CM10197: Analytical Methods for Applications

John Power

November 30, 2008

Introduction

There are several aspects of computer science that you need to master through the course of your studies. One of the primary themes of the subject is the modelling of the physical universe, and you need to understand it.

The idea is that you use a computer to simulate, or in some other way model, real activity such as movement in space. For instance, that is essential to any kind of animation. Even a restricted version of it, in which one has no movement, is essential to digital photography. So we shall use photography as a motivating example for much of the course, although the application of the work in the course extends well beyond photography to a remarkable diversity of computational settings. Much of what we study here will be used in your Graphics courses next year; and more will be used for those of you who study Computer Algebra in later years.

The issues surrounding photography are considerably more subtle than one might first imagine, involving sophisticated mathematics. Consider a person walking in a circle, and imagine photographing the person from directly above. From the photographer's perspective, where is the person's right-hand side? If the person is walking counter-clockwise, the person's right-hand side changes from being on the right of the picture, to being at the top of the picture, to being on the left of the picture, then at the bottom of the picture, then back to being on the right of the picture. The point here is that the location of the right-hand side of the person may be quite complex relative to the position of the camera even when one considers simple movements.

The situation becomes all the more complicated when one has more than one person in the frame, or when one allows three-dimensional movement, or when one has a more normal positioning of the camera than directly above the action. It has proved both helpful and convenient to use mathematical formalism to describe the situation, and it is not only convenient but necessary when one seeks to do animation.

A fundamental mathematical tool in this regard is linear algebra, which is essentially the mathematics of vectors and matrices. Linear algebra helps us to make precise the relationship between the perspectives of the camera and the people or objects within the frame, and the movement of both the camera and the people or objects within the frame. About half of this course will be devoted to linear algebra, with particular emphasis on two and three dimensions.

A second fundamental mathematical tool in this regard is calculus, which means differentiation and integration. For instance, in animation, the use of calculus allow us to ensure that movement is smooth rather than jerky, as it allows us to consider the speed of an action. Ultimately, in later courses, you will use quite complex calculus, more complex than we shall study here, as you will need to measure both speed and direction as a character walks in a circle, or perhaps a spiral, with varying speeds. So the other half of the course will be devoted to calculus.

The use you will make, in later courses, of these mathematical techniques will involve computational support. One of the most popular and useful tools in that regard, and one that is used prominently at this university, is MATLAB. There will be no examination question in this course that requires the use of MATLAB, but it will appear in some of your assignments and this would be a sensible time at which to familiarise yourselves with it. The reason we mention MATLAB here is to relate the mathematics of the course with computational practice, and to help to prepare you for your later courses.

Regarding texts, there are many fine texts on Linear Algebra. One good text available in the university library is

> Stanley I. Grossman, *Elementary Linear Algebra (5th ed)*, Saunders College Publishing, 1994, ISBN 0 03 097354 6

For calculus, the first nine chapters of the book

> G Stephenson, *Mathematical Methods for Science Students (2nd ed)*, Longman, 1983, ISBN 0 582 44416 0

also available in the university library, are very good, a little too hard for this course but in a direction that will suit you in future years. A recommended MATLAB book available in the university library is

> Brian D. Hahn, *Essential MATLAB for Engineers and Scientists*, Arnold, 1997, ISBN 0340691441

There are several other MATLAB books in the library: anything with a title including a work like mathematicians or scientists is likely to be worth a look to see whether it appeals to you.

Please note that these coursenotes do *not replace* the lectures. In particular, you will see very few worked examples in the coursenotes: we will work through many examples in lectures, largely taken from the exercises here. Also, you will see no curves plotted in the coursenotes: we will plot many curves on the blackboard; you can do them yourselves with MATLAB in your laboratory sessions; and you could consult other references for them, such as those listed above. I hope to place one example on Moodle.

1 LINEAR ALGEBRA

1.1 Matrices and Vectors

Definition 1.1.1 *Given natural numbers* m *and* n*, a* real $m \times n$ matrix *consists of a rectangular array of* $m \times n$ *real numbers, arranged in* m *rows and* n *columns as follows:*

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1j} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ij} & \cdots & a_{in} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mj} & \cdots & a_{mn} \end{pmatrix}$$

When we refer to the ij-th component of a matrix A as above, we mean a_{ij}, the real number that appears in the i-th row and j-th column of A. If we are simply told "A is a matrix", we often write A_{ij} to represent the ij-th component of A.

There are important special sorts of matrices.

Definition 1.1.2 *A real* $1 \times n$ *matrix* $(v_1, v_2, \cdots, v_n)$ *is called a* real n-dimensional row vector*.*

Definition 1.1.3 *A real* $m \times 1$ *matrix is called a* real m-dimensional column vector*.*

In this course, when we say *vector*, we shall usually mean a real m-dimensional column vector for some m, which we shall typically write as

$$v = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{pmatrix}$$

The main exception to that is in dimensions two and three, for which vectors are often written horizontally, e.g., as (a_1, a_2, a_3).

So, formally, vectors are special sorts of matrices. But the spirit of the two notions is quite different, as we shall see through the course. One can regard an $m \times n$ matrix as a list of m n-dimensional row vectors, typically denoting the i-th row by R_i, or alternatively as a list of n m-dimensional column vectors, typically denoting the j-th column by C_j.

We shall usually denote the particular n-dimensional column vector that has 1 in the i-th position and 0 everywhere else by ι_i. For example,

$$\iota_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Another important special sort of matrix is given by square matrices: those are matrices for which $m = n$, i.e., for which the number of columns is the same as the number of rows.

One can consider variants of the definition of real $m \times n$ matrix. For instance, one might allow the real numbers in the definition of real $m \times n$ matrix to be replaced by complex numbers. In that case, we would refer to a complex $m \times n$ matrix.

Exercise 1.1.4 *See how to use MATLAB to describe matrices. Then use it to describe some 2×2-matrices and 2-dimensional vectors, and see whether you can use MATLAB as a visual aide.*

1.2 Operations on Matrices

There are several basic operations and constructions one can make with matrices. A simple and important one is the sum.

Definition 1.2.1 *Given $m \times n$ matrices A and B, the* sum *of A and B, typically written as $A + B$, is given pointwise, i.e., $(A + B)_{ij} = A_{ij} + B_{ij}$.*

Spelling out the definition in the case of $m = n = 2$, the sum is given as follows:

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{pmatrix}$$

Proposition 1.2.2 *The sum of matrices is associative, commutative, has a unit, and has inverses, i.e.,*

- *for any $m \times n$ matrices A, B, and C,*

$$(A + B) + C = A + (B + C)$$

- *for any $m \times n$ matrices A and B,*

$$A + B = B + A$$

- *for any $m \times n$ matrix A,*

$$A = 0 + A = A + 0$$

  where, by mild overloading of notation, 0 denotes the $m \times n$ matrix for which every component is the real number 0

- *for any $m \times n$ matrix A, there is a matrix we denote by $-A$ for which*

$$A + (-A) = 0 = (-A) + A$$

Exercise 1.2.3 *Try to prove part of Proposition 1.2.2. Does Proposition 1.2.2 remind you of a definition given in a previous course?*

Definition 1.2.4 *Given an $m \times n$ matrix A and a real number r,* scalar multiplication *of A by r is given by the matrix whose ij-th component is rA_{ij}.*

Spelling out the definition in the case of $m = n = 2$, scalar multiplication is given as follows:

$$r \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} ra_{11} & ra_{12} \\ ra_{21} & ra_{22} \end{pmatrix}$$

Proposition 1.2.5 *Scalar multiplication is associative, has a unit, and respects the sum of matrices, i.e.,*

- *for any $m \times n$ matrix A and real numbers r and s,*

$$(rs)A = r(sA)$$

- *for any $m \times n$ matrix A,*

$$1A = A$$

- *for any $m \times n$ matrices A and B and any real number r,*

$$r(A + B) = rA + rB$$

Exercise 1.2.6 *Try to prove part of Proposition 1.2.5. Does this remind you of a previous course?*

Definition 1.2.7 *Given an $m \times n$ matrix A, the* transpose *of A, often written A_t, is the $n \times m$ matrix given by swapping the rows and columns of A.*

Spelling this out where $m = 2$ and $n = 3$, the transpose of the matrix

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}$$

is the matrix

$$\begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \\ a_{13} & a_{23} \end{pmatrix}$$

The transpose can be used to help to calculate inverses to square matrices if they exist.

Exercise 1.2.8 *Try to formulate and prove a proposition about transpose along the lines of Propositions 1.2.2 and 1.2.5, enunciating basic facts about the transpose and its interaction with sum and scalar multiplication.*

Exercise 1.2.9 *Let* $A = \begin{pmatrix} 1 & 3 \\ 2 & 5 \\ -1 & 2 \end{pmatrix}$, $B = \begin{pmatrix} -2 & 0 \\ 1 & 4 \\ -7 & 5 \end{pmatrix}$, *and* $C = \begin{pmatrix} -1 & 1 \\ 4 & 6 \\ -7 & 3 \end{pmatrix}$.

Calculate (and show your working)

1. $3A$
2. $A - C$
3. $2A - 3B + 4C$

Now find a matrix D for which $2A + B - D$ is the 3×2 zero matrix.

1.3 Matrix Multiplication

The most interesting and complex basic operation on matrices is the multiplication of matrices. The reason it is interesting is because it is not given component-wise, but rather involves a more complex interaction between a pair of matrices. Why do you think anyone might care about the multiplication operation we are about to define? An answer effectively appears in the next section, but here we just note the definition.

