LABORATORY MANUAL

ECE 403/503

OPTIMIZATION for MACHINE LEARNING

This manual was prepared by

Wu-Sheng Lu

University of Victoria
Department of Electrical and Computer Engineering

© University of Victoria
May 2019

A. SAFETY REGULATIONS

Safe practice in the laboratory requires an open attitude and knowledgeable awareness of potential hazards. Safety is a collective responsibility and requires full cooperation from everyone in the laboratory. This cooperation means each student and instructor must observe standard safety precautions and procedures and should:

1. Follow all safety instructions carefully.
2. Become thoroughly acquainted with the location and use of safety facilities such as fire extinguishers, first aid kits, emergency showers, eye- wash stations and exits. See marked floor plan posted near exit staircases for equipment locations.
3. Become familiar with experimental procedures and all potential hazards involved before beginning an experiment.

ELECTRICAL HAZARDS

General Safety Principles

Electrical currents of astonishingly low amperage and voltage under certain circumstances may result in fatal shock. Low-voltage dc circuits do not normally present a hazard to human life, although severe burns are possible. Voltage as low as 24 volts ac can be dangerous and present a lethal threat. The time of contact with a live circuit affects the degree of damage, especially where burns are concerned. Very small electrical shocks, even small "tingles", should serve as a warning that an electrical problem exists and that a potentially dangerous shock can occur. The equipment and circuits in use must be immediately disconnected and not re-used until the problem is discovered and corrected by the instructor and/or technologist.

1. When handling electric wires, never use them as supports and never pull on live wires.
2. Report and do not use equipment with frayed wires or cracked insulation and equipment with damaged plugs or missing ground prongs.
3. Report and do not use receptacles with loose mountings and/or weak gripping force.
4. Avoid pulling plugs out of receptacles by the cord and avoid rolling equipment over power cords.
5. Be sure line-powered equipment has 3-wire grounding cords and that you know how to use the equipment properly. Ask for help and instruction when needed.
6. Any electrical failure or evidence of undue heating of equipment should be reported immediately to the instructor and/or technologist. If you smell over-heating components or see smoke coming from any circuit or equipment, switch the power off immediately.
7. Ensure all equipment is powered-off at the end of each experiment.
8. Only qualified ECE personnel should maintain electric or electronic equipment.
9. Cardiopulmonary resuscitation (CPR) often will revive victims of high-voltage shock. Only qualified people should attempt CPR.

FIRST AID

1. For simple cuts or minor first aid use the First Aid Kits available in each room. The University Health Services may also be contacted at **8492**. All injuries, no matter how minor, should be reported to your instructor and/or technologist.
2. For medical emergencies call **911** and Traffic and Security at **7599**.

INDIVIDUAL RESPONSE PROCEDURES FOR FIRES

If **you discover** a fire, smoke or an explosion:

1. Shout for assistance.
2. Activate the nearest fire alarm.
3. If it is a small fire, attempt to put it out with available fire equipment. See the marked floor plan posted near exit staircases for equipment locations.
4. If the fire is out of control and it is too large to handle with one fire extinguisher, isolate the fire by closing the doors and windows behind you as you leave. Do **not** lock the doors.
5. Warn others and leave the building with reasonable speed using recommended exits. Assist disabled and injured persons in reaching assembly areas when conditions permit.
6. Stand by to identify yourself and provide information to fire personnel.

If a **Fire Alarm sounds**:

1. Secure any equipment you are using and switch the power off.
2. Close windows and doors behind you as you leave. Do **not** lock doors.
3. Leave the building with reasonable speed using recommended exit.
4. Follow instructions of your floor warden or deputy. Wardens will ensure evacuation of assigned rooms.
5. **DO NOT** use elevators for evacuation.
6. **DO NOT** re-enter the building until allowed to do so by the Fire Department.

INDIVIDUAL RESPONSE PROCEDURES FOR EARTHQUAKES

IF INDOORS:

Take action at the first indication of ground shaking.

1. Stay inside; move away from windows, shelves, heavy objects and furniture that may fall. Take cover under a table or a desk, or in a strong doorway (anticipate that doors may slam shut).
2. In halls, stairways or other areas where no cover is available, move to an interior wall. Turn away from windows, kneel alongside the wall, bend your head close to your knees, clasp your hands firmly behind your neck covering the sides of your head with your elbows.
3. Elevators must not be used. They are extremely vulnerable to damage from earthquakes. Ground shaking may cause counterweights and other components to be torn from their connections, causing extensive damage to the elevator cabs and operating mechanisms.
4. When exiting a building, move quickly through exits and away from buildings. Parapets and columns supporting roof overhangs may fall.
5. Assemble away from gas, sewer and power lines.

IF OUTDOORS:

1. Move to an open space away from buildings, trees and overhead power lines.
2. Lie down or crouch low to the ground (legs will not be steady) and constantly survey the area for additional hazards.

B. LABORATORY OPERATION GUIDELINES

During the operation of the laboratories, the following simple procedures and guidelines are essential and **must** be adhered to by all students.

- **FOOD, DRINKS and SMOKING** are **NOT** permitted in the laboratories.
- Before starting your experiment, make sure proper equipment and circuit connections are made as per instructions in the laboratory manual. Verify your set-up with your instructor.
- All damaged or missing equipment or parts should be reported as soon as possible to the technologist. A Repair Request form (available in each lab) must be completed before equipment can be serviced.
- Equipment should not be removed from the lab station. If equipment is required elsewhere, it is to be returned to the lab station once the requirement is finished.
- All electronic components, such as capacitors, resistors, transistors etc., must be returned to their respective storage trays when the experiment is finished.
- All leads and cables are to be returned to the wall racks when the lab is finished and oscilloscope probes are to remain with the oscilloscope.
- Benches and equipment set-ups are to be tidied up after each lab session. All garbage is to be placed in the garbage cans provided. No writing on equipment or benches is permitted.
- If students have any questions about the experiment, they should consult the instructor first and then the technologist.
- Abide by all safety rules and regulations of the laboratory.

ELECTRICAL AND COMPUTER ENGINEERING LABORATORIES

ACKNOWLEDGMENT FORM

NAME: ______________________________

COURSE: ______________________________

LAB INSTRUCTOR: ______________________________

Experiments conducted in the Electrical and Computer Engineering laboratories involve the handling of electric and electronic equipment and circuits. Failure to handle this equipment properly may lead to injury or even fatal shock. For the safety of everyone, it is required that you understand and follow the appropriate laboratory procedures as outlined in the laboratory manuals and by your laboratory instructor.

Your signature below is your acknowledgment that you have read the *Safety Regulations* and the *Laboratory Operation Guidelines* and agree to abide by them.

______________________________ ______________________________
STUDENT'S SIGNATURE DATE

____
STUDENT NUMBER

LIST OF EXPERIMENTS

Experiment 1 - Handwritten Digits Recognition Using PCA ……………………..……..... 1

Experiment 2 - Multi-Category Classification Using Binary Linear Classifiers ..……... 5

Experiment 3 - Predicting Energy Efficiency for Residential Buildings…..……………. 9

Experiment 4 - Breast Cancer Diagnosis via Logistic Regression ……………………13

Appendix A - An Introduction to MATLAB by D. F. Griffiths

Preparing Your Laboratory Report

The objective of the experiments described in this manual is to familiarize the student with computer simulations and implementation of several optimization as well as data processing techniques as they are applied to machine learning problems. The primary software tool required in all experiments is MATLAB to which the student can access during the laboratory sessions. An appendix, that introduces main functions in MATLAB and their usage, is included to facilitate the students in preparing their MATLAB code.

PREPARATION

Successful completion of an experiment depends critically on error-free MATLAB programming. Therefore, preparation prior to the laboratory period is essential. Specifically, the student should *study the description of the experiment and prepare useable MATLAB codes required in the preparation section(s)* ***before*** *the experiment is carried out*. The student will be required to present the preparation at the beginning of the lab session.

THE LABORATORY REPORT

A laboratory report is required from each group for each experiment performed. The lab report should be submitted **within one week after the experiment**. The front page of the report is shown on the next page and should be used for each laboratory report.

The report should be divided into the following parts:

(a) Objectives.
(b) Introduction.
(c) Results including relevant MATLAB programs and figures, and description of the implementations.
(d) Discussion.
(e) Conclusions.

UNIVERSITY OF VICTORIA
Department of Electrical and Computer Engineering

ECE 403/503 Optimization for Machine Learning

LABORATORY REPORT

Experiment No:

Title:

Date of Experiment:
(should be as scheduled)

Report Submitted on:

To:

Laboratory Group No.:

Name(s): (please print)

Experiment 1

Handwritten Digits Recognition Using PCA

1. Introduction and Objectives

Research for handwritten digit recognition (HWDR) by machine learning (ML) techniques has stayed active in the past several decades. The sustained interest in HWDR is primarily due to its broad applications in bank check processing, postal mail sorting, automatic address reading, and mail routing, etc. In these applications, both accuracy and speed of digit recognition are critical indicators of system performance. In a machine learning setting, we are given a training data set consisting of a number of samples of handwritten digits, each belongs to one of the ten classes $\{\mathcal{D}_j, j = 0, 1, \ldots, 9\}$ where class $\mathcal{D}_j$ collects all data samples labeled as digit j. The HWDR problem seeks to develop an approach to utilizing these known data classes to train a multi-category classifier so as to recognize handwritten digits *outside* the training data. The primary challenge arising from the HWDR problem lies in the fact that handwritten digits (within the same digit class) vary widely in terms of shape, line width, and style, even when they are properly centralized and normalized in size.

Classification of multi-category data can be accomplished by ML methods that are originally intended for classifying binary-class data using the so-called *one-versus-the-rest* approach. Alternatively, there are ML techniques that deal with multi-category data directly. A representative method of this type is based on principal component analysis (PCA). In Sec. 1.3 of the course notes, PCA is introduced as a means for feature extraction. The objective of this laboratory experiment is to learn PCA as a technique for pattern recognition and apply it to the HWDR problem.

2. The Dataset

The database MNIST for HWDR was introduced in Example 1.1 of Section 1.2 of the course notes. A full-scale training data from MNIST fit into the structure $\mathcal{D} = \{(\boldsymbol{x}_n, y_n), n = 1, 2, \ldots, N\}$ with $N = 60{,}000$, where, for each digit $\boldsymbol{x}_n$, a label is chosen from $y_n \in \{0, 1, \ldots, 9\}$ to match what $\boldsymbol{x}_n$ represents. Originally each $\boldsymbol{x}_n$ is a gray-scale digital image of 28×28 pixels, with components in the range $[0, 1]$ with 0 and 1 denoting most white and most black pixels, respectively. In this experiment, each $\boldsymbol{x}_n$ has been converted into a column vector of dimension $d = 784$ by stacking matrix $\boldsymbol{x}_n$ column by column. In this way, each digit from MNIST can be regarded as a "point" in the 784-dimensional Euclidean space. If we put the entire training data together, column by column, to form a matrix $\boldsymbol{X} = [\boldsymbol{x}_1 \quad \boldsymbol{x}_2 \quad \cdots \quad \boldsymbol{x}_m]$, then $m = 60{,}000$ and $\boldsymbol{X}$ has a size of 784×60000.

3. PCA for Multi-Category Classification

In the HWDR problem, we are given a training set that consists of 10 data classes of the form $\mathcal{D}_j = \{(\boldsymbol{x}_i^{(j)}, y_j), i = 1, 2, \ldots, n_j\}$ for $j = 0, 1, \ldots, 9$, where all samples $\{\boldsymbol{x}_i^{(j)}\}$ in class $\mathcal{D}_j$ are associated with the same label $y_j = j$. Our goal is to classify (i.e. recognize) a new digit $\boldsymbol{x}$ *outside*

the training data as one of 10 possible digits. To do so, PCA is applied to *each* data class $\{(\boldsymbol{x}_i^{(j)}, y_j), i = 1,2,...,n_j\}$ by first computing the mean vector and covariance matrix as

$$\boldsymbol{\mu}_j = \frac{1}{n_j}\sum_{i=1}^{n_j}\boldsymbol{x}_i^{(j)}, \quad \boldsymbol{C}_j = \frac{1}{n_j}\sum_{i=1}^{n_j}(\boldsymbol{x}_i^{(j)} - \boldsymbol{\mu}_j)(\boldsymbol{x}_i^{(j)} - \boldsymbol{\mu}_j)^T \quad \text{for } j = 0, 1, \ldots, 9 \tag{E1.1}$$

and then performing eigen-decomposition of $\boldsymbol{C}_j$ as

$$\boldsymbol{C}_j = \boldsymbol{U}_j\boldsymbol{S}_j\boldsymbol{U}_j^T \quad \text{for } j = 0, 1, \ldots, 9 \tag{E1.2}$$

Next, we construct a rank-q approximation for each $\boldsymbol{C}_j$ as

$$\boldsymbol{C}_j \approx \boldsymbol{U}_q^{(j)}\boldsymbol{S}_q^{(j)}\boldsymbol{U}_q^{(j)T}$$

where

$$\boldsymbol{U}_q^{(j)} = [\boldsymbol{u}_1^{(j)}\ \boldsymbol{u}_2^{(j)} \cdots \boldsymbol{u}_q^{(j)}] \qquad \text{for } j = 0, 1, \ldots, 9 \tag{E1.3}$$

consists of q eigenvectors of $\boldsymbol{C}_j$ corresponding to the q largest eigenvalues and $\boldsymbol{S}_q^{(j)}$ is a diagonal matrix that collects the q largest eigenvalues in descent order. So now we get 10 pairs $\{\boldsymbol{\mu}_j, \boldsymbol{U}_q^{(j)}\}$ for $j = 0, 1, \ldots, 9$, where $\boldsymbol{\mu}_j$ denotes the average character of the jth class $\mathcal{D}_j$, and $\boldsymbol{U}_q^{(j)}$ are q principal axes that define a q-dimensional subspace in which the variations of the individual members in the jth class from its mean $\boldsymbol{\mu}_j$ can be well represented. The dimension reduction becomes substantial when $q \ll d$. It is important to note that the quantities described above can be prepared off-line prior to the real-time task of classifying a new sample $\boldsymbol{x}$. The classification of a point $\boldsymbol{x}$ is carried out in the following steps:

(i) Project point $\boldsymbol{x} - \boldsymbol{\mu}_j$ into the jth data class for $j = 0, 1, \ldots, 9$. These projections are performed by computing principal components

$$\boldsymbol{f}_j = \boldsymbol{U}_q^{(j)T}(\boldsymbol{x} - \boldsymbol{\mu}_j) \quad \text{for } j = 0, 1, \ldots, 9 \tag{E1.4}$$

(ii) Represent point $\boldsymbol{x}$ in the jth class as

$$\hat{\boldsymbol{x}}_j = \boldsymbol{U}_q^{(j)}\boldsymbol{f}_j + \boldsymbol{\mu}_j \quad \text{for } j = 0, 1, \ldots, 9 \tag{E1.5}$$

(iii) Compute the Euclidean distance between $\boldsymbol{x}$ and its representation $\hat{\boldsymbol{x}}_j$ as

$$e_j = \| \boldsymbol{x} - \hat{\boldsymbol{x}}_j \|_2 \quad \text{for } j = 0, 1, \ldots, 9 \tag{E1.6}$$

(iv) Sample $\boldsymbol{x}$ is classified to class j^* if e_{j^*} reaches the minimum among $\{e_j, j = 0, 1, \ldots, 9\}$, that is,

$$\text{sample } \boldsymbol{x} \text{ is classified to class } j^* = \arg\min_{j=0,1,\ldots,9}\{e_j\} \tag{E1.7}$$

Below we summarize the complete procedure as an algorithm.

PCA-Based Multi-Category Classification Algorithm

Input: Training data that contains 10 classes $\mathcal{D}_0, \mathcal{D}_1, \ldots, \mathcal{D}_9$; the number q of principal axes to be used; and a new data point $\boldsymbol{x}$ to be classified.

Step 1 Compute $\boldsymbol{\mu}_j$ and $\boldsymbol{C}_j$ using (E1.1) for each data class $\mathcal{D}_j$, for $j = 0, 1, \ldots, 9$.

Step 2 Compute the q eigenvectors of $\boldsymbol{C}_j$ corresponding to the q largest eigenvalues and use them to construct matrix $\boldsymbol{U}_q^{(j)} = [\boldsymbol{u}_1^{(j)}\ \boldsymbol{u}_2^{(j)} \cdots \boldsymbol{u}_q^{(j)}]$. This step is performed for $j = 0, 1, \ldots, 9$ to get 10 matrices $\left\{\boldsymbol{U}_q^{(j)}, j = 0,1,\ldots,9\right\}$.

Step 3 Use (E1.4) to calculate principal components $\left\{\boldsymbol{f}_j, j = 0,1,\ldots,9\right\}$.

Step 4 Use (E1.5) to calculate $\left\{\hat{\boldsymbol{x}}_j, j = 0,1,\ldots,9\right\}$.

Step 5 Use (E1.6) to calculate $\left\{e_j, j = 0,1,\ldots,9\right\}$.

Step 6 Use (E1.7) to Identify the target class j^* for data point $\boldsymbol{x}$.

Remarks

(i) **`eigs`** is a convenient MATLAB function for implementing Step 2 of the algorithm: For a square and symmetric matric $\boldsymbol{C}$ and an integer $q > 0$, **`[Uq,Sq] = eigs(C,q)`** returns two items, namely **`Uq`** and **`Sq`**, where **`Uq`** contains q eigenvectors of $\mathbf{C}$ corresponding to the q largest eigenvalues.

(ii) It is important to realize that the 10 pairs $\{\boldsymbol{\mu}_j, \boldsymbol{U}_q^{(j)} \text{ for } j = 0,1,\ldots,9\}$ can be calculated *before* the real-time classification of new digits begins. Moreover, these pre-calculated items can be used not only for one single test sample but for the *entire* set of new digits.

4. Procedure

4.1 From the course website download the data and MATLAB function listed below.

- **`X1600.mat`** – contains a total of 16,000 handwritten digits selected at random from the training data of MNIST database. Note that **`X1600.mat`** is a matrix of size 784 × 16000, with each column representing a digit, and it contains 10 classes of digits, each class contains 1600 samples. **`X1600`** is structured so that its first 1600 columns contain 1600 digit "0", and the next 1600 columns contain 1600 digit "1", and so on.

- **`Te28.mat`** – contains 10,000 handwritten digits from testing purposes. **`Te28.mat`** is matrix of size 784 × 10,000, with each column representing a digit for testing the performance of the classifier you are to build.

- **`Lte28.mat`** – contains 10,000 integers between 0 and 9, each integer is the label of the corresponding digit from **`Te28.mat`**.

4.2 Prepare MATLAB code to implement the algorithm described in Sec. 3 above. It is desirable to prepare the code so that it can handle any number of test samples (see Remark (ii) above). It is recommended that the number q of principal axes be set to $q = 29$. Run the code with **`Te28`** as the test data.

4.3 Evaluate and report the performance of the PCA-based classifier by comparing the classification results with the labels in **`Lte28.mat`**. Report the total number of misclassifications as well as overall error rate in percentage.

4.4 Evaluate and report the efficiency of the PCA-classifier in terms of the CPU time (in seconds) consumed by the classifier per 1000 digits. Note that for a fair evaluation the CPU time for performing Steps 1 and 2 of the algorithm should *not* be counted. Instead, the CPU time should cover the time consumed by Steps 3-6.

Include your MATLAB code in the lab report please.

Experiment 2
Multi-Category Classification Using Binary Linear Classifiers

1. Introduction and Objective

In this experiment, we investigate a technique for multi-category classification based on binary classifications. The technique is then applied to Fisher's 3-class datasets of Iris plants to demonstrate its effectiveness. The dataset of Iris plants to be used in this experiment was created and published in 1936 by R. A. Fisher [R1]. Fisher's paper is a classic in the field and is referenced frequently to this day, as a matter of fact the dataset is arguably the best-known in the pattern recognition literature [R2]. The dataset includes features of 150 Iris plants of 3 species known as Setosa, Versicolor, and Virginica, where each sample Iris is represented by a 4-dimensional vector in terms of lengths and widths of the sepal and petal of the flower.

References

[R1] R. A. Fisher, "The use of multiple measurements in taxonomic problems", *Annual Eugenics*, vol. 7, part II, pp. 179-188, 1936.

[R2] UCI Machine Learning, http://archive.ics.uci.edu/ml, University of California Irvine, School of Information and Computer Science.

2. Background

2.1 The idea of multi-category classification using linear binary classifiers

Consider a labeled dataset $\mathcal{D}$ that contains a total of K data classes with $K > 2$, namely,

$$\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2 \cup \cdots \cup \mathcal{D}_K$$

The technique is known as *one-versus-the-rest*, and it is built on a simple process of converting the data classes at hand into a *two-class* scenario so that binary classification becomes applicable. The process is run K times, each time the original datasets are grouped into a "*positive class*" and a "*negative class*" appropriately so as to identify an optimal linear function for classifying these two classes. The K linear functions so obtained are then applied to the test data for multi-category classification.

