should be able to do the following exercises. The solutions are given on page 103.

1. Consider a triangle with sides a, b, and c and corresponding angles ∠α, ∠β, and ∠γ.
   (a) Use the law of cosines, i.e.,
   \[ c^2 = a^2 + b^2 - 2ab \cos \angle \alpha \]
to calculate c if \( a = 3 \), \( b = 5 \), and \( \angle \alpha = 70^\circ \).
   (b) Then show \( c \) to its full accuracy.
   (c) Use the law of sines, i.e.,
   \[ \frac{\sin \angle \alpha}{c} = \frac{\sin \angle \beta}{b}, \]
to calculate \( \angle \beta \) in degrees and show it in scientific notation.
   (d) What MATLAB command should you use first if you wanted to save these results to the file triangle.ans?

2. Calculate \( \frac{\sqrt{3} - 2}{\sqrt{3} + 1} \).
   3. Check whether this is also true numerically when using MATLAB by using a number of different values of \( \theta \). Use MATLAB statements which make it as easy as possible to do this.
   4. How would you find out information about the fix function?

2. Arrays: Vector and Matrix Calculations

In the previous section we discussed operations using single numbers, i.e., scalars. In this section we discuss operations on sets of numbers called arrays. Until the advanced subsection at the end, we restrict our attention to one-dimensional arrays, which are called vectors, and two-dimensional arrays, which are called matrices. In this section we will generally refer to these sets of numbers specifically as vectors or matrices rather than use the more inclusive term “arrays.” MATLAB was originally developed specifically to work with vectors and matrices, and that is still one of its primary uses.

Notation: We will always write matrices using capital letters and vectors using lower case letters.

This makes it much easier to understand MATLAB operations. This is also a good practice for you to use.

In addition, when we write “vector” we mean a column vector and so it is immediately obvious that \( A^T \) is a legitimate operation of a matrix times a vector as long as the number of columns of the matrix \( A \) equals the number of rows of the column vector \( x \). Also, \( A^T A \) is legitimate because the column vector \( x \) has only one column while the matrix \( A \) is expected to have more than one row. On the other hand, \( x^T A \) is legitimate (\( x^T \) denotes the conjugate transpose of the vector \( x \)) as long as the row vector \( x^T \) has the same number of columns as the number of rows of the matrix \( A \).

In addition, we have very specific notation for denoting vectors and matrices and the elements of each. We collect all this notation here.

Notation: \( R^m \) denotes all real column vectors with \( m \) elements and \( C^m \) denotes all complex column vectors with \( m \) elements.

\( R^{m \times n} \) denotes all real \( m \times n \) matrices (i.e., having \( m \) rows and \( n \) columns) and \( C^{m \times n} \) denotes all complex \( m \times n \) matrices.

In this overview the word “vector” means a column vector so that \( C^m = C^{1 \times m} \). Vectors are denoted by boldface letters, such as \( x \); we will write a row vector as, for example, \( x^T \), where \( x^T \) denotes the transpose of a matrix or vector (that is, the rows and columns are reversed).

Notation: \( \mathbf{A} = (a_{ij}) \) means that the \((i,j)\) element of \( \mathbf{A} \) (i.e., the element in the \( i^{th} \) row and the \( j^{th} \) column) is \( a_{ij} \).

\( x = (x_i) \) means the \( i^{th} \) element of \( x \). By the way MATLAB works with complex matrices as well as it does real matrices. To remind you of this fact, we will use \( C \) rather than \( R \) unless there is a specific reason not to. If there is a distinction between the real and complex case, we will first describe the real case and then follow with the complex case in parentheses.

2.1. Generating Matrices

To generate the matrix
\[
\begin{pmatrix}
1 & 2 & 3 \\
4 & 5 & 6 \\
7 & 8 & 9
\end{pmatrix}
\]
in MATLAB type
\[
\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}
\]
where \( \{,\} \) denotes one or more spaces) or
\[
\mathbf{A} = \begin{bmatrix} 1,2,3 & 4,5,6 & 7,8,9 \end{bmatrix}
\]
or
\[
\mathbf{A} = \begin{bmatrix} \{1,2,3, \} & \{4,5,6, \} & \{7,8,9, \} \end{bmatrix}
\]
In other words, either spaces or commas can be used to delineate the elements of each row of a matrix; semicolons are required to separate rows. (Any number of spaces can be put around commas or semicolons to improve the readability of the expression.)

Notation: Since we prefer spaces, we will generally use them rather than commas to separate elements in a row.

Rors can also be separated by beginning each on a separate line. For example, the matrix \( A \) can also be entered by
\[
\mathbf{A} = \begin{bmatrix} 1,2,3 \\ 4,5,6 \\ 7,8,9 \end{bmatrix}
\]
However, we do not use it because there is no way to correct an element on a previous line which you have just noticed has been entered incorrectly. The more complicated matrix
\[
\begin{pmatrix}
1 & 2 + \sqrt{7} & 3 \sin 1 \\
17/3 & \pi + 3 \\
1/3 & 2 - \sqrt{7} & -7 \cos \pi/7
\end{pmatrix}
\]
can be entered by typing
\[ C = \begin{bmatrix} 1 & 2 -\sqrt{3} & 3 \times \sin(1) \end{bmatrix}; \quad \exp(2) = 17/3 \pi^3; \quad \frac{1}{3} \times 2 -\sqrt{3}, -7 \times \cos(\pi/7) \]

or
\[ C = \begin{bmatrix} 1 & 2 -\sqrt{3}, 3 \times \sin(1) \end{bmatrix}; \quad \exp(2) = 17/3 \pi^3; \quad 1/3, 2 -\sqrt{3}, -7 \times \cos(\pi/7) \]

Warning: When an element of a matrix consists of more than one term, it is important to enter all the terms without spaces — unless everything is enclosed in parentheses. For example,
\[ x_1 = \begin{bmatrix} 1 \pi^3 \end{bmatrix} \]
is the same as
\[ x_2 = \begin{bmatrix} 1 \pi^3 \end{bmatrix} \]
and is the same as
\[ x_3 = \begin{bmatrix} 1 (\pi^3 + 3) \end{bmatrix} \]
but is not the same as
\[ x_4 = \begin{bmatrix} 1 \pi^3 + 3 \end{bmatrix} \]
\( \% \) not the same as the previous three statements (Try it!) In other words, MATLAB tries to understand what you mean, but it does not always succeed.

Definition

The transpose of a matrix \( A \in \mathbb{C}^{m \times n} \), denoted by \( A^T \), is obtained by reversing the rows and columns of \( A \). That is, if \( A = (a_{ij}) \) then \( A^T = (a_{ji}) \). (For example, the \((2,4)\) element of \( A^T \), i.e., \( i = 2 \) and \( j = 4 \), is \( a_{42} \)).

A square matrix \( A \) is symmetric if \( A^T = A \).

The conjugate transpose of a matrix \( A \in \mathbb{C}^{m \times n} \), denoted by \( A^* \), is obtained by reversing the rows and columns of \( A \) and then taking the complex conjugates of all the elements. That is, if \( A = (a_{ij}) \) then \( A^* = (\bar{a}_{ji}) \), where \( \bar{a}_{ji} \) denotes the complex conjugate of \( a_{ji} \). A square matrix \( A \) is Hermitian if \( A^* = A \).

Note: In MATLAB \( A^T \) is calculated by \( A' \) (i.e., a period followed by a single quote mark), while \( A^* \) is calculated by \( A \) (i.e., just a single quote mark).

A vector can be entered in the same way as a matrix. For example, the vector
\[ x = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \end{bmatrix} \]
can be entered as
\[ x = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \end{bmatrix} \]
However, this requires many semicolons; instead, take the transpose of a row vector by entering
\[ x = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \end{bmatrix}^T \]
where the MATLAB symbol for the transpose, i.e., \( \text{T} \), is \( \text{.'} \) (i.e., a period followed by a single quote mark). There is one further simplification that is usually observed when entering a vector. The MATLAB symbol for the conjugate transpose, i.e., \( \text{~} \), or of a matrix is \( \text{.'} \) (i.e., just a single quote mark), which requires one less character than the symbol for the transpose. Thus, \( x \) is usually entered as
\[ x = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \end{bmatrix} \]
There is a simpler way to generate \( x \), namely using the colon operator. It won’t be discussed in detail for a few pages, but in its simplest form
\[ x = \begin{bmatrix} 1:6 \end{bmatrix} \]
That is, \( n \) is the same as \( 1, 2, 3, \ldots, n \).

Warning: \( x^T \rightarrow x \). \text{while} \( x^T \rightarrow x' \) \( \text{so that you can only calculate} \ x^T \ \text{by} \ x' \ \text{if} \ x \ \text{is real} \) This has been a constant problem! Sometimes the elements of a matrix are complicated enough that you will want to simplify the process of generating the matrix. For example, the vector \( x = \begin{bmatrix} 1 & \sqrt{2} & \sqrt{3} & \sqrt{4} & \sqrt{5} \end{bmatrix} \) can be entered by typing
\[ x = \begin{bmatrix} s & 2 & \sqrt{3} & s & 3 \end{bmatrix} \]
or, as we will see later, by
\[ x = \begin{bmatrix} s & 2 & 3 & 6 & 2 & 3 \end{bmatrix} \]
We have now discussed how to enter matrices into MATLAB by using square parentheses, i.e., \(...\). You work with individual elements of a matrix by using round parentheses, i.e., \(...\). For example, the element \( a_{ij} \) of the matrix \( A \) is \( A(i,j) \) in MATLAB. Suppose you want to create the matrix
\[ B = \begin{bmatrix} 1 & 2 & 3 \[4 & 5 & 6 \[7 & 8 & 10 \end{bmatrix} \]
without having to enter all nine elements. If \( A \) (see the beginning of this section) has already been generated, the simplest way is to type
\[ B = A; \]
\[ B = \begin{bmatrix} 1 & 2 & 3 \[4 & 5 & 6 \[7 & 8 & 10 \end{bmatrix} \]
or
\[ B = A; \]
\[ B = \begin{bmatrix} 1 & 2 & 3 \[4 & 5 & 6 \[7 & 8 & 10 \end{bmatrix} \]
That is, an element of an array
Also, the element \( x_1 \) of the vector \( x \) is \( x(1) \) in MATLAB. For example, to create the column vector
\[ x = \begin{bmatrix} 1, 2, 3, \ldots, 47, 48, 49, 51 \end{bmatrix} \]
can be entered by typing
\[ x = \begin{bmatrix} 1:50 \end{bmatrix}; \]
\[ x(50) = 51 \]
or
\[ x = \begin{bmatrix} 1:50 \end{bmatrix}; \]
\[ x(50) = 51 \]
or
\[ x = \begin{bmatrix} 1:50 \end{bmatrix}; \]
\[ x(\text{length}(x)) = x(\text{length}(x)) + 1 \]
where \( \text{length} \) returns the number of elements in a vector.
MATLAB also has a number of functions that can generate matrices. For example,
\[ C = \begin{bmatrix} 0 \end{bmatrix} \]
or
\[ C = \begin{bmatrix} 0 \end{bmatrix} \]
generates a \( n \times 5 \) zero matrix. Also,
\[ C = \begin{bmatrix} 0 \end{bmatrix} \]
generates a \( 5 \times 8 \) zero matrix. Finally, you can generate a zero matrix \( C \) with the same size as an already existing matrix, such as \( A \), by
\[ C = \begin{bmatrix} \text{zeros}(A) \end{bmatrix} \]
where \( \text{zeros}(A) \) is a row vector consisting of the number of rows and columns of \( A \). This function is frequently used to preallocate a matrix of a given size so that MATLAB does not need to “guess” how large to make it. Similarly, you can generate a matrix with all ones by \( \text{ones}(m,n) \), \( \text{ones}(m,n) \), or \( \text{ones}(m,n) \)
You can also generate the identity matrix, i.e., the matrix with ones on the main diagonal and zeroes off of it, by using the function \( \text{eye} \) with the same arguments as above.

Another useful matrix is a random matrix, i.e., a matrix whose elements are all random numbers.

The two most commonly used random numbers are uniformly distributed random numbers and normally distributed random numbers. Uniformly distributed random numbers in \([0,1]\) are generated by the \( \text{rand} \) function, which takes the same arguments as above, i.e.,
\[ r = \begin{bmatrix} \text{rand}(1) \end{bmatrix} \]
To generate uniformly distributed random numbers in \([a,b]\) use
\[ r = \begin{bmatrix} a + (b-a) \times \text{rand}(n) \end{bmatrix} \]
To be precise, these are \( \text{pseudo} \) random numbers because they are calculated by a deterministic formula which begins with an initial “seed” — which is called the \( \text{state} \), not the seed. Every time that a new MATLAB session is started, the default seed is set, and so the same sequence of random numbers will be generated. However, every time that this function is executed during a session, a different sequence of random numbers is generated. If desired, a different seed can be set at any time by entering
\[ \text{rng}('\text{default integer state number}') \]
To use a different sequence of random numbers every time you run MATLAB, begin your session with
\[ \text{rng}('\text{shuffle}') \]
To generate normally distributed random numbers with mean 0 and standard deviation \( \sigma \) use
\[ r = \begin{bmatrix} \text{randn}(m,n) \end{bmatrix} \]
To obtain mean \( \mu \) and standard deviation \( \sigma \) use
> R = mu + sig*randn(m, n)

The normal distribution (or, at least, a reasonable approximation to it) occurs much more frequently in the physical world than the uniform distribution. However, the uniform distribution has many uses, including generating the normal distribution!

There are two other probability distributions which are frequently needed. Random matrices of integers are generated by

> r = randi(k)

which generates one integer in the interval [1, k] or

> R = randi(k, n, m) % or randi(k, n, m)

which generates a matrix of integers. Similarly,

> r = randi([k1 k2]) % or randi([k1 k2], m)

which generates one integer in the interval [k1, k2].

There is another “random” function which is useful if you want to reorder a sequence, rather than just generate random elements from a set. The function randperm(n) generates a random permutation of the integers 1, 2, …, n. This is called a random selection without replacement.

MATLAB also makes it convenient to assemble matrices in “pieces”, that is, to put matrices together to make a larger matrix. That is, the original matrices are submatrices of the final matrix. For specificity, let us continue with A (see the beginning of this section). Suppose you want to assemble a 5×3 matrix whose first three rows are the rows of A and whose last two rows are all ones. This is easily generated by

> [ A ; ones(2, 3) ]

(The semicolon indicates that a row has been completed and so the next rows consist of all ones. The fact that A is a matrix in its own right is immaterial. All that is necessary is that the number of columns of A be the same as the number of columns of ones(2, 3).) This matrix could also be generated by

> [ A ; [1 1 1] ; [1 1 1] ]

or by

> [ A ; [1 ; 1 1 ; 1 1] ]

or even by

> [ A ; [1 1 1 1] ]

Similarly, to generate a 5×3 matrix whose first three columns are the columns of A and whose last column is (1, 1, 1, 1, 1)T type,

> [ A ; [1 1 1 1] ]

(The space following the A indicates that the next column is to follow. The fact that the next entry is a column vector is immaterial. All that is necessary is that the number of rows of A be the same as the number of rows in the new last column.)

2.2. The Colon Operator

For real numbers a and b the MATLAB notation

> [a:b] or, more simply,

> a:b

generates the row vector (a, a+1, a+2, …, a+b) where the integer k satisfies a + b ≤ k and a + (k + 1) > b. Thus, the vector x = (1, 2, 3, 4, 5, 6)T should be entered into MATLAB as

> x = [1:6]T

or even as

> x = [1:6]T

(although we can’t imagine why you would want to do it this way). If c is also a real number the MATLAB notation

> [a:c:b] or

> a:c:b

generates a row vector where the difference between successive elements is c. Thus, we can generate numbers in any arithmetic progression using the colon operator. For example, typing

> [18:3:2]

generates the row vector (18, 15, 12, 9, 6, 3). While typing

> [ pi : -2*pi : 0 ]

<table>
<thead>
<tr>
<th>Element Matrices</th>
</tr>
</thead>
<tbody>
<tr>
<td>zeros(n)</td>
</tr>
<tr>
<td>zeros(m, n)</td>
</tr>
<tr>
<td>zeros(size(A))</td>
</tr>
<tr>
<td>size(A)</td>
</tr>
<tr>
<td>eye</td>
</tr>
<tr>
<td>eye</td>
</tr>
<tr>
<td>rand</td>
</tr>
<tr>
<td>rand</td>
</tr>
<tr>
<td>randi</td>
</tr>
<tr>
<td>randn</td>
</tr>
<tr>
<td>rng</td>
</tr>
<tr>
<td>randperm(n)</td>
</tr>
<tr>
<td>size(A)</td>
</tr>
<tr>
<td>length(a)</td>
</tr>
<tr>
<td>numel(A)</td>
</tr>
<tr>
<td>A.'</td>
</tr>
<tr>
<td>A.'</td>
</tr>
</tbody>
</table>
generates the row vector \([1 \ 2 \ 3 \ 4 \ 5 \ 6]\).

Occasionally your fingers will go crazy and you will type something like

```matlab
>> [10:1]
```

(rather than \([1:10]\)). MATLAB will not complain; it will simply return \([1\] \) which is an empty matrix. **Warning:** There is a slight danger if \(c\) is not an integer. As an oversimplified example, entering

```matlab
>> x = [0.2 : 0.01 : 0.98]'
```

**should** generate the column vector \((0.02, 0.021, 0.022, \ldots, 0.979, 0.98)\). However, because of round-off errors in storing floating-point numbers, there is a possibility that the last element in \(x\) will be 0.979. The MATLAB package was written specifically to minimize such a possibility, but it still remains. We will discuss the function `linspace` which avoids this difficulty in Section 4.1. An easy “fix” to avoid this possibility is to calculate \(x\) by

```matlab
>> x = [20:0.001]/1000
```

2.3. Manipulating Vectors and Matrices

For specificity in this subsection we will mainly work with the 5x6 matrix

\[
\begin{bmatrix}
1 & 2 & 3 & 4 & 5 & 6 \\
7 & 8 & 9 & 10 & 11 & 12 \\
13 & 14 & 15 & 16 & 17 & 18 \\
19 & 20 & 21 & 22 & 23 & 24 \\
25 & 26 & 27 & 28 & 29 & 30
\end{bmatrix}
\]

which can be generated by

```matlab
```

**Note:** Spaces will frequently be used in generating vectors and matrices in this subsection for readability. You can use the colon notation to extract submatrices from \(E\). For example,

```matlab
>> F = E([1 3 5], [2 4 5])
```

extracts the elements in the first, third, and fifth rows and the second, third, fourth, and fifth columns of \(E\). Thus,

\[
F = \begin{bmatrix}
2 & 3 & 4 & 5 \\
14 & 15 & 16 & 17 \\
26 & 27 & 28 & 29
\end{bmatrix}
\]

You can generate this submatrix more easily by typing

```matlab
>> F = E(:,[2:5])
```

There is an additional shortcut you can use: in a matrix \(E\) by itself represents an entire row or column. For example, the second column of \(F\) is \(F(:,1)\) and the second row is \(F(2,:).\) To replace the second column of \(F\) by two times the present second column minus four times the fourth column enter

```matlab
>> F(:,2) = 2*F(:,2) - 4*F(:,4)
```

And suppose you now want to double all the elements in the last two columns of \(F\). Simple type

```matlab
>> F(:,3:4) = 2*F(:,3:4)
```

There is a last-additional shortcut you can use. Suppose you want the matrix \(F\) to consist of the odd rows of \(E\) and the second to the last column, as we did above. You might be changing the size of \(E\), and not want to have to remember how large it is. This can be easily done by

```matlab
>> F = E([1:2:end] , [2:end])
```

The keyword `end` designates the last element of the dimension: 5 for the rows of \(E\) and 6 for the columns.

**Note:** The keyword `end` has a number of meanings. It also ends a block of code begun with a `if`, `for`, `while`, or `switch` (as we will see later). Finally, it can also terminate a primary function or a subfunction, and it must terminate a nested function (as we will also see later).

Returning to \(x\), entering \(E(:,:)\) prints out exactly the same matrix as entering \(E\) (as does

\[
E(1:end,1:end) .
\]

This is not a very useful way of entering \(E\), but it shows how the colon operator can work. On the other hand, entering

```matlab
>> G = E(:,1:4)
```

generates a matrix with the same size as \(E\) but with the columns reversed, i.e.,

\[
\begin{bmatrix}
6 & 5 & 4 & 3 & 2 & 1 \\
12 & 11 & 10 & 9 & 8 & 7 \\
18 & 17 & 16 & 15 & 14 & 13 \\
24 & 23 & 22 & 21 & 20 & 19 \\
30 & 29 & 28 & 27 & 26 & 25
\end{bmatrix}
\]

It is also very easy to switch rows in a matrix. For example, to switch the third and fifth rows of \(G\), enter the single line

```matlab
>> G(3, :) = G(5, :)
```

which is much simpler, and more transparent, than the three lines

```matlab
>> temp = G(3, :)
>> G(3, :) = G(5, :)
>> G(5, :) = temp
```

**Note:** There is a more general function which can reverse two variables: `scalars` or `vector` or `matrices`. For example,

```matlab
>> [y, x] = deal(x, y); % or [Y, X] = deal(X, Y);
```

reverses the values in these two variables. They can even have different sizes! (This is a very specialized function, but it is annoying to need three statements to reverse two variables.)

Finally, there is one more use of a colon. Entering

```matlab
>> F = E(:,1)
```

generates a column vector consisting of the columns of \(E\) (i.e., the first five elements of \(F\) are the first column of \(E\), the next five elements of \(F\) are the second column of \(E\), etc.).

**Note:** On the right side of an equation, \(E\) is a column vector with the elements being the columns of \(E\) in order. On the left side of an equation, \(E(\) reshapes a matrix. However, we will not discuss this reshaping further because the `reshape` function described below is easier to understand.

There is also a practical example for reordering the elements of a vector. The random function `rand` can also be used to randomly generate any set of values (integer or real or even complex) by, for example,

```matlab
>> last_prime = 20;
>> p = primes(last_prime); % the prime numbers up to last_primes
>> ir = randi(length(p), n, 1)
```

All the primes up to 20 are calculated (there are 8) and \(n\) random integers in 1, 2, 3, …, \(\text{length}(p)\) are calculated. Then `ir` determines which primes to put into \(r\). This is called a random selection with replacement since a prime can occur more than once. If you do not want the primes to occur more than once, use `randperm` instead.

The colon operator works on rows and/or columns of a matrix. A different function is needed to work on the diagonals of a matrix. For example, you extract the main diagonal of \(E\) by typing

```matlab
>> d = diag(E)
```

(\(d\) is the column vector \((1, 8, 15, 22, 29)^T\)), one above the main diagonal by typing

```matlab
>> d1 = diag(E, 1)
```

(\(d1\) is the column vector \((2, 9, 16, 23, 30)^T\)), and two below the main diagonal by typing

```matlab
>> d2 = diag(E, -2)
```

(\(d2\) is the column vector \((-13, 20, 27)^T\)).

The MATLAB function `diag` transforms a matrix (i.e., a non-vector) into a column vector. The converse also holds: when `diag` is applied to a vector, it generates a symmetric matrix. The function

```matlab
>> F = diag(d)
```

generates a \(5 \times 5\) matrix whose main diagonal elements are the elements of \(d\), i.e., \(1, 8, 15, 22, 29,\) and whose off-diagonal elements are zero. Similarly, entering

```matlab
>> F1 = diag(d1, 1)
```

generates a \(6 \times 6\) matrix whose first diagonal elements (i.e., one above the main diagonal) are the elements
of \( d_1 \), i.e., 2, 9, 16, 23, 30, and whose other elements are zero, that is,

\[
F_1 = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 \\
0 & 9 & 0 & 0 & 0 \\
0 & 0 & 16 & 0 & 0 \\
0 & 0 & 0 & 23 & 0 \\
0 & 0 & 0 & 0 & 30
\end{pmatrix}.
\]

Finally, typing

\[
F_2 = \text{diag}(42, -2)
\]

generates a 5 x 5 matrix whose \(-2^{nd}\) diagonal elements (i.e., two below the main diagonal) are the elements of \( d_2 \), i.e., 13, 20, 27, and whose other elements are zero, i.e.,

\[
F_2 = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
13 & 0 & 0 & 0 & 0 \\
0 & 20 & 0 & 0 & 0 \\
0 & 0 & 27 & 0 & 0
\end{pmatrix}.
\]

The Toeplitz matrix is a very special matrix whose values are constant along each diagonal. For example,

\[
\begin{pmatrix}
7 & 4 & 3 & 1 \\
-2 & 7 & 4 & 3 \\
-5 & -2 & 7 & 4 \\
-1 & -5 & -2 & 7
\end{pmatrix}
\]

is generated by

\[
\text{toeplitz}(\{7, -2, -5, 1\}, \{7, 4, 3\}).
\]

If the Toeplitz matrix is symmetric, only the row or the column elements need be entered (i.e., only one argument is required).

You can also extract the upper triangular or the lower triangular part of a matrix. For example,

\[
G_1 = \text{triu}(F)
\]

constructs a matrix which is the same size as \( F \) and which contains the same elements as \( F \) on and above the main diagonal; the other elements of \( G_1 \) are zero. This function can also be applied to any of the diagonals of a matrix. For example,

\[
G_2 = \text{triu}(F, 1)
\]

constructs a matrix which is the same size as \( F \) and which contains the same elements as \( F \) on and above the first diagonal, i.e.,

\[
G_2 = \begin{pmatrix}
0 & 2 & 3 & 4 & 5 & 6 \\
0 & 0 & 9 & 10 & 11 & 12 \\
0 & 0 & 0 & 16 & 17 & 18 \\
0 & 0 & 0 & 0 & 23 & 24 \\
0 & 0 & 0 & 0 & 0 & 30
\end{pmatrix}.
\]

The similar function \( \text{tril} \) extracts the lower triangular part of a matrix.

As an example of the relationship between these three functions, consider the square random matrix \( F \) generated by

\[
F = \text{rand}(6)
\]

All the following MATLAB statements calculate \( F \) anew:

\[
\begin{align*}
\text{triu}(F) + \text{tril}(F) - \text{diag}(\text{diag}(F)) \\
\text{triu}(F, 1) + \text{diag}(\text{diag}(F)) + \text{tril}(F, -1) \\
\text{triu}(F) + \text{tril}(F, -1) \\
\text{triu}(F, 2) + \text{diag}(\text{diag}(F), 1) + \text{tril}(F)
\end{align*}
\]

Note: Numerically the first statement might not generate exactly the same matrix as the following three because of round-off errors.

It is important to note that \( \text{diag}, \text{triu} \) and \( \text{tril} \) cannot appear on the left-hand side of an equation. Instead, to zero out all the diagonals above the main diagonal of \( F \) enter

\[
F = F - \text{triu}(F, 1)
\]

do zero out just the first diagonal above the main diagonal enter

\[
F = F - \text{tril}(\text{triu}(F, 1), 1)
\]

What if you want to insert numbers from the upper right-hand corner of a matrix to the lower left-hand corner? There is no explicit function which does this but there are a number of indirect functions:

- \( \text{flipud}(A) \) flips the matrix up and down, i.e., reverses the rows of the matrix;
- \( \text{rot90}(A) \) rotates the matrix \( 90^\circ \); and
- \( \text{rot90}(A, k) \) rotates the matrix \( k \times 90^\circ \).

MATLAB has a function which is useful in changing the shape of a matrix while keeping the same numerical values. The statement

\[
E = \text{reshape}(R, n, n)
\]

reshapes the matrix \( R \in \mathbb{C}^{m \times n} \) into \( K \in \mathbb{C}^{m \times n} \) where \( m \) and \( n \) must satisfy \( mn = pq \) (or an error message will be generated). A column vector is generated from \( R \), as in \( R(:, 1) \), and the elements of \( K \) are taken columnwise from this vector. That is, the first \( m \) elements of this column vector go in the first column of \( K \), the second \( m \) elements go in the second column, etc. For example, the matrix \( K \) which has been used throughout this subsection can be easily (and quickly) generated by

\[
E = \text{reshape}([1:30], 6, 5)^
\]

Occasionally, there is a need to delete elements of a vector or rows or columns of a matrix. This is easily done by using the null matrix \([\]\). For example, entering

\[
\begin{align*}
\text{x} &= \{1, 2, 3, 4\}' \\
\text{y} &= \text{zeros}(2, 2)
\end{align*}
\]

results in \( x = (1, 3, 4)' \). As another example, you can delete the even columns of \( G \) by

\[
G(:, 2:2:6) = []
\]

The result is

\[
G = \begin{pmatrix}
6 & 4 & 2 \\
12 & 16 & 8 \\
18 & 22 & 20 \\
24 & 28 & 26
\end{pmatrix}
\]

Also, occasionally, there is a need to duplicate or tile a matrix to form a larger matrix. The statement

\[
B = \text{repmat}(A, m, n)
\]

generates a matrix \( B \) which contains \( m \) rows and \( n \) columns of copies of \( A \). (If \( m = n \) then \( \text{repmat}(A, m, n) \) is sufficient.) If \( A \) is a \( p \times q \) matrix, then \( B \in \mathbb{R}^{mp \times nq} \). This even works if \( A \) is a scalar, in which case this is the same as

\[
B = \text{ones}(m, n)
\]

(but it is much faster if \( m \) and \( n \) are large since no multiplication is involved).

One frequent use of \( \text{repmat} \) is when a specific operation is to be applied to each row or to each column of a matrix. For example, suppose that the column vectors \( \{x_1, x_2, \ldots, x_m\} \) have been combined into the matrix \( X \) and we want to calculate the corresponding matrix for the vectors \( y_j = x_j + a \) for all \( j \in [1, n] \). This can be easily done by

\[
Y = X + \text{repmat}(a, 1, n); \quad \text{which, unfortunately, requires that the new matrix } A = \text{repmat}(a, 1, n) \text{ be created. We would prefer to simply enter}
\]

\[
Y = X + a \quad \text{(WRONG)};
\]

However, we can enter
which, incidentally, is much faster than using repeat. This function can actually be applied to multidimensional matrices, but we will only describe it for repeat(<function handle>, A, b). <function> is a function handle which operates on the matrix \( A \in \mathbb{R}^{m \times n} \) and the column vector \( b \in \mathbb{R}^n \) or the row vector \( b \in \mathbb{R}^{m \times 1} \). The simplest operation is to let the function handle be one of the following:

- \( \sin \) or one column vector and one scalar
- \( \cos \) it then outputs a column vector of the same size as the input.

 INCIDENTALLY, it is possible to write your own function which inputs either two column vectors of the same size or one column vector and one scalar; it then outputs a column vector of the same size as the input.

\[
\begin{array}{|c|c|}
\hline
\text{Built-in Functions for \texttt{bsxfun}} & \text{Function} \\
\hline
\text{plus} & \text{\texttt{bsxfun}} \\
\text{minus} & \text{\texttt{bsxfun}} \\
\text{times} & \text{\texttt{bsxfun}} \\
\text{divide} & \text{\texttt{bsxfun}} \\
\text{power} & \text{\texttt{bsxfun}} \\
\text{max} & \text{\texttt{bsxfun}} \\
\text{min} & \text{\texttt{bsxfun}} \\
\text{rem} & \text{\texttt{bsxfun}} \\
\text{atan} & \text{\texttt{bsxfun}} \\
\text{atan2} & \text{\texttt{bsxfun}} \\
\text{besselj} & \text{\texttt{bsxfun}} \\
\text{besselk} & \text{\texttt{bsxfun}} \\
\text{bessely} & \text{\texttt{bsxfun}} \\
\text{besselh} & \text{\texttt{bsxfun}} \\
\text{gamma} & \text{\texttt{bsxfun}} \\
\text{log} & \text{\texttt{bsxfun}} \\
\text{sqrt} & \text{\texttt{bsxfun}} \\
\end{array}
\]

Of course, it is possible to write your own function which inputs either two column vectors of the same size or one column vector and one scalar; it then outputs a column vector of the same size as the input.

\[
\begin{align*}
A(i,j) &= a_{ij} \\
A(:,:j) &= \text{the } j\text{th column of } A \\
A(:,i) &= \text{the } i\text{th row of } A \\
A(1,:) &= A \\
A(1:7,1:2) &= \text{There are many more choices than we care to describe:} \\
&= 71 \text{ can be } 1 \text{ or } 11:12 \text{ or } 11:13:12 \text{ or } ... \text{ or } [11 \ 12 \ ... \ 19] \\
&= 72 \text{ can be } 1 \text{ or } 11:12 \text{ or } 11:13:12 \text{ or } ... \text{ or } [11 \ 12 \ ... \ 20] \\
A(:) &= \text{On the right-hand side of an equation, this is a column vector containing} \\
&= \text{the columns of } A \text{ one after the other.} \\
\text{diag}(A) &= \text{A column vector of the } k\text{th diagonal of the matrix (i.e., non-vector) } A. \text{ If } k \\
&= \text{is not given, then } k = 0. \\
\text{diag}(A,k) &= \text{A square matrix with the } k\text{th diagonal being the vector d. If } k \text{ is not} \\
&= \text{given, then } k = 0. \\
\text{triu}(A) &= \text{A matrix which is the same size as } A \text{ and consists of the elements on and} \\
&= \text{above the } k\text{th diagonal of } A. \text{ If } k \text{ is not given, then } k = 0. \\
\text{tril}(A) &= \text{The same as the function triu except it uses the elements on and below} \\
&= \text{the } k\text{th diagonal of } A. \text{ If } k \text{ is not given, then } k = 0. \\
\text{flipud}(A) &= \text{Flips a matrix left to right.} \\
\text{fliplr}(A) &= \text{Flips a matrix up and down.} \\
\text{rot90}(A) &= \text{Rotates a matrix } k\times 90^\circ. \text{ If } k \text{ is not given, then } k = 1. \\
\text{repmat}(A,n,m) &= \text{Generates a matrix with } m \text{ rows and } n \text{ columns of copies of } A. \text{ (If } n = m \text{ the} \\
&= \text{third argument is not needed.)} \\
\text{bsxfun}(\text{func}, A, b) &= \text{Perform the operation given by the function handle on all the columns} \\
&= \text{of the matrix } A \text{ using the column vector } b \text{ or on all the rows using the row} \\
&= \text{vector } b. \\
\text{reshape}(A,m,n) &= \text{Generates an } m \times n \text{ matrix whose elements are taken columnwise from } A. \text{ Note:} \\
&= \text{The number of elements in } A \text{ must be } mn. \\
[] &= \text{The null matrix. This is also useful for deleting elements of a vector} \\
&= \text{or rows or columns of a matrix.} \\
\text{toeplitz}(c,r) &= \text{Generates a Toeplitz matrix where the elements along each diagonal} \\
&= \text{are constant. } c \text{ and } r \text{ are the values on the first diagonal and the first row} \\
&= \text{respectively.}
\end{align*}
\]

2.4. Simple Arithmetical Operations

Matrix Addition:
If \( A, B \in \mathbb{C}^{m \times n} \) then the MATLAB operation
\[
\gg A = B
\]
means \( A + B = (a_{ij} + (b_{ij}) = (a_{ij} + b_{ij}) \). That is, the \((i,j)\)th element of \( A + B \) is \( a_{ij} + b_{ij} \).

Matrix Subtraction:
If \( A, B \in \mathbb{C}^{m \times n} \) then the MATLAB operation
\[
\gg A = B
\]
means \( A - B = (a_{ij} - (b_{ij}) = (a_{ij} - b_{ij}) \).

Matrix Multiplication by a scalar:
If \( A \in \mathbb{C}^{m \times n} \) then for any scalar \( c \) the MATLAB operation
\[
\gg c \times A
\]
means \( cA = (ca_{ij}) = (ca_{ij}) \). For example, the matrix \( q = (0, 1\pi, 2\pi, 3\pi, 4\pi, 5\pi)^T \) can be generated by
\[
\gg q = [0\ :\ .1\* \pi: .5* \pi]'
\]
but more easily by
\[
\gg q = [0\ :\ .1\ :\ .5]* \pi
\]
or
\begin{itemize}
\item q = [0;5]' * 1* p1 \\
\end{itemize}

Matrix Multiplication:
If A ∈ C^{m×n} and B ∈ C^{n×p} then the MATLAB operation
\[ A * B \]
meets AB = (akj) = \left( \sum_{i=1}^{n} a_{ij} b_{ij} \right). That is, the (i,j)th element of AB is a_{ij} b_{ij} + a_{ik} b_{kj} + \cdots + a_{ip} b_{pj}.

Matrix Exponentiation:
If A ∈ C^{n×n} and p is a positive integer, then the MATLAB operation
\[ A^p \]

Matrix Exponentiation is also defined when p is not an integer. For example,
\[ A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, B = A^{-1/2} \]
calculates a complex matrix B whose square is A. (Analytically, B^2 = A, but numerically.
\[ B^2 = A \]
returns a non-zero matrix — however, all of its elements are less than 10^eps in magnitude.)

Note: For two values of p, there are equivalent MATLAB expressions:
\[ A^{1/2} \text{ can also be calculated by } \text{sqrtm}(A) \text{ and } A^{1/2} \text{ can also be calculated by } \text{inv}(A). \]

Matrix Division:

The expression
\[ A \div B \]
makes no sense in linear algebra: if B is a square non-singular matrix it might mean B^{-1}A or it might mean AB^{-1}. Instead, use the expression
\[ A \div B \]
which is defined to be the solution of the linear system Ax = b (where A must be a square non-singular matrix) by Gaussian elimination. This is much faster computationally than calculating the solution of x = A^{-1}b by
\[ x = \text{inv}(A) \times b \]
Similarly,
\[ A \times B \]
solves Ax = B by repeatedly solving Ax = b where b is each column of B in turn and is the corresponding column of X.

It is also possible to solve x^TA = b^T for x^T by
\[ b^T = (A^\dagger)^T \text{ (BUT DON'T)} \]
and to solve Ax = B by
\[ B = A \times x \text{ (BUT DON'T)} \]
\[ X = \text{inv}(A) \times b \]

Elementwise Multiplication:
If A, B ∈ C^{m×n}, then the MATLAB operation
\[ A \odot B \]
means (a_{ij}b_{ij}). That is, the (i,j)th element of A \odot B is a_{ij}b_{ij}. Note that this is not a matrix operation, but it is sometimes a useful operation. For example, suppose \( y \in \mathbb{R}^n \) has been defined previously and you want to generate the vector \( z = (y_1, y_2, y_3, \ldots, y_n)^T \). You merely type
\[ z = [1:n]' * y \]
or
\[ z = y' * [1:n] \]
(where the spaces are for readability). Recall that if y \in \mathbb{C}^n you will have to enter
\[ z = y \cdot [1:n] \]
because you do not want to take the complex conjugate of the complex elements of y.

Elementwise Division:
If A, B ∈ C^{m×n}, then the MATLAB operation
\[ A \oslash B \]
means (a_{ij}/b_{ij}).

Elementwise Left Division:
If A, B ∈ C^{m×n}, then the MATLAB operation
\[ A \oslash B \]
means the same as A \oslash B.

Elementwise Exponentiation:
If A ∈ C^{n×n}, then
\[ A \oslash p \]
means (a_{ij}^p) and
\[ A \oslash (-p) \]
means (a_{ij}^{-p}). Also, if A, B ∈ C^{n×n}, then
\[ A \oslash B \]
meets (a_{ij}/b_{ij}).

Where needed in these arithmetic operations, MATLAB checks that the matrices have the correct size. For example,
\[ A + B \]
will return an error message if A and B have different sizes, and
\[ A \oslash B \]
will return an error message if the number of columns of A is not the same as the number of rows of B.

Note: There is one exception to this rule. When a scalar is added to a matrix, as in A + c, the scalar is
promoted to the matrix cJ where J has the same size as A and all its elements are 1. That is,
\[ A + c \]
evaluated as
\[ A + c \text{ones(size(A))} \]
This is not a legitimate expression in linear algebra, but it is a very useful expression in MATLAB.
For example, you can represent the function
\[ y = 2 \sin(3x + 4) - 5 \text{ for } x \in [2, 3] \]
by 101 data points using
\[ x = [2:0.01:3]' ; \]
\[ y = 2 \sin(3x + 4) - 5 \]
This is much more intelligible than calculating y using
\[ y = 2 \sin(3x + 4 \cos(101, 1)) - 5 \cos(101, 1) \]

In some courses that use vectors, such as physics classes, the dot product of the real vectors \( \vec{a} \) and \( \vec{b} \) is defined by
\[ \vec{a} \cdot \vec{b} = \sum_{i=1}^{n} a_i b_i \]
In linear algebra this is called the \textit{inner product} and is defined for vectors \( \vec{a}, \vec{b} \in \mathbb{R}^n \) by \( \vec{a} \cdot \vec{b} \). It is calculated by
\[ \vec{a} \cdot \vec{b} \]
If (a, b) ∈ \mathbb{C}^n the inner product is a^b\overline{b} and is calculated by \( a^b \). The \textit{outer product} of these two vectors is defined to be ab^T and is calculated by
\[ \vec{a} \otimes \vec{b} \]
If (a, b) are complex the outer product is ab^H and is calculated by \( a^b \). It is important to keep these two products separate: the \textit{inner product} is a scalar, i.e., \( a^b \in \mathbb{R} \) (if complex, \( a^b \in \mathbb{C} \)), while the \textit{outer product} is an n×n matrix, i.e., \( ab^T \in \mathbb{R}^{n×n} \) (if complex, \( ab^H \in \mathbb{C}^{n×n} \)).

In linear algebra we often work with "large" matrices and are interested in the amount of "work" required to perform some operation. In the distant past, MATLAB kept track of the number of flops, i.e., the number of floating-point operations, performed during the MATLAB session. Unfortunately, this disappeared in version 6. Instead, we can calculate the amount of CPU time! required to execute a statement by using \texttt{cputime}. This function returns the CPU time in seconds that have been used since you began your MATLAB session. This time is frequently difficult to calculate, and is seldom more accurate
than to \( \frac{1}{100} \) of a second. Here is a simple example to determine the CPU time required to invert a matrix.

\[
\text{m = input('n = ')}; \text{time = cputime; inv(rand(n)); cputime - time}
\]

**Warning:** Remember that you have to subtract the CPU time used before the operation from the CPU time used after the operation.

You can also calculate the wall clock time required for some sequence of statements by using tic and toc. For example,

\[
\text{tic; \{sequence of statements\}; toc}
\]

returns the time in seconds for this sequence of statements to be performed.

**Note:** This is very different from using cputime. tic followed by toc is exactly the same as if you had used a stopwatch to determine the time. Since a timesharing computer can be running many different processes at the same time, the elapsed time might be much greater than the CPU time. On the other hand, on a multiprocessor computer, the elapsed time might be much less.

### Operator Precedence

The unary plus and minus are the plus and minus signs in

\[
+ \text{ and } -
\]

because "\+" has higher precedence than "./". Also, \( \backslash \text{ and } / \) because ".\*" and ".\^" have the same precedence and so the operations are evaluated from left to right.

### 2.6. Be Careful!

Be very careful occasionally you might misinterpret how MATLAB displays the elements of a vector or matrix.

For example, the MATLAB function \( \text{eig(hilb(5))} \) calculates the eigenvalues of a square matrix. (We discuss eigenvalues in Section 7.) To calculate the eigenvalues of the Hilbert matrix of order 5, i.e.,

\[
\begin{bmatrix}
1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \\
\frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} \\
\frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} \\
\frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} & \frac{1}{8} \\
\frac{1}{5} & \frac{1}{6} & \frac{1}{7} & \frac{1}{8} & \frac{1}{9}
\end{bmatrix}
\]

(we discuss this matrix in detail in Section 5.2) enter

\[
\text{>> format short} \\
\text{>> eig(hilb(5))}
\]

MATLAB displays the eigenvalues as the column vector

\[
\text{ans} =
\begin{bmatrix}
0.0000 \\
-0.0003 \\
0.0114 \\
0.2085 \\
1.5671
\end{bmatrix}
\]

You might think the the first element of this vector is 0. However, if it was zero MATLAB would display 0 and not 0.0000. Entering

\[
\text{>> format short e} \\
\text{>> ans}
\]

\[
\begin{bmatrix}
0.00000 \\
-0.00003 \\
0.000114 \\
0.0002085 \\
0.0015671
\end{bmatrix}
\]

---

1On the other hand, in the statistical computer languages R and S (which are somewhat similar to MATLAB), ".^\times" has higher precedence than ".^\star" and so \(1:n\times1\) is \((1:n)^\times1\) = \(2:(n+1)\).
displays
ans =
  3.2879e-06
  3.0590e-04
  1.1407e-02
  2.0853e-01
  1.5671e+00
which makes it clear that the smallest eigenvalue is far from zero.

On the other hand, if you enter

```matlab
>> format short
disp(k(1:3; 4 5 6; 7 8 9])
ans =
     16.1168
     -1.1168
     -0.0000
```

It might appear from our previous discussion that the last eigenvalue is not zero, but is simply too small to appear in this format. However, entering

```matlab
>> eig(k)
ans =
     16.1168
     -1.1168
     -0.0000
```

The last eigenvalue is not zero, but is simply too small to appear in the format. However, entering

```matlab
>> format short e
ans =
     1.6117e+01
     -1.1168e+00
     -0.0000e+00
```

Since the last eigenvalue is close to `eps`, but all the numbers in the matrix `A` are of “reasonable size”, you can safely assume that this eigenvalue is zero analytically. It only appears to be nonzero when calculated by MATLAB because the computer cannot add, subtract, multiply, or divide correctly!

As another example of how you might misinterpret the display of a matrix, consider the Hilbert matrix of order two:

```matlab
H = [1 1/2; 1/2 1/3]
```

We write `H^n` as

```
\begin{pmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{pmatrix}^n
```

while in MATLAB entering

```matlab
>> format short
disp(H^n)
```

It is very easy to miss the term “1.0e+10 *”. It becomes small — in fact, it should be zero.

```
ans =
  1.0e+10 *
    1.5437 0.8262
    0.8262 0.4421
```

It is very easy to miss the term “1.0e+10 *”, because it stands apart from the elements of the matrix.

**Note:** Use “format short g” so you will not have this problem.

Similarly, entering

```matlab
>> format short e
```

should result in the zero matrix, since `(H^n)^2 = H`. However, MATLAB displays

```
ans =
     1.0e-15 *
    0.2220 0
    0 0
```

where, again, it is easy to miss the term “1.0e-15 *” and not realize that this matrix is very small — in fact, it should be zero.