Definition 1.3.1 *Given an $m \times n$ matrix A and an $n \times p$ matrix B, the matrix AB is the $m \times p$ matrix defined as follows:*

$$(AB)_{ij} = A_{i1}B_{1j} + A_{i2}B_{2j} + \cdots + A_{in}B_{nj}$$

For example, given a 3×2 matrix

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix}$$

and a 2×2 matrix

$$\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$$

the composite is the 3×2 matrix given as follows:

$$\begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \\ a_{31}b_{11} + a_{32}b_{21} & a_{31}b_{12} + a_{32}b_{22} \end{pmatrix}$$

Proposition 1.3.2 *Matrix multiplication is associative, with left and right identities, and it respects sum and scalar multiplication, i.e.,*

- *for any matrices A, B and C for which multiplication exists,*

$$(AB)C = A(BC)$$

- *for any natural number n, there is a matrix I_n such that, for any $m \times n$ matrix A and any $n \times p$ matrix B*

$$AI_n = A \qquad I_nB = B$$

- *for any matrices A, B and C for which the various multiplications exist,*

$$(A + B)C = AC + BC \qquad\qquad C(A + B) = CA + CB$$

- *for any matrices A and B for which the multiplication exists, and for any real number r,*

$$r(AB) = (rA)B = A(rB)$$

Exercise 1.3.3 *Try to prove part of Proposition 1.3.2. This might remind you of a definition you have seen before, but look a little different: here, we need to keep asking whether a multiplication exists or not. That concern motivates the definition of category, which those of you who study theoretical options might see in future years.*

Exercise 1.3.4 *Is it true that for every pair of* 2×2 *matrices* A *and* B*, that* $AB = BA$*? If so, can you give a proof of it? And if not, can you give a counter-example, i.e., a pair of such matrices for which it is not true?*

Exercise 1.3.5 *Try to find out what* dynamic programming *is, e.g., by doing a web-search or visiting the library, and find out why a computer scientist might care about it? and why you think I have mentioned it just after defining matrix multiplication?*

I again ask why would anyone be interested in the multiplication of matrices? The answer comes in the next section. For the moment, observe that matrix multiplication allows us to give an uncluttered expression of simultaneous equations. For instance, in two variables, the simultaneous equations

$$a_1x + b_1y = c_1$$

$$a_2x + b_2y = c_2$$

are expressible as the matrix equation

$$\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

If one has three characters in the frame of a camera shot, each occupying a three-dimensional position, moving with a three-dimensional velocity, one would have simultaneous equations involving 18 variables. And three is not many people to have in one camera shot. So the efficiency of the matrix notation is of considerable value. The constructions one needs to make with matrices in succeeding sections make it all the more useful as variables, especially their repetition, clutter the presentation.

Exercise 1.3.6 *Calculate the following (and show your working):*

1. $\begin{pmatrix} 2 & 3 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 4 & 1 \\ 0 & 6 \end{pmatrix}$
2. $\begin{pmatrix} 1 & 6 \\ 0 & 4 \\ -2 & 3 \end{pmatrix} \begin{pmatrix} 7 & 1 & 4 \\ 2 & -3 & 5 \end{pmatrix}$
3. $\begin{pmatrix} -4 & 5 & 1 \\ 0 & 4 & 2 \end{pmatrix} \begin{pmatrix} 3 & -1 & 1 \\ 5 & 6 & 4 \\ 0 & 1 & 2 \end{pmatrix}$

Exercise 1.3.7 *Are* $\begin{pmatrix} 1 \\ 4 \\ -7 \end{pmatrix}$ *and* $\begin{pmatrix} 2 \\ 3 \\ 2 \end{pmatrix}$ *orthogonal? Show your working.*

1.4 Linear Transformations

Matrices allow us to characterise, in computationally accessible terms, fundamental ways in which we can move points uniformly about two- or three-dimensional space, while keeping the origin, which you might regard as the position of a putative camera, fixed. In particular, they include rotations in any axis, dilations along any axis or combination of axes, and projections, as, for instance, one uses in taking a photograph as one makes a two-dimensional image of a three-dimensional scene. In this section, we characterise those functions that matrices represent.

Definition 1.4.1 *Every* $m \times n$ *matrix* A *determines a function* $L_A : R^n \longrightarrow R^m$ *as follows:*

$$L_A \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1j} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mj} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

i.e., the function L_A *is defined by the matrix multiplication of the* $m \times n$ *matrix* A *with an element of* R^n *seen as an* n*-dimensional column vector.*

Example 1.4.2 *For any real number* r*, consider the matrix*

$$\begin{pmatrix} r & 0 \\ 0 & 1 \end{pmatrix}$$

Premultiplying by this matrix sends the vector $\begin{pmatrix} x \\ y \end{pmatrix}$ *to* $\begin{pmatrix} rx \\ y \end{pmatrix}$*. So, the matrix dilates the* X *axis of the plane by a factor of* r*. One can do similarly for the* Y *axis. And one can generalise to arbitrary dimensions.*

Example 1.4.3 *For any real number* θ*, consider the matrix*

$$\begin{pmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{pmatrix}$$

This is sometimes denoted by R_θ *because pre-multiplication by it rotates a vector by the angle* θ *counter-clockwise around the origin.*

Example 1.4.4 *For any real number* r*, consider the matrix*

$$\begin{pmatrix} r & 0 & 0 \\ 0 & r & 0 \\ 0 & 0 & r \end{pmatrix}$$

Pre-multiplication by this matrix sends a point $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ *in space to a point in the same direction from the origin but at* r *times the length of* $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$. *If you consider a pin-hole camera situated at the origin, photographing a two-dimensional scene, with the image to be formed at distance a factor of* r *from the origin, this linear transformation becomes useful. This amounts to perspective.*

In Section 1.2, we defined the sum and scalar multiplication of matrices. One can do likewise for functions between sets of the form R^n, as follows:

Definition 1.4.5 *Given functions* $f, g : R^n \longrightarrow R^m$, *their* sum *is the function that sends* v *in* R^n *to the element* $f(v) + g(v)$ *of* R^m.

Definition 1.4.6 *Given a function* $f : R^n \longrightarrow R^m$ *and a real number* r, *the* scalar multiplication *of* r *with* f *is defined to be the function that sends* v *in* R^n *to the element* $rf(v)$ *of* R^m.

Using Definition 1.4.1, we can compare the sum and scalar multiplication of matrices with the sum and scalar multiplication of functions. The result is as follows:

Proposition 1.4.7 *For every pair* A *and* B *of* $m \times n$ *matrices,*

$$L_{A+B} = L_A + L_B$$

And for every matrix A *and real number* r,

$$L_{rA} = rL_A$$

Exercise 1.4.8 *Prove Proposition 1.4.7.*

And now for the piece de resistance:

Theorem 1.4.9 *For every* $m \times n$ *matrix* A *and every* $n \times p$ *matrix* B,

$$L_A L_B = L_{AB}$$

The theorem tells us that matrix multiplication yields exactly composition of functions, which is an eminently natural way in which one combines the movements about the plane or about space: one first makes one move, then one makes another.

Exercise 1.4.10 *Prove Theorem 1.4.9.*

One can characterise the functions from R^n to R^m that arise from matrices as follows.

Definition 1.4.11 *A function $T : R^n \longrightarrow R^m$ is* linear *if*

1. *T(v+v') = T(v) + T(v')*
2. *rT(v) = T(rv)*

for all vectors v and v' and all real numbers r.

Proposition 1.4.12 *For every $m \times n$ matrix A, the function $L_A : R^n \longrightarrow R^m$ is linear.*

Proof This is an immediate consequence of Proposition 1.3.2 and the fact that an element of R^n, i.e., an n-dimensional vector, is a particular sort of matrix.

The linear transformations are a restrictive, but very useful, class of functions.

Exercise 1.4.13 *Prove that if T is a linear transformation, then $T(0) = 0$. (Hint: Use the first condition in the definition of linear transformation, remember that $0 + 0 = 0$, and use subtraction.)*

Exercise 1.4.14 *Consider Examples 1.4.2 and 1.4.3. Consider the linear transformations generated by the two classes of matrices. Can you describe their behaviour? Can you use MATLAB to visualise them?*

Recall that, in Section 1.1 that we defined the vector ι_i to be the vector with 1 in the i-th position and 0 everywhere else. There is a precise sense in which such vectors generate all vectors in a way that is consistent with our definition of linear transformation.

Definition 1.4.15 *Let $v_1, v_2, \cdots, v_k$ be a list of vectors. Then any vector of the form*

$$r_1 v_1 + r_2 v_2 + \cdots r_k v_k$$

is called a linear combination *of $v_1, v_2, \cdots, v_k$.*

Exercise 1.4.16 *Prove that any n-dimensional vector is uniquely expressible as a linear combination of $\iota_1, \iota_2, \cdots, \iota_n$.*

Proposition 1.4.17 *Every linear transformation $T : R^n \longrightarrow R^m$ is determined by its value on each ι_i.*

Proof Every vector

$$v = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

is uniquely expressible as a linear combination of ι_i's, the linear combination being given by

$$v = v_1 \iota_1 + v_2 \iota_2 + \cdots v_n \iota_n$$

Since T is linear, it follows that

$$T(v) = v_1 T(\iota_1) + v_2 T(\iota_2) + \cdots v_n T(\iota_n)$$

Theorem 1.4.18 *A function $T : R^n \longrightarrow R^m$ is linear if and only if there is an $m \times n$ matrix A for which $T = L_A$.*

Proof We need to show that for any linear function T, there exists a matrix A_T such that $T = L_{A_T}$. For any matrix A, consider the matrix multiplication $A\iota_i$. The result is exactly the i-th column of A. So if we *define* A_T to be given by treating the sequence of m-dimensional column vectors $T(\iota_1), T(\iota_2), \cdots, T(\iota_n)$ as an $m \times n$ matrix A_T, it would follow that for each i, we would have $A_T \iota_i = T(\iota_i)$. But, since both T and every L_B are linear, it follows that $T = L_{T_A}$.

Exercise 1.4.19 *Recall Exercise 1.3.4. asking about commutativity of matrix multiplication. Now that we know that matrices are equivalent to linear transformations, we can reformulate the question: is it true that for every pair of linear transformations T and S from R^2 to itself, the composition of functions TS agrees with ST? If so, give a proof, and if not, give a counter-example.*

We need two final definitions to complete our analysis here.

Definition 1.4.20 *Given a linear transformation $T : R^n \longrightarrow R^m$,*

- *the* kernel *of T, denoted $kerT$, is the set of those $v \epsilon R^n$ such that $T(v) = 0$.*
- *the* range *of T, denoted $RangeT$, is the set of those m-dimensional vectors w for which there exists $v \epsilon R^n$ for which $T(v) = w$.*

You would be familiar with the notion of range from your previous course. So we shall focus on the kernel.