To start, we single out class $\mathcal{D}_1$ and assign label $y_n = 1$ to all its samples and re-name the data class as "positive class" $\mathcal{P}$. And we combine the rest of the classes into *one* class and assign label $y_n = -1$ to all its samples and call it "negative class" $\mathcal{N}$. That is,

$$\begin{cases} \mathcal{P} = \mathcal{D}_1 \text{ with } y_n \equiv 1 \\ \mathcal{N} = \mathcal{D}_2 \cup \cdots \cup \mathcal{D}_K \text{ with } y_n \equiv -1 \end{cases}$$

In this way, one deals with a two-class dataset so that a linear model $f_1(\boldsymbol{x}, \boldsymbol{w}_1, b_1) = \boldsymbol{w}_1^T \boldsymbol{x} + b_1$ is obtained for binary classification by doing its best to separate class $\mathcal{P}$ from class $\mathcal{N}$.

Next, the process described above is repeated in that we single out class $\mathcal{D}_2$ and assign label $y_n = 1$ to all its samples and re-name it as "positive class" $\mathcal{P}$, while combining the rest of the classes into *one* class and assign label $y_n = -1$ to all its samples and call it "negative class" $\mathcal{N}$. A linear model $f_2(\boldsymbol{x}, \boldsymbol{w}_2, b_2) = \boldsymbol{w}_2^T \boldsymbol{x} + b_2$ is obtained based on the updated data sets $\mathcal{P}$ and $\mathcal{N}$.

The process continues in a similar fashion until K such linear models $f_i(\boldsymbol{x}, \boldsymbol{w}_i, b_i) = \boldsymbol{w}_i^T \boldsymbol{x} + b_i$ for $i = 1, 2, \ldots, K$ are obtained. Classification of a new sample $\boldsymbol{x}$ outside the dataset is then performed by the classifier below:

$\boldsymbol{x}$ belongs to class i^* if $f_{i^*}(\boldsymbol{x}, \boldsymbol{w}_{i^*}, b_{i^*})$ reaches maximum among $\{f_i(\boldsymbol{x}, \boldsymbol{w}_i, b_i) \text{ for } i = 1, 2, \ldots, K\}$ (E2.1)

2.2 Linear classifier for binary classification

The linear model for binary (i.e. two-class) classification was studied in Sec. 1.5, see pages 24-26 of the course notes. Let the two data classes $\mathcal{P}$ and $\mathcal{N}$ assume the form

$$\mathcal{P} = \{(\boldsymbol{x}_i^{(P)}, 1),\ i = 1, 2, \ldots, N_p\} \text{ and } \mathcal{N} = \{(\boldsymbol{x}_i^{(N)}, -1),\ i = 1, 2, \ldots, N_n\}$$

respectively.

From Eq. (1.34) of the course notes, it follows that the optimal parameters of linear function $f(\boldsymbol{x}, \boldsymbol{w}, b) = \boldsymbol{w}^T \boldsymbol{x} + b$ that best separate class P from class N can be found by solving the system of linear equations

$$\hat{\boldsymbol{X}}^T \hat{\boldsymbol{X}} \hat{\boldsymbol{w}} = \hat{\boldsymbol{X}}^T \boldsymbol{y} \tag{E2.2a}$$

where

$$\hat{\boldsymbol{w}} = \begin{bmatrix} \boldsymbol{w} \\ b \end{bmatrix}, \quad \hat{\boldsymbol{X}} = \begin{bmatrix} \boldsymbol{x}_1^{(P)T} & 1 \\ \vdots & \vdots \\ \boldsymbol{x}_{N_p}^{(P)T} & 1 \\ \boldsymbol{x}_1^{(N)T} & 1 \\ \vdots & \vdots \\ \boldsymbol{x}_{N_n}^{(N)T} & 1 \end{bmatrix}, \quad \boldsymbol{y} = \begin{bmatrix} 1 \\ \vdots \\ 1 \\ -1 \\ \vdots \\ -1 \end{bmatrix} \tag{E2.2b}$$

・Note that in this experiment parameter $\boldsymbol{w}$ is a vector of length 4, and b is a scalar; the sizes of classes $\mathcal{P}$ and $\mathcal{N}$ are set to $N_p = 40$ and $N_n = 80$; and $\boldsymbol{y}$ is a constant vector of length 120, with its first 40 components being 1 and the rest of 80 components being -1.

2.3 The dataset

Fisher's dataset is available from the course website as **`D_iris.mat`**. To be precise, **`D_iris`** is a matrix of size 5×150 whose first 4 rows are the data of the 150 samples, and the last row contains labels of the 150 flowers. Be aware that matrix **`D_iris`** was constructed so that its first 50 columns are associated with Iris Setosa, the next 50 columns are for Iris Versicolor, and the last 50 columns are for Iris Virginica.

Once **`D-iris.mat`** is downloaded, the code below prepares the datasets.

```
D = D_iris(1:4,:);
X1 = D(:,1:50);
X2 = D(:,51:100);
X3 = D(:,101:150);
```

For each data class of 50 samples, 40 samples are used for training and the remaining 10 samples are used for testing. Random selection of the samples is made for all three data classes as follows.

```
rand('state',111)
r1 = randperm(50);
Xtr1 = X1(:,r1(1:40));
Xte1 = X1(:,r1(41:50));
```

```
rand('state',112)
r2 = randperm(50);
Xtr2 = X2(:,r2(1:40));
Xte2 = X2(:,r2(41:50));
rand('state',113)
r3 = randperm(50);
Xtr3 = X3(:,r3(1:40));
Xte3 = X3(:,r3(41:50));
```

· Data sets `Xtr1`, `Xtr2`, and `Xtr3` are train data from Setosa, Versicolor, and Virginica, respectively, each contains 40 samples. Data sets `Xte1`, `Xte2`, and `Xte3` are test data from Setosa, Versicolor, and Virginica, respectively, each contains 10 samples.

3. Procedure

3.1 From the course website download data matrix `D_iris.mat`

3.2 Follow Sec. 2.3 above to prepare training and testing datasets.

3.3 Based on the material in Secs. 2.1 and 2.2 above to prepare MATLAB code to carry out the procedure below. Note that in this lab experiment the number of classes is $K = 3$. Consequently, 3 binary classifications are required to produce linear models

$$\begin{cases} f_1(\boldsymbol{x}, \boldsymbol{w}_1, b_1) = \boldsymbol{w}_1^T \boldsymbol{x} + b_1 \\ f_2(\boldsymbol{x}, \boldsymbol{w}_2, b_2) = \boldsymbol{w}_2^T \boldsymbol{x} + b_2 \\ f_3(\boldsymbol{x}, \boldsymbol{w}_3, b_3) = \boldsymbol{w}_3^T \boldsymbol{x} + b_3 \end{cases}$$

Also note that vector $\boldsymbol{y}$ in (E2.2b) is a constant vector of size 120 $\times$ 1 throughout the procedure, and is given by

```
y = [ones(40,1); -ones(80,1)];
```

(i) Use dataset `Xtr1` as class $\mathcal{P}$ and datasets `Xtr2`, and `Xtr3` combined as class $\mathcal{N}$ to prepare matrix $\hat{X}$. Compute optimal parameters $\boldsymbol{w}_1$ and b_1 by solving the equation in (E2.2a). This allows to construct a linear model $f_1(\boldsymbol{x}, \boldsymbol{w}_1, b_1)$.

(ii) To obtain the second linear model, repeat Step (i) above, where classes $\mathcal{P}$ and $\mathcal{N}$ need to be re-constructed, see Sec. 2.1 above. This will let you prepare correct matrix $\hat{X}$ and hence a correct model $f_2(\boldsymbol{x}, \boldsymbol{w}_2, b_2)$.

(iii) Repeat Step (ii) above with appropriately re-constructed classes $\mathcal{P}$ and $\mathcal{N}$ to establish the third linear model $f_3(\boldsymbol{x}, \boldsymbol{w}_3, b_3)$.

(iv) Apply the classifier in (E2.1) to the test data sets {`Xte1`,`Xte2`, `Xte3`} by performing the following steps:

Step 0: Set counter `mis_class = 0;` Assign label $y_k = 1$ to all samples in `Xte1;` label $y_k = 2$ to all samples in `Xte2;` and label $y_k = 3$ to all samples of `Xte3.`

Step 1: For a given test sample $\boldsymbol{x}$, apply the 3 linear models obtained above. This yields

three real values $f_i = \boldsymbol{w}_i^T \boldsymbol{x} + b_i$ for $i = 1, 2, 3$. Then identify index i^* such that f_{i^*} is the largest among $\{f_1, f_2, f_3\}$. Classify sample $\boldsymbol{x}$ to class i^*. To record your result, it is recommended to use a 3-component column vector whose i^*th component is set to value 1 while the other two components are set to value 0. For the kth test sample, call this vector $\boldsymbol{e}_k$. In addition, compare value i^* with the label of sample $\boldsymbol{x}$ being tested, if they are not equal, add 1 to counter **`mis_class`**.

Step 2: Repeat Step 1 to cover the entire test dataset. This should yield a total of 30 vectors $\{\boldsymbol{e}_k,\ k = 1, 2, \ldots, 30\}$ and **`mis_class`** tells you the total number of misclassifications made. If you combine all 30 $\{\boldsymbol{e}_k\}$ to form a matrix

$$\boldsymbol{E} = \begin{bmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \cdots & \boldsymbol{e}_{30} \end{bmatrix}_{3\times 30}$$

then $\boldsymbol{E}$ provides sufficient information to allow you to produce a 3×3 confusion matrix (see Example 1.11 of the course notes) that summarizes the performance of the classification technique applied.

Report the confusion matrix obtained as well as the error rate which is given by **`mis_class/30`** in this case.

Include your MATLAB code in the lab report.

Experiment 3
Predicting Energy Efficiency for Residential Buildings

1. Introduction and Objective

In this experiment, we build a multi-output linear model to predict energy efficiency of residential buildings in terms of heating and cooling loads. The database, from which the model in question learns, was created by A. Tsanas and A. Xifara in 2012 [R1] and has since been popular for performance evaluation of various prediction as well as classification techniques [R2].

In [R1], an energy analysis using 12 different building shapes was carried out, where the buildings differ with respect to a total of eight features including relative compactness, surface area, wall area, glazing area, and glazing area distribution, and so on. The data set contains 768 samples, each sample is characterized by a vector $\boldsymbol{x}$ with 8 components which are numerical values of the eight features mentioned above. Also associated with each sample is a 2-component output vector $\boldsymbol{y}$ representing heating and cooling loads of the building. Here we take the term "output vector" to mean a functional mapping from a set of eight features of a building as seen in a vector $\boldsymbol{x}$ to the building's heating and cooling loads. Clearly, we are dealing with a dataset of the form $\{(\boldsymbol{x}_n, \boldsymbol{y}_n), n = 1, 2, \ldots, N\}$ with $\boldsymbol{x}_n \in R^{8\times 1}$, $\boldsymbol{y}_n \in R^{2\times 1}$ and $N = 768$. The objective of the experiment is to develop a 2-output linear model that predicts heating and cooling loads for an "unseen" residential building characterized by a new feature vector $\boldsymbol{x}$.

References

[R1] A. Tsanas and A. Xifara, "Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools", *Energy and Buildings*, vol. 49, pp. 560-567, 2012.

[R2] UCI Machine Learning, http://archive.ics.uci.edu/ml, University of California Irvine, School of Information and Computer Science.

2. Background

2.1 Multi-output linear model for prediction

Multi-output linear model for prediction is studied in Sec. 1.5, see pages 22-24 of the course notes. Below we summarize the method where the quantities involved are explicitly specified in accordance with the above dataset. The linear model of interest is given by

$$\boldsymbol{y} = \boldsymbol{W}^T \boldsymbol{x} + \boldsymbol{b} \tag{E3.1}$$

where $\boldsymbol{y} \in R^{2\times 1}$, $\boldsymbol{W} \in R^{8\times 2}$ and $\boldsymbol{b} \in R^{2\times 1}$. With train data $\{(\boldsymbol{x}_n, \boldsymbol{y}_n), n = 1, 2, ..., 640\}$ (see Sec. 2.2 below) , we construct matrices

$$\hat{\boldsymbol{X}} = \begin{bmatrix} \boldsymbol{x}_1^T & 1 \\ \boldsymbol{x}_2^T & 1 \\ \vdots & \vdots \\ \boldsymbol{x}_{640}^T & 1 \end{bmatrix} \tag{E3.2}$$

and follow Eq. (1.30) from the course notes to compute the optimal parameters $\boldsymbol{W}^* = \begin{bmatrix} \boldsymbol{w}_1^* & \boldsymbol{w}_2^* \end{bmatrix}$ and $\boldsymbol{b}^* = \begin{bmatrix} b_1^* \\ b_2^* \end{bmatrix}$ as

$$\begin{bmatrix} \boldsymbol{w}_1^* & \boldsymbol{w}_2^* \\ b_1^* & b_2^* \end{bmatrix} = \left(\hat{\boldsymbol{X}}^T \hat{\boldsymbol{X}} + \varepsilon \boldsymbol{I}_9 \right)^{-1} \hat{\boldsymbol{X}}^T \begin{bmatrix} \boldsymbol{y}_1^T \\ \boldsymbol{y}_2^T \\ \vdots \\ \boldsymbol{y}_{768}^T \end{bmatrix} \tag{E3.3}$$

where we have include a term $\varepsilon \boldsymbol{I}_9$ to assure inverse of matrix $\hat{\boldsymbol{X}}^T \hat{\boldsymbol{X}} + \varepsilon \boldsymbol{I}_9$ does exist. Here scalar $\varepsilon > 0$ shall be small and we recommend that $\varepsilon = 0.01$ be used for this experiment.

2.2 The dataset

Listed below are eight features collected in the dataset created by Tsanas and Xifara:

- Relative compactness as component x_1
- Surface area as component x_2
- Wall area as component x_3
- Roof area as component x_4
- Overall height as component x_5
- Orientation as component x_6
- Glazing area as component x_7
- Glazing area distribution as component x_8

The measurements of the two outputs corresponding to a given set of these building features are:

- Heating load as component y_1
- Cooling load as component y_2

The dataset is available from the course website as matrix **`D_building`** of size 10×768 whose first 8 rows constitute data matrix **`X`** and the last 2 rows form the 2-component output matrix **`Y`** which contains measurements of heating and cooling loads for the corresponding 768 buildings. Once **`D_building`** is down-loaded, data matrix **`X`** and output matrix **`Y`** can be obtained as

```
X = D_building(1:8,:);
Y = D_building(9:10,:);
```

Among the 768 samples, we select 128 samples at random for test purposes and the remaining 640 samples will be used for training.

In addition, the samples in the test set are re-organized so that the output values (i.e. y_1 and y_2), when depicted in a figure as curves, roughly go upwards. This helps avoid unnecessary oscillations in these curves so as to make better visual inspection of prediction performance. The code below does all these things:

```
[d,N] = size(X);
rand('state',2)
r = randperm(N);
Xtr = X(:,r(129:N));
Ytr = Y(:,r(129:N));
Xte1 = X(:,r(1:128));
Yte1 = Y(:,r(1:128));
ym1 = mean(Yte1);
[~,ind1] = sort(ym1);
Xte = Xte1(:,ind1);
Yte = Yte1(:,ind1);
```

・In the above, `Xtr` contains 640 training samples and `Ytr` contains the corresponding outputs; `Xte` is the re-organized test data of size 8×128, and `Yte` contains the corresponding outputs.

3. Procedure

3.1 From the course website download data matrix `D_building.mat`

3.2 Use Eq. (E3.2) to prepare matrix $\hat{X}$, where $\{\boldsymbol{x}_n, n = 1, 2, ..., 640\}$ are from the train data `Xtr`, see Sec. 2.2 above.

3.3 Use Eq. (3.3) to prepare MATLAB code to compute optimal parameters $\boldsymbol{W}^*$ and $\boldsymbol{b}^*$ for the model in (E3.1).

3.4 Apply the optimized model

$$\boldsymbol{y} = \boldsymbol{W}^{*T}\boldsymbol{x} + \boldsymbol{b}^*$$

to the test data (i.e. `Xte`, see Sec. 2.2 above). This yields 128 predicted output vectors which we denote by $\{\boldsymbol{y}_n^{(p)}, n = 1, 2, ..., 128\}$. Evaluate the prediction performance as follows.

(i) Use $\{\boldsymbol{y}_n^{(p)}, n = 1, 2, ..., 128\}$ to form matrix

$$\boldsymbol{Y}^{(p)} = \begin{bmatrix} \boldsymbol{y}_1^{(p)} & \boldsymbol{y}_2^{(p)} & \cdots & \boldsymbol{y}_{128}^{(p)} \end{bmatrix}$$

and compute the overall relative prediction error as

$$e_p = \frac{\| \boldsymbol{Y}_{te} - \boldsymbol{Y}^{(p)} \|_F}{\| \boldsymbol{Y}_{te} \|_F}$$

where matrix `Yte` contains 128 ture output vectors (see Sec. 2.2 above), $\| \cdot \|_F$ denotes

Frobenius norm which is defined by $\| A \|_F = \left(\sum_i \sum_j a_{i.j}^2 \right)^{1/2}$.

(ii) For comparison, plot the first row of `Yte` and first row of $Y^{(p)}$ in the same figure, highlighted the two curves with different color. In another figure, plot the second row of `Yte` and second row of $Y^{(p)}$, highlighted these curses with different color. Comment on your visual inspection of the two figures.

Include your MATLAB code in the lab report.

Experiment 4

Breast Cancer Diagnosis via Logistic Regression

1. Objective

The goal of this experiment is to develop a computer program for automatic diagnosis of breast cancer based on logistic regression where the logistic loss function is defined by a dataset provided by Dr. Wolberg from General Surgery Department, University of Wisconsin, Madison, WI in 1990's. The dataset contains 30 carefully selected features from each of 569 patients [R1],[R2]. The same dataset has also been made available from the UCI Machine Learning Repository [R3]. The parameters in the logistic regression model are optimized by minimizing the logistic loss function mentioned above. In this experiment, the optimization is performed using the gradient descent (GD) algorithm, see Sec. 2.3 of the course notes.

References

[R1] W. N. Street, W. H. Wolberg, and O. L. Mangasarian, "Nuclear feature extraction for breast tumor diagnosis," in *IS?T/SPIE Int. Symp. Electronic Imaging: Science and Technology*, vol. 1905, pp. 861-870, San Jose, CA., 1993.

[R2] O. L. Mangasarian, W. N. Street, and W. H. Wolberg, "Breast cancer diagnosis and prognosis via linear programming," AAAI Tech. Report SS-94-01, 1994.

[R3] UCI Machine Learning, http://archive.ics.uci.edu/ml, University of California Irvine, School of Information and Computer Science.

2. Background and Data Pre-Processing

2.1 Background

As described in the literature [R1][R2], early detection of breast cancer is enhanced and unnecessary surgery avoided by diagnosing breast masses from Fine Needle Aspirates (FNA's). A graphical interface has been developed to compute nuclear features interactively. In order to obtain objective and precise measurements, a small region of each breast FNA is digitized, resulting in a 640×400, 8-bit-per-pixel gray scale image. An image analysis program uses a curve-fitting program to determine the boundaries of nuclei from initial dots placed near these boundaries by a mouse. A portion of one such processed image is shown in Figure E4.1. Ten features are computed for each nucleus that include area, radius, perimeter, symmetry, number and size of concavities, fractal dimension (of the boundary), compactness, smoothness (local variation of radial segments), and texture (variance of gray levels inside the boundary). The mean value, extreme value, and standard error of each of these cellular features are computed, resulting in a total of 30 real-valued features for each image. All feature values are recorded with four significant digits.

The dataset to be used in this experiment collects these 30 feature values from each of 569 patients of which 357 patients were diagnosed as benign while the other 212 patients were diagnosed as malignant. In what follows the dataset will be referred to as WDBC which stands for Wisconsin Diagnostic Breast Cancer.

Using this dataset, the objective of this lab experiment is to develop a classifier to classify a new and unused sample either to benign or to malignant.

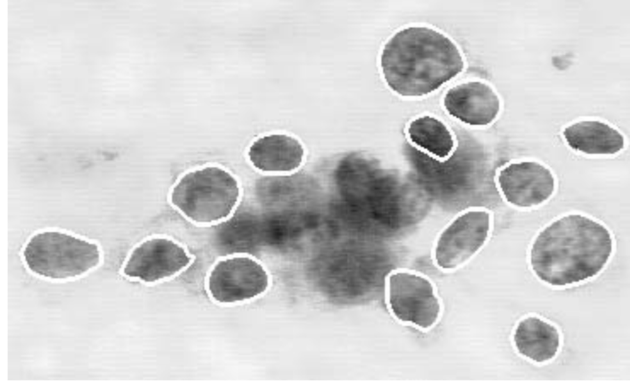


Figure E4.1 (from [R2]) A sample image where the cell nucleus boundaries are identified using an active contour model known as a "snake".

2.2 Pre-processing the data

Before a useful loss function can be defined, it is often necessary to pre-process the original dataset acquired from raw measurements. In real-world problems, numerical ranges of different features can be vastly different because different features represent different physical or artificial characteristics of the data. It turns out that a loss function defined by a dataset whose features vary in widely different numerical ranges is often more difficult to handle relative to that defined in a domain with a normalized scale for all features. In the case of WDBC, its 1st, 3rd, and 4th features represent radius (which is the mean value of distances from center to points on the perimeter), perimeter, and area of a cell nuclei, respectively. The value of a radius is typically in the range 20, while the perimeter and area are in the vicinity of 120 and 1,200, respectively. Under these circumstances, the data in WDBC need to be normalized first.