**Be careful:** MATLAB has finite memory. You should have no problem creating a matrix by

```matlab
>> A = zeros(1000)
```

but you might well have a problem if you enter

```matlab
>> A = zeros(10000)
```

The amount of memory available is dependent on the computer and the operating system and is very hard to determine. Frequently it is much larger than the amount of physical memory on your computer. But even if you have sufficient memory, MATLAB may slow to a crawl and become unusable. The `who` command will tell you how much memory you are using and show you the size of all your variables. If you have large matrices which are no longer needed, you can reduce their sizes by equating them to the null matrix, i.e., `[]`, or remove them entirely by using `clear`.

**Warning:** Recall that the `clear` command is very dangerous because `clear A` deletes the variable `A` but `clear` (without anything following) deletes all variables!

### 2.7. Common Mathematical Functions

In linear algebra mathematical functions cannot usually be applied to matrices. For example, `A` and `sin A` have no meaning unless `A` is a square matrix. (We will discuss their mathematical definitions in Section 1.5.)

Here we are interested in how MATLAB applies common mathematical functions to matrices and vectors. For example, you might want to take the sine of every element of the matrix `A = (a_i) \in \mathbb{C}^{m \times n}`, i.e., `B = \{ \sin a_i \}`. This is easily done in MATLAB by

```matlab
>> B = sin(A)
```

Similarly, if you want `C = \{ e^{a_i} \}`, enter

```matlab
>> C = exp(A)
```

Also, if you want `D = \{ \sqrt{a_i} \}` type

```matlab
>> C = sqrt(A)
```

or

```matlab
>> C = A.^(1/2)
```

All the common mathematical functions in the table entitled “Some Common Real Mathematical Functions” in Section 1.5 can be used in this way.

As we will see in the section on graphics, this new interpretation of mathematical functions makes it easy in MATLAB to graph functions without having to use the MATLAB programming language.

### 2.8. Data Manipulation Functions

MATLAB has a number of “simple” functions which are used quite frequently. Since many of them are quite useful in analyzing data, we have grouped them around this common theme.

To calculate the maximum value of the vector `x`, type

```matlab
>> x = max(x)
```

If you also want to know the element of the vector which contains this maximum value, type

```matlab
>> [n, i] = max(x)
```

If the elements of the vector are all real, the result of this statement is the element which has the maximum value. However, if any of the elements of `x` are complex (i.e., non-real), this statement has no mathematical meaning. MATLAB defines it to determine the element of the vector which has the maximum
absolute value of the elements of \( x \).

Warning: Make sure you understand the description of \( \max \) if you every apply it to non-real vectors. For example, if \( x = (-2, i)^T \) then \( \max(x) \) returns \(-2\). This is because the element which has the largest absolute value is \(-2\).

Thus, if \( x \) is a non-real vector, then \( \max(x) \) is \( \text{real} \) the same as \( \max(\abs(x)) \).

Since the columns of the matrix \( A \) can be considered to be vectors in their own right, this command can also be applied to matrices. Thus,

\[
\text{max}(A)
\]

returns a row vector of the maximum element in each of the columns of \( A \) if all the elements of \( A \) are real. If any of the elements of \( A \) are non-real, this function returns the element in each column which has the maximum \( \text{absolute value} \) of all the elements in that column. And to calculate a column vector of the maximum element in each of the rows of \( A \) use

\[
\text{max}(A)'\]

which switches the rows and columns of \( A \) so it calculates the maximum element in each row as a row vector, and then converts the result to a column vector.

To find the maximum value of an entire real matrix, type

\[
\text{max}(\text{max}(A))
\]

or

\[
\text{max}(A(:))
\]

and to find the maximum \( \text{absolute value} \) of an entire real or complex matrix, type

\[
\text{max}(\text{abs}(\text{abs}(A)))
\]

or

\[
\text{max}(\text{abs}(A(:)))
\]

There is also another use for \( \max \). If \( A \) and \( B \) are matrices which either have the same size, or one or both is a scalar, then

\[
\text{max}(A, B)
\]

returns a matrix which is the same size as \( A \) and \( B \) (or the size of the larger if one is a scalar) and which contains the larger of the values in each element. For example,

\[
A = \text{max}(A, 0)
\]

replaces all negative elements of \( A \) with zeroes.

Note: If \( \max \) has one argument, then it determines the maximum value of a vector or the maximum value in each column of a matrix. If it has two arguments, it determines the maximum value of each element of the two matrices.

Not surprisingly, the function \( \text{sum} \) acts similarly to \( \max \) except that it finds the minimum value (or element with the minimum \( \text{absolute value} \)) of the elements of a vector or the columns of a matrix.

To calculate the sum of the elements of the vector \( x \), type

\[
\text{sum}(x)
\]

\( \text{sum} \) behaves similarly to \( \max \) when applied to a matrix. That is, it returns the row vector of the sums of each column of the matrix. This sum is sometimes useful in adding a deterministic series. For example,

\[
1 + 1/2 + 1/3 + 1/4 + 1/5 + 1/6 + 1/7 + 1/8 + 1/9 + 1/10 + 1/11 + 1/12 + \ldots
\]

is entered much more easily as

\[
\text{sum(ones(1, 20))/[1:20])}
\]

or even as

\[
\text{sum(1./[1:20])}
\]

The mean, or average, of these elements is calculated by

\[
\text{mean}(x)
\]

where \( \text{mean}(x) = \text{sum}(x)/\text{length}(x) \).

\( \text{std} \) calculates the standard deviation of the elements of a vector. The standard deviation is a measure of how much a set of numbers “vary”, and it is defined as

\[
\text{std}(x) = \left( \frac{1}{n} \sum_{i=1}^{n} (x_i - \langle x \rangle)^2 \right)^{1/2}
\]

where (\( \langle x \rangle \) is the mean of the elements.

MATLAB can also sort the elements of the vector \( x \) in increasing order by

\[
\text{sort}(x)
\]

If the vector is non-real, the elements are sorted in increasing absolute value. (If two elements have the same absolute value, the one with the smaller absolute angle in polar coordinates is used.)

The MATLAB function \( \text{diff} \) calculates the difference between successive elements of a vector. For example, if \( x \in \mathbb{R} \) then the function

\[
x = \text{diff}(x)
\]

generates the vector \( s \in \mathbb{R}^{n-1} \) which is defined by \( s_i = x_{i+1} - x_i \). There are a number of uses for this function.

- If \( x \) has been sorted, then if any element of \( \text{diff}(x) \) is 0, an element of \( s \) has been repeated — and we can even determine the number of times it has been repeated.
- Similarly, if all the elements of \( \text{diff}(x) \) are positive, then all the elements of \( s \) are monotonically increasing.
- A numerical approximation to the derivative of \( y = f(x) \) can be calculated by \( \text{diff}(y)/\text{diff}(x) \).

(The functions \( \text{any} \) and \( \text{all} \), which are discussed in Section 8.2, are used to determine if any the elements of a vector satisfy some property and/or if all the elements satisfy it.)

The MATLAB function which is almost the inverse of \( \text{diff} \) is \( \text{cumsum} \). It calculates the cumulative sum of the elements of a vector or matrix. For example, if \( s \in \mathbb{R}^{n-1} \) has been generated by \( s = \text{diff}(x) \), then

\[
x = \text{cumsum}(s)
\]

generates the vector \( c \in \mathbb{R}^n \) where \( c_i = \sum_{j=1}^{i} s_j \). We can recover \( x \) by

\[
x = \text{cumsum}(s) - \text{cumsum}(s(1:n-1))
\]

and \( (1, 1, 2, 3, \ldots, n) \) by

\[
\text{cumprod}(1:n)
\]

All of these functions can be applied to matrices, in which case they act on each column of the matrix separately. They can also act on each row of a matrix separately by taking the transpose of the matrix.

However, there is always an optional argument — often the second, but sometimes the third — which can change the function so that it acts on each row of the matrix. The main difficulty with using another argument is remembering if it is the second or the third!

There are also a number of MATLAB functions which are particularly designed to plot data. The functions we have just discussed, such as the average and standard deviation, give a coarse measure of the distribution of the data. To actually “see” what the data looks like, it has to be plotted. Two particularly useful types of plots are histograms (which show the distribution of the data) and plots of data which include error bars. These are both discussed in Section 4.1.

Although it does not quite fit here, sometimes you want to know the length of a vector \( x \), which is \( N \). (Note that this is \( \text{norm}(x) \) which returns the number of elements in \( x \), i.e., \( n \). ) This length, which is often called the Euclidean length, can be calculated by entering

\[
\text{sqrt}(x'x)
\]

but it can be entered more easily by

\[
\text{norm}(x)
\]

(As we discuss in Section 7, the norm of a vector is a more general concept than simply the Euclidean length.)

Warning: The number of elements in the vector \( x \) is calculated by \( \text{length}(x) \) while the (Pythagorean) length of the vector is calculated by \( \text{norm}(x) \).
2.9. Advanced Topic: Multidimensional Arrays

We have already discussed 1-D arrays (i.e., vectors) and 2-D arrays (i.e., matrices). Since these are two of the most fundamental objects in linear algebra, there are many operations and functions which can be applied to them. In MATLAB you can also use multidimensional arrays (i.e., n-D arrays).

A common use for multidimensional arrays is simply to hold data. For example, suppose a company produces three products and we know the amount of each product produced each quarter; the data naturally fits in a 2-D array, i.e., (product, region, amount). Now suppose the company has five sales regions so we split the amount of each product into these regions; the data naturally fits in a 3-D array, i.e., (product, region, amount).

For another example, a 3-D array might be the time evolution of 2-D data. Suppose we record a grey scale digital image of an experiment every minute for an hour. Each image is stored as a matrix $M$ with $m_{i,j}$ denoting the value of the pixel positioned at $(x_i, y_j)$. The 3-D array $M_{11}$ can contain all these images: $M_{11}(i,j,k)$ denotes the value of the pixel positioned at $(x_{i,j})$ in the $k$th image. The entire $k$th image is $M_{11}(i,j,:)$ and it is filled with the $k$th image $M_{11}$.

If you want to multiply $M_{11}$ by another matrix $A$, you can use $M_{11} \times A$ or $M_{11}(i,j,: \times A)$; if you want to average the first two images you can use $0.5 \times (M_{11}(i,j,: \times 2) = A)$.

Many MATLAB functions can be used in n-D such as ones, rand, sum, and size. The cat function is particularly useful in generating higher-dimensional arrays. For example, suppose we have four matrices $A, B, C$, and $D \in \mathbb{R}^{11 \times 11}$ which we want to put into a three-dimensional array. This is easily done by

\[
\text{ABC}D = \text{cat}(3, A, B, C, D)
\]

which concatenates the four matrices using the third dimension of $\text{ABC}D$. Thus the "3" denotes the third dimension of $\text{ABC}D$. And it is much easier than entering

\[
\text{ABC}D(:,1) = A;
\]
\[
\text{ABC}D(:,2) = B;
\]
\[
\text{ABC}D(:,3) = C;
\]
\[
\text{ABC}D(:,4) = D;
\]

If instead, we enter

\[
\text{ABC}D = \text{cat}(j, A, B, C, D)
\]

then the four matrices are concatenated along the $j$th dimension of $\text{ABC}D$. That is, $\text{cat}(1, A, B, C, D)$ is the same as $[A, B, C, D]$ and $\text{cat}(2, A, B, C, D)$ is the same as $[A; B; C; D]$.

Another useful function is squeeze which squeezes out dimensions which only have one element. For example, if we enter

\[
E = \text{squeeze}(A)
\]

then the four matrices are concatenated along the $j$th dimension of $\text{ABC}D$. That is, $\text{cat}(1, A, B, C, D)$ is the same as $[A, B, C, D]$ and $\text{cat}(2, A, B, C, D)$ is the same as $[A; B; C; D]$.

Note: the norm of a matrix is not the Euclidean length of each column in the matrix.

2.10. Be Able To Do

After reading this section you should be able to do the following exercises. The solutions are given on page 163.

1. Consider the matrix

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

(a) Enter it in the following three ways:

(i) type in all 16 elements directly.

(ii) since each row is in arithmetic progression, use the colon operator to enter each row.

(iii) since each column is in arithmetic progression, use the colon operator (and the transpose operator) to enter each column.

(b) Multiply the second row of $A$ by $-\frac{1}{2}$, add it to the third row, and put the result back in the second
row. Do this all using \texttt{cpp MATLAB} statement.

2. Generate the tridiagonal matrix

$$A = \begin{pmatrix}
1 & -1 & 0 \\
-1 & 4 & -1 \\
0 & -1 & 4
\end{pmatrix} \in \mathbb{R}^{3 \times 3},
$$

where the value of \( n \) has already been entered into MATLAB.

3. Generate the tridiagonal matrix

$$A = \begin{pmatrix}
1 & -1 & 0 & \ldots & 0 \\
-1 & 4 & -1 & \ldots & 0 \\
0 & -1 & 4 & \ldots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & \ldots & 1 & -1
\end{pmatrix} \in \mathbb{R}^{n \times n},
$$

where the value of \( n \) has already been entered into MATLAB.

4. Consider the matrix

$$A = \begin{pmatrix}
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & -5 \\
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
5 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}.
$$

(a) Enter it using as few keystrokes as possible. (In other words, don’t enter the elements individually.)
(b) Zero out all the elements of \( A \) below the diagonal.

5. Enter the column vector

$$x = (0, 1, 4, 9, 16, 25, \ldots, 841, 900)^T$$

using as few keystrokes as possible. (In other words, don’t enter the elements individually.)

6. (a) Generate a random 5×5 matrix \( R \).
(b) Determine the largest value in each row of \( R \) and the element in which this value occurs.
(c) Determine the average value of all the elements of \( R \).
(d) Generate the matrix \( R \) where every element of \( A \) is the sine of the corresponding element of \( R \).
(e) Put the diagonal elements of \( R \) into the vector \( x \).

7. Generate the matrix

$$A = \begin{pmatrix}
1 & 2 & 3 \\
4 & 5 & 6 \\
7 & 8 & 10
\end{pmatrix}.
$$

(a) Calculate a matrix \( B \) which is the square root of \( A \). That is, \( B^2 = A \). Also, calculate a matrix \( C \) each of whose elements is the square root of the corresponding element of \( A \).
(b) Show that the matrices you have obtained in (a) are correct by substituting the results back into the original formulas.

3. Anonymous Functions, Strings, and Other Data Types

Now that we have discussed scalar and matrix calculations, the next important topic is graphics. However, there are a number of minor topics which are useful in graphics and so we collect them here. First, anonymous functions allow us to easily define a function which we can then plot. Second, some graphics functions require that the name of a function be passed as an argument. Third, character strings are necessary in labelling plots. And, finally, cell arrays are occasionally helpful in labelling plots. Cell arrays are generally used to manage data and since structures are also used to manage data we also include them here. Another reason is that there are a number of data types in MATLAB including floating-point variables, integers, text variables, cell arrays, structures, and logical variables. We might as well get all these out of the way at once.

3.1. Anonymous Functions

In MATLAB it is common to define a mathematical function in a separate file as we discuss in Section 8.3. (This is similar to writing a function or subroutine in a high-level computer language.) However, if the mathematical function is particularly simple, that is, it can be written as one simple expression, we can define it in MATLAB using an anonymous function. If our function is

$$f(< \text{arg}_1 >, < \text{arg}_2 >, \ldots) = < \text{expression} >$$

the MATLAB statement is

$$f = @(<\text{arg}_1>, <\text{arg}_2>, \ldots) \text{"expression"}$$

For example, we can define the function

$$f(t) = t^5 e^{-2t} \cos(3t)$$

by

$$f = @(t) t^5 \cdot e^{-2t} \cdot \cos(3t)$$

and then evaluate it by

$$x = [0:0.01:1]$$
$$fx = f(x)$$
$$A = \text{rand}(5)$$
$$fA = f(A)$$

More generally, we can define

$$g(x, y, a, b, c) = x^2 e^{-y} \cos(cy)$$

by

$$g = @(x, y, a, b, c) x^2 \cdot a \cdot \exp(-b \cdot x) \cdot \cos(c \cdot y)$$

and then evaluate it by

$$x = (x, y, a, b, c)^T$$
$$g = @(x) x(1) \cdot x(2) \cdot \exp(-x(3) \cdot x(4)) \cdot \cos(x(5) \cdot x(2))$$

(In this example there is no advantage to using \( \text{\texttt{\textbackslash}} \text{\texttt{arg}} \text{\texttt{T}} \) — there is no left-hand side to this expression — we use \( \text{\texttt{\textbackslash}} \text{\texttt{arg}} \text{\texttt{T}} \) to let \( g \) have one vector argument, say \( x \), \( x \), \( a \), \( b \), \( c \)) by

$$g = @(x) x(1) \cdot x(2) \cdot \exp(-x(3) \cdot x(4)) \cdot \cos(x(5) \cdot x(2))$$

Here is an example of an anonymous function.

$$f(t) = t^5 e^{-2t} \cos(3t)$$

by

$$f = @(t) t^5 \cdot e^{-2t} \cdot \cos(3t)$$

then

$$f(3)$$

is allowed, but not

$$f([1:10])$$

Be careful!

The syntax for defining an anonymous function is

$$f = @(\text{\texttt{arglist}}) \text{"expression"}$$

(There is no \texttt{left-hand-side} to this expression, the name of this function is \texttt{f}.) The symbol \( \texttt{f} \) is the MATLAB operator that constructs a function handle. This is similar to a pointer in C which gives the address of a variable or of a function. The name "handle" is used in MATLAB to denote a variable which refers to some "object" which has been created. Thus, we can think of an anonymous function as being created by

$$f = @(\text{\texttt{arglist}}) \text{"expression"}$$

and the handle to the function (in C, the address of the function) being returned by using \( \texttt{f} \). By the way, we can create a function handle to a MATLAB function by, for example,

$$f = \text{\texttt{\textbackslash}} \text{\texttt{cos}}$$

so that \( f(3) \) is the same as \( \text{\texttt{cos}}(3) \). We give an example where this is very useful in Section 3.4.

It is even possible to define a piecewise function in this way. For example, the piecewise function

$$f(x) = \begin{cases} 
1 - |x| & \text{for } x \in [-1, 1] \\
0 & \text{otherwise}
\end{cases}$$
i.e., an isosceles triangle with the length of the base 2 and the height 1, is

\[ t(x) = 2 \text{abs}(x^2 - 1) \]

in MATLAB.

Note: The x in "f(x)" is a dummy variable, i.e., the x is independent of any other x which appears in the code. Or, in other words, the function could have been defined equally well by

\[ t = 2 \text{abs}(x^2 - 1) \]

It is important to understand that all user-defined variables which appear in <expression> must otherwise appear in the argument list or be defined before the function is defined. If the variable does not appear in the argument list, then its value is fixed when the function is defined. For example, if a very simple function is defined by

\[ r = 10 \]

then the function is \( h(x) = 10x \) even if \( r \) is modified later. Thus,

\[ h(5) \]

returns 50 and so does

\[ h(0) \]

Warning: Don’t forget that if a variable does not appear in the argument list, then its value is fixed when the function is defined.

A function can also be defined — but don’t do it — by the `inline` function. For example, the function \( f \) defined above can also be defined by

\[ f = \text{inline}(t.5 \cdot \exp(-2t) \cdot \cos(3t)), \text{’t’} \]

In general, if our function is

\[ f(\text{<arg1>}, \text{<arg2>}, \ldots) = <\text{expression}> \]

the MATLAB statement is

\[ f = \text{inline}(\text{’<expressions>’}, \text{’<arg1>’}, \text{’<arg2>’}, \ldots) \]

Since it is quite easy to forget to put dots (i.e., .) before the mathematical operations of multiplication (i.e., *), division (i.e., /), and exponentiation (i.e., ^), the MATLAB command `vectorize` does it for you. To continue the first example,

\[ f = \text{vectorize}(\text{inline}‘t.5 \cdot \exp(-2t) \cdot \cos(3t)), \text{’t’} \]

is equivalent to the \( f \) defined above but does not require you to remember all the dots.

Warning: The `inline` function is obsolete and should not be used because it is slow and also because it is rather difficult to read, i.e.,

\[ f = \text{inline}(t \cdot .5 \cdot \exp(-2t) \cdot \cos(3t)), \text{’t’} \]

is “more similar” to \( f(t) = t^2e^{-2t}\cos(3t) \). `inline` is mentioned here only because it is often found in “old” codes.

3.2. Passing Functions as Arguments

Warning: This is a very important section — read it carefully and understand it. If you try to pass a function as an argument to another function, something like

\[ \text{expplot}(\sin) \]

you are trying to generate an “easy plot” of the sine function and you get a `MATLAB` error message, something like

Error using `sin`
Not enough input arguments.

you will know what you did wrong.

It is sometimes necessary to pass the name of a function into a MATLAB function or a function m-file created by the user. For example, as we discuss in Section 4.1, we can plot the function \( y = f(x) \) in the interval \([-5, 5]\) by

\[ \text{fplot(’function name’, [-5 5])} \]

But how do we pass this “name”? We put the `name` in quotes because we do not pass the name of the function, but its handle. If \( f \) has been defined by an anonymous function, then we enter

\[ \text{fplot(\{\’name\’\}, \{-5 5\})} \]

because \( f \) is a variable which we have already defined. If `fnc` is a MATLAB function or a user-defined function m-file, then it is not known in the MATLAB workspace so

\[ \text{fplot(fnc, [-5 5])} \]

will not work. Instead, we use

\[ \text{fplot(’fnc’, [-5 5])} \]

Note: There are a number of “older” ways to pass function names. For example,

\[ \text{fplot(’fnc’, [-5 5])} \]

will also work. We can even pass a “simple” function by, for example,

\[ \text{fplot(’(x^2sin(x) + sqrt(1 + cos(x)))/(x^2 + 1)'), [-5 5])} \]

but don’t do it! Instead, use anonymous functions.

3.3. Strings

Character strings are a very minor part of MATLAB, which is mainly designed to perform numerical calculations. However, they perform some very useful tasks which are worth discussing now.

It is often important to combine text and numbers on a plot. Since we discuss graphics in the next section, now is a good time to discuss how characters are stored in MATLAB variables. A string variable, such as

\[ \text{str} = \text{’And now for something completely different’} \]

is a row vector with each character (actually its ASCII representation as shown on page 167) being a single element. MATLAB knows that this is a text variable, not a “regular” row vector, and so converts the numerical value in each element into the corresponding character when it is printed out. For example, to see what is actually contained in the vector `str` enter

\[ \text{str} = 0 \]

or

\[ \text{str} = \text{’And now for something completely different’} \]

Since `str` is a row vector, a substring can be easily extracted. For example,

\[ \text{str(1:7)} \]

returns And now and

\[ \text{str([9:11, 34])} \]

returns ford.

Character variables are handled the same as vectors or matrices. For example, to generate a new text variable which adds “by Monty Python” to `str`, i.e., to concatenate the two strings, enter

\[ \text{str2} = [\text{str} \ ’’ by Monty Python’’} \]

or

\[ \text{str2} = [\text{str}, \ ’’ by Monty Python’’} \]

(you might be easier to read). To convert a scalar variable, or even a vector or a matrix, to a character variable use the function `num2str`. For example, suppose you enter

\[ x = \text{inspace}(0, 2\pi, 100) \]

\[ c1 = 2 \]

\[ c2 = -3 \]

\[ y = c1+\sin(x) + c2\cos(x) \]

and want to put a description of the function into a variable. This can be done by

\[ s = \{ \text{num2str(c1), ’sin(x) + ’, num2str(c2), ’cos(x)’} \} \]

without explicitly having to enter the values of `c1` and `c2`. (An optional second argument to `num2str` determines exactly how the number or numbers are to be printed, but this is usually not needed.)

A text variable can also contain more than one line if it is created as a matrix. For example,

\[ \text{str} = [\’\’And now’’} \]

\[ \’’for’’} \]

\[ \’’something’’} \]

\[ \’’completely’’} \]

\[ \’’different’’} \]

is four lines long. Since `str` is a matrix, each row must have the same number of elements and so we have to pad all but the longest row. (Using cell arrays, we will shortly show how to avoid this requirement.)

Note: We do not usually enter matrices this way, i.e., one column per line. Instead, we simply use “’” to
separate columns. However, here we need to make sure that each row has exactly the same number of characters — or else a fatal error message will be generated.

If desired, you can have more control over how data is stored in strings by using the \texttt{sprintf} function which behaves very similarly to the \texttt{C} commands \texttt{printf}, \texttt{fprintf}, and \texttt{printf}. It is also very similar to the \texttt{fprintf} function in MATLAB which is discussed in detail in Section 6. Note that the data can be displayed directly on the screen by using \texttt{disp}. That is, \texttt{sprintf(\ldots)} generates a character string and \texttt{disp(sprintf(\ldots))} displays it on the screen.

There also is a \texttt{str2num} function to convert a text variable to a number and \texttt{sscanf} to do the same with more control over how the data is read. (This is also very similar to the \texttt{C} command, as discussed in Section 6.) Occasionally, there may be a worry that a string has leading or trailing blanks. These can be removed by

\begin{verbatim}
  str = blanks(n);  \Rightarrow  str_i = num2str( randint(1000) );
  str("n-length(str_i)+1:1") = str_i;
\end{verbatim}

Here some unknown integer, with unknown length, will appear at the end of the string.

One useful string function is \texttt{cellstr}, which can also be used with string matrices if there leading or trailing blanks in all the rows. (It can also be used with cell arrays of strings, which we discuss next.)

Conversely, you can create a string with blanks in it by beginning with all blank characters and then putting non-blanks wherever you desire by

\begin{verbatim}
  n = 10;
  str = blanks(n);
  str_i = num2str( randint(1000) );
  str("n-length(str_i)+1:1") = str_i;
\end{verbatim}

3.4. Cell Arrays and Structures

It is occasionally useful in MATLAB to have a single variable contain all the data which is related to a specific task — and this data might well consist of scalars, vectors and/or matrices, and text variables.

One simple reason for this is that it is easier to pass all the data into and out of functions. A cell array generalizes the “standard” arrays which were discussed in the previous section. The elements of a “standard” array are numbers, either real or complex, whereas the elements of a cell array can be any data type. The primary difference between a cell array and a structure is that in a structure the elements are named rather than numbered. We consider this an advanced topic not because it is complicated but because it is so seldom necessary.

A simple example of a cell array is

\begin{verbatim}
  C = {{[2.0000 + 3.0000i]} 'go cells'; [1 2 3], hilb(5) }
\end{verbatim}

and the output is

\begin{verbatim}
  C =  
  [2.0000 + 3.0000i] 'go cells'
  {3x1 double} [5x5 double]
\end{verbatim}

The only difference between this and a “standard” array is that here curly braces, i.e., \{\ldots\}, enclose the elements of the array rather than brackets, i.e., [\ldots\]. Note that only the scalar and the text variable are shown explicitly. The other elements are only described. A second way to generate the same cell array is

by

\begin{verbatim}
  C(1,1) = {[2+3i]}
  C(1,2) = {'go cells'}
  C(2,1) = {[1 2 3]}
  C(2,2) = {hilb(5)}
\end{verbatim}

and a third way is by

\begin{verbatim}
  C(1,1) = 2+3i
  C(1,2) = 'go cells'
  C(2,1) = [1 2 3]
  C(2,2) = hilb(5)
\end{verbatim}

It is important to understand that there is an important difference between \texttt{C(1,j)} and \texttt{C[1,j]}. The former is the cell containing element in the \texttt{(1,j)}th location whereas the latter is the element itself. For example,

\begin{verbatim}
  C(1,1)=5  \% WRONG
\end{verbatim}

returns an error message because a cell cannot be raised to a power whereas

\begin{verbatim}
  C(1,1)=5  \% CORRECT
\end{verbatim}

returns \texttt{1.2200000000000000e+006}. All the contents of a cell can be displayed by using the \texttt{celldisp} function. In addition, just as a “standard” array can be preallocated by using the \texttt{zeros} function, a cell array can be preallocated by using the \texttt{cell} function. We will not discuss cells further except to state that cell array manipulation is very similar to “standard” array manipulation.

Warning: In MATLAB you can change a variable from a number to a string by simply putting it on the left-hand side of equal signs. For example,

\begin{verbatim}
  c = pi  \Rightarrow  c = '4.141593...
\end{verbatim}

redefines \texttt{c} twice without any difficulty. However, this cannot be done with cells. If you now try

\begin{verbatim}
  c = [3]  \Rightarrow  c = hilb(5)
\end{verbatim}

MATLAB will return with the error message

\begin{verbatim}
  ??? Cell contents assignment to a non-cell array object.
\end{verbatim}

In order to use \texttt{c} as a cell (if it has been previously used as a non-cell), you have to either clear it using \texttt{clr} or explicitly redefine it by using the \texttt{cell} function.

One particularly useful feature of cell arrays is that a number of text variables can be stored in one cell array. We previously defined a “standard” array of strings in \texttt{str} on page 42 where each string had to have the same length. Using a cell array we can simply enter

\begin{verbatim}
  >> Str_cell = {'And'; 'now'; 'for'; something'; 'completely'; 'different'}
  >> Str_cell(1)
\end{verbatim}

or

\begin{verbatim}
  >> Str_cell = {'And'; 'now'; 'for'; something'; 'completely'; 'different'}
  >> Str_cell(1)
\end{verbatim}

and obtain the 1st row by

\begin{verbatim}
  >> Str_cell(1,1)
\end{verbatim}

Note: It is even possible to sort a number of strings in lexicographical ordering by putting each string in a separate row of a cell array and using the \texttt{sort} function.

Structures can store different types of data similarly to cell arrays, but the data is stored by name, called \texttt{fields}, rather than by number. Structures are very similar to structures in C and C++. The cell array we have been using can be written as a structure by

\begin{verbatim}
  >> Cs.scalar = 2+3i
  >> Cs.text = 'go cells'
  >> Cs.vector = [1 2 3]
  >> Cs.matrix = hilb(5)
\end{verbatim}

Typing
Note: is allowed — but probably not very interesting.

functions are discussed in detail in the section. This provides a means to call a function indirectly.

3.5. Advanced Topic: Data Types and Classes

A MATLAB variable can have a large number of different types of values. Those values used to be called data types, but they are now commonly called classes — which is a central concept in object-oriented programming (OOP). The OOP capabilities of MATLAB are similar to those in C++ and Java. We will discuss some of these capabilities at the end of this subsection. However, first we discuss the “fundamental” classes in MATLAB. An important point to remember is that a variable in most programming languages is a single quantity, whereas in MATLAB it is a vector or a matrix or an array of quantities.

We have already described a number of fundamental classes, and we first discuss those which occur frequently in this tutorial.

double: By default any variable that is given a numerical value is a double precision floating-point number. For example, the variable \( x \) which is defined by \( x = 1 \) is an instance of the double class, not an integer class.

char: All strings are instances of the char class. Each character in a string is represented by two bytes because it can represent any Unicode UTF-16 character (although here we are only interested in ASCII characters).

cell: A cell variable itself is an instance of the cell class, but it can contain any number of elements, which can all be instances of different classes.

struct: Similarly, a structure variable is an instance of the struct class, but it can contain any number of fields, which can all be instances of different classes.

function_handle: This provides a means to call a function indirectly.

logical: We will discuss logical variables in Section 8.1, but a simple example is

\[
A = \text{rand}(2)
\]

\[
C = (A > .5)
\]
Each element of $C \in \mathbb{R}^{2 \times 2}$ has the logical value 1, i.e., true, if the corresponding element of $A$ is > 5 and value 0, i.e., false, otherwise. These logical values are each 1 byte.

MATLAB has 15 fundamental classes, each in the form of a matrix (from a 0 × 0 matrix up to an $n$ dimensional matrix for any $n$). For completeness, we now list them all, separating the ones discussed in this document with the rest.

<table>
<thead>
<tr>
<th>Fundamental Classes</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>cell</td>
<td>its elements can be instances of any classes and sizes</td>
</tr>
<tr>
<td>char</td>
<td>string (each character is 2 bytes)</td>
</tr>
<tr>
<td>double</td>
<td>double precision floating-point number (8 bytes)</td>
</tr>
<tr>
<td>function_handle</td>
<td>allows indirect references to functions</td>
</tr>
<tr>
<td>logical</td>
<td>logical (true or false) (1 byte)</td>
</tr>
<tr>
<td>struct</td>
<td>its fields can be instances of any classes and sizes</td>
</tr>
<tr>
<td>single</td>
<td>single precision floating-point number (4 bytes)</td>
</tr>
<tr>
<td>int8</td>
<td>integer in the range $-127$ to $128$ (1 byte)</td>
</tr>
<tr>
<td>uint8</td>
<td>unsigned integer in the range $0$ to $255$ (1 byte)</td>
</tr>
<tr>
<td>int16</td>
<td>integer in the range $-2^{15}$ to $2^{15}$ (2 bytes)</td>
</tr>
<tr>
<td>uint16</td>
<td>unsigned integer in the range $0$ to $2^{16}$ (1 byte)</td>
</tr>
<tr>
<td>int32</td>
<td>integer in the range $-2^{31}$ to $2^{31}$ (4 bytes)</td>
</tr>
<tr>
<td>uint32</td>
<td>unsigned integer in the range $0$ to $2^{32}$ (4 bytes)</td>
</tr>
<tr>
<td>int64</td>
<td>integer in the range $-2^{63}$ to $2^{63}$ (8 bytes)</td>
</tr>
<tr>
<td>uint64</td>
<td>unsigned integer in the range $0$ to $2^{64}$ (8 bytes)</td>
</tr>
</tbody>
</table>

For example, to obtain an instance of the single class you can enter

```matlab
A = single(rand(5))
```

The same technique holds for all the numerical classes.

**Warning: Caveat Emptor!** The procedures for combining different numerical classes is very different from other programming languages. For example, the results of

```matlab
a = 5.5
i = int32(3)        % integer in the range [-2^31 to 2^31]
j = int8(127)      % integer in the range [-127 to 128]
ai = a + i          % double precision floating-point number
aj = a + j          % double precision floating-point number
```

are $ai = 9$, which is an instance of the int32 class, and $aj = 127$, which is an instance of the int8 class.

To determine the class of a variable, use the function `class`. You can also determine if a variable has a particular class by using `isa`. Continuing the previous example,

```matlab
class(1)                 % returns 'int32' and
detects the class of a variable.
isa(1, 'int8')            % returns 0.
```

In addition, MATLAB has user-defined classes, similar to classes in object-oriented programming languages. A simple template for generating a new class is

```matlab
classdef <class name> % name the class
    properties % determine the properties and set their access
    %
    methods % define the methods used in this class
    function ??? %
    %
end
end
```

To define a class which is a subclass of another class, enter

```matlab
classdef <class name> < <superclass name>
end
```

There are two kinds of classes in MATLAB: handle classes and value classes. In a handle class all instances of the class refer to the same data, i.e., any and all copies of an object use the same fields as the original object. In a value class each object has its own unique fields. You create a handle class by `classdef <class name> < handle`

where `handle` is an abstract class. Only a `handle` class can define events and listeners. (An event is some change or action that occurs in an object of a `handle` class. A particular event is attached to an event name, i.e., `inside events` . . . . . . , and when an event name occurs, a notification is broadcast using the event name. Listener objects then execute functions, called callbacks, when an event name is broadcast.)

```matlab
class Determine the class of a variable.
isa Determine whether a variable is an instance of a particular class.
```

### 3.6. Be Able To Do

After reading this section you should be able to do the following exercises. The solutions are given on page 163.

1. Generate a structure with the fields “name”, “rank”, and “serial_number”. Put something appropriate in each in two ways:
   (1) directly, i.e., `s.name = ???`, and
   (2) using the `struct` function.
   Then add one to the serial number.
2. (a) Generate a 2 × 3 cell array with the following elements:
   (1.1): a uniform random matrix of size 5.
   (1.2): the string “Hilbert”.
   (1.3): this is the square of the matrix in (1.1).
   (2.2): the function handle for the function `sin(x)`. 
   (1.3): this is the square of the matrix in (1.1).
   (2.3): the logical value `true`.
   (b) Square the value of the element (1, 2) which is in the cell element (1, 1).
Evaluate the function in the cell element (2.2) at \( x = 5 \).

3. Use anonymous functions to define
   (a) \( f(x, y) = \sin(x) + y \cos(x - y) \),
   (b) \( f(x) = x_1 - e^{x_2} + \cos x_3 \),
   (c) \( f(x) = \frac{1 - \cos x}{x} \) for \( x \in [0, 2\pi] \), otherwise.

   **Hint:** Use the `heaviside` function.

4. Graphics

   A very useful feature of MATLAB is its ability to generate high quality two- and three-dimensional plots using simple and flexible functions. All graphical images are generated in a “graphics window” which is completely separate from the “text window” in which MATLAB statements are typed. Thus, non-graphical and graphical statements can be completely intermixed.

   Graphical images can be generated both from data calculated in MATLAB and from data which has been generated outside of MATLAB. In addition, these images can be output from MATLAB and printed on a wide variety of output devices, including color ink-jet printers and black-and-white and color laser printers.

   There are a number of demonstrations of the graphical capabilities in MATLAB which are invoked by

   ```matlab
   >> demo
   ```

   Since the MATLAB statements which generate the plots are also shown, this demo makes it quite easy to generate your own graphics. You also can have very fine control over the appearance of the plots. We begin by considering only the basic functions; more advanced graphics functions are discussed in the next section.

   **Note:** Most MATLAB functions which take vectors as arguments will accept either row or column vectors.

4.1. Two-Dimensional Graphics

   The MATLAB function `plot` is used to constructing basic two-dimensional plots. For example, suppose you want to plot the functions \( y_1 = \sin x \) and \( y_2 = e^{-x} \) for \( x \in [0, 2\pi] \); also, you want to plot \( y_3 = \sin(\cos(x^2 - x)) \) for \( x \in [0, 8] \). First, generate \( n \) data points on the curve by

   ```matlab
   >> n = 100;
   >> x = linspace(0, n, 2*pi/n); %/n+1;
   >> y1 = sin(x);
   >> y2 = exp(-x);
   >> x = [0:0.01:0.02-eps];
   >> y3 = sin(x).
   >> plot(x, y1);
   >> plot(x, y2);
   >> plot(x, y3)
   ```

   Note that the axes are changed for every plot so that the curve just fits inside the axes. We can generate the \( x \) coordinates of the data points more easily by

   ```matlab
   >> x = linspace(0, 2*pi, n);
   >> xx = linspace(0, 8, n);
   ```

   The `linspace` function has two advantages over the colon operator:
   (1) the endpoints of the axis and the number of points are entered directly as

   ```matlab
   >> x = linspace(0, 2*pi, n);
   >> linspace(0, 2*pi/n-1, 2*pi/n); %/n+1;
   ```

   The `linspace` function has two advantages over the colon operator:
   (1) the endpoints of the axis and the number of points are entered directly as

   ```matlab
   >> x = linspace(first point, last point, number of points)
   ```

   (2) built-in functions are minimalized so you are guaranteed that \( x \) has exactly \( n \) elements, and its first and last elements are exactly the values entered into the function.\(^1\)

   To put all the curves on one plot, type

   ```matlab
   >> plot(x, y1, x, y2, x, y3)
   ```

   Each curve will be a different color — but this will not be visible on a black-and-white output device. Instead, you can change the type of lines by

   ```matlab
   >> plot(x, y1, x, y2, '-+', xx, y3, ':')
   ```

   where `-` means a dashed line and `+` means a dotted line. (We list all these symbols in the following table.) In addition, you can use small asterisks to show the locations of the data points for the \( y_3 \) curve by

   ```matlab
   >> plot(x, y1, x, y2, '-+', xx, y3, 'x')
   ```

   These strings are used to modify the color of the line, to put markers at the nodes, and to modify the type of line as shown in the table below. (As we discuss later in this section, the colors are defined by giving the intensities of the red, green, and blue components in that order.)

   **Note:** The plot function can even have only one argument. For example, entering

   ```matlab
   >> plot(y1)
   ```

   will result in a plot which is equivalent to

   ```matlab
   >> plot([1:length(y1)], y1)
   ```

   ![](table.png)

   For example, using the commands

   ```matlab
   >> x = linspace(0, 2*pi, n);
   >> y = sin(x).
   >> z = cos(x).
   >> plot(x, y, 'r', x, y2, 'g--', x, y3, 'sp')
   ```

   plots three curves: the first is a red, solid line; the second is a green dashed line with circles at the data points; the third has magenta pentagrams at the data points but no line connecting the points. We can also plot the first curve, and then add the second, and then the third by

   ```matlab
   >> hold off
   >> hold on
   >> plot(x, y1)
   >> hold on
   >> plot(x, y2)
   >> hold on
   >> plot(x, y3)
   ```

   Note that the axes can change for every new curve. However, all the curves appear on the same plot.

   **Warning:** Do not place the `hold on` command before the first plot. This can lead to very strange results.

   ```matlab
   >> x = linspace(0, 2*pi, n);
   ```

   to return \( n-1 \) elements rather than \( n \). (For example, the output of `[0 : 0.01 : 0.02-eps]` is `0 0.0100`.)

   This is why we used the statement

   ```matlab
   >> x = linspace(0, 2*pi/n-1, n)
   ```

   above, which does not suffer from round-off errors because the colon operator is only applied to integers.

\(^1\)As we discussed previously, it is very unlikely (but it is possible) that round-off errors might cause the
because certain parameters have already been set. For example, entering
   >> xx = linspace(0, 10, 1001);
   >> hold on
   >> semilogy(xx, exp(xx))
results in a linear plot, i.e., plot(xx, exp(xx)), not the semilog plot which is desired.
We briefly digress to present a technical — but very important — detail. If a plot is hidden (possibly
because you have been typing in the workspace and have raised the MATLAB window above the graphics
window, or because there are some other windows visible on your terminal), you may not be able to see it.
The command
   >> shg
raises the current graphics window above all other windows. This is a very useful command because plots
are frequently hidden. It is particularly beneficial to put this before the first plot function in a script.
M-file: it cannot do any harm, and it will make the plot visible if it is already hidden.\footnote{If there is no graphics window then this command is equivalent to figure, while if there is a graphics window,
then this command is equivalent to figure(gcf). We discuss figure and gcf in advanced subsections.}
We also briefly digress to suggest another possible use for the plot function: making an animation. In
Section 4.6 we will discuss how to make a real movie. However, it is easy in MATLAB to simulate the
time evolution of some function by repeatedly using plot. We have just discussed how to put multiple
curves on a plot, but this rapidly becomes unwieldly. Instead, just replace one plot by another. For exam-
ple, time evolution of some function by repeatedly using
   >> t = 0:del_time:max_time
   >> for t = 0:del_time:max_time
       >> drawnow
   >> end
   >> movie
the command
   >> movie
raises the current graphics window above all other windows. This is a very useful command because plots
are frequently hidden. It is particularly beneficial to put this before the first plot function in a script.
M-file: it cannot do any harm, and it will make the plot visible if it is already hidden.\footnote{Alternatively, you can use the pause function (discussed in Section 8.3) to make all the plots visible.
drawnow draws each new plot as fast as possible, whereas pause(t) pauses execution of the code for the given
time.}
while if there is a graphics window
   >> for t = 0:del_time:max_time
       >> drawnow
   >> end
   >> movie
is the same in all the plots. Otherwise your brain has to try to do the rescaling “on the fly” — which
is very difficult. Of course, you frequently do not know how large the axes need to be until you
have filled up the entire graphics window. The axis function (discussed below) can then be used to
rescale all the plots.
In addition, you can determine the endpoints of the current plot by
   >> v = axis
v(1) and v(2) are the minimum and maximum values on the x axis, and v(3) and v(4) are the corres-
ponding values on the y axis. You can change the endpoints of the axes by, for example,
   >> axis([-1 10 -4 4])
The general form of this function is axis([xmin xmax ymin ymax]). If you only want to set one of the
axes, use
   >> xlim([<minimum x>, <maximum x>])
for the x axis or
   >> ylim([<minimum y>, <maximum y>])
for the y axis.

XXX script m-file: running_gaussians
a = .5;
L = 10;
max_time = 100;
del_time = .01;
max_vert_axis = 3;
f = @(x) exp(-((x - L) - L/2).^2 / a^2);
nc_nodes = 20;
c = [1:nc_nodes]';
nr_points = 1001;
x = linspace(0, 10, nr_points);
X, C = meshgrid(x, c);
R = repmat(linspace(0:(2*pi-1), 1, nr_points));
g = @(t) sum(R.*((x-C(t)));
ast_y = linspace(., max_vert_axis, .95*max_vert_axis, nr_nodes);
for t = 0:del_time:max_time
    >> ast_x = mod(c*t + L/2, L);
    >> plot(x, g(t), ast_x, ast_y, 'r*')
    >> axis([0 L 0 max_vert_axis])
    >> title(t)
end

Suppose that $H$ for some strange reason you want to plot the periodic function
curves on a plot. But this rapidly becomes unwieldly. Instead, just replace one plot by another. For exam-
ple, time evolution of some function by repeatedly using plot.

\texttt{figure(gcf)}
The function axis is given only in effect for one plot. Every new plot turns it off, so it must be called for every plot (unless hold on has been invoked).

The plot function generates linear axes. To generate logarithmic axes use semilogx for a logarithmic axis in x and a linear axis in y, semilogy for a linear axis in x and a logarithmic axis in y, and loglog for logarithmic axes in both x and y.

MATLAB has two different functions to plot a function directly rather than plotting a set of points.

Warning: These functions do not always generate the correct curve (or curves) because they know nothing of the actual behavior of the function. They can have problems with sharp peaks and asymptotes and other “strange behavior.” We will show some examples shortly.

The first function we discuss is fplot, which can be executing by simply entering

```matlab
fplot(<function handle>, <limits>)
```

where the function is usually generated as an anonymous function or a MATLAB function or a user-generated function m-file (as described in Section 8.3). The limits are either

* [xmin xmax] in which case the y-axis just encloses the curve or
* [xmin ymin ymax] in which case you are also specifying the endpoints on the y-axis.

Note: Recall in Section 3.2.2 we discussed how to pass a function as an argument.