Exercise 1.4.21 *Consider the function from R^3 to R^2 that sends (x, y) to x, i.e., the first projection. Check that it is a linear transformation. What is its kernel? and what is its range? Can you visualise this using MATLAB? Can you imagine why someone might be interested in this in regard to photography?*

Exercise 1.4.22 *Are the following functions linear? Show your working.*

1. *the function $T : R^3 \longrightarrow R^2$ that sends $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ to $\begin{pmatrix} 1 \\ z \end{pmatrix}$*

2. *the function $T : R \longrightarrow R^3$ that sends x to $\begin{pmatrix} x \\ x \\ x \end{pmatrix}$*

Exercise 1.4.23 *Let A_θ denote the matrix*

$$\begin{pmatrix} cos\theta & -sin\theta & 0 \\ sin\theta & cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Describe geometrically the linear transformation given by $L_{A_\theta} : R^3 \longrightarrow R^3$.

1.5 Gaussian Elimination

You know from your previous courses that not all functions are invertible. In particular, not all linear transformations are invertible.

Example 1.5.1 *Consider the matrix*

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

Consider the function $L_A : R^2 \longrightarrow R^2$. It acts as the first projection, which is not invertible.

MATLAB will tell you whether a given square matrix is invertible, but how does it do it? It needs an algorithm, the standard algorithm being given by *Gaussian elimination* or a variant. Gaussian elimination has broader uses too that we shall investigate in later sections. It works as follows.

Definition 1.5.2 *Given a real $m \times n$ matrix A, the* elementary row operations *consist of the following three sorts of operations:*

1. *multiplying one row by a non-zero real number $R_i \mapsto rR_i$*
2. *adding a multiple of one row to another row $R_i \mapsto R_i + rR_j$*
3. *interchanging two rows $R_i \leftrightarrow R_j$*

Definition 1.5.3 *A matrix A is in* reduced row echelon form *if the following four conditions all hold:*

1. *all rows (if any) consisting entirely of zeros appear at the bottom of the matrix*
2. *the first non-zero number (starting from the left) in any row not consisting entirely of zeros is* 1
3. *if two successive rows do not consist entirely of zeros, then the first* 1 *in the lower row occurs further to the right than the first* 1 *in the higher row*
4. *any column containing the first* 1 *in a row has zeros everywhere else.*

A matrix is in row echelon form *if it satisfies the first three conditions.*

Gaussian elimination amounts to the systematic application of the elementary row operations to convert an arbitrary matrix A into one in row echelon form, and Gauss-Jordan elimination amounts to the systematic application of the elementary row operations to convert a matrix into one in reduced row echelon form. Both forms are important, for somewhat different reasons.

Theorem 1.5.4 *Any matrix A can be reduced to a matrix in reduced row echelon form by the elementary row operations.*

Spelling out Gaussian elimination applied to a matrix A

1. If C_1, the first column (from the left), consists only of 0's, move to C_2. Proceed inductively.
2. Suppose R_i is the first row for which C_1 has a non-zero entry. If $i \neq 1$, apply the third operation to exchange R_i with R_1. Then apply the first operation to force the first entry in (what is then) R_1 to become 1.
3. Use the second operation applied to each R_i together with R_1 to force the first entry of R_i to become 0.
4. Proceed inductively.

This is most easily seen by examples, which we shall discuss.

For Gauss-Jordan elimination, first do Gaussian elimination as above. Then systematically use the second operation to reduce to 0 all non-zero entries lying above the first 1 in any row, starting on the left. Again, this is most easily seen by examples.

Exercise 1.5.5 *Use Gaussian elimination to reduce the following matrices to row echelon form:*

$$\begin{pmatrix} 2 & -4 & 8 \\ 3 & 5 & 8 \\ -6 & 0 & 4 \end{pmatrix} \qquad \begin{pmatrix} 2 & -7 \\ 3 & 5 \\ 4 & -3 \end{pmatrix} \qquad \begin{pmatrix} 0 & 2 & 3 \\ 2 & -6 & 7 \\ 1 & -2 & 5 \end{pmatrix}$$

Now apply (the rest of) Gauss-Jordan elimination to reduce the two matrices to reduced row echelon form.

1.6 Inverses

Definition 1.6.1 *An $n \times n$ matrix A has an* inverse *if there exists a matrix A^{-1} for which $AA^{-1} = I_n = A^{-1}A$.*

Exercise 1.6.2 *Prove that an inverse is necessarily unique: but first, try to understand exactly what this statement means.*

Exercise 1.6.3 *Prove that a matrix A has an inverse if and only if the linear transformation $L_A : R^n \longrightarrow R^n$ is invertible. (Hint: the hard part of this is to prove that if L_A is invertible, so is A. In order to see that, suppose L_A has an inverse f, then prove that f is linear, then deduce that f must be of the form L_B for some matrix B, and finally deduce that B must be an inverse of A.*

Proposition 1.6.4 *If A and B are invertible $n \times n$ matrices,*

$$(AB)^{-1} = B^{-1}A^{-1}$$

One can use Gauss-Jordan elimination to give an algorithm to check whether a given matrix is invertible, and if so, to calculate the inverse. The algorithm is as follows:

1. write the $n \times 2n$ matrix given by writing I_n immediately to the right of A
2. use row reduction to reduce the matrix A to reduced row echelon form, applying the same operations to I_n.

The matrix A is invertible if and only if the reduced row echelon form of A is exactly I_n, and if A is invertible, A^{-1} is given by the matrix to which I_n is reduced by the above algorithm.

Theorem 1.6.5 *An $n \times n$ matrix A is invertible if and only if, for any n-dimensional vector b, the matrix equation $Ax = b$ has a unique solution.*

Proof If A is invertible, the one and only solution is given by $x = A^{-1}b$. Conversely, the unique solution of $Ax = \iota_k$ determines the k-th column of A^{-1}.

An invertible linear transformation amounts to a change of basis, as we shall discuss later.

Exercise 1.6.6 *Use Gauss-Jordan elimination to check whether the following matrices are invertible, and if so, to calculate an inverse for them:*

$$\begin{pmatrix} 3 & 2 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & -1 \end{pmatrix} \qquad \begin{pmatrix} 1 & 6 & 2 \\ -2 & 3 & 5 \\ 7 & 12 & -4 \end{pmatrix} \qquad \begin{pmatrix} 2 & 0 & 4 \\ 0 & 1 & 1 \\ 3 & -1 & 1 \end{pmatrix} \qquad \begin{pmatrix} 1 & 2 & 3 \\ 1 & 1 & 2 \\ 0 & 1 & 2 \end{pmatrix}$$

1.7 Determinants

Every square matrix has a determinant. But historically, the notion of determinant preceded that of matrix. Their central practical use for us is that it is easier to calculate a determinant than it is to calculate the inverse of a matrix, and one can tell from the determinant whether a matrix has an inverse. So they provide a fast way to check whether a matrix is invertible, although not actually giving the inverse.

One easily calculable definition by induction is as follows:

Notation 1.7.1 *Given a square n-dimensional matrix A, let M_{ij} denote the the square $(n-1)$-dimensional matrix determined by deleting the i-th row and j-th column of A.*

Definition 1.7.2 *We define the* determinant *of a square n-dimensional matrix by induction on n. It is sometimes denoted by $det(A)$ and sometimes by $|A|$. It is defined as follows:*

- *if $n = 1$, $det(A) = a_{11}$, the single entry of A.*
- *if $n > 1$, define*

$$det(A) = a_{11}det(M_{11}) - a_{12}det(M_{12}) + \cdots + a_{1n}(-1)^{1+n}(M_{1n})$$

Two fundamental facts about determinants are as follows:

Theorem 1.7.3 *1. If A and B are square n-dimensional matrices,*

$$det(AB) = (detA)(detB)$$

2. A is invertible if and only if $det(A) \neq 0$.

Exercise 1.7.4 *Is $det(A + B) = detA + detB$ in general. If so, give a proof, and if not, give a counter-example.*

Exercise 1.7.5 *Calculate the determinants of the following matrices, showing your working:*

$$\begin{pmatrix} 1 & -1 & 2 & 4 \\ 0 & -3 & 5 & 6 \\ 1 & 4 & 0 & 3 \\ 0 & 5 & -6 & 7 \end{pmatrix} \qquad \begin{pmatrix} -2 & 0 & 0 & 7 \\ 1 & 2 & -1 & 4 \\ 3 & 0 & -1 & 5 \\ 4 & 2 & 3 & 0 \end{pmatrix} \qquad \begin{pmatrix} 1 & 1 & -1 & 0 \\ -3 & 4 & 6 & 0 \\ 2 & 5 & -1 & 3 \\ 4 & 0 & 3 & 0 \end{pmatrix}$$

1.8 Lines and Planes in 3-space

We have considered matrices and vectors almost entirely in terms of abstract algebra so far. But it is vital that you come to some understanding of them as they relate to the plane and to space, i.e., in the cases of R^2 and R^3, as that is what one does when taking photographs or composing 2- or 3-dimensional images.

In two and three dimensions, by convention, one typically writes vectors horizontally rather than vertically, as we have done so far. Two fundamental notions are the *dot product* and the *cross product* of vectors. The dot product can usefully be defined in arbitrary dimensions, but the cross product is useful only in three dimensions.

Definition 1.8.1 *Given n-dimensional column vectors u and v, the* dot product*, denoted $u \cdot v$, is defined to be the matrix multiplication $u_t v$, i.e., the matrix multiplication*

$$(u_1, \cdots, u_n) \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$$

Exercise 1.8.2 *Calculate dot products of some simple two-dimensional vectors, e.g., the dot product of $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ with itself, with $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$, or with an arbitrary vector $\begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$.*

In general, for n-dimensional vectors u and v for which neither is a scalar multiple of the other, u and v determine a plane embedded in R^n: you should be able to picture this in the cases of n being either 2 or 3, and you should be able to use MATLAB to visualise it. The dot product $u \cdot v$ can be described in more familiar terms as follows:

$$u \cdot v = |u||v|cos\theta$$

where θ is the angle between u and v. So, two n-dimensional vectors u and v are called *orthogonal* if $u \cdot v = 0$.