Below we describe a simple method to normalize a dataset $\mathcal{D} = \{(\boldsymbol{x}_n, y_n), n = 1, 2, ..., N\}$ where $\boldsymbol{x}_n \in R^{d\times 1}$. For WDBC, $d = 30$ and $N = 569$. The label y_n is set to $y_n = -1$ if $\boldsymbol{x}_n$ is associated with a patient diagnosed as benign, and y_n is set to $y_n = 1$ otherwise.

Step 1 Form data matrix $\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_N \end{bmatrix}$ of size d by N and denote its ith row by $\boldsymbol{z}_i$, namely,

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{z}_1 \\ \boldsymbol{z}_2 \\ \vdots \\ \boldsymbol{z}_d \end{bmatrix}$$

Note that vectors $\boldsymbol{z}_i$'s are the ith features of the N data samples from $\mathcal{D}$.

Step 2 Write the ith row vector $\boldsymbol{z}_i$ as $\boldsymbol{z}_i = \begin{bmatrix} z_1^{(i)} & z_2^{(i)} & \cdots & z_N^{(i)} \end{bmatrix}$ and compute its mean and variance as

$$m_i = \frac{1}{N}\sum_{k=1}^{N} z_k^{(i)}, \quad \sigma_i^2 = \frac{1}{N-1}\sum_{k=1}^{N}(z_k^{(i)} - m_i)^2$$

This step is carried out for each i for $i = 1, 2, \ldots, d$.

Step 3 Normalize z_i as

$$\hat{z}_i = \frac{1}{\sigma_i}\begin{bmatrix} z_1^{(i)} - m_i & z_2^{(i)} - m_i & \cdots & z_N^{(i)} - m_i \end{bmatrix}$$

so that the normalized $\hat{z}_i$ has zero mean and unity variance.

Step 4 Finally, the normalized dataset is constructed as

$$\hat{X} = \begin{bmatrix} \hat{z}_1 \\ \hat{z}_2 \\ \vdots \\ \hat{z}_d \end{bmatrix} \quad \text{(E4.1)}$$

Given data matrix **X** (refer to Sec. 3.1 below for the steps to get this matrix), its normalization can be performed in MATLAB as follows:

```
Xh = zeros(30,569);
for i = 1:30,
    xi = X(i,:);
    mi = mean(xi);
    vi = sqrt(var(xi));
    Xh(i,:) = (xi - mi)/vi;
end
```

3. Solving Logistic Regression Problem for Breast Cancer Diagnosis

3.1 Data for training and testing

WDBC is available from the course website named **D_wdbc.mat** which is a matrix of size 31×569 whose first 30 rows constitute data matrix **X** and the last row contains the labels of the 569 examples. Label "-1" is assigned to the samples associated with those diagnosed as benign, while label "1" is assigned to the samples associated with those diagnosed as malignant.

Once **D_wdbc** is loaded into your computer, data matrix **X** and its labels **y** can be obtained as

```
X = D_wdbc(1:30,:);
y = D_wdbc(31,:);
```

The data for training and testing can then be constructed in a few steps described below.

- Perform the procedure in Sec. 2.2 above to normalize data matrix **X**. Denote the normalized data matrix by **Xh**.
- Next, the examples having same label are put together as one subset. In this experiment, we have two subsets. The subset with label "-1" is called **Xn**, while the subset with label "1" is called **Xp**. These subsets can be generated as follows.

```
indn = find(y == -1);
indp = find(y == 1);
Xn = Xh(:,indn);
Xp = Xh(:,indp);
```

• The numbers of samples contained in **Xn** and **Xp** are 357 and 212, respectively. The code given below selects 300 examples at random from **Xn** for training and remaining 57 examples for testing while selects 180 examples at random from **Xp** for training and remaining 32 examples for testing.

```
rand('state',8)
r1 = randperm(357);
Xntr = Xn(:,r1(1:300));
Xnte = Xn(:,r1(301:357));
rand('state',7)
r2 = randperm(212);
Xptr = Xp(:,r2(1:180));
Xpte = Xp(:,r2(181:212));
Xtrain = [Xntr Xptr];
ytrain = [-ones(1,300), ones(1,180)];
Xtest = [Xnte Xpte];
ytest = [-ones(1,57), ones(1,32)];
```

• The code generates two useful datasets:
Xtrain is a matrix of size 30 by 480 containing features from 480 patients as training data;
ytrain contains 480 labels corresponding to the samples in **Xtrain**;
Xtest is a matrix of size 30 by 89 containing features from 89 patients as test data;
ytest contains 89 labels corresponding to the samples in **Xtest**.

3.2 Logistic Regression

Given a dataset $\mathcal{D}$, logistic regression studied in Sec. 3.1 of the course notes is applicable to two-category classification problems. In what follows, it is assumed that dataset has been normalized, and for notation simplicity it is still denoted by $\mathcal{D} = \{(\boldsymbol{x}_n, y_n) \text{ for } n = 1, 2, ..., N\}$ with $N = 569$. The classification is performed in two steps.

(i) Minimize the objective function

$$f(\hat{\boldsymbol{w}}) = \frac{\mu}{2} \| \hat{\boldsymbol{w}} \|_2^2 + \frac{1}{N} \sum_{n=1}^{N} \ln\left(1 + e^{-y_n \hat{\boldsymbol{w}}^T \hat{\boldsymbol{x}}_n}\right) \quad \text{(E4.2)}$$

with respect to parameter $\hat{\boldsymbol{w}} \in R^{31\times 1}$ where

$$\hat{\boldsymbol{w}} = \begin{bmatrix} b \\ \boldsymbol{w} \end{bmatrix} \text{ and } \hat{\boldsymbol{x}}_n = \begin{bmatrix} 1 \\ \boldsymbol{x}_n \end{bmatrix}$$

and $\boldsymbol{x}_n$ and y_n for $n = 1, 2, \ldots, N$ are provided by dataset $\mathcal{D}$.

To apply GD algorithm for minimizing $f(\hat{\boldsymbol{w}})$, the gradient $\nabla_{\hat{\boldsymbol{w}}} f(\hat{\boldsymbol{w}})$ is evaluated in closed-form as

$$\nabla_{\hat{\boldsymbol{w}}} f(\hat{\boldsymbol{w}}) = \mu \hat{\boldsymbol{w}} - \frac{1}{N} \sum_{n=1}^{N} \frac{y_n e^{-y_n \hat{\boldsymbol{w}}^T \hat{\boldsymbol{x}}_n}}{\left(1 + e^{-y_n \hat{\boldsymbol{w}}^T \hat{\boldsymbol{x}}_n}\right)} \hat{\boldsymbol{x}}_n \quad \text{(E4.3)}$$

Note that objective function $f(\hat{\boldsymbol{w}})$ is strictly convex, hence it admits a unique global minimizer and this minimizer is characterized by its gradient $\nabla_{\hat{\boldsymbol{w}}} f(\hat{\boldsymbol{w}})$ being zero. The convexity of $f(\hat{\boldsymbol{w}})$ also assures that the GD algorithm is *insensitive* to the choice of initial point $\hat{\boldsymbol{w}}_0$.

(ii) If we call the subset of training samples with label "–1" class $\mathcal{N}$ and the subset of training samples with label "1" class $\mathcal{P}$, then minimizer $\{\boldsymbol{w}^*, b^*\}$ obtained from step (i) can be used to classify a new data point $\boldsymbol{x}$ outside the training data to class $\mathcal{P}$ or class $\mathcal{N}$ in accordance with

$$\begin{cases} \boldsymbol{x} \in \mathcal{N} & \text{if } \boldsymbol{w}^{*T}\boldsymbol{x} + b^* < 0 \\ \boldsymbol{x} \in \mathcal{P} & \text{if } \boldsymbol{w}^{*T}\boldsymbol{x} + b^* > 0 \end{cases} \tag{E4.4}$$

4. Procedure

4.1 From the course website download data matrix **`D_wdbc.mat`**

4.2 Follow Sec. 3.1 above to generate training and testing datasets and their labels.

4.3 Prepare two MATLAB functions, one for evaluating the objective function in (E4.2) and another for evaluating its gradient in (E4.3).

4.4 Prepare MATLAB code to minimize $f(\hat{\boldsymbol{w}})$ in (E4.2) using the GD algorithm (see Sec. 2.3 of the course notes).

Remark

- To prepare the code, it is important to distinguish between $\boldsymbol{w}$ and $\hat{\boldsymbol{w}}$ as well as between $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}$.

4.5 Perform the minimization using the code prepared in Item 4.4 above. Set the regularization parameter $\mu = 0.02$ and use the initial point generated below to start the algorithm:

```
randn('state',9);
w0 = randn(31,1);
```

Remarks

- To run the GD algorithm, the MATLAB function for back-tracking line search, **`bt_lsearch.m`**, is required. Make sure to download it into the appropriate directory where you intend to run GD.
- You may terminate the algorithm either by using a small convergence tolerance ε or set a number of iterations. For the problem at hand, it is a good idea to just let the algorithm run K iterations and stop. In case your data and code are all well prepared, running the GD algorithm with $K = 150$ shall produce a good classifier.

4.6 Use the solution $\{\boldsymbol{w}^*, b^*\}$ obtained from Item 4.5 above to specify the classifier in (E4.4) and apply it to the test data from Item 4.2 above. Evaluate and report the performance of the classifier in terms of misclassification rate in percentage.

4.7 (Optional) The GD algorithm will need to run a much greater number of iterations to yield a classifier of comparable quality if the regularization term is switched off by setting $\mu = 0$. Try this to see what happens.

Include your MATLAB code in the lab report please.

APPENDIX A

An Introduction to MATLAB

Version 2.2

David F. Griffiths
Department of Mathematics
The University
Dundee DD1 4HN

With additional material by Ulf Carlsson
Department of Vehicle Engineering
KTH, Stockholm, Sweden

Contents

1 MATLAB **2**

2 Starting Up **2**
- 2.1 Windows Systems . . . 2
- 2.2 Unix Systems . . . 2
- 2.3 Command Line Help . . . 2
- 2.4 Demos . . . 3

3 Matlab as a Calculator **3**

4 Numbers & Formats **3**

5 Variables **3**
- 5.1 Variable Names . . . 3

6 Suppressing output **3**

7 Built–In Functions **4**
- 7.1 Trigonometric Functions . . . 4
- 7.2 Other Elementary Functions . . . 4

8 Vectors **4**
- 8.1 The Colon Notation . . . 5
- 8.2 Extracting Bits of a Vector . . . 5
- 8.3 Column Vectors . . . 5
- 8.4 Transposing . . . 5

9 Keeping a record **6**

10 Plotting Elementary Functions **6**
- 10.1 Plotting—Titles & Labels . . . 6
- 10.2 Grids . . . 7
- 10.3 Line Styles & Colours . . . 7
- 10.4 Multi–plots . . . 7
- 10.5 Hold . . . 7
- 10.6 Hard Copy . . . 7
- 10.7 Subplot . . . 8
- 10.8 Zooming . . . 8
- 10.9 Formatted text on Plots . . . 8
- 10.10 Controlling Axes . . . 9

11 Keyboard Accelerators **9**

12 Copying to and from Word and other applications **9**
- 12.1 Window Systems . . . 10
- 12.2 Unix Systems . . . 10

13 Script Files **10**

14 Products, Division & Powers of Vectors **10**
- 14.1 Scalar Product (`*`) . . . 10
- 14.2 Dot Product (`.*`) . . . 11
- 14.3 Dot Division of Arrays (`./`) . . . 12
- 14.4 Dot Power of Arrays (`.^`) . . . 12

15 Examples in Plotting **12**

16 Matrices—Two–Dimensional Arrays **13**
- 16.1 Size of a matrix . . . 13
- 16.2 Transpose of a matrix . . . 14
- 16.3 Special Matrices . . . 14
- 16.4 The Identity Matrix . . . 14
- 16.5 Diagonal Matrices . . . 14
- 16.6 Building Matrices . . . 15
- 16.7 Tabulating Functions . . . 15
- 16.8 Extracting Bits of Matrices . . . 15
- 16.9 Dot product of matrices (`.*`) . . . 16
- 16.10 Matrix–vector products . . . 16
- 16.11 Matrix–Matrix Products . . . 17
- 16.12 Sparse Matrices . . . 17

17 Systems of Linear Equations **17**
- 17.1 Overdetermined system of linear equations . . . 18

18 Characters, Strings and Text **19**

19 Loops **20**

20 Logicals **21**
- 20.1 While Loops . . . 22
- 20.2 `if...then...else...end` . . . 22

21 Function m–files **23**
- 21.1 Examples of functions . . . 24

22 Further Built–in Functions **25**
- 22.1 Rounding Numbers . . . 25
- 22.2 The `sum` Function . . . 25
- 22.3 `max` & `min` . . . 25
- 22.4 Random Numbers . . . 26
- 22.5 `find` for vectors . . . 26
- 22.6 `find` for matrices . . . 26

23 Plotting Surfaces **27**

24 Timing **28**

25 On–line Documentation **28**

26 Reading and Writing Data Files **29**
- 26.1 Formatted Files . . . 29
- 26.2 Unformatted Files . . . 30

27 Graphic User Interfaces **30**

28 Command Summary **31**

1 MATLAB

- Matlab is an interactive system for doing numerical computations.
- A numerical analyst called Cleve Moler wrote the first version of Matlab in the 1970s. It has since evolved into a successful commercial software package.
- Matlab relieves you of a lot of the mundane tasks associated with solving problems numerically. This allows you to spend more time thinking, and encourages you to experiment.
- Matlab makes use of highly respected algorithms and hence you can be confident about your results.
- Powerful operations can be performed using just one or two commands.
- You can build up your own set of functions for a particular application.
- Excellent graphics facilities are available, and the pictures can be inserted into LaTeX and Word documents.

These notes provide only a brief glimpse of the power and flexibility of the Matlab system. For a more comprehensive view we recommend the book

Matlab Guide
D.J. Higham & N.J. Higham
SIAM Philadelphia, 2000, ISBN: 0-89871-469-9.

2 Starting Up

2.1 Windows Systems

On Windows systems MATLAB is started by double-clicking the MATLAB icon on the desktop or by selecting MATLAB from the start menu.
The starting procedure takes the user to the Command window where the Command line is indicated with '`>>`'. Used in the calculator mode all Matlab commands are entered to the command line from the keyboard.
Matlab can be used in a number of different ways or modes; as an advanced calculator in the calculator mode, in a high level programming language mode and as a subroutine called from a C-program. More information on the first two of these modes is given below.
Help and information on Matlab commands can be found in several ways,

- from the command line by using the '`help topic`' command (see below),
- from the separate Help window found under the Help menu or
- from the Matlab helpdesk stored on disk or on a CD-ROM.

Another useful facility is to use the '`lookfor keyword`' command, which searches the help files for the keyword. See Exercise 16.1 (page 17) for an example of its use.

2.2 Unix Systems

- You should have a directory reserved for saving files associated with Matlab. Create such a directory (`mkdir`) if you do not have one. Change into this directory (`cd`).
- Start up a new `xterm` window (do `xterm &` in the existing `xterm` window).
- Launch Matlab in one of the `xterm` windows with the command

  ```
  matlab
  ```

  After a short pause, the logo will be shown followed by

  ```
  >>
  ```

  where `>>` is the Matlab prompt.

  Type `quit` at any time **to exit from Matlab.**

2.3 Command Line Help

Help is available from the command line prompt. Type `help help` for "help" (which gives a brief synopsis of the help system), `help` for a list of topics. The first few lines of this read

```
HELP topics:

matlab/general -  General purpose commands.
matlab/ops     -  Operators and special char...
matlab/lang    -  Programming language const...
matlab/elmat   -  Elementary matrices and ma...
matlab/elfun   -  Elementary math functions.
matlab/specfun -  Specialized math functions.
```

(truncated lines are shown with ...). Then to obtain help on "Elementary math functions", for instance, type

```
>> help elfun
```

This gives rather a lot of information so, in order to see the information one screenful at a time, first issue the command `more on`, i.e.,

```
>> more on
>> help elfun
```

Hit any key to progress to the next page of information.

2.4 Demos

Demonstrations are invaluable since they give an indication of Matlabs capabilities. A comprehensive set are available by typing the command

```
>> demo
```

(Warning: this will clear the values of all current variables.)

3 Matlab as a Calculator

The basic arithmetic operators are `+ - * / ^` and these are used in conjunction with brackets: `( )`. The symbol `^` is used to get exponents (powers): `2^4=16`.
You should type in commands shown following the prompt: `>>`.

```
>> 2 + 3/4*5
ans =
    5.7500
>>
```

Is this calculation `2 + 3/(4*5)` or `2 + (3/4)*5`? Matlab works according to the priorities:

1. quantities in brackets,
2. powers `2 + 3^2` ⇒`2 + 9 = 11`,
3. `* /`, working left to right (`3*4/5=12/5`),
4. `+ -`, working left to right (`3+4-5=7-5`),

Thus, the earlier calculation was for `2 + (3/4)*5` by priority 3.

4 Numbers & Formats

Matlab recognizes several different kinds of numbers

Type	Examples
Integer	$1362, -217897$
Real	$1.234, -10.76$
Complex	$3.21 - 4.3i \quad (i = \sqrt{-1})$
Inf	Infinity (result of dividing by 0)
NaN	**N**ot **a** **N**umber, $0/0$

The "`e`" notation is used for very large or very small numbers:
`-1.3412e+03` $= -1.3412 \times 10^3 = -1341.2$
`-1.3412e-01` $= -1.3412 \times 10^{-1} = -0.13412$
All computations in MATLAB are done in double precision, which means about 15 significant figures. The format—how Matlab prints numbers—is controlled by the "format" command. Type `help format` for full list. Should you wish to switch back to the default format then `format` will suffice.
The command

```
format compact
```

is also useful in that it suppresses blank lines in the output thus allowing more information to be displayed.

Command	Example of Output
`>>format short`	`31.4162`(4–decimal places)
`>>format short e`	`3.1416e+01`
`>>format long e`	`3.141592653589793e+01`
`>>format short`	`31.4162`(4–decimal places)
`>>format bank`	`31.42`(2–decimal places)

5 Variables

```
>> 3-2^4
ans =
   -13
>> ans*5
ans =
   -65
```

The result of the first calculation is labelled "`ans`" by Matlab and is used in the second calculation where its value is changed.
We can use our own names to store numbers:

```
>> x = 3-2^4
x =
   -13
>> y = x*5
y =
   -65
```

so that `x` has the value -13 and $y = -65$. These can be used in subsequent calculations. These are examples of **assignment statements**: values are assigned to variables. Each variable must be assigned a value before it may be used on the right of an assignment statement.

5.1 Variable Names

Legal names consist of any combination of letters and digits, starting with a letter. These are allowable:

```
NetCost, Left2Pay, x3, X3, z25c5
```

These are **not** allowable:

```
Net-Cost, 2pay, %x, @sign
```

Use names that reflect the values they represent.
Special names: you should avoid using
`eps = 2.2204e-16` $= 2^{-54}$ (The largest number such that `1 + eps` is indistinguishable from 1) and
`pi = 3.14159...` $= \pi$.
If you wish to do arithmetic with complex numbers,both `i` and `j` have the value $\sqrt{-1}$ unless you change them

```
>> i,j, i=3
ans = 0 + 1.0000i
ans = 0 + 1.0000i
i   = 3
```

6 Suppressing output

One often does not want to see the result of intermediate calculations—terminate the assignment statement or expression with semi–colon

```
>> x=-13; y = 5*x, z = x^2+y
y =
   -65
z =
   104
>>
```

the value of x is hidden. Note also we can place several statements on one line, separated by commas or semi–colons.

Exercise 6.1 *In each case find the value of the expression in Matlab and explain precisely the order in which the calculation was performed.*

i)	`-2^3+9`	ii)	`2/3*3`
iii)	`3*2/3`	iv)	`3*4-5^2*2-3`
v)	`(2/3^2*5)*(3-4^3)^2`	vi)	`3*(3*4-2*5^2-3)`

7 Built–In Functions

7.1 Trigonometric Functions

Those known to Matlab are
`sin, cos, tan`
and their arguments should be in radians.
e.g. to work out the coordinates of a point on a circle of radius 5 centred at the origin and having an elevation $30^o = \pi/6$ radians:

```
>> x = 5*cos(pi/6), y = 5*sin(pi/6)
x =
    4.3301
y =
    2.5000
```

The inverse trig functions are called `asin, acos, atan` (as opposed to the usual arcsin or $\sin^{-1}$ etc.). The result is in radians.

```
>> acos(x/5), asin(y/5)
ans = 0.5236
ans = 0.5236
>> pi/6
ans = 0.5236
```

7.2 Other Elementary Functions

These include `sqrt, exp, log, log10`

```
>> x = 9;
>> sqrt(x),exp(x),log(sqrt(x)),log10(x^2+6)
ans =
     3
ans =
   8.1031e+03
ans =
    1.0986
ans =
    1.9395
```

`exp(x)` denotes the exponential function $\exp(x) = e^x$ and the inverse function is `log`:

```
>> format long e, exp(log(9)), log(exp(9))
ans = 9.000000000000002e+00
ans = 9
>> format short
```

and we see a tiny rounding error in the first calculation. `log10` gives logs to the base 10. A more complete list of elementary functions is given in Table 2 on page 32.