This function uses adaptive step control to generate as many data points as it considers necessary to plot the function accurately. You can also store the data points calculated by

```matlab
[x, y] = fplot(<function handle>, <limits>)
```

rather than having the function plotted directly. You then have complete control over how to plot the curve using the plot function.

The other function which can plot a function is ezplot, which is more general than fplot. To plot a function on the interval \([-2\pi, 2\pi]\) enter

```matlab
ezplot(<function handle>)
```

To include limits (as with fplot) enter

```matlab
ezplot(<function handle>, <limits>)
```

In addition, a parametrically defined function can be plotted by

```matlab
ezplot(<function 1>, <func 2>, <limits>)
```

Finally, this function can also plot an implicitly defined function, i.e., \( f(x, y) = 0 \), by

```matlab
ezplot(<2D func>, <limits>)
```

For example,

```matlab
f = @(x, y) (x^2 + y^2)^2 - (x^2 - y^2);
```

```matlab
ezplot(f)
```

plots the lemniscate of Bernoulli (basically an “u” symbol). Warning: Be particularly careful when plotting implicit functions because they can be REALLY HARDY and occasionally ezplot may not get it right.

There is an important difference between

```matlab
fplot(f, [-5 5])
```

and

```matlab
ezplot(f, [-5 5])
```

In the former \( f(x) \) is only evaluated for scalar values of \( x \), while in the latter \( f(x) \) is evaluated for vector values of \( x \). Thus, when using ezplot care must be taken if \( f \) is evaluated in a function m-file. If \( f(x) \) cannot be evaluated for vector values, the error message

Warning: Function failed to evaluate on array inputs; vectorizing the function may speed up its evaluation and avoid the need to loop over array elements

will be generated.
you now rerun this code you will see a much smoother histogram.

A histogram shows the frequency of values in a vector, say \( x \) again, but suppose we want to compare this histogram to an actual probability density function. For example, we have just discussed the Gaussian distribution. If it has mean \( \mu \) and standard deviation \( \sigma \), then the density function is

\[
  f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( -\frac{(x-\mu)^2}{2\sigma^2} \right)
\]

where \( \int_{-\infty}^{\infty} f(x)\,dx = 1 \). To generate a random sequence \( x \):

- \( \mu = 5; \)
- \( \sigma = 2; \)
- \( \text{randn} = 100000; \quad \Rightarrow \quad x = \mu + \sigma \text{randn}(100000, 1); \)
- \( \text{histogram}(x) \)

In order to compare the histogram with the density function, we must rescale the histogram so that its area is 1 by

- \( \text{histogram}(x, 'Normalisation', 'pdf') \)

There are a number of other normalizations that can be used, including a cumulative probability distribution (CDF). It is also possible to only draw the top of the histogram by using the property name DisplayStyle and property value stairs.

In addition, it is possible to count the number in each bin rather than plotting them by

- \( \text{nrdata b SRRRRR} \quad \text{bin b histcounts}(x, e) \)
- \( \text{nr} = \text{histcounts}(x) \)

if you don’t want to determine the edges of the bins yourself.

We have already seen how to plot the vector \( x \) vs. the vector \( y \) by using the plot function. If, additionally, you have an error bar of size \( e_i \) for each point \( y_i \), you can plot the curve connecting the data points along with the error bars by

- \( \text{errorbar}(x, y, e) \)

Sometimes the error bars are not symmetric about the \( y \) values. In this case, you need vectors \( 1 \) and \( u \) where at \( x_i \) the error bars extend from \( y_i - u_i \) to \( y_i + u_i \). This is done by

- \( \text{errorbar}(x, y, 1, u) \)

Note: All the elements of \( 1 \) and \( u \) are non-negative.

Data can also be entered into MATLAB from a separate data file. For example,

- \( M = \text{csvread}'<\text{file name}>\) \)

reads data from a file one row per line of input. The numbers in each line must be separated by commas. The data can then be plotted as desired. The function \( \text{csvwrite} \) writes the elements of a matrix into a file using the same format. (If desired, you can have much more control over how data is input and output by using the \( \text{fscanf} \) and \( \text{fprintf} \) functions, which are similar to their C counterparts. These functions are discussed in detail in Section 6.)

The \( \text{load} \) function can also be used to read a matrix into MATLAB from a separate data file. The data must be stored in the data file one row per line. The difference between this function and \( \text{csvread} \) is that the numbers can be separated by commas or semicolons or by spaces. The matrix is input by entering

- \( \text{load} <\text{file name}> \)
- \( \text{load } ' <\text{file name}>' \)
- \( \text{load} '<\text{file name}>' \)

and it is stored in the matrix named \( <\text{file name}-no \text{ extension}> \) (i.e., drop the extension, if any, in the file name). Or you can enter

\section*{Note:}

The advantage of using the \( \text{print} \) function is that the arguments can be variables. An oversimplified example is

- \( \text{device} = ' -deps'; \)
- \( \text{file} = '<\text{file name} \text{ b&w}>'; \)

The advantage of using the \( \text{print} \) function is that the arguments can be variables. An oversimplified example is

- \( \text{device} = ' -deps'; \)
- \( \text{file} = '<\text{file name} \text{ b&w}>'; \)
- \( \text{print} \quad \text{device}, \text{file} \)

It is oversimplified because there is no need to use three lines when one will do. However, if many plots are to be printed then the print device can be changed once rather than in every print function. Also, if you are printing many plots then you can easily modify the file names as in

- \( i = 1; \)
- \( \text{file} = ['fiddlededum', num2str(i), '.eps']; \)
- \( \text{print} \quad \text{device}, \text{file} \)
- \( i = i + 1; \)
- \( \text{file} = ['fiddlededum', num2str(i), '.eps']; \)
- \( \text{print} \quad \text{device}, \text{file} \)

You can load variables from a file containing all the variables to the file \( \text{allvariables.mat} \) in binary format and then delete all the variables. Entering

- \( \text{load} \quad \text{allvariables} \)
- \( \text{save} \quad \text{allvariables.mat} \)

loads all these variables back into MATLAB using the binary format. On the other hand, if you create a file, say \( \text{symmatrix.dat} \), containing the elements of a matrix and enter it into MATLAB using

- \( \text{load} \quad \text{symmatrix} \)

you obtain a new matrix, called \( \text{symmatrix} \), which contains these elements. Thus, the load function determines how to read a file depending on the extension.
csvread('<file name>')  Reads data into MATLAB from the named file, one row per line of input; the numbers in each line can be separated by spaces or commas. The name of the resulting matrix is '<file name>'.

load('<file name>')    Reads data into MATLAB from the named file, one row per line of input; the numbers in each line can be separated by spaces or commas.

importdata('<file name>') Similar to load but there need not be the same number of elements in each row.

csvwrite('<file name>', A) Writes out the elements of a matrix to the named file using the same format as csvread.

print
Prints a plot or saves it in a file using various printer specific formats. For example, print -deps '<file name>' saves the plot in the file using encapsulated PostScript (so it can be plotted on a PostScript laser printer).

plot(x, y)  Plots the data points in Cartesian coordinates. The general form of this function is plot(x1, y1, x2, y2, ... ) where x1, y2, ... are optional character strings containing information about the type of line, mark, and color to be used. Some additional arguments that can be used:

- plot(x) plots x vs. the index number of the elements.
- plot(Y) plots each column of Y vs. the index number of the elements.
- plot(x,Y) plots each column of Y vs. x.

If z is complex, plot(z) plots the imaginary part of z vs. the real part.

semilogx
The same as plot but the x axis is logarithmic.

semilogy
The same as plot but the y axis is logarithmic.

loglog
The same as plot but both axes are logarithmic.

fplot(<function handle>, [limits])
Plots the specified function within the limits given. The limits can be [xmin xmax] or [xmin xmax ymin ymax].

ezplot(<function handle>)
Generates an "easy" plot (similar to fplot) given the function f(x). It can also plot a parametric function, i.e., (x(t), y(t)), or an implicit function, i.e., f(x,y) = 0. Limits can also be specified if desired.

ex2fplot(<2D fnc>)
Plots the data points in polar coordinates.

ezpolar(<function handle>)
Generate an "easy" polar plot of r = <function handle>(θ).

xlabel(<string>)
Puts a label on the x-axis.

ylabel(<string>)
Puts a label on the y-axis.

title(<string>)
Puts a title on the top of the plot.

axis
Controls the scaling and the appearance of the axes. axis equal and axis([xmin xmax ymin ymax]) are two common uses of this function. The endpoints of the current plot are returned by axis (i.e., with no arguments).

xlim
Sets the endpoints of the x axis.

ylim
Sets the endpoints of the y axis.

hold
Holds the current plot (hold on) or release the current plot (hold off).

linspace(a, b, n)
Generates n equally-spaced points between a and b (inclusive).

logspace(a, b, n)
Generates n logarithmically-spaced points between 10^a and 10^b.

histcounts
Returns the number in each bin.

histogram(x)
Plots a histogram of the data in a vector using an appropriate number of bins.

hist(x, <number of bins>) changes the number of bins.

errorbar(x, y, e)
The first plots the data points x vs. y with error bars given by e.

errorbar(x, y, 1, u)
The second plots error bars which need not be symmetric about y.

subplot(m, n, p)
Divides the graphics window into m x n rectangles and selects the pth rectangle for the current plot. In addition, rectangles can be placed anywhere inside the window by subplot(‘Position’, ...).

shg
Raises the current graphics window so that it is visible, or creates a new graphics window if none exists.

drawnow
Update the current figure (which is frequently not done if MATLAB is executing further statements).
4.2. Three-Dimensional Graphics

The MATLAB function `plot3` plots curves in three-dimensions. For example, to generate a helix enter

```matlab
>> t = linspace(0, 20*pi, 1001);
>> c = cos(t);
>> s = sin(t);
>> plot3(c, s, t);
```
and to generate a conical helix enter

```matlab
>> t = linspace(0, 20*pi, 2001);
>> c = cos(t);
>> s = sin(t);
>> plot3(t.*c, t.*s, t);
```
Also, you can put a label on the z-axis by

```matlab
>> zlabel('string')
```
There is also an “easy” `plot3` function. It generates the curve \((x(t), y(t), z(t))\) for \(t \in (0,2\pi)\) by

```matlab
>> ezplot3(x, y, z)
```
if \(x, y, \) and \(z\) have been defined using anonymous functions. Again, you change the domain of \(t\) by specifying the additional argument \([\text{tmin}, \text{tmax}]\).

MATLAB also plots surfaces \(z = f(x,y)\) in three-dimensions with the hidden surfaces removed. First, the underlying mesh must be created. The easiest way is to use the function `meshgrid`. This combines a discretization of the \(x\) axis, i.e., \(\{x_1, x_2, \ldots, x_m\}\), and the \(y\) axis, i.e., \(\{y_1, y_2, \ldots, y_n\}\), into the rectangular mesh \(\{(x_i, y_j) | i = 1,2, \ldots, m, j = 1,2, \ldots, n\}\) in the \(x-y\) plane. The function \(f\) can then be evaluated at these mesh nodes. For example,

```matlab
>> x = linspace(-3, 3, 61); % generates a colored, wire-frame surface whereas
>> y = linspace(-2, 2, 41); % generates a colored, filled-in surface. We discuss how to change the colors, and even how to use the colors as another variable, in the next section.
>> [X, Y] = meshgrid(x, y);
>> F = (X + Y).*exp( -X.^2 - 2*Y.^2 );
>> mesh(X, Y, F)
```
generates a colored, wire-frame surface whereas

```matlab
>> surf(X, Y, F)
```
generates a colored, filled-in surface. We discuss how to change the colors, and even how to use the colors as another variable, in the next section.

You can change the view of a three-dimensional plot by clicking on the menu item which shows a contour lines are plotted at specific values by

```matlab
>> contourf(X, Y, Z)
```
where \(Z\) is the function evaluated at these mesh nodes. For example,\(f(r, \theta) = \frac{r + \cos(2\theta)}{1 + 2 \cdot r \cdot \cos(3\theta)}\) for \(r \leq 1\)

```matlab
>> f = @(r, th) ( r + exp(2*r.*sin(2*th)) ) ./ ( 1.2 - r.*cos(3*th) );
>> r = linspace(0, 1, 61);
>> th = linspace(0, 2*pi, 61);
>> [R, TH] = meshgrid(r, th);
>> [X, Y] = pol2cart(TH, R);
>> surf(X, Y, f(R, th))
```
The function `pol2cart` transforms the polar coordinates into cartesian coordinates which can be understood by `surf`, `mesh`, or `contour`.

We close with an additional detail about `meshgrid`. It can also generate a grid in three dimensions by, for example,

```matlab
>> x = linspace(-3, 3, 61)';
>> y = linspace(-2, 2, 41)';
>> z = linspace(-1, 1, 11)';
>> [X, Y, Z] = meshgrid(x, y, z);
```
Three dimensions is the highest we can go with `meshgrid`. However, a multidimensional grid can also be generated by

```matlab
>> [X, Y, Z] = ndgrid(x, y, z); % and ndgrid can be used in any number of dimensions. The difference between the two functions is that the order of the first two arguments is reversed. For example,
```

```matlab
>> X = meshgrid(1:3, 4:7) % returns
   1  2  3
   1  2  3
   1  2  3
>> Y = [ 4  4  4
       5  5  5
       6  6  6
       7  7  7 ]
```
while

```matlab
>> X = ndgrid(1:3, 4:7) % returns
   1  1  1
   2  2  2
   3  3  3
>> Y = [ 4  5  6
       4  5  6
       4  5  6 ]
```
If you type

```matlab
>> contour(X, Y, F)
```
you will see contour plots of the surface. That is, you will be looking down the \(z\) axis at curves which represent lines of constant elevation (i.e., constant \(z\) values). If we type

```matlab
>> contour3(X, Y, F)
```
you will see contour plots of the surface in three dimensions. You can again change your view of these curves by using the `view` function. These contour lines are labelled by

```matlab
>> [C, h] = contour(X, Y, F)
```
and `clabel(C, h)`

Also, contour lines are plotted at specific values by

```matlab
>> contour(X, Y, F, v)
```
where \(v\) is a vector of the values. To obtain a filled contour plot enter

```matlab
> contourf(X, Y, F)
```
If you do not want to bother with generating the mesh explicitly, you can generate “easy” plots by

```matlab
>> ezcontour, ezcontour3, ezmesh, and ezsurf.
```
A surface can also be plotted in polar coordinates. For example, the code which plots

```matlab
f(r, \theta) = \frac{r + \cos(2\theta)}{1 + 2 \cdot r \cdot \cos(3\theta)}\)
```
is

```matlab
>> f = @(r, th) ( r + exp(2*r.*sin(2*th)) ) ./ ( 1.2 - r.*cos(3*th) )
>> r = linspace(0, 1, 61);
>> th = linspace(0, 2*pi, 61)
>> R = linspace(0, 1, 61)
>> [R, TH] = meshgrid(r, th)
>> [X, Y] = pol2cart(TH, R)
>> surf(X, Y, f(R, th))
```
and `ezsurf` can be used in any number of dimensions. The difference between the two functions is that the order of the first two arguments is reversed. For example,
Three-Dimensional Graphics

plot3(x, y, z) Plots the data points in Cartesian coordinates. The general form of this function is plot3(x1, y1, z1, x2, y2, z2, ...) where x1, y1, z1, ... are optional character strings containing information about the type of line, marker, and color to be used.

Default is size = 0.5
- 'MarkerSize', size - where size is approximately the diameter of the marker. The size of the point marker, i.e., ".", is \( \frac{1}{2} \) this size, and you can draw a point which is approximately one pixel in diameter by letting size = 1.

It is possible to obtain the current position of the cursor within a plot by using the ginput function.
Example: To collect any number of points enter

```
>> [x, y] = ginput
```
Each position is entered by pressing any mouse button or any key on the keyboard except for the carriage return (or enter) key. To terminate this function press the return key. To enter exactly n positions, use

```
>> [x, y] = ginput(n)
```
You can terminate the positions at any time by using the return key. Finally, to determine which mouse button or which key was used, enter

```
>> [x, y, button] = ginput(n)
```

The vector button contains integers specifying which mouse button (1 = left, 2 = center, and 3 = right) or which key (its ASCII representation) was pressed.

Labels can also be added to a plot. Text can be placed anywhere inside the plot using

```
>> text(xpt, ypt, 'string')
```
The text is placed at the point (xpt,ypt) in units of the current plot. The default is to put the center of the left-hand edge of the text at this point. You can also use the mouse to place text inside the plot using

```
>> gtext('string')
```
The text is fixed by depressing a mouse button or any key.

If more than one curve appears on a plot, you might want to label each curve. This can be done directly using the text or gtext function. Alternatively, a legend can be put on the plot by

```
>> legend('string1', 'string2', ...)
```
Each string appears on a different line preceded by the type of line so you should use as many strings as there are curves. The entire legend is put into a box and it can be moved within the plot by using the left mouse button.

To add labels to a plot, you can use

```
>> xlabel('string')
>> ylabel('string')
```

Note: For you TeX users note the funny control sequence \beta(x) which generates \( \beta(x) \). If you would have typed \( \beta(x) \) you would have obtained \( \beta(x) \) because MATLAB preserves spaces. If typing \( \beta(x) \) sets your teeth on edge, try \( \beta(x) \) instead.

It is often essential for the title to include important information about the plot (which would, otherwise, have to be written down "somewhere" and connected to this specific plot). For example, suppose you enter

```
>> x = linspace(0, 2*pi, 100)
>> c1 = 2
>> c2 = -3
>> p1 = 1
>> p2 = 3
>> y = c1*sin(x)-p1 + c2*cos(x)-p2
>> plot(x,y)
```
and you want to "play around" with the two coefficients to obtain the most "pleasing" plot. Then you probably should have the title include a definition of the function — and you should not have to modify the title by hand every time you change the coefficients. This can be done by

```
>> title('y = c1sin(x) - p1 + c2cos(x) - p2')
```

1On a Macintosh computer you probably do not have a center mouse button.

4.3. Advanced Topic: Functions

In the previous subsections we have discussed how to use "simple" graphics functions to generate basic plots. MATLAB can also do much more "interesting" graphics, and even publication quality graphics. Here we discuss some of the more useful advanced features. We divide the topic into two subsections: the first discusses the functions themselves and the second discusses how to change some of the properties of these functions.

Note: The demonstration program shows many more of the graphics capabilities of MATLAB. Enter demo and then in Help Navigator click on Graphics.
A useful color map for outputting to laser printers is 'gray'. In this colormap all three components of a color are the same. For example:

```matlab
f = 0.5*(sin(2*pi*x.*y.*z)+2+1.5*sin(2*pi*abs(x.*y.*z)+2));
```

```matlab
colormap(f)
```

The 'gray' colormap is useful for outputting to laser printers. It is a simple colormap where each channel has the same value, and it is often used for grayscale images.
the values in the contour plot.

Note: There is a linear scaling between the value of \( f \) and the corresponding color in the colormap \( C \in \mathbb{R}^{n \times k} \). For \( \text{caxis}(\{f_{\text{min}}, f_{\text{max}}\}) \), the value \( f \) is first modified to lie in the interval \( [f_{\text{min}}, f_{\text{max}}] \) by \( f_{\text{mod}} = \frac{\text{max}(\text{min}(f, f_{\text{max}}), f_{\text{min}}) - f_{\text{min}}}{f_{\text{max}} - f_{\text{min}}} \). Then the row \( i \) of \( C \) is calculated by

\[
  i = \text{fix} \left( \frac{f_{\text{mod}} - f_{\text{min}}(n-1)}{f_{\text{max}} - f_{\text{min}}} \right) + 1.
\]

MATLAB can also plot a two-dimensional image (i.e., a picture) which is represented by a matrix \( X \in \mathbb{R}^{m \times n} \). The \((i,j)\)th element of \( X \) specifies the color to use in the current colormap. This color appears in the \((i,j)\)th rectilinear patch in the plot. For example, to display the color image of a clown enter

- \text{clf}
- \text{imshow(clown)}
- \text{colorbar}

The image function inputs the matrix \( X \) and the colormap \( C \) from \text{clown.mat}. Then the image is displayed using the new colormap. Similarly,

- \text{load earth}
- \text{imagesc(earth)}
- \text{clim([-100 100])}

An image is a collection of RGB values for each pixel. RGB images can be imported into MATLAB by

- \text{imread(filename)}
- \text{imread(filename, 'id')}

where \( \text{id} \) is the associated colormap for the image. The types of files which can be imported are: "bmp", "cur", "txt", "gif", "bat", "ico", "jpe", "jif", "jpt", "jpa", "jp2", "jpg", "pbm", "pcx", "pgm", "png", "pam", "ppm", "ras", "tif", "tiff", or "xwd". They can be shown by \text{image} or \text{imshw}.

They can be exported by \text{imwrite}.

4.4. Advanced Topic: Handles and Properties

In this subsection we briefly discuss \textit{handle graphics}. This is a collection of low-level graphics functions which do the actual work of generating graphics. In the previous parts of this section we have mainly discussed “high-level” graphics functions which allow us to create useful and high-quality graphical images very easily. The low-level functions allow us to customize these graphical images, but at the cost of having to get much more involved in how graphical images are actually created. This subsection will be quite short because we do not want to get bogged down in this complicated subject. Instead, we will only discuss a few of what we consider to be — the more useful customizations.

In handle graphics we consider every component of a graphical image to be an \textit{object}, such as a subplot, an axis, a piece of text, a line, a surface, etc. Each object has \textit{properties} and we customize an object by changing its properties. Of course, we have to be able to refer to a particular object to change its properties, and a \textit{handle} is the unique identifier which refers to a particular object. (Each handle is a unique floating-point number.)

We will use a small number of examples to explain handle graphics. There are many properties of the text that can be changed in the \textit{text} function by

- \text{text(xpt, ypt, 'string', 'Prop 1', 'Value 1', 'Prop 2', 'Value 2', ...)}
- \text{h = text(xpt, ypt, 'string'); set(h, 'Prop 1', 'Value 1', 'Prop 2', 'Value 2', ...)}

where \text{Prop ?} is the name of one of the properties for the text object and \text{Value ?} is one of the
allowed values. (We show some names and values in the following table.) We have shown two ways to customize the properties. In the former all the properties are set in the text function. In the latter the text function creates an object, using its default properties, with handle h. The set function then changes some of the properties of the object whose handle is h. For example, entering

```matlab
set(h, 'Color', 'r', 'FontSize', 16, 'Rotation', 90)
```
results in a large, red text which is rotated 90°. You can also change the default properties for gtext, xlabel, ylabel, zlabel, and title.

### Text Properties

<table>
<thead>
<tr>
<th>Property</th>
<th>Default</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clipping</strong></td>
<td><strong>on</strong></td>
<td>Any portion of the text that extends outside the axes rectangle is clipped.</td>
</tr>
<tr>
<td><strong>Color</strong></td>
<td><strong>none</strong></td>
<td>A three-element vector specifying a color in terms of its red, blue, and</td>
</tr>
<tr>
<td></td>
<td></td>
<td>given components, or a string of the predefined colors.</td>
</tr>
<tr>
<td><strong>FontName</strong></td>
<td><strong>Helvetica</strong></td>
<td>The name of the font to use. (The default is Helvetica.)</td>
</tr>
<tr>
<td><strong>FontSize</strong></td>
<td><strong>10</strong></td>
<td>The font point size. (The default is 10 point.)</td>
</tr>
<tr>
<td><strong>HorizontalAlignment</strong></td>
<td><strong>center</strong></td>
<td>Text is centered.</td>
</tr>
<tr>
<td><strong>Rotation</strong></td>
<td><strong>right</strong></td>
<td>Text is right justified.</td>
</tr>
<tr>
<td><strong>VerticalAlignment</strong></td>
<td><strong>top</strong></td>
<td>The top of the text rectangle is at the point.</td>
</tr>
</tbody>
</table>

The more common way of customizing parameters is by using the set function. The two functions `get` and `set` are used to obtain the value of one parameter and to set one or more parameters. For example, to get the font which is presenting being used enter

```matlab
s = get(h, 'FontName')
```
and the string s now contains the name of the font. The two arguments to `get` are the handle of the object desired and the name of the property.

There are two other functions which can obtain a handle: `h = gca` returns the handle of the current axes figure and `ha = gca` returns the handle of the current axes in the current figure. For a simple example which uses handle graphics, suppose we want to plot the function \( y = e^{\sin x} \) for \( -\pi / 2 \leq x \leq \pi / 2 \) and we want the horizontal axis to have major tick marks at every \( \pi / 2 \) and we want these tick marks labelled. We do this by

```matlab
x = linspace(0,2*pi,101);
y = exp(sin(x));
set(gca, 'XTick', [0:pi/2:2*pi])
set(gca, 'XTickLabel', {'0', '^0.5\pi', '\pi', '1.5\pi', '2\pi'})
set(gca, 'MinorTick', 'on')
set(gca, 'TickDir', 'out')
```
where the third line puts major tick marks at multiples of \( \pi / 2 \), the fourth line puts the labels shown at each tick mark, the fifth line adds small tick marks between the labelled ticks, and the last line puts the tick marks outside the plot border area. (At present, we cannot use \( \pi \) characters in `XTickLabel` to generate a Greek \( \pi \).

There is one case where we frequently use handle graphics. If a figure is printed, the graphical images do not fill the entire page. The default size is approximately 6.5 inches wide and 5.5 inches high. When we want to use the full size of a sheet of paper we use

```matlab
figure('PositionPaper', [0 0 8.5 11])
```

### Advanced Topic: GUIs (Graphical User Interfaces)

Warning: The codes in this subsection use for loops and if tests which are discussed in Section 8.1. They also contain primary functions, subfunctions, and nested functions which are discussed in Section 8.3.

MATLAB makes it very easy to create a graphical user interface for a program. Of course, it is usually even easier for the programmer to have the users interact with a program by entering data (i.e., typing) in the workspace. However, this is often not easier for the users. Thus, the programmer must make the decision whether or not to spend the time required to write a GUI. Note that this not only includes the time required to code the GUI but also the — possibly much longer — time required to design the GUI. In this subsection we are only concerned with how to write a GUI — not how to write a good one.

A GUI is a graphical display which allows a user to interact "pictorially" with a program. It usually consists of three elements:

1. **Components**: These are the tools which enable the user to interact with a program, and include push buttons, sliders, radio buttons, check boxes, editable text, pop-up menus, listboxes, and toggle buttons. Such an interaction is called an event, and a program which responds to events is called driven. Components can also include plots and/or tables which allow the program to interact with the user, i.e., show results of the program.

2. **Figures**: All of these components must be arranged within a figure, which is a graphical window which is separate from the MATLAB window which arises when MATLAB is first executed.

3. **Call-backs**: These are pieces of code which enable the user to interact with the program. For example, when a user uses a mouse to click a button, this triggers an event, but it is not directly connected with the program. It is the responsibility of the programmer to write the code which specifies how each event affects the program.

The simple way to write a GUI is to type

```matlab
figure
```
This launches a GUI which you can use to write your own GUI. The largest part of this window is a rectangular and gridded canvas into which the various components are placed. To the left of this canvas is a list of all the possible components, either using icons or their actual names. You build your own GUI by positioning any or all of these components in the canvas. When done, you save your GUI under its own
name. You have created two files: the fig file (i.e., the extension is fig) and the function file (i.e., the extension is m). So far you have only created the first two elements of a GUI: the components and the figure. The third part, i.e., the callbacks, are your responsibility. Your program must be integrated into this function file and you must also write the code to connect an event with its effect on your now integrated program.

We will not discuss guide any further, and we will not discuss how to integrate your program and code the callbacks. This is well-documented in MATLAB, including an explicit example which is called simplegui. Read this documentation carefully and work through this explicit example yourself.

What we will discuss is how to use the non-simple way to write a GUI, namely by writing all the actual functions which setup the figure, the components, and the callbacks. Since this can get quite complicated, we only show two very simple examples. We will not write any GUIs from scratch. Instead, we will show how to take a simple figure which is created by the plot function and add “something” to it. Thus, we will not attempt to describe all the functions which are involved in generating a GUI, but only a very small subset.

For the first example consider the function

```matlab
function rippling(a, b)
    n = 101;
x = linspace(-2, 2, n);
y = x;
[X, Y] = meshgrid(x, y);
for t = 0:0.01:10
    at = a + 0.4*t;
    bt = b + 0.4*t;
    Z = sin((at.*x + 0.2*Yyz).*e.^-2 * ...
                 bt*sin(2*Yyz) + 0.2*Yyz).*Yyz.^2);
surf(X, Y, Z);
view([-90, 60]);
xlabel('x')
ylabel('y')
zlabel('z')
drawnow;
end
```

which generates the “interesting” surface

\[
z = f(x, y, t, a, b) = \sin\left((a + 0.8t + 0.1x + 0.2y)2 + (b + 0.4t)\sin 2\pi y + 0.2y)\right)2(y)
\]

where

\[
\eta(y) = y + 0.25\sin 2(\pi/2 - 1).
\]

It is “rippling” in time (and the time is shown at the top of the 3D surface). Suppose we would like to stop the time evolution briefly to admire the surface. This is easily done by

```matlab
function pause_rippling
    uicontrol(gcf, ...
        'Style', 'PushButton', ...
        'String', 'Pause', 'callback', 'button_callback';
        rippling(0, 0) % the above function which generates the surface
        XXXX, ... subfunction
        function button(hObject, event)
            set(hObject, 'String', 'Continue')
            uiwait
        else
            uiresume
            set(hObject, 'String', 'Pause')
        end
    end
```

The actual figure is shown below when the push button has been clicked.

The discussion of this code follows. amsg either raises the current already-existing figure, or creates a new figure. The uicontrol function (user interface control) positions a component (check box, editable text field, list box, pop-up menu, push button, radio button, slider, toggle buttons, or static text field) in the figure. This component must be attached to something which is called its parent. (uipanel can be used to split the GUI into different regions; a uicontrol can then be attached to any of these panels or the original figure.) The first argument to the uicontrol function in the code determines its parent. (It is not actually needed because the default parent is the current figure, but it never hurts.) The remaining arguments occur in pairs with the first being the property name and the second its value (which might be a string or a scalar or a matrix or a structure or a figure handle or a . . .). The first pair of arguments makes control a push button, the second that the word Pause should appear on the button, the third that the units are pixels, and the fourth the position of the button in the figure (using these units). The button is to be positioned beginning at the lower left hand corner of the figure and extend 80 pixels to the right and 20 pixels upward.1 However, the push button still does not do anything. The last pair of arguments makes the event of pushing a button execute the function button. The next line in the code runs the function rippling (which has been discussed previously) using the parameters \( a = b = 0 \).

As long as the user does not click the push button, everything is as if rippling was run directly from the workspace. Clicking the push button (i.e., triggering an event) causes the function button to be executed where the first argument hObject is a handle to the particular event. (The second argument is unused.) get returns the value of the property name String which is attached to the handle hObject. This value is the string Pause and so the if statement is true. The function set is executed and it replaces the push button label by Continue. uiresume then blocks execution of the program (i.e., plotting the surface). The program is now in a wait state and will remain there until the user clicks the push button (whose label is now Continue) again. button is again executed and now the if statement is false because hObject contains Continue. uiresume is executed so the program resumes execution. The following set function changes the label on the push button back to Pause.

The above GUI does not allow us to modify \( a \) or \( b \). The following one does.

---

1The button could have been positioned in a “more pleasing” location in the GUI because there is lots of avai space in a three-dimensional plot. However, in a two-dimensional plot this is a “safe” location.
function pause_rippling2
shg
h_pause = uicontrol(gcf, 'Style', 'PushButton', 'String', 'Start', ...
    'Units', 'Pixels', 'Position', [0 0 80 20], ...
    'Callback', @button);
a = 0;
b = 0;
ucicontrol(gcf, 'Style', 'Text', 'String', 'a = ', 'Position', [90 0 30 20]);
ucicontrol(gcf, 'Style', 'Edit', 'String', num2str(a, ',4.2f'), ...
    'Position', [120 0 60 20], 'Callback', @edit_a);
ucicontrol(gcf, 'Style', 'Text', 'String', 'b = ', 'Position', [190 0 30 20]);
ucicontrol(gcf, 'Style', 'Edit', 'String', num2str(b, ',4.2f'), ...
    'Position', [220 0 60 20], 'Callback', @edit_b);
drawnow

uicontrol

function button(hObject, eventdata)
    %%%%% nested functions follow
    function edit_a(hObject, eventdata)
        num = str2num(get(hObject, 'String'));
        if strcmp(eventData, 'Start')
            uiresume
            set(hObject, 'String', 'Pause')
        elseif strcmp(eventData, 'Pause')
            uiresume
            set(hObject, 'String', 'Continue')
        else
            uiresume
            set(hObject, 'String', 'Pause')
        end
    end
    function edit_b(hObject, eventdata)
        num = str2num(get(hObject, 'String'));
        if strcmp(eventData, 'Start')
            uiresume
            set(hObject, 'String', 'Pause')
        elseif strcmp(eventData, 'Pause')
            uiresume
            set(hObject, 'String', 'Continue')
        else
            uiresume
            set(hObject, 'String', 'Pause')
        end
    end

This is shown below where both a and b have been modified and the push button has been clicked.

The first uicontrol is the same as before, except that the word Start initially appears on the push button. The function rippling does not begin running until it is pushed. Afterwards, the label alternates between Pause and Continue. The next two uicontrols put in the static text field “a = ” and the editable text field which contains the default value of a. Note that the value of a is contained in the String property name and its string value is num2str(a, ',4.2f'). 1 If the number (actually the string) is modified, the function edit_a is executed; in this function the value of a is changed by getting the character variable and changing it to a number. Note that this is a nested function (to be discussed in Section 8.3) so that the value of a which is modified in this function is also modified in the primary function pause_rippling2. The last two uicontrols do the same for b.

The drawnow command, which immediately follows the creation of all the components, is necessary so that the components actually appear in the figure. The following function, i.e., uistat, puts the GUI into a wait state. This state continues until the user clicks the push button, which causes the function button to be executed. The string contains Start so the if statement is true so the GUI resumes execution, i.e., the function rippling is executed, and the label is changed to Pause.

1The second argument to num2str causes the number to appear as ±×××.

2The function button is also a nested function, whereas it was a subfunction in pause_rippling. Since it does not modify a variable which is needed by the primary function, it can be either. Since edit_a and edit_b need to be nested function, we also made button one.
4.6. Advanced Topic: Making Movies

There are a number of different ways to make movies in MATLAB. We show some of them in the following code which is a modification of the MATLAB function life (which is copyrighted by The MathWorks, Inc). The documentation inside life follows.

**LIFE** MATLAB’s version of Conway’s Game of Life:

"Life" is a cellular automaton invented by John Conway that involves live and dead cells in a rectangular, two-dimensional universe. In MATLAB, the universe is a sparse matrix that is initially all zero.

Whether cells stay alive, die, or generate new cells depends upon how many of their eight possible neighbors are alive. By using sparse matrices, the calculations required become astonishingly simple. We use periodic (torus) boundary conditions at the edges of the universe. Pressing the "Start" button automatically seeds this universe with several small random communities. Some will succeed and some will fail.

C. Moler, 7-11-92, 8-7-92
Adapted by Ned Gulley, 6-21-93
Copyright 1984-2004 The MathWorks, Inc.
$Revision: 5.10.4.1 $ $Date: 2004/08/16 01:38:30 $ $

This code strips out the part of **life** which does the actual iterations and shows how to save the results in an **avi** movie. Two different methods are shown explicitly in the code. The first initializes the movie, adds frames explicitly to the movie, and then closes it. The second only saves each frame in a **jpg** file; an external software package must then be used to combine these files into a movie. A third method, which is a slight modification of the first, is then discussed.

The advantage of the second method is that no additional memory is required; the disadvantage is that each frame must be individually saved to disk. Another disadvantage of the first and third methods in Linux and Mac is that the movie file which is generated is not compressed. An external software package must be used to compress it.

**Warning:** This code uses a number of MATLAB functions which have not been yet discussed: **sparse**, **uicontrol**, **uiwait**.
for. if, while, break, find, and logical expressions.

script m-file: sample_movie
% modified from the MATLAB function life which is
% Copyright 1984-2004 The MathWorks, Inc.
seed = 3343; % modify seed to obtain different cells
rand(‘state’, seed);
M = avifile(‘life_movie.avi’);
m = 26;
max_nr_iter = 10000;
X = sparse(m, m);
% Huhis “lowMlevel” function generates a new plot much faster than using
% ‘plot’ function rather than the
p = 311; % Huhis “lowMlevel” function generates a new plot much faster than using
for count = 1:15
kk = floor((m - 4)*rand(1) + 2);
ky = floor(m*(m - 4)*rand(1) + 2);
X(kk,ky) = (rand(3) > 0.5),
end;
[l, j] = find(X);
nr_iter = 0;
plothandle = plot(l, j, ’’, ’Color’, ’Red’, ’MarkerSize’, 12);
title(’$n$ iter = ’, num2str(nr_iter))
axis([0 m+ 0 nr])
drawnow
F = getframe(gcf);
M = addframe(F, F);
file_name = [’life_movie’, num2str(nr_iter, ’%06d’), ’.jpg’];
saveas(gcf, file_name);
Xprev = X;
Xprev2 = X;
if nr_iter == 1
plot(handle, ’Color’, ’Red’, ’MarkerSize’, 12);
title(’$n$ iter = ’, num2str(nr_iter))
drawnow
F = getframe(gcf);
M = addframe(F, F);
file_name = [’life_movie’, num2str(nr_iter, ’%06d’), ’.jpg’];
saveas(gcf, file_name);
if X == Xprev2
break
else
Xprev2 = Xprev;
Xprev = X;
end
if nr_iter > max_nr_iter
disp(’EXIT - MAXIMUM NUMBER OF ITERATIONS EXCEEDED’)
break
end
end
N = close(0); % Huhis “lowMlevel” function generates a new plot much faster than using
% ‘plot’ function rather than the

A discussion of this code follows using the line numbers.
1–2: Initializes a “reasonably short” iteration sequence on my computer of 52 iterations — but this is not necessarily true on yours. (Modify the seed if necessary on your computer.)
3, 20–21, 38–39, 53: Creates a movie directly. The first line initializes the object M, the second captures the first frame and the third appends it to the object M. This is then repeated for every iteration. The last line is necessary for clearing the buffers and closing the file life_movie.avi.
Note: Lines 22–23 and 40–41 are not needed to create the movie using this method. Note: The movie is compressed by default in Microsoft Windows. See the parameter name compression in the documentation and try the various codecs to see which is “best.” (You can even supply your own.) Warning: The movie cannot exceed 2GB.
4. The size of the grid is $m \times m$.
5, 48–51: The maximum number of iterations is needed since with certain initial conditions the game will never end.
6: Creates a $m \times m$ sparse matrix with all zero elements. (See Section 9.)
7–12: Calculates the initial configuration.
13, 33: Determines the nonzero elements of X.
14: If there is a current figure, this makes it visible both as a figure and on the computer screen.
15, 34: The current iteration number.
16–18: Plots the initial configuration, its color, and the size of the dots; puts a title on it; and sets the limits on the axes. The plot is given the handle plothandle.
19, 37: Updates the current figure since if a number of plots are drawn in immediate succession only the last is made visible.
22–23, 39–41: Shows an indirect way to make a movie. Each frame is given a name with a unique number identifying the iteration number. It is immediately saved to disk. After this MATLAB function has ended, all these .jpg files can be combined into a movie using various software packages. SAY MORE???
Note: The second argument in num2str is essential so that the files appear in the correct order in the directory.
Note: Lines 3, 20–21, 38–39, and 54 are not needed.
24–25, 45–46: Saves the previous iterate and its previous iterate in order to determine if the game has settled down.
26–29: Initializes the variables needed to update the iterates.
30, 52: Calculates and plots each successive iteration.
31–32: Generates the next iteration.
35–37: Plots the next iterate using the set function rather than the plot function, puts a title on it, and makes it visible. (This “low-level” function generates a new plot much faster than using the plot function directly but, almost always, the plot function is “fast enough.”)
42–47: Compares the current and the second previous iteration. If they are identical, the iterations have settled down and the run ends.
There is another slightly different way to make a movie. Lines 3, 21–23, and 39–41 are deleted. Lines 20 and 38 are modified to
F(nr_iter+1) = getframe(gcf);
and line 53 is modified to
movie2avi(F,’life_movie.avi’)
This movie is also compressed in Microsoft Windows. Incidentally, you can view this movie in MATLAB by using the movie function.

avifile
Create a new avi file.
addframe
Add a frame to the avi file.
getframe
Get the current frame.
close(’avifile’)
Close the file opened with avifile.
movie
Play movie frames.
movie2avi
Save the current movie frames to an avi file.
saveas
Save a figure to disk.

Movie
4.7. Be Able To Do

After reading this section you should be able to do the following exercises. The solutions are given on page 163.

1. Plot $e^x$ and one of its Taylor series approximations.
   (a) Begin by plotting $e^x$ for $x \in [-1,1]$.
   (b) Then plot
   \begin{equation*}
   p_n(x) = 1 + \frac{x^3}{1!} + \frac{x^5}{3!} + \frac{x^7}{5!}
   \end{equation*}
on the same graph.
   (c) Also plot the difference between $e^x$ and this cubic polynomial, i.e., $e^x - p_2(x)$ on the same graph.
   (d) Next, generate a new graph containing all three curves by using only one plot function, force the axes to be to the same scale, and let all three curves have different colors. Put labels on the $x$ and $y$ axes and a title on the entire plot.
   (e) The above plot is not very instructive because $e^x - p_2(x)$ is much smaller than either $e^x$ or $p_2(x)$. Instead, use two plots. The first plot contains $e^x$ and $p_2(x)$ and the second plot, which is immediately below the first, contains $e^x - p_2(x)$. These two plots should fill an entire sheet of paper.

2. Consider the function
   \begin{equation*}
   f(x,y) = (x^2 + 4y^3)\sin(2\pi x)\sin(2\pi y).
   \end{equation*}
   (a) Plot this function for $x,y \in [-2,2].$
   Note: Make sure you use the “*” operator in front of each sine term. What does the surface look like if you don’t?
   (b) This surface has high peaks which interfere with your view of the surface. Change your viewpoint so you are looking down at the surface at such an angle that the peaks do not block your view of the central valley.
   Note: There are an infinite number of answers to this part.

3. Explain how the local function running_gaussians works. In particular, explain how the sum of all the modes is calculated in the function $g$.

5. Solving Linear Systems of Equations

One of the basic uses of MATLAB is to solve the linear system

\begin{equation*}
\begin{pmatrix}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\
\vdots \\
a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m
\end{pmatrix}
\end{equation*}
or the equivalent matrix equation

\begin{equation*}
Ax = b.
\end{equation*}

Note that there are $m$ equations in $n$ unknowns so that there may be zero solutions to this linear system, one solution, or an infinite number of solutions. We will discuss the case where $m \neq n$ in detail in Section 5.3. Here we concentrate on $m = n$.

5.1. Square Linear Systems

As we discussed previously, when $m = n$ the MATLAB operation
\begin{equation*}
x = \text{linsolve}(A, b)
\end{equation*}
calculates the unique solution $x$ by Gaussian elimination when $A$ is nonsingular. We can also solve it by

\begin{equation*}
x = \text{linsolve}(A, b)
\end{equation*}

The advantage of using \text{linsolve} is that it can be much faster when $A$ has a particular property. The third argument to \text{linsolve} gives the particular property. For our purposes the most important properties are lower triangular, upper triangular, symmetric, and positive definite. Enter

\begin{equation*}
x = \text{linspace}(A, b, \text{prop})
\end{equation*}

where \text{prop} is a \text{logical} structure with the following elements:

- \text{LT} – the matrix is lower triangular,
- \text{UT} – the matrix is upper triangular,
- \text{SYM} – the matrix is symmetric triangular, and
- \text{POSDEF} – the matrix is positive definite.

Note:

- \text{Aug} this function for $x,y \in [-2,2].$
- \text{Aug} this function for $x,y \in [-2,2].$
- \text{Aug} this function for $x,y \in [-2,2].$
- \text{Aug} this function for $x,y \in [-2,2].$

We discuss logical variables in Section 5.1. If $A$ has one (or more) of these properties, it can be solved much faster than using $x = Ab$.

When $A$ is singular there are either zero solutions or an infinite number of solutions to this equation and a different approach is needed. The appropriate MATLAB function is now \text{rref}. It begins by applying Gaussian elimination to the linear system of equations. However, it doesn’t stop there; it continues until it has zeroed out all the elements it can, both above and the main diagonal as well as below it. When done, the linear system is in \text{reduced row echelon form}:

- \text{The first nonzero coefficient in each linear equation is a 1 (but a linear equation can be simply $0 = 0$, in which case it has no nonzero coefficient).}
- \text{The first nonzero term in a particular linear equation occurs later than in any previous equation.}
- \text{The linear system must be written in \text{augmented matrix form}, i.e.,}$

\begin{equation*}
\begin{pmatrix}
x_1 & x_2 & \cdots & x_n & -rhs \\
a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\
a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
a_{m1} & a_{m2} & \cdots & a_{mn} & b_m
\end{pmatrix}
\end{equation*}

\text{where} $A$ is a \text{matrix} and $b$ is a \text{vector}.