Exercise 1.8.3 *Check, in the case of $n = 2$, that a pair of orthogonal vectors, i.e., vectors at right-angles to each other, have dot product* 0*. What about $n = 3$?*

For arbitrary n, one can sensibly define the *magnitude* of an n-dimensional vector by

$$|v| = \sqrt{v \cdot v}$$

What might one mean by the *direction* of a n-dimensional vector? The definition cannot, for arbitrary n, simply be an angle, as one has too many dimensions for the information given by an angle alone to determine everything. In general, direction is given by a vector: it is the unit vector, i.e., the vector of unit length, that is parallel to the given vector and pointing in the same direction. Formally, that is given as follows.

Definition 1.8.4 *The* direction *of a non-zero vector v in R^3 is defined to be the unit vector $u = v/|v|$.*

So the dot product, in arbitrary dimensions, yields the notions of magnitude and direction of a vector, and can be used to check whether a pair of vectors are orthogonal to each other.

In three dimensions, one has the particular notion of the *cross product* of vectors. The definition of cross product only makes sense for R^3, so is an artefact rather than being a natural vector construction. Nevertheless, it is useful and is defined as follows.

Definition 1.8.5 *Given* 3*-dimensional vectors* u *and* v*, their cross product is the vector*

$$\begin{pmatrix} u_2v_3 - u_3v_2 \\ u_3v_1 - u_1v_3 \\ u_1v_2 - u_2v_1 \end{pmatrix}$$

The cross product can be seen as an instance of a mild generalisation of the notion of determinant as follows:

Theorem 1.8.6 *The cross product of* u *and* v *is given by the determinant of*

$$\begin{pmatrix} i & j & k \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{pmatrix}$$

One can prove assorted facts about cross products, but the most useful one is that the cross product $u \times v$ is orthogonal to both u and v.

Theorem 1.8.7 *Given* 3*-dimensional vectors* u *and* v*, if* θ *is the angle between* u *and* v*, then*

$$|u \times v| = |u||v|sin\theta$$

Example 1.8.8 *Take three* 3*-dimensional vectors. Imagine they are the edges of a parallelepiped. Take the dot product of any one with the cross product of the other two. This is a measure of volume of the parallelepiped. It is also the determinant of the matrix built from the three vectors.*

This works in other dimensions by analogy. In higher dimension, one must use *Grassman algebra* since cross products are ad-hoc devices that work only in three dimensions.

To give a line in R^3, it is necessary and sufficient to specify a point on the line and a vector parallel to the line. One can also give equational presentations of a line embedded in 3-space.

Suppose one has a point (x_0, y_0, z_0) that is on the putative line, and a vector (a, b, c) that determines the direction of the line. Then a point (x, y, z) lies on the line if and only if, for some real number r, one has

$$(x, y, z) = (x_0, y_0, z_0) + r(a, b, c)$$

This is called the *vector equation* for the line; the corresponding three simultaneous equations are called *parametric equations* of the line; and, if a, b, and c are all non-zero, the corresponding equations

$$\frac{x - x_0}{a} = \frac{y - y_0}{b} = \frac{z - z_0}{c}$$

are called *symmetric equations* of the line. If any of a, b or c is zero, the relevant symmetric equation is given by stating what is the constant value of x, y or z respectively.

Note that the parametric and symmetric equations for a line are not unique: to see that, consider starting with another point on the line.

Exercise 1.8.9 *To give a plane in R^3 that goes through the origin is equivalent to giving a line in R^3 that goes through the origin. Think about why this is true. It underlies the general phenomenon of* duality.

To give a plane in R^3 is equivalent to giving a point in the plane and a vector $\mathbf{n}$ that is orthogonal to every every vector in the plane. The vector $\mathbf{n}$ is said to be *normal* to the plane. If $P = (x_0, y_0, z_0)$ is a point in the plane, and $\mathbf{n} = (a, b, c)$, then the plane is given by the set of points $Q = (x, y, z)$ satisfying the equation

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0$$

the reason being that the vector from P to Q must be orthogonal to $\mathbf{n}$, and this equation defines orthogonality. Observe that, as x_0, y_0, and z_0 are all fixed real numbers, determined by P, we can reorganise the equation by putting $d = ax_0 + by_0 + cz_0$ and moving it to the right hand side. That induces the following definition.

Definition 1.8.10 *The* standard equation of a plane *is an equation of the form*

$$ax + by + cz = d$$

The set of vectors (x, y, z) satisfying the equation necessarily form a plane, and all planes in R^3 are characterised by such equations.

Exercise 1.8.11 *Try to see the above using MATLAB.*

Exercise 1.8.12 *Calculate the cross product of* $(2, -3, 1)$ *and* $(1, 2, 1)$.

Exercise 1.8.13 *Find a vector equation, parametric equations, and symmetric equations of the line containing the two points* $(2, 1, 3)$ *and* $(1, 2, -1)$.

Exercise 1.8.14 *Find a vector equation, parametric equations, and symmetric equations of the line that contains* $(-1, -2, 5)$ *and is parallel to the vector* $(0, -3, 7)$.

Exercise 1.8.15 *Describe the equations of the planes determined by the following:*

1. $P = (-4, -7, 5)$ *and* $\mathbf{n} = (-3, -4, 1)$
2. *the following three points lie in the plane:* $(-7, 1, 0)$, $(2, -1, 3)$, *and* $(4, 1, 6)$
3. *the following three points lie in the plane:* $(1, 2, -4)$, $(2, 3, 7)$, *and* $(4, -1, 3)$.

1.9 Span, Linear Independence and Basis

Recall the definition of a linear combination of vectors, Definition 1.4.15. We observed that every n-dimensional vector is uniquely expressible as a linear combination of $\iota_1, \iota_2, \cdots, \iota_n$. That is true and helpful, but note that use of it depends upon our particular frame of reference: in three dimensions, it implies we have a clear idea of which directions are right, straight ahead, and up. But if are composing a photograph, the camera's perspective on such directions are not necessarily the same as those of a character within the frame.

The situation can be complex. For instance, some characters, e.g., some sorts of robots, may not be able to look at any angle upwards or downwards, but can only swivel. And if one introduces mirrors, right-hands can become skewed. To support such activity, we need a more subtle understanding of the ways in which perspective can change, and that is provided by a delicate analysis of linear combinations.

Definition 1.9.1 *Given vectors $v_1, v_2, \cdots, v_k$ in R^n, the* span *of $\{v_1, v_2, \cdots, v_k\}$ is the set of linear combinations of $v_1, v_2, \cdots, v_k$.*

Example 1.9.2 *The span of the vectors $\iota_1, \iota_2, \iota_3$ in R^3 is the whole of R^3.*

Exercise 1.9.3 *The span of any two non-zero vectors in R^3 that are not parallel to each other is a plane passing through the origin. First, write an interesting example of two such vectors. Then use MATLAB to visualise the situation. Now try to prove the result. Can you write an equation for the plane?*

Definition 1.9.4 *Given vectors $v_1, v_2, \cdots, v_k$ in R^n, the vectors are* linearly independent *if, whenever $r_1, r_2, \cdots r_k$ are real numbers for which*

$$r_1 v_1 + r_2 v_2 + \cdots r_k v_k = 0$$

it follows that $r_1 = r_2 = \cdots = r_k = 0$.

Example 1.9.5 *The vectors $\iota_1, \iota_2, \iota_3$ in R^3 are linearly independent.*

Exercise 1.9.6 *Prove that* $\begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix}$ *and* $\begin{pmatrix} 2 \\ 5 \\ -3 \end{pmatrix}$ *are linearly independent.*

Theorem 1.9.7 *For any square n-dimensional matrix A, the matrix A is invertible if and only if the columns of A are n linearly independent vectors.*

Exercise 1.9.8 *Try to prove the theorem.*

The theorem gives us an algorithm to check whether n n-dimensional vectors are linearly independent: line up the vectors in n columns, and check whether the resulting matrix is invertible. We do not even need to work that hard: we only need to check whether the determinant of the matrix is non-zero.

Exercise 1.9.9 *Write down two non-zero* 3*-dimensional vectors, and visualise them using MATLAB. Can you see a geometric condition that is equivalent to linear independence. You should see that they are linearly independent if and only if they are not parallel.*

What about three non-zero 3*-dimensional vectors. They are linearly independent if and only if they do not all lie in the same plane.*

Linear independence and span work together as follows.

Theorem 1.9.10 *Any set of n linearly independent vectors in R^n spans R^n.*

Exercise 1.9.11 *Try to prove the theorem. (Hint: given linearly independent vectors $v_1, v_2, \cdots, v_n$, you need to show that for all v, there are real numbers $r_1, r_2, \cdots, r_n$ for which*

$$v = r_1 v_1 + r_2 v_2 + \cdots + r_n v_n$$

Express this as a matrix equation, and use Theorem 1.9.7 to solve the matrix equation.)

The notions of span and linear independence combine to give the notion of basis. In this course, we will not make substantial use of the notion of basis as we have not substantially considered subspaces of a space, which is where the idea becomes interesting. Formally, a basis is a set of linearly independent vectors that span a space: but Theorem 1.9.10 means that the concept does not have substance unless we consider subspaces.

For any invertible matrix A, application of L_A sends the standard basis $\iota_1, \iota_2, \cdots, \iota_n$ to the basis given by the columns of A. So, while an arbitrary linear transformation moved points about space, an invertible linear transformation gives a change of basis, making precise the idea that a character in the frame of a camera can have different notions of right, straight ahead, and up to those of the camera.

There are special sorts of basis, notably orthogonal bases, in which any two different vectors in the basis are orthogonal to each other. And among the orthogonal bases are the orthonormal bases, in which each basis vector is of unit length.

Exercise 1.9.12 *Play with MATLAB using the notion of a basis. Consider some non-orthonormal bases, such as $\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$ and then some bases that are not even orthogonal, such as $\begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}$*

The notion of change of basis is fundamental because of the difference in perspective between a camera and a character within the frame. But we also need to incorporate change of basis into our analysis of linear transformations. That now amounts simply to a multiplication of matrices: the only need for care is the danger of writing a matrix instead of its inverse by mistake.