8 Vectors

These come in two flavours and we shall first describe **row vectors**: they are lists of numbers separated by either commas or spaces. The number of entries is known as the "length" of the vector and the entries are often referred to as "elements" or "components" of the vector.The entries must be enclosed in square brackets.

```
>> v = [ 1 3, sqrt(5)]
v =
    1.0000    3.0000    2.2361
>> length(v)
ans =
     3
```

Spaces can be vitally important:

```
>> v2 = [3+ 4 5]
v2 =
     7     5
>> v3 = [3 +4 5]
v3 =
     3     4     5
```

We can do certain arithmetic operations with vectors of the same length, such as `v` and `v3` in the previous section.

```
>> v + v3
ans =
    4.0000    7.0000    7.2361
>> v4 = 3*v
v4 =
    3.0000    9.0000    6.7082
>> v5 = 2*v -3*v3
v5 =
   -7.0000   -6.0000  -10.5279
>> v + v2
??? Error using ==> +
Matrix dimensions must agree.
```

i.e. the error is due to `v` and `v2` having different lengths. A vector may be multiplied by a scalar (a number—see `v4` above), or added/subtracted to another vector of the **same** length. The operations are carried out elementwise.
We can build row vectors from existing ones:

```
>> w  = [1 2 3],  z = [8 9]
>> cd = [2*z,-w], sort(cd)
w =
     1     2     3
z =
     8     9
```

```
cd =
    16    18    -1    -2    -3
ans =
    -3    -2    -1    16    18
```

Notice the last command `sort`'ed the elements of `cd` into ascending order.
We can also change or look at the value of particular entries

```
>> w(2) = -2, w(3)
w =
     1    -2     3
ans =
     3
```

8.1 The Colon Notation

This is a shortcut for producing row vectors:

```
>>  1:4
ans =
     1     2     3     4
>> 3:7
ans =
     3     4     5     6     7
>> 1:-1
ans =
     []
```

More generally $a : b : c$ produces a vector of entries starting with the value a, incrementing by the value b until it gets to c (it will not produce a value beyond c). This is why `1:-1` produced the empty vector `[]`.

```
>> 0.32:0.1:0.6
ans =
    0.3200    0.4200    0.5200
>> -1.4:-0.3:-2
ans =
   -1.4000   -1.7000   -2.0000
```

8.2 Extracting Bits of a Vector

```
>>  r5 = [1:2:6, -1:-2:-7]
r5 =
     1     3     5    -1    -3    -5    -7
```

To get the 3rd to 6th entries:

```
>> r5(3:6)
ans =
     5    -1    -3    -5
```

To get alternate entries:

```
>> r5(1:2:7)
ans =
     1     5    -3    -7
```

What does `r5(6:-2:1)` give?
See `help colon` for a fuller description.

8.3 Column Vectors

These have similar constructs to row vectors. When defining them, entries are separated by `;` or "newlines"

```
>> c = [ 1; 3; sqrt(5)]
c =
    1.0000
    3.0000
    2.2361
>> c2 = [3
4
5]
c2 =
     3
     4
     5
>> c3 = 2*c - 3*c2
c3 =
   -7.0000
   -6.0000
  -10.5279
```

so column vectors may be added or subtracted **provided that they have the same length**.

8.4 Transposing

We can convert a row vector into a column vector (and vice versa) by a process called *transposing*—denoted by `'`.

```
>> w, w', c, c'
w =
     1    -2     3
ans =
     1
    -2
     3
c =
    1.0000
    3.0000
    2.2361
ans =
    1.0000    3.0000    2.2361
>> t = w + 2*c'
t =
    3.0000    4.0000    7.4721
>> T = 5*w'-2*c
T =
    3.0000
  -16.0000
   10.5279
```

If `x` is a *complex vector*, then `x'` gives the *complex conjugate transpose* of `x`:

```
>> x = [1+3i, 2-2i]
ans =
   1.0000 + 3.0000i   2.0000 - 2.0000i
>> x'
ans =
   1.0000 - 3.0000i
   2.0000 + 2.0000i
```

Note that the components of x were defined without a * operator; this means of defining complex numbers works even when the variable i already has a numeric value. To obtain the plain transpose of a complex number use .' as in

```
>> x.'
ans =
   1.0000 + 3.0000i
   2.0000 - 2.0000i
```

9 Keeping a record

Issuing the command

```
>> diary mysession
```

will cause all subsequent text that appears on the screen to be saved to the file mysession located in the directory in which Matlab was invoked. You may use any legal filename *except* the names on and off. The record may be terminated by

```
>> diary off
```

The file mysession may be edited with emacs to remove any mistakes.
If you wish to quit Matlab midway through a calculation so as to continue at a later stage:

```
>> save thissession
```

will save the current values of all variables to a file called thissession.mat. **This file cannot be edited.** When you next startup Matlab, type

```
>> load thissession
```

and the computation can be resumed where you left off. A list of variables used in the current session may be seen with

```
>> whos
```

See help whos and help save.

```
>> whos
Name  Size Elements  Bytes  Density Complex
ans   1 by 1  1        8     Full      No
  v   1 by 3  3       24     Full      No
 v1   1 by 2  2       16     Full      No
 v2   1 by 2  2       16     Full      No
 v3   1 by 3  3       24     Full      No
 v4   1 by 3  3       24     Full      No
  x   1 by 1  1        8     Full      No
  y   1 by 1  1        8     Full      No

Grand total is 16 elements using 128 bytes
```

10 Plotting Elementary Functions

Suppose we wish to plot a graph of $y = \sin 3\pi x$ for $0 \le x \le 1$. We do this by sampling the function at a sufficiently large number of points and then joining up the points (x, y) by straight lines. Suppose we take $N + 1$ points equally spaced a distance h apart:

```
>> N = 10; h = 1/N; x = 0:h:1;
```

defines the set of points $x = 0, h, 2h, \ldots, 1-h, 1$. The corresponding y values are computed by

```
>> y = sin(3*pi*x);
```

and finally, we can plot the points with

```
>> plot(x,y)
```

The result is shown in Figure 1, where it is clear that the value of N is too small.

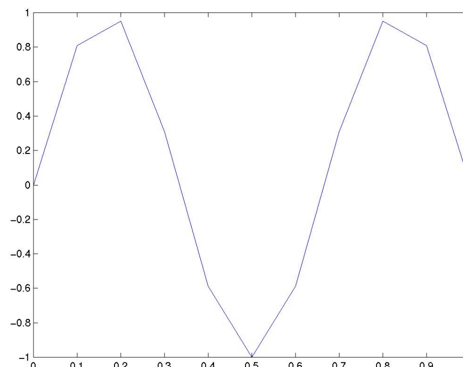


Figure 1: Graph of $y = \sin 3\pi x$ for $0 \le x \le 1$ using $h = 0.1$.

On changing the value of N to 100:

```
>> N = 100; h = 1/N; x = 0:h:1;
>> y = sin(3*pi*x);  plot(x,y)
```

we get the picture shown in Figure 2.

10.1 Plotting—Titles & Labels

To put a title and label the axes, we use

```
>> title('Graph of y = sin(3pi x)')
>> xlabel('x axis')
>> ylabel('y-axis')
```

The *strings* enclosed in single quotes, can be anything of our choosing. Some simple LaTeX commands are available for formatting mathematical expressions and Greek characters—see Section 10.9.

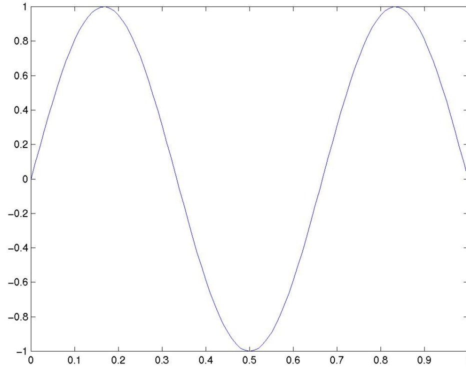


Figure 2: Graph of $y = \sin 3\pi x$ for $0 \le x \le 1$ using $h = 0.01$.

10.2 Grids

A dotted grid may be added by

```
>> grid
```

This can be removed using either `grid` again, or `grid off`.

10.3 Line Styles & Colours

The default is to plot solid lines. A solid white line is produced by

```
>> plot(x,y,'w-')
```

The third argument is a string whose first character specifies the colour(optional) and the second the line style. The options for colours and styles are:

Colours		Line Styles	
y	yellow	.	point
m	magenta	o	circle
c	cyan	x	x-mark
r	red	+	plus
g	green	-	solid
b	blue	*	star
w	white	:	dotted
k	black	-.	dashdot
		--	dashed

The number of available plot symbols is wider than shown in this table. Use `help plot` to obtain a full list. See also `help shapes`.

10.4 Multi–plots

Several graphs may be drawn on the same figure as in

```
>> plot(x,y,'w-',x,cos(3*pi*x),'g--')
```

A descriptive legend may be included with

```
>> legend('Sin curve','Cos curve')
```

which will give a list of line–styles, as they appeared in the plot command, followed by a brief description. Matlab fits the legend in a suitable position, so as not to conceal the graphs whenever possible.
For further information do `help plot` etc.
The result of the commands

```
>> plot(x,y,'w-',x,cos(3*pi*x),'g--')
>> legend('Sin curve','Cos curve')
>> title('Multi-plot ')
>> xlabel('x axis'),  ylabel('y axis')
>> grid
```

is shown in Figure 3. The legend may be moved manually by dragging it with the mouse.

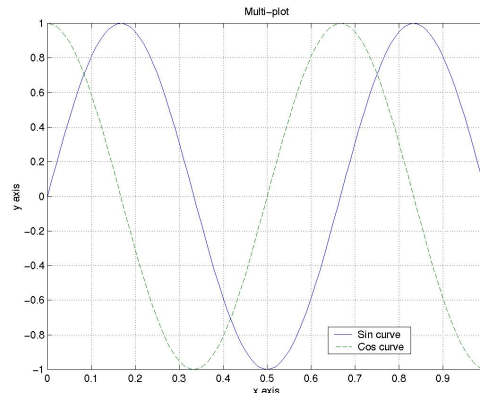


Figure 3: Graph of $y = \sin 3\pi x$ and $y = \cos 3\pi x$ for $0 \le x \le 1$ using $h = 0.01$.

10.5 Hold

A call to `plot` clears the graphics window before plotting the current graph. This is not convenient if we wish to add further graphics to the figure at some later stage. To stop the window being cleared:

```
>> plot(x,y,'w-'), hold
>> plot(x,y,'gx'), hold off
```

"`hold on`" holds the current picture; "`hold off`" releases it (but does not clear the window, which can be done with `clf`). "`hold`" on its own toggles the hold state.

10.6 Hard Copy

To obtain a printed copy select Print from the File menu on the Figure toolbar.
Alternatively one can save a figure to a file for later printing (or editing). A number of formats is available (use `help print` to obtain a list). To save a file in "Encapsulated PostScript" format, issue the Matlab command

```
print -deps fig1
```

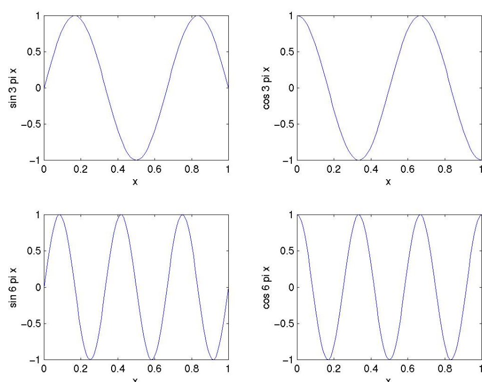
which will save a copy of the image in a file called `fig1.eps`.

10.7 Subplot

The graphics window may be split into an $m \times n$ array of smaller windows into which we may plot one or more graphs. The windows are counted 1 to mn row–wise, starting from the top left. Both `hold` and `grid` work on the current subplot.

```
>> subplot(221), plot(x,y)
>>   xlabel('x'),ylabel('sin 3 pi x')
>> subplot(222), plot(x,cos(3*pi*x))
>>   xlabel('x'),ylabel('cos 3 pi x')
>> subplot(223), plot(x,sin(6*pi*x))
>>   xlabel('x'),ylabel('sin 6 pi x')
>> subplot(224), plot(x,cos(6*pi*x))
>>   xlabel('x'),ylabel('cos 6 pi x')
```

`subplot(221)` (or `subplot(2,2,1)`) specifies that the window should be split into a 2×2 array and we select the first subwindow.



10.8 Zooming

We often need to "zoom in" on some portion of a plot in order to see more detail. This is easily achieved using the command

```
>> zoom
```

Pointing the mouse to the relevant position on the plot and clicking the left mouse button will zoom in by a factor of two. This may be repeated to any desired level.
Clicking the right mouse button will zoom out by a factor of two.
Holding down the left mouse button and dragging the mouse will cause a rectangle to be outlined. Releasing the button causes the contents of the rectangle to fill the window.
`zoom off` turns off the zoom capability.

Exercise 10.1 *Draw graphs of the functions*

$$\begin{aligned} y &= \cos x \\ y &= x \end{aligned}$$

for $0 \le x \le 2$ *on the same window. Use the zoom facility to determine the point of intersection of the two curves (and, hence, the root of* $x = \cos x$*) to two significant figures.*

The command `clf` clears the current figure while `close 1` will close the window labelled "Figure 1". To open a new figure window type `figure` or, to get a window labelled "Figure 9", for instance, type `figure (9)`. If "Figure 9" already exists, this command will bring this window to the foreground and the result subsequent plotting commands will be drawn on it.

10.9 Formatted text on Plots

It is possible to change to format of text on plots so as to increase or decrease its size and also to typeset simple mathematical expressions (in LaTeX form).
We shall give two illustrations.
First we plot the first 100 terms in the sequence $\{x_n\}$ given by $x_n = \left(1 + \frac{1}{n}\right)^n$ and then graph the function $\phi(x) = x^3 \sin^2(3\pi x)$ on the interval $-1 \le x \le 1$. The commands

```
>> set(0,'Defaultaxesfontsize',16);
>> n = 1:100;  x = (1+1./n).^n;
>> subplot (211)
>> plot(n,x,'.',[0 max(n)],exp(1)*[1 1],...
     '--','markersize',8)
>> title('x_n = (1+1/n)^n','fontsize',12)
>> xlabel('n'), ylabel('x_n')
>> legend('x_n','y = e^1 = 2.71828...',4)
>> %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
>> subplot (212)
>> x = -2:.02:2; y = x.^3.*sin(3*pi*x).^2;
>> plot(x,y,'linewidth',2)
>> legend('y = x^3sin^2 3\pi x',4)
>> xlabel('x')
```

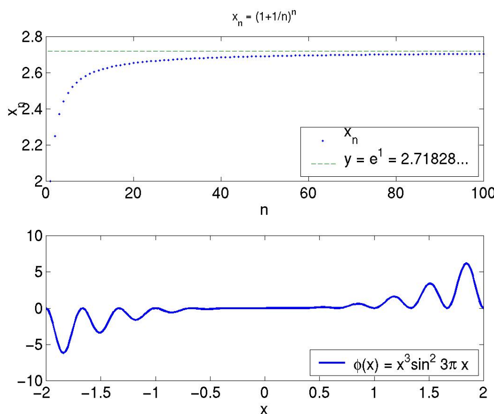
produce the graph shown below. The salient features of these commands are

1. The first line increases the size of the default font size used for the axis labels, legends and titles.
2. The size of the plot symbol "." is changed from the default (6) to size 8 by the additional string followed by value "`'markersize',8`".
3. The strings '`x_n`' are formatted as x_n to give subscripts while '`x^3` leads to superscripts x^3.
   Note also that $sin^2 3\pi x$ translates into the Matlab command `sin(3*pi*x).^2`—the position of the exponent is different.
4. Greek characters $\alpha, \beta, \ldots, \omega, \Omega$ are produced by the strings '`\alpha`', '`\beta`', ...,'`\omega`', '`\Omega`'. the integral symbol: $\int$ is produced by '`\int`'.
5. The thickness of the line used in the lower graph is changed from its default value (0.5) to 2.
6. Use `help legend` to determine the meaning of the last argument in the `legend` commands.

One can determine the current value of any plot property by first obtaining its "handle number" and then using the `get` command such as

```
>> handle = plot (x,y,'.')
>> get (handle,'markersize')
ans =
     6
```

Experiment also with `set (handle)` (which will list possible values for each property) and `set (handle,'mark` Also, all plot properties can be edited from the Figure window by selecting the Tools menu from the toolbar. For instance, to change the `linewidth` of a graph, first select the curve by double clicking (it should then change its appearance) and then select Line Properties... from the Tools. This will pop up a dialogue window from which the width, colour, style,... of the curve may be changed.



10.10 Controlling Axes

Once a plot has been created in the graphics window you may wish to change the range of x and y values shown on the picture.

```
>> clf, N = 100; h = 1/N; x = 0:h:1;
>> y = sin(3*pi*x); plot(x,y)
>> axis([-0.5 1.5 -1.2 1.2]),  grid
```

The `axis` command has four parameters, the first two are the minimum and maximum values of x to use on the axis and the last two are the minimum and maximum values of y. Note the square brackets. The result of these commands is shown in Figure 4. Look at `help axis` and experiment with the commands `axis equal`, `axis verb`, `axis square`, `axis normal`, `axis tight` in any order.

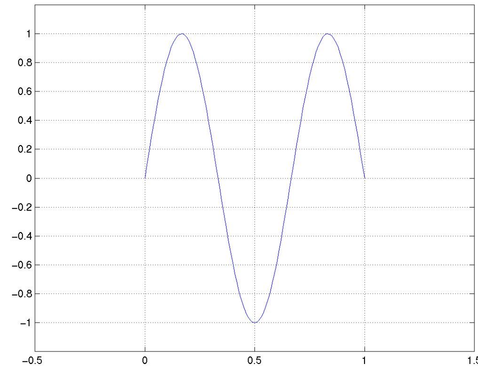


Figure 4: The effect of changing the axes of a plot.

11 Keyboard Accelerators

One can recall previous Matlab commands by using the ↑ and ↓ cursor keys. Repeatedly pressing ↑ will review the previous commands (most recent first) and, if you want to re-execute the command, simply press the return key.
To recall the most recent command starting with `p`, say, type `p` at the prompt followed by ↑. Similarly, typing `pr` followed by ↑ will recall the most recent command starting with `pr`.
Once a command has been recalled, it may be edited (changed). You can use ← and → to move backwards and forwards through the line, characters may be inserted by typing at the current cursor position or deleted using the `Del` key. This is most commonly used when long command lines have been mistyped or when you want to re–execute a command that is very similar to one used previously.
The following emacs–like commands may also be used:

`cntrl a`	move to start of line
`cntrl e`	move to end of line
`cntrl f`	move forwards one character
`cntrl b`	move backwards one character
`cntrl d`	delete character under the cursor

Once you have the command in the required form, press return.

Exercise 11.1 *Type in the commands*

```
>> x = -1:0.1:1;
>> plot(x,sin(pi*x),'w-')
>> hold on
>> plot(x,cos(pi*x),'r-')
```

Now use the cursor keys with suitable editing to execute:

```
>> x = -1:0.05:1;
>> plot(x,sin(2*pi*x),'w-')
>> plot(x,cos(2*pi*x),'r-.'), hold off
```

12 Copying to and from Word and other applications

There are many situations where one wants to copy the output resulting from a Matlab command (or commands) into a Windows application such as Word or into a Unix file editor such as “emacs” or “vi”.

12.1 Window Systems

Copying material is made possible on the Windows operating system by using the Windows clipboard.
Also, pictures can be exported to files in a number of alternative formats such as encapsulated postscript format or in jpeg format. Matlab is so frequently used as an analysis tool that many manufacturers of measurement systems and software find it convenient to provide interfaces to Matlab which make it possible, for instance, to import measured data directly into a `*.mat` Matlab file (see `load` and `save` in Section 9).

Example 12.1 *Copying a figure into Word.*

Diagrams prepared in Matlab are easily exported to other Windows applications such as Word. Suppose a plot of the functions $\sin(2\pi ft)$ and $\sin(2\pi ft + \pi/4)$, with $f = 100$, is needed in a report written in Word.
We create a time vector, t, with 500 points distributed over 5 periods and then evaluate and plot the two function vectors.

```
>> t = [1:1:500]/500/20;
>> f = 100;
>> y1 = sin(2*pi*f*t);
>> y2 = sin(2*pi*f*t+pi/4);
>> plot(t,y1,'-',t,y2,'--');
>> axis([0 0.05 -1.5 1.5]);
>> grid
```

In order to copy the plot into a Word document

- Select "Copy Figure" under the Edit menu on the figure windows toolbar.
- Switch to the Word application if it is already running, otherwise open a Word document.
- Place the cursor in the desired position in the document and select "Paste" under the "Edit" menu in the Word tool bar.

12.2 Unix Systems

In order to carry out the following exercise, you should have Matlab running in one window and either Emacs or Vi running in another.
To copy material from one window to another, (here L means click Left Mouse Button, etc)
First select the material to copy by L on the start of the material you want and then either dragging the mouse (with the buttom down) to highlight the text, or R at the end of the material. Next move the mouse into the other window and L at the location you want the text to appear. Finally, click the M.
When copying from another application into Matlab you can only copy material to the prompt line. On Unix systems figures are normally saved in files (see Section 10.6) which are then imported into other documents.