\text{Warning: It is very important to understand that an augmented matrix is \text{not} a matrix (because the operations we apply to augmented matrices are not the operations we apply to matrices). It is simply a linear system of equations written in shorthand: the first column is the coefficients of the $x_1$ term, the second column is the coefficients of the $x_2$ term, etc., and the last column is the coefficients on the right-hand side. The vertical line between the last two columns represents the equal sign. Normally, an augmented matrix is written without explicitly writing the header information; however, the vertical line representing the equal sign should be included to explicitly indicate that this is an augmented matrix.}\n
\text{rref} operates on this augmented matrix to make as many of the elements as possible zero by using allowed operations on linear equations — these operations are not allowed on matrices, but only on linear systems of equations. The result is an augmented matrix which, when written back out as a linear system of equations, is particularly easy to solve. For example, consider the system of equations

\begin{equation*}
\begin{pmatrix}
x_1 + 2x_2 + 3x_3 &=& -1 \\
4x_1 + 5x_2 + 6x_3 &=& -1 \\
7x_1 + 8x_2 + 10x_3 &=& 0
\end{pmatrix}
\end{equation*}

which is equivalent to the matrix equation
\begin{equation*}
Ax = b \text{ where }
\begin{pmatrix}
1 & 2 & 3 \\
4 & 5 & 6 \\
7 & 8 & 10
\end{pmatrix}
\begin{pmatrix}
x_1 \\
x_2 \\
x_3
\end{pmatrix}
= \begin{pmatrix}
-1 \\
-1 \\
0
\end{pmatrix}
\end{equation*}

\text{Warning: It is very important to understand that an augmented matrix is \text{not} a matrix (because the operations we apply to augmented matrices are not the operations we apply to matrices). It is simply a linear system of equations written in shorthand: the first column is the coefficients of the $x_1$ term, the second column is the coefficients of the $x_2$ term, etc., and the last column is the coefficients on the right-hand side. The vertical line between the last two columns represents the equal sign. Normally, an augmented matrix is written without explicitly writing the header information; however, the vertical line representing the equal sign should be included to explicitly indicate that this is an augmented matrix.}\n
\text{rref} operates on this augmented matrix to make as many of the elements as possible zero by using allowed operations on linear equations — these operations are not allowed on matrices, but only on linear systems of equations. The result is an augmented matrix which, when written back out as a linear system of equations, is particularly easy to solve. For example, consider the system of equations

\begin{equation*}
\begin{pmatrix}
x_1 + 2x_2 + 3x_3 &=& -1 \\
4x_1 + 5x_2 + 6x_3 &=& -1 \\
7x_1 + 8x_2 + 10x_3 &=& 0
\end{pmatrix}
\end{equation*}

which is equivalent to the matrix equation
\begin{equation*}
Ax = b \text{ where }
\begin{pmatrix}
1 & 2 & 3 \\
4 & 5 & 6 \\
7 & 8 & 10
\end{pmatrix}
\begin{pmatrix}
x_1 \\
x_2 \\
x_3
\end{pmatrix}
= \begin{pmatrix}
-1 \\
-1 \\
0
\end{pmatrix}
\end{equation*}
The augmented matrix for this linear system is
\[
\begin{pmatrix}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & -1
\end{pmatrix}
\]

(We have included the header information for the last time.) Entering
\[
\text{rref}([A\ b])
\]
returns the augmented matrix
\[
\begin{pmatrix}
1 & 0 & 0 & 2
\end{pmatrix}
\]

Clearly, the solution of the linear system is \(x_1 = 2, x_2 = -3,\) and \(x_3 = 1.\)

Of course, you could just as easily have found the solution by
\[
x = A\ b
\]
so let us now consider the slightly different linear system
\[
egin{align*}
x_1 + 2x_2 + 3x_3 &= -1 \\
x_2 + 5x_3 &= -1 \\
x_1 + 6x_2 + 9x_3 &= -1.
\end{align*}
\]

This is equivalent to the matrix equation \(Ax = b\) where
\[
A = \begin{pmatrix} 1 & 2 & 3 \end{pmatrix}
\quad \text{and} \quad b = \begin{pmatrix} -1 \end{pmatrix}.
\]

Since \(A\) is a singular matrix, the linear system has either no solutions or an infinite number of solutions.

The augmented matrix for this linear system is
\[
\begin{pmatrix}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & -1
\end{pmatrix}
\]

Entering
\[
\text{rref}([A\ b])
\]
returns the augmented matrix
\[
\begin{pmatrix}
1 & 0 & -1 & 1
\end{pmatrix}
\]

so the solution of the linear system is \(x_1 = 1 + x_3\) and \(x_2 = -1 - 2x_3\) for any \(x_3 \in \mathbb{R}\) (or \(C\) if desired).

In vector form, the solution is
\[
\begin{pmatrix}
x_1 \\
x_2 \\
x_3
\end{pmatrix} = \begin{pmatrix}
1 + x_3 \\
-1 - 2x_3 \\
0
\end{pmatrix} = \begin{pmatrix}
x_3 \\
-2x_3 \\
0
\end{pmatrix} = \begin{pmatrix}
1 \\
0 \\
1
\end{pmatrix} + x_3 \begin{pmatrix}
1 \\
-2 \\
1
\end{pmatrix}.
\]

Suppose you modify the matrix equation slightly by letting \(b = (-1, -1, 0)^T.\) Now entering
\[
\text{rref}([A\ b])
\]
results in the augmented matrix
\[
\begin{pmatrix}
1 & 0 & -1 & 1
\end{pmatrix}
\]

Since the third equation is \(0 = 1,\) there is clearly no solution to the linear system.

Warning: The function \texttt{rref} does not always give correct results. For example, if
\[
C = \begin{pmatrix}
0.95 & 0.03 \\
0.05 & 0.97
\end{pmatrix}
\]

then the matrix \(I - C\) is singular (where \(I\) is the identity matrix). However, if you solve
\[
(I - C)x = 0
\]
by
\[
\text{rref}([\text{eye(size(C))}-C\ [0\ 0])
\]
MATLAB displays
\[
\begin{pmatrix}
1 & 0 & 0 \\
0 & 1 & 0
\end{pmatrix}
\]

which indicates that the only solution is \(x = 0.\) On the other hand, if you enter
\[
\text{rref}([\text{eye(size(C))}-C\ [b\ 0])
\]
then MATLAB realizes that \(I - C\) is singular. Clearly there is some value of \(b\) between 0 and 1 where MATLAB switches between believing that \(I - C\) is non-singular and singular.\footnote{To understand this “switch”, look at the actual coding of \texttt{rref}. It uses the variable \texttt{tol} to determine whether an element of the augmented matrix is “small enough” that it should be set to 0. \texttt{tol} is (essentially) calculated by}\[
\text{tol} = \text{max}([\text{size(singular matrix)}] + \epsilon) * \text{norm(singular matrix)} * \text{inf}.
\]
The maximum of the number of rows and columns of the augmented matrix, \(i.e., \text{max}([\text{size}])\), is multiplied by \(\epsilon\) and this is multiplied by the “size” of the augmented matrix. \(\text{norm}([\text{size}])\) depends on the size of the elements of \(b.\) Thus, the determination whether or not \(b\) should be set to 0 depends on the magnitude of the elements of \(b.\)

You can obtain the correct answer to the homogeneous equation by entering
\[
\text{rref}([\text{eye(size(C))}-C\ [0\ 0])
\]
which decreases the tolerance to \(\epsilon.\)

5.2. Catastrophic Round-Off Errors

We have mentioned repeatedly that \texttt{computers cannot add, subtract, multiply, or divide correctly!} Up until now, the errors that have resulted have been very small. Now we present two examples where the errors are very large.

In this first example, the reason for the large errors is easy to understand. Consider the matrix
\[
A = \begin{pmatrix} 1 & 2 & 3 \\
4 & 5 & 6 \\
7 & 8 & 9
\end{pmatrix}.
\]

which is singular when \(c = 0\) and nonsingular otherwise. But how well does MATLAB do when \(c = 1?\)

Entering
\[
\text{rref}([A\ b])
\]
returns the augmented matrix
\[
\begin{pmatrix}
1 & 0 & -1 & 1
\end{pmatrix}
\]

so that the final matrix should be \(0.\) (Do not use \(\text{eps}\) for the name of this variable or you will change the predefined variable \(\text{eps}\). Other possible names are \(\text{ep}\) or \(\text{equil}.)\) Begin by letting \(r = 0\) and

\footnote{To understand this “switch”, look at the actual coding of \texttt{rref}. It uses the variable \texttt{tol} to determine whether an element of the augmented matrix is “small enough” that it should be set to 0. \texttt{tol} is (essentially) calculated by}\[
\text{tol} = \text{max}([\text{size(singular matrix)}] + \epsilon) * \text{norm(singular matrix)} * \text{inf}.
\]
The maximum of the number of rows and columns of the augmented matrix, \(i.e., \text{max}([\text{size}])\), is multiplied by \(\epsilon\) and this is multiplied by the “size” of the augmented matrix. \(\text{norm}([\text{size}])\) depends on the size of the elements of \(b.\) Thus, the determination whether or not \(b\) should be set to 0 depends on the magnitude of the elements of \(b.\)

You can obtain the correct answer to the homogeneous equation by entering
\[
\text{rref}([\text{eye(size(C))}-C\ [0\ 0])
\]
which decreases the tolerance to \(\epsilon.\)
observe that the result displayed is nowhere close to the zero matrix! However, note that MATLAB is warning you that it thinks something is wrong with the statement

Warning: Matrix is close to singular or badly scaled.

Results may be inaccurate. RCOND = 1.641976e-18.

(\textbf{RCOND} is its estimate of the inverse of the condition number. See cond in Section 7 for more details.)

Now choose some small nonzero values for \( \epsilon \) and see what happens. How small can \( \epsilon \) be before MATLAB warns you that the matrix is "close to singular or badly scaled?" In this example, you know that the matrix is "close to singular" if \( \epsilon \) is small (but nonzero) even if MATLAB does not. The next example is more interesting.

For the second example, consider the Hilbert matrix of order \( n \), i.e.,

\[
H_n = \begin{pmatrix}
1 & 1/2 & 1/3 & \cdots & 1/n \\
1/2 & 1/3 & 1/4 & \cdots & 1/(n+1) \\
1/3 & 1/4 & 1/5 & \cdots & 1/(n+2) \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1/n & 1/(n+1) & 1/(n+2) & \cdots & 1/(2n-1)
\end{pmatrix}
\]

which is generated in MATLAB by

\[
H = \text{hilb}(n)
\]

Here does not seem to be anything particularly interesting, or strange, about this matrix; after all, \( h_{ij} = 1/(i+j-1) \) so the elements are all of "reasonable" size. If you type

\[
\begin{align*}
\text{m} &= 10; \quad H = \text{hilb}(m); \quad (H^\text{\textdagger})(1/2) \rightarrow 2 - H \\
\end{align*}
\]

the result is not particularly surprising. (Recall that \( H^\text{\textdagger} \) is very different from \( H^-1 \).) The resulting matrix should be the zero matrix, but, because of round-off errors, it is not. However, every element is in magnitude less than \( 10^{-15} \), so everything looks fine.

However, suppose you solve the matrix equation

\[
Rx = b
\]

for a given \( b \). How close is the numerical solution to the exact solution? Of course, the problem is: how can you know what the analytical solution is for a given \( b \)? The answer is to begin with \( x \) and calculate \( b \) by \( b = Rx \). Then solve \( Rx = b \) for \( x \) and compare the final and initial values of \( x \). Do this in MATLAB by

\[
\begin{align*}
\text{m} &= 10; \quad x = \text{rand}(m, 1); b = H \cdot x; \quad \text{xnum} = H \cdot b \\
\text{x} &= \text{norm}(x - \text{xnum})
\end{align*}
\]

The result is not very satisfactory: the maximum difference in the elements of the two vectors is usually somewhere between 10^{-5} and 10^{-3}. That is, even though all the calculations have been done to approximately 16 significant digits, the result is only accurate to three to five significant digits! (To see how much worse the result can be, repeat the above statements for \( m = 12 \).)

It is important to realize that most calculations in MATLAB are very accurate. It is not that solving a matrix equation necessarily introduces lots of round-off errors; instead, Hilbert matrices are very "unstable" matrices — working with them can lead to inaccurate results. On the other hand, most matrices are quite "stable." For example, if you repeat the above sequence of steps with a random matrix, you find that the results are quite accurate. For example, enter

\[
\begin{align*}
\text{m} &= 1000; \quad H = \text{rand}(m, 1); b = H \cdot x; \quad \text{xnum} = H \cdot b; \quad \text{x} = \text{norm}(x - \text{xnum})
\end{align*}
\]

The results are much more reassuring, even though \( m \) is 100 times as large for this random matrix as for the Hilbert matrix — and even though there are over 600,000 times as many floating point operations needed to calculate \( x \) by Gaussian elimination for this random matrix!

Note: By entering all the statements on one line, it is easy to repeat this experiment many times for different random numbers by simply resubmitting this one line.

5.3. Overdetermined and Underdetermined Linear Systems

If \( A \in \mathbb{C}^{m \times n} \) where \( m > n \), \( Ax = b \) is called an overdetermined system because there are more equations than unknowns. In general, there are no solutions to this linear equation. However, you can find a "best" approximation by finding the solution for which the vector

\[
\| Ax - b \|_2
\]

is minimized. (The \texttt{norm} function is discussed in Sections 2.8 and 7.) This is called the least-squares solution.

Algebraically, the approximation can be calculated by solving

\[
A^T A x = A^T b.
\]

which is called the residual, is smallest in Euclidean length; that is,

\[
\| Ax - b \|_2 \equiv \left( \sum_{i=1}^{n} (Ax)_i - b_i \right)^{1/2}
\]

is minimized. The \texttt{norm} function is discussed in Sections 2.8 and 7. This is called the least-squares solution. This best approximation is calculated in MATLAB by typing

\[
Ax = \text{pinv}(A)
\]

However, numerically this is less accurate than the method used in MATLAB.

Note that this is the same operator used to find the solution to a square linear system. This cannot be the intent here since \( A \) is not a square matrix. Instead, MATLAB interprets this operator as asking for the least-squares solution. Again, this operator only makes sense if there is a unique solution which minimizes the length of the vector \( Ax - b \). If there are an infinite number of least-squares solutions, MATLAB warns you of this fact and then returns one of the solutions. For example, if

\[
A = \begin{pmatrix}
1 & 2 & 3 \\
4 & 5 & 6 \\
7 & 8 & 9 \\
10 & 11 & 12
\end{pmatrix}
\]

and \( b = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix} \),

then \( Ax = b \) has no solutions, but has an infinite number of least-squares approximations. If you enter

\[
Ax = \text{pinv}(A)
\]

the response is

\[
\text{warning: rank deficient, rank = 2 tois = 1.4594e-14.}
\]

It also returns the solution \((-1/3, 0, 0/\text{tois})^T\) (after using the MATLAB function \texttt{rats} which we discuss below), which is one particular least-squares approximation. (If \( A \) is complex, solve \( A^H b \).

Occasionally, if there are an infinite number of least-squares approximations, the solution desired is the "smallest" one, i.e., the \( x \) for which the length of the vector \( Ax - b \) is minimized. This can be calculated using the \texttt{pseudoinverse of} \( A \), also called the Moore-Penrose pseudoinverse, which is denoted by \( A^+ \). Since \( A \) is not square, it cannot have an inverse. However, the pseudoinverse is the unique \( m \times n \) matrix which satisfies the Moore-Penrose conditions:

\cdot \quad A^+ A = A

\cdot \quad A A^+ = A^+

\cdot \quad (A^+)^T = A^+

\cdot \quad (A A^+)^T = A A^+

In particular, if \( A \) is a square nonsingular matrix, then \( A^+ = A^{-1} \). This pseudoinverse is calculated in MATLAB by entering

\[
\text{pinv}(A)
\]

The reason for mentioning the pseudoinverse of \( A \) is that the least-squares approximation to \( Ax = b \) can also be calculated by

\[
\text{pinv}(A) b
\]

If there are an infinite number of least-squares approximations, this returns the one with the smallest length. (In the previous example it is \((-13/45, 7/90, 4/9)^T\).

Next, suppose that \( A \in \mathbb{C}^{m \times n} \) with \( m < n \). \( Ax = b \) is called an underdetermined system because there are less equations than unknowns. In general, there are an infinite number of solutions to this equation. We can find one particular solution by entering

\[
Ax = b
\]

This solution will have many of its elements being 0. We can also find the solution with the smallest length by entering

\[
\text{pinv}(A) b
\]

Warning: It is possible for an overdetermined system to have one or even an infinite number of solutions (not least-squares approximations). It is also possible for an underdetermined system to have no solutions.

One function which is occasionally useful is \texttt{rats}. If all the elements of \( A \) and \( b \) are rational numbers, then the solution and/or approximation obtained is usually a rational number, although stored as
a floating-point number. This function displays a “close” rational approximation to the floating-point number, which may or may not be the exact answer. For example, entering

```matlab
rats(1/3 - 1/17 * 1/5)
```

results in the text variable 121/285, which is the correct answer.

Warning: Be careful when using this function. `rats(sqrt(2))` makes no sense (as was known to Pythagoras).

### Solving Linear Systems

\[ A \backslash b \]

When \( A = b \) is an over-determined system, i.e., \( m > n \) where \( A \in \mathbb{C}^{m \times n} \), there are usually no solutions, and this calculates the least-squares approximation; when \( A \) is an under-determined system, i.e., \( m < n \), there are usually an infinite number of solutions, and this calculates the solution with the maximal number of zeros.

\texttt{pinv(A)} Calculates the pseudoinverse of \( A \).

\texttt{rata(x)} Calculates a “close” approximation to the floating-point number \( x \). This is frequently the exact value.

### 6. File Input-Output

In Section 4.1 we discussed the \texttt{csvread} and \texttt{csvwrite} function which allow simple input from and output to a file. The MATLAB functions \texttt{fscanf} and \texttt{fprintf}, which behave very similarly to their C counterparts, allow much finer control over input and output. Thus, using them a file has to be opened by

```matlab
fid = fopen(’<filename>’, ’permission strings’)
```

where the file identifier \( fid \) is a unique non-negative integer attached to the file. (Three file identifiers always exist as in C: 0 is the standard input, 1 is the standard output, and 2 is the standard error.) The permission string specifies how the file is to be accessed:

- ’r’ read only from the file.
- ’w’ write only to the file (anything previously contained in the file is overwritten).
- ’a’ append to the end of the file (everything previously contained in the file is retained).
- ’r+’ read from and write to the file (anything previously contained in the file is overwritten).
- ’w+’ write from and write to the file (anything previously contained in the file is overwritten). 

If the \texttt{fopen} function fails, \( -1 \) is returned in the file identifier. Enter

```matlab
fclose(fid)
```

if a file needs to be closed.

To write formatted data to a file, enter

```matlab
fprintf(fid, ’format strings’, variable 1, variable 2, ...)
```

The elements contained in the variables are written to the file specified in a previous \texttt{fopen} function according to the format string. These variables are printed out in order and matrices are converted to column vectors (i.e., \( \mathbf{A} \rightarrow \mathbf{A}^\top \)). If \( \texttt{fprintf} \) is omitted, the output appears on the screen. The format string is very similar to that of C, with the exception that the format string is cycled through until the end of the file is reached or the number of elements specified by \texttt{size} is attained.

To briefly review some of the C format specifications, the conversion characters are:

- \( \texttt{d} \) The argument is converted to decimal notation, and, if possible, to integer notation.
- \( \texttt{c} \) The argument is a single character.
- \( \texttt{s} \) The argument is a string.
- \( \texttt{e} \) The argument is a floating-point number in ’E’ format.
- \( \texttt{f} \) The argument is a floating-point number in decimal notation.
- \( \texttt{g} \) The argument is a floating-point number in either ’E’ or decimal notation.

Each conversion character is preceded by ’\%’. The following may appear between the ’\%’ and the conversion character:

- A minus sign which specifies left adjustment rather than right adjustment.
- An integer which specifies a minimum field width.
- If the maximum field width is larger than the minimum field width, the minimum field width is preceded by an integer which specifies the maximum field width, and the two integers are separated by a period.

\texttt{fprintf} can also be used to format data on the screen by omitting the \( \texttt{fid} \) at the beginning of the argument list. Thus, it is possible to display a variable using as little or as much control as desired. For example, if \( x \) contains \(-23.6\) three different ways to display it are

```matlab
x = -23.6000
```

```matlab
type(x)
```

```matlab
v = num2str(x)
```

\texttt{fprintf} and \texttt{disp} are equivalent in their effects.

### Advanced Input-Output

As in C, the \texttt{fopen} function returns the file identifier, which is a unique non-negative integer attached to the file. The size argument, which puts an upper limit on the amount of data to be read, is optional. If it is a scalar, or is not used at all, \( A \) is actually a vector. If it is \([m, n] \), then \( A \) is a matrix of this size.

```matlab
fopen(’<filename>’, ’permission strings’)
```

Opens the file with the permission string determining how the file is to be accessed. The function returns the file identifier, which is a unique non-negative integer attached to the file.

```matlab
fclose(fid)
```

Closes the file with the given file identifier.

```matlab
fprintf(fid, ’format strings’, variable 1, variable 2, ...)
```

Behaves very similarly to the \texttt{C} command in writing data from a file using any desired format.

```matlab
fscanf(fid, ’format strings’, variable 1, variable 2, ...)
```

Behaves very similarly to the \texttt{C} command in reading data to a file using any desired format.

```matlab
csvread(’<filename>’)
```

Reads from the file specified in a previous \texttt{fopen} function according to the format string and puts the data into the matrix \( A \). The size argument, which sets an upper limit on the amount of data to be read, is optional. If it is a scalar, or is not used at all, \( A \) is actually a vector. If it is \([m, n] \), then \( A \) is a matrix of this size.

```matlab
csvwrite(’<filename>’, A)
```

Write formatted data to a file.

### 7. Some Useful Linear Algebra Functions

We briefly describe in alphabetical order some of the MATLAB functions that are most useful in linear algebra. Most of these discussions can be read independently of the others. Where this is true, we indicate which should be read first.

**Note:** A few of these functions can only be applied to full matrices, and others only to sparse matrices.

There is a mathematical “definition” for these terms, but in MATLAB a matrix is full if it is created using the methods described in Section 2, while it is sparse if it is created using the methods...
\textbf{chol}

Let $A \in \mathbb{R}^{n \times n}$ be symmetric and positive definite. Then there exists an upper triangular matrix $R$ such that $R^T R = A$. $R$ is calculated by

$R = \text{chol}(A)$

If $A$ is not positive definite, an error message is printed. (If $A \in \mathbb{C}^{n \times n}$ then $R^T R = A$.)

\textbf{cond}

\textbf{Note:} Read the discussion on \texttt{norm} below first.

The condition number of $A \in \mathbb{C}^{m \times n}$, which is denoted by $\text{cond}(A)$, is a positive real number which is always $\geq 1$. It measures how "stable" $A$ is: if $\text{cond}(A) \rightarrow \infty$ the matrix is singular, while if $\text{cond}(A) = 1$ the matrix is as nice a matrix as you could hope for — in particular, $\text{cond}(I) = 1$. To estimate the number of digits of accuracy you might lose in solving the linear system $Ax = b$, enter

\texttt{log10(\text{cond}(A))}

In Section 5.2 we discussed the number of digits of accuracy you might lose in solving, $R \mathbf{x} = \mathbf{b}$ where $R$ is the Hilbert matrix of order 10. In doing many calculations it was clear that the solution was only accurate to 3 or 5 significant digits. Since $\text{cond}(R)$ is $1.6 \times 10^{13}$, it is clear that you should lose about 13 of the 16 digits of accuracy in this calculation. Thus, everything fits.

If $A$ is nonsingular, the condition number is defined by

\[ \text{cond}_p(A) = ||A||_p ||A^{-1}||_p \quad \text{for } p \in [1, \infty] \]

\[ \text{cond}_2(A) = ||A||_2 ||A^{-1}||_2 \]

It is calculated in MATLAB by

\texttt{cond(A, p)}

where $p = 1$, 2, \texttt{Inf}, or \texttt{'fro'}. If $p = 2$ the function can be shortened to

\texttt{cond(A)}

Note that the calculation of the condition number of $A$ requires the calculation of the inverse of $A$. The MATLAB function \texttt{condest} approximates the condition number without having to calculate this inverse. See the discussion of this function below for further information on when it might be preferable.

\textbf{Note:} Sometimes we want to solve, or find the "best" approximation to, $A \mathbf{x} = \mathbf{b}$ where $A \in \mathbb{C}^{m \times n}$ is not a square matrix. (This is discussed in detail in Section 3.3.) Since we still want to know the accuracy of any solution, we want to generalize the condition number to nonsquare matrices. This is done by defining the condition number of a nonsquare matrix in the 2-norm to be the ratio of the largest to the smallest singular value of $A$, i.e., $\sigma_1/\sigma_{\text{min}(m,n)}$.

\textbf{condest}

\textbf{Note:} Read the discussion on \texttt{cond} above first.

The calculation of the condition number of $A \in \mathbb{C}^{m \times n}$ requires the calculation of its inverse. There are two reasons this might be inadvisable:

- The calculation of $A^{-1}$ requires approximately $2n^3$ flops, which might take too long if $n$ is very large.
- If $A$ is a sparse matrix (i.e., most of its elements are zero), we discuss in Section 9 how to store only the nonzero elements of $A$ to conserve storage. (For example, if $n = 10,000$ and $A$ is tridiagonal, the number of nonzero elements in $A$ is approximately 30,000 but the total number of elements in $A$ is 100,000,000.) Since the inverse of a sparse matrix is generally much less sparse (in fact it may have no zero elements at all), MATLAB may not be able to store $A^{-1}$.

The function \texttt{condest} calculates a lower bound to the condition number of a matrix in the 1-norm without having to determine its inverse. This approximation is almost always within a factor of ten of the exact value.

When MATLAB calculates $A \mathbf{b}$ or $\text{inv}(A)$, it also calculates $\text{condest}(A)$. It checks if its estimate of the condition number is large enough that $A$ is likely to be singular. If so, it returns an error message such as

\texttt{Warning: Matrix is close to singular or badly scaled. Results may be inaccurate. RCOND = 2.055969e-18.}

where $\text{RCOND}$ is the inverse of $\text{condest}(A)$.

\textbf{det}

Let $A \in \mathbb{C}^{n \times n}$. The determinant of $A$ is calculated by

\texttt{det(A)} = 0 if and only if $A$ is singular. However, due to round-off errors it is very unlikely that you will obtain 0 numerically unless all the entries to $A$ are integers. For example, consider the matrix

\[ C = \begin{bmatrix} 0.95 & 0.03 \\ 0.05 & 0.97 \end{bmatrix} \]

$1 - C$ is singular (where $1$ is the identity matrix) but

\[ \text{det}(A) = 0.01 \]

This is due to round-off errors. However, the number it returns is much smaller than eps and so it seems "reasonable" that $1 - C$ is singular. On the other hand,

\[ \text{det(hilb(10))} \]

returns $2.3 \times 10^{-35}$ but the Hilbert matrix is not singular for any $n$. Similarly,

\[ \text{det(0.1*eye(1000))} \]

returns $10^{-100}$, but it is also not singular. (The singular value decomposition, which is described below, is a much better method for determining if a square matrix is singular.)

\textbf{eig}

Let $A \in \mathbb{C}^{n \times n}$. A scalar $\lambda \in \mathbb{C}$ is an eigenvalue of $A$ if there exists a nonzero vector $\mathbf{v} \in \mathbb{C}^n$ such that

$A \mathbf{v} = \lambda \mathbf{v}$

$\mathbf{v}$ is called the eigenvector corresponding to $\lambda$. There are always $n$ eigenvalues of $A$, although they need not all be distinct. MATLAB will very happily calculate all the eigenvalues of $A$ by

\texttt{eig(A)}

It will also calculate all the eigenvectors by

\texttt{[Y, D] = eig(A)}

$D \in \mathbb{C}^{n \times n}$ is a diagonal matrix containing the $n$ eigenvalues on its diagonal and the corresponding eigenvectors are found in the same columns of the matrix $Y \in \mathbb{C}^{n \times n}$.

\textbf{Note:} This is the first time we have had a function return more than one argument. We discuss this notation in detail in Section 8.3. For now, we simply state that when $[Y, D]$ occurs on the right side of the equal sign it means the matrix whose first column comes from $Y$ and whose last columns come from $D$. However, on the left side of the equal sign it means that the function returns two arguments where the first is stored in the variable $Y$ and the second in $D$.

\texttt{eig} can also calculate all the eigenvalues of the generalized eigenvalue problem

$A \mathbf{v} = \lambda B \mathbf{v}$

by

\texttt{eig(A, B)}

A matrix is defective if it has less eigenvectors than eigenvalues. MATLAB normally cannot determine when this occurs. For example, the matrix

$B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$

\[ A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \]
is defective since it has two eigenvalues, both of which are 1, but only has one eigenvector, namely $(1.0)^T$. If you enter

```matlab
B = [1 1; 0 1]; [Y, D] = eig(B)
```
MATLAB calculates the two eigenvalues correctly, but it finds the two eigenvectors $(1.0)^T$ and $(-1.2220 \times 10^{-16})^T$. Clearly the latter eigenvector should be $(-1.0)^T$ so that, in fact, there is only one eigenvector.

**Note:** If $A$ is a sparse matrix, you cannot use `eig`. You either have to use the function `eigs` or do `eig(full(A))`.

`eigs`

**Note:** Read the discussion on `eig` above first.

Frequently, you do not need all the eigenvalues of a matrix. For example, you might only need the largest ten in magnitude, or the five with the largest real part, or the one which is smallest in magnitude, or ... Or you might only need a few of the generalized eigenvalues of $AX = 

$x$. `eigs` can do all of this. Of course, this means that there are numerous possible arguments to this function so read the documentation carefully.

Why not just use `eig` anyway? Calculating all the eigenvalues of a nonsymmetric $A \in \mathbb{R}^{n \times n}$ requires (very) approximately $10n^3$ flops, which can take a very long time if $n$ is very large. On the other hand, calculating only a few eigenvalues requires many, many fewer flops. If $A$ is a full matrix, it requires $cn^2$ flops where $c$ is of “reasonable” size; if $A$ is a sparse matrix, it requires $cn$ flops.

**Note:** If $A$ is sparse, you cannot use `eig` — you will first have to do `eig(full(A))`.

Also, this function generates lots of diagnostic output. To calculate the largest 3 eigenvalues of $A$ in matrix without generating any diagnostics, enter

```matlab
>> op.disp = 0
>> eigs(A, 3, 'LM', op)
(op.disp is a structure, which was discussed in Section 3.4.)
```

### inv

To calculate the inverse of the square matrix $A \in \mathbb{C}^{n \times n}$ enter

```matlab
inv(A)
```

The inverse of $A$, denoted by $A^{-1}$, is a matrix such that $AA^{-1} = A^{-1}A = I$, where $I \in \mathbb{R}^{n \times n}$ is the identity matrix. If such a matrix exists, it must be unique.

MATLAB cannot always tell whether this matrix does, in fact, exist. For example, the matrix

\[
A = \begin{pmatrix}
1 & 2 \\
2 & 4
\end{pmatrix}
\]

does not have an inverse. If you try to take the inverse of this matrix, MATLAB will complain that

**Warning:** Matrix is singular to working precision.

It will display the inverse matrix, but all the entries will be Inf.

The above matrix was very simple. The matrix

\[
A = \begin{pmatrix}
1 & 2 & 3 \\
4 & 5 & 6 \\
7 & 8 & 9
\end{pmatrix}
\]

(7.1)

also does not have an inverse. If you ask MATLAB to calculate the inverse of $A$, it will complain that

**Warning:** Matrix is close to singular or badly scaled.

**Results may be inaccurate. RCOND = 2.055969e-16.**

(`RCOND` is the inverse of a numerical approximation to the condition number of $A$; see `condest` above.) That is, MATLAB is not positive that $A$ is singular, because of round-off errors, but it thinks it is likely. However, MATLAB still does try to calculate the inverse. Of course, if you multiply this matrix by $A$ the result is nowhere close to $I$. (Try it!) In other words, be careful — and read (and understand) all warning messages.

### lu

Let $A \in \mathbb{C}^{n \times n}$. Then there exists an upper triangular matrix $U$, a unit lower triangular matrix $L^{-1}$, and a permutation matrix $P$ such that

\[
LU = PA.
\]

The MATLAB function `lu` calculates these matrices by entering

```matlab
>> [L, U, P] = lu(A)
```

If $A$ is invertible, all the elements of $U$ on the main diagonal are nonzero. If you enter

```matlab
>> A = [1 2 3; 4 5 6; 7 8 9]; [L, U, P] = lu(A)
```

where $A$ is the singular matrix defined earlier, $w_{31}$ should be zero. Entering

```matlab
>> U(3,3)
```
displays $1.1102e-16$, which clearly should be zero as we discussed in Section 1.5.

### norm

The norm of a vector or matrix is a nonnegative real number which gives some measure of the “size” of the vector or matrix. (It was briefly discussed in Section 2.8.) The $p$th norm of a vector is defined by

\[
\|x\|_p = \begin{cases} \left( \sum_{i=1}^{n} |x_i|^p \right)^{1/p} & \text{if } p \in [1, \infty) \\ \max_{i \leq n} |x_i| & \text{if } p = \infty. \end{cases}
\]

For $p = 1, 2, \infty$ it is calculated in MATLAB by entering

```matlab
>> norm(x, p)
```

where $p = 1, 2, \infty$. If $p = 2$ the function can be shortened to

```matlab
>> norm(x)
```

The $p$th norm of a matrix is defined by

\[
\|A\|_p = \max_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p} \quad \text{for } p \in [1, \infty]
\]

and is calculated in MATLAB by entering

```matlab
>> norm(A, p)
```

where again $p = 1, 2, \infty$. If $p = 2$ the function can be shortened to

```matlab
>> norm(A)
```

There is another matrix norm, the Frobenius norm, which is defined for $A \in \mathbb{C}^{m \times n}$ by

\[
\|A\|_F = \left( \sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 \right)^{1/2}
\]

and is calculated in MATLAB by entering

```matlab
>> norm(A, 'Fro')
```

It is equivalent to `norm(A(:))`.

### null

Let $A \in \mathbb{C}^{n \times n}$. We can calculate an orthonormal basis for the null space of $A$ by

```matlab
>> null(A)
```

### orth

Let $A \in \mathbb{C}^{n \times n}$. We can calculate an orthonormal basis for the columns of $A$ by

```matlab
>> orth(A)
```

A unit lower triangular matrix is lower triangular and, in addition, all the elements on the main diagonal are 1.

$p$ is a permutation matrix if its columns are a rearrangement of the columns of $I$.
Let $A \in \mathbb{R}^{m \times n}$. Then there exists an orthogonal matrix $Q \in \mathbb{R}^{m \times m}$! and an upper triangular matrix $R \in \mathbb{R}^{r \times n}$ such that $A = QR$.

If $A \in \mathbb{C}^{m \times n}$ then there exists a unitary matrix $Q \in \mathbb{C}^{m \times m}$ and an upper triangular matrix $R \in \mathbb{C}^{r \times n}$ such that $A = QR$. We calculate $Q$ and $R$ in MATLAB by entering $[Q, R] = qr(A)$. It is frequently preferable to add the requirement that the diagonal elements of $R$ be decreasing in magnitude, i.e., $|r_{i,i+1}| \leq |r_{i,i}|$ for all $i$. In this case $AE = QR$ for some permutation matrix $E$ and $[Q, R, E] = qr(A, 0)$.

One reason for this additional requirement on $R$ is that you can immediately obtain an orthonormal basis for the range of $A$ and the null space of $A^T$. If $r_{ik}$ is the last nonzero diagonal element of $R$, then the first $k$ columns of $Q$ are an orthonormal basis for the range of $A$ and the final $n-k$ columns are an orthonormal basis for the null space of $A^T$. The function $orth$ is preferable if all you want is an orthonormal basis for $R(A)$.

There is another use for the $qr$ function. There exists a matrix $Q \in \mathbb{R}^{m \times n}$ with orthonormal columns and an upper triangular matrix $R \in \mathbb{R}^{r \times n}$ such that $A = QR$.

It is calculated by $[Q, R, E] = qr(A, 0)$.

This is equivalent to applying the Gram-Schmidt algorithm to $A$.

The Moore-Penrose pseudoinverse has already been discussed in Section 5.3. We include it here for completeness. It is calculated by using the singular value decomposition, which we discuss below.

### rank

Let $A \in \mathbb{C}^{m \times n}$. The rank of $A$ is the number of linearly independent columns of $A$ and is calculated by $\text{rank}(A)$.

This number is calculated by using the singular value decomposition, which we discuss below.

### svd

Let $A \in \mathbb{R}^{m \times n}$. $A$ can be decomposed into

$$A = U \Sigma V^T$$

where $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ are orthogonal matrices and $\Sigma \in \mathbb{R}^{m \times n}$ is a diagonal matrix (although not necessarily square) with real nonnegative elements in decreasing order. That is,

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min(m,n)} \geq 0.$$

If $A \in \mathbb{C}^{m \times n}$ then $U \in \mathbb{C}^{m \times m}$ and $V \in \mathbb{C}^{n \times n}$ are unitary matrices and $\Sigma \in \mathbb{R}^{m \times n}$ is again a diagonal matrix with real nonnegative elements in decreasing order. These matrices are calculated by

$$[U, \Sigma, V] = \text{svd}(A)$$

The diagonal elements of $\Sigma$ are called the singular values of $A$. Although $A$ need not be a square matrix, both $A^T A$ and $A A^T$ are square symmetric matrices. (If $A$ is complex, $A^T A$ and $A A^T$ are both square Hermitian matrices.) Thus, their eigenvalues are nonnegative.

Their nonzero eigenvalues are the squares of the singular values of $A$. In addition, the eigenvectors of $A^T A$ are the columns of $U$ and those of $A A^T$ are the columns of $V$. (If $A$ is complex, the eigenvectors of $A^T A$ are the columns of $V$ and those of $A A^T$ are the columns of $U$.)

The best numerical method to determine the rank of $A$ is to use its singular values. For example, to see that $A$ has rank 2, use the svd function to find that the singular values of $A$ are $25.4398, 1.7226$, and $8.1099 \times 10^{-16}$. Clearly the third singular value should be 0 and so $A$ has 2 nonzero singular values and so has a rank of 2. On the other hand, the Hilbert matrix of order 15 has singular values

$$1.8 \times 10^8, 4.3 \times 10^{-1}, 5.7 \times 10^{-2}, 5.6 \times 10^{-3}, 4.3 \times 10^{-4}, 2.7 \times 10^{-5}, 1.3 \times 10^{-6}, 5.5 \times 10^{-8}, 1.8 \times 10^{-9}, 4.7 \times 10^{-11}, 9.3 \times 10^{-13}, 1.4 \times 10^{-14}, 1.4 \times 10^{-16}, 1.2 \times 10^{-17}, 2.4 \times 10^{-18}$$

according to MATLAB. Following Principle 1.2, you can see there is no separation between the singular values which are clearly not zero and the ones which are “close to” zero. Thus, you cannot conclude that any of these singular values should be set to 0. Our “best guess” is that the rank of this matrix is 15.

Some Useful Functions in Linear Algebra

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>chol(A)</td>
<td>Calculates the Cholesky decomposition of a symmetric, positive definite square matrix.</td>
</tr>
<tr>
<td>cond(A)</td>
<td>Calculates the condition number of a square matrix.</td>
</tr>
<tr>
<td>cond(A, p)</td>
<td>Calculates the condition number in the $p$-norm.</td>
</tr>
<tr>
<td>condest(A)</td>
<td>Calculates a lower bound to the condition number of $A$ in the 1-norm.</td>
</tr>
<tr>
<td>det(A)</td>
<td>Calculates the determinant of a square matrix.</td>
</tr>
<tr>
<td>eig(A)</td>
<td>Calculates the eigenvalues, and eigenvectors if desired, of a square matrix.</td>
</tr>
<tr>
<td>eigs</td>
<td>Calculates some eigenvalues, and eigenvectors if desired, of a square matrix. There are numerous possible arguments to this function so read the documentation carefully.</td>
</tr>
<tr>
<td>inv(A)</td>
<td>Calculates the inverse of a square invertible matrix.</td>
</tr>
<tr>
<td>lu(A)</td>
<td>Calculates the LU decomposition of a square invertible matrix.</td>
</tr>
<tr>
<td>norm(v)</td>
<td>Calculates the norm of a vector.</td>
</tr>
<tr>
<td>norm(v, p)</td>
<td>Calculates the $p$-norm.</td>
</tr>
<tr>
<td>norm(A)</td>
<td>Calculates the norm of a matrix.</td>
</tr>
<tr>
<td>norm(A, p)</td>
<td>Calculates the $p$-norm.</td>
</tr>
<tr>
<td>null(A)</td>
<td>Calculates the null space of a matrix.</td>
</tr>
<tr>
<td>orth(A)</td>
<td>Calculates an orthonormal basis for the null space of a matrix.</td>
</tr>
<tr>
<td>rank(A)</td>
<td>Estimates the rank of a matrix.</td>
</tr>
<tr>
<td>svd(A)</td>
<td>Calculates the singular value decomposition of a matrix.</td>
</tr>
</tbody>
</table>

1. The eigenvalues of a real square symmetric matrix are nonnegative. (The eigenvalues of a complex square Hermitian matrix are real and nonnegative.)
2. For example, if $m > n$, there are $n$ singular values and their squares are the eigenvalues of $A^T A$. The $n$ eigenvalues of $A^T A$ consist of the squares of those $n$ singular values and $m-n$ additional zero eigenvalues.
3. In fact, it can be proven that the Hilbert matrix of order $n$ is nonsingular for all $n$, and so its rank is truly $n$. However, if you enter $\text{rank(hilb(15))}$ you obtain 12, so that MATLAB is off by three.
8. Programming in MATLAB

Using the functions we have already discussed, MATLAB can do very complicated matrix operations. However, sometimes there is a need for finer control over the elements of matrices and the ability to test, and branch on, logical conditions. Although prior familiarity with a high-level programming language is useful, MATLAB’s programming language is so simple that it can be learned quite easily and quickly.

8.1. Control Flow and Logical Variables

MATLAB has four flow of control and/or branching instructions: for loops, while loops, if-else branching tests, and switch branching tests.

Notation: All of these instructions end with an end statement, and it is frequently difficult to determine the extent of these instructions. Thus, it is very important to use indentation to indicate the structure of a code, as we do in the remainder of this tutorial. This greatly increases the readability of the code for human beings.

The general form of the for loop is

```matlab
for <variable> = <expression>
    <statement>
end
```

where the variable is often called the loop index. The elements of the row vector `expression` are stored one at a time in the variable and then the statements up to the end statement are executed. For example, you can define the vector \( x \in \mathbb{R}^n \) where \( x_i = i \sin(i^2 \pi/n) \) by

```matlab
x = zeros(n, 1);
for i = 1:n
    x(i) = i * sin( i^2 * pi/n);
end
```

(The first line is not actually needed, but it allows MATLAB to know exactly the size of the final vector before the for loops begin. This saves computational time and makes the code more understandable; it is discussed in more detail in Section 8.5.) In fact, the entire for loop could have been entered on one line as

```matlab
for i = 1:n
    x(i) = i * sin( i^2 * pi/n);
end
```

However, for readability it is best to split it up and to indent the statements inside the loop. Of course, you can also generate the vector by

```matlab
x = [1:n]’ .* sin( [1:n]’ .* 2 * pi/n )
```

which is certainly “cleaner” and executes much faster in MATLAB.

Warning: In using `i` as the index of the for loop, `i` has just been redefined to be `n` instead of \( \sqrt{n} \).

A more practical example of the use of a for loop is the generation of the Hilbert matrix of order `n`, which we have already discussed a number of times. This is easily done using two for loops by

```matlab
H = zeros(n);
for i = 1:n
    for j = 1:n
        H(i,j) = 1/(i + j - 1);
    end
end
```

Warning: In using `i` and `j` as the indices of the for loops, `i` and `j` have just been redefined to be `n` instead of \( \sqrt{n} \).

for loops often have branches in them. For this we need the if branch, which we now describe. The simplest form of the if statement is

```matlab
if <logical expression>
    <statement>
    ...
end
```

where the statements are evaluated as long as the `logical expression` is true. The `logical expression` is generally of the form

```matlab
<arithmetic expression-left> <operator> <arithmetic expression-right>
```

where `op` is one of the relational operators shown below. Some examples of logical expressions are

- `1 = 5`
- `x(1) >> 1`
- `imag(A(1,1)) <= 0`
- `sin(i) - 1 < x(i) + x(i)-3`

Is `i` equal to `5`? Is `x_2` greater than `7`? Is the imaginary part of `a_{ij}` nonzero? Is \( \sin 1 - 1 < x_1 + x_1^2 \)?

We can compare complex numbers to determine whether or not they are equal. However, mathematically we cannot apply the other four relational operators. For example, what does \( 2 + 3i < 3 + 2i \)? However, in MATLAB only the real parts of complex numbers are compared so

```matlab
>> 2 + 3i < 3 + 2i
returns 1.
```

Warning: String variables cannot be compared by `==` or `==`. Instead, if `a` and `b` are text variables, enter

```matlab
>> strcmp(a, b)
```

The result is true if the two character strings are identical and false otherwise.

<table>
<thead>
<tr>
<th>Relational Operator</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>&lt;</code></td>
<td>Less than.</td>
</tr>
<tr>
<td><code>&lt;=</code></td>
<td>Less than or equal to.</td>
</tr>
<tr>
<td><code>==</code></td>
<td>Equal.</td>
</tr>
<tr>
<td><code>&gt;=</code></td>
<td>Greater than or equal to.</td>
</tr>
<tr>
<td><code>&gt;</code></td>
<td>Greater than.</td>
</tr>
<tr>
<td>`!='</td>
<td>Not equal to.</td>
</tr>
<tr>
<td><code>strcmp(a, b)</code></td>
<td>Compares strings.</td>
</tr>
</tbody>
</table>

A second form of the if statement is

```matlab
if <logical expression>
    <statement group 1>
else
    <statement group 2>
end
```

where statement group 1 is evaluated if the `logical expression` is true and statement group 2 is evaluated if it is false. The final two forms of the if statement are

```matlab
if <logical expression>
    <statement group 1>
elseif <logical expression 2>
    <statement group 2>
elseif <logical expression 3>
    <statement group 3>
... elseif <logical expression r>
    <statement group r>
end
```

if `<expression>` can be a matrix in which each column vector is stored one at a time in `i`.

---

\[\hat{t}\]
if <logical expression>
<statement group>
else <logical expression>
<statement group>
... else <logical expression>
<statement group>
end

where statement group 1 is evaluated if the <logical expression> is true, statement group 2 is evaluated if the <logical expression> is true, etc. The final else statement is not required. If it occurs and none of the previous logical expressions is true, statement group r+1 is evaluated. If it does not occur and if none of the logical expressions are true, then none of the statement groups are executed.

When a logical expression such as

\[ a \land b \]

is evaluated, the result is either the logical value “TRUE” or “FALSE”. MATLAB calculates this as a numerical value which is returned in the variable ans. The value is 0 if the expression is false and 1 if it is true.