Exercise 1.9.13 *Check whether the following vectors span R^3 (and show your working):*

$$\begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix} \qquad \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} \qquad \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

Exercise 1.9.14 *Check whether the following vectors are linearly independent (and show your working):*

$$\begin{pmatrix} -3 \\ 4 \\ 2 \end{pmatrix} \qquad \begin{pmatrix} 7 \\ -1 \\ 3 \end{pmatrix} \qquad \begin{pmatrix} 1 \\ 1 \\ 8 \end{pmatrix}$$

Exercise 1.9.15 *Check whether the following vectors are linearly independent (and show your working):*

$$\begin{pmatrix} 1 \\ -2 \\ 1 \\ 1 \end{pmatrix} \qquad \begin{pmatrix} 3 \\ 0 \\ 2 \\ -2 \end{pmatrix} \qquad \begin{pmatrix} 0 \\ 4 \\ -1 \\ 1 \end{pmatrix} \qquad \begin{pmatrix} 5 \\ 0 \\ 3 \\ -1 \end{pmatrix}$$

Exercise 1.9.16 *Find a basis in R^3, i.e., a spanning pair of vectors in R^3, for the set of vectors in the plane*

$$2x - y - z = 0$$

1.10 Homogeneous Forms

Linear transformations allow us to explain rotations around any axis, and dilation in any direction, e.g., when altering the shape of a window on a computer screen. But they do not allow us to analyse translation, i.e., they keep the origin fixed, $T(0) = O$, so one cannot use them directly to model moving a window from one part of the screen to another for example.

Another way of expressing this is by saying that we have studied equations between vectors of the form $y = Ax$, but we have not studied equations between vectors of the form $y = Ax + c$.

We would like to include translation in our analysis, specifically in order to model phenomena such as moving a window. But we would also like to do so without having to go through our entire analysis again, uniformly dealing with a constant c. We can do that by means of homogeneous coordinates.

Compare

$$\begin{pmatrix} a_{11} & a_{12} & c_1 \\ a_{21} & a_{22} & c_2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ 1 \end{pmatrix}$$

with

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

The two vector equations convey essentially the same information. This motivates homogeneous coordinates as follows.

Definition 1.10.1 *For any vector* $v = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$, *a* corresponding homogeneous vector *is an* $(n+1)$*-dimensional vector of the form* $\begin{pmatrix} rv_1 \\ rv_2 \\ \vdots \\ rv_n \\ r \end{pmatrix}$ *for any non-zero real number* r.

Given a homogeneous vector, one can recover the original vector by taking a scalar multiplication by $1/r$, then projecting out the final coordinate.

Exercise 1.10.2 *What is the behaviour of* $L_A : R^3 \longrightarrow R^3$ *given by*

$$\begin{pmatrix} 1 & 0 & c_1 \\ 0 & 1 & c_2 \\ 0 & 0 & 1 \end{pmatrix}$$

applied to a homogeneous vector? Can you explain this behaviour in terms of the 2*-dimensional vector corresponding to the homogeneous vector?*

Homogeneous vectors are fundamental to our understanding of cameras, in particular how a camera converts a three-dimensional scene into a two-dimensional photograph.

Consider a collection of rays of light emanating from the origin. The origin is the focus of a camera. Placing a plane in this space allows one to capture an image as the rays intersect the plane. The easiest way to do this is to divide by the last element, which is called the *homogeneous depth*.

If (x, y, z) is a point in three dimensions, then $(x/z, y/z)$ is its observed two-dimensional location in the *canonical camera*, which has its focus at the origin, looks along the z-axis, with its image being the plane at $z = 1$. So z is taken as the homogeneous depth because any point on the same ray of light as (x, y, z) projects to the same place. So (hx, hy, hz) will project to $(x/z, y/z)$ too.

2 CALCULUS

2.1 Differentiation

Differentiation is a fundamental concept that you will need in modelling any number of real world activities on a computer Broadly speaking, it is about "rate of change." For instance, if you are given a function stating how far one has travelled over a period of time, the derivative tells you the velocity of your travel.

As we shall see later in the course, notably in the section on Taylor and Maclaurin series, derivatives allow us to approximate many functions by polynomial functions, which are easy for us to manipulate. In later courses, you will see that many functions may be modelled by trigonometric functions, i.e., functions built out of *sin* and *cos*: such approximations are particularly helpful if one consider actions of waves of various kinds.

But in order to do any of that, you need to understand the basic notions associated with differentiation. So we study that in this section. You will need to remember a number of basic derivatives and two rules for evaluating more complicated ones: the product rule and the chain rule.

Definition 2.1.1 *A function* $f : R \longrightarrow R$ *is* continuous at a *if for all* $\epsilon > 0$*, there exists* $\delta > 0$ *such that, if* $|x - a| < \delta$*, it follows that* $|f(x) - f(a)| < \epsilon$*.*

The definition can equivalently be expressed as the assertion that

$$\lim_{x \to a} f(x)$$

exists and equals $f(a)$.

Definition 2.1.2 *A function* $f : R \longrightarrow R$ *is* continuous *if it is continuous at every element* a *of* R*.*

Example 2.1.3 *Examples of continuous functions include all polynomial functions, i.e., all functions of the form*

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0$$

and the usual trigonometric functions sin*,* cos*, and* tan*.*

Exercise 2.1.4 *For an example of a function that is not continuous, consider the function that sends* x *to* $\cos(1/x)$ *if* $x \neq 0$ *and sends* 0 *to* 0*. Try to picture this using MATLAB, and try to write a proof that it is not continuous at* 0*.*

Proposition 2.1.5 *Suppose* $f, g : R \longrightarrow R$ *are continuous at* a*. Then,*

$$\lim_{x \to a} (f(x) + g(x)) = \lim_{x \to a} f(x) + \lim_{x \to a} g(x).$$

$$\lim_{x \to a} (f(x)g(x)) = \lim_{x \to a} f(x). \lim_{x \to a} g(x)$$

$$\lim_{x \to a} \frac{f(x)}{g(x)} = \frac{\lim_{x \to a} f(x)}{\lim_{x \to a} g(x)} \quad \text{if} \quad \lim_{x \to a} g(x) \neq 0.$$

Definition 2.1.6 *A function* $f : R \longrightarrow R$ *is* differentiable at *a if the limit*

$$\lim_{x \to a} \frac{f(x) - f(a)}{x - a}$$

exists. If the limit exists, it is called the derivative *of f at a and is denoted by* $f'(a)$.

Definition 2.1.7 *A function* $f : R \longrightarrow R$ *is* differentiable *if f is differentiable at every element a of R. The function from R to R determined by taking the derivative at each element a of R is denoted by* f' *or by* $\frac{d}{dx}f$ *or sometimes by* $\frac{df}{dx}$.

Example 2.1.8 *Important basic examples of derivatives are as follows:*

1. $\frac{d}{dx}(x^n) = nx^{n-1}$ *for any natural number n*
2. $\frac{d}{dx}(x^r) = rx^{r-1}$ *for any non-zero real number r*
3. $\frac{d}{dx}(sin\, x) = cos\, x$
4. $\frac{x}{dx}(cos\, x) = -sin\, x$
5. $\frac{d}{dx}(ln\, x) = \frac{1}{x}$ *if* $x \neq 0$
6. $\frac{d}{dx}(e^x) = e^x$

There are many interesting functions whose domain is not quite the whole of R. For example, the function $1/x$ is defined on $R - \{0\}$. Moreover, derivatives exist for all elements of its domain. So we need to refine Definitions 2.1.6 and 2.1.7 a little in order to be able to include such possibilities within our analysis. I shall not repeat all the details, but the central point is as follows:

Definition 2.1.9 *Given a subset S of R, a function* $f : S \longrightarrow R$ *is* differenentiable *if f is differentiable at every element a of S.*

More generally again, we are not only interested in functions whose codomain is R. Just as was the case in the Linear Algebra part of the course, a key use of this work is in photography, in which one needs to model activity in space or at least in the plane. So often the codomain of a function we want to integrate is $R \times R$, generally written $R^{"}$, or $R \times R \times R$, typically written R^3. Again, I shall not spell out details here.

Exercise 2.1.10 *Consider a person walking in a circle. Can you see how to describe his or her behaviour in terms of a function from R to* R^2*? Try to write down such a function. Now consider what is the velocity of the person. Can you write that down as a function? Why might an animator be interested in this? Can you recall such activity in movies, either old ones in which it was simply photographed, or new ones involving computer-generated animation?*

What if the person moves faster or slower depending upon where he or she is in the circle? Perhaps more realistically, you might consider the velocity of a person or character while walking on a contoured surface, i.e., up and down hills.

Does MATLAB help you to visualise some of this?

Two general laws are particularly useful in calculating derivatives. The first is as follows:

Theorem 2.1.11 (The product rule) *Given differentiable functions $f, g : R \longrightarrow R$, define $h : R \longrightarrow R$ by $h(x) = f(x)g(x)$. Then*

$$\frac{d}{dx}h = f\frac{d}{dx}g + g\frac{d}{dx}f$$

The key fact about differentiation you will need in later courses is as follows:

Theorem 2.1.12 (The chain rule) *Given differentiable functions $f, g : R \longrightarrow R$ and given an element a of R,*

$$(fg)'(a) = f'(g(a)).g'(a)$$

The theorem is often expressed as the equation

$$\frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx}$$

This situation will arise when you reparametrise cubic curves in differential geometry, as you will do in later courses when you model movement in animation or photography.

Exercise 2.1.13 *Consider a person walking in a circle counter-clockwise. Can you see how to model that by a function from R to $R \times R$?*

Now consider a person standing still but swinging his or her arm. Can you see how to model that by a function from R either into $R \times R$ or, more simply, if you consider photographing the activity from above, into R?

Now consider a person who is walking in a circle and is swinging his or her arm while doing so. Can you see how to combine the above two functions to model that?

If you find that difficult, consider the simpler situation of a person walking in a straight line while swinging his or her arm.

Now think about the velocity of the person's hand, at least when photographed from above. How would you calculate that? Think about whether the chain rule for differentiation might be relevant.

Also see whether MATLAB can help you to model this behaviour.

Exercise 2.1.14 *Differentiate*

1. $ln(cos(\frac{1}{x}))$

2. $e^{sin^2 x}$

3. $sin(cosx)$

on their domains of definedness.