13 Script Files

Script files are normal ASCII (text) files that contain Matlab commands. It is essential that such files have names having an extension `.m` (e.g., `Exercise4.m`) and, for this reason, they are commonly known as *m-files*. The commands in this file may then be executed using
`>> Exercise4`
Note: the command does not include the file name extension `.m`.
It is only the output from the commands (and not the commands themselves) that are displayed on the screen.
Script files are created with your favourite editor under Unix while, under Windows, click on the "New Document" icon at the top left of the main Matlab window to pop up a new window showing the "M-file Editor". Type in your commands and then save (to a file with a `.m` extension).
To see the commands in the command window prior to their execution:
`>> echo on`
and `echo off` will turn echoing off.
Any text that follows `%` on a line is ignored. The main purpose of this facility is to enable comments to be included in the file to describe its purpose.
To see what m-files you have in your current directory, use
`>> what`

Exercise 13.1
1. *Type in the commands from §10.7 into a file called `exsub.m`.*
2. *Use `what` to check that the file is in the correct area.*
3. *Use the command `type exsub` to see the contents of the file.*
4. *Execute these commands.*

See §21 for the related topic of function files.

14 Products, Division & Powers of Vectors

14.1 Scalar Product (*)

We shall describe two ways in which a meaning may be attributed to the product of two vectors. In both cases the vectors concerned must have the same length.
The first product is the standard scalar product. Suppose that $\underline{u}$ and $\underline{v}$ are two vectors of length n, $\underline{u}$ being a **row** vector and $\underline{v}$ a **column** vector:

$$\underline{u} = [u_1, u_2, \ldots, u_n], \qquad \underline{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}.$$

The scalar product is defined by multiplying the corresponding elements together and adding the results to give a single number (scalar).

$$\underline{u}\,\underline{v} = \sum_{i=1}^{n} u_i v_i.$$

For example, if $\underline{u} = [10, -11, 12]$, and $\underline{v} = \begin{bmatrix} 20 \\ -21 \\ -22 \end{bmatrix}$ then $n = 3$ and

$$\underline{u}\,\underline{v} = 10 \times 20 + (-11) \times (-21) + 12 \times (-22) = 167.$$

We can perform this product in Matlab by

```
>> u = [ 10, -11, 12], v = [20; -21; -22]
>> prod = u*v     % row times column vector
```

Suppose we also define a row vector `w` and a column vector `z` by

```
>> w = [2, 1, 3], z = [7; 6; 5]
w =
     2     1     3
z =
     7
     6
     5
```

and we wish to form the scalar products of `u` with `w` and `v` with `z`.

```
>> u*w
??? Error using ==> *
Inner matrix dimensions must agree.
```

an error results because `w` is not a column vector. Recall from page 5 that transposing (with `'`) turns column vectors into row vectors and vice versa.

So, to form the scalar product of two row vectors or two column vectors,

```
>> u*w'      % u & w are row vectors
ans =
    45
>> u*u'      % u is a row vector
ans =
   365
>> v'*z      % v & z are column vectors
ans =
   -96
```

We shall refer to the Euclidean length of a vector as the **norm** of a vector; it is denoted by the symbol $\|\underline{u}\|$ and defined by

$$\|\underline{u}\| = \sqrt{\sum_{i=1}^{n} |u_i|^2},$$

where n is its dimension. This can be computed in Matlab in one of two ways:

```
>> [ sqrt(u*u'), norm(u)]
ans =
   19.1050    19.1050
```

where `norm` is a built–in Matlab function that accepts a vector as input and delivers a scalar as output. It can also be used to compute other norms: `help norm`.

Exercise 14.1 *The angle, θ, between two column vectors $\underline{x}$ and $\underline{y}$ is defined by*

$$\cos\theta = \frac{\underline{x}'\underline{y}}{\|\underline{x}\|\,\|\underline{y}\|}.$$

Use this formula to determine the cosine of the angle between

$$\underline{x} = [1, 2, 3]' \quad \text{and} \quad \underline{y} = [3, 2, 1]'.$$

Hence find the angle in degrees.

14.2 Dot Product (.*)

The second way of forming the product of two vectors of the same length is known as the Hadamard product. It is not often used in Mathematics but is an invaluable Matlab feature. It involves vectors of the same type. If $\underline{u}$ and $\underline{v}$ are two vectors of the same type (both row vectors or both column vectors), the mathematical definition of this product, which we shall call the **dot product**, is the **vector** having the components

$$\underline{u} \cdot \underline{v} = [u_1 v_1, u_2 v_2, \ldots, u_n v_n].$$

The result is a vector of the same length and type as $\underline{u}$ and $\underline{v}$. Thus, we simply multiply the corresponding elements of two vectors.
In Matlab, the product is computed with the operator `.*` and, using the vectors `u, v, w, z` defined on page 11,

```
>> u.*w
ans =
      20 -11 36
>> u.*v'
ans =
      200 231 -264
>> v.*z, u'.*v
ans =
      140 -126 -110
ans =
      200 231 -264
```

Example 14.1 *Tabulate the function $y = x \sin \pi x$ for $x = 0, 0.25, \ldots, 1$.*

It is easier to deal with column vectors so we first define a vector of x-values: (see Transposing: §8.4)
`>> x = (0:0.25:1)';`
To evaluate y we have to multiply each element of the vector x by the corresponding element of the vector $\sin \pi x$:

$x\ \times$	$\sin \pi x =$	$x \sin \pi x$
$0\ \times$	$0 =$	0
$0.2500\ \times$	$0.7071 =$	0.1768
$0.5000\ \times$	$1.0000 =$	0.5000
$0.7500\ \times$	$0.7071 =$	0.5303
$1.0000\ \times$	$0.0000 =$	0.0000

To carry this out in Matlab:

```
>>  y = x.*sin(pi*x)
y =
         0
    0.1768
    0.5000
    0.5303
    0.0000
```

Note: a) the use of `pi`, b) `x` and `sin(pi*x)` are both column vectors (the `sin` function is applied to each element of the vector). Thus, the dot product of these is also a column vector.

14.3 Dot Division of Arrays (./)

There is no mathematical definition for the division of one vector by another. However, in Matlab, the operator `./` is defined to give element by element division—it is therefore only defined for vectors of the same size and type.

```
>> a = 1:5, b = 6:10, a./b
a =
     1     2     3     4     5
b =
     6     7     8     9    10
ans =
 0.1667  0.2857  0.3750  0.4444  0.5000
>> a./a
ans =
     1     1     1     1     1
>> c = -2:2, a./c
c =
    -2    -1     0     1     2
Warning: Divide by zero
ans =
 -0.5000  -2.0000  Inf   4.0000   2.5000
```

The previous calculation required division by 0—notice the `Inf`, denoting infinity, in the answer.

```
>> a.*b -24, ans./c
ans =
   -18   -10     0    12    26

Warning: Divide by zero
ans =
     9    10   NaN    12    13
```

Here we are warned about 0/0—giving a `NaN` (**N**ot **a** **N**umber).

Example 14.2 *Estimate the limit*

$$\lim_{x\to 0}\frac{\sin \pi x}{x}.$$

The idea is to observe the behaviour of the ratio $\frac{\sin \pi x}{x}$ for a sequence of values of x that approach zero. Suppose that we choose the sequence defined by the column vector

```
>> x = [0.1; 0.01; 0.001; 0.0001]
```

then

```
>> sin(pi*x)./x
ans =
    3.0902
    3.1411
    3.1416
    3.1416
```

which suggests that the values approach π. To get a better impression, we subtract the value of π from each entry in the output and, to display more decimal places, we change the format

```
>> format long
>> ans -pi
ans =
  -0.05142270984032
  -0.00051674577696
  -0.00000516771023
  -0.00000005167713
```

Can you explain the pattern revealed in these numbers? We also need to use `./` to compute a scalar divided by a vector:

```
>> 1/x
??? Error using ==> /
Matrix dimensions must agree.
>> 1./x
ans =
   10    100    1000    10000
```

so `1./x` works, but `1/x` does not.

14.4 Dot Power of Arrays (.^)

To square each of the elements of a vector we could, for example, do `u.*u`. However, a neater way is to use the `.^` operator:

```
>> u.^2
ans =
   100   121   144
>> u.*u
ans =
   100   121   144
>> u.^4
ans =
       10000        14641        20736
>> v.^2
ans =
   400
   441
   484
>> u.*w.^(-2)
ans =
    2.5000   -11.0000    1.3333
```

Recall that powers (`.^` in this case) are done first, before any other arithmetic operation.

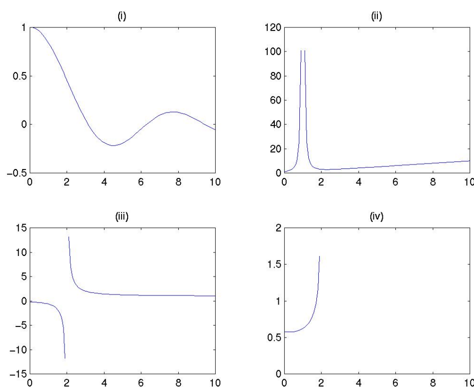
15 Examples in Plotting

Example 15.1 *Draw graphs of the functions*

i) $y = \frac{\sin x}{x}$ *ii)* $u = \frac{1}{(x-1)^2} + x$
iii) $v = \frac{x^2+1}{x^2-4}$ *iv)* $w = \frac{(10-x)^{1/3}-2}{(4-x^2)^{1/2}}$

for $0 \leq x \leq 10$.

```
>> x = 0:0.1:10;
>> y = sin(x)./x;
>>     subplot(221), plot(x,y), title('(i)')
Warning: Divide by zero
>> u = 1./(x-1).^2 + x;
>>     subplot(222),plot(x,u), title('(ii)')
Warning: Divide by zero
>> v = (x.^2+1)./(x.^2-4);
>>     subplot(223),plot(x,v),title('(iii)')
Warning: Divide by zero
>> w = ((10-x).^(1/3)-1)./sqrt(4-x.^2);
Warning: Divide by zero
>>     subplot(224),plot(x,w),title('(iv)')
```



Note the repeated use of the "dot" operators.
Experiment by changing the axes (page 9), grids (page 7) and hold(page 7).

```
>>  subplot(222),axis([0 10 0 10])
>>  grid
>>  grid
>>  hold on
>>  plot(x,v,'--'), hold off, plot(x,y,':')
```

Exercise 15.1 *Enter the vectors*

$$\underline{U} = [6, 2, 4], \quad \underline{V} = [3, -2, 3, 0],$$
$$\underline{W} = \begin{bmatrix} 3 \\ -4 \\ 2 \\ -6 \end{bmatrix}, \quad \underline{Z} = \begin{bmatrix} 3 \\ 2 \\ 2 \\ 7 \end{bmatrix}$$

into Matlab.

1. *Which of the products*

   `U*V,  V*W,  U*V', V*W', W*Z', U.*V`

   `U'*V, V'*W, W'*Z, U.*W, W.*Z, V.*W`

   is legal? State whether the legal products are row or column vectors and give the values of the legal results.

2. *Tabulate the functions*

$$y = (x^2 + 3)\sin \pi x^2$$

and

$$z = \sin^2 \pi x/(x^{-2} + 3)$$

for $x = 0, 0.2, \ldots, 10$. *Hence, tabulate the function*

$$w = \frac{(x^2 + 3)\sin \pi x^2 \sin^2 \pi x}{(x^{-2} + 3)}.$$

Plot a graph of w *over the range* $0 \leq x \leq 10$.

16 Matrices—Two–Dimensional Arrays

Row and Column vectors are special cases of **matrices**. An $m \times n$ matrix is a rectangular array of numbers having m rows and n columns. It is usual in a mathematical setting to include the matrix in either round or square brackets—we shall use square ones. For example, when $m = 2, n = 3$ we have a 2×3 matrix such as

$$A = \begin{bmatrix} 5 & 7 & 9 \\ 1 & -3 & -7 \end{bmatrix}$$

To enter such an matrix into Matlab we type it in row by row using the same syntax as for vectors:

```
>> A = [5 7 9
        1 -3 -7]
A =
     5     7     9
     1    -3    -7
```

Rows may be separated by semi-colons rather than a new line:

```
>> B = [-1 2 5; 9 0 5]
B =
    -1     2     5
     9     0     5
>> C = [0, 1; 3, -2; 4, 2]
C =
     0     1
     3    -2
     4     2
>> D = [1:5; 6:10; 11:2:20]
D =
     1     2     3     4     5
     6     7     8     9    10
    11    13    15    17    19
```

So `A` and `B` are 2×3 matrices, `C` is 3×2 and `D` is 3×5. In this context, a row vector is a $1 \times n$ matrix and a column vector a $m \times 1$ matrix.

16.1 Size of a matrix

We can get the size (dimensions) of a matrix with the command `size`

```
>> size(A), size(x)
ans =
     2     3
ans =
     3     1
>> size(ans)
ans =
     1     2
```

So A is 2×3 and x is 3×1 (a column vector). The last command `size(ans)` shows that the *value* returned by `size` is itself a 1×2 matrix (a row vector). We can save the results for use in subsequent calculations.

```
>> [r c] = size(A'), S = size(A')
r =
     3
c =
     2
S =
     3     2
```

16.2 Transpose of a matrix

Transposing a vector changes it from a row to a column vector and vice versa (see §8.4). The extension of this idea to matrices is that transposing interchanges rows with the corresponding columns: the 1st row becomes the 1st column, and so on.

```
>> D, D'
D =
     1     2     3     4     5
     6     7     8     9    10
    11    13    15    17    19
ans =
     1     6    11
     2     7    13
     3     8    15
     4     9    17
     5    10    19
>> size(D), size(D')
ans =
     3     5
ans =
     5     3
```

16.3 Special Matrices

Matlab provides a number of useful built–in matrices of any desired size.
`ones(m,n)` gives an $m \times n$ matrix of 1's,

```
>> P = ones(2,3)
P =
     1     1     1
     1     1     1
```

`zeros(m,n)` gives an $m \times n$ matrix of 0's,

```
>> Z = zeros(2,3), zeros(size(P'))
Z =
     0     0     0
     0     0     0
ans =
     0     0
     0     0
     0     0
```

The second command illustrates how we can construct a matrix based on the size of an existing one. Try `ones(size(D))`.
An $n \times n$ matrix that has the same number of rows and columns and is called a **square** matrix.
A matrix is said to be **symmetric** if it is equal to its transpose (i.e. it is unchanged by transposition):

```
>> S = [2 -1 0; -1 2 -1; 0 -1 2],
S =
     2    -1     0
    -1     2    -1
     0    -1     2
>> St = S'
St =
     2    -1     0
    -1     2    -1
     0    -1     2
>> S-St
ans =
     0     0     0
     0     0     0
     0     0     0
```

16.4 The Identity Matrix

The $n \times n$ **identity** matrix is a matrix of zeros except for having ones along its leading diagonal (top left to bottom right). This is called `eye(n)` in Matlab (since mathematically it is usually denoted by I).

```
>> I = eye(3), x = [8; -4; 1], I*x
I =
     1     0     0
     0     1     0
     0     0     1
x =
     8
    -4
     1
ans =
     8
    -4
     1
```

Notice that multiplying the 3×1 vector `x` by the 3×3 identity `I` has no effect (it is like multiplying a number by 1).

16.5 Diagonal Matrices

A diagonal matrix is similar to the identity matrix except that its diagonal entries are not necessarily equal to 1.

$$D = \begin{bmatrix} -3 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

is a 3×3 diagonal matrix. To construct this in Matlab, we could either type it in directly

```
>> D = [-3 0 0; 0 4 0; 0 0 2]
D =
    -3     0     0
     0     4     0
     0     0     2
```

but this becomes impractical when the dimension is large (e.g. a 100×100 diagonal matrix). We then use the `diag` function.We first define a vector `d`, say, containing the values of the diagonal entries (in order) then `diag(d)` gives the required matrix.

```
>>  d = [-3 4 2], D = diag(d)
d =
    -3     4     2
D =
    -3     0     0
     0     4     0
     0     0     2
```

On the other hand, if `A` is any matrix, the command `diag(A)` extracts its diagonal entries:

```
>> F = [0 1 8 7; 3 -2 -4 2; 4 2 1 1]
F =
     0     1     8     7
     3    -2    -4     2
     4     2     1     1
>> diag(F)
ans =
     0
    -2
     1
```

Notice that the matrix does not have to be square.

16.6 Building Matrices

It is often convenient to build large matrices from smaller ones:

```
>> C=[0 1; 3 -2; 4 2]; x=[8;-4;1];
>> G = [C x]
G =
     0     1     8
     3    -2    -4
     4     2     1
>> A, B, H = [A; B]
A =
     5     7     9
     1    -3    -7
B =
    -1     2     5
     9     0     5
ans =
     5     7     9
     1    -3    -7
    -1     2     5
     9     0     5
```

so we have added an extra column (`x`) to `C` in order to form `G` and have stacked `A` and `B` on top of each other to form `H`.

```
>> J = [1:4; 5:8; 9:12; 20 0 5 4]
J =
     1     2     3     4
     5     6     7     8
     9    10    11    12
    20     0     5     4

>> K = [ diag(1:4) J; J' zeros(4,4)]
K =
    1    0    0    0    1    2    3    4
    0    2    0    0    5    6    7    8
    0    0    3    0    9   10   11   12
    0    0    0    4   20    0    5    4
    1    5    9   20    0    0    0    0
    2    6   10    0    0    0    0    0
    3    7   11    5    0    0    0    0
    4    8   12    4    0    0    0    0
```

The command `spy(K)` will produce a graphical display of the location of the nonzero entries in `K` (it will also give a value for `nz`—the number of nonzero entries):

```
>> spy(K), grid
```

16.7 Tabulating Functions

This has been addressed in earlier sections but we are now in a position to produce a more suitable table format.

Example 16.1 *Tabulate the functions* $y = 4 \sin 3x$ *and* $u = 3 \sin 4x$ *for* $x = 0, 0.1, 0.2, \ldots, 0.5$.

```
>> x = 0:0.1:0.5;
>> y = 4*sin(3*x); u = 3*sin(4*x);
>> [ x' y' u']
ans =
         0         0         0
    0.1000    1.1821    1.1683
    0.2000    2.2586    2.1521
    0.3000    3.1333    2.7961
    0.4000    3.7282    2.9987
    0.5000    3.9900    2.7279
```

Note the use of transpose (') to get column vectors. (we could replace the last command by `[x; y; u;]'`) We could also have done this more directly:

```
>>  x = (0:0.1:0.5)';
>> [x 4*sin(3*x) 3*sin(4*x)]
```

16.8 Extracting Bits of Matrices

We may extract sections from a matrix in much the same way as for a vector (page 5).
Each element of a matrix is indexed according to which row and column it belongs to. The entry in the ith row and jth column is denoted mathematically by $A_{i,j}$ and, in Matlab, by `A(i,j)`. So

```
>> J
J =
     1     2     3     4
     5     6     7     8
```

```
      9    10    11    12
     20     0     5     4
>> J(1,1)
ans =
      1
>> J(2,3)
ans =
      7
>> J(4,3)
ans =
      5
>> J(4,5)
???  Index exceeds matrix dimensions.
>> J(4,1) = J(1,1) + 6
J =
      1     2     3     4
      5     6     7     8
      9    10    11    12
      7     0     5     4
>> J(1,1) = J(1,1) - 3*J(1,2)
J =
     -5     2     3     4
      5     6     7     8
      9    10    11    12
      7     0     5     4
```

In the following examples we extract i) the 3rd column, ii) the 2nd and 3rd columns, iii) the 4th row, and iv) the "central" 2×2 matrix. See §8.1.

```
>> J(:,3)          % 3rd column
ans =
      3
      7
     11
      5
>> J(:,2:3)        % columns 2 to 3
ans =
      2     3
      6     7
     10    11
      0     5
>> J(4,:)          % 4th row
ans =
      7     0     5     4
>> J(2:3,2:3)      % rows 2 to 3 & cols 2 to 3
ans =
      6     7
     10    11
```

Thus, : on its own refers to the entire column or row depending on whether it is the first or the second index.

16.9 Dot product of matrices (.*)

The dot product works as for vectors: corresponding elements are multiplied together—so the matrices involved must have the same size.

```
>> A, B
A =
      5     7     9
      1    -3    -7
B =
     -1     2     5
      9     0     5
>> A.*B
ans =
     -5    14    45
      9     0   -35
>> A.*C
??? Error using ==> .*
Matrix dimensions must agree.
>> A.*C'
ans =
      0    21    36
      1     6   -14
```

16.10 Matrix–vector products

We turn next to the definition of the product of a matrix with a vector. This product is only defined for **column vectors** that have the same number of entries as the matrix has columns. So, if A is an $m \times n$ matrix and $\underline{x}$ is a column vector of length n, then the matrix–vector $A\underline{x}$ is legal.
An $m \times n$ matrix times an $n \times 1$ matrix $\Rightarrow$ a $m \times 1$ matrix.
We visualise A as being made up of m row vectors stacked on top of each other, then the product corresponds to taking the **scalar** product (See §14.1) of each row of A with the vector $\underline{x}$: The result is a column vector with m entries.

$$\begin{aligned} A\underline{x} &= \left[\begin{array}{c} \boxed{5 \quad 7 \quad 9} \\ \boxed{1 \quad -3 \quad -7} \end{array}\right] \left[\boxed{\begin{array}{r} 8 \\ -4 \\ 1 \end{array}}\right] \\ &= \left[\begin{array}{c} 5 \times 8 + 7 \times (-4) + 9 \times 1 \\ 1 \times 8 + (-3) \times (-4) + (-7) \times 1 \end{array}\right] \\ &= \left[\begin{array}{c} 21 \\ 13 \end{array}\right] \end{aligned}$$

It is somewhat easier in Matlab:

```
>> A = [5 7 9; 1 -3 -7]
A =
      5     7     9
      1    -3    -7
>> x = [8; -4; 1]
x =
      8
     -4
      1
>> A*x
ans =
     21
     13
```

$$(m \times \boxed{n) \text{ times } (n} \times 1) \Rightarrow (m \times 1).$$

```
>> x*A
??? Error using ==> *
Inner matrix dimensions must agree.
```

Unlike multiplication in arithmetic, **`A*x` is not the same as `x*A`**.