MATLAB also contains the logical operators “AND” (denoted by “\&”), “OR” (denoted by “|”), “NOT” (denoted by “~”), and “EXCLUSIVE OR” (invoked by the function xor). These set on false or true statements which are represented by numerical values: zero for false statements and nonzero for true statements. Thus, if a and b are real numbers then:

- the relational equation
  \[ c = a \land b \]
  means that \( c \) is true (i.e., 1) if both \( a \) and \( b \) are true (i.e., nonzero); otherwise \( c \) is false (i.e., 0).
- the relational equation
  \[ c = a \lor b \]
  means that \( c \) is true (i.e., 1) if \( a \) and/or \( b \) is true (i.e., nonzero); otherwise \( c \) is false (i.e., 0).
- the relational equation
  \[ c = a \land b \]
  means that \( c \) is true (i.e., 1) if \( a \) is false (i.e., 0); otherwise \( c \) is false (i.e., 0).
- the relational equation
  \[ c = \text{xor}(a, b) \]
  means that \( c \) is true (i.e., 1) if exactly one of \( a \) and \( b \) is true (i.e., nonzero); otherwise \( c \) is false (i.e., 0).

In the above statements \( c \) is a logical variable which has the logical value “TRUE” or “FALSE”. Frequently — but not always — the variable can be set by \( c = 1 \) or \( c = 0 \); but \( c \) is not a logical variable, but a numerical variable. Frequently — but not always — a numerical variable can be used instead of a logical variable. The preferred ways to set a logical variable are the following. The logical variable can be set by \( c = \text{logical}(1) \) or \( c = \text{logical}(0) \) — and now \( c \) is a logical variable. A simpler way to set the logical variable \( c \) is \( c = \text{true} \) or \( c = \text{false} \).

There are second logical operators “AND” (denoted by \&\&) and “OR” (\|\|) which are sometimes used. The statement

\[ c = a \land b \]

returns a (scalar) logical true if both inputs evaluate to true (so that if they are variables they both must be scalars). The difference from \& is that if \( a \) is false then \( b \) is not evaluated. Similarly,

\[ c = a \land b \]

returns a (scalar) logical true if either input evaluates to true. If \( a \) is true then \( b \) is not evaluated (as in C, C++, and Java). This is called short-circuiting the AND and OR operators.

### Logical Operators

<table>
<thead>
<tr>
<th>Operator</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A &amp; B</td>
<td>AND</td>
</tr>
<tr>
<td>A | B</td>
<td>OR</td>
</tr>
<tr>
<td>~A</td>
<td>NOT</td>
</tr>
<tr>
<td>xor(A, B)</td>
<td>EXCLUSIVE OR</td>
</tr>
</tbody>
</table>

The second MATLAB loop structure is the while statement. The general form of the while loop is

```matlab
while <logical expression>
    <statement>
    ...  
end
```

where the statements are executed repeatedly as long as the <logical expression> is true. For example, \( \text{eps} \) can be calculated by

```matlab
ep = 1;  
while 1 + ep > 1  
    ep = ep/2;  
end  
eps = 2*ep
```

It is possible to break out of a for loop or a while loop from inside the loop by using the break statement as in C. This terminates the execution of the innermost for loop or while loop.

The `continue` statement is related to `break`. It causes the next iteration of the for or while loop to begin immediately.

The `switch` function executes particular statements based on the value of a variable or an expression. Its general form is

```matlab
switch <variable or expression>
    case <Value 1>
        <statement group 1>
    case <Value 2>, <Value 2b>, <Value 2c>, ..., <Value 2n>
        <statement group 2>
    ...  
    case <Value n>
        <statement group r>
    otherwise
        <statement group r+1>
end
```

where statement group 1 is evaluated if the variable or expression has <Value 1>, where statement group 2 is evaluated if the variable or expression has values <Value 2a> or <Value 2b> or <Value 2c>, etc. (Note that if a case has more than one value, then all the values must be surrounded by curly brackets.) The final otherwise is not required. If it occurs and if none of the values match the variable or expression, then statement group r+1 is evaluated. If it does not occur and if none of the values match, then none of the statement groups are executed.

**Warning:** The `switch` function is different in MATLAB than in C in two ways:

- First, in MATLAB the `case` statement can contain more than one value; in C it can only contain one.
- And, second, in MATLAB only the statements between the selected case and the following one or the following `otherwise` or `end` (whichever occurs first) are executed; in C all the statements following the selected case are executed up to the next `break` or the end of the block.
8.2. Matrix Relational Operators and Logical Operators

Although MATLAB does have a quite powerful programming language, it is needed much less frequently than in typical high-level languages. Many of the operations and functions that can only be applied to scalar quantities in other languages can be applied to vector and matrices in MATLAB. For example, MATLAB’s relational and logical operators can also be applied to vectors and matrices. In this way, algorithms that would normally require control flow for coding in most programming languages can be coded using simple MATLAB functions.

If \( A, B \in \mathbb{R}^{m \times n} \) then the relational equation

\[
C = A \text{ rel} B
\]

is evaluated as \( c_{ij} = a_{ij} \text{ rel} b_{ij} \), where \text{ rel} is one of the relational operators defined previously. \( C \) is a logical array, that is, its data type is “logical” not “numeric”. The elements of \( C \) are all \( 0 \) or \( 1 \) if \( a_{ij} \text{ rel} b_{ij} \) is a false statement and \( 1 \) if it is a true one. Also, the relational equation

\[
C = k \text{ rel} A
\]

defined when \( c \) is a scalar. It is evaluated as if we had entered

\[
C = A \text{ rel} \text{ ones(size(A))}
\]

Similar behavior holds for logical operators:

\[
C = A \text{ rel} B
\]

means \( c_{ij} = a_{ij} \text{ rel} b_{ij} \),

\[
C = A \text{ rel} 1
\]

means \( c_{ij} = a_{ij} \text{ rel} 1 \),

\[
C = -A
\]

means \( c_{ij} = -a_{ij} \), and

\[
C = \text{ xor}(A, B)
\]

means \( c_{ij} = a_{ij} \text{ xor} b_{ij} \). Again the elements of \( C \) are all \( 0 \) or \( 1 \).

To show the power of these MATLAB functions, suppose we have entered

\[
F = \text{ rand}(m, n)
\]

and now we want to know how many elements of \( F \) are greater than 0.5. We can code this as

\[
r Elements = 0;
\]

\[
\text{for } i = 1:m
\]

\[
\text{for } j = 1:n
\]

\[
\text{if } F(i,j) > 0.5
\]

\[
r Elements = nr Elements + 1;
\]

\[
\text{end end}
\]

\[
r Elements
\]

However, it can be coded much more simply, quickly, and efficiently since the relational expression

\[
C = F > 0.5;
\]

or, to make the meaning clearer,

\[
C = (F > 0.5);
\]

generates the matrix \( C \) where

\[
C_{ij} = \begin{cases} 1 & \text{if } f_{ij} > 0.5 \\ 0 & \text{otherwise.} \end{cases}
\]

Since the number of ones is the result we want, simply enter

\[
\text{sum(sum(C > 0.5))}
\]

\[
\text{sum(sum(C))}
\]

And suppose we want to replace all the elements of \( F \) which are \( \leq 0.5 \) by zero. This is easily done by

\[
F = F .* (F > 0.5)
\]

The relational expression \( F > 0.5 \) generates a matrix with zeroes in all the locations where we want to zero the elements of \( F \) and ones otherwise. Multiplying this new matrix elementwise with \( F \) zeroes out all the desired elements of \( F \). We can also replace all the elements of \( F \) which are \( \leq 0.5 \) by \(-\pi\) using

\[
C = (F < 0.5)
\]

\[
F = F .* (C \neq 0)
\]

Shortly we will present two easier ways to do this.

There is even a MATLAB function which determines the locations of the elements of a vector or a matrix where some property is satisfied. The function

\[
\text{find}(A)
\]

generates a column vector containing the indices of \( x \) which are nonzero. (Recall that nonzero can also mean “TRUE” so that this function finds the elements where some condition is true.) For example, if \( x = (0, 4, 0, 1, -1, 0, \pi)^T \) then the resulting vector is \((2, 4, 5, 7)^T\). We can add \( 10 \) to every nonzero element of \( x \) by

\[
\text{ix} = \text{find}(x);
\]

\[
\text{x(ix) = x(ix) + 10;}
\]

\textbf{Note:} If no element of the vector \( x \) is nonzero, the result is the empty matrix \([\ ]\) and the following statement is not executed.

\textbf{Note:} There is a similar function which finds if a substring is contained in a string.

\[
k = \text{strfind(str, pattern)}
\]

returns the starting index for any and all occurrence of \text{pattern} in \text{str}.

\text{find} can also be applied to a matrix. The function

\[
\text{find}(A)
\]

first transforms \( A \) to a column vector (i.e., \( A(:) \)) and then determines the locations of the nonzero elements. As a simple example of the power of this function we can add \( 10 \) to every nonzero element of \( A \) by

\[
jA = \text{find}(A);
\]

\[
A(jA) = A(jA) + 10
\]

Or we can work with the matrix directly by entering

\[
A = \text{find}(A)
\]

The two column vectors \( iA \) and \( jA \) contain the rows and columns, respectively, of the nonzero elements. We can also find the locations of the nonzero elements and their values by
We can also convert a non-logical variable to a logical one by using the MATLAB function logical. To explain logical arrays more clearly, we take a specific and very simple example. Enter

```
>> v = [0; .25; .5];
>> c = (v >= .5);
```

so that \( v = [0 \ 0 \ .25 \ .5 \ 1.0] \) and \( c = [0 \ 1 \ 1 \ 1 \ 1] \) where "0" denotes false and "1" denotes true. The result of

```
>> v(c)
```

is \( [.5 \ .75 \ 1.0] \). That is, \( c \) is a logical vector and \( v(c) \) deletes the elements of \( v \) which are "false". On the other hand

```
>> iv = find(v >= .5);
returns iv = [1 2] and
```

```
>> v(iv) = [3];
returns v = [.5 .75 1.0]. The difference between \( c \) and \( iv \) is that \( c \) is a logical vector and \( iv \) is a scalar vector. If you enter

```
>> v([0 0 1 1 1]) % WRONG
```

instead of

```
>> v(c)
```
you obtain the error message

??? Subscript indices must either be real positive integers or logicals.

because \([0 \ 0 \ 1 \ 1 \ 1]\) is a numeric vector and so must contain the numbers of the elements of \( v \) which are desired — but there is no element "0".

MATLAB also has two functions that test vectors and matrices for logical conditions. The function

```
>> any(z)
```

returns 1 if any element of the vector \( z \) is nonzero (i.e., "TRUE"); otherwise 0 is returned. When applied to a matrix, it operates on each column and returns a row vector. For example, we can check whether or not a matrix is tridiagonal by

```
>> any( any( triu(A, 2) + triu(A, -2)) )
```

Here we check all the elements of \( A \) except those on the main diagonal and on the two adjacent ones. A result of 1 means that at least one other element is nonzero. If we want a result of 0 to mean that \( A \) is tridiagonal we can use

```
>> any( any( triu(A, 2) + triu(A, -2)) )
```

instead. The function

```
>> any(A)
```

operates columnwise and returns a row vector containing the result of \( any \) as applied to each column. The complementary function \( all \) behaves the same as \( any \) except it returns 1 if all the entries are nonzero (i.e., "TRUE"). For example, you can determine if a matrix is symmetric by

```
>> all( all(A == A') )
```

A result of 1 means that \( A \) is identical to \( A' \). It is also easy to check if two arrays (including structures and cells) have exactly the same contents. The function

```
>> isequal(A, B)
```

returns true if all the contents are the same and false otherwise. This means that (nonempty) arrays must have the same data type and be the same size.

For completeness we mention that MATLAB has a number of other functions which can check the status of variables, the status of the elements of vectors and matrices, and even of their existence. For example, you might want to zero out all the elements of a matrix \( A \) which are Inf or NaN. This is easily done by

```
>> A(find( infinite(A) ) == false ) = 0
```
or

```
>> A(find( ~infinite(A) ) == true ) = 0
```

where \( infinite(A) \) generates a matrix with 1 in each element for which the corresponding element of \( A \) is finite. To determine if the matrix \( A \) even exists, enter

```
>> exist('A')
```

See the table below for more details and more functions.
Logical Functions

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>all</td>
<td>True if all the elements of a vector are true; operates on the columns of a matrix.</td>
</tr>
<tr>
<td>any</td>
<td>True if any of the elements of a vector are true; operates on the columns of a matrix.</td>
</tr>
<tr>
<td>exist('name')</td>
<td>False if this name is not the name of a variable or a file. If it is, this function returns: 1 if this is the name of a variable, 2 if this is the name of an m-file, 5 if this is the name of a built-in MATLAB function. The indices of a vector or matrix which are nonzero.</td>
</tr>
<tr>
<td>find</td>
<td>The indices of a matrix &amp; from A(:,), the column vector form, to A, the matrix form.</td>
</tr>
<tr>
<td>ind2sub</td>
<td>Converts indices of a matrix &amp; from A(:,), the column vector form, to A, the matrix form.</td>
</tr>
<tr>
<td>logical</td>
<td>Converts a numeric variable to a logical one.</td>
</tr>
<tr>
<td>strfind</td>
<td>Find any and all occurrences of a substring in a string.</td>
</tr>
<tr>
<td>{Cpt}.iscell</td>
<td>True for a cell array.</td>
</tr>
<tr>
<td>isschar</td>
<td>True for a character array.</td>
</tr>
<tr>
<td>iscolumn</td>
<td>True for a column vector.</td>
</tr>
<tr>
<td>isempty</td>
<td>True if the array is empty; i.e., [ ].</td>
</tr>
<tr>
<td>isequal</td>
<td>Tests if two (or more) arrays have the same contents.</td>
</tr>
<tr>
<td>isfield</td>
<td>True if the argument is a structure field.</td>
</tr>
<tr>
<td>isfinite</td>
<td>Generates an array with 1 in all the elements which are finite (i.e., not Inf or NaN) and 0 otherwise.</td>
</tr>
<tr>
<td>isfloat</td>
<td>True if a floating-point array.</td>
</tr>
<tr>
<td>isinf</td>
<td>Generates an array with 1 in all the elements which are Inf and 0 otherwise.</td>
</tr>
<tr>
<td>isslogical</td>
<td>True for a logical array.</td>
</tr>
<tr>
<td>ismember</td>
<td>Generates an array with 1 in all the elements which are contained in another array.</td>
</tr>
<tr>
<td>isman</td>
<td>Generates a matrix with 1 in all the elements which are NaN and 0 otherwise.</td>
</tr>
<tr>
<td>isnumeric</td>
<td>True for a floating-point array.</td>
</tr>
<tr>
<td>isprime</td>
<td>Generates an array with 1 in all the elements which are prime numbers. Only non-negative integers are allowed in the elements.</td>
</tr>
<tr>
<td>isnan</td>
<td>True for a real array, as opposed to a complex one.</td>
</tr>
<tr>
<td>isrow</td>
<td>True for a row vector.</td>
</tr>
<tr>
<td>isscalar</td>
<td>True for a scalar variable.</td>
</tr>
<tr>
<td>isstruct</td>
<td>True for a sparse array.</td>
</tr>
<tr>
<td>issparse</td>
<td>True for a structure array.</td>
</tr>
<tr>
<td>isvector</td>
<td>True for a vector, as opposed to a matrix.</td>
</tr>
</tbody>
</table>

8.3. Function M-files

We have already discussed script m-files, which are simply an easy way to collect a number of statements and execute them all at once. Function m-files, on the other hand, are similar to functions or procedures or subroutines or subprograms in other programming languages. Ordinarily, variables which are created in a function file exist only inside the file and disappear when the execution of the file is completed. These are called local variables. Thus you do not need to understand the internal workings of a function file; you only need to understand what the input and output arguments represent.

Note: The generic term for script files and function files is m-files, because the extension is “m”.

Unlike script files, function files must be constructed in a specific way. The first line of the file <file name>.m must begin with the keyword function. Without this word, the file is a script file. The complete first line, called the function definition line, is

\[ \text{function } <\text{out} > = \text{<function name}> (<\text{in} > 1, \ldots, <\text{in} > n) \]

or

\[ \text{function } [<\text{out} > 1, \ldots, <\text{out} > m] = \text{<file name}> (<\text{in} > 1, \ldots, <\text{in} > n) \]

where the name of the function must be the same as the name of the file (but without the extension). The input arguments are <\text{in} > 1, <\text{in} > 2, \ldots . The output arguments appear in the left of the equal sign: if there is only one output argument, i.e., <\text{out} >, it appears by itself; if there is more than one, i.e., <\text{out} > 1, etc., they must be separated by commas and must be enclosed in square brackets.

Variables in MATLAB are stored in a part of memory called a workspace. The base workspace contains all the variables created during the interactive MATLAB session, which includes all variables created in script m-files which have been executed. Each function m-file contains its own function workspace which is independent of the base workspace and every other function workspace. The only way to “connect” these workspaces is through the arguments of a function or by using the global command (which we will discuss shortly).

There is great flexibility in the number and type of input and output arguments; we discuss this topic in great detail later. The only detail we want to mention now is that the input arguments are all passed “by value” as in C. (That is, the values of the input arguments are stored in temporary variables which are local to the function.) Thus, the input arguments can be modified in the function without affecting any input variables in the calling statement.

Warning: The name of the file and the name of the function must agree. This is also the name of the function that executes the function.

Comment lines should immediately follow. A comment line begins with the percent character, i.e., “%”. Comment lines, which immediately follow the function definition line constitute the documentation for this function; these lines are called the online help entry for the function. When you type

\[ \text{help <function name>} \]

all these lines of documentation are typed out. If you type

\[ \text{type <function name>} \]

the entire file is printed out. In addition, the first line of documentation, i.e., the second line of the file, can be searched for keywords by entering

\[ \text{lookfor <keyword>} \]

Make sure this first comment line contains the name of the function and important keywords which describe its purpose.

Note: Comments can be placed anywhere in an m-file, including on a line following a MATLAB statement. The initial comment lines in a script file and the comment lines in a function file which immediately follow the first line are special: they appear on the screen when you type

\[ \text{help <function name>} \]

Before discussing functions at great length, there is one technical detail it is important to consider before you use up your CPU. How does MATLAB find the m-files you have created? Since MATLAB contains thousands of functions, this is not an easy task. Once MATLAB has determined that the word is not a variable, it searches for the function in a particular order. We show the order here and then discuss the items in detail throughout this subsection.

1. It checks if <function name> is a built-in function (i.e., coded in C).
2. It checks if <function name> is a function, i.e., the primary function, a subfunction, or a nested function in the current scope, in the current file. (We discuss all these terms shortly.)
3. It checks if the file <function name>.m exists in the current directory.
4. It checks if the current directory has a subdirectory called “private”: if it does, MATLAB checks if the file <function name>.m exists in this subdirectory.
5. It searches the directories in the search path for the file <function name>.m.

Note from (3) that MATLAB searches in the current directory for the function by searching for the m-file with the same name. If the m-file is not in the current directory, the simplest way to enable MATLAB

\[ \text{function } <\text{out} > = \text{<function name}> (<\text{in} > 1, \ldots, <\text{in} > n) \]

\[ \text{function } [<\text{out} > 1, \ldots, <\text{out} > m] = \text{<file name}> (<\text{in} > 1, \ldots, <\text{in} > n) \]

If you are worried because passing arguments by value might drastically increase the execution time of the function, we want to reassure you that this does not happen. To be precise, MATLAB does not actually pass all the input arguments by value. Instead, an input variable is only passed by value if it is modified by the function. If an input variable is not modified, it is passed “by reference”. (That is, the input argument is the actual variable used in the calling statement and not a local copy.) In this way you get the benefit of “call by value” without any unnecessary overhead. And how does MATLAB know if an input argument is modified? It can only be modified if it appears on the left-hand side of an equal sign inside the function!
to find it is have the subdirectory in your search path. If you type

```path```

you will see all the directories that are searched. If you have created a subdirectory called “matlab” in your main directory, this is usually the first directory searched (unless the search path has been modified).

Thus, you can put your m-files in this subdirectory and be sure that MATLAB will find them. You can also add directories to the search path by

```addpath('new_directory', 'new_directory2', ...)```

(The former puts “new_directory” at the beginning of the search path while the latter puts it at the end.)

Alternately, you can add one or more directories at the beginning of the search path by

```addpath('new_directory #1', 'new_directory #2', ...)```

**Warning:** When you begin a MATLAB session, it always checks if the subdirectory matlab or

```Documents/MATLAB```

exists in your main directory. If you create this subdirectory after you start a MATLAB session, it will not be in the search path.

Now we return to our discussion of creating functions. We begin with a simple example of a function file which constructs the Hilbert matrix (which we have already used a number of times).

```function```

```H = hilb_local(n)```

```% hilb_local: Hilbert matrix of order n (not from MATLAB)```

```% This is one of the most famous examples of a matrix which is```

```% nonsingular, but which is very badly conditioned.```

```H = zeros(n);```

```for i = 1:n```

```for j = 1:n```

```H(i,j) = 1/(i+j-1);```

```end```

```end```

The input argument is n and the output argument is H. The first line of the documentation includes the name of the function as well as a brief description that `lookfor` uses. The following lines of documentation also appear on the screen if we enter `help hilb_local`.

**Note:** The above code is not presently used in MATLAB (although it was in early versions.)

The actual MATLAB code for this function is shown in Section 8.5.

We follow by defining H to be an n-by-m matrix. Although not essential, this statement can greatly increase the speed of the function because space can be preallocated for the matrix. For example, consider the following code:

```function```

```a = prealloc(m, initialize)```

```% prealloc: testing how well preallocating a vector works```

```% m = the size of the vector```

```% initialize = true - preallocate the vector```

```% = false - do not```

```if initialize == true```

```a = zeros(n,1);```

```end```

```a(1) = 1;```

```for i = 2:n```

```a(i) = a(i-1) + 1;```

```end```

If `initialize` is set to false the vector a is not preallocated, while if `initialize` is true it is. We find that

```prealloc(100000, true);```

runs over 400 (that’s right, four hundred) times as fast as

```prealloc1(100000, false);```

Note that 1 and 3 are removed from $\sqrt{-1}$ since they appear as for loop indices. However, since 1 and 3 are local to this function, this does not have any effect on the calling code when this function is executed. Also, the variable H is local to the function. If we type

```Z = hilb_local(12)```

then the matrix Z contains the Hilbert matrix and H is undefined.

Normally functions are completed when the end of the file is reached (au above). If the control flow in a function file is complicated enough, this might be difficult to accomplish. Instead, you can use the `return` statement, which can appear anywhere in the function and force an immediate end to the function.

**Note:** The return function also works in script m-files.

Alternatively, you can force the function to abort by entering `error(strerror)`

If the string is not empty, the string is displayed on the terminal and the function is aborted; on the other hand, if the string is empty, the statement is ignored.

One feature of function files which is occasionally very useful is that they can have a variable number of input and output variables. For example, the norm of a vector x can be calculated by entering

```norm(x, p)```

if $p = 1$, 2, or `Inf`, or, more simply, by

```norm(x)```

if $p = 2$. Similarly, if only the eigenvalues of a matrix $A \in \mathbb{C}^{n \times n}$ are desired, enter

```eigval = eig(A)```

However, if both the eigenvalues and eigenvectors are desired, enter

```[V, D] = eig(A)```

where $D \in \mathbb{C}^{n \times n}$ is a diagonal matrix containing the $n$ eigenvalues on its diagonal and the corresponding eigenvectors are found in the same columns of the matrix $V \in \mathbb{C}^{n \times n}$.

**Note:** On the right side of an equation, [$V$] or [$V, D$] is the matrix whose initial columns come from $V$ and whose final columns come from $D$. This requires that $V$ and $D$ be matrices which have the same number of rows. On the left side, $[V, D]$ denotes the two output arguments which are returned by a function. $V$ and $D$ can be completely different variables. For example, one can be a character variable and the other a matrix.

MATLAB can also determine the number of input and output arguments: `nargin` returns the number of input arguments and `nargout` returns the number of output arguments. For example, suppose we want to create a function file which calculates

$$f(x, \xi, \alpha) = e^{-(\xi - C)^T \sin x}.$$  

We can “spruce” this function up to have default values for $\xi$ and $\alpha$ and also to calculate its derivative with the following function file.

```function```

```out1, out2 = spruce(x, xi, a)```

```% spruce: a silly function to make a point, f(x,xi,a) = sin(x)*exp(-a*(x-xi)^2)```

```% if only x is input, xi = 0 and a = 1```

```% if only x and xi are input, a = 1```

```% if only one output argument, f(x,xi,a) is calculated```

```% if two output arguments, f(x,xi,a) and f'(x,xi,a) are calculated```

```if nargin == 1```

```x = 0;```

```a = 1;```

```else```

```nargin = 2```

```a = 1;```

```end```

```out1 = exp(-a.*(x-xi).^2).*sin(x);```

```if nargout == 2```

```out2 = exp(-a.*(x-xi).^2).*cos(x) - 2.*a.*.(x-xi).*sin(x);```

```end```

If there is only one input argument then $\xi$ is set to 0 and $\alpha$ is set to 1 (which are useful default values) while if there are only two input arguments then $\alpha$ is set to 1. If there is only one output argument then only $f(x)$ is calculated, while if there are two output arguments then both $f(x)$ and $f'(x)$ are
calculated. Also, note that $x$ can be a scalar (i.e., a single value) or it can be a vector. Similarly, $\xi$ and $a$ can
each be a scalar or a vector. If $x$ is a vector, i.e., $(x_1, x_2, \ldots, x_T)^T$, while $\xi$ and $a$ are scalars, then the function is
\[ f(x_i, \xi, a) = e^{-a(x_i - \xi)^2} \sin(x_i) \quad \text{for } i = 1, 2, \ldots, n, \]
and all the values can be calculated in one call to `spruce`. If, on the other hand, $x$, $\xi$, and $a$ are all vec-
tors, then the function is
\[ f(x_i, \xi_i, a_i) = e^{-a_i(x_i - \xi_i)^2} \sin(x_i) \quad \text{for } i = 1, 2, \ldots, n, \]
and again, all the values can be calculated in one call to `spruce`.

A common error which writing a function m-file is forgetting that the argument(s) might be vectors or matrices. For example, we mentioned the Heaviside step function in *Some Common Real Mathematical Functions* on page 13, but pointed out that it is in a toolbox which you might not have. We could write it as

```matlab
function Y = myheaviside(x)
    if x < 0
        y = 0;
    elseif x == 0
        y = 0.5;
    else
        y = 1;
    end
    return
ans = 1
```

Note that the function does not return an error message — it simply returns an incorrect result. (Inciden-
tially, `myheaviside([-6:0])` also returns 1.) The following function, however, will work in all cases.

```matlab
function Y = myheaviside(x)
    Y = zeros(size(x));
    Y(x==0) = 1;
    Y(x>0) = 0.5;
end
```

((The input and output arguments are capitalized to indicate that they can be matrices.) For example,

```matlab
>> myheaviside([-3:3])
returns
ans =
   0   0   0   0.5000   1.0000   1.0000   1.0000
```

We have now presented all the essential features of the MATLAB programming language, and it cer-
tainly is a "minimal" language. MATLAB can get away with this because most matrix operations can be
performed directly — unlike in most other programming languages. You only need to write your own function if MATLAB cannot already do what you want. If you want to become proficient in this language, simply use the `type` command to look at the coding of some functions.

Now that we have discussed the essentials of programming in MATLAB, it is necessary to discuss how
to program correctly. When you are entering one statement at a time in the text window, you immediately
see the result of your calculation and you can determine whether or not it is correct. However, in an m-file
you have a sequence of statements which normally end with semicolons so that you do not see the inter-
mediate calculations. What do you do if the result is incorrect? In other words, how do you debug your
m-file?

There are a number of simple techniques you can use and we discuss them in turn. In a script m-file
intermediate calculations are normally not printed out, but they are still available to look at. This can fre-
cently lead to an understanding of where the calculation first went wrong. However, this is not true of
function m-files since all the local variables in the function disappear when the function ends. Of course,
with any m-file you can selectively remove semicolons so that intermediate results are printed out. This is
probably the most common method of debugging programs — no matter what programming language is
being used.

When loops are involved (either using `for` or `while` statements), the printed output can seem to be
crassless (and it is endless if you are in an infinite loop). And it is usually impossible to read the output
since it is zipping by at (what appears to be) nearly the speed of light! The `pause` function can slow
down or even stop this output. By itself `pause` stops the m-file until some key is pressed while
`pause(<floating-point number>)`
pauses execution for this many seconds — or fractions of a second. (This is computer dependent, but
`pause(0.01)` should be supported on most platforms.) You can even turn these pauses on and off by
using `pause on` and `pause off`.

`Note`: Occasionally, you will forget that you have put the function
`pause`
into your code and wonder why it is taking so long to execute. The alternative
`input(‘Press Enter to continue’, ’s’);`
`pauses` the code and also reminds you that it has been paused.

The `echo` command is also useful for debugging script and function m-files, especially when if state-
ments are involved. Typing

```matlab
>> echo on
```
turns on the echoing of statements in all script files (but not printing the results if the statements end
with semicolons), and `echo off` turns echoing off again. However, this does not affect function files. To
turn echoing on for a particular function, type

```matlab
>> echo <function name> on
```
and to turn echoing on for all functions, type

```matlab
>> echo on all
```

The `keyboard` command is also very useful for debugging m-files. It stops execution of the m-file, simi-
lar to the `pause` function. However, it returns complete control to the user to enter any and all MATLAB
statements. In particular, you can examine any variables in the function’s workspace. If desired, you can
also change the value of any of these variables. The only way you will recognize this is not a “standard”
MATLAB session is that the prompt is

```matlab
K>
```
for Keyboard. To terminate the “keyboard” session and return control to the m-file, enter

```matlab
K> return
```
To terminate both the “keyboard” session and the execution of the m-file, enter

```matlab
K> quit
```
When using the debugger, you are not running your program. Instead, you are running the debugger
— it is running your program. Thus, many of the commands you enter are commands for the debug-
ger, to distinguish these commands from “normal” MATLAB commands/functions they begin with `db.`
There are two ways to run the debugger: you can type the debugger commands into the workspace, or in
a MATLAB window you can use the mouse and click on the corresponding icons.

In addition, you are still in the workspace so many MATLAB statements can still be executed. For ex-
ample, to see the values of variables just type the variable name into the workspace. (Alternatively, you
can move the mouse over the variable name in the window and its description and value(s) will be shown.)
Also, you can run most MATLAB statements as long as you do not try to create new variables (but you
can modify existing variables).

We will not discuss the commands in this debugger in detail, but only provide a brief description of
each one, because these are similar to commands in any debugger. If you have experience with using a
debugger, `doc` will give you complete details.
Some, but not all, of these commands can be found in the MATLAB editor window when you are editing a MATLAB script or function. It is necessary to set one or more breakpoints to stop the execution of the code so that you can take control. This is easily done by using the left mouse button to click on the dashes just to the right of the line numbers, which will cause a small red circle to appear to remind you of the breakpoints.

Now we want to discuss the arguments in a MATLAB function, since they are used somewhat differently than in other programming languages. For example, in

```
function out = funct1(a, t)
```

a and t are the input arguments and out is the output argument. Any and all input variables are local to the function and so can be modified without affecting the arguments when the function `funct1` is called. (This is true no matter what type of variables they are.) In

```
function [out1, out2, out3] = funct2(z)
```

z is the only input argument and there are three output arguments, each of which can be any type of variable. There is no requirement that all three of these output arguments actually be used. For example, the calling statement might be any of the following:

```
>> art = funct2(1.6)
[physics, chemistry] = funct2([1 2 3])
```

(just to be somewhat silly).

As an aside, it is even possible to not output some of the former arguments. Some examples are:

```
>> [a, - chemistry] = funct2([1 2 3])
```

and

```
>> [a, - chemistry, horticulture] = funct2([1 30; 6, 5])
```

where ‘-’ is a placeholder for an output argument. It is never necessary to use it in this way, but it makes a code easier to read. When an output argument is given a name, anyone reading the code will expect it to be used later in the code — and so, if it isn’t, why isn’t it? I used to use the variable name “dummy” for an unneeded output argument, but this is far preferable.

In MATLAB input arguments occur on the right side of the equal sign and output arguments occur on the left. Arguments which are to be modified by the function must occur on both sides of the equal sign in the calling statement. For example, in

```
function out2 if z is modified and returned in out1 then the calling sequence should be
```

```
>> [z, b, c] = funct2(z)
```

where z appears on both sides of the equal sign. (There is an alternative to this awkward use of parameters which are modified by the function: you can make a variable global, as we discuss at the end of this section. However, this is not usually a good idea.)

There is another difference between MATLAB and most other programming languages where the type of each variable has to be declared; either explicitly or implicitly. For example, a variable might be an integer, a single-precision floating-point number, a double-precision floating-point number, a character string, etc. In MATLAB, on the other hand, there is no such requirement. For example, the following statements can follow one another in order and define \( x \) to be a string variable, then a vector, then a scalar, and finally a matrix.

```
>> x = ’qwerty’
>> x = x * 0
>> x = sum(z)
>> x = x + [1; 2; 3 4]
```

It is particularly important to understand this “typelessness” when considering output arguments. For example, there are three output arguments to `funct2` and any of them can contain any type of variable. In fact, you can let the type of these arguments depend on the value or type of the input argument. This is probably not something you should want to do frequently, but it is sometimes very useful.

Occasionally, there is a need to pass values from the workspace to a function or to pass values between different functions without using the input arguments. (As we discussed earlier, this may be desirable if a variable is modified by a function.) In C this is done by using global variables. MATLAB also has global variables which are defined by declaring the variables to be global using

```
global variable 1 variable 2 variable 3 ...
```

By the way, a variable is a global variable if it appears in a global statement and a local variable if it does not. (Note that a variable can be a local variable in one function and a global variable in another.) This statement must appear in every function which is to share the variables. If the workspace is also to share these variables, you must type this statement (or be put into a script file which you execute) before these variables are used.

Warning: Spaces, not commas, must separate the variables in a global statement. Using global statements is generally considered to be very bad practice.

Instead of using a global variable, it is frequently preferable to save the value of a local variable between calls to the function. Normally, local variables come into existence when the function is called and disappear when the function ends. Sometimes it is very convenient to be able to “save” the value of a local variable so that it will still be in existence when the function is next called. In C, this is done by declaring the variable static. In MATLAB it is done by declaring the variable persistent using

```
persistent variable 1 variable 2 variable 3 ...
```

Warning: Spaces, not commas, must separate the variables.

Note: The first time you enter the function, a persistent variable will be empty, i.e., [ ] , and you can test for this by using isempty.

We now present a simple example where persistent variables are very helpful. Suppose we want to write a function m-file to evaluate

```
h(y) = \left( y_1 \right) \left( y_1^2 \right) \left( -y_2 \right) \left( \cos(y_3) \right)
```

where, \( \alpha, \beta, \Gamma, \Sigma \) are parameters which will be set initially and then left unchanged during a run.

(The parameter Gamma is capitalized even though it is not a matrix because \( \gamma \) is very different from \( \Gamma \).)

We might be studying a mathematical model where this function will be evaluated many, many times for different values of \( y \). For each experiment these parameters will be fixed, but they will be different for each experiment. We do not want to “hardcode” the values in the function because we would have to repeatedly change the function — which is very undesirable. Certainly we can write the function as

```
function z = func2(y, param)
```

but then we have to include these four parameters in each call to the function. We can always simplify this function by combining the four parameters into one structure by

```
function z = func2(y, param) = param.gamma+cos(param.omega*t1) );
```

but then it is harder to read the equation. (If this function was more complicated it would be much harder to read.) To make this last function easier to read we could write it as

```
function z = func2(y, param) = param.gamma+cos(param.omega*t1) );
```

...
function z = fnc23(y, param)
alpha = param.alpha
beta = param.beta
Gamma = param.Gamma
omega = param.omega

z = (y(2) - y(1)+(1-beta*y(1)^2)-alpha*y(2)+Gamma*cos(omega*t) );

end

Warning: Recall that the kth argument in varargin is varargin[k], NOT varargin(k).

The final — but very important — technical detail about function m-files concerns an important element of programming style in any computer language. It frequently happens that a block of code occurs two or more times in a function. Sometimes these blocks can be combined by using a loop, but, even if possible, this often makes the code unwieldy. Instead, this block of code can be put into a new function and called from the original function. Another reason for splitting a block of code off into a new function is when the function has grown large enough to be hard to comprehend. The remedy is to split the code up into a number of functions, each of which can be easily understood and debugged. In MATLAB functions normally have to be separated into different files so that each function and its file name agree; otherwise, MATLAB cannot find the function. This can be annoying if a number of files have to be created; for example, it can be difficult to remember the purpose of all these functions, and it can be difficult to debug the primary function. MATLAB has a feature to handle this proliferation of files: function m-files can contain more than one function. The first function in the file is called the primary function and its name must agree with the name of the file. There are two further types of functions: Any remaining functions are called subfunctions or nested functions. (At the end of Section 10.2 we code the function gravity using a number of nested functions.)

Note: The primary function or a subfunction begins with the function definition line (i.e., the line which begins with the keyword function). It is possible to end the primary function and each subfunction with the statement end; but this is not necessary because MATLAB recognizes that a function has ended when it encounters the next function statement. However, if a nested function is used then it — and all other functions — must end with the end statement.

First, we discuss subfunctions, which are quite simple. They are placed following the primary function and between or following other subfunctions. For example,

function primary_function
% code for the primary function
function subfunction1
% code for the first subfunction
function subfunction2
% code for the second subfunction

These subfunctions are only visible to the primary function and to other subfunctions in the same file. Thus, different m-files can contain subfunctions with the same name. Also, the help, lookfor, and type commands can only access the primary file. It is crucial to understand that variables in the primary function or in a subfunction are local to that function and unknown outside it (unless they are declared to be global). The only way to pass variables between these functions is through the argument list. Usually, subfunctions are sufficient — and they are much easier to describe. When they are not sufficient, we have nested functions. Its main advantage (as far as we are concerned) is that variables can be passed into and out of a nested function without being in the argument list. Nested functions are more complicated than subfunctions and we will only provide a brief discussion.†

To make this discussion specific, consider the following function m-file:

function z = fnc23(y, param)
alpha = param.alpha
beta = param.beta
Gamma = param.Gamma
omega = param.omega

z = (y(2) - y(1)+(1-beta*y(1)^2)-alpha*y(2)+Gamma*cos(omega*t) );

end

† They are similar to internal functions in Fortran 95, and they are somewhat related to inner classes in Java — but not in C++.
function [p1, p2, p3] = nested_ax(x, y, z) % 1
p1 = x; % 2
p2 = y; % 3
r = k; % 4
p1 = p1 + nested_1(20); % 5
p2 = p1 + nested_2(40); % 6

%%%% nested functions follow

function out1 = nested_1(r) % 7
nla = p1 + z + r; % 8
nib = p2 + z; % 9
p = p1 + p2; % 10
p1 = nla + nib; % 11
p2 = snavlhb; % 12
p3 = sub_1(p, p2); % 13
out1 = nested_2(1); % 14
end

function out2 = nested_2(r) % 15
n2a = p1 + r; % 16
p = p + p1 + p2; % WRONG % 17
p1 = n2a + z + nib; % 18
p2 = p2 + p2; % 19
p3 = n2a; % 20
out2 = p1 + p2; % 21
end

disp(['r' = ', num2str(r)]) % 22
disp(['nib' = ', num2str(nib)]) % 23
end

%%%% subroutine

function s3 = sub_1(a, b) % 24
s1 = 10; % 25
s2 = a + b; % 26
s2 = s2 + p2; % WRONG % 27
s3 = s2 + nested_3(s1); % 28
end

%%%% nested function follows

function out3 = nested_3(t) % 29
n3h = s1 + t; % 30
out3 = n3h; % 31
end

% disp(out3) % WRONG % 32
end % 33

end % 34

A nested function is within another function. For example, the nested functions nested_1 and nested_2 are nested within the primary function nested_ex, and the nested function nested_ex is nested within the subroutine sub_1. (Nested functions can have other nested functions within them, but enough is enough.)

The important concept to understand when using nested functions is the scope of variables in the function m-file. The scope of a variable is the context within which it is defined, i.e., where it can be set, modified, and used. Now let us consider a function workspace. The workspace of the primary function is also independent of the workspace of each subroutine. However, since a nested function is within one or more other functions, it is within the workspace of this function or these functions. In the function nested_ex the nested functions nested_1 (lines 7–15) and nested_2 (lines 16–23) have access to the variables p1, p2, and p3 of the primary function (lines 1–26). They also have access to the subroutine sub_1 (lines 27–37) (but not any of its variables) as shown in line 13. Note that nested_2 also has access to nested_1 as shown in line 14.

To begin, a nested function can have local variables. Any variable in the argument list of a nested function is local to that function, and the same is true for any variable which contains values returned by the function. Thus, the variable out1 is local to the function nested_1 and out2 is local to nested_2. Also, the variable r is local to nested_1 and also local to nested_2. The variable which is defined on line 4 is unchanged by the calls to the two nested functions. The value returned in line 24 will always be 5. Similarly, the variables t and out3 are local to nested_3.

What about the remaining variables? The variables nla and p contain values returned by the function and are defined on line 8 and 9. p1 and p2 are defined in nested_ex and so can be used in nested_1 and nested_2. This is also true for p3 which is an output variable in nested_ex. Finally, note that nib is not used in nested_ex until line 25, after nested_1 and nested_2 have been called, but it can still be used in nested_2 even though it was defined in nested_1.

Just for completeness, none of the variables in nested_ex can be accessed in sub_1 and vice versa. For instance, p2 cannot be accessed in the subroutine sub_1 as we show on line 30. Now let us return to the topic of how MATLAB finds a function. As we stated previously (but did not discuss), when a function is called from within an m-file, MATLAB first checks if the function named is the primary function or a subroutine in the current file. If it is not, MATLAB searches for a private function by the same name (described below). Only if all this fails does MATLAB use your search path to find the function. Because of the way that MATLAB searches for functions, you can replace a MATLAB function by a subroutine in the current directory using private functions.

In the previous paragraph we described how to create a subroutine to replace one function by another of the same name. There is another, more general, way to handle this replacement: you can create a subroutine in your current directory with the special name "private". Any routines in this subdirectory are visible only to functions in the current directory. The functions in this subdirectory are called private functions.

For example, suppose we are working in the directory "personal" and have created a number of files which use rref to solve linear systems. And suppose we have written our own version of this function, because we think we can calculate the reduced row echelon of a matrix more accurately. The usual way to test our new function would be to give it a new name, say myrref, and to change the call to rref in every file in this directory to myrref. This would be quite time-consuming, and we might well miss some. Instead, we can code and debug our new function in the subdirectory "private", letting the name of our new function be rref and the name of the m-file be rref.m. All calls in the directory to rref will use the new function we are testing in the subdirectory "private", rather than MATLAB’s function. Even more important, any function in any other directory which calls rref will use the MATLAB function and not our "new, improved version".

The final topic we will briefly discuss involves recursion. It is possible — and sometimes useful — for a function to call itself. As a simple example, consider the Fibonacci sequence
\[ f_{n+2} = f_{n+1} + f_n \] for \( n \geq 0 \)

---

**WARNING**
A common mistake when using nested functions is defining what you think is a local variable in a nested function and forgetting that the same variable name is used in the outer function. It is hard to imagine making that mistake here because the code is so short — but it can easily happen in a real code. One solution is to append a special character to all local variables in nested functions (for example, append an underscore, i.e., "_", to the end of the name of each local variable).
with initial values
\[ f_0 = 1 \quad \text{and} \quad f_1 = 1 \]
This sequence, i.e., 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, ..., can be coded as

```
function y = fibonacci(n)
    \% WARNING: VERY VERY INEFFICIENT CODE
    \% \% must be a nonnegative integer \%
    if n == 0
        y = 1;
    elseif n == 1
        y = 1;
    else
        y = fibonacci(n-1) + fibonacci(n-2); \% two recursive calls for n > 1
    end
end
```
A recursive code must be able to stop, and this code stops if \( n = 0 \) or \( n = 1 \). For larger values of \( n \), the function is called recursively.

**Warning:** This code is very, very inefficient. We have provided it simply to show how recursion can lead to a very compact code. The reason this code is very inefficient is because it uses an incredibly large amount of CPU time for "large" \( n \) (and 50 is very, very, very large). In fact, in order to calculate \( f_5 \), the function fibonacci is called recursively \( 2^5 - 2 \) times — and \( f_5 \) grows exponentially with \( n \).

There are examples for which recursion is advantageous. However, our aim is simply to show how to use it — not whether to use it.

---

**Function Commands**

- **function** Begins a MATLAB function.
- **end** Ends a function. This statement is only required if the function m-file contains a nested function, in which case it must be used by all the functions in the file.
- **error('message')** Displays the error message on the screen and terminates the m-file immediately.
- **echo** Turns echoing of statements in m-files on and off.
- **global** Defines a global variable (i.e., it can be shared between different functions and/or the workspace). Use global variables only when absolutely necessary — and it almost never is.
- **persistent** Defines a local variable whose value is to be saved between calls to the function.
- **return** Terminates the function or the script m-file immediately.
- **nargin** Number of input arguments supplied by the user.
- **nargout** Number of output arguments supplied by the user.
- **pause** Halts execution until you press some key.
- **varargin** "Groups" input arguments together.
- **varargout** "Groups" output arguments together.

---

**Where to Search for Functions**

- **path** View or change the search path.
- **addpath** Add directories to the current search path.

---

### 8.4. Odds and Ends

In MATLAB it is possible for a program to create or modify statements "on the fly", i.e., as the program is running. Entering

```
   >> eval('<string>)
```
executes whatever statement or statements are contained in the string. For example, entering

```
   >> a = 'x = linspace(0, 10, n); y = x.*sin(x).*exp(x/5); plot(x, y)';
   >> eval(a)
```
executes all three statements contained in the string \( a \). In addition, if an executed statement generates output, this is the output of \( \text{eval} \). For example, if we type

```
   >> A = zeros(5,6);
   >> [n, n] = eval('size(A)');
   then \( n = 5 \) and \( n = 6 \).
```
There is a very practical application for this function since it can combine a number of statements into one. For example, suppose we want to work with the columns of the Hilbert matrix of size \( n \) and we want to create variables to hold each column, rather than using \( H(:,1) \). We can do this by hand by typing

```
   >> C = H(:,1);
   >> C2 = H(:,2);
   ... which gets tiring very quickly. Instead, we can do this by typing
   >> for i = 1:n
        eval(['C' num2str(i) ' = H(:,i) '])
   end
```
This requires some explanation. We might be a little clearer if we separate the statement inside the for loop into two statements by

```
   >> s = ['C', num2str(i), ' = H(:,i) ']
   >> eval(s)
```
(we include commas in the first statement for readability). \( s \) is a text variable which contains \( c1=H(:,1) \) the first time the loop is executed, then \( c2=H(:,2) \) the second time, etc. (To understand how \( s \) is created, recall that \( s \) is really just a row vector with each element containing the ASCII representation of the corresponding character.)