2.2 Partial Differentiation

In considering photography, as has been a theme of the course, we often consider functions of two variables. That has been clear in the Linear Algebra section of the course, where we invariably studied matrices of dimension greater than one. But it applies equally when we turn to integration.

Consider, for instance, a black-and-white photograph, and consider how we might express a measure of the intensity of light at any point in the photograph: it is a function of two variables as the light intensity varies from point to point of the surface to be exposed. You can consider it as a three-dimensional map, with the height of a particular point determined by the intensity of light at that point, putting the height equal to 0 where there is no light and putting it equal to 1 where the surface is exposed to light of maximum intensity.

One can measure the variation in light intensity as a derivative in any direction. This leads us to consider partial differentiation, where one fixes one axis and considers a derivative in the other, and where one calculates an overall derivative from the partial derivatives.

The details are actually very easy, once one has understood derivatives; the main complication in the literature is that the terminology tends to be complicated. Essentially, you just pretend that every variable other than that for which you need to take a derivative is a constant. The details are as follows.

Definition 2.2.1 *Given a function* $f : R^2 \longrightarrow R$, *the* partial derivative *of* $f(x, y)$ *with respect to* x *at* a *is defined to be the limit*

$$\left(\frac{\partial f}{\partial x}\right)_y (a) = lim_{x \to a} \frac{f(x, y) - f(a, y)}{x - a}$$

if the limit exists.

One duly refers to f as *differentiable in* x if the partial derivatives with respect to x exist for every a, and one can dually speak of f being differentiable in y.

There are many different forms of notation for the partial derivative in the literature, but it should be clear from context what is intended. For instance, there seems to be no strong reason why the symbol ∂ often appears rather than the symbol d as we have been using, but ∂ generally seems more common. A convenient compact notation replaces $\left(\frac{\partial f}{\partial x}\right)_y$ simply by f_x, with the subscript here denoting the variable with respect to which f is differentiated. One is supposed to remember that a particular y has been chosen and is implicit in f_x.

To calculate a partial derivative f_x is easy: you pretend each occurrence of y in $f(x, y)$ is a constant, and differentiate $f(x, y)$ as though it was a function with x the only variable. Given $f : R^2 \longrightarrow R$, taking the partial derivative with respect to x duly yields a function from R to R, sending y to the value of f_x determined by y when seen as a constant.

Exercise 2.2.2 *Calculate the two partial derivatives, i.e., the partial derivatives with respect to x and y, of the function*

$$f(x, y) = sin^2(x)cos(y) + \frac{x}{y^2}$$

The chain rule extends to partial differentiation without fuss as follows:

Theorem 2.2.3 (The chain rule for partial differentiation) *Given functions $f : R \longrightarrow R$ and $u : R^2 \longrightarrow R$ for which the various derivatives exist, for every y in R,*

$$\left(\frac{\partial(fu)}{\partial x}\right)_y (x) = \left(\frac{df}{du}\right)(u(x, y)) \cdot \left(\frac{\partial u}{\partial x}\right)_y (x)$$

and dually for f_y.

Exercise 2.2.4 *Given a function $f : R^2 \longrightarrow R$ for which all the various derivatives exist and are continuous, is it true that*

$$f_{xy} = f_{yx}$$

i.e., if one first takes the partial derivative with respect to x, then takes the derivative of the resulting function from R to R with respect to y, does one obtain the same result as doing so in the opposite order? If not, try to find a counter-example; but if it is true, try to construct a proof.

Exercise 2.2.5 *Can you understand a pair of partial derivatives in terms of vectors? Does this relate to photography and to the Linear Algebra part of the course?*

Partial derivatives can be combined to give what is called the *total derivative*. One has a function $u : R^2 \longrightarrow R$ together with functions $x, y : R \longrightarrow R$ and considers the composite $u(x, y) : R^2 \longrightarrow R$. For instance, the function u might tell you the height (or light intensity) of a given point in the plane, while, given a time t, the real numbers $x(t)$ and $y(t)$ might give the position in the plane, or in the frame, of a person after time t. One could calculate the derivative of the composite directly, or one could compose it from the two derivatives by using the total derivative, which is defined as follows.

Definition 2.2.6 *Given functions $f : R^2 \longrightarrow R$ and $x, y : R \longrightarrow R$, the* total derivative *of the composite function $u = f(x, y) : R \longrightarrow R$ is given by the formula*

$$\frac{du}{dt} = \frac{\partial f}{\partial x} \cdot \frac{dx}{dt} + \frac{\partial f}{\partial y} \cdot \frac{dy}{dt}$$

In the situation described above, the total derivative describes the derivative of the variation of height relative to time.

Exercise 2.2.7 *Use the notion of total derivative applied to* $u(x, y) = x^2 + y^2$, $x = sin(t)$, *and* $y = t^2$, *to calculate* $\frac{du}{dt}$.

Exercise 2.2.8 *Suppose one wants to consider double derivatives, e.g., as one uses in calculating acceleration. Can you use the formula in Definition 2.2.6 to give a formula for the double derivative of* u *with respect to* t*? That will become important (not in this course) when one calculate Taylor expansions in several variables, which one uses to approximate complex functions, e.g., in Numerical Methods.*

Exercise 2.2.9

1. *Calculate* f_{xyz}, f_{yzx}, *and* f_{zxy} *where* $f(x, y, z) = e^{xyz}$ *and prove that they are equal.*
2. *Find* $\frac{du}{dt}$ *in two ways given that* $u = x^n y^n$ *and* $x = cos(at)$, *and* $y = sin(bt)$, *where* a, b *and* n *are constants.*
3. *Calculate* $xf_x + yf_y$ *where* $f(x, y) = xy - \frac{1}{x+y}$.

2.3 Integration

Just as you will need to recall or learn differentiation in order to study Vision and Graphics in particular in your later courses, you will also need to recall or learn some integration. Integration gives you a way to measure area or volume. And it acts as a kind of inverse to differentiation.

You will need to remember the basic integrals, such as those for polynomials and trigonometric functions. And you will need to remember the basic methods of calculating more complex integrals, the most complicated and important of which is integration by parts.

The fundamental fact about integration is as follows:

Definition 2.3.1 *If* $f : R \longrightarrow R$ *and* $F : R \longrightarrow R$ *are two functions for which* $\frac{d}{dx}F = f$, *then* F *is called an* indefinite integral *of* f *and is written as*

$$F(x) = \int f(x)dx$$

A function f *is called* integrable *if a function* F *satisfying the condition of the definition exists.*

If f is integrable, there is generally more than one function F satisfying the condition of the definition, because if F satisfies the condition, so does the function sending x to $F(x) + c$ for any real number c.

Almost always, one is interested in *definite* integrals, rather than indefinite integrals, and they are uniquely determined if they exist. We shall not give the formal definition here, as you only need to know the following fact about definite integrals:

Theorem 2.3.2 *Given a function $f : R \longrightarrow R$ for which an indefinite integral F exists, and given real numbers a and b, the definite integral $\int_a^b f(x)dx$ is uniquely determined by the equation*

$$\int_a^b f(x)dx = F(b) - F(a)$$

You will need to extend this to allow for a or b to be infinite. That is achieved by using the following formula and the corresponding one for a if the following limit exists:

$$\int_a^\infty f(x)dx = lim_{b\to\infty}\left(\int_a^b f(x)dx\right)$$

Ignoring that for simplicity, but with obvious generalisations, the following are basic facts you need to know about definite integrals:

Theorem 2.3.3

1. $\int_a^b f(x)dx = -\int_b^a f(x)dx$
2. $\int_a^b f(x)dx + \int_b^c f(x)dx = \int_a^c f(x)dx$
3. $\int_a^b (f(x)+g(x))dx = \int_a^b f(x)dx = \int_a^b g(x)dx$
4. *for any real number r,* $\int_a^b rf(x)dx = r\int_b^a f(x)dx$
5. $\int_0^a f(x)dx = \int_0^a f(a-x)dx$

Exercise 2.3.4 *Try to prove the various parts of the theorem. The last item is quite tricky, so do not worry if you struggle with it.*

The trickiest but most helpful way to integrate a complicated function is by use of integration by parts. This amounts to the inverse of the product rule for differentiation and it works as follows.

Definition 2.3.5 *Given differentiable functions f and g, the equation for* integration by parts *for indefinite integrals is*

$$\int f(x)g'(x)dx = f(x)g(x) - \int f'(x)g(x)dx$$

The equation is often written in the form

$$\int udv = uv - \int vdu$$

Exercise 2.3.6 *Write down an equation for integration by parts for definite integrals.*

Exercise 2.3.7 *Try to prove the equation for integration by parts from what you know about differentiation. Specifically, try to see how it relates to the product rule for differentiation.*

Using integration by parts is quite tricky, because in order to use it, you need to see a given function as a product of two less complicated functions, in a way that makes the integral part of the right-hand side of the equation less complicated than the integral you were originally trying to calculate.

Example 2.3.8 *Consider $\int x\cos(x)dx$. It is not one of the basic integrals we know, so we need to use some technique in order to try to calculate the integral. Trying integration by parts, the most obvious first possibility is to consider putting $u = x$. So $\frac{du}{dx} = 1$.*

We now need to consider the formula $\frac{dv}{dx} = cos(x)$. This yields $v = sin(x)$. That does lead to a simpler right-hand side as it yields

$$\int x\cos(x)dx = x\sin(x) - \int \sin(x)dx = x\sin(x) + \cos(x) + c$$

for any real number c.

In the example, we choose the way to integrate by parts correctly. Sometimes, one needs to apply integration by parts more than once in order to reduce an integral to one of the basic ones. And sometimes one needs to try a few different options before one finds one that works.

For a relatively hard exercise, consider the following.

Exercise 2.3.9 *Try to calculate $\int e^x\cos(x)dx$.*

Exercise 2.3.10 *Calculate the following integrals:*

1. $\int x\sin(x)dx$
2. $\int\left(\frac{x^5}{x^3-1}\right)dx$
3. $\int x^3e^{x^2}dx$
4. $\int_0^1\left(\frac{x^2}{\sqrt{(1-x^2)}}\right)dx$

The last of the integrals above is tricky: you will need to make real use of the fact that it is a definite integral rather than an indefinite one: at the heart of it lies the fact that as θ goes from 0 to $\pi/2$, the function $sin(\theta)$ goes from 0 to 1; if you think hard about that for a while, you should be able to see how to reparametrise the integral in terms of θ at some point in your attempt to evaluate it. If this question encourages you to consult a book, splendid!