16.11 Matrix–Matrix Products

To form the product of an $m \times n$ matrix A and a $n \times p$ matrix B, written as AB, we visualise the first matrix (A) as being composed of m row vectors of length n stacked on top of each other while the second (B) is visualised as being made up of p column vectors of length n:

$$A = m \text{ rows} \left\{ \begin{bmatrix} \boxed{} \\ \boxed{} \\ \vdots \\ \boxed{} \end{bmatrix} \right. , \quad B = \underbrace{\begin{bmatrix} \Big| \Big| \cdots \Big| \end{bmatrix}}_{p \text{ columns}},$$

The entry in the ith row and jth column of the product is then the **scalar**product of the ith row of A with the jth column of B. The product is an $m \times p$ matrix:

$$(m \times \boxed{n) \text{ times } (n} \times p) \Rightarrow (m \times p).$$

Check that you understand what is meant by working out the following examples by hand and comparing with the Matlab answers.

```
>> A = [5  7  9;  1 -3 -7]
A =
     5     7     9
     1    -3    -7
>> B = [0, 1; 3, -2; 4, 2]
B =
     0     1
     3    -2
     4     2
>> C = A*B
C =
    57     9
   -37    -7
>> D = B*A
D =
     1    -3    -7
    13    27    41
    22    22    22
>>  E = B'*A'
E =
    57   -37
     9    -7
```

We see that `E = C'` suggesting that

(A*B)' = B'*A'
Why is $B * A$ a 3×3 matrix while $A * B$ is 2×2?

Exercise 16.1 *It is often necessary to factorize a matrix, e.g., $A = BC$ or $A = S^T XS$ where the factors are required to have specific properties. Use the '`lookfor keyword`' command to make a list of factorizations commands in Matlab.*

16.12 Sparse Matrices

Matlab has powerful techniques for handling sparse matrices — these are generally large matrices (to make the extra work involved worthwhile) that have only a very small proportion of non–zero entries.

Example 16.2 *Create a sparse 5×4 matrix S having only 3 non–zero values: $S_{1,2} = 10$, $S_{3,3} = 11$ and $S_{5,4} = 12$.*

We first create 3 vectors containing the i–index, the j–index and the corresponding values of each term and we then use the `sparse` command.

```
>> i = [1, 3, 5]; j = [2,3,4];
>> v = [10 11 12];
>> S = sparse (i,j,v)
S =
   (1,2)        10
   (3,3)        11
   (5,4)        12
>> T = full(S)
T =
     0    10     0     0
     0     0     0     0
     0     0    11     0
     0     0     0     0
     0     0     0    12
```

The matrix `T` is a "full" version of the sparse matrix `S`.

Example 16.3 *Develop Matlab code to create, for any given value of n, the sparse (tridiagonal) matrix*

$$B = \begin{bmatrix} 1 & n & & & & \\ -2 & 2 & n-1 & & & \\ & -3 & 3 & n-2 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -n+1 & n-1 & 1 \\ & & & & -n & n \end{bmatrix}$$

We define three COLUMN vectors, one for each "diagonal" of non–zeros and then assemble the matrix using `spdiags` (short for sparse diagonals). The vectors are named `l`, `d` and `u`. They must all have the same length and only the **first** $n-1$ terms of `l` are used while the **last** $n-1$ terms of `u` are used. `spdiags` places these vectors in the diagonals labelled `-1`, `0` and `1` (`0` defers to the leading diagonal, negatively numbered diagonals lie below the leading diagonal, etc.)

```
>> n = 5;
>> l = -(2:n+1)'; d = (1:n )'; u = ((n+1):-1:2)';
>> B = spdiags([l' d' u'],-1:1,n,n);
>> full(B)
ans =
     1     5     0     0     0
    -2     2     4     0     0
     0    -3     3     3     0
     0     0    -4     4     2
     0     0     0    -5     5
```

17 Systems of Linear Equations

Mathematical formulations of engineering problems often lead to sets of simultaneous linear equations. Such

is the case, for instance, when using the finite element method (FEM).

A general system of linear equations can be expressed in terms of a coefficient matrix A, a right-hand-side (column) vector $\mathbf{b}$ and an unknown (column) vector $\mathbf{x}$ as

$$A\mathbf{x} = \mathbf{b}$$

or, componentwise, as

$$\begin{aligned} a_{1,1}x_1 + a_{1,2}x_2 + \cdots a_{1,n}x_n &= b_1 \\ a_{2,1}x_1 + a_{2,2}x_2 + \cdots a_{2,n}x_n &= b_2 \\ &\vdots \\ a_{n,1}x_1 + a_{n,2}x_2 + \cdots a_{n,n}x_n &= b_n \end{aligned}$$

When A is non-singular and square ($n \times n$), meaning that the number of *independent* equations is equal to the number of unknowns, the system has a unique solution given by

$$\mathbf{x} = A^{-1}\mathbf{b}$$

where A^{-1} is the inverse of A. Thus, the solution vector $\mathbf{x}$ can, in principle, be calculated by taking the inverse of the coefficient matrix A and multiplying it on the right with the right-hand-side vector $\mathbf{b}$.

This approach based on the matrix inverse, though formally correct, is at best inefficient for practical applications (where the number of equations may be extremely large) but may also give rise to large numerical errors unless appropriate techniques are used. These issues are discussed in most courses and texts on numerical methods. Various stable and efficient solution techniques have been developed for solving linear equations and the most appropriate in any situation will depend on the properties of the coefficient matrix A. For instance, on whether or not it is symmetric, or positive definite or if it has a particular structure (sparse or full). Matlab is equipped with many of these special techniques in its routine library and they are invoked automatically.

The standard Matlab routine for solving systems of linear equations is invoked by calling the matrix left-division routine,

```
>> x = A \ b
```

where "\" is the matrix left-division operator known as "backslash" (see `help backslash`).

Exercise 17.1 *Enter the symmetric coefficient matrix and right–hand–side vector* $\mathbf{b}$ *given by*

$$A = \begin{bmatrix} 2 & -1 & 0 \\ 1 & -2 & 1 \\ 0 & -1 & 2 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

and solve the system of equations $A\mathbf{x} = \mathbf{b}$ *using the three alternative methods:*

i) $\mathbf{x} = A^{-1}\mathbf{b}$, *(the inverse* A^{-1} *may be computed in Matlab using* `inv(A)`*.)*

ii) `x = A \ b`*,*

iii) $\mathbf{x}^T = \mathbf{b}^t A^T$ *leading to* `xT = b' / A` *which makes use of the "slash" or "right division" operator "/". The required solution is then the transpose of the row vector* `xT`*.*

Exercise 17.2 *Use the backslash operator to solve the complex system of equations for which*

$$A = \begin{bmatrix} 2+2\mathrm{i} & -1 & 0 \\ -1 & 2-2\mathrm{i} & -1 \\ 0 & -1 & 2 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 1+\mathrm{i} \\ 0 \\ 1-\mathrm{i} \end{bmatrix}$$

Exercise 17.3 *Find information on the matrix inversion command '*`inv`*' using each of the methods listed in Section 2 for obtaining help.*

What kind of matrices are the '`inv`*' command applicable to?*

Obviously problems may occur if the inverted matrix is nearly singular. Suggest a command that can be used to give an indication on whether the matrix is nearly singular or not. [Hint: see the topics referred to by '`help inv`*'.]*

17.1 Overdetermined system of linear equations

An overdetermined system of linear equations is a one with more equations (m) than unknowns (n), i.e., the coefficient matrix has more rows than columns ($m > n$). Overdetermined systems frequently appear in mathematical modelling when the parameters of a model are determined by fitting to experimental data. Formally the system looks the same as for square systems but the coefficient matrix is rectangular and so it is not possible to compute an inverse. In these cases a solution can be found by requiring that the magnitude of the residual vector $\mathbf{r}$, defined by

$$\mathbf{r} = A\mathbf{x} - \mathbf{b},$$

be minimized. The simplest and most frequently used measure of the magnitude of $\mathbf{r}$ is require the Euclidean length (or norm—see Section 14.1) which corresponds to the sum of squares of the components of the residual. This approach leads to the least squares solution of the overdetermined system. Hence the least squares solution is defined as the vector $\mathbf{x}$ that minimizes

$$\mathbf{r}^T\mathbf{r}.$$

It may be shown that the required solution satisfies the so–called "normal equations"

$$C\mathbf{x} = \mathbf{d}, \text{ where } C = A^T A \text{ and } \mathbf{d} = A^T\mathbf{b}.$$

This system is well–known that the solution of this system can be overwhelmed by numerical rounding error in practice unless great care is taken in its solution (a large part of the difficulty is inherent in computing the matrix–matrix product $A^T A$). As in the solution of square systems of linear equations, special techniques have been developed to address these issues and they have been incorporated into the Matlab routine library. This means that a direct solution to the problem of overdetermined equations is available in Matlab through its left division operator "\". When the matrix A is not square, the operation

```
x = A\b
```

automatically gives the least squares solution to $A\mathbf{x} = \mathbf{b}$. This is illustrated in the next example.

Example 17.1 *A spring is a mechanical element which, for the simplest model, is characterized by a linear force-deformation relationship*

$$F = kx,$$

F being the force loading the spring, k the spring constant or stiffness and x the spring deformation. In reality the linear force–deformation relationship is only an approximation, valid for small forces and deformations. A more accurate relationship, valid for larger deformations, is obtained if non–linear terms are taken into account. Suppose a spring model with a quadratic relationship

$$F = k_1 x + k_2 x^2$$

is to be used and that the model parameters, k_1 and k_2, are to be determined from experimental data. Five independent measurements of the force and the corresponding spring deformations are measured and these are presented in Table 1.

Force F [N]	Deformation x [cm]
5	0.001
50	0.011
500	0.013
1000	0.30
2000	0.75

Table 1: Measured force-deformation data for spring.

Using the quadratic force-deformation relationship together with the experimental data yields an overdetermined system of linear equations and the components of the residual are given by

$$\begin{aligned} r_1 &= x_1 k_1 + x_1^2 k_2 - F_1 \\ r_2 &= x_2 k_1 + x_2^2 k_2 - F_2 \\ r_3 &= x_3 k_1 + x_3^2 k_2 - F_3 \\ r_4 &= x_4 k_1 + x_4^2 k_2 - F_4 \\ r_5 &= x_5 k_1 + x_5^2 k_2 - F_5. \end{aligned}$$

These lead to the matrix and vector definitions

$$A = \begin{bmatrix} x_1 & x_1^2 \\ x_2 & x_2^2 \\ x_3 & x_3^2 \\ x_4 & x_4^2 \\ x_5 & x_5^2 \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} F_1 \\ F_2 \\ F_3 \\ F_4 \\ F_5. \end{bmatrix}$$

The appropriate Matlab commands give (the components of `x` are all multiplied by `1e-2`, i.e., 10^{-2}, in order to change from cm to m)

```
>> x = [.001 .011 .13 .3 .75]*1e-2;
>> A = [x' (x').^2]
A =
    0.0000    0.0000
    0.0001    0.0000
    0.0013    0.0000
    0.0030    0.0000
    0.0075    0.0001
>> b = [5 50 500 1000 2000];
```

and the least squares solution to this system is given by

```
>> k = A\b'
k =
  1.0e+07 *
    0.0386
   -1.5993
```

Thus, $\mathbf{k} \approx \begin{bmatrix} 0.39 \\ -16.0 \end{bmatrix} \times 10^6$ and the quadratic spring force-deformation relationship that optimally fits experimental data in the least squares sense is

$$F \approx 38.6 \times 10^4 x - 16.0 \times 10^6 x^2.$$

The data and solution may be plotted with the following commands

```
>> plot(x,f,'o'), hold on      % plot data points
>> X = (0:.01:1)*max(x);
>> plot(X,[X' (X.^2)']*k,'-')  % best fit curve
>> xlabel('x[m]'), ylabel('F[N]')
```

and the results are shown in Figure 5.

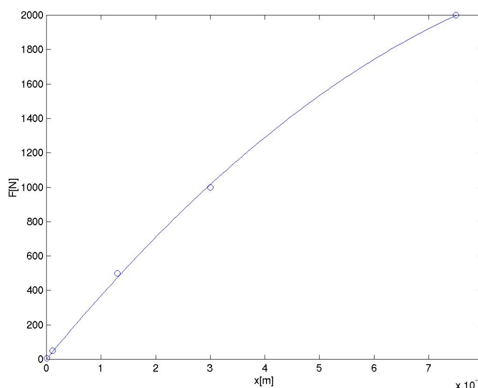


Figure 5: Data for Example 17.1 (circles) and best least squares fit by a quadratic model (solid line).

Matlab has a routine `polyfit` for data fitting by polynomials: see "`help polyfit`". It is not applicable in this example because we require that the force – deformation law passes through the origin (so there is no constant term in the quadratic model that we used).

18 Characters, Strings and Text

The ability to process text in numerical processing is useful for the input and output of data to the screen or to disk-files. In order to manage text, a new datatype of "character" is introduced. A piece of text is then simply a string (vector) or array of characters.

Example 18.1 *The assignment,*

```
>> t1 = 'A'
```

assigns the value `A` to the `1-by-1` character array `t1`. The assignment,

```
>> t2 = 'BCDE'
```

assigns the value BCDE to the 1-by-4 character array t2.

Strings can be combined by using the operations for array manipulations.
The assignment,

```
>> t3 = [t1,t2]
```

assigns a value `ABCDE` to the `1-by-5` character array `t3`.
The assignment,

```
>> t4 = [t3,' are the first 5       ';...
'characters in the alphabet.']
```

assigns the value
`'ABCDE are the first 5 '`
`'characters in the alphabet.'`
to the `2-by-27` character array `t4`. It is essential that the number of characters in both rows of the array `t4` is the same, otherwise an error will result. The three dots `...` signify that the command is continued on the following line
Sometimes it is necessary to convert a character to the corresponding number, or vice versa. These conversions are accomplished by the commands '`str2num`'—which converts a string to the corresponding number, and two functions, '`int2str`' and '`num2str`', which convert, respectively, an integer and a real number to the corresponding character string. These commands are useful for producing titles and strings, such as '`The value of pi is 3.1416`'. This can be generated by the command `['The value of pi is ',num2str(pi)]`.

```
>> N = 5; h = 1/N;
>> ['The value of N is ',int2str(N),...
', h = ',num2str(h)]
ans =
The value of N is 5, h = 0.2
```

19 Loops

There are occasions that we want to repeat a segment of code a number of different times (such occasions are less frequent than other programming languages because of the `:` notation).

Example 19.1 *Draw graphs of* $\sin(n\pi x)$ *on the interval* $-1 \le x \le 1$ *for* $n = 1, 2, \ldots, 8$.

We could do this by giving 8 separate `plot` commands but it is much easier to use a loop. The simplest form would be

```
>> x = -1:.05:1;
>> for n = 1:8
     subplot(4,2,n), plot(x,sin(n*pi*x))
   end
```

All the commands between the lines starting "`for`" and "`end`" are repeated with `n` being given the value `1` the first time through, `2` the second time, and so forth, until $n = 8$. The `subplot` constructs a 4×2 array of subwindows and, on the nth time through the loop, a picture is drawn in the nth subwindow.

The commands

```
>> x = -1:.05:1;
>> for n = 1:2:8
     subplot(4,2,n), plot(x,sin(n*pi*x))
     subplot(4,2,n+1), plot(x,cos(n*pi*x))
   end
```

draw $\sin n\pi x$ and $\cos n\pi x$ for $n = 1, 3, 5, 7$ alongside each other.
We may use any legal variable name as the "loop counter" (`n` in the above examples) and it can be made to run through all of the values in a given vector (`1:8` and `1:2:8` in the examples).
We may also use `for` loops of the type

```
>> for counter = [23  11  19  5.4  6]
      .......
   end
```

which repeats the code as far as the `end` with `counter=23` the first time, `counter=11` the second time, and so forth.

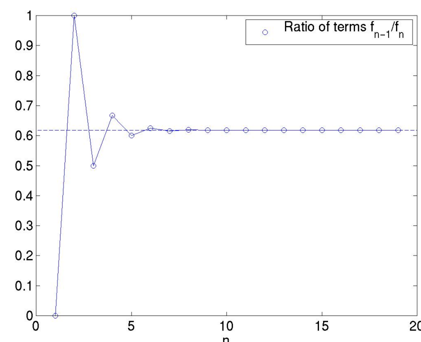
Example 19.2 *The Fibonnaci sequence starts off with the numbers* 0 *and* 1*, then succeeding terms are the sum of its two immediate predecessors. Mathematically,* $f_1 = 0$*,* $f_2 = 1$ *and*

$$f_n = f_{n-1} + f_{n-2}, \qquad n = 3, 4, 5, \ldots.$$

Test the assertion that the ratio f_{n-1}/f_n *of two successive values approaches the golden ratio* $(\sqrt{5}-1)/2$ $= 0.6180\ldots$.

```
>> F(1) = 0; F(2) = 1;
>> for i = 3:20
     F(i) = F(i-1) + F(i-2);
   end
>> plot(1:19, F(1:19)./F(2:20),'o' )
>> hold on,  xlabel('n')
>> plot(1:19, F(1:19)./F(2:20),'-' )
>> legend('Ratio of terms f_{n-1}/f_n')
>> plot([0 20], (sqrt(5)-1)/2*[1,1],'--')
```

The last of these commands produces the dashed horizontal line.



Example 19.3 *Produce a list of the values of the sums*

$$
\begin{aligned}
S_{20} &= 1+\tfrac{1}{2^2}+\tfrac{1}{3^2}+\cdots+\tfrac{1}{20^2}\\
S_{21} &= 1+\tfrac{1}{2^2}+\tfrac{1}{3^2}+\cdots+\tfrac{1}{20^2}+\tfrac{1}{21^2}\\
&\vdots\\
S_{100} &= 1+\tfrac{1}{2^2}+\tfrac{1}{3^2}+\cdots+\tfrac{1}{20^2}+\tfrac{1}{21^2}+\cdots+\tfrac{1}{100^2}
\end{aligned}
$$

There are a total of 81 sums. The first can be computed using `sum(1./(1:20).^2)` (The function `sum` with a vector argument sums its components. See §22.2].) A suitable piece of Matlab code might be

```
>> S = zeros(100,1);
>> S(20) = sum(1./(1:20).^2);
>> for n = 21:100
>>   S(n) = S(n-1) + 1/n^2;
>> end
>> clf; plot(S,'.',[20 100],[1,1]*pi^2/6,'-')
>> axis([20 100 1.5 1.7])
>> [ (98:100)'  S(98:100)]
ans =
   98.0000    1.6364
   99.0000    1.6365
  100.0000    1.6366
```

where a column vector `S` was created to hold the answers. The first sum was computed directly using the `sum` command then each succeeding sum was found by adding $1/n^2$ to its predecessor. The little table at the end shows the values of the last three sums—it appears that they are approaching a limit (the value of the limit is $\pi^2/6 = 1.64493\ldots$).

Exercise 19.1 *Repeat Example 19.3 to include 181 sums (i.e., the final sum should include the term* $1/200^2$*.)*

20 Logicals

Matlab represents `true` and `false` by means of the integers 0 and 1.

`true = 1,    false = 0`

If at some point in a calculation a scalar `x`, say, has been assigned a value, we may make certain logical tests on it:

`x == 2`	is `x` equal to 2?
`x ~= 2`	is `x` **not** equal to 2?
`x > 2`	is `x` greater than 2?
`x < 2`	is `x` less than 2?
`x >= 2`	is `x` greater than or equal to 2?
`x <= 2`	is `x` less than or equal to 2?

Pay particular attention to the fact that the test for equality involves two equal signs `==`.

```
>> x = pi
x =
    3.1416
>> x ~= 3,  x ~= pi
ans =
     1
ans =
     0
```

When `x` is a vector or a matrix, these tests are performed elementwise:

```
x =
   -2.0000    3.1416    5.0000
   -1.0000         0    1.0000
>> x == 0
ans =
     0     0     0
     0     1     0
>> x > 1,  x >=-1
ans =
     0     1     1
     0     0     0
ans =
     0     1     1
     1     1     1
>> y = x>=-1, x > y
y =
     0     1     1
     1     1     1
ans =
     0     1     1
     0     0     0
```

We may combine logical tests, as in

```
>> x
x =
   -2.0000    3.1416    5.0000
   -5.0000   -3.0000   -1.0000
>> x > 3 & x < 4
ans =
     0     1     0
     0     0     0
>> x > 3 | x == -3
ans =
     0     1     1
     0     1     0
```

As one might expect, `&` represents **and** and (not so clearly) the vertical bar `|` means **or**; also `~` means **not** as in `~=` (not equal), `~(x>0)`, etc.