Finally, there is a very esoteric application for this function that allows it to catch errors. This is similar to the "catch" and "throw" commands in C++ and Java. To use this feature of \( \text{eval} \), call it using two arguments as:

```
   >> eval('try_string', 'catch_string')
```
The function executes the contents of \( \text{try_string} \) and ignores the second argument if the execution succeeds. However, if it fails then the contents of \( \text{catch_string} \) are executed. (This might be a call to a function which can handle the error.) If there is an error, the command \( \text{lasterr} \) returns a string containing the error message generated by MATLAB.

A MATLAB function which is occasionally useful in a function is \( \text{feval} \). It executes a function, usually defined by an m-file, whose name is contained in a string by

```
   >> feval('string', x1, x2, ..., xn)
```
(See below for other ways to pass the function in the argument list.) Here \( x1, x2, ..., x_n \) are the arguments to the function. For example, the following two statements are equivalent

```
   >> A = zeros(5,6)
   >> A = feval('zeros', 5, 6)
```
Suppose that in the body of one function, say \( \text{sample} \), we want to execute another function whose name we do not know. Instead, the name of the function is to be passed as an argument to \( \text{sample} \). Then \( \text{feval} \) can be used to execute this text variable. For example, suppose in function \( \text{sample} \) we want to generate either linear or logarithmic plots. We can input the type of plot to use by
function sample(<type of plot>)
  ...
  feval(<type of plot>, x, y1, x, y2, '--, xx, y3, ':')
  ...
There are two common ways to pass the function <type of plot> in the argument list:
1. use a character string, e.g., feval('loglog',...), or
2. use a function handle, e.g., feval(@logval,...), or
Note: eval and feval serve similar purposes since they both evaluate something. In fact, feval can always be replaced by eval since, for example, feval('zeros',5,6) can always be replaced by eval(['zeros(5,6)']). However, there is a fundamental difference between them: eval requires the MATLAB interpreter to completely evaluate the string, whereas feval only requires MATLAB to evaluate an already existing function. feval is much more efficient, especially if the string must be evaluated many times inside a loop.

```
function m = hilb2(n)
J = 1:n; % J is a row vector
I = J'; % I is an n by n matrix with each column being 1:n
E = ones(n, n);
H = E./(I*J+1);
end
```

In the past this code ran nearly 20 times as fast as hilb_local. However, now hilb_local they are “roughly” comparable—depending on the computer and operating system. The reason is that MATLAB has greatly improved its handling of for and while statements. Thus, it is frequently not necessary to convert simple loops into complicated vector code.

As a realistic example of method (2), suppose you have a large vector y which is the discretization of a smooth function and you want to know some information about it. In particular, consider the intervals in y where y_i > R. What is the average length of these intervals and what is their standard deviation? Also, only include intervals which lie completely within y (i.e., ignore any intervals which begin or end y). It is not difficult to write such a code using control flow statements:

```
function H = hilb2(n)
J = 1:n; % J is a row vector
J = J(ones(n,1,:)); % J is now an n by n matrix with each row being 1:n
I = J'; % I is an n by n matrix with each column being 1:n
E = ones(n, n);
H = E./(I*J+1);
end
```

We have put “quickly” and “slowly” in quotes because this is quite subjective. Remember that your time is valuable: if it takes you longer to optimize your code than you will save in running it more quickly, still the urge to tinker around with it. Also remember that the amount of time it actually takes to optimize a code is usually a factor of two or three or... longer than the time you think it will take before you get started.

8.5. Advanced Topic: Vectorizing Code

As long as your MATLAB code executes “quickly”, there is no need to try to make it faster. However, if your code is executing “slowly”, you might be willing to spend some time trying to speed it up. There are three standard methods to speed up a code:

1. Preallocate matrices as shown in the function prealloc on page 101. This is very simple and very effective if the matrices are “large”.
2. Use MATLAB functions, whenever possible, rather than writing your own. If a MATLAB function is built-in, then it has been written in C and is faster than anything you can do. Even if it is not, much time has been spent optimizing the functions that come with MATLAB; you are unlikely to do better.
3. Replace control flow instructions with vector operations. We have already discussed this topic at length in Section 8.2. Here we will focus on some advanced techniques.

Of course, we should first determine how much CPU time is really being expended in our program—and that is the profile command. It will happily show you the percentage of CPU time spent in each line of each function in your program! And it will return the results in a very readable fashion! The statements
```
profile on
profile viewer
```
begin and end profiling and display the results in the Profiler window.

As a simple example of method (0), consider the function hilb on page 101. hilb_local(2000) runs over 300 times slower if the line m = zeros(n) is omitted. Continuing with this example, currently the MATLAB function hilb is written as

```
function H = hilb2(n)
J = 1:n; % J is a row vector
J = J(ones(n,1,:)); % J is now an n by n matrix with each row being 1:n
I = J'; % I is an n by n matrix with each column being 1:n
E = ones(n, n);
H = E./(I*J+1);
end
```

Odd and Ends

```matlab
eval
  Executes MATLAB statements contained in a text variable. Can also “catch” an error in a statement and try to fix it.
eval
  Executes a function specified by a string. (Can be used to pass a function name by argument.)
lasterr
  If eval “catches” an error, it is contained here.
```
function ylen_intvl = get_intervals_slow(y, R)
    n = length(y);
    if y(1) > R % check if the first point is in an interval
        in_intvl = 1; % yes
        yin = (intvl, nr) = 1;
    else
        in_intvl = 0; % no
        yintvl_nr = 0;
    end
    for i = 1:2: n % check the rest of the points
        if in_intvl == 1 % we are currently in an interval
            if y(i) <= R % check if this point is also in the interval
                yout = (intvl, nr) = 1; % no, so end the interval
                in_intvl = 0;
            else
                % we are currently not in an interval
                if y(i) > R % check if this point is in the next interval
                    yin = (intvl, nr) = intvl, nr + 1; % yes, so begin a new interval
                    in_intvl = 1;
                end
            end
        end
    end
    for i = 1:2: n % check if we have begun in an interval
        if y(i) > R % check if we have begun an interval
            yin(i) = []; % yes, so delete it
            yout(i) = [];
        end
    end
    ylen = length(yin) > length(yout) % check if we have ended in an interval
    yin(length(yin)) = []; % yes, so delete it
    ylen_intvl = yout - yin;
end

When completed, yin and yout contain the element numbers where an interval begins and where it ends, respectively. This is straightforward — but very slow if y has millions of elements.

To write a vectorized code, we have to think about the problem differently:

1. We do not care about the actual values in y, only whether they are greater than R or not. So we construct a logical matrix corresponding to y by
   yr = (y > R).
2. We do not actually care about the 0's and 1's — only about where the value changes because these mark the boundaries of the intervals. So we take the difference between adjacent elements of yr by
   yd = diff(yr).
3. We actually only need to know the elements which contain nonzero values so we find the element numbers by
   ye = find(yd). i.e., ye = find(yd=0).
4. We do not care about the actual locations of the beginning and end of each interval, only the lengths of these intervals. So we take the difference again by
   ylen = diff(ye).
5. Finally, ylen contains the lengths of both the intervals and the distances between successive intervals. So we take every other element of ylen. We also have to be a little careful and check whether y begins and/or ends in an interval.

Here is the code:

function ylen_intvl = get_intervals_fast(y, R)
    yr = (y > R); % (1)
    yd = diff(yr); % (2)
    ye = find(yd); % (3)
    ylen = diff(ye); % (4)
    if y(1) > R % (5), check if we begin in an interval
        ylen(1) = []; % yes
    end
    ylen_intvl = ylen(1:2:length(ylen)); % get every other length
end

Finally, the question remains: is the time savings significant? For “large” y the CPU time is reduced by over 20 (but this can vary greatly depending on y and R).

Note: In get_intervals_slow we did not preallocate the vectors yin and yout. Since we have no idea how many intervals there are, we have no way to preallocate these vectors to a “reasonable” size. We could preallocate them to a large size, say length(y)/4, and then strip out the unused elements at the end of the code. However, unless the number of intervals is in the tens of thousands, this will probably not save any time.

### Improving Efficiency of Function

profile Profile the execution time of a MATLAB code. This is very useful for improving the performance of a code by determining where most of the CPU time is spent.

### 9. Sparse Matrices

Many matrices that arise in applications only have a small proportion of nonzero elements. For example, if $\mathbf{T} \in \mathbb{C}^{n \times n}$ is a tridiagonal matrix, then the maximum number of nonzero elements is $3n-2$. This is certainly a small proportion of the total number of elements, i.e., $n^2$, if $n$ is “large” (which commonly means in the hundreds or thousands or ...)

For full matrices (i.e., most of the elements are nonzero) MATLAB stores all the elements, while for sparse matrices (i.e., most of the elements are zero) MATLAB only stores the nonzero elements: their locations (i.e., their row numbers and column numbers) and their values. Thus, sparse matrices require much less storage space in the computer. In addition, the computation time for matrix operations is significantly reduced because zero elements can be ignored.

Once sparse matrices are generated, MATLAB is completely responsible for handling all the details of their use: there are no special functions needed to work with sparse matrices. However, there are a number of functions which are inappropriate for sparse matrices, and MATLAB generally generates a warning message and refers you to more appropriate functions. For example, cond(S) has to calculate $S^{-1}$, which is generally a full matrix; instead, you can use condest which estimates the condition number by using Gaussian elimination. You have two alternatives: first, use full to generate a full matrix and use the desired function; or, second, use the recommended alternative function.

There are three common function in MATLAB for creating sparse matrices. The first is to use speye to create a sparse identity matrix instead of using eye which creates a full identity matrix. The second is to enter all the nonzero elements of $\mathbf{S} \in \mathbb{C}^{n \times n}$ individually by

\[ \mathbf{S} = \text{sparse}(i, j, s, m, n) \]

where i and j are vectors which contain the row and column indices of nonzero elements and s is the vector which contains the corresponding values. For example, the square bidirectional matrix

\[
\begin{bmatrix}
0 & -2 & -4 & -6 & \cdots & 0 \\
-2 & 0 & -4 & -6 & \cdots & 0 \\
-4 & -2 & 0 & -4 & \cdots & 0 \\
-6 & -4 & -2 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & -2 & -4 & -6 & \cdots & 0 \\
\end{bmatrix}
\]

In the above matrix, the nonzero elements are located at the positions specified by the vectors i and j, and the corresponding values are given by the vector s. The resulting matrix \( \mathbf{S} \) is sparse and can be used in various matrix operations, such as multiplication, addition, and solving linear systems, without the overhead associated with full matrices.
has the following nonzero elements

<table>
<thead>
<tr>
<th>i</th>
<th>j</th>
<th>s_{ij}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>n/2</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>n</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>n-2</td>
<td>3</td>
</tr>
<tr>
<td>n-1</td>
<td>n-1</td>
<td>2</td>
</tr>
<tr>
<td>n</td>
<td>n+1</td>
<td>n</td>
</tr>
</tbody>
</table>

A simple way to generate this matrix is by entering

\[ A = \text{sparse}(1:n, 1:n) \]

and check explicitly that \( A \) is exactly what is shown in the above matrix.

In addition, a full (or even an already sparse) matrix \( A \) can be converted to sparse form with all zero elements removed by

\[ A = \text{sparsify}(A) \]

Finally, a zero \( m \times n \) matrix can be generated by

\[ A = \text{sparse}(m, n) \]

which is short for

\[ A = \text{sparse}([], [], 0, m, n) \]

The third common function for generating sparse matrices is

\[ B = \text{spdiags}([B, \ldots, B], [n-1, n, \ldots, 1], m, n) \]

which works with entire diagonals of \( B \). \( B \) is a \( m \times n \) matrix and its \text{colnums} become the diagonals of \( B \) specified by \( d \in \mathbb{C}^n \). (For example, if \( d = (0:1)' \) then the first column of \( B \) contains the elements on the main diagonal and the second column contains the elements on the diagonal which is one above the main diagonal.) Thus, we can also generate the matrix \( B \) given above by

\[ B = \text{spdiags}([[B, \ldots, B], [n-1, n, \ldots, 1], m, n]) \]

\[ B = \text{spdiags}([[B, \ldots, B], [0:1]', [n-1:2:2*(n-2)]'), m, n] \]

Warning: Be Careful! The function \text{spdiags} is somewhat similar to \text{diag} but must be handled more carefully. Note that the element \( b_{1,3} \) is 0, which does not appear in \( B \). The difficulty is that the number of rows of \( B \) is generally larger than the lengths of the diagonals into which the columns of \( B \) are to be placed and so some padding is required in \( B \). The padding is done so that all the elements in the \( k^{th} \) row of \( B \) come from the \( k^{th} \) column of \( B \).

For example, the matrix

\[ B = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 \\
0 & 2 & 0 & 0 & 0 \\
0 & 0 & 3 & 0 & 0 \\
0 & 0 & 0 & 4 & 0
\end{bmatrix} \]

can be generated as a sparse matrix by

\[ B = \text{spdiags}([1:4], [-1]) + \text{diag}([6:8], 2) \]

or by

\[ B = \text{spdiags}([1:4], 0, 0, 0, [6:8])' \]

In the latter case note that the columns of \( B \) have to be padded with zeroes so that each column has five elements, whereas in the former case the vector which becomes the particular diagonal precisely fits into the diagonal. The element \( b_{1,3} \) of \( B \) contains the value 6. It appears in the \( 3^{rd} \) row of \( B \) because it occurs in the \( 3^{rd} \) column of \( B \). Note that the element \( b_{1,2} \) is not used since it would go into the element \( s_{n,n+1} \).
Before discussing how to solve Duffing’s equation specifically, we discuss the functions which solve time-evolution odes. To obtain a numerical solution to a time-evolution first-order ode system, enter

\[
\text{ode solver}(\text{function handle}, \text{tspan}, y0)
\]

or

\[
[t, y] = \text{ode solver}(\text{function handle}, \text{tspan}, y0)
\]

or

\[
sol = \text{ode solver}(\text{function handle}, \text{tspan}, y0)
\]

First, we have to choose which ode solver to use; this is discussed in detail shortly. It would be possible for MATLAB itself to decide which numerical method to employ. However, there are good reasons why the decision should be left in the hands of the user. Warning: Make sure you understand how to enter the name of the function handle. This is discussed at length in Section 3.2, and we also briefly discuss it below.

All of the solvers use the same input and output arguments, which we now discuss. The input parameters are:

- **Function**: The name of the function handle that calculates \( f(t, y) \).
- **Tspan**: The vector that specifies the time interval over which the solution is to be calculated. If this vector contains two elements, these are the initial time and the final time; in this case the ode solver determines the times at which the solution is output. If this vector contains more than two elements, these are the only times at which the solution is output. Note: the final time can be less than the initial time, in which case the trajectory is moving backwards in time.
- **Y0**: The vector of the initial conditions for the ode.

If there are no output parameters, the individual elements of the solution, \( y_1(t), y_2(t), \ldots, y_n(t) \), are plotted vs. \( t \) on a single plot. The circles on the trajectories show the actual times at which the solution is calculated.

If there are two output parameters, these are:

- \( t \) The column vector of the times at which the solution is calculated.\(^1\)
- \( Y \) The matrix which contains the numerical solution at the times corresponding to \( t \). The first column of \( Y \) contains \( y_1 \), the second column \( y_2 \), etc.\(^3\)

If there is one output parameter, then it is a structure. The output is now

\[
sol.x \quad \text{the column vector of the times at which the solution is calculated.}
\]

\[
sol.y \quad \text{the matrix which contains the numerical solution at the times corresponding to } t.
\]

There are seven distinct ode solvers which can be used, as shown in the table below. All these ode solvers use an adaptive step size to control the error in the numerical solution. Each time step is chosen to try to keep the local error within the prescribed bounds as determined by the relative error and the absolute error tolerances (although it does not always succeed). That is, \( e_r \), which is the error in \( y_1 \), is supposed to satisfy

\[
e_r \leq \max([\text{RelTol} \cdot |y_1|, \text{AbsTol}(1)])
\]

where the default value of RelTol is \( 10^{-3} \) and of the vector AbsTol is \( 10^{-6} \) for each element. (However, there is no guarantee that the error in the numerical calculation actually satisfies this bound.)

<table>
<thead>
<tr>
<th>DDE Solver</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ode45</td>
<td>Non-stiff ode solver; fourth-order, one-step method.</td>
</tr>
<tr>
<td>ode23</td>
<td>Non-stiff ode solver; second-order, one-step method.</td>
</tr>
<tr>
<td>ode113</td>
<td>Non-stiff ode solver; variable-order, multi-step method.</td>
</tr>
<tr>
<td>ode23s</td>
<td>Stiff ode solver; second-order, one-step method.</td>
</tr>
<tr>
<td>ode23t</td>
<td>Stiff ode solver; trapezoidal method.</td>
</tr>
<tr>
<td>ode23tb</td>
<td>Stiff ode solver; second-order, one-step method.</td>
</tr>
</tbody>
</table>

\(^1\)The \( t \) in \( [t, y] \) is unrelated to the \( t \) argument in the function **duffing**.

\(^3\)We have capitalized the \( y \) in \( \text{tspan}, y0 \) to indicate that the output is a matrix whereas the argument \( y \) is a vector in the function.
It is up to you to decide which ode solver to use. As a general rule, unless you believe that the ode is stiff (which we discuss in the next paragraph), try ode45 or ode113. For a given level of accuracy, these methods should run “reasonably fast.” (Which one runs faster is very dependent on the ode.) If you know (or believe) that the ode is stiff, or if these two non-stiff solvers fail, then try ode15s.

And what is a stiff ode? There is no precise definition. Instead, we say it is stiff if the time step required to obtain a stable and accurate solution is “unreasonably small.” The best way to explain this rather vague impression is through some simple examples.

Consider the second-order time-evolution ode

$$y'' + 999y' + 1000y = 0 \quad \text{for } t \geq 0$$

with the initial conditions $y(0) = \eta_1$ and $y'(0) = \eta_2$. The solution to this ode is

$$y(t) = c_1 e^{\eta_1 t} + c_2 e^{-1000t}$$

where

$$c_1 = \frac{1}{1001}(\eta_1 - \eta_2) \quad \text{and} \quad c_2 = \frac{1}{1001}(1000\eta_1 + \eta_2).$$

There are two time scales in this solution: there is a rapid decay due to the $e^{-1000t}$ term and there is a slow growth due to the $e^{\eta_1 t}$ term. Initially, the time step will be “very small” so that the rapid decay is calculated accurately (i.e., $\Delta t \ll \frac{1}{1001}$). However, soon it will be negligible and the time step should increase so that it calculates the slow growth accurately (i.e., $\Delta t \ll 1$). However, if a non-stiff solver, such as ode45 or ode23, is used, the time step must always be “very small.” That is, it must accurately track the rapidly decaying term — even after this term has disappeared in the numerical solution. The reason is that a numerical instability will cause the trajectory to blow up if the time step increases. However, if a stiff solver is used, the time step can increase by many orders of magnitude when the rapidly decaying term has disappeared.

The same is true for the ode

$$y'' + 1011y' + 1000y = 0$$

whose solution is

$$y(t) = c_1 e^{t} + c_2 e^{-1000t}.$$  

Initially, the time step will be “very small” so that the rapid decay is calculated accurately (i.e., $\Delta t \ll \frac{1}{1001}$). However, soon it will be negligible and the time step should increase so that it calculates the slowly decaying term accurately (i.e., $\Delta t \ll 1$).

On the other hand, consider the ode

$$y'' - 1001y' + 1000y = 0$$

whose solution is

$$y(t) = c_1 e^{t} + c_2 e^{1000t}.$$  

The time step must always be “very small” so that the rapidly growing mode $e^{1000t}$ is calculated accurately (i.e., $\Delta t \ll \frac{1}{1000}$). Thus, this is not a stiff ode.

The above examples are very simple. They are only designed to show that an ode is stiff if there is a rapidly decaying mode and any growth in the solution occurs on a much slower time scale. (This frequently happens in chemical reaction models, where some reactions occur on a very fast time scale and other occur on a much slower time scale.) In the next subsection we discuss van der Pol’s equation, a second-order ode which is either non-stiff or stiff depending on the value of one parameter. You can plot the solution and observe the separation of the fast scale and the slow scale as this parameter increases.

One difficulty with a stiff ode solver is that you might have to supply the Jacobian of the ode yourself if the ode is **STIFFLY** **STATY**: The Jacobian of $f(t, y)$ is the $n \times n$ matrix

$$J(t, y) = \left( \frac{\partial f_i}{\partial y_j}(t, y) \right),$$

i.e., the element in the $i^{th}$ row and $j^{th}$ column of $J$ is

$$\frac{\partial f_i}{\partial y_j}.$$  

Any of the stiff methods can approximate this matrix numerically. However, if the ode is “bad” enough, this may not be enough. You may have to calculate all these partial derivatives yourself and include them in your function file. (We show an example of this later.)

The reason for this large choice of ode solvers is that some odes are very, very, very **PARTY**. It is possible that most of the ode solvers will fail and only one, or maybe two, will succeed. SAY MORE??

To conclude this subsection, we return to Duffing’s equation. Suppose we want to solve the ode for $t \in [0, 100]$ with initial conditions $y = (2, 1)^T$ and plot the results. Since this is a very well-behaved ode for the parameters given, we can use ode45. The simplest approach is to use an anonymous function to input the right-hand side:

- $\alpha = 0.05$;
- $\beta = 1.0$;
- $\Gamma = 0.5$;
- $\omega = 1.0$;
- $\text{duffing}_a = @(t, y)\{y(2); y(1)\{1-beta*y(1)^2-\alpha*y^2(2)+\Gamma*y+\omega*\cos(\omega*t)\};$
- $\text{ode45}(\text{duffing}_a, [0 100], [2 1]);$

(The ‘a’ denotes the fact that $\text{duffing}_a$ is an anonymous function handle.) The solution will now be plotted as $y_1$ and $y_2$ vs. $t$. (This plot is rather “chattered” because, not only is the trajectory plotted, but in addition markers are put at each of the points of the numerical solution.)

**Warning:** There are a number of parameters which are needed by the function and these must be defined before the function is created. Also, the function handle $\text{duffing}_a$ will always use these parameters, even if they are later changed.

**Note:** Since $\text{duffing}_a$ is already a function handle, we merely need to use its name as the first argument to ode45.

To obtain complete control over what is plotted, you should let ode45 output the trajectory and do the plots yourself. This is easily accomplished by changing the last line of the previous code to

- $\{t, y\} = \text{ode45}(\text{duffing}_a, [0 100], [2 1]);$
- $\text{figure}(1)$
- $\text{subplot}(2, 1, 1)$
- $\text{plot}(t, y(:,1))$
- $\text{subplot}(2, 1, 2)$
- $\text{plot}(t, y(:,2))$
- $\text{figure}(2)$
- $\text{plot}(y(:,1), y(:,2))$

This results in a plot of $y_1$ vs. $t$ and a separate plot of $y_2$ vs. $t$, so that both plots are visible even if they have vastly different scales. There is also a separate plot of $y_2$ vs. $y_1$, which is called a phase plane.

The next simplest approach is to use a nested function, and so there must also be a primary function.

```
function duffing_ode(alpha, beta, Gamma, omega, y0, final_time)
\text{ode45}(\text{duffing}_a, [0 final_time], y0);
\text{\texttt{XXXX nested function} follows}
\text{function deriv = duffing_a(t, y)}
\text{deriv} = \{ y(2); y(1)\{1-beta*y(1)^2-\alpha*y^2(2)+\Gamma*y+\omega*\cos(\omega*t)\} \};
\text{end$
```

(The ‘a’ denotes the fact that $\text{duffing}_a$ is a nested function.) Note that the parameters are input to the primary function and so are immediately accessible to the nested function. Clearly, this second approach (of using a nested function) requires more coding that the first approach (of using an anonymous function). However, the first approach only works if the right-hand side can be defined using one MATLAB statement. If the right-hand side is more complicated, then a nested function is the simplest choice.

**Note:** Since $\text{duffing}_a$ is a function, and not a function handle, we have to include “a” before the name of the function.

The third, and oldest, approach is to create a separate function m-file (i.e., a primary function) which calculates the right hand side of this ode system.
function deriv = duffing_p(t, y)
% duffing_p: Duffing's equation, primary function
alpha = 0.06;
beta = 1.0;
Gamma = 0.8;
omega = 1.0;
deriv = [ y(2) ; y(1)*(1-beta*y(1)^2-alpha*y(2)+Gamma*cos(omega*t) ) ];
% % deriv = [ y(2) ; y(1)*(1-beta*y(1)^2-alpha*y(2)+Gamma*cos(omega*t) ) ];
% (This "p" denotes the fact that duffing_p is a primary function.) Note that all the parameters are defined in the m-file so that it will have to be modified whenever we want to modify the parameters. This is a very bad approach because this file will have to be repeatedly modified.

Note: Since duffing_p is a function, and not a function handle, we have to include "@" before the name of the function.

Finally, it is very inconvenient that the parameters in Duffing's equation are determined in the function itself. We should be able to "explore" the rich behavior of Duffing's equation without having to constantly modify the function — in fact, once we have the function exactly as we want it, we should never touch it again. (This is not only true for esthetic reasons; the more we fool around with the function, the more likely we are to screw it up!)

This is easily done by adding parameters to the function file:

function deriv = duffing_p2(t, y, alpha, beta, Gamma, omega)
% duffing_p2: Duffing's equation, primary function
% with coefficients passed through the argument list
deriv = [ y(2) ; y(1)*(1-beta*y(1)^2-alpha*y(2)+Gamma*cos(omega*t) ) ];
% (This "p" denotes the fact that duffing_p2 is another primary function.) However, this function cannot be called directly by the ode solver. Instead it is called indirectly using closure by
% closure function
alpha = 0.06;
beta = 1.0;
Gamma = 0.8;
omega = 1.0;
duffing_c = @(t, y) duffing_p2(t, y, alpha, beta, Gamma, omega);
ode45(duffing_c, [0 100], [1 1]);
% (which is contained in the accompanying zip file). Notice that the function duffing_c takes only two arguments: t and y. But the function it invokes is duffing_p2 which takes six arguments. Thus, node45 thinks it is only passing two arguments to duffing_c, but it is actually passing six arguments to duffing_p2.

To see a sampling of the different type of behavior in Duffing's equation, let $\alpha = 0.15$, $\beta = 1$, $\Gamma = 0.3$ and $\omega = 1$, and let the initial condition be $y(0) = (0,1)^T$. After a short initial transient, the solution settles down and appears to be "regular" by $t = 100$; in fact, it appears to be exactly periodic with a period of 2$t$ due to the $0.3 cos t$ term. (In fact, to the accuracy of the computer it is exactly periodic.) However, if we merely change the initial condition to $y(0) = (1,0)^T$ the behavior appears to be chaotic, even at $t = 1000$. Here is an example of a saddle which has periodic motion for one initial condition and is chaotic for another! If we now change $\alpha$ from 0.15 to 0.22 we find periodic motion with a period of 6$t$. This is just a sampling of the behavior of Duffing's equation in different parameter regions.

By the way, to separate the initial transient behavior from the long-time behavior, you can use the script m-file:

```
initial_time = ???
final_time = ???
y0 = ???
alpha = ???;
beta = ???;
Gamma = ???;
omega = ???;
duffing_a = @(t, y)[ y(2) ; y(1)*(1-beta*y(1)^2-alpha*y(2)+Gamma*cos(omega*t) ) ];
[t, Y] = ode45(duffing_a, [0 initial_time], y0);
figure(1)
plot(Y(:,1), Y(:,2));
figure(2)
plot(Y(:,1), Y(:,2));
```

10.2. Advanced Functions

There are a number of parameters that we can use to "tune" the particular ode solver we choose. The MATLAB function ode45 is used to change these parameters from their default values by

```matlab
params = ode45(ode45 function name, [tspan, y0, params]);
```

where each parameter has a particular name and it is followed by the desired value. The result of this function is that the parameters are contained in the variable params. You include these parameters in the ode solver by adding this variable to the argument list of the ode solver function as

```matlab
[t, Y] = ode45(ode solver function handle, tspan, y0, params);
```

Some of the more common parameters are shown in the table below; they will be discussed further later.

To determine all the parameters, their possible values and the default value, enter

```matlab
ode45set
```
The absolute error tolerance. This can be a scalar in which case it applies to all the elements of \( y \) or it can be a vector where each element applies to the corresponding element of \( y \). (Default value: \( 10^{-6} \)).

Events A handle to a function which determines when an event occurs.

Jacobian A handle to a function which returns the Jacobian.

JPattern A sparse matrix whose nonzero elements (which should be 1) correspond to the possible nonzero elements of the Jacobian. This is only used when the Jacobian is calculated numerically, i.e., when the Jacobian property is not used.

OutputFcn A handle to a function which is called after each successful time step. For example, a plot of the trajectory can be generated automatically as it is being calculated. Useful MATLAB functions are:

- \( @odeplot \) which generates a plot of time versus all the components of the trajectory, i.e. \( t \) vs. \( y_1, y_2, \ldots, y_n \).
- \( @odephas \) which generates a plot of a vector of \( y_j \) vs. \( y_k \), i.e. \( Y(:,1) \) vs. \( Y(:,2) \).
- \( @odephas2 \) which generates a plot of \( y_j \) vs. \( y_k \) vs. \( y_l \), i.e. \( Y(:,1) \) vs. \( Y(:,2) \) vs. \( Y(:,3) \).

It is possible to plot different components of \( y \) using OutputSel.

OutputSel A vector containing the components of \( y \) which are to be passed to the function specified by the OutputFcn parameter.

Refine Refines the times which are output in \( t \). This integer value increases the number of times by this factor. (Default value: 1 for all ode solvers except \( \text{ode}45 \) for \( \text{ode}45 \).)

RelTol The relative error tolerance. (Default value: \( 10^{-3} \)).

Stats Whether statistics about the run are output on the terminal (value: ‘on’) after the trajectory is calculated or they are not (value: ‘off’). (Default value: ‘off’).

For example, if you want to use \( \text{ode}45 \) with the relative error tolerance set to \( 10^{-6} \) for Duffing’s equation, enter:

```matlab
params = odeset(’RelTol’, 1e-6); % Set the relative tolerance to 10^-6
[t, y] = ode45(@duffing, tspan, y0, params); % Solve Duffing’s equation with the specified parameters
```

The trajectory will be more accurate — but the function will run slower. If you also want the statistics on the performance of the particular ode solver used, enter:

```matlab
params = odeset(’RelTol’, 1e-6, ’Stats’, ’on’); % Set the relative tolerance and enable statistics
[t, y] = ode45(@duffing, tspan, y0, params); % Solve Duffing’s equation with the specified parameters
```

and the number of successful steps, the number of failed steps, and the number of times \( f(t) \) was evaluated will be printed on the terminal. This might be useful in “optimizing” the performance of the ode solver if the function seems to be running excessively slowly. For implicit methods where the Jacobian needs to be calculated, the number of times the Jacobian was evaluated, the number of LU decompositions, and the number of times the linear system was solved will also be returned.

The ode solver can also record the time and the location when the trajectory satisfies a particular condition: this is called an event. For example, if we are calculating the motion of the earth around the sun, we can determine the position of the earth when it is closest to the sun and/or farthest away; or, if we are following the motion of a ball, we can end the calculation when the ball hits the ground — or we can let it continue bouncing. Enter:

```matlab
balleod % to see a simple example
```

For example, suppose we want to record where and when a trajectory of Duffing’s equation passes through \( y_1 = \pm 0.5 \). That is, we define an ‘event’ to be whenever the first component of \( y \) passes through \(-0.5\) or \(+0.5\). This can be done by modifying the primary function \text{duffing ode} and replacing the \text{ode45} statement by:

```matlab
params = odeset(’RelTol’, 1e-6, ’Events’, @duffing_event); % Set the relative tolerance and define the event function
[t, y] = ode45(@duffing, tspan, y0, params); % Solve Duffing’s equation with the specified parameters
```

where \( \text{function} \) is a new nested function in the primary function \text{duffing ode}.

```matlab
function [value, isterminal, direction] = duffing_event(t, y)
value = (y(1)+0.5); % check whether y(1) passes through ±0.5
  if (y(1) > 0.5) % % do not halt when this occurs
    isterminal = [0; 0]; direction = [0; 0]; % an event occurs when y(1) passes through ±0.5
  end % zero in either direction
end
```

Note that we can define the right-hand side of Duffing’s equation by using \( \text{duffing_a} \) and \( \text{duffing_p} \) and \( \text{duffing_p2} \) and \( \text{duffing_c} \). We have chosen \( \text{duffing a} \) since we have created the nested function \text{duffing_event}. (We could let \text{duffing_event} be a primary function, but there is no reason to do so.)

There are a number of steps we have to carry out to turn “events” on. First, we have to use the \text{odeset} function. However, this only tells the ode solver that it has to watch for one or more events; it does not state what event or events to watch for. Instead, we describe what an event is in this new function. Three vector arguments are output:

- \( \text{value} \) — A column vector of values which are checked to determine if they pass through zero during a time step. No matter how we describe the event, as far as the ode solver is concerned an event only occurs when an element of this vector passes through zero. In some cases, such as this example it is easy to put an event into this form. In other cases, such as determining the apogee and perigee of the earth’s orbit, the calculation is more complicated.
- \( \text{isterminal} \) — A column vector determining whether the ode solver should terminate when this particular event occurs. \( 1 \) means yes and \( 0 \) means no.
- \( \text{direction} \) — A column vector determining how the values in \( \text{value} \) should pass through zero for an event to occur:
  - \( 1 \) means the value must be increasing through zero for an event to occur,
  - \( -1 \) means the value must be decreasing through zero for an event to occur, and
  - \( 0 \) means that either direction triggers an event.

The final step is that the left-hand side of the calling statement must be modified to:

```matlab
[t, y, tevent, yevent, indexevent] = ode45(…)
```

Any and all events that occur are output by the ode solver through these three additional variables:

- \( \text{tevent} \) is a vector containing the time of each event.
- \( \text{yevent} \) is a matrix containing the location of each event, and
- \( \text{indexevent} \) is a vector containing which value in the vector \( \text{value} \) passed through zero.

If the result is stored in the structure \( \text{sol} \), the new output is:

- \( \text{sol.xe} \) is a vector containing the time of each event,
- \( \text{sol.ye} \) is a matrix containing the location of each event, and
- \( \text{sol.ie} \) is a vector containing which value in the vector \( \text{value} \) passed through zero.

Since the function \text{duffing_event} might appear confusing, we now discuss how an event is actually calculated. At the initial time, \( t \) and \( y \) are known and \( \text{duffing_event} \) is called so that the vector \( e^{(0)} = (0.5, 0.5) \), i.e., \( \text{value} \) is calculated. In addition, \( \text{isterminal} \) and \( \text{direction} \) are returned. Next, \( \text{duffing} \) is called and the solution \( y^{(1)} \) is calculated at time \( t^{(1)} \). \( \text{duffing_event} \) is called again and \( e^{(1)} \) is calculated and compared elementwise to \( e^{(0)} \). If the values have different signs in some row, then \( \text{direction} \) is checked to determine if the values are passing through zero in the correct direction or if either direction is allowed. If so, the time at which the element is zero is estimated and the ode is solved again to obtain a more accurate estimate. This procedure continues until the zero is found to the desired accuracy. Then \( \text{isterminal} \) is checked to see if the run should be continued or should be stopped.

Another interesting ode is van der Pol’s equation

\[
y'' - \mu (1 - y^2) y' + y = 0
\]

where \( \mu > 0 \) is the only parameter. As a first order system it is

\[
\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} y_2 \\ (\mu (1 - y^2) y_2 - y_1) \end{bmatrix}
\]
and its Jacobian is
\[
J = \begin{pmatrix}
0 & 1 \\
-2y & -1
\end{pmatrix}.
\]

The right-hand side can be coded as a nested function inside a primary function by

```python
function vdp ode(speed, angle, height)
    % gravity_ode. The trajectory of a ball thrown from (0,0) with initial % speed and angle (in degrees) given.
    % times: (1) = time ball at peak, (2) = time ball hits ground % values: (1) = z value at peak, (2) = x value when ball hits ground % (3,4) = x value when ball reaches height % note: (3,4) will not be used if height > z value at peak % g = 9.8; % gravity_init()
    [t, y, event, Yevent, index_event] = ode45(@gravity_ode, tspan, y0, params);
    if length(event) == 2 % times = event;
        values = [Yevent(1,3) Yevent(2,1)];
    else
        times = event([2 4 1 3]); % values = [Yevent(2,3) Yevent(4,1) Yevent(1,1) Yevent(3,1)];
end

%%%% nested functions follow %
function gravity_init()
    % gravity_init: Initialize everything
    tspan = [0 10];speed=[g];
    y0 = [0; speed*cos(angle*pi/180); 0; speed*sin(angle*pi/180)];
    params = odeset('RelTol', 1e-6, ...
        'Events', @gravity_event, ...
        'Refine', 20, ...
        'OutputFcn', @odephas2, ...
        'OutputSel', [1 3],);
end
function deriv = gravity(t, y)
% gravity: Calculates the right-hand side of the ode deriv = [y(2); 0; y(4); -y(1)];
end
function [value, isterminal, direction] = gravity_event(t, y)
% gravity_event: determines the events value = [y(3); y(3)-height; y(4)]; % z = 0, z-height = 0, z' = 0 isterminal = [1; 0; 0]; % halt only when z = 0 direction = [-1; 0; -1]; % an event occurs when z or z' decrease through 0 % or z-height passes through 0 in either direction end
```

One complication with solving this system numerically is that we do not know when the ball will hit the ground, so we cannot give the final time. Instead, we use a time, 10s/g which is much greater than needed and we let the program stop itself when the ball hits the ground. In addition, we want the relative error to be 10^-6. Finally, we want the trajectory (i.e. z vs. x) to be plotted automatically.

The following is a completely self-contained example using nested functions.

```python
function [times, values] = gravity_ode(speed, angle, height)
    % gravity_ode. The trajectory of a ball thrown from (0,0) with initial % speed and angle (in degrees) given.
    % times: (1) = time ball at peak, (2) = time ball hits ground % values: (1) = z value at peak, (2) = x value when ball hits ground % (3,4) = x value when ball reaches height % note: (3,4) will not be used if height > z value at peak % g = 9.8; % gravity_init()
    [t, y, event, Yevent, index_event] = ode45(@gravity_ode, tspan, y0, params);
    if length(event) == 2 % times = event;
        values = [Yevent(1,3) Yevent(2,1)];
    else
        times = event([2 4 1 3]); % values = [Yevent(2,3) Yevent(4,1) Yevent(1,1) Yevent(3,1)];
end

%%%% nested functions follow %
function gravity_init()
    % gravity_init: Initialize everything
    tspan = [0 10];speed=[g];
    y0 = [0; speed*cos(angle*pi/180); 0; speed*sin(angle*pi/180)];
    params = odeset('RelTol', 1e-6, ...
        'Events', @gravity_event, ...
        'Refine', 20, ...
        'OutputFcn', @odephas2, ...
        'OutputSel', [1 3],);
end
function deriv = gravity(t, y)
% gravity: Calculates the right-hand side of the ode deriv = [y(2); 0; y(4); -y(1)];
end
function [value, isterminal, direction] = gravity_event(t, y)
% gravity_event: determines the events value = [y(3); y(3)-height; y(4)]; % z = 0, z-height = 0, z' = 0 isterminal = [1; 0; 0]; % halt only when z = 0 direction = [-1; 0; -1]; % an event occurs when z or z' decrease through 0 % or z-height passes through 0 in either direction end
```

Note that the parameters \( g \), \( speed \), \( angle \), and \( height \) do not need to be passed into the nested functions. Similarly, \( tspan \), \( y0 \), and \( params \) do not need to be passed out. 

Analytically: We convert Newton’s laws to the first-order system

\[
\begin{pmatrix}
y_1' \\
y_2' \\
y_3' \\
y_4'
\end{pmatrix} =
\begin{pmatrix}
y_2 \\
-g \\
y_4 \\
y_3 - y_4^2 \\
\end{pmatrix},
\]

by letting \( y_1 = x, y_2 = y_1, y_3 = y_2, \) and \( y_4 = z' \). The initial conditions are \( y_1(0) = 0, y_2(0) = x \cos \alpha, y_3(0) = 0, \) and \( y_4(0) = x \sin \alpha \).
MATLAB also has the function ode15s which solves fully implicit odes. It is very similar to the functions we have already discussed, but there is one important difference. Although it is a very powerful function, we only provide a very simple example which uses it.

We consider a linear second-order ode in a neighborhood of a regular singular point. Consider the ode

\[ P(t)y''(t) + Q(t)y'(t) + R(t)y(t) = 0 \]

where \( P(t) \), \( Q(t) \), and \( R(t) \) are polynomials with no common factors. The singular points of this ode are the values of \( t \) for which \( P(t) = 0 \). If \( t_0 \) is a singular point, it is a regular singular point if \( \lim_{t \to t_0^-}(t - t_0)Q(t)/P(t) \) and \( \lim_{t \to t_0^+}(t - t_0)^2R(t)/P(t) \). A "common" ode of this type is Bessel’s equation

\[ t^2y''(t) + ty'(t) + (t^2 - n^2)y(t) = 0 \]

where \( n \) is a nonnegative integer and the initial condition is given at \( t = 0 \). The solution is denoted by \( J_n(t) \) and, for specificity, we will concentrate on \( n = 1 \). At \( t = 0 \) the ode reduces to \(-y(0) = 0\) and so we require \( y(0) = 0 \). The free initial condition is \( y'(0) = 1 \). If we write Bessel’s equation as

\[ y''(t) + \frac{1}{t} y'(t) + \left(1 - \frac{n^2}{t^2}\right) y(t) = 0 \]

we clearly have a problem at \( t = 0 \) and for \( t \approx 0 \). The ode solvers we discussed previously can handle (10.2) for \( t \geq 1 \) with the initial conditions that \( y(1) \) and \( y'(1) \) are given. However, a completely different method of solution is required for \( t \geq 0 \) and the form (10.1) is preferred to (10.2).

When we convert Bessel’s equation to the first order system we again let \( \gamma_1(t) = y(t) \) and \( \gamma_2(t) = y'(t) \) and leave the \( t^2 \) in the numerator to obtain

\[
\begin{pmatrix}
\frac{d\gamma_1}{dt} \\
\frac{d\gamma_2}{dt}
\end{pmatrix} = 
\begin{pmatrix}
-\frac{\gamma_2}{t} \\
-t\gamma_2 + \left(1 - \frac{n^2}{t^2}\right) \gamma_1
\end{pmatrix}.
\]

Previously, we have always written the first-order system as \( \gamma' = f(t, \gamma) \), but this form has a problem when \( t = 0 \). Instead, we write it as \( g(t, y, \gamma') = 0 \) so that

\[
g(t, y, \gamma') = 
\begin{pmatrix}
\frac{\gamma_1}{t} \\
\frac{\gamma_2}{t} + \gamma_1 + \left(1 - \frac{n^2}{t^2}\right) \gamma_1
\end{pmatrix} = 0.
\]

Finally, we not only have to input the initial condition \( \gamma(0) = (0, 1)^T \), but we also have to input \( \gamma'(0) = (y(0), y'(0)) \). It is easy to calculate \( y'(0) = y_0(0) \), but \( y'(0) \) is more complicated. Differentiate (10.1) with respect to \( t \) to obtain

\[ t^2y'''(t) + 3ty''(t) + 2ty'(t) = 0 \]

and differentiate it again to obtain

\[ t^2y''''(t) + 5ty'''(t) + (t^2 + 3)yt''(t) + 4ty'(t) + 2ty(t) = 0. \]

Now set \( t = 0 \) to obtain \( y''(0) = 0 \). We can solve Bessel’s equation for \( t \in [0, 1] \) by

```
% g = @(t, y, yp) [yp(1)-y(2); t^2 * y(2) + y(2)*t + (t-1)*y(1)];
% tspan = [0 1];
% y0 = [0; 1];
% yp0 = [1; 0];
% [t, Y] = odes15i(g, tspan, y0, yp0);
% plot(t, Y(:,1))
```

### 11. Boundary-Value Ordinary Differential Equations

In addition to initial-value ordinary differential equations there is a second type of odes that MATLAB can solve numerically. Boundary-value odes are also odes of the form

\[
d\frac{dy}{dx} = f(x, y) \quad \text{for} \quad x \in [a, b]
\]

where \( y \in \mathbb{R}^n \) but conditions are given at both ends of the interval. If the boundary conditions are separated, then \( k \) conditions are given at \( x = a \) and \( n - k \) other conditions are given at \( x = b \). If the boundary conditions are non-separated, then the conditions at \( x = a \) and \( x = b \) are related. To allow any of these boundary conditions we write the boundary conditions as \( \phi(y(a), y(b)) = 0 \) where \( \phi \in \mathbb{R}^n \).