2.4 Convergence of Infinite Series

A fundamental technique that is used not only in animation but in a wide variety of computational situations, including computation that supports engineering and physics, involves the approximation of complicated functions by polynomial or trigonometric functions. We shall deal with the former when we consider

Taylor Series; you will see the latter in later courses when you study Fourier Series.

In order to understand the approximations, you first need to come to some understanding of sequences and series in general. Although closely related, sequences and series are different to each other.

Definition 2.4.1 *A* sequence *of real numbers is a function $a : Nat \longrightarrow R$, often written as (a_n) or as $a_1, a_2, \cdots$. A sequence a_n* converges *to a if for all $\epsilon > 0$, there exists a natural number N such that, for all $n > N$, one has $|a_n - a| < \epsilon$. Then a is called the* limit *of the sequence (a_n), and one sometimes writes $a_n \to a$ or sometimes $lim_{n\to\infty} a_n = a$.*

Definition 2.4.2 *The n-th* partial sum *of a sequence (a_n) is the sum*

$$S_n = a_1 + a_2 + \cdots + a_n = \sum_{k=0}^{k=n} a_k$$

The series

$$a_1 + a_2 + \cdots = \sum_{n=1}^{\infty} a_n$$

converges *to S if the sequence $S_1, S_2, \cdots, S_n$ converges to S. Then S is called the* sum *of the series.*

Exercise 2.4.3 *Suppose (a_n) is a sequence for which $a_n \to a$. Do you think that the series $a_1 + a_2 + \cdots$ must converge, i.e., must there exist a real number S satisfying the condition of the definition? If you think so, try to give a proof; and if not, try to give a counter-example.*

Example 2.4.4 *Consider the sequence*

$$1, \frac{1}{2}, \frac{1}{4}, \cdots, \frac{1}{2^n}, \cdots$$

The sequence converges to 0, and the series generated by the sequence converges to 2.

Example 2.4.5 *The geometric series*

$$\sum_{n=0}^{\infty} ak^n = a(1 + k + k^2 + \cdots)$$

has an n-th partial sum given by

$$S_n = a\frac{1 - k^n}{1 - k}$$

So if $|k| < 1$, the series converges to $\frac{a}{1-k}$.

Theorem 2.4.6 *1. If the series $\sum_{n=0}^{\infty} a_n$ converges, it follows that $lim_{n\to\infty} a_n = 0$.*

2. If $\sum_{n=)}^{\infty} a_n = S$, then for all real numbers r, one has $\sum_{n=0}^{\infty} ka_n = k\sum_{n=)}^{\infty} a_n$.

3. $\sum_{n=0}^{\infty}(a_n + b_n) = \sum_{n=0}^{\infty} a_n + \sum_{n=0}^{\infty} b_n$

4. If (a_n) and (b_n) are both sequences of non-negative real numbers for which $a_n \leq b_n$ for all n, and if $\sum_{n=0}^{\infty} b_n$ converges, it follows that $\sum_{n=0}^{\infty} b_n$ converges.

5. $\sum_{n=0}^{\infty} a_n = a_0 + \sum_{n=1}^{\infty} a_n$.

Exercise 2.4.7 *Try to prove parts of the theorem.*

Exercise 2.4.8 (The Harmonic Series) *Consider the series*

$$1 + 1/2 + 1/3 + \cdots 1/n \cdots$$

Does it converge? (Hint: compare it with the function sending X to $1/x$, and recall that the integral of the latter is given by $ln(x)$, which tends to inf *as x tends to* inf.

Theorem 2.4.9 (Alternating series) *If $\sum_{n=0}^{\infty} a_n$ is a series of alternating positive and negative real numbers, and if $|a_{n+1}| < |a_n|$ for all n, with $lim_{n\to\infty} a_n = 0$, the series must converge.*

Exercise 2.4.10 *Compare the Alternating Series Theorem with the Harmonic Series.*

Theorem 2.4.11 *Suppose $\sum_{n=0}^{\infty} a_n$ and $\sum_{n=0}^{\infty} b_n$ are convergent series of positive real numbers such that for all n, one has*

$$\frac{a_{n+1}}{a_n} \leq \frac{b_{n+1}}{b_n}$$

Then if $\sum_{n=0}^{\infty} b_n$ converges, it follows that $\sum_{n=0}^{\infty} a_n$ converges.

Proof Observe that, by finite induction, for each n,

$$a_n \leq \frac{b_n a_1}{b_1}$$

Now use Theorem 2.4.6.

Theorem 2.4.12 (Cauchy's Integral Test) *If $\sum_{n=0}^{\infty} a_n$ is a series of positive decreasing real numbers, and if there exists a positive, monotonic decreasing, integrable function f such that $f(n) = a_n$ for each n, then $\sum_{n=0}^{\infty} a_n$ converges if and only if the integral $\int_0^{\infty} f(x)dx$ exists.*

As you will have noticed by the theorems and exercises we have been considering, there is an important difference between the arbitrary series and series in which all numbers are positive. That motivates the following definition.

Definition 2.4.13 *A series* $\sum_{n=0}^{\infty} a_n$ *is* absolutely convergent *if the series given by* $\sum_{n=0}^{\infty} |a_n|$ *converges.*

If a series is absolutely convergent, it is necessarily convergent, as follows from earlier work. But it is possible for a series to converge without converging absolutely: an example of such is the harmonic series $1, 1/2, 1/3, \cdots$. By the Alternating Series theorem, it converges, but by Cauchy's Integral Test, it does not absolutely converge. So we say a series is *conditionally convergent* if it converges but does not absolutely converge.

As you have probably realised, you need to be careful with conditionally convergent series, as it is easy to make mistakes. In particular, if you make any rearrangement of an absolutely convergent series, it will still converge to the same limit, but that is not true for conditionally convergent series. Similar remarks apply to bracketing.

A particularly important way in which to combine two series is by taking their *Cauchy product.* It is determined by the following theorem.

Theorem 2.4.14 (Cauchy Product) *If* $\sum_{n=0}^{\infty} a_n$ *and* $\sum_{n=0}^{\infty} b_n$ *are absolutely convergent, then the* Cauchy product*, which is defined to be the series* $\sum_{n=0}^{\infty} c_n$ *where*

$$c_n = a_0 b_n + a_1 b_{n-1} + \cdots a_n b_0$$

is also absolutely convergent. Moreover

$$\sum_{n=0}^{\infty} c_n = \left(\sum_{n=0}^{\infty} a_n \right) \left(\sum_{n=0}^{\infty} b_n \right)$$

Perhaps the single most important sort of series you will encounter, with the possible exception of trigonometric series, is the power series

$$\sum_{n=0}^{\infty} a_n x^n = a_0 + a_1 x + a_2 x^2 + \cdots$$

Exercise 2.4.15 (The exponential series) *Check that*

$$1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + \cdots$$

is absolutely convergent for all x.

Exercise 2.4.16 *Check that*

$$x - \frac{x^3}{3} + \frac{x^5}{5} - \cdots + (-1)^n \frac{x^{2n+1}}{2n+1} + \cdots$$

converges for all $|x| < 1$.

Theorem 2.4.17 *Suppose the power series*

$$\sum_{n=0}^{\infty} a_n x^n = a_0 + a_1 x + a_2 x^2 + \cdots$$

converges for all $|x| < R$. Then, both its derivative

$$\frac{dS}{dx} = \sum_{n=0}^{\infty} n a_n x^{n-1} = a_1 + 2a_s x + 3a_3 x^2 + \cdots$$

and its integral

$$\int S dx = \sum_{n=0}^{\infty} \frac{a_n x^{n+1}}{n+1} = a_0 x = \frac{a_1 x^2}{2} + \frac{a_2 x^3}{3} + \cdots$$

both also converge for all $|x| < R$.

Exercise 2.4.18 *Check which of the following series converge, and which diverge:*

1. $\sum_{n=0}^{\infty} \frac{1}{2n(n+1)}$
2. $\sum_{n=0}^{\infty} \frac{1}{(n+\frac{1}{2})^2}$
3. $\sum_{n=0}^{\infty} n^2 x^n$ *where $x > 0$. [For which x's does this converge?]*

2.5 Taylor and Maclaurin Series

The work of previous sections, notably on differentiation and series, allows us to approximate well-behaved functions, specifically those functions for which derivatives exist, as do derivatives of derivatives, etcetera, by polynomial functions. That is mightily useful in regard to computation.

In later courses, you will also see how to approximate well-behaved functions by trigonometric functions. In some circumstances, they are more useful than polynomial functions, but they are beyond the scope of this course. Here, we just study polynomial approximations, the key construct being that of a Taylor series.

The notions of Taylor series and Maclaurin series are essentially the same: the latter focuses on the behaviour of a function at 0, whereas the former allows one to take an arbitrary real number. But one can pass between the two notions without fuss.

Theorem 2.5.1 (Taylor's Theorem) *If f is a continuous real-valued function for which all the derivatives f', f", etcetera, up to $f^{(n)}$ exist in an interval $a \leq x \leq b$ and $f^{(n+1)}$ exists in $a < x < b$, then*

$$f(x) = f(a) + \frac{(x-a)}{1!} f'(a) + \frac{(x-a)^2}{2!} f"(a) + \cdots + \frac{(x-a)^n}{n!} f^{(n)}(a) + E_n(x)$$

where

$$E_n(x) = \frac{(x-a)^{(n+1)}}{(n+1)!} f^{(n+1)}(\xi)$$

for some $a < \xi < x$.

The term E_n is a remainder term, so if we can prove that it tends to 0 as n tends to inf, it follows that $f(x)$ is represented by the power series

$$f(x) = f(a) + \frac{(x-a)}{1!} f'(a) + \frac{(x-a)^2}{2!} f"(a) + \cdots = \sum_{b=0}^{\infty} \frac{(x-a)^n}{n!} f^{(n)}(a)$$

which is called the *Taylor series* for $f(x)$ at a.