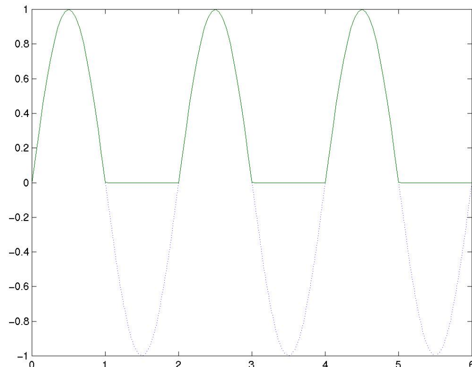
```
>> x > 3 | x == -3 | x <= -5
ans =
     0     1     1
     1     1     0
```

One of the uses of logical tests is to "mask out" certain elements of a matrix.

```
>> x, L =  x >= 0
x =
   -2.0000    3.1416    5.0000
   -5.0000   -3.0000   -1.0000
L =
     0     1     1
     0     1     1
>> pos = x.*L
pos =
         0    3.1416    5.0000
         0         0         0
```

so the matrix `pos` contains just those elements of `x` that are non–negative.

```
>> x = 0:0.05:6; y = sin(pi*x); Y = (y>=0).*y;
>> plot(x,y,':',x,Y,'-' )
```



20.1 While Loops

There are some occasions when we want to repeat a section of Matlab code until some logical condition is satisfied, but we cannot tell in advance how many times we have to go around the loop. This we can do with a `while...end` construct.

Example 20.1 *What is the greatest value of n that can be used in the sum*

$$1^2 + 2^2 + \cdots + n^2$$

and get a value of less than 100?

```
>> S = 1; n = 1;
>> while S+ (n+1)^2 < 100
    n = n+1;   S = S + n^2;
   end
>> [n, S]
ans =
     6    91
```

The lines of code between `while` and `end` will only be executed if the condition `S+ (n+1)^2 < 100` is true.

Exercise 20.1 *Replace 100 in the previous example by 10 and work through the lines of code by hand. You should get the answers $n = 2$ and $S = 5$.*

Exercise 20.2 *Type the code from Example20.1 into a script–file named `WhileSum.m` (See §13.)*

A more typical example is

Example 20.2 *Find the approximate value of the root of the equation $x = \cos x$. (See Example 10.1.)*

We may do this by making a guess $x_1 = \pi/4$, say, then computing the sequence of values

$$x_n = \cos x_{n-1}, \qquad n = 2, 3, 4, \ldots$$

and continuing until the difference between two successive values $|x_n - x_{n-1}|$ is small enough.

Method 1:

```
>> x = zeros(1,20); x(1) = pi/4;
>> n = 1;  d = 1;
>> while d > 0.001
     n = n+1; x(n) = cos(x(n-1));
     d = abs( x(n) - x(n-1) );
   end
   n,x
n =
    14
x =
Columns 1 through 7
0.7854 0.7071 0.7602 0.7247 0.7487 0.7326 0.7435
Columns 8 through 14
0.7361 0.7411 0.7377 0.7400 0.7385 0.7395 0.7388
Columns 15 through 20
   0    0    0    0    0    0
```

There are a number of deficiencies with this program. The vector `x` stores the results of each iteration but we don't know in advance how many there may be. In any event, we are rarely interested in the intermediate values of `x`, only the last one. Another problem is that we may never satisfy the condition $d \le 0.001$, in which case the program will run forever—we should place a limit on the maximum number of iterations.
Incorporating these improvements leads to

Method 2:

```
>> xold = pi/4; n = 1;  d = 1;
>> while d > 0.001 &  n < 20
     n = n+1;  xnew = cos(xold);
     d = abs( xnew - xold );
     xold = xnew;
   end
>> [n, xnew, d]
ans =
   14.0000    0.7388    0.0007
```

We continue around the loop so long as $d > 0.001$ **and** $n < 20$. For greater precision we could use the condition $d > 0.0001$, and this gives

```
>> [n, xnew, d]
ans =
   19.0000    0.7391    0.0001
```

from which we may judge that the root required is $x = 0.739$ to 3 decimal places.
The general form of `while` statement is

```
while a logical test
                Commands to be executed
                when the condition is true
end
```

20.2 `if...then...else...end`

This allows us to execute different commands depending on the truth or falsity of some logical tests. To test whether or not π^e is greater than, or equal to, e^π:

```
>> a = pi^exp(1); c = exp(pi);
>> if a >= c
     b = sqrt(a^2 - c^2)
   end
```

so that `b` is assigned a value only if $a \geq c$. There is no output so we deduce that $a = \pi^e < c = e^\pi$. A more common situation is

```
>> if a >= c
     b = sqrt(a^2 - c^2)
   else
     b = 0
   end
b =
     0
```

which ensures that `b` is always assigned a value and confirming that $a < c$.
A more extended form is

```
>> if a >= c
     b = sqrt(a^2 - c^2)
   elseif a^c > c^a
     b = c^a/a^c
   else
     b = a^c/c^a
   end
b =
    0.2347
```

Exercise 20.3 *Which of the above statements assigned a value to* `b`*?*

The general form of the `if` statement is

> `if` *logical test 1*
>   *Commands to be executed if test 1 is true*
> `elseif` *logical test 2*
>   *Commands to be executed if test 2 is true but test 1 is false*
>   ⋮
>
> `end`

21 Function m–files

These are a combination of the ideas of script m–files (§7) and mathematical functions.

Example 21.1 *The area, A, of a triangle with sides of length a, b and c is given by*

$$A = \sqrt{s(s-a)(s-b)(s-c)},$$

where $s = (a+b+c)/2$. Write a Matlab function that will accept the values a, b and c as inputs and return the value of A as output.

The main steps to follow when defining a Matlab function are:

1. Decide on a name for the function, making sure that it does not conflict with a name that is already used by Matlab. In this example the name of the function is to be `area`, so its definition will be saved in a file called `area.m`
2. The first line of the file must have the format:
   `function [`*list of outputs*`]`
   `=` *function_name*`(`*list of inputs*`)`
   For our example, the output (A) is a function of the three variables (inputs) a, b and c so the first line should read

   ```
   function [A] = area(a,b,c)
   ```
3. Document the function. That is, describe briefly the purpose of the function and how it can be used. These lines should be preceded by `%` which signify that they are comment lines that will be ignored when the function is evaluated.
4. Finally include the code that defines the function. This should be interspersed with sufficient comments to enable another user to understand the processes involved.

The complete file might look like:

```
function  [A] = area(a,b,c)
% Compute the area of a triangle whose
% sides have length a, b and c.
% Inputs:
%    a,b,c:  Lengths of sides
% Output:
%        A: area of triangle
% Usage:
%      Area = area(2,3,4);
% Written by dfg, Oct 14, 1996.
s = (a+b+c)/2;
A = sqrt(s*(s-a)*(s-b)*(s-c));
%%%%%%%%% end of area %%%%%%%%%%%
```

The command
`>> help area`
will produce the leading comments from the file:

```
  Compute the area of a triangle whose
  sides have length a, b and c.
  Inputs:
     a,b,c:  Lengths of sides
  Output:
         A: area of triangle
  Usage:
       Area = area(2,3,4);
  Written by dfg, Oct 14, 1996.
```

To evaluate the area of a triangle with side of length 10, 15, 20:

```
>> Area = area(10,15,20)
Area =
   72.6184
```

where the result of the computation is assigned to the variable `Area`. The variable `s` used in the definition of the function above is a "local variable": its value is local to the function and cannot be used outside:

```
>> s
??? Undefined function or variable s.
```

If we were to be interested in the value of s as well as A, then the first line of the file should be changed to

```
function [A,s] = area(a,b,c)
```

where there are two output variables.
This function can be called in several different ways:

1. No outputs assigned

   ```
   >> area(10,15,20)
   ans =
      72.6184
   ```

   gives only the area (first of the output variables from the file) assigned to `ans`; the second output is ignored.

2. One output assigned

   ```
   >> Area = area(10,15,20)
   Area =
      72.6184
   ```

   again the second output is ignored.

3. Two outputs assigned

   ```
   >> [Area, hlen] = area(10,15,20)
   Area =
      72.6184
   hlen =
      22.5000
   ```

Exercise 21.1 *In any triangle the sum of the lengths of any two sides cannot exceed the length of the third side. The function* `area` *does not check to see if this condition is fulfilled (try* `area(1,2,4)`*). Modify the file so that it computes the area only if the sides satisfy this condition.*

21.1 Examples of functions

We revisit the problem of computing the Fibonnaci sequence defined by $f_1 = 0, f_2 = 1$ and

$$f_n = f_{n-1} + f_{n-2}, \qquad n = 3, 4, 5, \ldots.$$

We want to construct a function that will return the nth number in the Fibonnaci sequence f_n.

- **Input**: Integer n
- **Output**: f_n

We shall describe four possible functions and try to assess which provides the best solution.

Method 1: File `Fib1.m`

```
function f = Fib1(n)
% Returns the nth number in the
% Fibonacci sequence.
F=zeros(1,n+1);
F(2) = 1;
   for i = 3:n+1
     F(i) = F(i-1) + F(i-2);
   end
f = F(n);
```

This code resembles that given in Example 19.2. We have simply enclosed it in a function m–file and given it the appropriate header,

Method 2: File `Fib2.m`

The first version was rather wasteful of memory—it saved all the entries in the sequence even though we only required the last one for output. The second version removes the need to use a vector.

```
function f = Fib2(n)
% Returns the nth number in the
% Fibonacci sequence.
if n==1
   f = 0;
elseif n==2
   f = 1;
else
   f1 = 0; f2 = 1;
   for i = 2:n-1
     f = f1 + f2;
     f1=f2; f2 = f;
   end
end
```

Method 3: File: `Fib3.m`

This version makes use of an idea called "recursive programming"— the function makes calls to itself.

```
function f = Fib3(n)
% Returns the nth number in the
% Fibonacci sequence.
if n==1
   f = 0;
elseif n==2
   f = 1;
else
   f = Fib3(n-1) + Fib3(n-2);
end
```

Method 4: File `Fib4.m`

The final version uses matrix powers. The vector `y` has two components, $\underline{y} = \begin{bmatrix} f_n \\ f_{n+1} \end{bmatrix}$.

```
function f = Fib4(n)
% Returns the nth number in the
% Fibonacci sequence.
A = [0 1;1 1];
y = A^n*[1;0];
f=y(1);
```

Assessment: One may think that, on grounds of style, the 3rd is best (it avoids the use of loops) followed by the second (it avoids the use of a vector). The situation is much different when it cames to speed of execution. When $n = 20$ the time taken by each of the methods is (in seconds)

Method	Time
1	0.0118
2	0.0157
3	36.5937
4	0.0078

It is impractical to use Method 3 for any value of n much larger than 10 since the time taken by method 3 almost doubles whenever n is increased by just 1. When $n = 150$

Method	Time
1	0.0540
2	0.0891
3	—
4	0.0106

Clearly the 4th method is much the fastest.

22 Further Built–in Functions

22.1 Rounding Numbers

There are a variety of ways of rounding and chopping real numbers to give integers. Use the definitions given in the table in §28 on page 32 in order to understand the output given below:

```
>> x = pi*(-1:3),  round(x)
x =
  -3.1416  0    3.1416  6.2832   9.4248
ans =
    -3     0     3     6     9
>> fix(x)
ans =
    -3     0     3     6     9
>> floor(x)
ans =
    -4     0     3     6     9
>> ceil(x)
ans =
    -3     0     4     7    10
>> sign(x),  rem(x,3)
ans =
    -1     0     1     1     1
ans =
   -0.1416  0    0.1416   0.2832   0.4248
```

Do "`help round`" for help information.

22.2 The `sum` Function

The "`sum`" applied to a vector adds up its components (as in `sum(1:10)`) while, for a matrix, it adds up the components in **each column** and returns a row vector. `sum(sum(A))` then sums all the entries of `A`.

```
>>  A = [1:3; 4:6; 7:9]
A =
     1     2     3
     4     5     6
     7     8     9
>> s = sum(A), ss = sum(sum(A))
s =
    12    15    18
ss =
    45

>> x = pi/4*(1:3)';
>> A = [sin(x), sin(2*x), sin(3*x)]/sqrt(2)
>> A =
    0.5000    0.7071    0.5000
    0.7071    0.0000   -0.7071
    0.5000   -0.7071    0.5000

>> s1 = sum(A.^2), s2 = sum(sum(A.^2))
s1 =
    1.0000    1.0000    1.0000
s2 =
    3.0000
```

The sums of squares of the entries in each column of `A` are equal to 1 and the sum of squares of all the entries is equal to 3.

```
>> A*A'
ans =
    1.0000         0         0
         0    1.0000    0.0000
         0    0.0000    1.0000
>> A'*A
ans =
    1.0000         0         0
         0    1.0000    0.0000
         0    0.0000    1.0000
```

It appears that the products AA' and $A'A$ are both equal to the identity:

```
>> A*A' - eye(3)
ans =
   1.0e-15 *
   -0.2220         0         0
         0   -0.2220    0.0555
         0    0.0555   -0.2220
>> A'*A - eye(3)
ans =
   1.0e-15 *
   -0.2220         0         0
         0   -0.2220    0.0555
         0    0.0555   -0.2220
```

This is confirmed since the differences are at round–off error levels (less than 10^{-15}). A matrix with this property is called an *orthogonal* matrix.

22.3 `max` & `min`

These functions act in a similar way to `sum`. If `x` is a vector, then `max(x)` returns the largest element in `x`

```
>> x = [1.3 -2.4 0 2.3], max(x), max(abs(x))
x =
    1.3000   -2.4000         0    2.3000
ans =
    2.3000
ans =
```

```
    2.4000
>> [m, j] = max(x)
m =
    2.3000
j =
     4
```

When we ask for two outputs, the first gives us the maximum entry and the second the index of the maximum element.
For a matrix, `A`, `max(A)` returns a row vector containing the maximum element from each column. Thus to find the largest element in `A` we have to use `max(max(A))`.

22.4 Random Numbers

The function `rand(m,n)` produces an $m \times n$ matrix of random numbers, each of which is in the range 0 to 1. `rand` on its own produces a single random number.

```
>> y =  rand, Y = rand(2,3)
y =
    0.9191
Y =
    0.6262    0.1575    0.2520
    0.7446    0.7764    0.6121
```

Repeating these commands will lead to different answers.
Example: *Write a function–file that will simulate n throws of a pair of dice.*
This requires random numbers that are integers in the range 1 to 6. Multiplying each random number by 6 will give a real number in the range 0 to 6; rounding these to whole numbers will not be correct since it will then be possible to get 0 as an answer. We need to use

```
floor(1 + 6*rand)
```

Recall that `floor` takes the largest integer that is smaller than a given real number (see Table 2, page 32).
File: `dice.m`

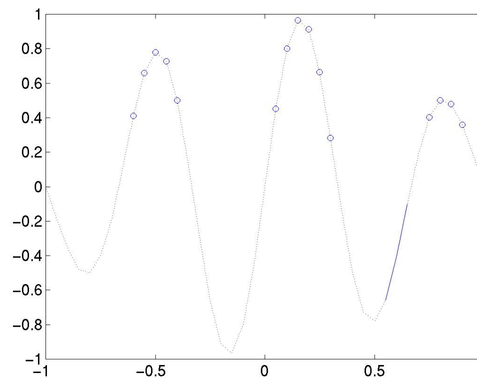
```
function  [d] = dice(n)
% simulates "n" throws of a pair of dice
% Input:     n, the number of throws
% Output:    an n times 2 matrix, each row
%            referring to one throw.
%
% Useage:  T = dice(3)
    d = floor(1 + 6*rand(n,2));
%% end of dice

>> dice(3)
ans =
     6     1
     2     3
     4     1
>> sum(dice(100))/100
ans =
    3.8500    3.4300
```

The last command gives the average value over 100 throws (it should have the value 3.5).

22.5 `find` for vectors

The function "`find`" returns a list of the positions (indices) of the elements of a vector satisfying a given condition. For example,

```
>> x = -1:.05:1;
>> y = sin(3*pi*x).*exp(-x.^2); plot(x,y,':')
>> k = find(y > 0.2)
 k =
  Columns 1 through 12
   9  10  11  12  13  22  23  24  25  26  27  36
  Columns 13 through 15
  37  38  39
>> hold on, plot(x(k),y(k),'o')
>> km = find( x>0.5 & y<0)
km =
    32    33    34
>> plot(x(km),y(km),'-')
```



22.6 `find` for matrices

The `find`–function operates in much the same way for matrices:

```
>> A = [ -2 3 4 4; 0 5 -1 6; 6 8 0 1]
A =
    -2     3     4     4
     0     5    -1     6
     6     8     0     1
>> k = find(A==0)
k =
     2
     9
```

Thus, we find that `A` has elements equal to 0 in positions 2 and 9. To interpret this result we have to recognize that "`find`" first reshapes `A` into a column vector—this is equivalent to numbering the elements of `A` by columns as in

1	4	7	10
2	5	8	11
3	6	9	12

```
>> n  = find(A <= 0)
n =
```

```
     1
     2
     8
     9
>> A(n)
ans =
    -2
     0
    -1
     0
```

Thus, n gives a list of the locations of the entries in A that are ≤ 0 and then A(n) gives us the values of the elements selected.

```
>> m = find( A' == 0)
m =
     5
    11
```

Since we are dealing with A', the entries are numbered by rows.

23 Plotting Surfaces

A surface is defined mathematically by a function $f(x, y)$—corresponding to each value of (x, y) we compute the height of the function by

$$z = f(x, y).$$

In order to plot this we have to decide on the ranges of x and y—suppose $2 \leq x \leq 4$ and $1 \leq y \leq 3$. This gives us a square in the (x, y)–plane. Next, we need to choose a grid on this domain; Figure 6 shows the grid with intervals 0.5 in each direction. Finally, we have to evaluate the function at each point of the grid and "plot" it.

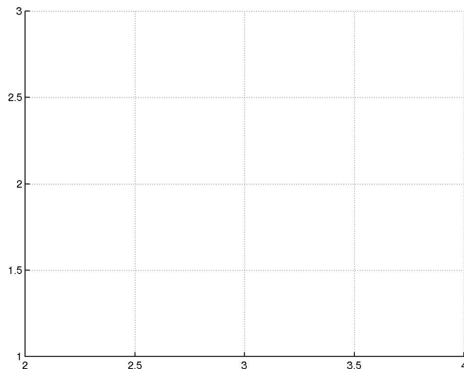


Figure 6: An example of a 2D grid

Suppose we choose a grid with intervals 0.5 in each direction for illustration. The x– and y–coordinates of the grid lines are

```
x = 2:0.5:4;   y = 1:0.5:3;
```

in Matlab notation. We construct the grid with meshgrid:

```
>> [X,Y] = meshgrid(2:.5:4, 1:.5:3);
>> X
X =
   2.0000   2.5000   3.0000   3.5000   4.0000
   2.0000   2.5000   3.0000   3.5000   4.0000
   2.0000   2.5000   3.0000   3.5000   4.0000
   2.0000   2.5000   3.0000   3.5000   4.0000
   2.0000   2.5000   3.0000   3.5000   4.0000
>> Y
Y =
   1.0000   1.0000   1.0000   1.0000   1.0000
   1.5000   1.5000   1.5000   1.5000   1.5000
   2.0000   2.0000   2.0000   2.0000   2.0000
   2.5000   2.5000   2.5000   2.5000   2.5000
   3.0000   3.0000   3.0000   3.0000   3.0000
```

If we think of the ith point along from the left and the jth point up from the bottom of the grid) as corresponding to the (i, j)th entry in a matrix, then (X(i,j), Y(i,j)) are the coordinates of the point. We then need to evaluate the function f using X and Y in place of x and y, respectively.

Example 23.1 *Plot the surface defined by the function*

$$f(x, y) = (x - 3)^2 - (y - 2)^2$$

for $2 \leq x \leq 4$ *and* $1 \leq y \leq 3$.

```
>> [X,Y] = meshgrid(2:.2:4, 1:.2:3);
>> Z = (X-3).^2-(Y-2).^2;
>> mesh(X,Y,Z)
>> title('Saddle'), xlabel('x'),ylabel('y')
```

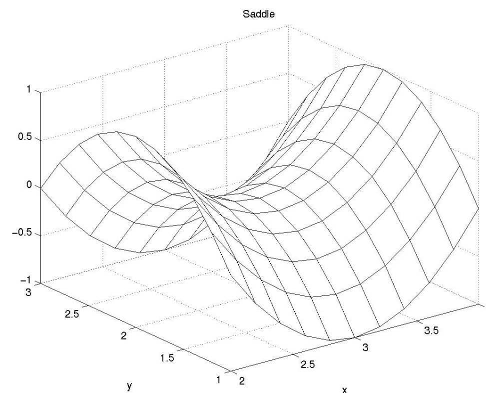


Figure 7: Plot of Saddle function.