For simplicity, we will only consider two closely related second-order odes, i.e., \( n = 2 \). This example should enable you to study any boundary-value ode. Consider the two nonlinear boundary-value ordinary differential equations

\[
d\frac{dy}{dx} = y, \quad \frac{d^2y}{dx^2} + 2\frac{dy}{dx} - y^4 = 0 \quad (11.1a)
\]

\[
d\frac{dy}{dx} = y, \quad \frac{d^2y}{dx^2} + 3\frac{dy}{dx} - e^{y(x)} = 0 \quad (11.1b)
\]

for \( x \in [0, 1] \) where \( e > 0 \). Our boundary conditions are

\[
\phi(y(0), y(1)) = \begin{pmatrix} y(0) - 1 \\ y(1) - 1 \end{pmatrix} = 0.
\]

which are called Dirichlet boundary conditions. These two odes are quite simple, but also quite interesting and challenging to solve for certain intervals in \( e \). We could use the Neumann boundary conditions \( y'(0) = 4 \) and \( y'(1) = -7 \) by

\[
\phi(y(0), y(1)) = \begin{pmatrix} y(0) - 4 \\ y(1) + 7 \end{pmatrix} = 0.
\]

Or we could use the mixed boundary conditions \( y(0) - y'(0) = 1 \) and \( y(1) + 2y'(1) - 3 \) by

\[
\phi(y(0), y(1)) = \begin{pmatrix} y(0) - y'(0) - 1 \\ y(1) + 2y'(1) - 3 \end{pmatrix} = 0.
\]

Finally, we could use periodic boundary conditions, which are non-separated, by

\[
\phi(y(0), y(1)) = \begin{pmatrix} y(1) - y(0) \\ y(1) - y(0) \end{pmatrix} = 0.
\]

The primary MATLAB function is bvp4c. However, the functions bvpinit and deval are also needed. We solve the boundary value problem by

```
% sol = bvp4c([right-hand side], <boundary conditions>, <initial guess>))
```

There are two functions we need to write: odefun is \( f(x, y) \) and bcfun is the boundary conditions. For our example the ode given by

```
function yp = nlsode(x, y)
    global which_ode eps
    if which_ode == 1
        yp = [y(2); -eps*exp(y(1))-2*y(2)];
    else
        yp = [y(2); -exp(y(1))-2*y(2)/eps];
    end
end
```

where we use global to input which odes to use and \( \epsilon \). The boundary condition is given by

```
function bc = nlsode_bc(ya, yb)
    bc = [ya(1); yb(1)];
```

```
Since these boundary conditions are particularly simple, we also include the function
function bc = nnode_bc2(ya, yb)
bc = [ya(1)-ya(2)-1; yb(1)+2*yb(2)-3];
end

for mixed boundary conditions (11.4). In addition, we have to choose an initial guess for y(x) using
bvpinit by either
> bvpinit(x, y_init)
or
> bvpinit(x, <initial guess function>)

For example, if we want the initial iterate to be a parabola which is zero at \( x = 0 \) and has maxi-

mum value \( A \) then \( y(x) = y_1(x) = 4Ax(1-x) \) and \( y'(x) = y_2(x) = 4A(1-2x) \) then we can write
> x = linspace(0, 1, 21);
solinit = bvpinit(x, nnode_y_ic);

where \( \text{nnode}_y\_ic \) is written as
function y_ic = nnode_y_ic(x)
global A
y_ic = [4*A*x(1-x); 4*A*(1-2*x)];
The only alternative is to write
> x = linspace(0, 1, 21);
> y1_val = 777;
> y2_val = 777;
solinit = bvpinit(x, [y1_val; y2_val]);

where \( y1_{\_val} \) and \( y2_{\_val} \) are scalar values. Thus the initial guess is \( y_1 = y_1_{\_val} = \text{ones} \_size(x) \) and \( y_2 = y_2_{\_val} \_size(x) \). This is often unacceptable because constant initial guesses may be so far from
the solution that convergence cannot be obtained. What we would like to do is
> x = linspace(0, 1, 21);
> y1 = 4*Ax(1-x);
> y2 = 4*A*(1-2x);
solinit = bvpinit(x, [y1; y2]); % WRONG

This fails because \( y_1 \) and \( y_2 \) must be scalar variables and not vectors. If you really really need \( y_1 \) and
\( y_2 \) to be vectors, then do not use bvpinit. Instead, specify the structure solinit directly by entering
> x = linspace(0, 1, 21);
> y1 = 4*A*x(1-x);
> y2 = 4*A*(1-2x);
> solinit = bvpinit(x, [y1; y2]);
solinit.x = x;
solinit.y = [y1; y2];

Warning: This is dangerous because future versions of Matlab might change the fieldnames of the struc-
ture solinit. However, it works for now.

We are finally ready to solve this ode by

\[
\text{function } \text{sol} = \text{nnode}\_all(\text{which\_ode}, \text{e}, A, \text{nr\_points}) \\
\text{nnode}\_all: \text{boundary-value solver using bvp4c} \\
\text{which\_ode} = 1 \quad \text{y'} + 2 \quad \text{y} + \text{e} \cdot \text{exp}(y) = 0, \quad y(0) = y(1) = 0 \\
\text{which\_ode} = 2 \quad \text{y} + 2 \quad \text{y} + \text{e} \cdot \text{exp}(y) = 0, \quad y(0) = y(1) = 0 \\
A = \text{the initial guess is } y = 4A \cdot x (1-x) \\
\text{nr\_points} = \text{number of equally spaced points in initial guess} \\
x = \text{linspace}(0, 1, \text{nr\_points}); \\
solinit = bvpinit(x, nnode_y_ic); \\
sol = bvp4c(solnode, \text{nnode}_bc, \text{solinit}); \\
xp = \text{linspace}(0, 1, 101); \\
yp = \text{deval}(sol, xp);
\]

Having done all this work, we now combine everything into the function m-file nnode\_all to show how
much easier it is to use nested functions and to combine everything into one primary function.

function sol = nnode_all(which_ode, e, A, nr_points)

% nnode_all: boundary-value solver using bvp4c
% which_ode = 1 \ y' + 2 \ y + e \cdot \exp(y) = 0, \ y(0) = y(1) = 0
% which_ode = 2 \ y + 2 \ y + e \cdot \exp(y) = 0, \ y(0) = y(1) = 0
% A = the initial guess is y = 4A \cdot x (1-x)
% nr_points = number of equally spaced points in initial guess
% x = linspace(0, 1, nr_points);
solinit = bvpinit(x, nnode_y_ic);
sol = bvp4c(solnode, nnode_bc, solinit);
xp = linspace(0, 1, 101);
yp = deval(sol, xp);
plot(xp, yp(1,:), xp, yp(2,:), ’r’)
and we want to find the "best" fit by a straight line. To define the data points more abstractly as

\[ f(x) = ax + b \]

with given boundary conditions and an initial guess for \( y \). The user supplied functions are

\[ f(x), \quad \text{right hand side}(x,y) \]

and \( \text{boundary conditions}(y_1, y_2) \) where \( y_1 = y(a) \) and \( y_2 = y(b) \).

\[ bvpset \]

Assigns values to properties; these are passed to \( bvp4c \) when it is executed.

\[ bvpinit(x, y) \]

\[ bvpinit(x, \text{<initial guess function>}) \]

Interpolate to determine the solution at \( x \).

12. Polynomials and Polynomial Functions

In MATLAB the polynomial

\[ p(x) = c_0 x^{n-1} + c_1 x^{n-2} + \cdots + c_{n-1} x + c_n \]

is represented by the vector \( q = (c_1, c_2, \ldots, c_n)^T \). You can easily calculate the roots of a polynomial by

\[ r = \text{roots}(q) \]

Conversely, given the roots of a polynomial you can recover the coefficients of the polynomial by

\[ q = \text{poly}(r) \]

**Warning:** Note the order of the coefficients in the polynomial. \( c_1 \) is the coefficient of the highest power of \( x \) and \( c_n \) is the coefficient of the lowest power, i.e., \( 0 \).

The polynomial can be evaluated at \( x \) by

\[ y = \text{polyval}(q, x) \]

where \( x \) is a scalar, a vector, or a matrix. If \( A \) is a square matrix, then

\[ p(A) = c_1 A^{n-1} + c_2 A^{n-2} + \cdots + c_{n-1} A + c_n \]

is calculated by

\[ \text{polyvalm}(q, A) \]

(See Section 15 for more details on this type of operation.)

A practical example which uses polynomials is to find the "best" fit to data by a polynomial of a particular degree. Suppose the data points are

\[ \{( -3, -2), (1.2, -1), (0, -0.5), (1, 1), (1.8, 2) \} \]

and we want to find the "best" fit by a straight line. Defining the data points more abstractly as

\[ \{( x_i, y_i) \mid i = 1, 2, \ldots, n \} \]

and the desired straight line by \( y = c_0 x + c_1 \), the matrix equation for the straight line is

\[
\begin{bmatrix}
  x_1 & 1 \\
  x_2 & 1 \\
  \vdots & \vdots \\
  x_n & 1 \\
\end{bmatrix}
\begin{bmatrix}
  c_0 \\
  c_1 \\
\end{bmatrix}
= 
\begin{bmatrix}
  y_1 \\
  y_2 \\
  \vdots \\
  y_n \\
\end{bmatrix}.
\]

In general, there is no solution to this overdetermined linear system. Instead, we find the least-squares solution \( c = (c_0, c_1)^T \) by

\[ x = [xdp \ \text{ones}(n, 1)] \backslash ydp \]

We can plot the data points along with this straight line by

\[ [x = \text{linspace}(\min(x), \max(x), 100); \]

\[ y = \text{polyval}(c, x); \]

\[ \text{plot}(x, y, \text{xdp}, \text{ydp}, \text{'o'}) \]

**Notation:** When discussing interpolation at a point \( x \) with data points \( \{( x_i, y_i) \mid i = 1, 2, \ldots, n \} \), there are too many \( x \)'s and \( y \)'s. To make it clear which are the data points and which are the evaluation points, we use \( xdp \) and \( ydp \) for the data points, and \( x \) for evaluation points with values \( y \).

We can find the ‘best’ fit by a polynomial of degree \( m < n \), i.e., \( y = c_1 x^m + c_2 x^{m-1} + \cdots + c_{m+1} \), by calculating the least-squares solution to

\[ \forall c = y \]

where

\[ V = \begin{bmatrix}
  x_1^m & x_1^{m-1} & \cdots & x_1 & 1 \\
  x_2^m & x_2^{m-1} & \cdots & x_2 & 1 \\
  \vdots & \vdots & \ddots & \vdots & \vdots \\
  x_n^m & x_n^{m-1} & \cdots & x_n & 1 \\
\end{bmatrix} \]

and \( c = \begin{bmatrix}
  c_1 \\
  c_2 \\
  \vdots \\
  c_{m+1} \\
\end{bmatrix} \).

The matrix \( V \) is called a Vandermonde matrix. The statement

\[ V = \text{vander}(xdp); \]

generates the square Vandermonde matrix with \( m = n - 1 \). To generate the \( n \times (m - 1) \) Vandermonde matrix we want, enter

\[ V = \text{vander}(x) \]

\[ V(:, 1:m-1) = [] \]

This entire procedure can be carried out much more easily by entering

\[ q = \text{polyfit}(xdp, ydp, m-1) \]

where the third argument is the order of the polynomial (i.e., the number of coefficients in the polynomial).

**Warning:** The Vandermonde matrix is approximately as badly conditioned as the Hilbert matrix which was discussed in Section 5.2. For example, \( \text{cond}(\text{vander}(1:10)) = 2 \times 10^{12} \) whereas \( \text{cond}(\text{hilb}(10)) = 2 \times 10^7 \).

You can also find a local maximum or minimum of the polynomial \( p(x) \) by finding the zeroes of \( p'(x) \). The coefficients of \( p'(x) \) are calculated by

\[ q.derv = \text{polyder}(q) \]

where \( q \) is the vector of the coefficients of \( p(x) \). You can also integrate a polynomial by

\[ q.int = \text{polyint}(q) \]

in which case the constant term is \( 0 \), or by

\[ q.int = \text{polyint}(q, c) \]

for some constant \( c \) in which case the constant term is \( c \).

Given a set of data points \( \{( x_i, y_i) \mid i = 1, 2, \ldots, n \} \) there is sometimes a need to estimate values that lie within these data points (this is called interpolation) or outside them (this is called extrapolation). This estimation is generally done by fitting data which is "near" the desired value to a polynomial and then evaluating this polynomial at the value.

There are a number of functions to interpolate data points in any number of dimensions. The simplest function in one dimension to interpolate the points \( \{( x_i, y_i) \mid 1 \leq i \leq n \} \) is

\[ y = \text{interp1}(xdp, ydp, x, \text{method}) \]

where \( x \) is a vector of the values to be interpolated, \( y \) is the vector of the interpolated values, and \( \text{method} \) is an optional argument specifying the method to be used. One additional requirement for this function is that the elements of \( x \) are monotonic, i.e., either all in increasing order or in decreasing order, to make it easy for the function to determine which data points are "near" the desired value. Five of the interpolation methods which can be used are the following:

'nearest': The interpolated value is the value of the nearest data point.
'linear': Linear splines are used to connect the given data points. That is, straight lines connect
each pair of adjacent data points. (This is the default.)

'spline': Cubic splines are used to connect the given data points. That is, cubic polynomials connect each pair of adjacent data points. The additional constraints needed to obtain unique polynomials are that the the two polynomials which overlap at each interior data point have the same first and second derivatives at this point.

'pchip': Piecewise cubic Hermite polynomials connect each pair of adjacent data points. This is similar to spline but the second derivatives need not be continuous at the interior data points. Instead, this interpolation is better at preserving the shape of the data. In particular, on intervals where the data is monotonic so is the piecewise polynomial, and on intervals where the data is concave up or down so is the piecewise polynomial.

'cubic': The same as pchip. Incidentally, interpolation really means extrapolation. If a value lies outside the interval $[x_1, x_n]$ then, by default, NaN is returned. This can be changed by adding a fifth argument:

- If the fifth argument is a number, this value is returned whenever the value lies outside the interval.
- If the fifth argument is 'extrap', extrapolation (using the same method) is used.

An alternate way to interpolate these points is by using the two functions:

- $pp = 	ext{spline}(xdp, ydp)$
- $yvalues = 	ext{ppval}(pp, xvalues)$

to generate and interpolate the cubic spline or

- $pp = 	ext{pchip}(xdp, ydp)$
- $yvalues = 	ext{ppval}(pp, xvalues)$

to generate and interpolate the piecewise cubic Hermite polynomials. The first function generates the structure $pp$ which contains all the information required to obtain a unique piecewise polynomial. The second function interpolates the piecewise cubic polynomial at the $x$ values given by the vector $xvalues$.

These two functions allow extrapolation, so that the elements of $x$ can be outside the interval $[x_1, x_n]$. The function spline has an additional use: with it you can specify precisely the boundary conditions to use. That is, if $ydp$ has two more elements than $xdp$, the first element is the slope at $x_1$ and the last is the slope at $x_n$.

For a discussion of how $fzero$ works and

help $fzero$
mous function the result of a zero of by using an anonymous function
\[
\text{function } y = \text{fcos}(x) = \cos(ax) + bx;
\]
\[
y = \cos(x) + bx; \quad a = 777;
\]
Then we can use closure, as already discussed in Section 10.1, so that the parameters can be set outside of fzero. This is easily done by entering

```matlab
function y = fcos(x, a, b)

y = a*cos(x) + b*x;
end
```

As with fzero, the full argument list is

```
% fminsearch(f, xstart)
% [x, fval, err] = fminsearch(f, xstart, options)
```

where iterate0 is a vector specifying where to begin searching for a local minimum. For example, if we enter

```matlab
% fminsearch(f, 0, n, options)
% [x] = fminsearch(f, xstart, options)
```

by

```
f = @(x) 1/(1 + x.^2);
```

and the result is

```
x = optimset('Display','iter');
```

by

```
f = @(x) 1/(1 + x.^2);
```

and the result is

```
options = optimset('FunValTol',1e-8);
```

by

```
f = @(x) 1/(1 + x.^2);
```

and the result is

```
options = optimset('FunValTol',1e-8);
```

by

```
f = @(x) 1/(1 + x.^2);
```

and the result is

```
options = optimset('FunValTol',1e-8);
```

by

```
f = @(x) 1/(1 + x.^2);
```

and the result is

```
options = optimset('FunValTol',1e-8);
```

by

```
f = @(x) 1/(1 + x.^2);
```

and the result is

```
options = optimset('FunValTol',1e-8);
```

by

```
f = @(x) 1/(1 + x.^2);
```

and the result is

```
options = optimset('FunValTol',1e-8);
```

by

```
f = @(x) 1/(1 + x.^2);
```

and the result is

```
options = optimset('FunValTol',1e-8);
```

by

```
f = @(x) 1/(1 + x.^2);
```

and the result is

```
options = optimset('FunValTol',1e-8);
```

by

```
f = @(x) 1/(1 + x.^2);
```

and the result is

```
options = optimset('FunValTol',1e-8);
```

by

```
f = @(x) 1/(1 + x.^2);
```

and the result is

```
options = optimset('FunValTol',1e-8);
```

by

```
f = @(x) 1/(1 + x.^2);
```

and the result is

```
options = optimset('FunValTol',1e-8);
```

by

```
f = @(x) 1/(1 + x.^2);
```

and the result is

```
options = optimset('FunValTol',1e-8);
```

by

```
f = @(x) 1/(1 + x.^2);
```

and the result is

```
options = optimset('FunValTol',1e-8);
```

by

```
f = @(x) 1/(1 + x.^2);
```

and the result is

```
options = optimset('FunValTol',1e-8);
```

by

```
f = @(x) 1/(1 + x.^2);
```

and the result is

```
options = optimset('FunValTol',1e-8);
```

by

```
f = @(x) 1/(1 + x.^2);
```

and the result is

```
options = optimset('FunValTol',1e-8);
```

by

```
f = @(x) 1/(1 + x.^2);
```

and the result is

```
options = optimset('FunValTol',1e-8);
```

by

```
```
14. Discrete Fourier Transform

There are a number of ways to define the discrete Fourier transform; we choose to define it as the discretization of the continuous Fourier series. In this section we show exactly how to discretize the continuous Fourier series. We present the definition which follows directly from the discretization of the continuous Fourier series. In this section we show exactly how to discretize the continuous Fourier series. We present the definition which follows directly from the discretization of the continuous Fourier series. We present the definition which follows directly from the discretization of the continuous Fourier series.

### Numerical Operations on Functions

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>@bquad&lt;function handle&gt;, a, b, c, d</td>
<td>Numerically evaluates a double integral.</td>
</tr>
<tr>
<td>@fninbbox&lt;function handle&gt;, xmin, xmax</td>
<td>Numerically calculates a local minimum of a one-dimensional function given the endpoints of the interval in which to search.</td>
</tr>
<tr>
<td>@fninbbox&lt;function handle&gt;, iterate0</td>
<td>Numerically calculates a local minimum of a multidimensional function given the initial iterate vector.</td>
</tr>
<tr>
<td>@fzero&lt;function handle&gt;, x0</td>
<td>Numerically calculates a zero of a function given the initial iterate. x0 can be replaced by a 2-vector of the endpoints of the interval in which a zero lies.</td>
</tr>
<tr>
<td>optimset&lt;Prop 1&gt;,&lt;Value 1&gt;, ...</td>
<td>Allows you to modify the parameters used by fzero, fninbbox, and fninbbox.</td>
</tr>
<tr>
<td>@quad&lt;function handle&gt;, a, b</td>
<td>Numerically evaluates an integral using Simpson's method.</td>
</tr>
<tr>
<td>@quadk&lt;function handle&gt;, a, b</td>
<td>Numerically evaluates an integral using the adaptive Gauss-Kronrod method. The interval can be infinite and/or the function can have an integrable singularity.</td>
</tr>
<tr>
<td>@quadl&lt;function handle&gt;, a, b</td>
<td>Numerically evaluates an integral using the adaptive Gauss-Lobatto method.</td>
</tr>
</tbody>
</table>

The power in each mode, i.e., the power spectrum, is

\[ P_k = \begin{cases} |a_k|^2 & \text{if } k = 0 \\ \frac{1}{2}(|a_k|^2 + |b_k|^2) & \text{if } k > 0 \end{cases} \]

and the frequency of the \( k \)th mode is \( k/T \) cycles per unit time. Since

\[ \cos \omega t = \frac{e^{j\omega t} + e^{-j\omega t}}{2} \quad \text{and} \quad \sin \omega t = \frac{e^{j\omega t} - e^{-j\omega t}}{2j}, \]

we can rewrite the real Fourier series as the complex Fourier series

\[ f(t) = a_0 + \sum_{k=1}^{\infty} \left( a_k e^{j 2\pi k t / T} + b_k e^{-j 2\pi k t / T} \right) \]

so that

\[ f(t) = \sum_{k=-\infty}^{\infty} c_k e^{j 2\pi k t / T} \quad \text{for all } t \in [0,T] \]

where

\[ c_k = \begin{cases} a_k & \text{for } k \geq 0 \\ \frac{1}{2}(a_k - ib_k) & \text{for } k > 0 \end{cases} \quad \text{and} \quad c_{-k} = \frac{1}{2}(a_k + ib_k) \]

The coefficients \( c_{-k} \) are the complex Fourier coefficients of \( f \), and \( c_k \) and \( c_{-k} \) are the coefficients of the \( k \)th mode. (Note that these Fourier coefficients are generally complex.) We can also calculate \( c_k \) directly from \( f \) by

\[ c_k = \frac{1}{T} \int_0^T f(t) e^{-j 2\pi k t / T} \, dt \quad \text{for } k = -\infty, -2, -1, 0, 1, 2, \ldots \]

Note that if \( f \) is real, then \( c_{-k} = c_k \) (by replacing \( k \) by \(-k\) in the above equation). The power of \( f(t) \) is

\[ P = |a_0|^2 + \sum_{k=1}^{\infty} (|a_k|^2 + |b_k|^2) \]

and the power in each mode is

\[ P_k = \begin{cases} |a_k|^2 & \text{if } k = 0 \\ \frac{1}{2}(|a_k|^2 + |b_k|^2) & \text{if } k > 0 \end{cases} \]

work done to the particle is \( W = \int F \cdot v \, dt \) where \( F \) is the vector from \( P_0 \) to \( P_1 \). The power of the particle is the work done per unit time, i.e., \( P = \frac{1}{2} \int F \cdot v \, dt \).

Next, consider a charge \( q \) which is moving between two terminals having a potential difference of \( V \). The work done on the charge is \( W = qV = lR \), where \( l \) is the current and \( t \) is the time it takes for the change to move between the two terminals. If \( R \) is the resistance in the circuit, \( V \) and \( R \) are the power is

\[ P = \frac{W}{T} = lV = lR = \frac{1}{V} \]

Thus, if we consider \( f(t) \) to be the voltage or the current of some signal, the instantaneous power in the signal is proportional to \( f(t)^2 \) and the average power is proportional to

\[ \frac{1}{T} \int_0^T |f(t)|^2 \, dt. \]
We can only calculate a finite number of Fourier coefficients numerically and so we truncate the infinite series at the $M^{th}$ mode. We should choose $M$ large enough that

$$f(t) \approx \sum_{k=-M}^{M} c_k e^{2\pi i k t / T} \quad \text{for all } t \in [0, T].$$

There are now

$$N = 2M + 1$$

unknowns (which is an odd number because of the $k = 0$ mode). We require $N$ equations to solve for these $N$ unknown coefficients. We obtain these equations by requiring that the two sides of this approximation be equal at the $N$ equally spaced abscissas $t_j = jT/N$ for $j = 0, 1, 2, \ldots, N-1$ (so that $0 = t_0 < t_1 < \cdots < t_{N-1} < t_N = T$). That is,

$$f(t_j) = \sum_{k=-M}^{M} c_k e^{2\pi i k j / N} \quad \text{for } j = 0, 1, 2, \ldots, N-1$$

or, written as a first-order system,

$$f_j = \sum_{k=-M}^{M} c_k e^{2\pi i k j / N} \quad \text{for } j = 0, 1, 2, \ldots, N-1$$

(14.4)

where $f_j \equiv f(t_j)$. This linear system can be solved to obtain

$$c_k = \frac{1}{N} \sum_{j=0}^{N-1} f_j e^{-2\pi i k j / N} \quad \text{for } k = -M, -M+1, \ldots, M.$$  

(14.5)

The reason we have replaced the coefficients $c_{-M}, c_{-M+1}, \ldots, c_{-1}, c_0$ by $\gamma_{-M}, \gamma_{-M+1}, \ldots, \gamma_{-1}, \gamma_0$ is that the $c$’s are the coefficients in the continuous complex Fourier series, eq. (14.2), and are calculated by (14.3). The $\gamma$’s are the coefficients in the discrete complex Fourier series, eq. (14.4), and are calculated by (14.5).

**Note:** To repeat: the discrete Fourier coefficient $\gamma_k$ is a function of $M$, i.e., $\gamma_k(M)$, and is generally not equal to the continuous Fourier coefficient $c_k$. However, as $M \to \infty$ we have $\gamma_k(M) \to c_k$. For a fixed $M$ we generally only have $\gamma_k(M) \approx c_k$ as long as $|k|$ is "much less than" $M$. Of course, it takes practice and experimentation to determine what "much less than" means.

We define the discrete Fourier series by

$$f_{D}(t) = \sum_{k=-M}^{M} c_k e^{2\pi i k t / T} \quad \text{for all } t \in [0, T].$$

It is our responsibility (using our experience) to choose $M$ large enough that $f(t) \approx f_{D}(t)$. Given $f = (f_0, f_1, f_2, \ldots, f_{N-1})^T$, the Fourier coefficients are calculated in MATLAB by

$$c = fft(f) / N$$

where the coefficients of the discrete Fourier transform are contained in $c$ in the order

$$\left(\gamma_0, \gamma_1, \ldots, \gamma_{M-1}, \gamma_{-M}, \gamma_{-M+1}, \ldots, \gamma_{-1}\right)^T.$$

The function `ifftshift` changes the order to

$$\left(\gamma_{-M}, \gamma_{-M+1}, \ldots, \gamma_{-2}, \gamma_{-1}, \gamma_0, \gamma_1, \ldots, \gamma_{M-1}, \gamma_M\right)^T.$$

The original function, represented by the vector $f$, is recovered by

$$f = N \times ifft(ifftshift(c))$$

and the order is changed by `ifftshift`.

It is important to check the Fourier transform to make sure it returns the results we expect. We began with the real trigonometric series (14.1) and derived the complex trigonometric series (14.2) from it. The nonzero Fourier coefficients of $f(x) = \cos x = (e^{ix} + e^{-ix})/2$ are $a_1 = 1$ and $a_{-1} = 1$, whereas the nonzero Fourier coefficients of $f(x) = \sin x = (e^{ix} - e^{-ix})/(2i)$ are $a_1 = -1$ and $a_{-1} = i/2$ but $c_1 = -i/2$.

The code

```matlab
n = 10;
x = linspace(0, 2*pi, n+1);
x(n+1) = []; c = fft(cos(x));
d_c = ifftshift(c); c_m = fft(sin(x));
d_c_m = ifftshift(c_m);
c = ifft(cos(x));
d_c = ifftshift(c_c);
c1 = ifft(sin(x));
d_c_c1 = ifftshift(c1);
```

returns the vectors

$$c_0 = (0.45, 0.0, 0.0, 0.0, 0.45),$$

$$d_0 = (0.0, 0.45, 0.45, 0.0, 0.0),$$

$$c_1 = (0.0, 0.0, 0.0, 0.0, 0.45),$$

$$d_1 = (0.0, 0.0, 0.0, 0.0, 0.0),$$

$$c_{-1} = (0.0, 0.0, 0.0, 0.0, 0.0),$$

$$d_{-1} = (0.0, 0.0, 0.0, 0.0, 0.0).$$

Notice that `fft` and `ifft` both return the correct coefficients for $\cos x$ (up to the scaling of $n$), but only the `ifft` returns the correct coefficients for $\sin x$. Thus, the function `fft` is correct, but it multiplies the coefficients by $N$.

Also, notice that `ifftshift` correctly shifts the coefficients, whereas `ifftshift` does not. `ifftshift` correctly shifts the coefficients back. That is, `ifftshift` is the inverse of `ifftshift` so

$$\text{ifftshift}(\text{ifftshift}(c)) = c.$$  

**Warning:** One of the most common mistakes in using `fft` is forgetting that the input is in the order

$$f_0, f_1, f_2, \ldots, f_{N-1}$$

while the output is in the order

$$\gamma_0, \gamma_1, \ldots, \gamma_{M-1}, \gamma_{-M}, \gamma_{-M+1}, \ldots, \gamma_{-2}, \gamma_{-1}.$$  

There is only one difficulty with our presentation. As we have already stated, the vector $f$ has $N = 2M + 1$ elements, which is an odd number. The Fast Fourier Transform (`fft` for short), which is the method used to calculate the discrete Fourier coefficients by `fft` and also to recover the original function by `ifft`, generally works faster if the number of elements of $f$ is even, and particularly fast if it is a power of 2.

The figure below shows the eponential needed to calculate `fft(f)` as a function of $N$. Since the vertical axis is logarithmic, it is clear that there is a huge difference in the time required as we vary $N$. The dashed lines show the minimum and maximum asymptotic times as $e^{\alpha \log n}$.

For $N$ to be even, we have to drop one coefficient, and the one we drop is $\gamma_{M}$. Now

$$N = 2M$$

141
The original function represented by the vector \( b \) before given and the discrete Fourier coefficients are now contained in \( f \). The coefficients of the discrete Fourier transform are now contained in \( f \) as
\[
\hat{f}_k = \sum_{j=0}^{N-1} f_j e^{-2\pi i k j / N} 
\]
and the discrete Fourier coefficients are calculated by
\[
\hat{f}_k = \sum_{j=0}^{N-1} f_j e^{-2\pi i k j / N} 
\]
for \( k = -M, -M+1, \ldots, M-2, M-1 \).

As before, given \( f = (f_0, f_1, f_2, \ldots, f_{N-1})^T \), the Fourier coefficients are calculated by
\[
\hat{f}_k = \sum_{j=0}^{N-1} f_j e^{-2\pi i k j / N} 
\]

The coefficients of the discrete Fourier transform are now contained in \( f \) as
\[
\hat{f}_k = \left( \gamma_0, \gamma_1, \ldots, \gamma_{M-2}, \gamma_{M-1}, \gamma_{M-2}, \ldots, \gamma_1, \gamma_0 \right)^T.
\]

The original function, represented by the vector \( f \), is again recovered by
\[
f = \text{ifft}(\hat{f})
\]

Note: Since there are now an even number of Fourier coefficients, we can reorder them by using \text{ifftshift}, which switches the first half and the last half of the elements. The result is
\[
\text{ifftshift}(\hat{f}) = \left( \gamma_{M-2}, \gamma_{M-1}, \gamma_{M-2}, \gamma_{M-1}, \gamma_0, \gamma_1, \ldots, \gamma_{M-2}, \gamma_{M-1} \right).
\]

Also, \text{fftshift} is the same as \text{fftshift} if \( N \) is even.

\textbf{Warning}: Remember that if you reorder the elements of \( f \) by
\[
\text{fft_shift} = \text{fftshift}(f) 
\]
you will have to "reorder" the elements by applying
\[
f = \text{fftshift}(\text{fft_shift})
\]
again before you use \text{ifft}.

Note: When \( N \) is even we cannot recover \( \gamma_M \) and so we only know one of the two coefficients of the \( M^{th} \) mode. Thus, we cannot determine the \( M^{th} \) mode correctly. Although we cannot give a simple example, it occasionally happens that this causes difficulties. The solution is to set \( \gamma_M = 0 \) so that the \( M^{th} \) mode is dropped completely.

We show a simple example of the use of Fourier coefficients from The Student Edition of MATLAB: User's Guide. We begin with

```
%% script m-file: fft_ex1
time = 0.6;
N = 600;
t = linspace(0, time, N);
y0 = sin(2*pi*50*t) + sin(2*pi*120*t);  \% unperturbed signal
ypert = y0 + 2*randn(size(t));  \% perturbed signal
figure(1)
plot(t, y0, 'r'), axis([0 time -8 8])
hold on
plot(t, ypert, 'g')
```

(this is contained in the accompanying zip file). This is a signal at 50 and 120 hertz (cycles per unit time), \( y_0 \), which is perturbed by adding Gaussian noise, \( y_{pert} \). We plot the periodic unperturbed signal, and then the perturbed signal, vs. time. Clearly, once the random noise has been added, the original signal has been completely lost — or has it?

We now look at the Fourier spectrum of \( y_0 \) by plotting the power at each frequency in

```
%% script m-file: fft_ex2
f0 = fft(y0)/N;  \% Fourier spectrum of unperturbed signal
figure(2)
f0(N/2+1) = [];  \% delete k = N/2 +1 mode
power0 = abs(f0).^2;  \% amplitude of each mode
freq = [1:N]/time;  \% the frequency of each mode
plot(freq(1:N/2), power0, 'r'), axis([0 freq(N/2) 0 .5])
```

\( f_{pert} = \text{fft}(y_{pert})/N, \) \% Fourier spectrum of perturbed signal

```
hold on
powerpert = abs(fpert).^2;  \% amplitude of each mode
freqpert(1:N/2) = abs(freqpert(1:N/2));  \% amplitude of each mode
plot(freqpert(1:N/2), powerpert, 'g')
```

(this is contained in the accompanying zip file). First, we plot the unperturbed power, \( \text{power0} \), and then the perturbed power, \( \text{powerpert} \), vs. the frequency at each mode, \( \text{freq} \). The two spikes in the plot of the unperturbed power are precisely at 50 and 120 hertz, the signature of the two sine functions in \( y_0 \). (For simplicity in the discussion, we have deleted the power in the \( M^{th} \) mode by \( f_{pert}(N/2 +1) = [] \) so that \( \text{power0}(k) \) is the power in the \( (k-1)^{th} \) mode.) Clearly, the original spikes are still dominant, but the random noise has excited every mode.

To see how much power is in the unperturbed signal and then the perturbed signal, enter

```
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
cputime to calculate fft(f,n)
```
The perturbed signal has about five times as much power as the original signal, which makes clear how large the perturbation is. Let us see if we can reconstruct the original signal by removing any mode whose magnitude is "small". By looking at the power plots, we see that the power in all the modes, except for those corresponding to the spikes, have an amplitude $\leq 0.1$. Thus, we delete any mode of the perturbed Fourier spectrum, i.e., $f_{\text{perp}}$, whose power is less than this value. We call this new Fourier spectrum $f_{\text{chop}}$. We then construct a new signal $y_{\text{chop}}$ from this "chopped" Fourier spectrum and compare it with the original unperturbed signal in the accompanying zip file.

The discrete sine transform of $f$ is an odd function in $T$, or for $n$ and we discretize this interval by $P$. Then the inverse discrete Fourier transform of the Fourier coefficients $f_{\text{chop}}$ are the same if the number of elements is even.

Nor is $f(t)$ an odd function in $t$, i.e., $f(-t) = -f(t)$ for all $t$, then the real trigonometric series can be simplified to

$$f(t) = \sum_{k=1}^{\infty} b_k \sin \frac{2\pi k t}{T}$$

for $t \in [0, T]$ or for $t \in [-\frac{T}{2}, +\frac{T}{2}]$. Choosing the latter interval, we only need define $f(t)$ for $t \in [0, \frac{T}{2}]$ and, additionally, state that it is an odd function. We discretize this infinite series by

$$f_{\text{chop}}(t) = \sum_{k=1}^{N} b_k \sin \frac{2\pi k t}{T}$$

for $t \in [0, \frac{T}{2}]$ and we discretize this interval by $0 = t_0 < t_1 < \cdots < t_n < t_{n+1} = \frac{T}{2}$ where $\Delta t = \frac{1}{2} T/(n+1)$. We immediately have $f(t_0) = f(t_{n+1}) = 0$ so these two modes are not needed. This leaves us with the $n$ coefficients $b_k$ and $n$ data points $\{f(t), f(t_j)\}$ for $i=1, 2, \ldots, n$. Defining the vectors $b = (b_1, b_2, \ldots, b_n)^T$ and $f = (f(t_1), f(t_2), \ldots, f(t_n))^T$, the MATLAB functions $\text{dat}$ and $\text{idat}$ switch between them by

```matlab
>> f = dat(b)
>> b = idat(f)
```

### Discrete Fourier Transform

$\text{fft}(f)$ The discrete Fourier transform of $f$.
$\text{ifft}(fc)$ The inverse discrete Fourier transform of the Fourier coefficients $fc$.
$\text{ifftshift}(fc)$ Switches the first half and the second half of the elements of $fc$.
$\text{fftshift}(fc)$ Unswitches the first half and the second half of the elements of $fc$. $\text{fftshift}$ and $\text{ifftshift}$ are the same if the number of elements is even.
$\text{dat}(b)$ The discrete sine transform of $f$ if it is an odd function where $b$ are the coefficients and $f$ is the function. That is, $f = \text{dat}(b)$.
$\text{idat}(f)$ The inverse discrete sine transform so $b = \text{idat}(f)$.  

15. Mathematical Functions Applied to Matrices

As we briefly mentioned in Section 27, mathematical functions can generally only be applied to square matrices. For example, if $A \in \mathbb{C}^{m \times m}$ then $\hat{A}$ is defined from the Taylor series expansion of $e^A$. That is, since

$$e^A = 1 + \frac{A}{1!} + \frac{A^2}{2!} + \frac{A^3}{3!} + \cdots$$

we define $\hat{A}$ to be

$$\hat{A} = 1 + \frac{A}{1!} + \frac{A^2}{2!} + \frac{A^3}{3!} + \cdots$$

(Thus, if $A \in \mathbb{C}^{m \times n}$ where $m \neq n$ then $\hat{A}$ does not exist because $A^k$ does not exist if $k$ is not a square matrix.) If $A$ is a square diagonal matrix $\hat{A}$ is particularly simple to calculate since

$$A^p = \begin{pmatrix} a_{11}^p & a_{12} & \cdots & 0 \\ a_{21} & a_{22}^p & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n}^p \end{pmatrix}$$

Thus,

$$\hat{A} = \begin{pmatrix} e^{a_{11}} & e^{a_{12}} & \cdots & 0 \\ 0 & e^{a_{22}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{a_{n,n}} \end{pmatrix}$$

The MATLAB function

```matlab
>> expm(A)
```

calculates $\hat{A}$ if $A$ is a square matrix. (Otherwise, it generates an error message.)

A simple example where $\hat{A}$ occurs is in the solution of first-order ode systems with constant coefficients. Recall that the solution of

$$\frac{dy}{dt}(t) = ay(t) \quad \text{for } t \geq 0 \text{ with } y(0) = y_0$$

is

$$y(t) = y_0 e^{at}.$$  

Similarly, the solution of

$$\frac{dy}{dt}(t) = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} y_{11}(t) \\ y_{21}(t) \\ \vdots \\ y_{n,1}(t) \end{pmatrix} \quad \text{for } t \geq 0 \text{ with } y(0) = y_0$$

i.e., $y'(t) = A y(t)$, is

$$y(t) = \hat{A} y_0.$$  

To calculate $y(t)$ for any time $t$, you only need enter

```matlab
>> expm(A*t) * y0
```

**Note:** The above statement gives the exact solution to the ode system at $t = 10$ by

```matlab
>> expm(A*10) * y0
```

You could also use numerical methods, as discussed in Section 10, to solve it. However, you would have to solve the ode for all $t \in [0, 10]$ in order to obtain a numerical approximation at the final time. This would be much more costly than simply using the analytical solution.

Similarly, $\sqrt{A}$ is calculated in MATLAB by entering