The special case in which $a = 0$ is also called the *Maclaurin series.* We shall primarily consider Maclaurin series, for ease of exposition. It is easier (and important) to remember as it amounts to the following:

$$f(x) = f(0) + xf'(a) + \frac{x^2}{2!} f"(a) + \cdots + \frac{x^n}{n!} f^{(n)}(a) + \cdots$$

Example 2.5.2 *Consider the Maclaurin series for* $f(x) = e^{3x}$. *It yields*

$$e^{3x} = 1 + 3x + \frac{(3x)^2}{2!} + \cdots = \sum_{n=0}^{\infty} \frac{(3x)^n}{n!}$$

Exercise 2.5.3 *Use Maclaurin series to check the following:*

1. $sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots$ *for all* x
2. $cos(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots$ *for all* x
3. $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$ *for all* x.

These, and similar examples, can be combined using the various ways in which we combined power series in the previous section.

As you may recall, some important functions, such as $\frac{sin(x)}{x}$ are not defined at $x = 0$. For such functions, one naturally uses Taylor series for a point other than 0.

Exercise 2.5.4 *Calculate the first three non-zero terms of the Maclaurin series of the following:*

1. $\sqrt{}(1 + x)$
2. $cos(sin((x))$
3. $ln\left(\frac{1}{1-x}\right)$

and give a (reasonable) interval of x *on which the series converges.*

2.6 Least Squares Approximation

As we have seen, if enough derivatives exist, we can use Taylor series to approximate a function by a succession of increasingly accurate polynomials, given by allowing the degree of the polynomials to increase. So there is considerable value for us in focusing on polynomials. In particular, perhaps surprisingly, polynomials of degree 1 or 2, i.e., straight-line approximations or quadratic approximations, are remarkably useful, much more than one might imagine.

It is common, in modelling movement, to be given a number of discrete observations that need to be connected in some reasonable way to make a smooth curve: consider for instance the relationship between live action and the 18 frames per second taken by the usual movie camera.

For a first approximation to giving a smooth connection between observations, one tries to connect them as closely as possible by considering a line. Obviously, in general, a line cannot pass exactly through all of the points, but one can give a best approximation. The idea is as follows.

One has n data points $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ and one seeks a straight line $y = b+mx$ that best approximates them. To choose a best such line, a standard choice is the line that minimises the total of the differences in y-coordinate between the y-values of the n points and the y-values of the corresponding points on the line, i.e., we seek to minimise (the square root of)

$$(y_1 - (b + mx_1))^2 + (y_2 - (b + mx_2))^2 + \cdots (y_n - (b + mx_n))^2$$

The line determined by the choice of m and b that minimises this non-negative real number is called the *least squares straight-line approximation* to the data points $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$.

Put

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \qquad A = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \qquad u = \begin{pmatrix} b \\ m \end{pmatrix}$$

Then the problem can be expressed as the follows:

find a vector u such that $|y - Au|$ is a minimum.

Exercise 2.6.1 *Try to see why the above is a rephrasing of the least square problem for a straight line approximation. Use MATLAB to visualise some examples.*

Theorem 2.6.2 *The solution to the least square problem for a straight-line approximation is*

$$u = (A_t A)^{-1} A_t y$$

if $A_t A$ is not invertible, which is the case if not all the data points are collinear.

One can extend from straight-line approximation to quadratic approximation. There, one seeks optimal choices of a, b and c in describing the formula

$$y = a + bx + cx^2$$

In that case, one puts

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \qquad A = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{pmatrix} \qquad u = \begin{pmatrix} a \\ a \\ c \end{pmatrix}$$

and one seeks a vector u in R^3 that minimises $|y - Au|$.

Theorem 2.6.3 *The solution to the least square problem for a quadratic approximation is*

$$u = (A_t A)^{-1} A_t y$$

if $A_t A$ is invertible, which it is if at least three of the x_i's are distinct.

Exercise 2.6.4 *How would one address the least square problem for a cubic approximation? How about arbitrary n? Is there a relationship between the number of data points and the degree of a polynomial one might want to consider? Try using MATLAB to help you to visualise some of the possibilities here.*

Exercise 2.6.5

1. *Find the least squares straight-line approximation for the data points* $(1, 3), (-2, 4), (7, 0)$
2. *Find the least squares straight-line approximation for the data points* $(1, -3), (4, 6), (-2, 5), (3, -1)$
3. *Find the least squares quadratic approximation for the data points* $(-7, 3), (2, 8), (1, 5)$.

2.7 Complex Numbers

In the past few sections, we have gone to considerable effort to approximate complicated situations, such as functions that have all derivatives or sets of data points of arbitrary size and complexity, by polynomials, the latter being particularly well-behaved and well-understood.

In fact, in more sophisticated situations than we have studied in this course, but which you will see in later courses, it is more helpful to approximate complex functions or data by trigonometric functions, i.e., by functions involving *cos*, *sin*, etcetera, the key idea being that of a Fourier series. We shall not study that here, but you do need to know enough about complex numbers by the end of this course to be able to understand Fourier series in later courses.

One usually thinks of a complex number as a point in the plane: the horizontal axis of the plane is called the *real* axis, and the vertical axis is called the *imaginary* axis. There are two possible representations of complex numbers. In the first, a complex number z is represented as $z = (x, y) = x + iy$: the real number x is called the *real* part of z and the real number y is called the *imaginary* part of z. In the second, a real number z is represented as $z = (r, \theta)$, where r is the distance of the point z from the origin, and θ is the angle between z seen as a vector with the positive real axis. There are typically infinitely many choices of θ to represent the same complex number, as (r, θ) and $(r, 2\pi + \theta)$ always represent the same complex number.

The former representation of complex numbers is more convenient for describing addition of complex numbers, while the latter is more convenient for describing multiplication. One passes between the two by noting

$$x = rcos(\theta) \quad y = rsin(\theta)$$

As mentioned above, the real number θ is not uniquely determined by x and y, as $2\pi + \theta$ has the same *cos* and *sin* as θ, so to pass back and forth between the two representations of complex numbers, one usually makes the convention

$$-\pi < \theta \leq \pi$$

Complications will emerge in later courses, for instance when you study Computer Algebra, as functions are sometimes continuous or differentiable everywhere except along the negative x-axis, and that is precisely because of this convention regarding θ.

Given complex numbers $z = x + iy = (r, \theta)$ and $z' = x' + iy' = (r', \theta')$, the arithmetic operations are defined as follows:

- $(x + iy) + (x' + iy') = (x + x') + i(y + y')$
- $(x + iy) - (x' + iy') = (x - x') + i(y - y')$
- $(r, \theta)(r', \theta') = (rr', \theta + \theta')$
- $\frac{(r,\theta)}{(r',\theta')} = (r/r', \theta - \theta')$, defined if $r \neq 0$

If you want to multiply and divide complex numbers directly in terms of their $x+iy$ representation, the key fact to remember is that $i^2 = -1$. If you remember that, and also that

$$(x + iy)(x - iy) = x^2 + y^2$$

it is possible to calculate correct formulae using trigonometry: you do not need to learn the derivations; you just need to know how to calculate sums, etcetera.

By analogy with Maclaurin series, we can *define* constructions e^z and $sin(z)$ and $cos(z)$ for a complex number z as follows:

Definition 2.7.1

- $e^z = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots$
- $sinz = z - \frac{z^3}{3!} + \frac{z^5}{5!} \cdots$

- $cos z = 1 - \frac{z^2}{2!} + \frac{z^4}{4!} \cdots$

Theorem 2.7.2

$$e^{i\theta} = cos(\theta) + i\, sin(\theta)$$

Corollary 2.7.3

$$re^{i\theta} = rcos(\theta) + irsin(\theta)$$

You could regard either the theorem or the corollary as the key fact to remember about complex numbers. You can, of course, deduce formulae for *sin* and *cos* in terms of exponentials:

$$cos(\theta) = \frac{e^{i\theta} + e^{-i\theta}}{2} \qquad sin(\theta) = \frac{e^{i\theta} - e^{-i\theta}}{2i}$$

Exercise 2.7.4 *Use Theorem 2.7.2 to calculate the following:*

1. $e^{i\pi/2}$
2. $e^{i\pi}$
3. $e^{2i\pi}$
4. $e^{n\pi i}$ *for any integer* n
5. $e^{(2n+1)\pi i/2}$ *for any integer* n

Corresponding to our definition of e^z, we also define lnz as follows:

Definition 2.7.5 *If* $z = (r, \theta)$ *in polar coordinates,*

$$ln(z) = ln(r) + i(\theta + 2\pi k)$$

for any integer k*.*

Exercise 2.7.6 *Think about the relationship between the definitions of* e^z *and* $ln(z)$*.*

Theorem 2.7.7 (de Moivre) *For any natural number* n*, one has*

$$z^n = cos(n\theta) + i\, sin(n\theta)$$

Exercise 2.7.8 *Use de Moivre's Theorem to find formulae for* $cos(n\theta)$ *and* $sin(n\theta)$ *in terms of powers of* $cos(\theta)$ *and* $sin(\theta)$*.*

A central part of the interest in complex numbers lies in the fact that we can deduce such results about real numbers from results about complex numbers.

Example 2.7.9 *Let*

$$C = 1 + a\,cos(\theta) + a^2\,cos(2\theta) + \cdots + a^n\,cos(n\theta)$$

and let

$$S = a\,sin(\theta) + a^2\,sin(2\theta) + \cdots + a^n\,sin(n\theta)$$

So

$$C + iS = \frac{1 - (ae^{i\theta})^{n+1}}{1 - ae^{i\theta}}$$

whence one can deduce

$$C = \frac{1 - a\,cos(\theta) + a^{n+2}cos(n\theta) - a^{n+1}cos((n+1)\theta)}{1 - 2a\,cos(\theta) + a^2}$$

One can duly generalise differentiation from real variables to complex variables: the definition of the limit defining a derivative just requires care in that z can approach a particular complex number z_0 from any direction. This fact is fundamental in the later study of complex variables, but we shall not develop it here.

Exercise 2.7.10 *Calculate the real and imaginary parts of the following:*

1. $\frac{2+3i}{3+2i}$
2. $ln\left(\frac{1}{2}(\sqrt{3}+i)\right)$
3. $(1+i)^{iy}$
4. $\frac{1}{i^5}$