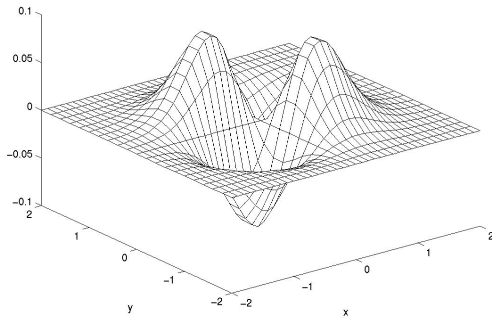
Exercise 23.1 *Repeat the previous example replacing* mesh *by* surf *and then by* surfl. *Consult the help pages to find out more about these functions.*

Example 23.2 *Plot the surface defined by the function*

$$f = -xye^{-2(x^2+y^2)}$$

on the domain $-2 \leq x \leq 2, -2 \leq y \leq 2$. *Find the values and locations of the maxima and minima of the function.*

```
>> [X,Y] = meshgrid(-2:.1:2,-2:.2:2);
>> f =  -X.*Y.*exp(-2*(X.^2+Y.^2));
>> figure (1)
>> mesh(X,Y,f), xlabel('x'), ylabel('y'), grid
>> figure (2),  contour(X,Y,f)
>> xlabel('x'), ylabel('y'), grid, hold on
```



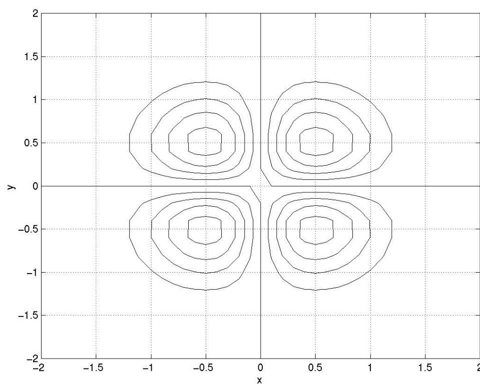


Figure 8: "`mesh`" and "`contour`" plots.

To locate the maxima of the "`f`" values on the grid:

```
>> fmax = max(max(f))
fmax =
    0.0886
>> kmax = find(f==fmax)
kmax =
   323
   539
>> Pos = [X(kmax), Y(kmax)]
Pos =
   -0.5000    0.6000
    0.5000   -0.6000
>> plot(X(kmax),Y(kmax),'*')
>> text(X(kmax),Y(kmax),' Maximum')
```

24 Timing

Matlab allows the timing of sections of code by providing the functions `tic` and `toc`. `tic` switches on a stopwatch while `toc` stops it and returns the CPU time

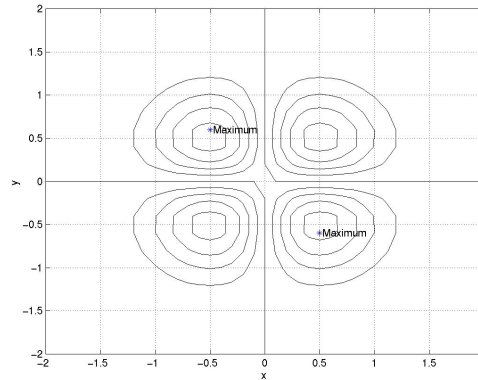


Figure 9: `contour` plot showing maxima.

(Central Processor Unit) in seconds. The timings will vary depending on the model of computer being used and its current load.

```
>> tic,for j=1:1000,x = pi*R(3);end,toc
elapsed_time =     0.5110
>> tic,for j=1:1000,x=pi*R(3);end,toc
elapsed_time =    0.5017
>> tic,for j=1:1000,x=R(3)/pi;end,toc
elapsed_time =    0.5203
>> tic,for j=1:1000,x=pi+R(3);end,toc
elapsed_time =    0.5221
>> tic,for j=1:1000,x=pi-R(3);end,toc
elapsed_time =    0.5154
>> tic,for j=1:1000,x=pi^R(3);end,toc
elapsed_time =    0.6236
```

25 On–line Documentation

In addition to the on–line help facility, there is a hypertext browsing system giving details of (most) commands and some examples. This is accessed by

```
>> doc
```

which brings up the *Netscape* document previewer (and allows for "surfing the internet superhighway"—the World Wide Web (WWW). It is connected to a worldwide system which, given the appropriate addresses, will provide information on almost any topic).
Words that are underlined in the browser may be clicked on with LB and lead to either a further subindex or a help page.
Scroll down the page shown and click on general which will take you to "General Purpose Commands"; click on clear. This will describe how you can clear a variable's value from memory.
You may then either click the "Table of Contents" which takes you back to the start, "Index" or the Back button at the lower left corner of the window which will take you back to the previous screen.
To access other "home pages", click on Open at the bottom of the window and, in the "box" that will open up, type

http://www.maths.dundee.ac.uk

or

http://www.maths.dundee.ac.uk/software/

and select Matlab from the array of choices.

26 Reading and Writing Data Files

Direct input of data from keyboard becomes impractical when

- the amount of data is large and
- the same data is analysed repeatedly.

In these situations input and output is preferably accomplished via data files. We have already described in Section 9 the use of the commands `save` and `load` that, respectively, write and read the values of variables to disk files.
When data are written to or read from a file it is crucially important that a correct data format is used. The data format is the key to interpreting the contents of a file and must be known in order to correctly interpret the data in an input file. There a two types of data files: formatted and unformatted. Formatted data files uses format strings to define exactly how and in what positions of a record the data is stored. Unformatted storage, on the other hand, only specifies the number format.
The files used in this section are available from the web site

http://www.maths.dundee.ac.uk/software/#matlab

Those that are unformatted are in a satisfactory form for the Windows version on Matlab (version 6.1) but not on Version 5.3 under Unix.

Exercise 26.1 *Suppose the numeric data is stored in a file '`table.dat`' in the form of a table, as shown below.*

```
100  2256
200  4564
300  3653
400  6798
500  6432
```

The three commands,

```
>> fid = fopen('table.dat','r');
>> a = fscanf(fid,'%3d%4d');
>> fclose(fid);
```

respectively

1. open a file for reading (this is designated by the string '`r`'). The variable `fid` is assigned a unique integer which identifies the file used (a file identifier). We use this number in all subsequent references to the file.
2. read pairs of numbers from the file with file identifier `fid`, one with 3 digits and one with 4 digits, and
3. close the file with file identifier `fid`.

This produces a **column vector** `a` with elements, `100 2256 200 4564 ...500 6432`. This vector can be converted to 5×2 matrix by the command
`A = reshape(2,2,5)';`.

26.1 Formatted Files

Some computer codes and measurement instruments produce results in formatted data files. In order to read these results into Matlab for further analysis the data format of the files must be known. Formatted files in ASCII format are written to and read from with the commands `fprintf` and `fscanf`.

> `fprintf(fid, 'format', variables)` writes variables an a format specified in string 'format' to the file with identifier fid
>
> `a = fscanf(fid, 'format',size)` assigns to variable `a` data read from file with identifier `fid` under format '`format`'.

Exercise 26.2 *Study the available information and help on `fscanf` and `fprintf` commands. What is the meaning of the format string,* `'%3d\n'`*?*

Example 26.1 *Suppose a sound pressure measurement system produces a record with 512 time – pressure readings stored on a file '`sound.dat`'. Each reading is listed on a separate line according to a data format specified by the string,* `'%8.6f %8.6f'`.

A set of commands reading time – sound pressure data from '`sound.dat`' is,

Step 1: Assign a namestring to a file identifier.

```
>> fid1 = fopen('sound.dat','r');
```

The string 'r' indicates that data is to be read (not written) from the file.

Step 2: Read the data to a vector named 'data' and close the file,

```
>> data = fscanf(fid1, '%f %f');
>> fclose(fid1);
```

Step 3: Partition the data in separate time and sound pressure vectors,

```
>> t = data(1:2:length(data));
>> press = data(2:2:length(data));
```

The pressure signal can be plotted in a `lin-lin` diagram,

```
>> plot(t, press);
```

The result is shown in Figure 10.

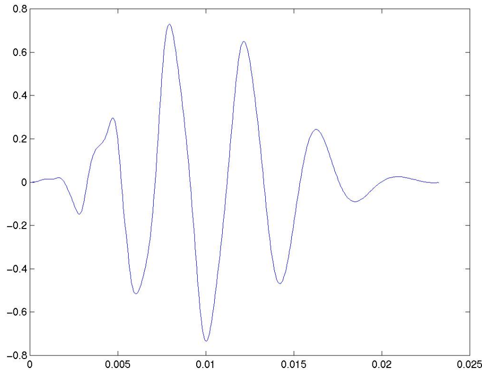


Figure 10: Graph of "sound data" from Example 26.1

26.2 Unformatted Files

Unformatted or binary data files are used when small-sized files are required. In order to interpret an unformatted data file the data precision must be specified. The precision is specified as a string, e.g., '`float32`', controlling the number of bits read for each value and the interpretation of those bits as character, integer or floating point values. Precision '`float32`', for instance, specifies each value in the data to be stored as a floating point number in 32 memory bits.

Example 26.2 *Suppose a system for vibration measurement stores measured acceleration values as floating point numbers using 32 memory bits. The data is stored on file '`vib.dat`. The following commands illustrate how the data may be read into Matlab for analysis.*

Step 1: Assign a file identifier, `fid`, to the string specifying the file name.

```
>> fid = fopen('vib.dat','rb');
```

The string `'rb'` specifies that binary numbers are to be read from the file.

Step 2 Read all data stored on file '`vib.dat`' into a vector vib.

```
>> vib = fread(fid, 'float32');
>> fclose(fid);
>> size(vib)
ans =
      131072
```

The `size(vib)` command determines the size, i.e., the number of rows and columns of the vibration data vector.

In order to plot the vibration signal with a correct time scale, the sampling frequency (the number of instrument readings taken per second) used by the measurement system must be known. In this case it is known to be 24000 Hz so that there is a time interval of 1/24000 seconds between two samples.

Step 3: Create a column vector containing the correct time scale.

```
>> dt = 1/24000;
>> t = dt*(1:length(vib))';
```

Step 4: Plot the vibration signal in a `lin-lin` diagram

```
>> plot(t,vib);
>> title('Vibration signal');
>> xlabel('Time,[s]');
>> ylabel('Acceleration, [m/s^2]');
```

27 Graphic User Interfaces

The efficiency of programs that are used often and by several different people can be improved by simplifying the input and output data management. The use of Graphic User Interfaces (GUI), which provides facilities such as menus, pushbuttons, sliders etc, allow programs to be used without any knowledge of Matlab. They also provides means for efficient data management.

A graphic user interface is a Matlab script file customized for repeated analysis of a specific type of problem. There are two ways to design a graphic user interface. The simplest method is to use a tool especially designed for the purpose. Matlab provides such a tool and it is invoked by typing '`guide`' at the Matlab prompt. Maximum flexibility and control over the programming is, however, obtained by using the basic user interface commands. The following text demonstrates the use of some basic commands.

Example 27.1 *Suppose a sound pressure spectrum is to be plotted in a graph. There are four alternative plot formats; lin-lin, lin-log, log-lin and log-log. The graphic user interface below reads the pressure data stored on a binary file selected by the user, plots it in a lin-lin format as a function of frequency and lets the user switch between the four plot formats.*

We use two m–files. The first (`specplot.m`) is the main driver file which builds the graphics window. It calls the second file (`firstplot.m`) which allows the user to select among the possible `*.bin` files in the current directory.

```
%  File: specplot.m
%
% GUI for plotting a user selected frequency spectrum
% in four alternative plot formats, lin-lin,
% lin-log, log-lin and log-log.
%
% Author: U Carlsson, 2001-08-22

      % Create figure window for graphs
figWindow = figure('Name','Plot alternatives');
      % Create file input selection button
fileinpBtn = uicontrol('Style','pushbutton',...
   'string','File','position',[5,395,40,20],...
   'callback','[fdat,pdat] = firstplot;');
      % Press 'File' calls function 'firstplot'
```

```
% Create pushbuttons for switching between four
% different plot formats. Set up the axis stings.
X = 'Frequency, [Hz]';
Y = 'Pressure amplitude, [Pa]';
linlinBtn = uicontrol('style','pushbutton',...
  'string','lin-lin',...
  'position',[200,395,40,20],'callback',...
  'plot(fdat,pdat);xlabel(X);ylabel(Y);');
linlogBtn = uicontrol('style','pushbutton',...
  'string','lin-log',...
  'position',[240,395,40,20],...
  'callback',...
  'semilogy(fdat,pdat);xlabel(X);ylabel(Y);');
loglinBtn = uicontrol('style','pushbutton',...
  'string','log-lin',...
  'position',[280,395,40,20],...
  'callback',...
  'semilogx(fdat,pdat);xlabel(X);ylabel(Y);');
loglogBtn = uicontrol('style','pushbutton',...
  'string','log-log',...
  'position',[320,395,40,20],...
  'callback',...
  'loglog(fdat,pdat);xlabel(X);   ylabel(Y);');

% Create exit pushbutton with red text.

exitBtn = uicontrol('Style','pushbutton',...
  'string','EXIT','position',[510,395,40,20],...
  'foregroundcolor',[1 0 0],'callback','close;');
```

```
% Script file: firstplot.m
% Brings template for file selection. Reads
% selected filename and path and plots
% spectrum in a  lin-lin diagram.
% Output data are frequency and pressure
% amplitude vectors: 'fdat' and  'pdat'.
% Author: U Carlsson, 2001-08-22

function [fdat,pdat] = firstplot

% Call Matlab function 'uigetfile' that
% brings file selction template.

[filename,pathname] = uigetfile('*.bin',...
  'Select binary data-file:');
% Change directory
cd(pathname);
% Open file for reading binary floating
% point numbers.
fid = fopen(filename,'rb');
data = fread(fid,'float32');
% Close file
fclose(fid);
% Partition data vector in frequency and
% pressure vectors
pdat = data(2:2:length(data));
fdat = data(1:2:length(data));
% Plot pressure signal in a lin-lin diagram
plot(fdat,pdat);
% Define suitable axis labels
xlabel('Frequency, [Hz]');
ylabel('Pressure amplitude, [Pa]');
```

Executing this GUI from the command line (`>> specplot`) brings the following screen.

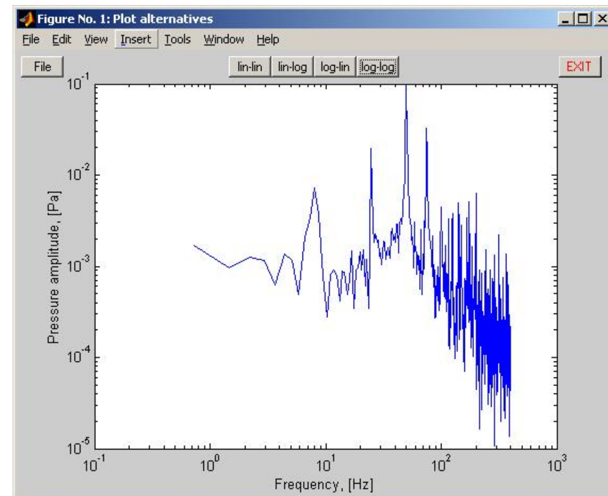


Figure 11: Graph of “vibration data” from Example 27.1

Example 27.1 illustrates how the '`callback`' property allows the programmer to define what actions should result when buttons are pushed etc. These actions may consist of single Matlab commands or complicated sequences of operations defined in various subroutines.

Exercise 27.1 *Five different sound recordings are stored on binary data files, `sound1.bin`, `sound2.bin`, ..., `sound5.bin`. The storage precision is '`float32`' and the sounds are recorded with sample frequency 12000 Hz.*
Write a graphic user interface that, opens an interface window and

- *lets the user select one of the five sounds,*
- *plots the selected sound pressure signal as a function of time in a lin-lin diagram,*
- *lets the user listen to the sound by pushing a 'SOUND' button and finally*
- *closes the session by pressing a 'CLOSE' button.*

28 Command Summary

The command
`>> help`
will give a list of categories for which help is available (e.g. `matlab/general` covers the topics listed in Table 3. Further information regarding the commands listed in this section may then be obtained by using:
`>> help` *`topic`*
try, for example,
`>> help help`

`abs`	Absolute value
`sqrt`	Square root function
`sign`	Signum function
`conj`	Conjugate of a complex number
`imag`	Imaginary part of a complex number
`real`	Real part of a complex number
`angle`	Phase angle of a complex number
`cos`	Cosine function
`sin`	Sine function
`tan`	Tangent function
`exp`	Exponential function
`log`	Natural logarithm
`log10`	Logarithm base 10
`cosh`	Hyperbolic cosine function
`sinh`	Hyperbolic sine function
`tanh`	Hyperbolic tangent function
`acos`	Inverse cosine
`acosh`	Inverse hyperbolic cosine
`asin`	Inverse sine
`asinh`	Inverse hyperbolic sine
`atan`	Inverse tan
`atan2`	Two–argument form of inverse tan
`atanh`	Inverse hyperbolic tan
`round`	Round to nearest integer
`floor`	Round towards minus infinity
`fix`	Round towards zero
`ceil`	Round towards plus infinity
`rem`	Remainder after division

Table 2: Elementary Functions

Managing commands and functions.	
`help`	On-line documentation.
`doc`	Load hypertext documentation.
`what`	Directory listing of M-, MAT- and MEX-files.
`type`	List M-file.
`lookfor`	Keyword search through the HELP entries.
`which`	Locate functions and files.
`demo`	Run demos.
Managing variables and the workspace.	
`who`	List current variables.
`whos`	List current variables, long form.
`load`	Retrieve variables from disk.
`save`	Save workspace variables to disk.
`clear`	Clear variables and functions from memory.
`size`	Size of matrix.
`length`	Length of vector.
`disp`	Display matrix or text.
Working with files and the operating system.	
`cd`	Change current working directory.
`dir`	Directory listing.
`delete`	Delete file.
`!`	Execute operating system command.
`unix`	Execute operating system command & return result.
`diary`	Save text of MATLAB session.
Controlling the command window.	
`cedit`	Set command line edit/recall facility parameters.
`clc`	Clear command window.
`home`	Send cursor home.
`format`	Set output format.
`echo`	Echo commands inside script files.
`more`	Control paged output in command window.
Quitting from MATLAB.	
`quit`	Terminate MATLAB.

Table 3: General purpose commands.

Matrix analysis.	
`cond`	Matrix condition number.
`norm`	Matrix or vector norm.
`rcond`	LINPACK reciprocal condition estimator.
`rank`	Number of linearly independent rows or columns.
`det`	Determinant.
`trace`	Sum of diagonal elements.
`null`	Null space.
`orth`	Orthogonalization.
`rref`	Reduced row echelon form.
Linear equations.	
\ and `/`	Linear equation solution; use "help slash".
`chol`	Cholesky factorization.
`lu`	Factors from Gaussian elimination.
`inv`	Matrix inverse.
`qr`	Orthogonal- triangular decomposition.
`qrdelete`	Delete a column from the QR factorization.
`qrinsert`	Insert a column in the QR factorization.
`nnls`	Non–negative least- squares.
`pinv`	Pseudoinverse.
`lscov`	Least squares in the presence of known covariance.
Eigenvalues and singular values.	
`eig`	Eigenvalues and eigenvectors.
`poly`	Characteristic polynomial.
`polyeig`	Polynomial eigenvalue problem.
`hess`	Hessenberg form.
`qz`	Generalized eigenvalues.
`rsf2csf`	Real block diagonal form to complex diagonal form.
`cdf2rdf`	Complex diagonal form to real block diagonal form.
`schur`	Schur decomposition.
`balance`	Diagonal scaling to improve eigenvalue accuracy.
`svd`	Singular value decomposition.
Matrix functions.	
`expm`	Matrix exponential.
`expm1`	M- file implementation of expm.
`expm2`	Matrix exponential via Taylor series.
`expm3`	Matrix exponential via eigenvalues and eigenvectors.
`logm`	Matrix logarithm.
`sqrtm`	Matrix square root.
`funm`	Evaluate general matrix function.

Table 4: Matrix functions—numerical linear algebra.

Graphics & plotting.	
`figure`	Create Figure (graph window).
`clf`	Clear current figure.
`close`	Close figure.
`subplot`	Create axes in tiled positions.
`axis`	Control axis scaling and appearance.
`hold`	Hold current graph.
`figure`	Create figure window.
`text`	Create text.
`print`	Save graph to file.
`plot`	Linear plot.
`loglog`	Log-log scale plot.
`semilogx`	Semi-log scale plot.
`semilogy`	Semi-log scale plot.
Specialized X-Y graphs.	
`polar`	Polar coordinate plot.
`bar`	Bar graph.
`stem`	Discrete sequence or "stem" plot.
`stairs`	Stairstep plot.
`errorbar`	Error bar plot.
`hist`	Histogram plot.
`rose`	Angle histogram plot.
`compass`	Compass plot.
`feather`	Feather plot.
`fplot`	Plot function.
`comet`	Comet-like trajectory.
Graph annotation.	
`title`	Graph title.
`xlabel`	X-axis label.
`ylabel`	Y-axis label.
`text`	Text annotation.
`gtext`	Mouse placement of text.
`grid`	Grid lines.
`contour`	Contour plot.
`mesh`	3-D mesh surface.
`surf`	3-D shaded surface.
`waterfall`	Waterfall plot.
`view`	3-D graph viewpoint specification.
`zlabel`	Z-axis label for 3-D plots.
`gtext`	Mouse placement of text.
`grid`	Grid lines.

Table 5: Graphics & plot commands.