```matlab
>> sqrtm(A)
```
Finally, \( \log \) is calculated in MATLAB by entering

\[
\log(A)
\]

These are the only explicit MATLAB function for applying mathematical functions to matrices. However, there is a general matrix function for the other mathematical functions. The function

\[
\text{funm}(A, \text{<function handle>})
\]
evaluates \(<\text{function name}\>(A)\) for the MATLAB functions \( \exp, \sin, \cos, \sinh, \) and \( \cosh \) as well as user-defined functions.

<table>
<thead>
<tr>
<th>Matrix Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>expm(A)</td>
</tr>
<tr>
<td>sqrtm(A)</td>
</tr>
<tr>
<td>logm(A)</td>
</tr>
<tr>
<td>funm(A, &lt;function handle&gt;)</td>
</tr>
</tbody>
</table>
### Appendix: Reference Tables

These tables summarize the functions and operations described in this tutorial. The number (or numbers) shown give the page number of the table where this entry is discussed.

#### Mathematical Operators

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Operation</th>
<th>Page(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>Addition</td>
<td>7, 31</td>
</tr>
<tr>
<td>-</td>
<td>Subtraction</td>
<td>7, 31</td>
</tr>
<tr>
<td>*</td>
<td>Scalar or matrix multiplication</td>
<td>7, 31</td>
</tr>
<tr>
<td>./</td>
<td>Elementwise division of matrices</td>
<td>31</td>
</tr>
<tr>
<td>\</td>
<td>Scalar left division, i.e., ( \frac{a}{b} )</td>
<td>7</td>
</tr>
<tr>
<td>.^</td>
<td>Scalar or matrix exponentiation</td>
<td>7, 31</td>
</tr>
<tr>
<td>.^</td>
<td>Elementwise exponentiation of matrices</td>
<td>31</td>
</tr>
</tbody>
</table>

#### Special Characters

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>:</td>
<td>Creates a vector by ( a:b ) or ( a:c:b ); subscripts matrices</td>
<td>28</td>
</tr>
<tr>
<td>;</td>
<td>Ends a statement without printing out the result; also, ends each row when entering a matrix</td>
<td>9</td>
</tr>
<tr>
<td>*</td>
<td>Ends a statement when more than one appear on a line and the result is to be printed out; also, separates the arguments in a function; also, can separate the elements of each row when entering a matrix</td>
<td>9</td>
</tr>
<tr>
<td>...</td>
<td>Continues a MATLAB statement on the next line</td>
<td>15</td>
</tr>
<tr>
<td>%</td>
<td>Begins a comment</td>
<td>15</td>
</tr>
<tr>
<td>↑</td>
<td>The up-arrow key moves backward in the MATLAB workspace, one line at a time</td>
<td>7</td>
</tr>
</tbody>
</table>

#### Predefined Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ans</td>
<td>The default variable name when one has not been specified</td>
<td>9</td>
</tr>
<tr>
<td>pi</td>
<td>( \pi )</td>
<td>9</td>
</tr>
</tbody>
</table>
| eps     | Approximately the smallest positive real number on the computer such that \( 1 + \text{eps} 
eq 1 \) | 9 |
| Inf     | \( \infty \) (as in \( 1/0 \)) | 9 |
| NaN     | Not-a-Number (as in \( 0/0 \)) | 9 |
| i       | \( \sqrt{-1} \) | 9 |
| j       | \( \sqrt{-1} \) | 9 |
| realmin | The smallest “usable” positive real number on the computer | 9 |
| realmax | The largest “usable” positive real number on the computer | 9 |

#### Format Options

<table>
<thead>
<tr>
<th>Format</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>format short</td>
<td>The default setting</td>
<td>11</td>
</tr>
<tr>
<td>format long</td>
<td>Results are printed to approximately the maximum number of digits of accuracy in MATLAB</td>
<td>11</td>
</tr>
<tr>
<td>format short e</td>
<td>Results are printed in scientific notation</td>
<td>11</td>
</tr>
<tr>
<td>format long e</td>
<td>Results are printed in scientific notation to approximately the maximum number of digits of accuracy in MATLAB</td>
<td>11</td>
</tr>
<tr>
<td>format short g</td>
<td>Results are printed in the best of either format short or format short e</td>
<td>11</td>
</tr>
<tr>
<td>format long g</td>
<td>Results are printed in the best of either format long or format long e</td>
<td>11</td>
</tr>
<tr>
<td>format compact</td>
<td>Omits extra blank lines in output</td>
<td>11</td>
</tr>
</tbody>
</table>
Some Common Mathematical Functions

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>abs</td>
<td>Absolute value (p. 13, 14)</td>
</tr>
<tr>
<td>acos</td>
<td>Inverse cosine (p. 13)</td>
</tr>
<tr>
<td>acosd</td>
<td>Inverse cosine (result in degrees) (p. 13)</td>
</tr>
<tr>
<td>acosh</td>
<td>Inverse hyperbolic cosine (p. 13)</td>
</tr>
<tr>
<td>acot</td>
<td>Inverse cotangent (p. 13)</td>
</tr>
<tr>
<td>acotd</td>
<td>Inverse cotangent (result in degrees) (p. 13)</td>
</tr>
<tr>
<td>acoth</td>
<td>Inverse hyperbolic cotangent (p. 13)</td>
</tr>
<tr>
<td>acsch</td>
<td>Inverse hyperbolic cosecant (p. 13)</td>
</tr>
<tr>
<td>acscd</td>
<td>Inverse hyperbolic cosecant (result in degrees) (p. 13)</td>
</tr>
<tr>
<td>acsch</td>
<td>Inverse hyperbolic cosecant (p. 13)</td>
</tr>
<tr>
<td>asec</td>
<td>Inverse secant (p. 13)</td>
</tr>
<tr>
<td>asecd</td>
<td>Inverse secant (result in degrees) (p. 13)</td>
</tr>
<tr>
<td>asech</td>
<td>Inverse hyperbolic secant (p. 13)</td>
</tr>
<tr>
<td>asinh</td>
<td>Inverse hyperbolic sine (p. 13)</td>
</tr>
<tr>
<td>atand</td>
<td>Inverse tangent (result in degrees) (p. 13)</td>
</tr>
<tr>
<td>atan2</td>
<td>Inverse tangent using two arguments</td>
</tr>
<tr>
<td>atan2d</td>
<td>Inverse tangent using two arguments</td>
</tr>
<tr>
<td>atanh</td>
<td>Inverse hyperbolic tangent (p. 13)</td>
</tr>
<tr>
<td>ceil</td>
<td>Round upward to the nearest integer (p. 13)</td>
</tr>
<tr>
<td>conj</td>
<td>Complex conjugate (p. 14)</td>
</tr>
<tr>
<td>cos</td>
<td>Cosine (p. 13)</td>
</tr>
<tr>
<td>cosd</td>
<td>Cosine (argument in degrees) (p. 13)</td>
</tr>
<tr>
<td>cosh</td>
<td>Hyperbolic cosine (p. 13)</td>
</tr>
<tr>
<td>cot</td>
<td>Cotangent (p. 13)</td>
</tr>
<tr>
<td>cotd</td>
<td>Cotangent (argument in degrees) (p. 13)</td>
</tr>
<tr>
<td>coth</td>
<td>Hyperbolic cotangent (p. 13)</td>
</tr>
<tr>
<td>csc</td>
<td>Cosecant (p. 13)</td>
</tr>
<tr>
<td>cscd</td>
<td>Cosecant (argument in degrees) (p. 13)</td>
</tr>
<tr>
<td>csch</td>
<td>Hyperbolic cosecant (p. 13)</td>
</tr>
<tr>
<td>exp</td>
<td>Exponential function (p. 13)</td>
</tr>
<tr>
<td>expm1</td>
<td>Exponential minus one (p. 13)</td>
</tr>
<tr>
<td>fix</td>
<td>Round toward zero to the nearest integer (p. 13)</td>
</tr>
<tr>
<td>floor</td>
<td>Round downward to the nearest integer (p. 13)</td>
</tr>
<tr>
<td>heaviside</td>
<td>The Heaviside step function (p. 13)</td>
</tr>
<tr>
<td>imag</td>
<td>Imaginary part of a complex number (p. 14)</td>
</tr>
<tr>
<td>log</td>
<td>Natural logarithm, i.e., the logarithm to the base e. (p. 13)</td>
</tr>
<tr>
<td>log10</td>
<td>Common logarithm, i.e., the logarithm to the base 10. (p. 13)</td>
</tr>
<tr>
<td>log1p</td>
<td>Log of (1 + x) (p. 13)</td>
</tr>
<tr>
<td>mod</td>
<td>The modulus after division (p. 13)</td>
</tr>
<tr>
<td>real</td>
<td>The real part of a complex number (p. 14)</td>
</tr>
<tr>
<td>rem</td>
<td>The remainder after division (p. 13)</td>
</tr>
<tr>
<td>round</td>
<td>Round to the closest integer (p. 13)</td>
</tr>
<tr>
<td>sec</td>
<td>Secant (p. 13)</td>
</tr>
<tr>
<td>secd</td>
<td>Secant (argument in degrees) (p. 13)</td>
</tr>
<tr>
<td>sech</td>
<td>Hyperbolic secant (p. 13)</td>
</tr>
<tr>
<td>sign</td>
<td>The sign of the real number (p. 13)</td>
</tr>
<tr>
<td>sinh</td>
<td>Hyperbolic sine (p. 13)</td>
</tr>
<tr>
<td>sqrt</td>
<td>Square root (p. 13)</td>
</tr>
<tr>
<td>tan</td>
<td>Tangent (p. 13)</td>
</tr>
<tr>
<td>tand</td>
<td>Tangent (argument in degrees) (p. 13)</td>
</tr>
<tr>
<td>tanh</td>
<td>Hyperbolic tangent (p. 13)</td>
</tr>
</tbody>
</table>

Input-Output Functions

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>csvread</td>
<td>Reads data into MATLAB from the named file, one row per line of input (p. 57)</td>
</tr>
<tr>
<td>csvwrite</td>
<td>Writes out the elements of a matrix to the named file using the same format as csvread (p. 57)</td>
</tr>
<tr>
<td>diary</td>
<td>Saves your input to MATLAB and most of the output from MATLAB to disk (p. 7)</td>
</tr>
<tr>
<td>fopen</td>
<td>Opens the file with the permission string determining how the file is to be accessed (p. 84)</td>
</tr>
<tr>
<td>fclose</td>
<td>Closes the file (p. 84)</td>
</tr>
<tr>
<td>fscanf</td>
<td>Behaves very similarly to the C command in reading data from a file using any desired format (p. 84)</td>
</tr>
<tr>
<td>fprintf</td>
<td>Behaves very similarly to the C command in writing data to a file using any desired format. It can also be used to display data on the screen (p. 84)</td>
</tr>
<tr>
<td>input</td>
<td>Displays the prompt on the screen and waits for you to enter whatever is desired (p. 10)</td>
</tr>
<tr>
<td>load</td>
<td>Reads data into MATLAB from the named file, one row per line of input (p. 57)</td>
</tr>
<tr>
<td>importdata</td>
<td>Similar to load but there need not be an equal number of elements in each row (p. 57)</td>
</tr>
<tr>
<td>named file</td>
<td>One row per line of input.</td>
</tr>
<tr>
<td>print</td>
<td>Prints a plot or saves it in a file using various printer-specific formats (p. 57)</td>
</tr>
</tbody>
</table>

Arithmetical Matrix Operations

<table>
<thead>
<tr>
<th>Operation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A + B</td>
<td>Matrix addition (p. 7, 31)</td>
</tr>
<tr>
<td>A - B</td>
<td>Matrix subtraction (p. 7, 31)</td>
</tr>
<tr>
<td>A * B</td>
<td>Matrix multiplication (p. 7, 31)</td>
</tr>
<tr>
<td>A^n</td>
<td>Matrix exponentiation (p. 7, 31)</td>
</tr>
<tr>
<td>A / B</td>
<td>The solution to A * x = b by Gaussian elimination when A is a square non-singular matrix (p. 31, 83)</td>
</tr>
<tr>
<td>A \ B</td>
<td>The solution to A * x = B by Gaussian elimination (p. 31)</td>
</tr>
<tr>
<td>b / A</td>
<td>The solution to x * A = b where x and b are row vectors (p. 7, 31)</td>
</tr>
<tr>
<td>B / A</td>
<td>The solution to x * A = B by Gaussian elimination (p. 31)</td>
</tr>
<tr>
<td>A .* B</td>
<td>Elementwise multiplication (p. 31)</td>
</tr>
<tr>
<td>A ./ B</td>
<td>Elementwise division (p. 31)</td>
</tr>
<tr>
<td>A .^ B</td>
<td>Elementwise exponentiation (p. 31)</td>
</tr>
<tr>
<td>A \ A</td>
<td>Elementwise left division, i.e., B \ A is exactly the same as A ./ B (p. 31)</td>
</tr>
</tbody>
</table>
### Elementary Matrices

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>eye</code></td>
<td>Generates the identity matrix (p. 22)</td>
</tr>
<tr>
<td><code>false</code></td>
<td>Generates a logical matrix with all elements having the value false (p. 95)</td>
</tr>
<tr>
<td><code>ones</code></td>
<td>Generates a matrix with all elements being 1 (p. 22)</td>
</tr>
<tr>
<td><code>rand</code></td>
<td>Generates a matrix whose elements are uniformly distributed random numbers in the interval (0, 1) (p. 22)</td>
</tr>
<tr>
<td><code>randi</code></td>
<td>Uniformly distributed random integers (p. 22)</td>
</tr>
<tr>
<td><code>randn</code></td>
<td>Generates a matrix whose elements are normally (i.e., Gaussian) distributed random numbers with mean 0 and standard deviation 1 (p. 22)</td>
</tr>
<tr>
<td><code>rng</code></td>
<td>Generates a seed for the random number generator (p. 9)</td>
</tr>
<tr>
<td><code>randperm(n)</code></td>
<td>Generates a random permutation of the integers 1, 2, ..., n (p. 22)</td>
</tr>
<tr>
<td><code>speye</code></td>
<td>Generates a sparse identity matrix (p. 119)</td>
</tr>
<tr>
<td><code>sprand</code></td>
<td>Sparse uniformly distributed random matrix (p. 119)</td>
</tr>
<tr>
<td><code>sprandn</code></td>
<td>Sparse normally distributed random matrix (p. 119)</td>
</tr>
<tr>
<td><code>true</code></td>
<td>Generates a logical matrix with all elements having the value true (p. 95)</td>
</tr>
<tr>
<td><code>zeros</code></td>
<td>Generates a zero matrix (p. 22)</td>
</tr>
</tbody>
</table>

### Specialized Matrices

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>hilb</code></td>
<td>Generates the Hilbert matrix (Defined on p. 81.)</td>
</tr>
<tr>
<td><code>vander</code></td>
<td>Generates the Vandermonde matrix. (Defined on p. 134.)</td>
</tr>
<tr>
<td><code>toeplitz</code></td>
<td>Generates a Toeplitz matrix (where the values are constant on each diagonal) (p. 28)</td>
</tr>
</tbody>
</table>

### Elementary Matrix Operations

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>size</code></td>
<td>The size of a matrix (p. 22)</td>
</tr>
<tr>
<td><code>numel(A)</code></td>
<td>The total number of elements in a vector or matrix (p. )</td>
</tr>
<tr>
<td><code>length</code></td>
<td>The number of elements in a vector (p. 22)</td>
</tr>
<tr>
<td><code>.'</code></td>
<td>The transpose of a matrix (p. 22)</td>
</tr>
<tr>
<td><code>.'</code></td>
<td>The conjugate transpose of a matrix (p. 22)</td>
</tr>
</tbody>
</table>

### Manipulating Matrices

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>cat</code></td>
<td>Concatenates arrays; this is useful for putting arrays into a higher-dimensional array (p. 38)</td>
</tr>
<tr>
<td><code>clear</code></td>
<td>Deletes a variable or all the variables. <strong>This is a very dangerous command.</strong></td>
</tr>
<tr>
<td><code>diag</code></td>
<td>Extracts or creates diagonals of a matrix (p. 28)</td>
</tr>
<tr>
<td><code>flipud</code></td>
<td>Flips a matrix up and down (p. 28)</td>
</tr>
<tr>
<td><code>fliplr</code></td>
<td>Flips a matrix left to right (p. 28)</td>
</tr>
<tr>
<td><code>ind2sub</code></td>
<td>Converts the indices of a matrix from column vector form to matrix form (p. 99)</td>
</tr>
<tr>
<td><code>ipermute</code></td>
<td>The inverse of <code>permute</code> (p. 38)</td>
</tr>
<tr>
<td><code>invperm</code></td>
<td>Reverses the dimensions of a multidimensional array (p. 38)</td>
</tr>
<tr>
<td><code>linalg</code></td>
<td>Generates a sparse matrix by diagonals (p. 119)</td>
</tr>
<tr>
<td><code>regat</code></td>
<td>Tiles a matrix with copies of another matrix (p. 28)</td>
</tr>
<tr>
<td><code>bsxfun</code></td>
<td>Performs an operation on a matrix and a vector (p. 28)</td>
</tr>
<tr>
<td><code>reshape</code></td>
<td>Reshapes the elements of a matrix (p. 28)</td>
</tr>
<tr>
<td><code>rot90</code></td>
<td>Rotates a matrix a multiple of 90° (p. 28)</td>
</tr>
<tr>
<td><code>squeeze</code></td>
<td>Removes (i.e., squeezes out) dimensions which only have one element (p. 38)</td>
</tr>
<tr>
<td><code>sub2ind</code></td>
<td>Converts the indices of a matrix from matrix form to column vector form (p. 99)</td>
</tr>
<tr>
<td><code>triu</code></td>
<td>Extracts the upper triangular part of a matrix (p. 28)</td>
</tr>
<tr>
<td><code>tril</code></td>
<td>Extracts the lower triangular part of a matrix (p. 28)</td>
</tr>
<tr>
<td><code>[]</code></td>
<td>The null matrix. This is also useful for deleting elements of a vector and rows or columns of a matrix (p. 28)</td>
</tr>
</tbody>
</table>

### Odds and Ends

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>path</code></td>
<td>View or change the search path (p. 111)</td>
</tr>
<tr>
<td><code>addpath</code></td>
<td>Add to the search path (p. 111)</td>
</tr>
<tr>
<td><code>edit</code></td>
<td>Create a new function or script m-file or edit an already existing one (p. 15)</td>
</tr>
<tr>
<td><code>type</code></td>
<td>Display the actual MATLAB code for a command/function (p. 15, 17)</td>
</tr>
<tr>
<td><code>cpumtime</code></td>
<td>Approximately the CPU time (in seconds) used during this session (p. 31)</td>
</tr>
<tr>
<td><code>tic, toc</code></td>
<td>Return the elapsed time between these two statements (p. 31)</td>
</tr>
<tr>
<td><code>pause</code></td>
<td>Halt execution until you press some key (p. 111)</td>
</tr>
<tr>
<td><code>primes(n)</code></td>
<td>Generate the first n prime numbers</td>
</tr>
<tr>
<td><code>rats</code></td>
<td>Convert a floating-point number to a “close” rational number, which is frequently the exact value (p. 83)</td>
</tr>
<tr>
<td><code>deal(x, y)</code></td>
<td>Reverse two variables.</td>
</tr>
</tbody>
</table>
Data Manipulation Functions

- `cumprod`: The cumulative product between successive elements of a vector or each column of a matrix (p. 37)
- `cussum`: The cumulative sum between successive elements of a vector or each column of a matrix (p. 37)
- `errorbar`: Plots a curve through data points and also the error bar at each data point (p. 58)
- `hist`: Plots a histogram of the elements of a vector (p. 37)
- `max`: The maximum element of a vector or each column of a matrix. Alternately, if it has two arguments, it returns a matrix of the same size which contains the larger of the elements in each matrix (p. 37)
- `min`: The minimum element of a vector or each column of a matrix (p. 37)
- `mean`: The mean, or average, of the elements of a vector or each column of a matrix (p. 37)
- `norm`: The norm of a vector or a matrix (p. 37)
- `prod`: The product of the elements of a vector or each column of a matrix (p. 37)
- `sort`: Sorts the elements of a vector or each column of a matrix in increasing order (p. 37)
- `std`: The standard deviation of the elements of a vector or each column of a matrix (p. 37)
- `sum`: The sum of the elements of a vector or each column of a matrix (p. 37)

Some Useful Functions in Linear Algebra

- `chol`: Calculates the Cholesky decomposition of a symmetric, positive definite matrix (p. 90)
- `cond`: Calculates the condition number of a matrix (p. 90)
- `condest`: Calculates a lower bound to the condition number of a square matrix (p. 90)
- `det`: Calculates the determinant of a square matrix (p. 90)
- `eig`: Calculates the eigenvalues, and eigenvectors if desired, of a square matrix (p. 90)
- `eigs`: Calculates some eigenvalues and eigenvectors of a square matrix (p. 90)
- `inv`: Calculates the inverse of a square invertible matrix (p. 90)
- `linvecsolve`: Solve a linear system of equations where the matrix can have certain properties to increase the CPU time (p. 80)
- `lu`: Calculates the LU decomposition of a square invertible matrix (p. 90)
- `norm`: Calculates the norm of a vector or matrix (p. 90)
- `null`: Calculates an orthonormal basis for the null space of a matrix (p. 90)
- `orth`: Calculates an orthonormal basis for the range of a matrix (p. 90)
- `pinv`: Calculates the pseudoinverse of a matrix (p. 83)
- `qr`: Calculates the QR decomposition of a matrix (p. 90)
- `rank`: Calculates the rank of a matrix (p. 90)
- `rref`: Calculates the reduced row echelon form of a matrix (p. 80)
- `svd`: Calculates the singular value decomposition of a matrix (p. 90)

Logical and Relational Operators

- `&&`: A short-circuiting logical AND (p. 94)
- `||`: A short-circuiting logical OR (p. 94)
- `&&`: Less than (p. 92)
- `<=`: Less than or equal to (p. 92)
- `==`: Equal to (p. 92)
- `>=`: Greater than (p. 92)
- `~`: Not equal to (p. 92)
- `~`: Logical NOT (p. 94)
- `<`: Greater than (p. 92)
- `<=`: Less than or equal to (p. 92)
- `~`: Greater than or equal to (p. 92)

Control Flow

- `break`: Terminates execution of a `for` or `while` loop (p. 95)
- `case`: Part of the `switch` command (p. 95)
- `continue`: Begins the next iteration of a `for` or `while` loop immediately (p. 95)
- `else`: Used with the `if` statement (p. 95)
- `elseif`: Used with the `if` statement (p. 95)
- `end`: Terminates the scope of the `for`, `if`, `switch`, and `while` statements (p. 95, 111)
- `error`: Displays the error message and terminates all flow of control statements (p. 111)
- `for`: Repeats statements a specific number of times (p. 95)
- `if`: Executes statements if certain conditions are met (p. 95)
- `otherwise`: Part of the `switch` command (p. 95)
- `switch`: Selects certain statements based on the value of the `switch` expression (p. 95)
- `while`: Repeats statements as long as an expression is true (p. 95)
Logical Functions

All
True if all the elements of a vector are true, operates on the columns of a matrix. (p. 99)
Any
True if any of the elements of a vector are true, operates on the columns of a matrix. (p. 99)
Exist
False if this name is not the name of a variable or a file. (p. 99)
Isequal
Tests if two (or more) arrays have the same contents. (p. 99)
Find
The indices of a vector or matrix which are nonzero. (p. 99)
Logical
Converts a numeric variable to a logical one. (p. 99)
IsCell
True for a cell array. (p. 99)
IsChar
True if a vector or array contains character elements. (p. 99)
IsColumn
True for a column vector. (p. 99)
Isequal
Tests if two (or more) arrays have the same contents. (p. 99)
IField
True if the argument is a structure field. (p. 99)
IsField
True if the matrix is empty. (p. 99)
IIsLogical
True for a logical variable or array. (p. 99)
IsMatrix
Generates an array with 1 in all the elements which are contained in another array. (p. 99)
IsMax
Generates a matrix with 1 in all the elements which are NaN and 0 otherwise. (p. 99)
IsMax
Generates a matrix with 1 in all the elements which are NaN and 0 otherwise. (p. 99)
IIsNumeric
True for a floating-point array. (p. 99)
IIsPrime
Generates an array with 1 in all the elements which are prime numbers. Only non-negative integers are allowed in the elements. (p. 99)
IIsReal
True for a real array, as opposed to a complex one. (p. 99)
IsRow
True for a row vector. (p. 99)
IIsScalar
True for a scalar variable. (p. 99)
IsSparse
True for a sparse array. (p. 99)
IsStruct
True for a structure array. (p. 99)
IsVector
True for a vector, as opposed to a matrix. (p. 99)

Programming Language Functions

Echo
Turns echoing of statements in m-files on and off. (p. 111)
End
Ends a function. Only required if the function m-file contains a nested function. (p. 95, 111)
Error
Displays the error message and terminates the function. (p. 111)
Eval
Executes MATLAB statements contained in a text variable. (p. 113)
Feval
Executes a function specified by a string. (p. 113)
Function
Begins a MATLAB function. (p. 111)
Global
Defines a global variable (i.e., it can be shared between different functions and/or the workspace). (p. 111)
Lasterr
If eval "catches" an error, it is contained here. (p. 113)
Persistent
Defines a local variable whose value is to be saved between calls to the function. (p. 111)
Keyboard
Stops execution in an m-file and returns control to the user for debugging purposes. (p. 105)
Margin
Number of input arguments supplied by the user. (p. 111)
Nargout
Number of output arguments supplied by the user. (p. 111)
Return
Terminates the function or script m-file immediately. (p. 105, 111)
Varargin
"Groups" input arguments together. (p. 111)
Varargout
"Groups" output arguments together. (p. 111)

Debugging Commands

Keyboard
Turns debugging on. (p. 105)
Dbstep
Execute one or more lines. (p. 105)
Dbcont
Continue execution. (p. 105)
Dbstop
Set a breakpoint. (p. 105)
Dbclear
Remove a breakpoint. (p. 105)
Dbup
Change the workspace to the calling function or the base workspace. (p. 105)
Dtdown
Change the workspace down to the called function. (p. 105)
Dbstack
Display all the calling functions. (p. 105)
Dbstatus
List all the breakpoints. (p. 105)
Dbtype
List the current function, including the line numbers. (p. 105)
DQuit
Quit debugging mode and terminate the function. (p. 105)
Return
Quit debugging mode and continue execution of the function. (p. 105, 111)

Discrete Fourier Transform

FFT
The discrete Fourier transform. (p. 145)
Fftshift
Switches the first half and the second half of the elements of a vector. (p. 145)
Ifft
The inverse discrete Fourier transform. (p. 145)
Ifftshift
Unswitches the first half and the second half of the elements of a vector. (p. 145)
Dat
The discrete sine transform. (p. 145)
Idat
The inverse discrete sine transform. (p. 145)
### Sparse Matrix Functions

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>spye</td>
<td>Generates a sparse identity matrix (p. 119)</td>
</tr>
<tr>
<td>sprand</td>
<td>Sparse uniformly distributed random matrix (p. 118)</td>
</tr>
<tr>
<td>sprandn</td>
<td>Sparse normally distributed random matrix (p. 118)</td>
</tr>
<tr>
<td>sparse</td>
<td>Generates a sparse matrix elementwise (p. 119)</td>
</tr>
<tr>
<td>spdiags</td>
<td>Generates a sparse matrix by diagonals (p. 119)</td>
</tr>
<tr>
<td>full</td>
<td>Converts a sparse matrix to a full matrix (p. 119)</td>
</tr>
<tr>
<td>find</td>
<td>Finds the indices of the nonzero elements of a matrix (p. 119)</td>
</tr>
<tr>
<td>nnz</td>
<td>Returns the number of nonzero elements in a matrix (p. 119)</td>
</tr>
<tr>
<td>spfun</td>
<td>Applies the function to a sparse matrix (p. 119)</td>
</tr>
<tr>
<td>spconvert</td>
<td>Generates a sparse matrix given the nonzero elements and their indices (p. 119)</td>
</tr>
</tbody>
</table>

### Numerical Operations on Polynomials

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>interp1</td>
<td>Does one-dimensional interpolation (p. 136)</td>
</tr>
<tr>
<td>interp2</td>
<td>Does two-dimensional interpolation (p. 136)</td>
</tr>
<tr>
<td>interp3</td>
<td>Does three-dimensional interpolation (p. 136)</td>
</tr>
<tr>
<td>pchip</td>
<td>Cubic Hermite interpolation (p. 136)</td>
</tr>
<tr>
<td>poly</td>
<td>Calculates the coefficients of a polynomial given its roots (p. 136)</td>
</tr>
<tr>
<td>polyder</td>
<td>Calculates the derivative of a polynomial (p. 136)</td>
</tr>
<tr>
<td>polyfit</td>
<td>Calculates the least-squares polynomial of a given degree which fits the given data (p. 136)</td>
</tr>
<tr>
<td>polyval</td>
<td>Calculates the integral of a polynomial (p. 136)</td>
</tr>
<tr>
<td>polyvalm</td>
<td>Evaluates a polynomial at a point (p. 136)</td>
</tr>
<tr>
<td>ppval</td>
<td>Evaluates a polynomial with a matrix argument (p. 136)</td>
</tr>
<tr>
<td>spval</td>
<td>Interpolates a piecewise polynomial calculated by pchip or spline (p. 136)</td>
</tr>
<tr>
<td>roots</td>
<td>Numerically calculates all the zeros of a polynomial (p. 136)</td>
</tr>
<tr>
<td>spline</td>
<td>Cubic spline interpolation (p. 136)</td>
</tr>
</tbody>
</table>

### Matrix Functions

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>expm</td>
<td>Matrix exponential (p. 147)</td>
</tr>
<tr>
<td>fsum</td>
<td>Evaluate general matrix function (p. 147)</td>
</tr>
<tr>
<td>logm</td>
<td>Matrix logarithm (p. 147)</td>
</tr>
<tr>
<td>sqrtm</td>
<td>Matrix square root (p. 147)</td>
</tr>
</tbody>
</table>

### Time Evolution ODE Solvers

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ode45</td>
<td>Non-stiff ode solver; fourth-order, one-step method for the ode ( y' = f(t, y) ). (p. 120)</td>
</tr>
<tr>
<td>ode23</td>
<td>Non-stiff ode solver; second-order, one-step method (p. 120)</td>
</tr>
<tr>
<td>ode113</td>
<td>Non-stiff ode solver; variable-order, multi-step method (p. 120)</td>
</tr>
<tr>
<td>ode15s</td>
<td>Stiff ode solver; variable-order, multi-step method (p. 120)</td>
</tr>
<tr>
<td>ode23s</td>
<td>Stiff ode solver; second-order, one-step method (p. 120)</td>
</tr>
<tr>
<td>ode23t</td>
<td>Stiff ode solver; trapezoidal method (p. 120)</td>
</tr>
<tr>
<td>ode23tb</td>
<td>Stiff ode solver; second-order, one-step method (p. 120)</td>
</tr>
<tr>
<td>ode15sii</td>
<td>Stiff ode solver; variable-order, multi-step method for the fully implicit ode ( f(t, y, y') = 0 ) (p. 129)</td>
</tr>
<tr>
<td>odeset</td>
<td>Assigns values to properties of the ode solver (p. 125)</td>
</tr>
</tbody>
</table>

### Boundary-Value Solver

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>bvp4c</td>
<td>Numerically solves ( y''(x) = f(x, y) ) for ( x \in [a, b] ) with given boundary conditions and an initial guess for ( y ) (p. 133)</td>
</tr>
<tr>
<td>bvpset</td>
<td>Assigns values to properties of bvp4c (p. 133)</td>
</tr>
<tr>
<td>bvpinit</td>
<td>Calculates the initial guess either by giving ( y ) directly or by using a function ( y = \text{initial guess function}(x) ) (p. 133)</td>
</tr>
<tr>
<td>deval</td>
<td>Interpolates the solution desired points (p. 133)</td>
</tr>
</tbody>
</table>

### Numerical Operations on Functions

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>dblquad</td>
<td>Numerically evaluates a double integral (p. 139)</td>
</tr>
<tr>
<td>fminbnd</td>
<td>Numerically calculates a local minimum of a one-dimensional function (p. 139)</td>
</tr>
<tr>
<td>fminsearch</td>
<td>Numerically calculates a local minimum of a multi-dimensional function (p. 139)</td>
</tr>
<tr>
<td>optimset</td>
<td>Allows you to modify the parameters used by fzero, fminbnd, and fminsearch (p. 139)</td>
</tr>
<tr>
<td>fzero</td>
<td>Numerically calculates a zero of a function (p. 139)</td>
</tr>
<tr>
<td>quad</td>
<td>Numerically evaluates an integral using Simpson's method (p. 139)</td>
</tr>
<tr>
<td>quadgk</td>
<td>Numerically evaluates an integral using the adaptive Gauss-Kronrod method. The interval can be infinite and/or the function can have an integrable singularity (p. 139)</td>
</tr>
<tr>
<td>quadl</td>
<td>Numerically evaluates an integral using the adaptive Gauss-Lobatto method (p. 139)</td>
</tr>
</tbody>
</table>
Solutions To Exercises

These are the solutions to the exercises given in subsections 1.9, 2.10, 3.6, and 4.7.

1.9.1a) \( \alpha = 3.7; \beta = 5.7; \text{deg} = \pi/180; \text{ab} = 79; \)
\( c = \sqrt{a^2 + b^2 - 2*a*b*cos(ab*deg)} \)
answer: 6.1751

b) \( \text{format long} \)
\( c \)
answer: 6.1750885147187636

c) \( \text{format short e} \)
\( \text{asind}((b/c)*sin(ab*deg)) / \text{deg} \)
\( \text{asind}((b/c)*sin(ab)) \)
answer: 4.9448e+01

d) \( \text{diary 'triangle.ans'} \)

1.9.2) \( x = 1.2e20 - 1i*12e20; y = (1/3) \)
answer: 1.3637e+07 - 7.6680e+061

1.9.3) \( \text{th = input('th = ')}; \text{cos(2*th) - (cos(2*t)*2 - 2) - 1} \)

1.9.4) \( \text{help fix or doc fix} \)

2.10.1a) \( A = [1 2 3 4; 5 6 7 8; 9 10 11 12; 13 14 15 16] \)
\( A = [1:4; 5:8; 9:12; 13:16] \)

b) \( A(2,:) = (1:-5)*K(2,:) + A(3,:); \)

2.10.2) \( A = 4*eye(n) - \text{diag(ones(n-1,1), 1) - diag(ones(n-1,1), -1)} \)

2.10.3) \( A = \text{diag([1:n].'-2) - diag(ones(n-1,1), 1) - diag( exp([1:n]), -1)} \)

2.10.4a) \( A = [\text{ones(6,4)} zeros(6)]; A(6,1) = 5; A(1,10) = -5 \)

b) \( A = A - \text{tril}(A,-1) \)

2.10.5) \( x = [0:30]'; -2 \quad \text{or} \quad x = [0:30].'-2' \)

2.10.6a) \( R = \text{rand}(R) \)

b) \( [m, in] = \text{max}(R') \)

c) \( \text{mean(mean(R))} \quad \text{or} \quad \text{mean(R(:,:,))} \)

d) \( S = \text{sin}(R) \)

e) \( r = \text{diag}(R) \)

2.10.7a) \( A = [1 2 3; 4 5 6; 7 8 10] \)
\( B = A' - 5 \quad \text{or} \quad B = \text{sqrtm}(A) \)
\( C = A - .5 \quad \text{or} \quad C = \text{sqrt}(A) \)

b) \( A = B + 2 \)
\( A = C - .2 \)

3.6.1) \( \text{si.name = 'Alfonso Bedoya'} \)
\( \text{si.rank = 'bandit'} \)
\( \text{si.serial_number = 1} \)
\( s2 = \text{struct('name', 'Alfonso Bedoya', 'rank', 'bandit', 'serial_number', 1)} \)
\( \text{si.serial_number = si.serial_number + 1} \)

3.6.2a) \( c(1,1) = \text{randn}() \)
\( c(2,1) = \text{pi/10} \)
\( c(2,2) = 0(x) \text{sin(exp}(x)) \)
\( c(1,3) = c(1,2)^2 \)
\( c(2,3) = \text{true} \)

b) \( c(1,1) = 1 \)
\( c(2,2) = 0 \)

3.6.3a) \( f = 0(x,y) \sin(x + y) .* \cos(x - y) \)

b) \( f = 0(x) \cos(2(x)) + \cos(3(x))./(abs(x(1) + x(2)) + 1) \)

3.6.4a) \( x = \text{linespace(-1, +1, 100)} \)
\( y = \text{exp}(x) \)
\( \text{plot}(x, y) \)

b) \( z = 1 + x + x.2 + x.6 /6 \)
\( \text{hold on} \)
\( \text{plot}(x, z) \)

c) \( \text{plot}(x, y-z) \)

3.6.4b) \( \text{hold off} \)
\( \text{plot}(x, y, 'r', x, z, 'g', x, y-x, 'm') \)

\( \text{axis equal} \)
\( \text{xlabel('x')} \)
\( \text{ylabel('y')} \)
\( \text{title('e^-\pi = -1 is profound')} \)

c) \( \text{subplot(2, 1, 1)} \)
\( \text{hold off} \)
\( \text{plot}(x, y, 'r', x, z, 'g'); \)

\( \text{axis equal} \)
\( \text{xlabel('x')} \)
\( \text{ylabel('y')} \)
\( \text{title('e^-\pi = -1 is profound')} \)
\( \text{subplot(2, 1, 2)} \)
\( \text{plot}(x, y-z) \)

4.7.1a) \( x = \text{linespace(-1, +1, 100)} \)
\( y = \text{exp}(x) \)
\( \text{plot}(x, y) \)

b) \( z = 1 + x + x.2 + x.6 /6 \)
\( \text{hold on} \)
\( \text{plot}(x, z) \)

c) \( \text{plot}(x, y-z) \)

4.7.2a) \( x = \text{linespace(-3, 3, 91)} \)
\( y = x \)
\( [X, Y] = \text{meshgrid}(x, y); \quad \text{or just do} \quad [X, Y] = \text{meshgrid}(x, y); \)
\( Z = (X.-2 + 4*Y.-2) .* \text{sin}(2*pi*X) .* \text{sin}(2*pi*Y); \)
\( \text{surf}(X, Y, Z); \)
b) One particular choice is
\[
\text{view}([1 \to 5]) \quad \text{or} \quad \text{view}([63 \to 66])
\]

4.7.3) \( \mathbf{x} \) is a row vector containing all the points at which the function is to be evaluated. \( \mathbf{c} \) is a column vector of the speeds. We can consider these to be the horizontal and vertical axes in two dimensions. We define \( h(x, c, t) = f(x - ct) \) so that \( g(x, t) = \sum_{k=1}^{m} h(x/c_k, t) \). \texttt{meshgrid} then turns these vectors into a mesh so that \( h(x_j, c_i, t) \) is evaluated by \( h(x(j), c(i), t) \), where \( i \) refers to the vertical axis and \( j \) the horizontal one. (Since the matrices \( \mathbf{X} \) and \( \mathbf{C} \) never change, they would not need to be actual arguments to \( h \).) Thus, \( h(t) \) (short for \( h(X, C, t) \)) is a matrix whose \((i, j)\)th element is \( f(x_j - c_i t) \). We need to multiply this by \( \mathbf{a} \) and then sum over \( i \). This is done by multiplying the matrix \( \mathbf{R} \), whose \((i, j)\)th element is \( a_i \), elementwise by the matrix \( h(X, C, t) \). The \texttt{sum} function then adds all the rows in each column.
### ASCII Table

<table>
<thead>
<tr>
<th>Octal</th>
<th>Decimal</th>
<th>Sequence</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>000</td>
<td>0</td>
<td></td>
<td>All symbols</td>
</tr>
<tr>
<td>001</td>
<td>1</td>
<td>-</td>
<td>Start of header</td>
</tr>
<tr>
<td>002</td>
<td>2</td>
<td>-</td>
<td>Start of text</td>
</tr>
<tr>
<td>003</td>
<td>3</td>
<td>-</td>
<td>End of text</td>
</tr>
<tr>
<td>004</td>
<td>4</td>
<td>-</td>
<td>End of transmission</td>
</tr>
<tr>
<td>005</td>
<td>5</td>
<td>-</td>
<td>Inquiry</td>
</tr>
<tr>
<td>006</td>
<td>6</td>
<td>-</td>
<td>Acknowledgment</td>
</tr>
<tr>
<td>007</td>
<td>7</td>
<td>-</td>
<td>Bell</td>
</tr>
<tr>
<td>010</td>
<td>8</td>
<td>*</td>
<td>Backspace</td>
</tr>
<tr>
<td>011</td>
<td>9</td>
<td>-</td>
<td>Vertical tab</td>
</tr>
<tr>
<td>012</td>
<td>10</td>
<td>_</td>
<td>Form feed</td>
</tr>
<tr>
<td>013</td>
<td>11</td>
<td>-</td>
<td>Line feed</td>
</tr>
<tr>
<td>014</td>
<td>12</td>
<td>-</td>
<td>Start of transmission</td>
</tr>
<tr>
<td>015</td>
<td>13</td>
<td>-</td>
<td>Ending text</td>
</tr>
<tr>
<td>016</td>
<td>14</td>
<td>~</td>
<td>Shift out</td>
</tr>
<tr>
<td>017</td>
<td>15</td>
<td>`</td>
<td>Shift in</td>
</tr>
<tr>
<td>020</td>
<td>16</td>
<td>`</td>
<td>Data link escape</td>
</tr>
<tr>
<td>021</td>
<td>17</td>
<td>&amp;</td>
<td>Device control 1 ( Often XON)</td>
</tr>
<tr>
<td>022</td>
<td>18</td>
<td>*</td>
<td>Device control 2</td>
</tr>
<tr>
<td>023</td>
<td>19</td>
<td>#</td>
<td>Device control 3 (Often XOFF)</td>
</tr>
<tr>
<td>024</td>
<td>20</td>
<td>$</td>
<td>Device control 4</td>
</tr>
<tr>
<td>025</td>
<td>21</td>
<td>-</td>
<td>Negative acknowledgement</td>
</tr>
<tr>
<td>026</td>
<td>22</td>
<td>–</td>
<td>Synchronous idle</td>
</tr>
<tr>
<td>027</td>
<td>23</td>
<td>°</td>
<td>End of transmissions block</td>
</tr>
<tr>
<td>030</td>
<td>24</td>
<td>-</td>
<td>Cancel</td>
</tr>
<tr>
<td>031</td>
<td>25</td>
<td>-</td>
<td>End of medium</td>
</tr>
<tr>
<td>032</td>
<td>26</td>
<td>&amp;</td>
<td>Substitute</td>
</tr>
<tr>
<td>033</td>
<td>27</td>
<td>-</td>
<td>Escape</td>
</tr>
<tr>
<td>034</td>
<td>28</td>
<td>^</td>
<td>File separator</td>
</tr>
<tr>
<td>035</td>
<td>29</td>
<td>-</td>
<td>Group separator</td>
</tr>
<tr>
<td>036</td>
<td>30</td>
<td>-</td>
<td>Record separator</td>
</tr>
<tr>
<td>037</td>
<td>31</td>
<td>&quot;</td>
<td>Unit separator</td>
</tr>
<tr>
<td>040</td>
<td>32</td>
<td>-</td>
<td>Space</td>
</tr>
<tr>
<td>041</td>
<td>33</td>
<td>_</td>
<td>Horizontal tab</td>
</tr>
<tr>
<td>042</td>
<td>34</td>
<td>&quot;</td>
<td>Double quote</td>
</tr>
<tr>
<td>043</td>
<td>35</td>
<td>＃</td>
<td>Name sign</td>
</tr>
<tr>
<td>044</td>
<td>36</td>
<td>$</td>
<td>Dollar sign</td>
</tr>
<tr>
<td>045</td>
<td>37</td>
<td>%</td>
<td>Percent</td>
</tr>
<tr>
<td>046</td>
<td>38</td>
<td>&amp;</td>
<td>Apostrophe</td>
</tr>
<tr>
<td>047</td>
<td>39</td>
<td>-</td>
<td>Cloning single quote (apostrophe)</td>
</tr>
<tr>
<td>050</td>
<td>40</td>
<td>{</td>
<td>Left parenthesis</td>
</tr>
<tr>
<td>051</td>
<td>41</td>
<td>]</td>
<td>Right parenthesis</td>
</tr>
<tr>
<td>052</td>
<td>42</td>
<td>*</td>
<td>Asterisk</td>
</tr>
<tr>
<td>053</td>
<td>43</td>
<td>+</td>
<td>Plus sign</td>
</tr>
<tr>
<td>054</td>
<td>44</td>
<td>,</td>
<td>Comma</td>
</tr>
<tr>
<td>055</td>
<td>45</td>
<td>-</td>
<td>Minus sign or dash</td>
</tr>
<tr>
<td>056</td>
<td>46</td>
<td>.</td>
<td>Period</td>
</tr>
<tr>
<td>057</td>
<td>47</td>
<td>/</td>
<td>Forward slash</td>
</tr>
<tr>
<td>060</td>
<td>48</td>
<td>0</td>
<td>Digit zero</td>
</tr>
<tr>
<td>061</td>
<td>49</td>
<td>1</td>
<td>Digit one</td>
</tr>
<tr>
<td>062</td>
<td>50</td>
<td>2</td>
<td>Digit two</td>
</tr>
<tr>
<td>063</td>
<td>51</td>
<td>3</td>
<td>Digit three</td>
</tr>
<tr>
<td>064</td>
<td>52</td>
<td>4</td>
<td>Digit four</td>
</tr>
<tr>
<td>065</td>
<td>53</td>
<td>5</td>
<td>Digit five</td>
</tr>
<tr>
<td>066</td>
<td>54</td>
<td>6</td>
<td>Digit six</td>
</tr>
<tr>
<td>067</td>
<td>55</td>
<td>7</td>
<td>Digit seven</td>
</tr>
<tr>
<td>068</td>
<td>56</td>
<td>8</td>
<td>Digit eight</td>
</tr>
<tr>
<td>069</td>
<td>57</td>
<td>9</td>
<td>Digit nine</td>
</tr>
<tr>
<td>070</td>
<td>58</td>
<td>_</td>
<td>Colon</td>
</tr>
<tr>
<td>071</td>
<td>59</td>
<td>-</td>
<td>Semicolon</td>
</tr>
<tr>
<td>072</td>
<td>60</td>
<td>≤</td>
<td>Less than or equal to</td>
</tr>
<tr>
<td>073</td>
<td>61</td>
<td>&gt;</td>
<td>Greater than</td>
</tr>
<tr>
<td>074</td>
<td>62</td>
<td>?</td>
<td>Question mark</td>
</tr>
</tbody>
</table>

American Standard Code for Information Interchange (ASCII) specifies a correspondence between bit patterns and character symbols. The octal and decimal representations of the bit patterns are shown along with a description of the character symbol. The first 32 codes (numbers 0–31 decimal) as well as the last (number 127 decimal) are non-printing characters which were initially intended to control devices or provide meta-information about data streams. For example, decimal 10 ended a line on a line printer and decimal 8 backspaced one character so that the preceding character would be overstruck. The control sequence column shows the traditional key sequences for inputting these non-printing characters where the caret (^) represents the “Control” or “Ctrl” key which must be held down while the following key is depressed.
Boundary-value ordinary differential equations. Sect. 11 (130), 133, 136, 140; 
continuous method, 132 
bread, 99, 132, 156
bessel, 27, 132, 156, 161
besselC, 130, 131, 134, 161
besselK, 130, 131, 134, 161
besselj, 130, 131, 134, 161
bessely, 130, 131, 134, 161
Binomial theorem, 132

C

see, 15, 17, 122
C (programming language), 7, 23, 40, 43, 44, 46, 55, 66,
83, 84, 94, 106, 132
C++ (programming language), 44, 112
Calculator. Subject, 1.1 (6), 2.4, 39, 42, 140, 152
case, 94, 95, 152
differentiation in C, 94
Case sensitive, 94
cat, 37, 38, 122
Catching errors, 112
caxis, 64, 66, 156
cell, 44, 46, 47, 157
cell array, 39, Subsect. 3.4 (43), 54
equal (see two cell arrays), 99
cedisp, 44, 46, 157
char, 47
character string, 7, 8, Subsect. 3.3 (42), 157
appending to, 42
converting strings, 29
concatenating, 42
casting, 111
find substring in, 96
multi-line, 42
putting a number of strings in lexicographical order, 44
TkC commands in, 62, 67
check, 85, 90, 156
Cholesky decomposition, 85
clabel, 59, 141, 156
class, 47, 49, 162
Class, Subsect. 3.5 (46)
callback, 48
event, 48
generate, 46
fundamental, 46
fundamental (table), 47
generate, 48
handle, 48
listener, 48
subclass, 48
user-defined, 47
value, 48
classdef, 48
clear, 9, 83, 12, 134, 161
danger in testing, 123
Clear (a figure), 63
See also clf and close
cif, 63, 66, 156
close, 64, 66, 156
Cloze, 107, 123, 137
Clown, 65
Colon operator (eyes), 44, 46
possible Floating-point errors in, 23, 49
See also linestack
Color map, 63
colorbar, 63, 66, 156
colormap, 63, 66, 156
Color, See RGB components
Command, 3
See also Function
Comment character, 15, 149
Complex conjugate, 14
Complex numbers, 6, Subsect. 1.6 (13)
Concrete of Nocedal
cond, 81, 85, 156
condct, 85, 156
Condition number, See Matrix conj., 14, 152
Conjugate transpose, 19
See also Transpose
Continuation (of a line), 15, 149
Continuation method, 132
continues, 94
contour, 59, 62, 157
Contour plot, 59
contourf, 59, 62, 157
Contourf, 59, 62, 157
Control flow, See Programming language
cov., 11, 157

disp., 14
dout, 14, 15, 157
draw, 14, 15, 157
drawnow, 59, 62, 157
drawnow, 59, 62, 157
D的眼睛, See Initial-value ordinary differential equations

E

 câ, 14
Earth, 65
echo, 154, 161, 169
edit, 14, 15, 154
err, 32, 38, 107, 112
Evalsh, 52, 55, 86, 97, 90, 162
See also: Matrix
eval, 86, 90, 102
evalc, 87, 92, 158
evald, 87, 95, 158
evalf, 87, 92, 94, 97, 108, 111, 158, 169
exp, 104, 109, 112
See also: Matrix
eval (a figure), 63
See also: M
error, 102, 111, 156, 160
Error bars, 54, 55
errorbar, 52, 55, 157
Euclidean length, See Length of a vector
eval, 111, 112, 113, 117
events, 105, 109
EXCLUSIVE OR (logical operator), 51, 65, 158

B

Ball, 127, 128
Ballot's equation. See Initial-value ordinary differential equations
Binary format, 16, 56
blank, 48, 157

D

D. best polynomial fit to, 131
closing files, 83
manipulation. Subsect. 2.8 (34), 157
opening files, 83
reading into MATLAB, 55, 57, 63, 118, 152
writing from MATLAB, 55, 57, 63, 118, 152
Data types, 40, Subsect. 3.5 (46)
fundamental, fundamental (table), 47
declear, 157, 158
decent, 157, 158
downd, 157, 158
dup, 157, 158
function [x] = fastinterp(xdata, ydata, xquery)
    % fast interpolation
    % 
    % Usage:
    % 
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % xdata: vector of x values
    % ydata: vector of y values
    % xquery: vector of query points
    % 
    % Output:
    % 
    % x: vector of interpolated values corresponding to xquery
    % 
    % The function fastinterp implements a fast interpolation method for
    % efficiently querying interpolated values from a given dataset.
    % It is designed to handle large datasets where speed is crucial.
    % 
    % Example:
    % 
    % xdata = [1:10];
    % ydata = [2:2:20];
    % xquery = [1:0.1:10];
    % 
    % x = fastinterp(xdata, ydata, xquery);
    % 
    % Notes:
    % 
    % This function assumes that xdata is always sorted in ascending order.
    % The interpolation method used is a simple linear interpolation.
    % 
    % See also:
    % 
    % References:
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % Documentation:
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterp(xdata, ydata, xquery)
    % 
    % fastinterpre...
Variables (Cont.)

conflict between variable and function name, 12
defined, 98
deleting, 7
global, 105, 106
inputting, 10
list of, 36
loading, 16
local, 99, 105, 106
logical, 97
modifying, 105, 106
overwriting, 7
persistent, 106
predefined, 7, 9, 150
ans, 8, 9, 150
eps, 2, 70, 94, 150
i, 9, 150
Inf, 2, 150
j, 6, 7, 150
ihat, 7, 150
overwriting, 7, 91
pi, 7, 9, 150
realmax, 7, 150
realm, 7, 150
reverse two, 10, 150
saving, 16
saving local variables in functions, 106
scope of, 159
special cases of vectors or matrices, 7
state, 106
string, 7; Subsect. 3.3 (42)
See also Character string
typeless, 7, 105
valid names, 8
Vector
average value of elements, 35
column vs. row, 18
deleting elements, 20
equal (see two vectors), 98
generating, Subsect. 2.1 (18)
individual elements, 20
“masking” elements of, 97
maximum value, 34
mean value of elements, 35
minimum value, 35
preallocation of, 20, 44, 101
repeated elements, testing for, 36
sort elements, 35
standard deviation of elements, 35
sum of elements, 35
vectorize, 41
Vectorizing code, Subsect. 8.5 (113)
profile execution time, 113
view, 59, 64, 155

W
while, 23, 94, 95, 155
who, 16, 21, 155
Workspace, 6, 104